Cálculo de los Estimadores de Borde

Por: Luis Pérez G. *
Germán Cabarcas I. *

RESUMEN

En este artículo se presentan algunos de los más importantes procedimientos utilizados para calcular los estimadores de borde. Un ejemplo muy ilustrativo de la Economía colombiana se emplea para explicar el método de la Traza de Borde. Se introduce el tema de la relación entre número condicional y estimadores de borde.

INTRODUCCION

Desde 1970, fecha en la cual Hoerl y Kennard[1] publicaron por primera vez sobre Regresión de Borde, son muchos los artículos que han aparecido sobre el tema. Sin embargo, su uso para enfrentar problemas prácticos afectados de multicolinealidad, es aún muy poco extendido. Ello puede deberse, en parte, al desconocimiento de los procedimientos computacionales existentes para el cálculo de los estimadores de borde.

El problema básico es el de estimar los coeficientes del modelo de regresión lineal estandarizado:

$$\underline{Y} = X\,\underline{b} + \underline{e} \qquad (1)$$

donde Y es un vector n x 1 de observaciones; X es una matriz n x p de rango p; b es un vector p x 1 compuesto por parámetros desconocidos $b_1, b_2, \ldots, b_p$; e es un vector de variables aleatorias independientes y normales con media cero y matriz de covarianzas $w^2 I_n$.

El estimador de borde introducido por Hoerl y Kennard[2] (1970) es del tipo:

$$\hat{\underline{b}}^* = (X'X + k\,I_p)^{-1}\,X'\,\underline{Y} \qquad (2)$$

* *Postgrado en Matemáticas, Universidad Nacional de Colombia, Seccional Medellín.*

donde k es un escalar positivo. Cuando k = 0 se tiene el estimador ordinario de mínimos cuadrados; si k > 0 el estimador obtenido es una contracción del estimador de mínimos cuadrados b en el siguiente sentido:

$$\hat{\underline{b}}^* = \left(I + k\,(X'X)^{-1}\right)^{-1}\,\underline{b},$$

con lo cual se comprueba que $\underline{b}^*$ es un estimador sesgado.

Por otra parte se demuestra que existe un k > 0 para el cual el error medio cuadrático MSE (k) es mínimo, ya que éste se puede expresar como la suma de una varianza decreciente y un sesgo creciente. Para detalles adicionales acerca del estimador de borde se recomienda Pérez y Cabarcas (1984)[7].

Desde el punto de vista práctico ese k sólo puede ser encontrado de una manera aproximada basándose en el hecho de que alrededor de su valor verdadero el sistema tiende a estabilizarse.

Son muchos los procedimientos propuestos para resolver el problema; a continuación se presentan algunos de los más importantes.

1. TRAZA DE BORDE

La Traza de Borde se define como el conjunto de variaciones de los coeficientes de borde $b_i^*(k)$ (i = 1, 2, ... p), cuando k toma valores positivos desde cero. Se puede obtener entonces una buena representación gráfica en dos dimensiones de la Traza de Borde si se toma un adecuado conjunto de valores de k en el intervalo (0,1) y se realiza una regresión por cada uno usando la ecuación (2). Al graficar en un mismo sistema de coordenadas resultan p curvas, una por cada $\hat{b}_i^*$ (i = 1,. . ., p) en función de k. Dicho diagrama se convierte en una verdadera radiografía del modelo de regresión: En presencia del fenómeno de multicolinealidad los coeficientes afectados tienen un comportamiento inestable para valores muy cercanos a cero, pero a medida que se alejan de cero, muestran una tendencia a la estabilización. Un valor de k a partir del cual se considere que existe estabilidad en todos los coeficientes puede ser escogido como el más adecuado.

La siguiente investigación práctica ilustra claramente este método. Se trata de explicar el comportamiento de las importaciones de Colombia en función de su consumo privado, del consumo de gobierno, del producto interno bruto y de las exportaciones en base a datos históricos comprendidos entre 1950 y 1980. Se tienen 31 observaciones de cada una de las variables anteriormente enumeradas, las cuales se presentan en la Tabla 1. Estos datos provienen de las Cuentas Nacionales y están convertidos todos en millones de pesos a precios corrientes del mercado.

TABLA 1. Datos para el modelo de las Importaciones*

AÑO	CPRIV	CGOB	P I B	EXPOR	IMPOR
1950	6044,7	433,2	7860,5	853,3	795,4
1951	6956,4	520,8	8940,9	1242,3	1138,9
1952	7482,0	581,8	9650,9	1286,5	1192,0
1953	8259,0	724,1	10734,7	1677,2	1565,7
1954	9711,8	856,0	12758,8	1907,5	1858,5
1955	10185,1	937,9	13249,8	1643,2	1897,5
1956	11232,3	959,4	14867,8	1846,0	1879,5
1957	12987,9	1021,6	17810,6	2702,6	2434,5
1958	15004,9	1196,1	20682,5	3889,9	3271,0
1959	17198,2	1369,5	23648,8	4069,8	3384,3
1960	19589,3	1659,3	26746,7	4163,9	4160,6
1961	22584,5	2016,0	30421,0	3920,2	4434,7
1962	25699,7	2356,0	34119,2	4146,6	4407,8
1963	33024,8	3149,0	43525,5	5173,5	5666,3
1964	41467,6	3483,6	53760,3	6376,5	7169,4
1965	45482,1	3954,3	60797,6	6043,5	6324,5
1966	55842,6	4910,4	73612,3	8916,5	11097,6
1967	61596,0	5716,8	83082,7	9950,3	9521,4
1968	70595,6	6579,8	96421,7	12519,6	13779,5
1969	81677,4	7832,8	110953,3	14675,1	15947,2
1970	93863,3	9961,6	130361,4	18515,8	20639,6
1971	110030,8	13428,6	152262,8	19151,2	24933,0
1972	134842,7	14649,2	186092,3	25217,1	26362,0
1973	180200,6	19014,3	243235,9	36186,4	32930,0
1974	237437,6	23158,3	329155,4	46794,5	52514,5
1975	306614,2	30424,9	412828,7	62242,6	60052,9
1976	373651,8	38730,4	534015,3	88047,8	77768,1
1977	474943,1	48153,9	718474,5	124727,2	98489,3
1978	615723,5	65724,3	916559,7	155180,8	129011,3
1979	805145,9	91120,0	1195379,5	193132,8	163211,3
1980	1056187,0	123692,1	1584274,4	258161,3	260610,8

* Datos en millones de pesos y a precios corrientes del mercado

El modelo a ser considerado es:

$$IMPOR = b_0 + b_1\ CPRIV + b_2\ CGOB + b_3\ PIB + b_4\ EXPOR$$

Siendo:

IMPOR = Importaciones
CPRIV = Consumo privado
CGOB = Consumo del gobierno
P I B = Producto interno bruto
EXPOR = Exportaciones.

Los resultados de la regresión usando el método ordinario de mínimos cuadrados, Tabla 2, revelan que a pesar del buen ajuste logrado, con un $R^2 = 0.9939$, los valores T calculados para b_1, b_3 y b_4 son pequeños y, según las pruebas de hipótesis no tienen influencia significativa en las importaciones. Esto parece contradecir la experiencia: un mayor consumo debe influir positiva y decisivamente en las importaciones. El coeficiente negativo correspondiente a la variable consumo privado hace duduar aún más de esta solución.

TABLA 2. Resumen de resultados obtenidos para el modelo de importaciones con el conjunto total de variables independientes.

Variable	Media	Desviación estándar	b	T
CPRIV	159721,25	255959,0	–0,1122	–0,566
CGOB	17042,45	29025,64	2,162	2,600
P I B	230847,75	382468,75	0,0602	0,226
EXPOR	36298,74	63240,97	0,0144	0,024
IMPOR	33820,93	58184,09	—	—
CONSTANTE	—	—	466,24	—

$R^2 = 0{,}9939$ S = 4855,8   D = 1,5453   N.C. = 70603,94

Siendo:
b = coeficiente de regresión estimado
T = valor T calculado
R^2 = coeficiente de determinación
S = desviación estándar del modelo
D = estadístico de Durbin — Watson
N.C. = número condicional

La presencia de multicolinealidad está comprobada por el tamaño de los coeficientes de correlación simple entre las variables explicativas como se muestra en la Tabla 3. En este caso se evidencia que dos a dos estas variables están altamente correlacionadas. Es importante observar que este criterio basado en los coeficientes de correlación simple no es siempre revelador de la multicolinealidad cuando más de dos variables se encuentran correlacionadas entre sí; en estos casos se recomienda usar otros criterios que incluyan coeficientes de correlación múltiples de cada variable explicativa respecto de las demás, u observar indicadores como el número condicional de la matriz $X^T X$, el cual, en el ejemplo presente, también es revelador por su tamaño.

TABLA 3. Coeficientes de correlación simple entre variables independientes. Modelo de importaciones.

	CPRIV	CGOB	P I B	EXPOR
CPRIV	1	0,998	0,999	0,987
CGOB		1	0,998	0,996
P I B			1	0,999
EXPOR				1

Luego de comprobar la inexistencia de otras dificultades en el modelo y en los datos, se presenta detalladamente la manera de usar el método de la Traza de Borde.

La ecuación (2) es usada, considerando 38 valores diferentes de k entre cero y uno, para calcular los correspondientes conjuntos de coeficientes de borde estandarizados que se muestran en la Tabla 4, en la cual se incluyen además los valores calculados para la suma de residuales cuadrados (SSE). El gráfico de la Traza de Borde y la variación de los SSE se muestran respectivamente en las Figuras 1 y 2. En ellas se puede apreciar la gran inestabilidad de los coeficientes de borde cuando se producen pequeñas perturbaciones en la diagonal principal de la matriz $X^T X$. Los coeficientes b_1^* y b_2^* aparecen sobreestimados en valor absoluto para k = 0; $b_1{}^*$ en las vecin-

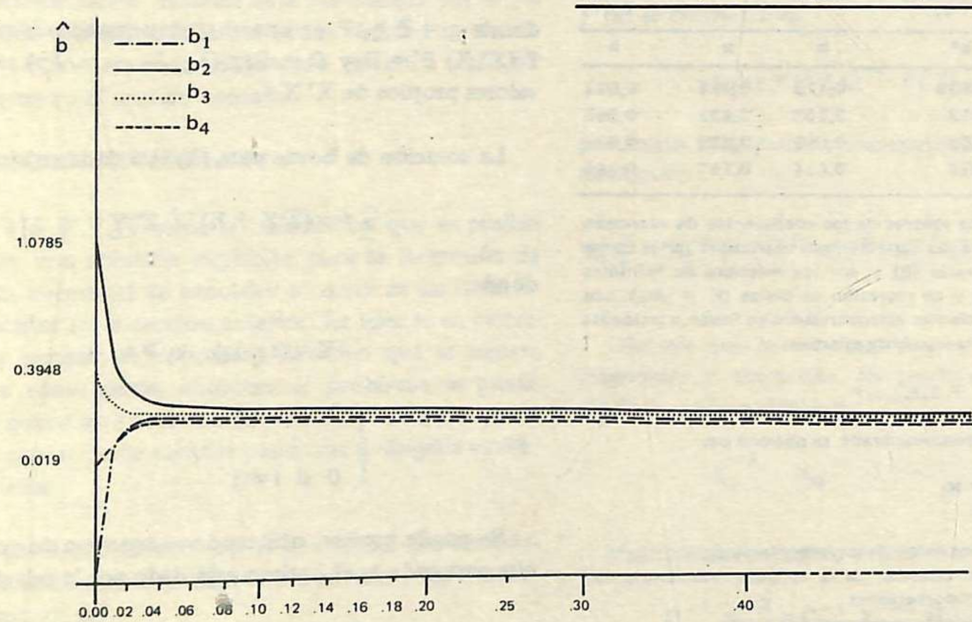

FIGURA 1. Traza de borde para los datos de importaciones.

TABLA 4. Coeficientes de borde estandarizados para los datos de importaciones.

k	b_1	b_2	b_3	b_4	SSE ($\times 10^9$)
0,000	−0,494	1,079	0,396	0,016	.6130
0,001	−0,095	0,862	0,161	0,068	.6391
0,002	0,014	0,729	0,178	0,075	.6701
0,003	0,071	0,646	0,190	0,089	.6956
0,004	0,106	0,588	0,199	0,102	.7164
0,005	0,129	0,545	0,205	0,115	.7338
0,006	0,147	0,512	0,210	0,126	.7477
0,007	0,159	0,486	0,214	0,135	.7597
0,008	0,169	0,464	0,217	0,143	.7706
0,009	0,177	0,447	0,219	0,150	.7790
0,010	0,184	0,431	0,222	0,157	.7866
0,011	0,189	0,418	0,224	0,162	.7938
0,012	0,194	0,407	0,226	0,167	.7991
0,013	0,198	0,397	0,227	0,171	.8050
0,014	0,201	0,388	0,228	0,175	.8106
0,015	0,204	0,381	0,229	0,178	.8145
0,016	0,207	0,374	0,230	0,182	.8179
0,017	0,209	0,367	0,231	0,185	.8225
0,018	0,211	0,362	0,232	0,188	.8248
0,019	0,213	0,357	0,232	0,190	.8283
0,020	0,214	0,352	0,233	0,192	.8320
0,040	0,229	0,302	0,239	0,216	.8700
0,060	0,234	0,284	0,240	0,224	.8901
0,080	0,236	0,273	0,240	0,228	.9145
0,100	0,236	0,266	0,240	0,299	.9451
0,120	0,236	0,261	0,239	0,231	.9701
0,140	0,236	0,257	0,238	0,231	1.004
0,160	0,235	0,254	0,238	0,231	1,035
0,180	0,235	0,251	0,237	0,231	1,069
0,200	0,234	0,249	0,236	0,230	1,115
0,250	0,232	0,244	0,233	0,229	1,234
0,300	0,229	0,240	0,231	0,227	1,377
0,400	0,225	0,232	0,226	0,223	1,719
0,500	0,220	0,226	0,221	0,219	2,126
0,600	0,215	0,221	0,216	0,214	2,614
0,700	0,211	0,215	0,212	0,209	3,154
0,800	0,207	0,210	0,207	0,206	3,698
0,900	0,203	0,206	0,203	0,202	4,264

dades de cero tiene un contradictorio signo negativo. Sin embargo, cuando se varía el valor de k desde cero, en sentido positivo, estos dos coeficientes decrecen rápidamente en valor absoluto y $b_1{}^*$ cambia de signo. En el intervalo (0.01, 0.02) las cuatro curvas empiezan a estabilizarse, es decir, las variaciones con k dejan de ser bruscas.

Bajo estas circunstancias se sugiere escoger k = 0.01 para determinar el mejor conjunto de estimadores de borde. Los resultados se muestran en la Tabla 5.

Es importante observar que para el conjunto de borde escogido el SSE tiene un valor de 0,7866 x 10^9 contra el SSE del estimador OMC 0,6130 x 10^9 que no es un incremento apreciable si se tiene en cuenta que en esas condiciones el coeficiente de determinación sólo disminuye de 0,9939 para el estimador OMC a 0,9922 para el conjunto de borde con k = 0.01. Así el modelo ajustado será:

IMPOR = 615,49 + 0,042. CPRIV + 0,865. CGOB + 0,034. PIB + 0,144. EXPOR

que da una más plausible representación de las relaciones entre importaciones y las variables independientes anteriormente descritas.

TABLA 5. Resumen de los coeficientes estimados para los datos de importaciones*

Variable	Mínimos cuadrados (k = 0)			Regresión de Borde (k = 0,01)
	b*	b	b	b
CONS-PRI	–0,4939	–0,112	0,184	0,042
CONS-GOB	1,079	2,162	0,431	0,865
PR-INT-B	0,396	0,061	0,222	0,034
EXPORTAC	0,016	0,014	0,157	0,144

* La tabla muestra los valores de los coeficientes de regresión estimados tanto para las variables estandarizadas (b*)) como para variables originales (b) y por los métodos de mínimos cuadrados (k = 0) y de regresión de borde (k = ,01). Los valores de los coeficientes estandarizados se llevan a unidades originales mediante la siguiente relación:

$$\hat{b}_i{}^* = \hat{b}_i \left(\frac{S_y}{S_{x_i}}\right), \quad i = 1,2,\ldots, P$$

El intercepto b_o desestandarizado se obtiene de:

$$\hat{b}_o{}^* = \overline{Y} - \sum_{i=1}^{p-1} \hat{b}_i{}^* X_i$$

Siendo:

S_y = la desviación estándar de las importaciones

S_{x_i} = la desviación estándar de la variable independiente i

Y = media de las importaciones

X_i = media de la variable independiente i

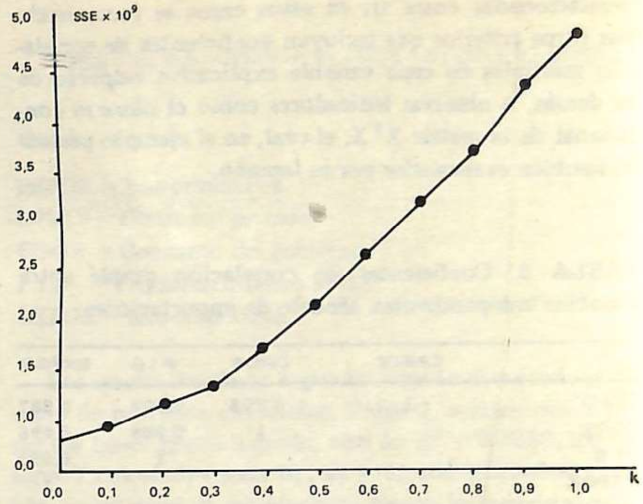

FIGURA 2. Comportamiento del SSE respecto a k.

2. PROCEDIMIENTO ITERATIVO

En un artículo posterior Hoerl y Kennard[2] sugieren un método iterativo a partir de la solución del modelo llamado "Generalizado".

Se parte de $\underline{Y} = X\,\underline{b} + \underline{e}$ como modelo estándar; es decir, X' X es la matriz de coeficientes de correlación simple entre los vectores regresores, E ($\underline{e}$) = $\underline{0}$ y E ($\underline{e}\,\underline{e}'$) = $w^2 I_n$. Y se construye el modelo:

$$\underline{Y} = Z\,\underline{a} + \underline{e} \tag{1}$$

donde $\underline{a}$ = P $\underline{b}$; P es una matriz normalizadora tal que P (X' X) P' = D; y D = diag ($\lambda_1, \lambda_2, \ldots, \lambda p$) matriz de valores propios de X' X.

La solución de borde para (1) está dada entonces por:

$$\hat{\underline{a}}^* = (Z'Z + K)^{-1}\, Z'\,\underline{Y} \tag{2}$$

donde:

$$K = (\delta_{ij}\, k_i), \quad k_i \geqslant o,$$

$$\text{y:} \quad \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Se puede probar, utilizando el teorema de existencia, que para cada k_i el óptimo está dado por la relación:

$$k_i = w^2 / a_i{}^2 \qquad i = 1, 2, \ldots, p.$$

Con base en lo anterior se propone entonces el siguiente procedimiento iterativo:

$$k_{i(j)} = w^2 / (a_i^*{}_{(j)})^2 \qquad i = 1, 2, \ldots, p \qquad (3)$$

donde el subíndice j se usa para denotar la j - ésima iteración.

Como valor inicial se usa:

$$a_i^*(0) = \hat{\underline{a}}_i \qquad i = 1, 2, \ldots, p \qquad (4)$$

donde $\hat{\underline{a}}_i$ es el estimador ordinario de mínimos cuadrados

Así:

$$k_{i(0)} = w^2 / a_i^*(0) \qquad i = 1, 2, \ldots, p,$$

y estos valores de k son usados en (2) a fin de obtener el siguiente $a_i^*{}_{(1)}$ que luego es usado en (3) para obtener $k_{i(1)}$.

En general, al encontrar un $k_{i(j)}$ se usa (2) para determinar $a_i^*{}_{(j+1)}$ y luego (3) para encontrar $k_{i(j+1)}$. Por su parte w^2 se estima por $\hat{\underline{w}}^2$ proporcionado por el método de mínimos cuadrados.

Ahora bien, el proceso iterativo concluye cuando la longitud cuadrada del vector de coeficientes de regresión estimados $(\hat{\underline{a}}^*{}_{(k)})'(\hat{\underline{a}}^*{}_{(k)})$ sea estable, es decir, cuando comparada con el obtenido en la iteración anterior la diferencia es menor que un error pequeño arbitrariamente escogido.

El procedimiento anterior está justificado por el hecho de que los estimadores de borde tienden a estabilizarse y en consecuencia el proceso iterativo debe ser convergente en el sentido anotado.

3. UNA SOLUCION EXPLICITA

En 1975 W.V. Hemmerle[4] demuestra que es posible encontrar una solución explícita para la Regresión de Borde sin necesidad de proceder a través de las iteraciones señaladas en la sección anterior. La idea es en principio muy simple: si el proceso iterativo que se sugiere arriba es convergente, entonces el problema se puede manejar como un problema de punto fijo siempre que la variable que se desea calcular pueda ser despejada en forma explícita.

He aquí cómo se enfrenta el problema. Las ecuaciones de iteración pueden expresarse en una forma matricial compacta, así:

$$A_{j+1} = (D + w^2 A_j^{-2})^{-1} B \qquad (5)$$

donde:

$$B = \begin{bmatrix} (Z'\underline{Y})_1 & 0 & \cdots & 0 \\ 0 & (Z'\underline{Y})_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & (Z'\underline{Y})_p \end{bmatrix}$$

$$A_j = \begin{bmatrix} a^*_{1(j)} & 0 & \cdots & 0 \\ 0 & a^*_{2(j)} & \cdots & \cdot \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a^*_{p(j)} \end{bmatrix}$$

La ecuación (5) se puede llevar, por medio de transformaciones algebraicas muy simples, a la forma:

$$A_{j+1} = (I + w^2 D^{-1} A_j^{-2})^{-1} A_0. \qquad (6)$$

Se hace luego:

$$C = \frac{D}{w^2} \qquad (7)$$

y (6) se convierte en:

$$A_{j+1} = (I + C^{-1} A_j^{-2})^{-1} A_o; \qquad (8)$$

por tanto la expresión correspondiente para A_{j+1}^{-2} está dada por:

$$A_{j+1}^{-2} = A_o^{-1} (I + C^{-1} A_{j+1}^{-2}) A_o^{-1} (I + C^{-1} A_j^{-2}) \qquad (9)$$

Observe que las matrices A_o, e, $I + C^{-1} A_j^{-2}$ son diagonales y conmutan. Se puede entonces simplificar (9) de la manera siguiente:

$$A_{j+1}^{-2} = A_o^{-2} (I + C^{-1} A_j^{-2})^2. \qquad (10)$$

Multiplicando ambos miembros por C^{-1} se tiene:

$$C^{-1} A_{j+1}^{-2} = C^{-1} A_o^{-2} (I + C^{-1} A_j^{-2})^2 \qquad (11)$$

y haciendo

$$E_j = C^{-1} A_j^{-2}, \tag{12}$$

el proceso iterativo se reduce a la fórmula simple:

$$E_{j+1} = E_o (I + E_j)^2. \tag{13}$$

Ahora bien, si el proceso iterativo dado por (13) es convergente la expresión siguiente es cierta:

$$\lim_{j \to \infty} E_j = \lim_{j \to \infty} E_{j+1} = E^*, \tag{14}$$

y en consecuencia, tomando límites a ambos lados de la ecuación (13) se obtiene:

$$E^* = E_o (I + E^*)^2. \tag{15}$$

Para encontrar E^* basta resolver las p ecuaciones cuadráticas de la forma:

$$(e_i^*)^2 + (2 - 1/e_o)\, e_i^* + 1 = 0,\ i = 1,2,\ldots, p \tag{16}$$

donde e_i^* y e_o son respectivamente las entradas de la diagonal de las matrices E^* y E_o.

Tal solución es equivalente a:

$$e_i^* = \frac{(1 - 2\, e_o) \pm \sqrt{(1 - 4\, e_o)}}{2\, e_o} \tag{17}$$

Resta sólo establecer las condiciones de convergencia del proceso iterativo dado por la ecuación (13). Para tal fin el siguiente teorema puede ser probado (Hemmerle, 1975)[4]:

"El proceso iterativo definido por $e_i\ (j + 1) = e_{i(0)}\ (1 + e_{i(j)})$ donde el subíndice j denota la j-ésima iteración, converge para $0 < e_o \leqslant 1/4$. El signo de la raíz cuadrada que debe ser escogido es el signo menos. Si $e_o > 1/4$ el proceso iterativo diverge y en este caso el correspondiente $a_i^*{}_{(j)}$ tiende a cero".

Los pasos a seguir utilizando este procedimiento se ilustran a continuación con el ejemplo tomado de Marquart[6].

La matriz X estandarizada y el vector observado $\underline{Y}$ son dados para este ejemplo:

$$X = \begin{bmatrix} 3\sqrt{2}/10 & 4\sqrt{2}/10 \\ 4\sqrt{2}/10 & 3\sqrt{2}/10 \\ 5\sqrt{2}/10 & 5\sqrt{2}/10 \end{bmatrix}$$

e,

$$Y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

1) Se calcula X' X y X' Y.

$$X'X = \begin{bmatrix} 1 & 49/50 \\ 49/50 & 1 \end{bmatrix}$$

y,

$$X'Y = \begin{bmatrix} \dfrac{26\sqrt{2}}{10} \\ \dfrac{25\sqrt{2}}{10} \end{bmatrix}$$

2) Se encuentran las matrices P y D.

$$P = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ \sqrt{2}/2 & -\sqrt{2}/2 \end{bmatrix},\ y,\ D = \begin{bmatrix} 99/50 & 0 \\ 0 & 1/50 \end{bmatrix}$$

En realidad P es una matriz formada por columnas con los vectores propios de la matriz X' X.

3) Se halla el estimador ordinario de mínimos cuadrados para $\underline{a}$.

$$Z'\,\underline{Y} = PX'\,\underline{Y} = \begin{bmatrix} 51/10 \\ 1/10 \end{bmatrix}$$

y,

$$\hat{\underline{a}} = (Z'Z)^{-1}\, Z'\,\underline{Y} = D^{-1}\, PX'\,\underline{Y}$$

o,

$$\hat{\underline{a}} = \begin{bmatrix} 85/33 \\ 5 \end{bmatrix}$$

4) Se estima w^2

$$\hat{w}^2 = \underline{Y}'\,\underline{Y} - \underline{a}'\,(Z'\,\underline{Y}) = 12/33$$

5) Los $e_{i(0)}$ para i = 1, 2 se calculan por la fórmula:

$$e_i\,(0) = \frac{\hat{w}^2}{\lambda_i\,(\hat{a}_i)^2}$$

$$e_1(0) = \frac{12/33}{99/50)\,(85/33)^2} = 8/289 < 1/4$$

$$e_2(0) = \frac{12/33}{1/50 \times 5^2} = 24/33 > 1/4$$

6) Cálculo de e_i^* para $i = 1, 2$

Como $e_{1(0)} < 1/4$ para la ecuación correspondiente el proceso iterativo converge y:

$$e_1^* = \frac{1 - 2\,e_o - \sqrt{(1 - 4\,e_o)}}{2\,e_o}$$

o,

$$e_1^* = 0,02932924$$

Como $e_{2(0)} > 1/4$ el proceso diverge.

7) Cálculo de $\hat{a}_i^*$.

Con e_1^* calculado en el paso 6), $\hat{a}_1^*$ se calcula por la fórmula:

$$\hat{a}_1^* = \frac{\hat{a}_1}{1 + e_1^*}$$

o,

$$\hat{a}_1^* = 2.50$$

y como $e_{2(0)} > 1/4$, $\hat{a}_2^* = 0$.

8) Cálculo de $\hat{b}_i^*$.

Como $\underline{a} = P\,\underline{b}$ entonces:

$$\begin{bmatrix} \hat{b}_1^* \\ \hat{b}_2^* \end{bmatrix} = p^{-1}\,\underline{\hat{a}}^*$$

o sea que:

$$\hat{b}_1^* = \hat{b}_2^* = 1,7694$$

4. OTROS PROCEDIMIENTOS

Wermuth (1972)[5] hace notar que la condición necesaria para minimizar Tr MSE (a(k)) con respecto a k es:

$$w^2 \sum_{i=1}^{P} \frac{\lambda_i}{(\lambda_i + k)^3} = k \sum_{i=1}^{P} \frac{\lambda_i\ a^2 i}{(\lambda_i + k)^3} \qquad (18)$$

y sugiere reemplazar w^2 y a_i^2 respectivamente por los correspondientes estimadores de mínimos cuadrados, para proceder luego a resolver la anterior ecuación para k.

Tal propuesta, aparentemente ingeniosa y simple, no es recomendable ya que los estimadores de mínimos cuadrados están afectados por la multicolinealidad presente en el modelo. Ya se ha visto cómo en la sección 2 se elude este facilismo y en su lugar se explora iterativamente el estimador de borde. La consecuencia que se deriva de su utilización es la subestimación de los valores de k con lo cual se estaría aún sobre-estimando la magnitud del vector de coeficientes. En el ejemplo que se viene analizando los coeficientes estimados por este método son $\hat{b}_1^* = 2,70$ $\hat{b}_2^* = 4,16$.

Hoerl y Kennard en 1975 retomaron esta solución y en compañía de Baldwin[3], proponen como alternativa el uso de una media armónica a partir de la solución de la ecuación (18). Todo esto con el fin de obtener un solo valor de k y contrarrestar al mismo tiempo el efecto que tiene en la estimación la presencia de coeficientes muy pequeños. Como (18) tiene por solución $k_i = w^2 / a_i^2$ entonces es apropiado calcular k_{HKB} así:

$$(1/k_{HKB}) = (1/p) \sum_{i=1}^{P} 1/k_i = (1/p) \sum_{i=1}^{P} a_i^2 \,/\, w^2 = 1/pw^2 \sum_{i=1}^{P} a_i^2 = \underline{a}'\,\underline{a}/pw^2$$

de donde:

$$k_{HKB} = p\,\hat{w}^2 \,/\, \underline{\hat{a}'\,\hat{a}} \qquad (19)$$

Es de anotar que esta propuesta fue comprobada por los autores usando simulación de una gran cantidad de posibilidades.

Por su parte, Thisted (1976)[5] encuentra que el anterior estimador HKB tiende a sobrecontraer el vector de coeficientes; él propone entonces modificarlo así:

$$K_T = (p - 2)\,\hat{w}^2 \,/\, \underline{\hat{a}'\,\hat{a}} \qquad (20)$$

Esta propuesta es válida especialmente para p grande y no tiene sentido usarla para p igual o cercano a 2.

5. NUMERO CONDICIONAL Y ESTIMADORES DE BORDE

Una interpretación del fenómeno de multicolinealidad indica que ésta es una enfermedad de la matriz de datos X. Concretamente se dice que al presentarse una alta correlación entre algunos de los vectores regresores la matriz X' X se comporta como matriz mal condicionada. Como consecuencia de ello la solución de mínimos cuadrados, que implica el cálculo de la inversa de

X' X, no es confiable: pequeñas perturbaciones en las entradas de la matriz X, o en las entradas del vector Y, producen grandes cambios en la estimación de los coeficientes de regresión.

Si por cualquier circunstancia se considera que el problema no es el de modelo adoptado, entonces la atención debe concentrarse en el análisis de los datos y muy especialmente en el análisis del mal condicionamiento de X'X. En este caso se sugiere hacer el análisis a partir de la medición del grado de condicionamiento de la matriz X'X.

Una medida de tal naturaleza es el conocido "número condicional" que se seguirá denotando aquí por la letra "C". En el caso de matrices simétricas y positivamente definidas, como X', es ampliamente aceptado definir C por la relación:

$$C = \frac{\lambda \max}{\lambda \min} \qquad (21)$$

es decir, como el cociente entre el valor propio máximo y el valor propio mínimo de la matriz en cuestión.

Este número es igual a 1 si la matriz es perfectamente ortogonal. Por el contrario, es mayor que 1 y tiende a ser muy grande en la medida en que la matriz se aleja mucho de la ortogonalidad. Además el número condicional es un factor de amplificación del error relativo que se comete al resolver la ecuación lineal $Ax = b$ en relación a la perturbación relativa del vector b. En la práctica este número es calculado sólo de manera aproximada.

En el problema de regresión de borde se trata de encontrar un valor de $k > 0$ para el cual la solución del sistema lineal $(X' X + kI_p) \underline{b} = X' \underline{Y}$ sea estable en el sentido siguiente: pequeñas perturbaciones en las observaciones de las variables que intervienen en el modelo NO deben producir grandes cambios en el cálculo de los coeficientes de regresión.

Ahora bien, si se calcula el número condicional C_k para cada una de las matrices (X'X + k Ip), variando k desde 0 hasta 1, es de esperar que el comportamiento de C_k tienda a la estabilización; de alguna manera esto es similar al comportamiento que tiene la traza de borde, ya que la estabilidad de ésta significa cambios pequeños en los coeficientes de borde. Los autores trabajan actualmente en este problema y se proponen publicar próximamente un artículo bajo el título: "Regresión de Borde y Número Condicional".

Agradecimiento especial a José Alejandro Gaviria por su colaboración en la consecución y el procesamiento de parte de la información utilizada.

BIBLIOGRAFIA

0 GAVIRIA, A. Desarrollo de un sistema para estudios modernos sobre regresión múltiple, trabajo de Grado, Universidad Nacional de Colombia. Medellín, 1985.

1 HOERL and KENNARD. Ridge Regression: Biased estimation for nonorthogonal problems. Technometrics. Vol. 12, pp. 55 - 67, 1970.

2. HOERL and KENNARD. Ridge Regression: Applications to nonorthogonal problems. Technometrics. Vol. 12, pp. 69-82, 1970.

3. HOERL, KENNARD and BALDWIN. Ridge Regression: Some simulations. Communications in Statistics. Vol. 4, No. 2, pp. 105 - 123, 1975.

4. HEMMERLE, William J. An explicit solution for generalized ridge. Technometrics. Vol. 17, No. 3, pp. 309 - 314, 1975.

5. LIN and KMENTA. Ridge regression under alternative loss criteria. The Review of Economics and Statistics, pp. 488 - 494, dic. 1, 1981.

6. MARQUARDT, D.W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics, vol. 12, pp. 591 - 612, 1970.

7. PEREZ, L y Cabarcas, G. Regresión de Borde. DYNA, 105. 1985.