PROBLEMAS DE COMPUTO DE LA CORRECCION POR SESGO EN EL CASO LOGNORMAL *

Eusebio Cleto del Rey **

1.- *Introducción*

En los trabajos empíricos se presenta, con cierta frecuencia la necesidad o conveniencia de formular relaciones del tipo:

$$Y = e^{\beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + u} \qquad (1)$$

Donde: Y es la variable dependiente; β_i son k constantes; X_i son k variables independientes (si se incluye a X_1, que toma siempre el valor 1); u es una variable aleatoria con distribución $N(0; \sigma^2)$.

Es ésta una formulación lognormal de Y.

Para estimar (1) se utiliza el modelo lineal general de regresión, luego de haber procedido a linealizar la expresión del siguiente modo:

$$\log Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_k X_k + u \qquad (2)$$

Donde: log simboliza logaritmo natural.

* Este trabajo surgió del Proyecto de Investigación "El Capital Humano Universitario de la Provincia de Salta", del Consejo de Investigación de la Universidad Nacional de Salta (UNSa). Tal Proyecto, que se lleva a cabo en el Departamento de Ciencias Económicas, Jurídicas y Sociales de esa Universidad, recibió el apoyo de la SECYT (hoy SUBCYT) y del CONICET.

El autor agradece los comentarios recibidos en la Reunión de Discusión No. 13, del Area de Economía, del mencionado Departamento de la UNSa, que tuvo lugar el 29 de Abril de 1982; como también los que desde Tucumán le hizo llegar el Dr. Victor Jorge Elías. Se reserva, sin embargo, la responsabilidad por el contenido resultante.

** Profesor Titular de Economía, Departamento de Ciencias Económicas, Jurídicas y Sociales, Universidad Nacional de Salta.

Una vez realizadas n observaciones sobre todas las variables involucradas, y obtenidas las estimaciones de las constantes β_i, a las cuales simbolizaremos $\hat{\beta}_i$, suele procederse, en muchos casos, a estimar valores de Y condicionales en determinados valores de las X_i que le interesan al investigador. Para ello se puede hacer:

$$\overline{Y} = e^{\hat{\beta}_1 + \hat{\beta}_2 X_2^* + \hat{\beta}_3 X_3^* + \ldots + \hat{\beta}_k X_k^*} \tag{3}$$

Donde: X_i^* es el valor que el investigador le asigna a X_i para estimar o predecir el valor de Y que le interesa.

Lo que se pretende con (3) es estimar la media de Y, condicional en los valores X^*_i de las variables independientes, a fin de tomarla como representativa de la distribución de la variable dependiente, dados esos valores de los X_i. Es posible demostrar, sin embargo, que Y no es un estimador insesgado de la media condicional, pero su sesgo puede ser corregido del siguiente modo:

$$Y = \overline{Y} \, g_m \, (t) \tag{4}$$

Donde: $g_m \, (t) = \sum_{P=0}^{\infty} \frac{m^P \, (m+2P)}{m \, (m+2)\ldots(m+2P)} \left(\frac{m}{m+1}\right)^P \frac{1}{P!} t^{P \; 1}$, es la función que corrige el sesgo, la cual depende de $m = n\text{-}k$, o sea de los grados de libertad, y de:

$$t = \frac{m+1}{2m} \left(\hat{\sigma}^2 - \hat{\sigma}^2_{\hat{\eta}}\right) \tag{5}$$

Donde, a su vez: $\hat{\sigma}^2$ es el estimador insesgado de la varianza de los u; $\hat{\sigma}^2_{\hat{\eta}}$ es el estimador insesgado de la varianza de $\hat{\eta} = \hat{\beta}_1 + \hat{\beta}_2 X_2^* + \hat{\beta}_3 X_3^* + \ldots + \hat{\beta}_k X^*_k$, o sea del exponente de la expresión (3).

No es nuestro propósito ir más allá en lo que a teoría de la estimación lognormal se refiere [2], pues nuestro interés se centra en ciertos problemas de cómputo que expondremos en las secciones siguientes. Lo dicho en esta Introducción trata de ubicar al lector en el tema general, del cual nuestro trabajo toca sólo un detalle.

1 BRADU, Dan and MUNDLAK, Yair: "Estimation in Lognormal Linear Models", *Journal of the American Statistical Association,* March, 1970, Vol. 65, No. 329, pág. 199.

2 Al lector interesado en profundizar el tema, en lo teórico, le sugerimos consultar: BRADU, D. and MUNDLAK, Y.: *Op. cit.,* pág. 198/211; JOHNSON, Norman L. and KOTZ, Samuel: *Distributions in Statistics - Continuous Univariate Distributions - 1,* John Wiley & Sons, New York, 1970; y la bibliografía allí mencionada.

2.- Simbología.

Definamos los símbolos a emplear en las secciones siguientes. A tal fin, adoptaremos el orden en que ellos aparecen en los desarrollos siguientes, a partir de la Sec. 3 [3] .

X^* es un vector de orden k por 1, cuyo primer elemento es 1 (uno) y los otros son los valores asignados por el investigador a cada variable X_i (para i = 2, ... , k), para obtener la deseada predicción de Y.

$(X'X)^{-1}$ es la matriz inversa de (X'X).

(X'X) es la matriz cuadrada, de orden k, simétrica, cuyos elementos son los momentos al origen de las variables independientes X_i (para i = 1, 2, ... , k). Así, un elemento de esta matriz está definido como: $\sum_{t=1}^{n} X_{it} X_{jt}$ cuando está ubicado en la posición (i,j).

X es la matriz, de orden n por k, que contiene las observaciones correspondientes a las variables independientes, X_i (para i = = 1, 2, ... k), con la que se computó la regresión. Todos los elementos de su primera columna son iguales a 1 (uno).

M es una matriz cuadrada, de orden (k-1), simétrica, cuyos elementos son los momentos respecto a las medias muestrales de las variables independientes X_i (para i = 2, ... , k). Así, un elemento de esta matriz está definido como:

$$m_{ij} = \sum_{t=1}^{n} (X_{it} - \bar{X}_i)(X_{jt} - \bar{X}_j).$$

(RR) es una matriz cuadrada, de orden (k - 1), simétrica, cuyos elementos son los coeficientes de correlación simple entre pares de las variables independientes X_i (para i = = 2, ... , k). Así, un elemento de esta matriz está definido como

3 Es aconsejable pasar directamente a leer la Secc. 3, y utilizar a la Secc. 2 a manera de un diccionario, para aclarar el significado de cada símbolo a medida que ellos aparezcan.

$r_{ij} = \dfrac{m_{ij}}{\sqrt{m_{ii}} \; \sqrt{m_{jj}}}$. Nótese que los elementos diagonales resultan todos iguales a 1 (uno).

M^{-1} es la matriz inversa de M.

$(RR)^{-1}$ es la matriz inversa de (RR).

S es una matriz diagonal, de orden (k-1), cuyos elementos diagonales están definidos como: $\dfrac{1}{\sqrt{m_{ii}}}$ (para i = 2, ... , k).

S^{-1} es la matriz inversa de S. Nótese que se trata de una matriz diagonal, cuyos elementos diagonales son: $\sqrt{m_{ii}}$ (para i = = 2, ... , k).

A_{11} es un escalar.

A_{21} es un vector de orden (k-1) por 1.

A_{12} es un vector de orden 1 por (k-1), y, por simetría de la matriz $(X'X)^{-1}$, es: $A_{12} = A'_{21}$.

A_{22} es una matriz cuadrada, de orden (k-1), simétrica (por simetría de $(X'X)^{-1}$).

B_{11} es un escalar, y es $B_{11} = n$, por ser el momento primero respecto al origen de la variable X_1, o sea la primera columna de la matriz X.

B_{21} es un vector de orden (k-1) por 1, cuyos elementos están definidos como $\sum_{t=1}^{n} X_{it}$, por ser los momentos cruzados respecto al origen de X_1 y cada una de las variables X_i (para i =2, ..., k).

B_{12} es un vector de orden 1 por (k-1), y, por simetría de la matriz (X'X), es: $B_{12} = B'_{21}$.

B_{22} es una matriz cuadrada, de orden (k-1), simétrica (por simetría de (X' X)), cuyos elementos pueden ser definidos como $\sum_{t=i}^{n} X_{it} X_{jt}$ (ver descripción de la matriz (X'X)).

I_k es una matriz unitaria de orden k.

I_{11} es un escalar, y es $I_{11} = 1$.

I_{12} es un vector de orden 1 por (k-1), cuyos elementos son todos iguales a 0 (cero).

I_{21} es un vector de orden (k-1) por 1, cuyos elementos son todos iguales a 0 (cero).

I_{22} es una matriz unitaria de orden (k-1), a la que podemos simbolizar $I_{22} = I_{(k-1)}$.

x^* es un vector de orden (k-1) por 1, cuyos elementos están definidos como la diferencia entre los valores asignados por el investigador a cada variable X_i (i= 2, ..., k), para obtener la deseada predicción de Y, y la media muestral de los datos utilizados en el cómputo de la regresión, correspondiente a cada una de esas variables X_i.

X_1^* es un vector de orden (k-1) por 1, cuyos elementos son los valores asignados por el investigador a cada variable X_i (i=2, ...,k), para obtener la deseada predicción de Y.

$\bar{X}$ es un vector de orden (k-1), cuyos elementos son las medias muestrales de los datos utilizados en el cómputo de la regresión, correspondientes a cada una de las variables X_i (para i = = 2, ..., k).

3.- Planteamiento del Problema.

Según las ecuaciones (4) y (5), la función que corrige el sesgo en la predicción lognormal por regresión lineal depende de m, $\hat{\sigma}^2$ y $\hat{\sigma}^2_{\hat{\eta}}$. Cualquier programa de cómputo de regresión nos permite conocer los grados de libertad y la varianza residual, pero no ocurre lo mismo con la estimación de la varianza de la predicción del logaritmo de Y. Esta última es:

$$\hat{\sigma}^2_{\hat{\eta}} = X^{*\prime} (X'X)^{-1} X^* \hat{\sigma}^2 \qquad (6)^4$$

Conocemos tres métodos de cómputo, para estimar una relación lineal entre variables por regresión múltiple: i) La operación principal es invertir la matriz $(X'X)$[5]; ii) La matriz que se invierte es

4 Ver: BRADU, D. and MUNDLAK, Y.: *Op.cit.*, pág. 203.

5 Véase, por ejemplo: JOHNSTON, J.: *Econometric Methods*, McGraw-Hill, Cap. 4 en la Primera Edición, 1960 y Cap. 5 en la Segunda Edición. Hay traducción al Castellano.

M^5 ; iii) La matriz que se invierte es $(RR)^6$.

En el caso i) no existe ningún problema, pues el programa de cómputo nos provee directamente la matriz $(X'X)^{-1}$, que es la que aparece en la fórmula (6).

El problema se plantea en los casos ii) y iii), en los que el programa computa M^{-1} y $(RR)^{-1}$, y no la matriz que necesitamos. Nos preocupa, pues, obtener el valor de la expresión $X^{*\prime} (X'X)^{-1} X^*$ (que es un escalar), cuando no conocemos $(X'X)^{-1}$, sino M^{-1} o $(RR)^{-1}$.

4.- Caso en el que se usa la Matriz de Correlación.

Supongamos que nuestro programa de regresión múltiple invierte la matriz de correlación (RR).

Por simple inspección del producto de matrices se puede probar que:

$$(RR) = SMS \tag{7}$$

De (7) surge directamente que:

$$(RR)^{-1} = S^{-1} M^{-1} S^{-1} \tag{8}$$

y que:

$$M^{-1} = S\,(RR)^{-1}\,S \tag{9}$$

Puesto que m_{ii} (para $i = 2, \ldots, k$) debe ser calculado como paso previo a la obtención de (RR), nuestro programa nos puede proveer tales datos. Con ellos formamos la matriz S, requerida por la expresión (9).

El caso iii) de la Sección 3 queda así reducido al caso ii), de la misma sección, al que consideraremos a continuación.

5.- Caso en el que se usa la Matriz de Momentos Respecto a las Medias Muestrales.

Pasamos a considerar el caso en el que el programa de cómputos invierte la matriz M, o aquel en el que hemos llegado a M^{-1} a partir de $(RR)^{-1}$, según vimos en la sección anterior. Estamos, por lo tanto, en los casos ii) y iii) de la Sección 3.

5 Véase, por ejemplo: JOHNSTON, J.: *Econometric Methods,* McGraw-Hill, Cap. 4 en la Primera Edición, 1960 y Cap. 5 en la Segunda Edición. Hay traducción al Castellano .

6 Véase, por ejemplo: MADRON, Thomas W.: "Multiple Regression for the TRS-80", *Byte*, October, 1981, pág. 430/47.

5.1. Demostración de que M^{-1} es una Sub-Matriz de $(X'X)^{-1}$.

Procedamos a particionar la matriz $(X'X)^{-1}$, del siguiente modo:

$$(X'X)^{-1} = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \tag{10}$$

A la matriz (X'X) la particionamos de modo que resulte conformable con la partición anterior:

$$(X'X) = \begin{vmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{vmatrix} \tag{11}$$

Por la definición de matriz inversa:

$$(X'X)^{-1} \ (X'X) = I_k \tag{12}$$

Podemos particionar:

$$I_K = \begin{vmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{vmatrix} \tag{13}$$

Utilizando las particiones de (10), (11) y (13) y el producto de matrices de (12) obtenemos:

$$A_{21} B_{11} + A_{22} B_{21} = I_{21} \tag{14}$$

$$A_{21} B_{12} + A_{22} B_{22} = I_{22} \tag{15}$$

Reemplazando algunas submatrices por sus valores, según sus definiciones, en (14) y (15):

$$nA_{21} + A_{22} B_{21} = 0 \tag{16}$$

$$A_{21} B_{12} + A_{22} B_{22} = I_{(k-1)} \tag{17}$$

De (16) obtenemos:

$$A_{21} = \frac{1}{n} A_{22} B_{21} \tag{18}$$

Premultiplicamos (17) por A^{-1}_{22}:

$A^{-1}_{22}\, A_{21} B_{12} + B_{22} = A^{-1}_{22}$ o, lo que es lo mismo:

$$A^{-1}_{22} = A^{-1}_{22}\, A_{21}\, B_{12} + B_{22} \tag{19}$$

Reemplazando (18) en (19) y operando:

$$A^{-1}_{22} = A^{-1}_{22} \left(-\frac{1}{n}\, A_{22}\, B_{21}\right) B_{12} + B_{22}$$

$$A^{-1}_{22} = B_{22} - \frac{1}{n}\, B_{21}\, B_{12} \tag{20}$$

Obsérvese que cada elemento de B_{22} es: $\sum_{t=1}^{n} X_{it}\, X_{jt}$ (para i, j = 2, . . . , k). Nótese, además que cada elemento de la matriz (de orden (k-1), cuadrada) $B_{21}\;\; B_{12}$ es igual a $\left(\sum_{t=1}^{n} X_{it} \sum_{t=1}^{n} X_{jt}\right)$ (para i, j = 2, . . ., k). Por lo tanto cada elemento de A^{-1}_{22} resulta:

$\sum_{t=1}^{n} X_{it}\; X_{jt} - \frac{1}{n} \left(\sum_{t=1}^{n} X_{it} \sum_{t=1}^{n} X_{jt}\right) = \sum_{t=1}^{n} (X_{it} - \overline{X}_i)(X_{jt} - \overline{X}_j) =$ $= m_{ij}$ (para i, j = 2, . . .,k). O sea que la matriz A^{-1}_{22} tiene sus elementos idénticos a los de M.

Por lo tanto

$$A^{-1}_{22} = M \tag{21}$$

Y, por ser única la matriz inversa:

$$A_{22} = M^{-1} \tag{22}$$

De (10) y (22) surge que M^{-1} es una submatriz de $(X'X)^{-1}$, que se obtiene eliminando la primera fila y la primera columna de esta última matriz.

5.2. *Estimación de la Varianza de la Predicción del Logaritmo de Y, cuando se utiliza la Inversa de M.*

Consideremos el vector :

$$\begin{vmatrix} 0 \\ x^* \end{vmatrix} = \begin{vmatrix} 1 \\ X^*_1 \end{vmatrix} - \begin{vmatrix} 1 \\ \overline{X} \end{vmatrix} \tag{23}$$

Particionado de un modo conformable con las matrices de (10) (11) y (13).

Resulta evidente que:

$$X^* = \begin{vmatrix} 1 \\ X_1^* \end{vmatrix} = \begin{vmatrix} 0 \\ x^* \end{vmatrix} + \begin{vmatrix} 1 \\ \bar{X} \end{vmatrix} \tag{24}$$

Utilizando (10) y (24) se obtiene:

$$X^{*'} (X'X)^{-1} X^* = (0\ x^{*'}) \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} 0 \\ x^* \end{vmatrix} + (0\ x^*) \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} 1 \\ \bar{X} \end{vmatrix} +$$

$$(1\ \bar{X}') \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} 0 \\ x^* \end{vmatrix} + (1\ \bar{X}') \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix} \begin{vmatrix} 1 \\ \bar{X} \end{vmatrix} \tag{25}$$

Realizando los productos de matrices particionadas planteados en (25), y operando, llegamos a:

$$X^{*'} (X'X)^{-1} X^* = x^* A_{22} x^* + \bar{X}' A_{22} \bar{X} + A_{11} + 2(x^* A_{21} + x^* A_{22} \bar{X} + A_{12} \bar{X}) \tag{26}$$

Para obtener los dobles productos de (26) se tuvo en cuenta que todos los sumandos son escalares, y que un escalar es siempre idéntico a su traspuesto.

Debido a que:

$$\bar{X} = \frac{1}{n} B_{21} \tag{27}$$

Resulta, aplicando (27) a (26):

$$X^{*'} (X'X)^{-1} X^* = x^{*'} A_{22} x^* + \left(\frac{1}{n}\right)^2 B_{12} A_{22} B_{21} + A_{11} + 2(x^{*'} A_{21} + \frac{1}{n} x^{*'} A_{22} B_{21} + \frac{1}{n} A_{12} B_{21}) \tag{28}$$

Pero, utilizando las particiones de (10), (11) y (13) y el producto de matrices de (12), obtenemos:

$$A_{11} B_{11} + A_{12} B_{21} = I_{11} \tag{29}$$

Reemplazando y operando en (29):

$$A_{12} B_{21} = 1 - n A_{11} \tag{30}$$

Además, de la ecuación (16) obtenemos, por pasaje de términos:

$$A_{22}\, B_{21} = -\, nA_{21} \tag{31}$$

Utilizando (28), (30) y (31) obtenemos, luego de operar:

$$X^{*'}\,(X'X)^{-1}\, X^{*} = x^{*'}\, A_{22}\, x^{*'} + \frac{1}{n} \tag{32}$$

Y según (22), será:

$$X^{*'}\,(X'X)^{-1}\, X^{*} = x^{*'}\, M^{-1} x^{*} + \frac{1}{n} \tag{33}$$

En conclusión, cuando nuestro programa de cómputo de regresión lineal múltiple trabaja con la matriz M, la fórmula (6) resulta:

$$\hat{\sigma}^2_{\hat{\eta}} = \left(x^{*'}\, M^{-1}\, x^{*} + \frac{1}{n}\right) \hat{\sigma}^2 \tag{34}$$

Que también es aplicable al caso en que se trabaja con la matriz (RR), una vez que su inversa, $(RR)^{-1}$, fue transformada en M^{-1} según lo establece la expresión (9).

6.- Aplicación a un Problema Práctico.

Presentamos a continuación, como ejemplo, una regresión múltiple que estima la siguiente relación:

$$Y = e^{\beta_1 + \beta_2 t + \beta_3 t^2 + u} \tag{35}$$

Donde: Y es la producción de soja en la República Argentina[7], $X_2 = t$ es el tiempo en años (tendencia) con origen en el año agrícola 1970/71; $X_3 = t^2$, y el resto de los símbolos tiene el mismo significado que en la ecuación (1).

7 Ver datos y fuente en el Apéndice de este trabajo.

Tal estimación tiene por fin predecir datos faltantes de la producción de soja, como ser el correspondiente a 1981/82, que es desconocido al momento de escribir estas líneas. Es posible que la especificación de la ecuación (35) no sea la más adecuada para ese fin, pues no hemos realizado la necesaria exploración de las alternativas existentes, en razón de que nuestro único interés es dar un ejemplo de la aplicación de lo que obtuvimos en forma algebraica.

Una vez linealizada la ecuación (35), según lo propusimos en la ecuación (2), aplicamos el modelo lineal general para obtener la siguiente regresión:

Variable	$\hat{\beta}$	Desvío Estándar de $\hat{\beta}$	t de Student Hipótesis: β=
X_1 (Coordenada al origen)	3,86607	0,33572	11,51586
$X_2 = t$	0,76979	0,14021	5,49025
$X_3 = t^2$	- 0,03226	0,01242	-2,59738

$R^2 = 0{,}96277$ $\qquad$ $\hat{\sigma}^2 = 0{,}081474114$

$n = 10$ $\qquad$ g.1. $= m = 7$

Nota: Salvo la estimación de la varianza residual, que es presentada con todos los decimales obtenidos, las cifras anteriores fueron redondeadas a los cienmilésimos. Los cálculos se realizaron con mayor número de decimales.

Presentamos a continuación la matriz (RR) y su inversa:

$$RR = \begin{vmatrix} 1 & 0{,}974558629 \\ 0{,}974558629 & 1 \end{vmatrix}$$

$$(RR)^{-1} = \begin{vmatrix} 19{,}90624992 & -19{,}39980763 \\ -19{,}39980763 & 19{,}90624992 \end{vmatrix}$$

De acuerdo con su definición, la matriz S es:

$$S = \begin{vmatrix} 0{,}110096376 & 0 \\ 0 & 0{,}0097541248 \end{vmatrix}$$

Los elementos diagonales de esta matriz pueden ser obtenidos a partir de los datos consignados en la matriz M, más adelante.

Calculemos, entonces:

$$S(RR)^{-1} S = \begin{vmatrix} 0{,}241287875 & -0{,}020833332 \\ 0{,}020833332 & 0{,}0018939393 \end{vmatrix}$$

Por otra parte, la matriz M y su inversa son:

$$M = \begin{vmatrix} 82{,}5 & 907{,}5 \\ 907{,}5 & 10510{,}5 \end{vmatrix}$$

$$M^{-1} = \begin{vmatrix} 0{,}241287878 & -0{,}020833333 \\ -0{,}020833333 & 0{,}0018939393 \end{vmatrix}$$

Por inspección podemos comprobar que la ecuación (9) se cumple con alto grado de precisión, o sea que $M^{-1} = S(RR)^{-1} S$.

Presentamos a continuación las matrices X'X y $(X'X)^{-1}$:

$$X'X = \begin{vmatrix} 10 & 55 & 385 \\ 55 & 385 & 3025 \\ 385 & 3025 & 25333 \end{vmatrix}$$

$$(X'X)^{-1} = \begin{vmatrix} 1{,}383333333 & -0{,}525 & 0{,}041666666 \\ -0{,}525 & 0{,}241287878 & -0{,}020833333 \\ 0{,}041666666 & -0{,}020833333 & 0{,}0018939393 \end{vmatrix}$$

Resulta fácil comprobar a simple vista que M^{-1} es la submatriz que se obtiene eliminando la primera fila y la primera columna de la matriz $(X'X)^{-1}$. Ello concuerda con lo que demostramos en la Secc. 5.1.

El siguiente paso es aplicar nuestros resultados para predecir la producción de soja en determinado año.

Antes debemos observar que las medias de los datos utilizados en la regresión, que corresponden a las variables independientes, son:

$$\overline{X} = \begin{vmatrix} 5{,}5 \\ 38{,}5 \end{vmatrix}$$

Trabajemos con el año agrícola 1981/82. Los vectores necesarios para calcular $\hat{\sigma}^2_{\hat{\eta}}$ son:

$$X^* = \begin{vmatrix} 1 \\ 11 \\ 121 \end{vmatrix} \qquad x^* = \begin{vmatrix} 5{,}5 \\ 82{,}5 \end{vmatrix}$$

Con ellos, y con las matrices anteriormente presentadas, obtenemos:

$$X^{*\prime}\ (X'X)^{-1}\ X^* = 1{,}383334246$$

$$x^{*\prime}\ M^{-1}\ x^* = 1{,}283333656$$

y recordando que $n = 10$, tendremos:

$$x^{*\prime}\ M^{-1}\ x^* + \frac{1}{n} = 1{,}383333656$$

que difiere de $X^{*\prime}\ (X'X)^{-1}\ X^*$, por razones de redondeo, en una cantidad de orden de 4,3 por diez millones. Este resultado concuerda con lo obtenido en la ecuación (33).

Para completar nuestro ejemplo, procedamos a predecir la producción de soja para 1981/82. A tal fin hacemos:

$$\hat{\eta} = X^{*\prime}\hat{\beta} = 8{,}42967727$$

$$\bar{Y} = e^{\hat{\eta}} = 4581$$

Esta es una estimación de la mediana condicional correspondiente al año 1981/82. Para obtener una estimación insesgada de la media condicional debemos corregir ese valor por sesgo logarítmico. Para ello calculamos, de acuerdo a la ecuación (6):

$$\hat{\sigma}^2_{\hat{\eta}} = 0{,}1127059$$

y, de acuerdo a la ecuación (5) calculamos t [8]:

$$t = -0{,}0178467$$

Obtenemos luego el valor de g_m (t) [9]:

$$g_7\,(-0{,}0178467) = 0{,}9843483$$

y multiplicando este valor por $\bar{Y}$, como lo señala la ecuación (4) obtenemos:

$$\hat{Y} = 4509$$

8 Para respetar la simbología más usual en cada caso, nos vimos obligados a emplear el símbolo t, en este ejemplo, con tres significados distintos. Por un lado es la variable tiempo, o tendencia, en la regresión. Por otro es t de Student. Y por último, en este lugar, es el número definido en ecuación (5), y que sirve para obtener g_m (t). Un poco de atención por parte del lector evitará confusiones.

9 El valor de $g_m(t)$ fue obtenido utilizando las tablas que presentan BRADU, D. and MUNDLAK, Y.: *Op. cit.*, pág. 200. Allí encontramos Table 1. g(t), que se emplea cuando t > 0, y Table 2. g(-t) que se usa con t negativo, como en el caso que nos ocupa, pero a la tabla se entra con el valor absoluto de t. La otra entrada corresponde a m. Debido a lo restringido de los valores de m que figuran en la tabla, y a que t avanza por décimos, es necesario interpolar frecuentemente, a fin de lograr mejores aproximaciones. Hemos utilizado interpolación logarítmica. Es nuestro propósito preparar un programa de cómputos que permita obtener $g_m(t)$ con gran aproximación en cada caso particular.

que es nuestra estimación insesgada de la producción de soja para la campaña 1981/82, medida en miles de toneladas.

7.- Consideraciones Finales.

Cualquiera de los tres métodos de cómputo de regresión lineal múltiple que hemos descripto, permite obtener los elementos necesarios para realizar la corrección por sesgo lognormal, cuando ello fuera requerido.

En los casos i) y ii) planteados en la Secc. 3, lo obtenido con álgebra matricial es perfectamente aplicable a los cómputos, y los errores de redondeo que resulten en $\hat{\sigma}^2_{\hat{\eta}}$ estarán, en forma casi exclusiva, relacionados con aquellos en los que se hubiera incurrido al invertir (X'X) o M.

En una versión anterior de este trabajo advertíamos al lector lo siguiente:

En el caso iii) el problema de los errores de redondeo se ve amplificado por el método aquí propuesto. Ya el programa de regresión comete algunos que no estarían presentes al invertir M, cuando primero háce (RR) = SMS y recién procede a obtener la inversa de esta última matriz, pues en cada una de esas operaciones redondea o (peor aún) trunca los resultados. Cuando luego computamos, de acuerdo a la ecuación (9), $M^{-1} = S(RR)^{-1}S$, amplificamos esos errores e introducimos otros nuevos. La acumulación puede conducir a que el resultado final sea, simplemente, "basura".

El ejemplo que aquí presentamos contradice al párrafo anterior, pues en él resulta $S(RR)^{-1}S = M^{-1}$ con alto grado de precisión, como puede ser observado en Secc. 6.

Debe tenerse en cuenta, sin embargo, que hemos trabajado en un caso de orden 2, que las inversiones se realizaron calculando primero los elementos de las matrices adjuntas y dividiéndolos luego por los correspondientes determinantes, que en todos los casos trabajamos con ocho o más decimales, etc. Muy distinto puede ser el resultado cuando la computadora trabaja con matrices de mayor orden, usa un método menos directo de inversión, su mantisa es de seis dígitos, etc.

Por ello, a pesar de nuestro contraejemplo, nos parece conveniente insistir en lo que aconsejamos a nuestra primitiva versión, con referencia a iii): Comprobar, en cada caso particular, el grado de exactitud de los cómputos, realizando por separado los productos $S^{-1}(RR)S^{-1}$ y $S(RR)^{-1}S$, y luego multiplicarlos entre sí, para observar cuan cercano a una matriz unitaria es el resultado. Si las diferencias fueran notables, más vale emplear otro método.

APENDICE

PRODUCCION DE SOJA EN LA REPUBLICA ARGENTINA

(en miles de toneladas)

Años	Producción
1971/72	78
1972/73	272
1973/74	496
1974/75	485
1975/76	695
1976/77	1.400
1977/78	2.500
1978/79	3.700
1979/80	3.500
1980/81	3.770

Fuente:

F.I.E.L. : *Indicadores de Coyuntura,* No. 194, Mayo de 1982, Cuadro 4.2.7., "Area Sembrada y Producción de Oleaginosos, Cultivos Industriales y Hortalizas", pág. 62.;

BIBLIOGRAFIA

1 - BRADU, Dan and MUNDLAK, Yair: "Estimation in Lognormal Linear Models", *Journal of the American Statistical Association*, March 1970, Vol. 65 No. 329.

2 - FUNDACION DE INVESTIGACIONES ECONOMICAS LATINOAMERICANAS (FIEL): *Indicadores de Coyuntura*, Mayo de 1982, No. 194.

3 - JOHNSON, Norman and KOTZ, Samuel: *Distributions in Statistics - Continuous Univariate Distributions - 1*, John Wiley & Sons, New York, 1970.

4 - JOHNSTON, J.: *Econometric Methods*, 2nd Edition, MacGraw - Hill Kogakusha, Ltd., Tokyo, 1972.

5 - MADRON, Thomas W.: "Multiple Regression for the TRS - 80", *Byte*, October 1981, Vol. 6, No. 10.

PROBLEMAS DE COMPUTO DE LA CORRECCION POR SESGO EN EL CASO LOGNORMAL

RESUMEN

Cuando tomamos logaritmos de la variable dependiente, a fin de linealizar su relación con las independientes y aplicar en la estimación el modelo lineal general de regresión, y suponemos que el término de error tiene distribución normal, estamos en el caso lognormal.

Las predicciones de valores de la variable dependiente, empleando tal estimación, resultan sesgadas, y es necesario proceder a corregirlas. Este trabajo se ocupa de los problemas de esa corrección, que se presentan cuando el programa de cómputos de la regresión invierte la matriz de momentos respecto a las medias muestrales, o la de correlación, y no la de momentos respecto al origen.

COMPUTATIONAL PROBLEMS OF THE BIAS CORRECTION IN THE LOGNORMAL CASE

SUMMARY

When a logarithmic transformation of the dependent variable is made in order to obtain a linear relation with the independent ones, to apply the general linear model the lognormal case is reached.

The predictions of the dependent variable values using such estimation are biased; so, it is necessary to correct them. This paper deals with the problems of such a correction, that appear when the computation program finds the inverse of the moments respect to the sample means matrix, or the correlation matrix instead of the matrix of the moments respect to the origin.