ESTIMACION DE UN MODELO DE REGRESION CON RESTRICCIONES DE ORDEN EN LOS PARAMETROS

Luis Borge González

RESUMEN.— El problema abordado en este trabajo consiste en estimar un modelo de regresión simple en K poblaciones bajo la hipótesis de que las pendientes de las rectas estimadas verifiquen una relación de orden, con el fin de posteriormente, contrastar dicha hipótesis.

1. Introducción

El interés en la estimación de modelos en los que los coeficientes de regresión están sujetos a restricciones con desigualdad es creciente cada día. Trabajos en modelos de desequilibrio, modelos Tobit y Probit y ciertos modelos de ecuaciones simultáneas muestran como aparecen las restricciones en forma de desigualdad en los parámetros. Su importancia ha hecho que tales problemas vengan ya enunciados en manuales sobre Métodos Econométricos.

El problema abordado en este trabajo consiste en estimar un modelo de regresión simple en K poblaciones bajo la hipótesis de que las pendientes de las rectas estimadas verifiquen una relación de orden, con el fin de posteriormente, contrastar dicha hipótesis. Plantearemos el problema con el siguiente modelo:

$$y_{i,j} = \alpha_i + \beta_i x_{i,j} + \varepsilon_{i,j}; \qquad \begin{array}{l} j = 1,2,\ldots,n_i \\ i = 1,2,\ldots.,k \end{array} \qquad (1.1)$$

siendo $x_{i,j}$ fijos, $\varepsilon_{i,j}$ las perturbaciones aleatorias, que supondremos independientes entre sí, de media cero, y

$$\beta_1 \leqslant \beta_2 \leqslant \ldots \leqslant \beta_k \qquad (1.2)$$

El problema así planteado, es bastante similar al de regresión isotónica, tratado por Barlow y otros (1972), y que podemos interpretar de la siguiente manera: dado un vector **a** = $(a_1, a_2,...,a_k)$ y unos pesos positivos w_i, encontrar el vector α^* que haga mínima la forma cuadrática

$$\sum_{i=1}^{k} (a_i - \alpha_i)^2 w_i$$

sujeto a:

$$\alpha_i \leqslant \alpha_{i+1}, \; i = 1,2,...,k-1$$

A dicha solución le daremos el nombre de regresión isotónica de **a**, con pesos w_i.

Este problema de regresión isotónica surge en la referencia anterior, al estimar la media de una población normal k-dimensional bajo la hipótesis de que las componentes siguen una relación de orden.

Nuestro planteamiento es similar suponiendo que las perturbaciones son normales, pero los parámetros vienen definidos por establecimiento del modelo (1.1) y con las restricciones (1.2).

La sección 2 la dedicaremos a la estimación y contrastes de las hipótesis planteadas, para pasar en la sección 3 a plantear un ejemplo en el cual podemos comprobar la aplicación de los resultados. El ejemplo se basa en el estudio del comportamiento de las Grandes Empresas Industriales Españolas frente al gasto en Tecnología de las mismas. Para adecuar dicho estudio al marco de los resultados teóricos tendremos que imponer fuertes hipótesis al modelo.

2. Estimación y contrastes de hipótesis

Supongamos el modelo de K poblaciones (1.1). Para efectuar los contrastes de hipótesis tendremos que suponer normalidad en las perturbaciones, y dado que bajo esta suposición los estimadores de los parámetros de las rectas de regresión coinciden con el estimador mínimo-cuadrático, en adelante supondremos que:

$$\varepsilon_{i,j} \sim N(0, \sigma^2), \; \forall \; i,j. \tag{2.1}$$

Si llamamos **b**,**a** y $\hat{\sigma}^2$, a los estimadores sin restricciones, y **b***, **a*** y σ^{*2}, a los estimadores con restricciones, dichos estimadores se obtienen al tratar de maximizar las correspondientes funciones de verosimilitud; en el caso sin restricciones tendremos:

$$b_i = \frac{\sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)'(y_{i,j} - \bar{y}_i)}{\sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2}$$

$$a_i = \bar{y}_i - b_i\bar{x}_i$$

$$\hat{\sigma}^2 = (1 / \sum_{i=1}^{k} {}_i) [\sum_{i=j}^{k} \sum_{j=1}^{n_i} (y_{i,j} - a_i - b_i x_{i,j})^2];$$

Para calcular los estimadores sujetos a restricciones, tendremos que maximizar la verosimilitud bajo las restricciones (1.2); puesto que estas no afectan a los parámetros de α, ni a σ^2, es fácil de comprobar que para un β fijo, los valores que maximizan la verosimilitud son de la forma:

$$a_i (\beta) = \bar{y}_i - \beta_i\bar{x}_i$$

$$\hat{\sigma}^2 (\beta) = (1/ \sum_{i=1}^{k} n_i) [\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{i,j} - a_i (\beta) - \beta_i x_{i,j})^2] \tag{2.3}$$

si llevamos estas condiciones a la función de verosimilitud, el estimador con restricciones de β, será el punto que hace mínima la expresión

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} [(y_{i,j} - \bar{y}_i) - \beta_i (x_{i,j} - \bar{x}_i)]^2 \tag{2.4}$$

sujeto a las restricciones (1.2).

Sumando y restando en la expresión (2.4) $b_i(x_{i,j} - \bar{x}_i)$, siendo b_i, las componentes del estimador sin restricciones, (2.4) se puede descomponer en dos términos de la forma:

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} [(y_{i,j} - \bar{y}_i) - b_i (x_{i,j} - \bar{x}_i)]^2 +$$

$$+ \sum_{i=1}^{k} [(b_i - \beta_i)^2 \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2] \tag{2.5}$$

el primer sumando no depende de los parámetros, luego tendremos que encontrar el valor de β que minimice el segundo sumando. Si llamamos:

$$w_i = \sum_{j=1}^{n_i} (x_{i,j} - \bar{x}_i)^2, \tag{2.6}$$

el problema será ahora el de encontrar el valor β que minimice:

$$\sum_{i=1}^{k} (b_i - \beta_i)^2 w_i \tag{2.7}$$

bajo las restricciones (1.2).

La solución del problema planteado es, como vimos en la introducción, la regresión isotónica de **b** con los pesos w_i definidos en (2.6).

La forma más rápida de obtener la regresión isotónica es como demuestran BARLOW y otros (1972) por medio del PAVA (Pool Adjacent Violators Algorithms). El algoritmo, esquemáticamente, consiste en lo siguiente:

Partiendo de **b**, se examinan sus componentes, y si verifican todas las restricciones, evidentemente él mismo es la solución. Si no lo es, es porque al menos para un j se verifica $b_j > b_{j+1}$. Estas dos componentes se sustituyen por el mismo:

$$b_j^1 = \frac{b_j w_j + b_{j+1} w_{j+1}}{w_j + w_{j+1}}$$

si en el primer paso, hubiera más restricciones abyacentes que no verifican las correspondientes restricciones, el «pooling» anterior lo haríamos con todas a la vez en el mismo bloque; y ésto para todos los bloques diferentes de restricciones no satisfechas. El nuevo punto puede no verificar alguna otra nueva restricción; se vuelve a hacer lo mismo que antes, en donde los pesos de los nuevos puntos distintos, serán la suma de los asociados al «pooling» del cual se obtuvo. Como el número de componentes es finito, se llegará a un punto, en un número finito de pasos que, en el caso más extremo, tendrá todas sus componentes iguales.

Antes de pasar a la siguiente etapa, la de hacer los contrastes de hipótesis, veamos el significado geométrico del estimador con restricciones, pues nos servirá de ayuda a la hora de buscar las distribuciones de los estadísticos que definen los contrastes. El problema planteado en (2.7) es el mismo que si queremos encontrar la proyección de un punto de R(k), **b**, sobre un subconjunto definido por las restricciones, que en nuestro caso forman un cono poliédrico convexo, utilizando la métrica definida por la matriz definida positiva y diagonal $W = (w_i)$. Un resultado bien conocido de programación convexa es que si el punto que se quiere proyectar no pertenece al cono (si así fuese, él mismo es la solución), la proyección, o punto más próximo a él, perteneciente al cono, está en la frontera; o dicho de otra manera, satisface una serie de restricciones con igualdad y el resto con desigualdad estricta, y se obtiene proyectando sobre el subespacio que definen las restricciones que ha de verificar con igualdad (ver BOOT (1968)). A dichas restricciones las vamos a llamar entrantes, y serán unas u otras dependiendo de la posición en la que se encuentre el punto que queremos proyectar.

Con esta idea vamos a definir una partición de R(K); para ello si x es un punto, llamaremos x* a la proyección de x sobre el cono de restricciones definido por (1.2):

sea $i_1, i_2, \ldots, i_s$, un subconjunto de restricciones,

$$x \in A_s^{i_1 \ldots i_s} \Longleftrightarrow \begin{cases} x_i^* = x_{i+1}^* & i = i_1 \ldots i_s \\ x_j^* < x_{j+1}^* & j \neq i_1, \ldots, i_s \end{cases} \tag{2.8}$$

En cuanto a notación, vamos a escribir el modelo en forma matricial haciendo:

$$Y_{ij} = (y_{ij} - \bar{y}_i), \qquad X_{ij} = (x_{ij} - \bar{x}_i),$$

$$Y = (Y_{11}, Y_{12}, \ldots, Y_{1,n_1}, \ldots Y_{k,n_k})'$$

$$X_i = (X_{i1}, \ldots, X_{in_i})'$$

$$X = \begin{pmatrix} X_1 & 0 & 0 & . & . & 0 \\ 0 & X_2 & 0 & . & . & 0 \\ . & . & . & . & . & . \\ 0 & . & . & . & . & X_k \end{pmatrix} \qquad (2,9)$$

El modelo en esta notación podemos escribirlo

$$Y = X\,\beta + \varepsilon,$$

sujeto a (2.10)

$$R\,\beta \leqslant 0$$

en donde ε tiene una distribución N $(0, \sigma^2\, I_{\Sigma n_i})$, y R es la matriz (K-1,K)

$$R = \begin{pmatrix} 1 & -1 & 0 & 0 & . & . & . & . & 0 \\ 0 & 1 & -1 & 0 & . & . & . & . & 0 \\ . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . \\ 0 & . & . & . & . & . & 0 & 1 & -1 \end{pmatrix} \qquad (2.11)$$

Vamos a llamar H_0 a la hipótesis de que todas las componentes de β son iguales, H_1 a la de que se verifiquen las restricciones del modelo (2.10), y H_2 a la de que los parámetros no están sometidos a restricciones. Los estimadores máximo-verosiles en cada hipótesis, que denotaremos $\mathbf{b}^o$, $\mathbf{b}^*$, y $\mathbf{b}$ y σ^{02}, σ^{*2}, $\hat{\sigma}^2$, para σ^2 en el caso que se desconozca, serán:

$$\mathbf{b}^o = \mathbf{b} - (X'X)^{-1}\, R'\, [R\,(X'X)^{-1}\, R'\,]^{-1}\, R\,\mathbf{b}; \qquad (2.12)$$

$\mathbf{b}^*$ será la regresión isotónica mencionada anteriormente, y que de acuerdo a la partición definida en (2.8), si $\mathbf{b}$ pertenece a un conjunto asociado a los subíndices i_i, i_2, ..., i_s, llamando R_s a la submatriz de R asociada a esos índices, entonces:

$$\mathbf{b}^* = \mathbf{b} - (X'X)^{-1}\, R_s'\, [R_s\,(X'X)^{-1}\, R_s']^{-1}\, R_s\,\mathbf{b}; \qquad (2.13)$$

y $\mathbf{b}$ será el estimador sin restricciones del modelo (2.10).

$$\mathbf{b} = (X'X)^{-1} X'Y;$$

cuando la varianza es desconocida, los estimadores son:

$$\sigma^{02} = (1/\sum_{i=1}^{k} n_i)\,(Y - X\mathbf{b}^{\circ})'(Y - X\mathbf{b}^{\circ}),$$

$$\sigma^{*2} = (1/\sum_{i=j}^{k} n_i)\,(Y - X\mathbf{b}^{*})'(Y - X\mathbf{b}^{*}), \qquad (2.14)$$

$$\hat{\sigma}^{2} = (1/\sum_{i=1}^{k} n_i)\,(Y - X\mathbf{b})'(Y - X\mathbf{b}).$$

Distinguiremos dos tipos de contrastes; aquéllos en que se supondrá la varianza conocida, y para los que se considera desconocida. A los estadísticos que definen la región crítica del test de razón de verosimilitud (TRV), les denominaremos, T_{01} cuando se contrasta H_0 frente a H_1, T_{12} para H_1 frente a H_2, en el caso de varianza conocida; cuando es desconocida a los estadísticos les denominaremos F_{01}, y F_{12}, respectivamente.

Se comprueba sin dificultad (ver L. BORGE 1985) que estos estadísticos son:

$$T_{01} = (\mathbf{b}^{\circ} - \mathbf{b}^{*})'X'X(\mathbf{b}^{\circ} - \mathbf{b}^{*}),$$

$$T_{12} = (\mathbf{b} - \mathbf{b}^{*})'X'X(\mathbf{b} - \mathbf{b}^{*}),$$

$$F_{01} = \frac{(\mathbf{b}^{\circ} - \mathbf{b}^{*})'X'X(\mathbf{b}^{\circ} - \mathbf{b}^{*})}{(Y - Y\mathbf{b}^{*})'(Y - X\mathbf{b}^{*})}$$

$$F_{12} = \frac{(\mathbf{b} - \mathbf{b}^{*})'X'X(\mathbf{b} - \mathbf{b}^{*})}{(Y - X\mathbf{b})'(Y - X\mathbf{b})}$$

El estudio de las distribuciones de estos estadísticos, está basado en la partición definida en (2.8). La distribución de los dos primeros puede verse en BARLOW y otros (1972), así como una tabla que, cuando la dimensión del espacio es menor o igual que cuatro, nos permite obtener los valores críticos para estos contrastes.

Para el segundo y cuarto se estudia la distribución bajo la hipótesis H_0, ya que es la que da mayor potencia dentro de H_1, (ROBERTSON y WEGMAN (1978)).

En el apéndice A damos la demostración sobre la distribución de F_{12}, y omitimos la de F_{01} por ser muy similar. Las distribuciones quedan caracterizadas con los teoremas que enunciamos a continuación.

Teorema 1

Sea $t > 0$,

$$P_{Ho}[T_{01} \geqslant t] = \sum_{m=0}^{k=2} P\,[\chi^2_{k-1-m} \geqslant t]\, P_{Ho}\,(m,k);$$

$$P_{Ho}\,[T_{01} = 0] = P_{Ho}\,(k\text{-}1,k).$$

Teorema 2

Sea $t > 0$,

$$P_{Ho}\,[T_{12} \geq t] = \sum_{m=1}^{k} P\,[\chi^2_m \geq t]\; P_{Ho}\,(m,k);$$

$$P_{Ho}\,[T_{12} = 0] = P_{Ho}\,(0,k).$$

Teorema 3

Sea $t > 0$,

$$P_{Ho}\,[F_{01} \geq t] = \sum_{m=0}^{k-2} P\,[F^{k-1-m}_{\Sigma n_i-k+m} \geq t]\; P_{Ho}\,(m,k);$$

$$P_{Ho}\,[F_{01} = 0] = P_{Ho}\,(k\text{-}1,k).$$

Teorema 4

Sea $t > 0$,

$$P_{Ho}\,[F_{12} \geq t] = \sum_{m=1}^{k-1} P[F^{m}_{\Sigma n_i-k} \geq t]\; P_{Ho}\,(m,k);$$

$$P_{Ho}\,[F_{12} = 0] = P_{Ho}\,(0,k).$$

Las probabilidades de la forma P_{Ho} (m,k) indican la probabilidad de que el estimador máximo verosímil sin restricciones **b**, pertenezca a la región del espacio R(k) tal que al proyectar un punto cualquiera de él sobre el cono, la proyección satisfaga m restricciones con igualdad; en la hipótesis nula son por tanto probabilidades de sucesos definidos por una variable aleatoria normal conocida, que en el caso de varianza conocida y $k=4$ en BARLOW y otros (1972) viene la forma de obtención, y una tabla de dichas probabilidades para distintos valores de los pesos w_i.

3. Ejemplo

En el ejemplo que hemos elegido para ilustrar el uso de alguno de los resultados teóricos expuestos anteriormente, vamos a estudiar, adecuado al problema conveniente, y con las hipótesis necesarias del modelo, la relación existente entre las variables gasto tecnológico y tamaño de una empresa.

Antes de entrar en el problema, queremos señalar que si bien el tamaño de las empresas no es un factor determinante para explicar la estructura del gasto tecnológico en nuestro país, sí que es el factor protagonista al menos en términos absolutos, como señalan LAFUENTE, SALAS y YAGUE (1985) en un reciente trabajo. Resultado además que corrobora el obtenido por otros países.

La idea Schumpeteriana, sostenida por KENNEDY and THIRLWALL (1972) entre otros muchos autores, de que el esfuerzo tecnológico tiende a crecer con el tamaño de las empresas hasta llega a un cierto umbral, es lo que se pretende contrastar.

El esfuerzo tecnológico en un grupo reducido de países (los más desarrollados), es una variable que está centrada en gran medida en las tareas de investigación, mientras que para el resto, entre los que se encuentra el nuestro, se ve complementado e incluso superado por la incorporación de tecnología del exterior. Si a ésto unimos la laguna existente en la medición de las actividades tecnológicas, nos vemos obligados a considerar como gasto tecnológico, la suma de los gastos en investigación y desarrollo (I + D), más los gastos derivados de la adquisición de tecnología procedente del exterior, identificados como los pagos por licencias y asistencia técnica.

En cuanto a la variable que representa el tamaño de las empresas, entre las opciones posibles, hemos elegido el valor añadido neto, dado que este agregado nos permite hacer hipótesis sobre un mayor o menor esfuerzo tecnológico sin discriminar entre las empresas, según pertenezcan a industrias pesadas o ligeras.

La fuente de datos utilizada es la Encuesta de las Grandes Empresas Industriales Españolas del Ministerio de Industria de 1980. Lamentablemente en esta encuesta no consta de forma individualizada los pagos tecnológicos al exterior por licencias y asistencia técnica, sino los saldos netos de Pagos menos Ingresos. No obstante, como en los datos sectoriales si presenta tal información por separado comprobamos que los ingresos suponen únicamente el 10% de los pagos y en treinta de los treinta y nueve sectores considerados los ingresos son prácticamente nulos. Por ello hemos considerado que podemos utilizar la variable pagos menos ingresos como sustitución de pagos por licencia y asistencia técnica.

Queremos también hacer constar que dado que la encuesta se refiere sólo a las grandes empresas industriales la diversificación sectorial, con respecto al esfuerzo tecnológico, de las pequeñas empresas hemos considerado que no se ve reflejada en los datos.

Una vez tenidas en cuenta todas estas salvedades, vamos a tratar de contrastar el supuesto de que las empresas a medida que aumentan su tamaño aumentan los gastos en tecnología, siendo este aumento más significativo en las empresas grandes que en las pequeñas.

Para realizar este contraste dentro del contexto del modelo establecido anteriormente, hemos dividido a las empresas en cuatro grupos según su tamaño de tal forma que en cada grupo hay el suficiente número de empresas y que los ajustes lineales en cada uno de ellos sean significativos.

El modelo que se ha estimado es:

$$GT = \alpha + \beta \, VAN + \varepsilon \qquad (3.1)$$

Para poder aplicar los resultados y obtener los valores críticos de los test a los que se refieren los teoremas 1 y 2, necesitamos conocer la varianza de ε. Puesto que los datos constituyen un «corte transversal» hemos constatado la presencia de heteroscedasticidad para cualquier nivel de significación con el test de correlación de Sperman, pues obtenemos un coeficiente de correlación de 0,668 entre el módulo de los resíduos y la variable VAN.

Para corregir la heteroscedasticidad, hemos planteado varios supuestos de crecimiento de la varianza de la forma

$$\sigma_i^2 = \sigma^2 (VAN_i)^a.$$

El valor de «a» para el que los residuos se ajustan más a una variable aleatoria Normalizada es 3/2, luego el modelo elegido para corregir la heteroscedasticidad, y construir un estimador de la varianza ha sido

$$\sigma_i^2 = 0.06 \, (VAN_i)^{3/2}.$$

Una vez conocido cual va a ser el estimador de la varianza que vamos a utilizar, hemos planteado el problema en los siguientes términos

$$gt_{ij} = \alpha_i + \beta_i \, van_{ij} + \varepsilon_{ij} \qquad \begin{array}{l} i=1,2,3,4. \\ j=1,\ldots,n_i \end{array}$$

siendo

$$\varepsilon_{\iota\xi} \quad N[0, 0.06(van_{ij})^{3/2}] \text{ e independientes}$$

En la hipótesis H_2 de que las pendientes son cualesquiera, se han obtenido como estimadores

$$\mathbf{b}_1 = 0.0173267$$
$$\mathbf{b}_2 = 0.0371382$$
$$\mathbf{b}_3 = 0.02530802$$
$$\mathbf{b}_4 = 0.0481182.$$

En la hipótesis H_1, se han obtenido los estimadores mediante la regresión isotónica de b, con pesos w_i, definidos en este caso como

$$w_i = \sum_{j=1}^{n_i} \frac{(van_{ij} - \overline{van_i})^2}{0.06 \, (van_{ij})^{3/2}}$$

en donde, debido a la heteroscedasticidad que se ha detectado, $\overline{van_i}$, es la media ponderada por $(van_{ij})^{3/2}$ en cada grupo, y siendo los pesos el vector

$$w = (16916, 1833, 6794, 35298)'.$$

La regresión isotónica obtenida es:

$$\mathbf{b}_1^* = 0.0173267$$
$$\mathbf{b}_2^* = 0.03085935$$
$$\mathbf{b}_3^* = 0.03085935$$
$$\mathbf{b}_4^* = 0.0481182.$$

En la hipótesis nula H_o, de que $\beta_i = \beta_{i+1}\ \forall\ i$, se obtiene un estimador con todas las componentes calculadas como:

$$b_i^o = \frac{\sum_{i=1}^{4} w_i b_i}{\sum_{i=1}^{4} w_i} = 0.03678074,$$

rechazaremos H_o, frente a H_1, para valores grandes del estadístico:

$$T_{01} = \sum_{i=1}^{4} w_i\,(b_i^* - b_i^o)^2,$$

que para nuestros datos tiene un valor de $11.8482 > 9.385$, valor crítico para este test, luego se rechaza la hipótesis nula de que las pendientes son iguales frente a la alternativa de que están ordenadas.

El estadístico que define el test para contrastar H_1 como hipótesis nula, frente a H_2, como alternativa es:

$$T_{12} = \sum_{i=1}^{4} w_i\,(b_i - b_i^*)^2$$

que tiene un valor de 0.96368, perteneciente a la región de aceptación, luego no podemos rechazar la hipótesis H_1 con una confianza del 99%.

4. Sumario y conclusiones

En este artículo se ha comentado cómo se resuelven ciertos problemas de regresión con restricciones de orden en los parámetros, así como las distribuciones de los estadísticos que definen los test para contrastar si los parámetros que representan las pendientes de rectas de regresión en diferentes poblaciones, siguen la mencionada relación de orden.

En el estudio práctico que se ha realizado a título ilustrativo, puede concluirse, que si bien las pendientes que dan máximo ajuste, esto es, sin restricciones, no están ordenadas, mediante los test descritos anteriormente, se comprueba que no se puede rechazar la hipótesis de que las pendientes siguen una relación de orden.

Este resultado debe ser analizado con ciertas reservas, pues como ya se ha dicho, el tamaño de las empresas, según los estudios aludidos, es el factor más importante a la hora de explicar el comportamiento del gasto tecnológico, que no es el único determinante en su explicación. Para ilustrar ésto, sirva el que los coeficientes R^2, con los que se ha trabajado, han oscilado entre 0.1 y 0,5, aunque con los contrastes de t para cada regresión siempre ha resultado significativo el modelo planteado.

5. Apéndice

Lema (de KUDO (1961))

Sea X N_k [0, A]. Entonces:

$$P\,[X'A^{-1}X \geqslant t;\quad C_i' X \geqslant 0,\ i=1,\dots,p] =$$

$$= P\,[X'A^{-1}X \geqslant t]\ P\,[C_i' X \geqslant 0,\ i=1,\dots,p]$$

En donde C_i', representan transformaciones lineales.

Teorema 4

Vamos a demostrar que se verifica el teorema 4; en primer lugar consideramos la partición (2.8) y un valor t 0; podemos escribir:

$$P_{Ho}\,[F_{12} \geqslant t] = \Sigma\, P_{Ho}\,[F_{12} \geqslant t,\quad b \in A_s^{i_1 \dots i_s}], \qquad \text{(A.1)}$$

en donde el sumatorio se extiende a todos los conjuntos de la partición; y de la definición de F_{12} es claro que será cero si b pertenece al cono.

Consideremos un sumando de (A.1); es decir

$$b \in A_s^{i_1 \dots i_s} \Longleftrightarrow \begin{cases} b_i^* = b_{i+1}^* & i=i_1,\dots i_s \\ b_j^* < b_{j+1}^* & j \neq i_1,\dots i_s \end{cases} \qquad \text{(A.2)}$$

Consideremos la siguiente transformación:

$$\begin{array}{l} \beta_{i_1} - \beta_{i_1+1} = \theta_1 \\ \beta_{i_2} - \beta_{i_2+1} = \theta_2 \\ \dots\dots\dots\dots\dots\dots \\ \beta_{i_s} - \beta_{i_s+1} = \theta_s \\ \beta_j - \beta_{j+1} = \theta_{s+j}, \quad j \neq i_1 \dots i_s, \quad 1 = 1,\dots,k\text{-}1\text{-}s \qquad \text{(A.3)} \\ \qquad \beta_K = \theta_K. \end{array}$$

En forma matricial:

$$P\beta = \theta$$

Donde P es la matriz que da el cambio de los parámetros; es inmediato que es una transformación no singular. Por la propiedad del estimador máximo verosímil de ser invariante ante transformaciones lineales, en el modelo con los nuevos parámetros, y las nuevas restricciones, los estimadores encontrados son los transformados de los antiguos por la matriz P; y si cuando $\mathbf{b} \in A_s^{i_1 \dots i_s}$, $\mathbf{b}^*$ se obtiene al proyectar sobre el subespacio asociado a las restricciones de subíndices $i_1,\dots,i_s$; en el nuevo modelo θ^*, se obtiene al proyectar $\hat{\theta}$, sobre el subespacio definido por:

$$\theta_1 = 0, \qquad \theta_2 = 0, \dots, \theta_s = 0; \tag{A.4}$$

es decir, que el vector θ^*, satisface

$$\begin{array}{ll} \theta_i^* = 0, & i = 1, 2, \dots, s \\ \theta_i^* \quad 0, & i = s+1,\dots,k\text{-}1. \end{array} \tag{A.5}$$

Sobre las distribuciones, tenemos que

$$\mathbf{b} \sim N\,[\beta, \sigma^2 (X'X)^{-1}],$$

y por tanto:

$$\hat{\theta} \sim N\,[\theta, \sigma^2\, P(X'X)^{-1}\, P'].$$

Esta última matriz de varianzas covarianzas, es simétrica y definida positiva de rango k; existe una matriz triangular inferior T de rango k que verifica:

$$T\, P\, (X'X)^{-1}\, P'T' = I_k. \tag{A.6}$$

haciendo ahora el cambio

$$\tau = T\,\theta, \tag{A.7}$$

y llamando

$B = X\, P^{-1}\, T^{-1}$, que verifica

$$B'B = I_k, \tag{A.8}$$

el modelo sería

$$Y = B\,\tau + \varepsilon; \tag{A.9}$$

y las restricciones, de haber elegido T triangular, serán k-1 de la forma:

$$r_{i1}\,\tau_1 + r_{i2}\,\tau_2 + \dots r_{ii}\,\tau_i \lessgtr 0, \tag{A.10}$$

Nuevamente por la invarianza del estimador máximo verosímil frente a transformaciones lineales, en el nuevo modelo los estimadores serán los transformados por la matriz T; y el estimador con restricciones que llamaremos **z***, se obtiene al proyectar con la métrica euclídea el sin restricciones **z**; para el caso que estamos considerando, cuando **b** $\varepsilon\ A_s^{i_1 \dots i_s}$, z* se obtiene al proyectar **z** sobre el subespacio definido por

$$\tau_1 = 0, \quad \tau_2 = 0 \quad \dots \quad \tau_s = 0. \tag{A.11}$$

Las coordenadas de **z*** serán entonces de la forma:

$$\mathbf{z}^* = (0, \dots, 0, z_{s+1}, \dots, z_k)'. \tag{A.12}$$

y las condiciones que teníamos en (A.2) y en (A.5) que definían una región del espacio con el estimador máximo verosimil, son ahora de la forma:

$$z_i > 0, \qquad i = 1, \dots, s$$

$$R_j P^{-1}\, T^{-1}\, (0,\dots 0\ z_{s+1},\dots z_k)' < 0,\ j = s+1,\dots,k\text{-}1 \tag{A.13}$$

Estas condiciones podemos escribirlas como k-1 restricciones lineales sobre el vector z, que denominaremos C_i, y donde las s primeras serán filas con un uno y el resto cero, y las restantes tendrán siempre los s primeros elementos nulos.

Las hipótesis nula para los nuevos parámetros, se traduce en que los componentes de τ, son todas nulas excepto la última; **z** será, por tanto, una variable aleatoria cuya distribución será

$$\mathbf{z} \sim N\,[\tau,\ \sigma^2\, I], \tag{A.14}$$

que en la hipótesis nula, el vector de medias tendrá todas sus componentes nulas excepto la última.

Podemos escribir

$$P_{Ho}\,[F_{12} \geqslant t, \qquad \mathbf{b}\ \varepsilon\ A_s^{i_1 \dots i_s}]$$

$$= P_{Ho}\ \frac{(\mathbf{z} - \mathbf{z}^*)'(\mathbf{z} - \mathbf{z}^*)}{(Y - B\,\mathbf{z})'(Y - B\,\mathbf{z})} \geqslant t,\ C_i\,\mathbf{z} \geqslant 0] \tag{A.15}$$

Por el lema enunciado en la primera parte del apéndice, podemos descomponer esta probabilidad en forma de producto; y el estadístico de (A.15) en la hipótesis nula vemos que el numerador es suma de s variables aleatorias normales de media cero e independientes. La distribución del denominador es un resultado bien conocido, es también una chi cuadrado. Comprobemos que son independientes; podemos escribir:

$$
\begin{aligned}
& (z - z^*)'(z - z^*) = \\
= \; & (z - \tau)' \, I_s \, (z - \tau) = \\
= \; & \varepsilon \, 'B \; I \; B' \, \varepsilon. && \text{(A.16)}
\end{aligned}
$$

$$
\begin{aligned}
& (Y - B\,z)'(Y - B\,z) = \\
= \; & \varepsilon' \, (I - BB') \, \varepsilon. && \text{(A.17)}
\end{aligned}
$$

las matrices en (A.16) y (A.17) son idempotentes y su producto es nulo. Por lo tanto, son independientes y queda probado el resultado del teorema, ya que sale la misma distribución del estadístico siempre que el conjunto $A_s^{i_1 \cdots i_s}$, tenga el mismo número de restricciones. Es a la suma de probabilidades de conjuntos de este tipo a lo que llamamos P_{Ho} (s,k).

6. Bibliografía

BARLOW, R.E., BARTHOLOMEW, D. J. BREMNER, J. M., and BRUNK, H. D. (1972): "Statistical Inference under Order Restrictions". Ed. Wiley.

BORGE, L. (1985): "Estimación y Contrastes de Hipótesis en el Modelo Lineal General con Restricciones de Desigualdad". Tesis Doctoral. Universidad de Valladolid.

BOOT, J. C. G. (1968): "Programmation quadratique". Ed. Dunod.

KENNEDY, Ch. and THIRLWALL, A. P. (1972): "Survey in applied Economics: Technical Progress". Economic Journal, Marzo, pp. 11-72.

KUDO, A. (1961): "A multivariate analogue of the one side test". Biometrica, 50, pp. 403-418.

LAFUENTE, A. SALAS, V. y YAGUE, M. J. (1985): "Productividad, Capital Tecnológico e Investigación en la Economía Española". Servicio de Publicaciones del Ministerio de Industria y Energía.

ROBERTSON, T. and WEGMAN, E. J. (1978): "Likelihood Ratio Test for Order Restrictions in Exponential Families". Annals of Statistics, 6 pp. 485-505.