ESTADISTICA ESPAÑOLA
núm. 116, 1988, págs. 75 a 85

EL CRITERIO AVE: UNA VERSION BAYESIANA

por
MARIA TERESA APARICIO ASPAS *
Departamento de Análisis Económico
Facultad de Ciencias Económicas y Empresariales.
Universidad de Zaragoza.

RESUMEN

Se trata de analizar desde una óptica bayesiana el criterio AVE de selección de modelos, cuando el objetivo es la predicción del valor esperado de la variable endógena. Con posterioridad se determina un conjunto suficiente de condiciones bajo las cuales ambos enfoques (bayesiano y el denominado como de "línea intermedia") dan lugar a idénticas conclusiones.

Palabras clave: Selección de modelos, función de pérdida, riesgo Bayes, distribución a priori informativa.

Clasificación AMS: 90A19, 62F15

1. INTRODUCCION

La "línea intermedia", tal y como se establece en APARICIO (1985), configura un proceso de selección de modelos basado en una premisa inicial según la cual todos los modelos pueden ser tomados como correctos, en el sentido de ser "a priori" adecuados para la consecución del fin establecido. O lo que, análogamente, AZNAR (1986)

* Trabajo realizado dentro del Proyecto CAYCIT PB85-0339. Agradezco las sugerencias y comentarios de Don Antonio Aznar y de un evaluador anónimo que han contribuído a mejorar el trabajo.

denomina como "situarse en una estrategia de contraste de hipótesis". De acuerdo con esto, y considerando una base informativa constituída por una función de pérdida e información muestral, se establece la estrategia de selección entre especificaciones alternativas concretada en adoptar aquél modelo con menor riesgo estimado.

Dentro de este marco de actuación expuesto cabe encuadrar al criterio PC (AMEMIYA (1980)), al criterio AIC (AKAIKE (1973)) y al criterio AVE (AZNAR (1984)).

El objeto de esta investigación es la fundamentación bayesiana del criterio AVE de selección de modelos. Se trata, por tanto, de establecer el conjunto de condiciones bajo las cuales el criterio mencionado es el resultado de un proceso de decisión bayesiano. A esta línea de trabajo –y en relación con el criterio AIC– responden los desarrollos de ZELLNER (1978) y de SMITH y SPIEGELHALTER (1980). No obstante, el tratamiento que en ellos se recoge adolece, desde mi punto de vista, de un defecto fundamental como es el de obviar el marco en el cual AKAIKE (1973) define su criterio, es decir, no se presta atención a la función de pérdida establecida por el autor y se plantea un esquema de razonamiento basado en una estructura de pérdidas diferente.

Un enfoque más coherente pasaría por respetar la función de pérdida establecida por el criterio correspondiente y articular su tratamiento dentro de un contexto bayesiano. Esta estrategia es la adoptada en esta investigación en relación con el criterio AVE. En este sentido, el apartado 2 presenta la versión bayesiana de este criterio en conformidad con la línea de actuación descrita y el apartado 3 particulariza el estudio al supuesto de asumir una distribución de probabilidad "a priori" de los parámetros informativa. En la última sección se exponen las principales conclusiones de estos desarrollos.

2. VERSION BAYESIANA DEL CRITERIO AVE

Supongamos un conjunto de modelos potenciales dado por:

$$\begin{aligned} M_1&: Y = X_1 \beta_1 + U_1 \qquad U_1 \sim N(0, \sigma_1^2 I) \\ M_2&: Y = X_2 \beta_2 + U_2 \qquad U_2 \sim N(0, \sigma_2^2 I) \end{aligned} \tag{2.1}$$

donde X_1 y X_2 son matrices de orden $T \cdot K_1$ y $T \cdot K_2$ respectivamente. En principio, los modelos de (2.1) no tienen porqué constituir modelos anidados.

Aún cuando todo el desarrollo posterior se podría extender, sin excesivas complicaciones, al caso de suponer una matriz de varianzas y covarianzas no escalar para el vector de perturbaciones, el hecho de considerar una matriz escalar responde a una estrategia global sobre como debería abordarse el proceso de selección de modelos (una amplia discusión sobre este punto puede verse en AZNAR (1986)). Desde esta perspectiva, la etapa previa a resolver el problema de elección entre especificaciones alternativas consistirá en determinar el conjunto de "modelos esféricos". Se entiende por "modelo esférico" aquel que presenta una diferenciación clara entre parte sistemática y parte aleatoria. De este modo, estos modelos pasarán a configurarse como el conjunto de especificaciones potenciales o estados de la naturaleza del problema de decisión, tal como se trata de expresar en (2.1).

El objetivo es seleccionar aquélla especificación más adecuada para predecir el valor esperado de la variable endógena, $E(Y_o)$, dados los valores futuros de las variables explicativas, x'_{oi} (i = 1, 2). El conjunto de información sobre el que se fundamenta la estrategia bayesiana está constituido por:

a) información a priori sobre los modelos, sintetizada en la distribución a priori $P(M_i)$ (i = 1, 2).

b) información a priori sobre los parámetros de los modelos explicitada a través de la distribución a priori $P(\theta_i / M_i)$ (i = 1, 2); siendo $\theta_i = (\theta_i, \sigma_i^2)$.

c) información muestral.

d) función de pérdida.

En concreto, el criterio AVE asume como función de pérdida para el problema de predicción planteado (véase AZNAR (1984)):

$$L(d_j, M_i, \theta_i) = L_{ji}(\hat{Y}_{oi} - E_i Y_o)^2 \quad (i, j = 1, 2) \tag{2.2}$$

siendo $\hat{Y}_{oi}$ la predicción de $E(Y_o)$ utilizando el modelo *i-ésimo*.

En la expresión anterior, E_i denota la esperanza respecto al modelo que se está considerando cada vez como correcto y d hace referencia al conjunto de decisiones posibles, es decir, d_j = predecir adoptando el modelo *j-ésimo*.

La función de pérdida establecida por el criterio AVE consta de dos elementos:

– L_{ji}, que trata de medir la importancia asignada a adoptar un modelo distinto del considerado como probable en cada caso.

– $(\hat{Y}_{oi} - E_i Y_o)$ $(\hat{Y}_{oi} - E_i Y_o)'$, que intenta medir la "distancia" entre la predicción y el valor futuro de la variable endógena a estimar.

Además, se asume que L_{ji} es la misma cantidad constante en el caso de que $i=j$ y que para el resto ($i \neq j$) las L_{ji} son grandes en relación con L_{ii}, (véase a este respecto APARICIO (1985) y AZNAR (1986)).

Esquemáticamente, la función de pérdida puede resumirse como:

Modelo verdadero Decisiones	M_1	M_2
d_1	$L_{11}(x'_{o1}\hat{\beta}_1 - x'_{o1}\beta_1)(x'_{o1}\hat{\beta}_1 - x'_{o1}\beta_1)'$	$L_{12}(x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)(x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)'$
d_2	$L_{21}(x'_{o2}\hat{\beta}_2 - x'_{o1}\beta_1)(x'_{o2}\hat{\beta}_2 - x'_{o1}\beta_1)'$	$L_{22}(x'_{o2}\hat{\beta}_2 - x'_{o2}\beta_2) \cdot (x'_{o2}\hat{\beta}_2 - x'_{o2}\beta_2)'$

CUADRO I

La estrategia de actuación bayesiana ante este planteamiento consiste en tomar aquella decisión que minimiza el riesgo Bayes.

La expresión general del riesgo Bayes viene dada por:

$$R.B.(d) = \sum_{M} \int_{S} \int_{\theta} f(Y/\theta)\, L(d, M, \theta)\, P(\theta/M)\, P(M)\, dY\, d\theta \qquad (2.3)$$

o, equivalentemente:

$$R.B.(d) = \sum_{M} P(M) \int_{\theta} P(\theta/M) \left[\int_{S} f(Y/\theta)\, L\,(d, M, \theta)\, dY \right] d\theta \rightarrow$$

$$\rightarrow \sum_{M} P(M) \int_{\theta} P(\theta/M)\, R\,(d, M, \theta)\, d\theta$$

siendo S el espacio de todas las muestras y $R(d, M, \theta)$ denota la función de riesgo clásico.

Teniendo en cuenta que:

$$\int_{\theta} P(\theta/M)\, R(d, M, \theta)\, d\theta = E_{\theta/M}\, R(d, M, \theta) \qquad (2.4)$$

donde (2.4) indica la esperanza del riesgo respecto de la distribución a priori de los parámetros, podemos concluir:

$$R.B.(d) = \sum_{M} P(M)\, E_{\theta/M}\, R(d, M, \theta) \qquad (2.5)$$

La concreción de la estrategia bayesiana, expuesta en (2.5), al planteamiento recogido en el cuadro I, permite establecer:

$$R.B.(d_1) = P(M_1)\, E_{\theta_1/M_1} \{ E\, [L_{11}\, (x'_{o1}\hat{\beta}_1 - x'_{o1}\beta_1)\, (x'_{o1}\hat{\beta}_1 - x'_{o1}\beta_1)'\,] \} +$$

$$+ P(M_2)\, E_{\theta_2/M_2} \{ E\, [L_{12}\, (x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)\, (x'_{01}\hat{\beta}_1 - x'_{o2}\beta_2)'\,] \}$$

Suponiendo σ_i^2 conocida se obtiene:

$$R.B.(d_1) = L_{11}P(M_1)E_{\beta_1/M_1}\{E(x'_{o1}\hat{\beta}_1 - x'_{o1}\beta_1)(x'_{o1}\hat{\beta}_1 - x'_{o1}\beta_1)'\} +$$

$$+ L_{12}P(M_2)\, E_{\beta_2/M_2}\{E\,(x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)(x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)'\} \tag{2.6}$$

Análogamente:

$$R.B.(d_2) = L_{21}P(M_1)E_{\beta_1/M_1}\{E(x'_{o2}\hat{\beta}_2 - x'_{o1}\beta_1)(x'_{o2}\hat{\beta}_2 - x'_{o1}\beta_1)'\} +$$

$$+ L_{22}P(M_2)\, E_{\beta_2/M_2}\{E\,(x'_{o2}\hat{\beta}_2 - x'_{o2}\beta_2)(x'_{o2}\hat{\beta}_2 - x'_{o2}\beta_2)'\} \tag{2.7}$$

donde E denota la esperanza respecto de la información muestral.

Calculando cada uno de los elementos de (2.6) por separado resulta:

1) $E_{\beta_2/M_2}\{E(x'_{o1}\hat{\beta}_1 - x'_{o1}\beta_1)(x'_{o1}\hat{\beta}_1 - x'_{o1}\beta_1)'\} =$

$= E_{\beta_1/M_1}\{x'_{o1}E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_1 - \beta_1)'\, x_{o1}\} = E_{\beta_1/M_1}[\, x'_{o1}V(\hat{\beta}_1)\, x_{o1}\,] =$

$$= x'_{o1}\, V(\hat{\beta}_1)\, x_{o1} \tag{2.8}$$

donde $V(\hat{\beta}_1)$ denota la matriz de varianzas y covarianzas del vector de estimadores $\hat{\beta}_1$.

2) $E_{\beta_2/\mu_2}\{E(x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)(x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)'\} =$

$= x'_{o1}E(\hat{\beta}_1 - \bar{\beta}_1)(\hat{\beta}_1 - \bar{\beta}_1)'\, x_{o1} + x'_{o2}V(\beta_2\, /\, M_2)\, x_{o2}$ -

$$- 2x'_{o1}E(\hat{\beta}_1 - \bar{\beta}_1)\,[\bar{\beta}'_2 x_{o2} - \bar{\beta}'_1 x_{o1}] + (x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)(x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)' \tag{2.9}$$

siendo $\bar{\beta}_i$ y $V(\beta_i\, /\, M_i)$ $(i = 1, 2)$ la correspondiente media y matriz de varianzas y covarianzas de la distribución a priori asumida para $\beta_i (i = 1, 2)$.

La demostración del resultado (2.9) es recogida en el Apéndice.

Llevando a cabo el mismo razonamiento para los elementos individuales de (2.7) se obtiene:

3) $E_{\beta_1/M_1}\{E(x'_{o2}\hat{\beta}_2 - x'_{o1}\beta_1)(x'_{o2}\hat{\beta}_2 - x'_{o1}\beta_1)'\} =$

$= x'_{o2}E(\hat{\beta}_2 - \bar{\beta}_2)(\hat{\beta}_2 - \bar{\beta}_2)'\, x_{o2} + x'_{o1}V(\beta_1\, /\, M_1)\, x_{o1} +$

$+ 2x'_{o2}E(\hat{\beta}_2 - \bar{\beta}_2)\,[\bar{\beta}'_2\, x_{o2} - \bar{\beta}'_1\, x_{o1}] +$

$$+ (x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)(x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)' \tag{2.10}$$

4) $$E_{\beta_2/M_2}\{ E(x'_{o2}\hat{\beta}_2 - x'_{o2}\beta_2)(x'_{o2}\hat{\beta}_2 - x'_{o2}\beta_2)' \} =$$

$$= x'_{o2} V(\hat{\beta}_2)\, x_{o2} \tag{2.11}$$

La sustitución de estos resultados parciales en (2.6) y (2.7) lleva a establecer:

$$\begin{aligned} R.B.(d_1) = {} & L_{11}P(M_1)\,[x'_{o1}V(\hat{\beta}_1)\,x_{o1}] + L_{12}P(M_2)\{ x'_{o1}E(\hat{\beta}_1-\bar{\beta}_1)(\hat{\beta}_1-\bar{\beta}_1)' \\ & x_{o1} + x'_{o2}V(\beta_2 / M_2)\,x_{o2} - 2x'_{o1}E(\hat{\beta}_1-\bar{\beta}_1)\,[\bar{\beta}'_2 x_{o2} - \bar{\beta}'_1 x_{o1}] + \\ & + (x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)(x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)' \} \end{aligned} \tag{2.12}$$

$$\begin{aligned} R.B.(d_2) = {} & L_{21}P(M_1)\{ x'_{o2}E(\hat{\beta}_2-\bar{\beta}_2)(\hat{\beta}_2-\bar{\beta}_2)'\,x_{o2} + x'_{o1}V(\beta_1 / M_1)\,x_{o1} + \\ & + 2x'_{o2}E(\hat{\beta}_2-\bar{\beta}_2)\,[\bar{\beta}'_2 x_{o2} - \bar{\beta}'_1 x_{o1}] + \\ & + (x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)(x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)' \} + L_{22}P(M_2)\{ x'_{o2}V(\hat{\beta}_2)\,x_{o2} \} \end{aligned} \tag{2.13}$$

La decisión adoptada será, por tanto, aquella con menor riesgo Bayes. En este sentido, se tomaría la decisión d_1 (elegir el modelo M_1) si y sólo si:

$$R.B.(d_1) < R.B.(d_2)$$

lo cual, bajo el supuesto de idénticas probabilidades a priori para los modelos [$P(M_1) = P(M_2) = P(M)$] se traduce en el cumplimiento de:

$$\begin{aligned} & P(M)\{ L_{11}\, x'_{o1}V(\hat{\beta}_1)\,x_{o1} - L_{22}\, x'_{o2}V(\hat{\beta}_2)\,x_{o2} \} < \\ & < P(M)\{ L_{21}\, x'_{o2}E(\hat{\beta}_2-\bar{\beta}_2)(\hat{\beta}_2-\bar{\beta}_2)'\,x_{o2} - L_{12}\, x'_{o1}E(\hat{\beta}_1-\bar{\beta}_1)(\hat{\beta}_1-\bar{\beta}_1)'\,x_{o1} \\ & + L_{21}\, x'_{o1}V(\beta_1 / M_1)\,x_{o1} - L_{12}\, x'_{o2}V(\beta_2 / M_2)\,x_{o2} + (\bar{\beta}'_2 x_{o2} - \bar{\beta}'_1 x_{o1}) \\ & \quad [2L_{21}\, x'_{o2}E(\hat{\beta}_2-\bar{\beta}_2) + 2L_{12}\, x'_{o1}E(\hat{\beta}_1-\bar{\beta}_1)\,] + \\ & + (x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)(x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)'\,[L_{21} - L_{12}] \}^{1} \end{aligned} \tag{2.14}$$

Siendo, por tanto, (2.14) la versión bayesiana del criterio AVE para la selección del modelo más adecuado al objeto de predecir el valor esperado de la variable endógena en un periodo futuro.

---

(1) Dadas las características asumidas para la función de pérdida (2.2), el primer miembro de la desigualdad (2.14) se puede expresar equivalentemente como

$$P(M)\{ L_{11}\,[x'_{o1}V(\hat{\beta}_1)\,x_{o1} - x'_{o2}V(\hat{\beta}_2)\,x_{o2}] \}$$

3. ANALISIS CON UNA DISTRIBUCION A PRIORI INFORMATIVA

La función de distribución de probabilidad $P(\theta / M)$ representa nuestro conocimiento a priori acerca de los parámetros del modelo. Cuando aquél es escaso podemos asumir como distribución a priori que sintetice tal grado de información (suponiendo el conocimiento de σ^2, la información previa sólo hará referencia al vector de parámetros de posición) a:

$$p(\beta_i / \sigma_i^2) \propto c \qquad (i = 1, 2) \tag{3.1}$$

donde c es un vector de constantes.

La combinación de (3.1) con la función de densidad conjunta de la muestra, que se supone normal, permite obtener la función de distribución a posteriori:

$$p(\beta_i / \sigma_i^2 Y_i) = \frac{| X_i' X_i |^{1/2}}{(2\pi\sigma_i^2)^{K_i/2}} \exp\left\{ - \frac{1}{2\sigma_i^2} (\beta_i - \hat{\beta}_i)' X_i' X_i (\beta_i - \hat{\beta}_i) \right\}$$

$$(i = 1, 2) \tag{3.2}$$

una vez introducida la constante apropiada para la normalización. De este modo, (3.2) está establecido que la distribución a posteriori de β_i es una normal multivariante con media el vector de estimadores m.c.o. y con matriz de varianzas y covarianzas $\sigma_i^2(X_i' X_i)^{-1}$.

En tanto en cuanto nuestro objetivo es la predicción del valor esperado de la variable endógena estamos centrando el análisis en un nuevo conjunto de información muestral generada por el mismo esquema de comportamiento. Para estos nuevos datos, puede considerarse a (3.2) como la distribución a priori de β_i, dado σ_i^2, lo que implica establecer para la muestra de valores futuros (véase ZELLNER (1971)) que:

$$p(\beta_i / \sigma_i^2) \sim N_K [\hat{\beta}_i, \sigma_i^2(X_i' X_i)^{-1}] \qquad (i = 1, 2) \tag{3.3}$$

Esta distribución a priori depende de la información muestral, via la matriz de segundos momentos de las variables explicativas. De este modo, responde al planteamiento de KLEIN Y BROW (1984) que propugna la necesidad de tal dependencia al objeto de poder asegurar que la distribución a posteriori es invariante ante transformaciones lineales de los datos.

De acuerdo con la distribución a priori informativa (3.3), la versión bayesiana del criterio AVE, dada en (2.14), se concreta en seleccionar el modelo M_1 si:

$$P(M) \{ L_{11}\, x_{01}' V(\hat{\beta}_1)\, x_{01} - L_{22}\, x_{02}' V(\hat{\beta}_2)\, x_{02} \} < P(M) \{ L_{21}\, \sigma_1^2 (X_1' X_1)^{-1} x_{01} -$$

$$- L_{12}\, \sigma_2^2 x_{02}'(X_2' X_2)^{-1} x_{02} + (x_{02}'\hat{\beta}_2 - x_{01}'\hat{\beta}_1)(x_{02}'\hat{\beta}_2 - x_{01}'\hat{\beta}_1)' [L_{21} - L_{12}] \} \tag{3.4}$$

Considerando que la magnitud constante (L_{ji}) de la función de pérdida establecida es la misma para todas las decisiones incorrectas, por lo que $L_{12} = L_{21}$, (3.4) equivale a:

$$P(M) \{ L_{11} x'_{01} V(\hat{\beta}_1) x_{01} - L_{22} x'_{02} V(\hat{\beta}_2) x_{02} \} <$$

$$< P(M) \{ L_{21} \sigma_1^2 x'_{01} (X'_1 X_1)^{-1} x_{01} - L_{12} \sigma_2^2 x'_{02} (X'_2 X_2)^{-1} x_{02} \} \Rightarrow$$

$$\Rightarrow P(M) \{ L_{11} - L_{21}) \sigma_1^2 x'_{01} (X'_1 X_1)^{-1} x_{01} - (L_{22} - L_{12}) \sigma_2^2 \; x'_{02} (X'_2 X_2)^{-1} x_{02} \} < 0$$

o equivalentemente:

$$P(M) (L_{11} - L_{21}) \sigma_1^2 x'_{01} (X'_1 X_1)^{-1} x_{01} < P(M) (L_{22} - L_{12}) \sigma_2^2 x'_{02} (X'_2 X_2)^{-1} x_{02} \Rightarrow$$

$$\Rightarrow \sigma_1^2 x'_{01} (X'_1 X_1)^{-1} x_{01} < \sigma_2^2 x'_{02} (X'_2 X_2)^{-1} x_{02} \qquad (3.5)$$

La regla de actuación (3.5), derivada bajo el planteamiento bayesiano, implica que el modelo más adecuado para predecir el valor esperado de la variable dependiente será aquél que minimice la varianza del predictor. Este es, precisamente, el resultado obtenido por el criterio AVE (AZNAR (1984)) cuando este se deriva de acuerdo con una base informativa que sólo considera información muestral y una función de pérdida.

Puede observarse asimismo, a partir de (3.5), que la acción o decisión óptima es invariante ante transformaciones lineales de la función de pérdida lo cual es siempre una propiedad deseable de un criterio de decisión.

El desarrollo efectuado ha permitido establecer los supuestos bajo los cuales las dos versiones del criterio AVE, la bayesiana y la correspondiente a la denominada "línea intermedia" (2), resultan equivalentes; supuestos que se concretan en:

1.– Idénticas probabilidades a priori para los modelos.

2.– Distribución a priori informativa sobre los parámetros de posición del modelo sintetizada en $N_{K_i}(\hat{\beta}_i, \sigma_i^2 (X'_i X_i)^{-1})$.

3.– Igualdad entre las constantes L_{ji} ($i \neq j$) que integran la función de pérdida establecida.

(2) Corriente metodológica que trata de conjugar los aspectos más relevantes de las dos aproximaciones tradicionales al tema de la selección de modelos: clásica y bayesiana (véase, al respecto, APARICIO (1985)).

4. CONCLUSIONES

El análisis del criterio AVE desde una óptica bayesiana, adoptando una estrategia de razonamiento basada en respetar la función de pérdida establecida por este método de selección, ha posibilitado presentar, con carácter general, la versión bayesiana del mismo. A partir de la regla de actuación (2.14), la consideración específica de los elementos bayesianos, en particular de la distribución a priori para los parámetros del modelo, da lugar a la expresión concreta del criterio de selección.

Una posible situación es aquella en que la información previa sobre β (todo el desarrollo se articula suponiendo conocido σ^2) se sintetiza a través de una distribución normal con media el vector de estimadores M.C.O. y matriz de varianzas y covarianzas $\sigma^2(X'X)^{-1}$. En tal caso, se demuestra como la estrategia de decisión resultante equivale a la establecida por el criterio AVE, cuando este se deriva a partir de una base informativa que combina sólo la información muestral y la función de pérdida.

De este modo, el método AVE queda formalmente justificado dentro de una estrategia bayesiana que asuma, entre otros supuestos, una distribución a priori informativa sobre los parámetros como la mencionada anteriormente.

5. APENDICE

Obtención de la expresión (2.9).

$$E_{\beta_2/M_2}\{ E(x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)(x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)' \} \tag{A.1}$$

Sumando y restando, en ambos paréntesis, los elementos $x'_{o1}\bar{\beta}_1$ y $x'_{o2}\bar{\beta}_2$, la expresión (A.1) es equivalente a:

$$E_{\beta_2/M_2}\{ E\,[x'_{o1}(\hat{\beta}_1-\bar{\beta}_1) - x'_{o2}(\beta_2-\bar{\beta}_2) + x'_{o1}\bar{\beta}_1 - x'_{o2}\bar{\beta}_2] \cdot$$

$$\cdot\,[x'_{o1}(\hat{\beta}_1-\bar{\beta}_1) - x'_{o2}(\beta_2-\bar{\beta}_2) + x'_{o1}\bar{\beta}_1 - x'_{o2}\bar{\beta}_2]' \} \Rightarrow$$

$$E_{\beta_2/M_2}\{ E\,[x'_{o1}(\hat{\beta}_1-\bar{\beta}_1)(\hat{\beta}_1-\bar{\beta}_1)'\,x_{o1} + x'_{o2}(\beta_2-\bar{\beta}_2)(\beta_2-\bar{\beta}_2)'\,x_{o2} +$$

$$+ x'_{o1}\bar{\beta}_1\bar{\beta}'_1\,x_{o1} + x'_{o2}\bar{\beta}_2\bar{\beta}'_2\,x_{o2} - 2x'_{o1}(\hat{\beta}_1-\bar{\beta}_1)(\beta_2-\bar{\beta}_2)'\,x_{o2} +$$

$$+ 2x'_{o1}(\hat{\beta}_1-\bar{\beta}_1)\bar{\beta}'_1\,x_{o1} - 2x'_{o1}(\hat{\beta}_1-\bar{\beta}_1)\bar{\beta}'_2\,x_{o2} - 2x'_{o2}(\beta_2-\bar{\beta}_2)\bar{\beta}'_1\,x_{o1} +$$

$$+ 2x'_{o2}(\beta_2-\bar{\beta}_2)\bar{\beta}'_2\,x_{o2} - 2x'_{o1}\bar{\beta}_1\bar{\beta}'_2\,x_{o2}\,] \}$$

Tomando esperanzas respecto a la distribución a priori de β_2 resulta:

$$x'_{o1}E(\hat{\beta}_1-\bar{\beta}_1)(\hat{\beta}_1-\bar{\beta}_1)'x_{o1} + x'_{o2}V(\beta_2 / M_2)x_{o2} + x'_{o1}\bar{\beta}_1\bar{\beta}'_1 x_{o1} +$$

$$+ x'_{o2}\bar{\beta}_2\bar{\beta}'_2 x_{o2} + 2x'_{o1}E(\hat{\beta}_1-\bar{\beta}_1)\bar{\beta}'_1 x_{o1} - 2x'_{o1}E(\hat{\beta}_1-\bar{\beta}_1)\bar{\beta}'_2 x_{o2} - 2x'_{o1}\bar{\beta}_1\bar{\beta}'_2 x_{o2}$$

Por tanto:

$$E_{\beta_2/M_2}\{E(x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)(x'_{o1}\hat{\beta}_1 - x'_{o2}\beta_2)'\} =$$

$$= x'_{o1}E(\hat{\beta}_1-\bar{\beta}_1)(\hat{\beta}_1-\bar{\beta}_1)'x_{o1} + x'_{o2}V(\beta_2 / M_2)x_{o2} - 2x'_{o1}E(\hat{\beta}_1-\bar{\beta}_1)$$

$$[\bar{\beta}'_2 x_{o2} - \bar{\beta}'_1 x_{o1}] + (x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)(x'_{o2}\bar{\beta}_2 - x'_{o1}\bar{\beta}_1)'$$

que es, precisamente, la expresión (2.9).

BIBLIOGRAFIA

AKAIKE H. (1973). "Information Theory and an Extensión of the Maximun Likelihood Principle". En *2nd International Symposium on Information Theory*, Ed. B. N. PETROV y F. CSAKY. Akademiai Kiado.

AMEMIYA T. (1980). "Selection of Regressors". *International Economic Review*. 21, 331-354.

APARICIO M. T. (1985). "Selección de Modelos Econométricos: Estudio comparado de un nuevo criterio". *Tesis Doctoral*. Departamento de Econometría. Facultad de Ciencias Economicas y Empresariales. Universidad de Zaragoza.

AZNAR A. (1984). "Buscando el Modelo Econométrico Util". *Estadística Española* 102, 5-11.

AZNAR A. (1986). "Econometric Model Selection: A New Approach". *Manuscrito*. Departamento Econometría. Facultad de Ciencias Económicas y Empresariales. Universidad de Zaragoza.

KLEIN R. W. y S. J. BROWN (1984). "Model Selection When There is «Minimal» Prior Information". *Econometrica* 52, 1291-1312.

SMITH A. F. M. y D. J. SPIEGELHALTER (1980). "Bayes Factor and Choice Criteria for Linear Models". *Journal of the Royal Statistical Society, serie B* 42, 213-220.

ZELLNER A. (1971). *"An Introduction to Bayesian Inference in Econometrics"*. New York: John Wiley and Sons.

ZELLNER A. (1978). "Jeffreys-Bayes Posterior Odds Ratio and the Akaike Information Criterion for Discriminating Between Models". *Economics Letters* 1, 337-342.

THE AVE CRITERIUM: A BAYES VERSION

SUMMARY

The point is to analyse from a Bayes approach, the models selection AVE criterium, when the purpose is the prediction of the expected value of the endogenous variable. After this, a sufficient set of conditions is determined, under which both approaches (the Bayes one and that known as the "intermediate line") give rise to identical conclusions.

Key words: Models selection, loss function, Bayes risk, a priori informative distribution.