COMPORTAMIENTO PARA MUESTRAS PEQUEÑAS DE UN CRITERIO DE SELECCIÓN CONSISTENTE*

M.ª Isabel Ayuda
Universidad de Zaragoza
Carmen García-Olaverri
Universidad Pública de Navarra

En este artículo se analiza el comportamiento en muestras pequeñas de un criterio de selección consistente, el criterio BEC propuesto por Geweke y Meese (1981). El análisis se lleva a cabo en un contexto de selección entre modelos anidados donde se considera como proceso generador de datos el modelo más amplio de los que se comparan. En el trabajo se determinan las situaciones bajo las que este criterio tiende a seleccionar modelos más restringidos que el proceso generador de datos (PGD), centrándonos principalmente en el comportamiento del criterio ante la proximidad entre los modelos comparados y ante la presencia de multicolinealidad entre los regresores del PGD. A través de un Estudio de Monte Carlo se estudia el comportamiento del criterio BEC en muestras pequeñas y se comprueba cómo, por ser consistente, tiende a seleccionar el 100% de las veces el PGD cuando este tamaño muestral aumenta y que debido a esta consistencia tiende a seleccionar generalmente el PGD independientemente del grado de multicolinealidad entre los regresores de éste y de la proximidad entre los modelos que se comparan.

Palabras clave: selección, BEC, SBIC, consistencia, Monte Carlo.

1. INTRODUCCIÓN

Un criterio de selección es, en general, una función C de ciertos estadísticos del modelo cuyo objetivo es seleccionar aquel modelo M_j que minimize $C(M_j)$. En la literatura sobre selección de modelos son varios los criterios que se han propuesto, y a su vez, éstos se han clasificado de muy distintas formas.

(*) Este trabajo ha sido financiado por el proyecto de investigación DGICYT PB94-0602. Una versión previa fue presentada en el XXII Congreso de Estadística e Investigación Operativa celebrado en Sevilla del 14 al 17 de noviembre de 1995. Agradecemos la colaboración y comentarios de Antonio Aznar en la elaboración de este artículo.

Una de las clasificaciones que se han hecho, y que destacamos, es entre criterios preferencialistas y verificacionistas. El interés en esta clasificación está basado en que entendemos que la calidad de un modelo no es algo inherente a cada especificación, sino que depende del objetivo para el que se ha diseñado. Es decir, un modelo puede ser el que mejor explique el comportamiento de ciertas observaciones muestrales, pero no tiene por qué ser el que proporcione mejores predicciones. Por ello, la clasificación que destacamos es la que distingue entre criterios que tienen en cuenta el uso posterior que se le va a dar al modelo (criterios preferencialistas) y los que no lo tienen en cuenta (criterios verificacionistas).

Dentro de los criterios verificacionistas lo que se pretende es buscar el modelo que mejor se adapta a los datos sin tener en cuenta, explícitamente, el uso futuro que se le vaya a dar al modelo seleccionado.

Los criterios preferencialistas tienen como objetivo seleccionar el *modelo útil*, entendiendo por *modelo útil* aquel que cumple adecuadamente los objetivos para los que ha sido propuesto. Estos objetivos se expresan en las funciones de pérdida y de riesgo. La estimación de la función de riesgo es la que nos sirve para discriminar entre los diferentes modelos.

En la literatura han aparecido diferentes trabajos dedicados a comparar los criterios. Entre estos trabajos, cabe destacar los artículos de Geweke y Meese (1981, págs. 55-70), Mills y Prasad (1992, págs. 201-233), García-Olaverri (1993), Aznar y García-Olaverri (1995, págs. 217-238) y Ayuda (1994), éste último referente a modelos no anidados. Dos de los criterios que han recibido mayor atención, son el SBIC propuesto en Schwarz (1978) y el BEC propuesto en Geweke y Meese (1981)[1].

En este artículo estudiamos el comportamiento del criterio BEC, en muestras pequeñas, en un contexto de selección entre dos modelos anidados, donde consideraremos únicamente el caso en que el Proceso Generador de Datos es el modelo amplio, ya que del cuadro que presentamos en el apartado 2 deducimos que ante dos modelos anidados cuando el PGD es el modelo restringido se debería seleccionar el restringido, y cuando el PGD es el modelo amplio, caben las dos posibilidades, dependiendo de una serie de factores que analizaremos en el apartado siguiente. Por ello, estudiaremos bajo qué circunstancias el criterio BEC selecciona modelos anidados en el PGD.

Este artículo se organiza de la siguiente forma: la sección segunda está dedicada a explicar cuál es el marco de trabajo; el papel de la consistencia en un proceso de selección de modelos es estudiado en la sección tercera, dedicando la cuarta a un análisis de los rasgos del criterio BEC. Los apartados 5 y 6 presentan los resultados de un estudio de Monte Carlo y las conclusiones se recogen en el último apartado.

(1) En este artículo nos dedicamos a analizar el comportamiento del criterio BEC ya que a este se le ha prestado menos atención en la literatura que al criterio SBIC. Además el criterio SBIC presenta algunos problemas de análisis en el caso de muestras pequeñas, y su comportamiento es prácticamente idéntico al del criterio BEC.

2. Marco de trabajo

Supongamos dos modelos lineales anidados:

(1) $$M_{k1}: y = X_1\beta_1 + u_1$$

(2) $$M_{k2}: y = X_2\beta_2 + u_2 = X_1\beta_1 + X^*\beta^* + u_2$$

donde y es un vector T x 1 de observaciones de la variable dependiente; X_1 es una matriz T x k_1 de observaciones de los k_1 regresores; X_2 es otra matriz T x k_2 de observaciones de los k_2 regresores; β_1, β_2 y β^* son vectores de k_1, k_2 y k_2-k_1 parámetros, respectivamente, y u_1 y u_2 son vectores T x 1 de perturbaciones aleatorias. Cuando M_{k1} es el PGD, $u_1 \sim N(0,\sigma^2_{k1}I)$ y cuando M_{k2} es el PGD, $u_2 \sim N(0,\sigma^2_{k2}I)$.

Sea x_{p1} el vector de observaciones de los k_1 regresores de M_{k1} correspondientes a un período extramuestral, que llamamos p, y x_{p2} el vector de observaciones de los k_2 regresores incluidos en M_{k2}, correspondientes a una observación extramuestral p. Para el modelo M_{k1}, el predictor y el error de predicción son, respectivamente:

$$\hat{y}_{p1} = x'_{p1}\hat{\beta}_1 \quad y \quad e_{p1} = y_p - \hat{y}_{p1}$$

donde $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'$ y e y_p es la observación extramuestral de la variable endógena.

Para el modelo M_{k2}, el predictor y el error de predicción son:

$$\hat{y}_{p2} = x'_{p2}\hat{\beta}_2 \quad y \quad e_{p2} = y_p - \hat{y}_{p2} \quad \text{con } \hat{\beta}_2 = (X_2'X_2)^{-1}X_2'y$$

Suponemos que incluimos tanto en X_i como en y, todas las observaciones previas a p.

En los resultados presentados en el cuadro 1 se observa que cuando el PGD es el modelo restrindigo (M_{k1}) el predictor y el error de predicción en el modelo restringido son insesgados y de menor varianza que en el modelo amplio que también son insesgados. Por lo tanto, siempre que genera el modelo restringido, debería de seleccionarse éste ya que proporciona predictores y errores de predicción insesgados y de menor varianza que con el modelo amplio. En cambio, cuando el PGD es el modelo amplio, M_{k2}, el predictor y el error de predicción para el modelo restringido son sesgados, siendo insesgados en el modelo amplio. Las varianzas siguen siendo menores para el primero que para el segundo, por ello, dependiendo de la importancia que se le dé al sesgo y a la varianza, caben las dos posibilidades, seleccionar el modelo M_{k1} o el M_{k2}. Debido a que cuando genera el restringido no hay duda de que éste debería ser elegido, y que cuando genera el amplio cabe la posibilidad de seleccionar modelos más restringidos, en este trabajo nos dedicamos únicamente a analizar bajo qué condiciones el criterio BEC va a seleccionar modelos con menos regresores que el PGD.

3. Consistencia de criterios

Sean M_1, M_2, ..., M_k una serie de modelos que compiten entre sí; el criterio de selección C consiste en elegir uno de ellos, M_j, de forma que:

(3) $$C(M_j) = \min C(M_i) \qquad i=1, ..., k$$

Cuadro 1
PROPIEDADES DE LOS PREDICTORES Y ERRORES DE PREDICCIÓN MCO

	Predicciones usando M_{k1}		Predicciones usando M_{K2}
	$E_1\hat{y}_{p1}=E_1y_p$, $E_1e_{p1}=0$ INSESGADOS		$E_1\hat{y}_{p2}=E_1y_p$, $E_1e_{p2}=0$ INSESGADOS
PGD $=M_{k1}$	$Var_1(\hat{y}_{p1})=\sigma_1^2x'_{p1}(X'_1X_1)^{-1}x_{p1}$	$<$	$Var_1(\hat{y}_{p2})=\sigma_1^2x'_{p2}(X'_2X_2)^{-1}x_{p2}$
	$Var_1(e_{p1})=\sigma_1^2(1+x'_{p1}(X'_1X_1)^{-1}x_{p1})$	$<$	$Var_1(e_{p2})=\sigma_1^2(1+x'_{p2}(X'_2X_2)^{-1}x_{p2})$
	$E_2\hat{y}_{p1}\neq E_2y_p$, $E_2e_{p1}\neq 0$ SESGADOS		$E_2\hat{y}_{p2}=E_2y_p$, $E_2e_{p2}=0$ INSESGADOS
PGD $=M_{k2}$	$Var_2(\hat{y}_{p1})=\sigma_2^2x'_{p1}(X'_1X_1)^{-1}x_{p1}$	$<$	$Var_2(\hat{y}_{p2})=\sigma_2^2x'_{p2}(X'_2X_2)^{-1}x_{p2}$
	$Var_2(e_{p1})=\sigma_2^2(1+x'_{p1}(X'_1X_1)^{-1}x_{p1})$	$<$	$Var_2(e_{p2})=\sigma_2^2(1+x'_{p2}(X'_2X_2)^{-1}x_{p2})$

Decimos que un criterio C es consistente si a medida que aumenta el tamaño muestral crece su tendencia a elegir el Proceso Generador de Datos. Es decir, si M_s es el PGD, decimos que el criterio es consistente si

$$\lim_{T\to\infty}\left[\Pr\left(C(M_s)<C(M_i)\right)\right]=1 \qquad \forall i\neq s \tag{4}$$

En general, la propiedad de consistencia puede considerarse como una cualidad deseable para un criterio de selección de modelos, especialmente si el objetivo que se persigue con la selección es el de acercarnos al PGD. Sin embargo, desde un marco preferencialista, la propiedad de consistencia puede estar, como señala Aznar (1989), en franca contradicción con el propio proceso de derivación del criterio, pues un modelo construido para seleccionar el modelo más útil no tiene por qué tender a seleccionar el Proceso Generador de Datos.

El siguiente párrafo de Shibata (1983) ilustra la anterior idea, «... sin embargo, si se desea tener un buen estimador o predictor, no es muy sabio insistir en la consistencia del procedimiento de selección. La inconsistencia no implica un defecto en el procedimiento sino la inevitable concomitancia entre los riesgos de sobredimensionar y subdimensionar». Así puede darse el caso de que un modelo con distinto número de regresores que el PGD tenga asociado el menor riesgo y por tanto sea el modelo seleccionado.

Un resultado standard en la literatura es que el criterio BEC (Geweke y Meese, 1981, págs. 55-70), es consistente, aunque en muestras pequeñas en ocasiones elige modelos anidados en el PGD. Nuestro objetivo es analizar de qué factores depende esa tendencia a elegir modelos diferentes al PGD; más concretamente pretendemos estudiar si la búsqueda de la consistencia provoca el que los criterios sean poco sensibles a determinados factores que, desde nuestro punto de vista, poseen un notable interés. Por ejemplo, ¿se debe tender a elegir el PGD independientemente de la varianza de la perturbación aleatoria del mismo?, ¿afecta a la selección el que los modelos que se comparan sean muy parecidos (en cuanto al número de variables) entre sí? En otras palabras, si admitimos el hecho de que existen casos aceptables de selección de modelos anidados en el PGD, queremos determinar si la consistencia de los criterios es compatible con esas situaciones (siempre para muestras pequeñas) en las que consideramos válido no elegir el PGD. Básicamente esas situaciones son dos:

a) Cuando la varianza del PGD es muy grande.

b) Cuando las variables que están en el PGD y no en el restringido (variables omitidas) tienen poco poder explicativo.

La situación a) es admisible siempre que en la selección se dé importancia al contenido informativo. En efecto, volviendo al cuadro 1 se observa que la varianza del predictor o del error de predicción es proporcional a la varianza del PGD, pero con factor de proporcionalidad mayor en el modelo amplio; si σ^2 aumenta, el seleccionar el PGD puede carecer de interés ya que a pesar de obtener predicciones insesgadas, éstas pueden llevar asociada una varianza tan grande que carezcan de contenido informativo alguno.

En cuanto a la situación b) —bajo poder explicativo de las variables omitidas— podemos considerar los dos casos siguientes:

b.1) Cuando los coeficientes asociados a las variables omitidas sean pequeños, en cuyo caso es más aceptable elegir el modelo restringido que cuando dichos coeficientes sean grandes.

b.2) Cuando los modelos a comparar son muy parecidos (en términos del número de variables).

Si pensamos, por ejemplo, en un PGD en el que aparecen 10 variables explicativas, no es lo mismo comparar este modelo con uno que tenga 9 de esas variables que hacerlo con uno que tenga sólo 1 variable explicativa. Los modelos con 9 y10 variables son relativamente parecidos y es aceptable seleccionar el de 9 alguna vez, aunque no sea el PGD. En cambio, el modelo con 1 variable está más alejado del PGD y la tendencia a alegirlo frente al de 10 debiera ser muy limitada.

b.3) Cuando las variables omitidas presentan alta correlación con otras variables del modelo. La explicación es obvia, si las variables omitidas presentan correlación alta con otras variables, su poder explicativo es escaso y por tanto el seleccionar modelos sin ellas (por ejemplo, modelos anidados en el PGD) es perfectamente admisible.

En este trabajo nos centraremos en las situaciones b.2 y b.3) en el marco de selección entre dos modelos anidados siendo el modelo amplio el PGD, pudiendo extenderse fácilmente al caso más general de comparación entre k modelos.

4. LA SELECCIÓN DE MODELOS ANIDADOS EN EL PGD MEDIANTE EL USO DEL CRITERIO BEC

El criterio BEC consiste en seleccionar el modelo M_j ($j=k_1$, k_2) (definidos en (1) y (2)) que minimice la expresión:

$$\text{BEC}\ (M_j)=\hat{\sigma}_j^2+j\cdot\hat{\sigma}_{k2}^2\frac{\ln T}{T-k_2} \tag{5}$$

con $\hat{\sigma}_j^2$ estimador máximo verosímil de σ^2, utilizando el modelo con j regresores y $\hat{\sigma}_{k2}^2$ el estimador máximo verosímil de σ^2 utilizando el modelo amplio con k_2 regresores.

Según el criterio BEC el modelo restringido será preferido al amplio si BEC $(M_{k1}) < BEC(M_{k2})$, es decir, si:

(6) $$\hat{\sigma}_{k1}^2 + k_1\,\hat{\sigma}_{k2}^2 \frac{\ln T}{T-k_2} < \hat{\sigma}_{k2}^2 + k_2\,\hat{\sigma}_{k2}^2 \frac{\ln T}{T-k_2}$$

El estimador de σ^2 a partir del modelo restringido es:

(7) $$\hat{\sigma}_{k1}^2 = \frac{\hat{u}_1'\hat{u}_1}{T}$$

Si el PGD es el modelo amplio, entonces se tiene que:

$$E(\hat{u}_1'\hat{u}_1) = (T-k_1)\sigma^2 + \beta^{*\prime}X^{*\prime}M_1X^*\beta^* \quad \text{con } M_1 = I - X_1(X_1'X_1)^{-1}X_1'$$

(8) $$\text{es decir, } E(\hat{\sigma}_{k1}^2) = \frac{(T-k_1)\sigma^2}{T} + \frac{\beta^{*\prime}X^{*\prime}M_1X^*\beta^*}{T}$$

Observemos que el segundo sumando a la expresión (8) es fijo pues consideramos el caso de regresores no estocásticos. No obstante para tener una aproximación del tipo de magnitud que representa dicho sumando basta con observar que la expresión $\beta^{*\prime}X^{*\prime}M_1X^*\beta^*$ representa la suma de los cuadrados de los residuos de una regresión de $X^*\beta^*$ sobre X_1:

(9) $$X^*\beta^* = X_1\beta_1 + \omega$$

por lo que:

(10) $$\beta^{*\prime}X^{*\prime}M_1X^*\beta^* = \hat{\omega}'\hat{\omega}$$

y por tanto:

(11) $$E(\beta^{*\prime}X^{*\prime}M_1X^*\beta^*) = \text{Var } \omega \cdot (T-k_1)$$

es decir, $\beta^{*\prime}X^{*\prime}M_1X^*\beta^*$ será una función decreciente de k_1; así, la expresión (8) puede aproximarse por:

(12) $$E(\hat{\sigma}_{k1}^2) = \frac{(T-k_1)\sigma^2}{T} + (T-k_1)\frac{\text{Var } \omega}{T}$$

con ω definido en (9).

Por otro lado, si la varianza se estimara utilizando el modelo amplio:

(13) $$\hat{\sigma}_{k2}^2 = \frac{\hat{u}_2'\hat{u}_2}{T}$$

su esperanza sería:

(14) $$E(\hat{\sigma}^2_{k2}) = \frac{(T-k_2)}{T}\sigma^2$$

Podemos concluir que la expresión (6) se satisfará en media si $E(BEC(M_{k1})) < E(BEC(M_{k2}))$, es decir si:

$$(15)\quad (T-k_1)\frac{\sigma^2}{T}+\frac{\beta^{*\prime}X^{*\prime}M_1X^*\beta^*}{T}+k_1\frac{(T-k_2)}{T}\sigma^2\frac{\ln T}{T-k_2}<\frac{(T-k_2)}{T}\sigma^2+k_2\frac{(T-k_2)}{T}\sigma^2\frac{\ln T}{T-k_2}$$

Simplificando la anterior desigualdad podemos concluir que existirá tendencia hacia modelos anidados en el PGD siempre que:

$$(16)\quad \frac{(k_2-k_1)\ \sigma^2+\beta^{*\prime}X^{*\prime}M_1X^*\beta^*}{\sigma^2}<(k_2-k_1)\ln T$$

o alternativamente, utilizando la aproximación dada en (11), la tendencia hacia modelos anidados en el PGD se dará si:

$$(17)\quad \frac{(k_2-k_1)\sigma^2+\mathrm{Var}\ \omega\ (T-k_1)}{\sigma^2}<(k_2k_1)\ \ln T$$

Analicemos de qué factores depende el que se cumpla o no la anterior desigualdad; el estudio lo llevamos a cabo bajo el supuesto de que modificamos alguna de las condiciones iniciales o parámetros del modelo, dejando fijo el resto.

1) Variaciones en el tamaño muestral

Cuando T aumenta, el lado izquierdo de la desigualdad (17) crece más rápidamente que el derecho, luego la desigualdad resulta más difícil de satisfacer, es decir, la tendencia a seleccionar modelos anidados en el PGD disminuye.

Este hecho no es sino la propiedad de consistencia que sabemos satisfacer el criterio BEC, en el sentido de que a medida que aumenta T, la selección converge hacia el PGD.

2) Variaciones en la varianza del PGD

Cuando σ^2 crece, el lado derecho de la desigualdad se mantiene constante mientras el izquierdo disminuye, es decir, la desigualdad resulta más fácil de cumplir y por tanto aumenta la tendencia a elegir el restringido.

Este hecho concuerda con la situación a) en la que admitíamos la tendencia a seleccionar el modelo restringido a medida que σ^2 aumenta.

3) Variaciones en los coeficientes de las variables que no están en el modelo restringido

Cuando β^* aumenta (si es escalar, consideramos aumento en su valor absoluto, si es vector, en su norma) el lado izquierdo de la desigualdad (16) aumenta, mientras el derecho permanece constante, es decir, la desigualdad resulta más difícil de cumplirse, esto es, disminuye la tendencia a seleccionar el PGD.

Este hecho concuerda con la situación b.1), en la que considerábamos aceptable que se tendiera a elegir más veces el modelo restringido a medida que β^* disminuyese.

4) Variaciones en la distancia relativa de los modelos

Para el caso que se analiza en este trabajo de comparación entre dos modelos anidados, donde el PGD tiene d regresores más que el modelo alternativo, la expresión (17) se reduce a:

(18) $$\frac{d\,\sigma^2+\mathrm{Var}\,\omega\ (T\text{-}k_1)}{d\,\sigma^2}<\ln T \qquad \text{con } d=k_2-k_1$$

Se observa que cuando aumenta k_1 la parte izquierda de la expresión (18) disminuye mientras que el lado derecho permanece constante, lo que indica que se tiende a elegir más veces el modelo restringido (obsérvese que de la aproximación dada en (10), cuando aumenta k_1, los residuos de esa regresión disminuyen).

Es decir, si por ejemplo se comparan dos modelos anidados, el primero con 10 regresores y el segundo con 9, donde el de 10 es el PGD, se tiende a elegir más veces el restringido que en el caso de comparar uno de 2 regresores con uno de 1.

5) Variaciones en el grado de colinealidad entre regresores

Si $X^*\beta^*$ presenta un alto grado de colinealidad con alguna de las variables en X_1, los residuos de la regresión de $X^*\beta^*$ sobre X_1 disminuirán; esto equivale a decir que Var (ω) disminuirá. Por tanto, en la desigualdad (17) el lado izquierdo disminuye, mientras el derecho permanece constante, por lo que, aumenta la tendencia a elegir el modelo restringido, lo que concuerda plenamente con la situación descrita en b.3). Sin embargo, este razonamiento es válido cuando los regresores en X^* son estocásticos.

Los casos 1, 2 y 3 han sido ampliamente estudiados en la literatura; en las secciones siguientes analizamos los casos 4 y 5.

5. Comportamiento del criterio ante la proximidad de los modelos

Uno de los objetivos perseguidos con este trabajo es el de determinar si el criterio BEC es o no sensible a la distancia relativa entre los modelos, entendiendo por tal el cociente de la diferencia entre el número de regresores de ambos modelos y el número de regresores del PGD. Es decir, consideramos que la distancia relativa es mayor cuando se compara un modelo de 2 regresores frente a uno de 1, que si se compara un modelo de 8 regresores frente a uno de 7.

Para determinar la sensibilidad o no del criterio BEC a la distancia entre los modelos, se realiza una comparación entre pares de modelos, que difieren únicamente en la presencia o no de la última variable:

(22) $$M_{k1}:\ y_t=\beta_0+\beta_1X_{1t}+\beta_2X_{2t}+\ldots+\beta_{k1}X_{k1t}+u_{1t}$$
$$M_{k2}:y_t=\beta_0+\beta_1X_{1t}+\beta_2X_{2t}+\ldots+\beta_{k2}X_{k2t}+u_{2t}$$

donde X_{it}~iiN(0, 1), $i=1, 2,\ldots k_2 u_{1t}$~$N(0, \sigma^2_{k1})$, $\beta_i=1$, $i=1, 2,\ldots,k_1$ donde $k_2=k_1+1$ y donde M_{k2} es el Proceso Generador de Datos.

El estudio se realizó con un programa en GAUSS, comparando los resultados de aplicar el criterio BEC a dos modelos con distinto número de regresores k_1 y k_2, respectivamente, para distintos tamaños muestrales T, distintos valores del coeficiente β_{k2}, varios valores de σ^2 y del coeficiente de correlación ρ entre X_{k2} y X_{k2-1}, determinando para cada combinación de pares de modelos el número de veces que el criterio BEC selecciona el PGD. La comparación se realizó para las distintas combinaciones de:

$T=25, 50, 75$

$k_2=2, 3, 4, ..., 8$

$\beta_{k2}=0.2, 0.5, 0.7, 0.9$

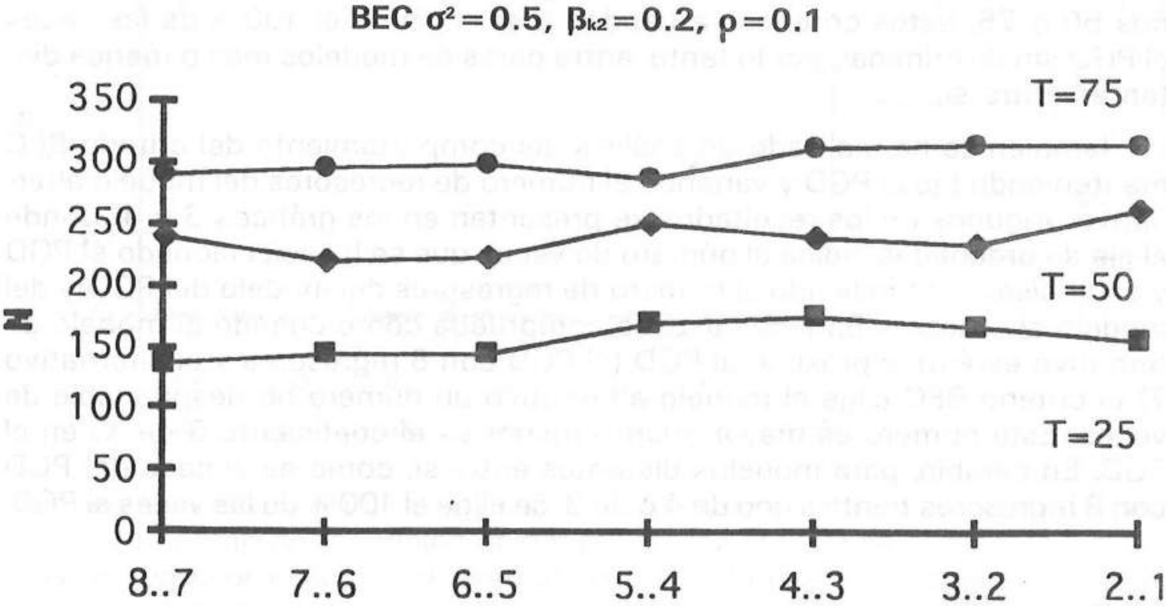
$\rho=0.1, 0.5, 0.99$

$\sigma^2=0.5, 1, 1.5$

Se han elegido estos valores del tamaño muestral por que para tamaños mayores que 75 el criterio ya tiende a seleccionar el 100% de las veces el PGD. Los valores de ρ se han elegido con objeto de abarcar todo el intervalo de valores posibles de correlación entre dos variables, es decir, un valor pequeño, uno intermedio y uno alto. Los valores de la varianza σ^2 se han elegido aleatoriamente ya que el analizar el comportamiento de ésta en la selección a través del criterio BEC no es un objetivo primordial en este trabajo. Se han realizado pruebas para otros valores de estas variables, pero para simplificar la presentación de los resultados y no ampliar en exceso el trabajo aquí se presentan sólo algunos de los resultados. En los gráficos 1 y 2, se presentan algunos de éstos, donde en el eje de ordenadas se indica el número de veces que ha sido seleccionado el PGD para distintos tamaños muestrales, indicando en el eje de abcisas el número de regresores de los pares de modelos entre los que se lleva a cabo la selección.

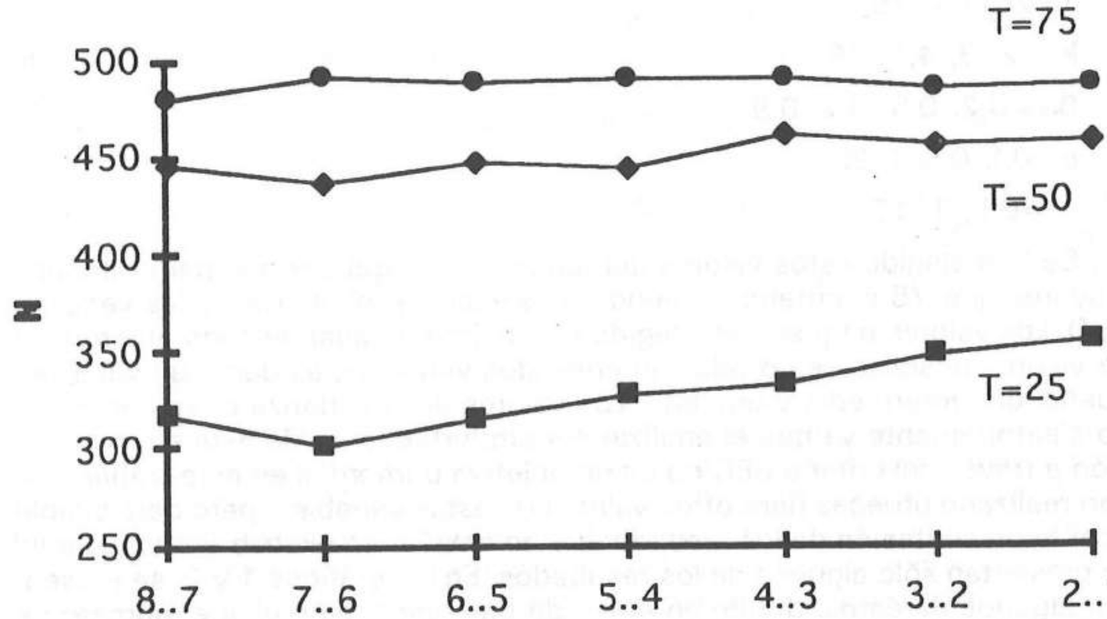
Gráfico 1

BEC $\sigma^2=0.5$, $\beta_{k2}=0.2$, $\rho=0.1$



Se ha demostrado analíticamente que se tiende a elegir más veces el PGD cuando disminuye k_1 (en el caso analizado de $k_2=k_1+1$), como era de esperar, en cambio, gráficamente no se comprueba que los criterios sean muy sensibles a la distancia entre los modelos. Únicamente, en el gráfico 2 y para un tamaño muestral muy pequeño, se observa cómo se selecciona más veces

Gráfico 2
BEC $\sigma^2 = 1$, $\beta_{k2} = 0.5$, $\sigma = 0.5$



el PGD, cuando éste tiene dos regresores frente a uno con un regresor, que en el caso de un modelo de 8 regresores frente a uno de 7.

Tenemos que señalar que debido a la propiedad de consistencia propia de estos criterios, incluso en muestras no grandes como pueden ser los tamaños 50 ó 75, estos criterios ya tienden a seleccionar el 100% de las veces el PGD sin discriminar, por lo tanto, entre pares de modelos más o menos distantes entre sí.

También se ha realizado un análisis del comportamiento del criterio BEC manteniendo fijo el PGD y variando el número de regresores del modelo alternativo. Algunos de los resultados se presentan en los gráficos 3 y 4, donde el eje de ordenadas indica el número de veces que se ha seleccionado el PGD y en abcisas está indicado el número de regresores del modelo del PGD y del modelo alternativo. En este caso, se comprueba cómo cuando el modelo alternativo está muy próximo al PGD (el PGD con 8 regresores y el alternativo 7) el criterio BEC elige el modelo alternativo un número no despreciable de veces. Este número es mayor cuanto menor es el coeficiente β de X_8 en el PGD. En cambio, para modelos distantes entre sí, como es el caso del PGD con 8 regresores frente a uno de 4 o de 3, se elige el 100% de las veces el PGD.

6. Comportamiento del criterio ante la colinealidad de los regresores

En esta sección, tratamos de determinar si el criterio BEC es sensible ante la presencia de altas correlaciones entre las variables o por el contrario su marcada tendencia a capturar la presencia de toda y cada una de las variables del PGD hace que se elijan especificaciones con altas correlaciones entre los

Gráfico 3
BEC $\sigma^2 = 1$, $\beta_{k2} = 0.2$, $\rho = 0.1$

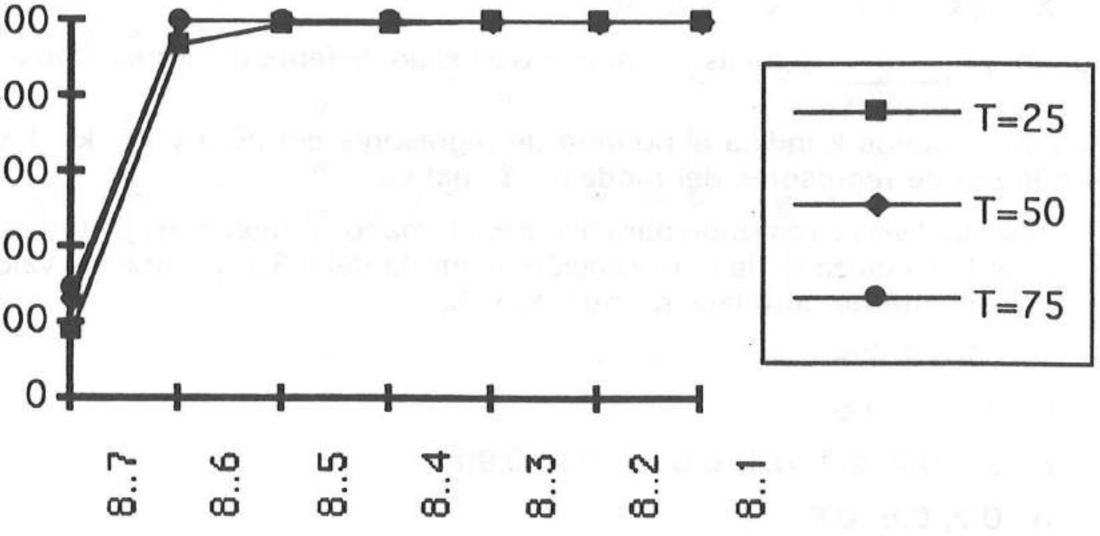
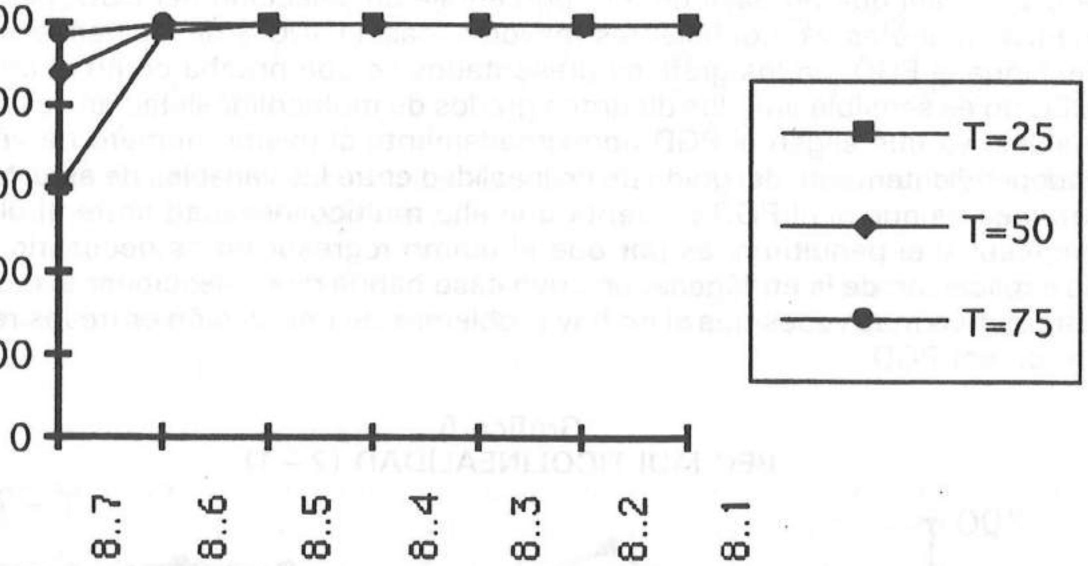


Gráfico 4
BEC $\sigma^2 = 1$, $\beta_{k2} = 0.5$, $\rho = 0.5$



regresores. En concreto, se analiza el comportamiento del criterio BEC cuando se comparan dos modelos anidados, donde el PGD es el modelo amplio, con un regresor más que el alternativo, y donde el último regresor del PGD presenta distintos grados de multicolinealidad con el anterior. Para ello, llevamos a cabo un estudio de Monte Carlo donde analizamos 3 pares de modelos. Dentro de cada par, el PGD tiene un regresor más que el modelo alternativo. El PGD en cada caso ha sido:

1) PGD: $y_t = 1 + X_{1t} + \beta X_{2t} + u_t$

2) PGD: $y_t = 1 + X_{1t} + X_{2t} + \ldots + \beta X_{5t} + u_t$

3) PGD: $y_t = 1 + X_{1t} + X_{2t} + ... + \beta X_{8t} + u_t$

donde: $X_{it} \sim N(0, 1)$ $i = 1, ..., k-1$

$X_{kt} = yX_{k-1,t} + v_t$ $v_t \sim N(0, 1)$

donde: $y = \frac{\rho}{(1-\rho^2)^{1/2}}$ para asegurar que ρ es el coeficiente de correlación entre X_k y X_{k-1}, donde k indica el número de regresores del PGD y $k_0 = k-1$ será el número de regresores del modelo alternativo.

Este análisis es realizado para distintos tamaños muestrales, distintos valores de la varianza de la perturbación aleatoria del PGD, y distintos valores del coeficiente de correlación entre X_k y X_{k-1}:

T=25, 50, 75

$\sigma^2 = 0.5, 1, 1.5$

$\rho = 0.1, 0.2, 0.3, 0.4, 0.5, ..., 0.9, 0.99$

$\beta = 0.2, 0.5, 0.8$

En los gráficos 5 y 6 se indica en el eje de abcisas el grado de correlación entre los dos últimos regresores del PGD y en el eje de ordenadas se indica el número de veces que es seleccionado el PGD. Todos estos gráficos se han obtenido para una varianza de la perturbación igual a 1 y un coeficiente β igual a 0.2, de ahí que no haya un alto porcentaje de selección del PGD, para varianzas menores y/o coeficientes mayores, casi el 100% de las veces se selecciona el PGD. En los gráficos presentados se comprueba cómo el criterio BEC no es sensible ante los distintos grados de multicolinealidad entre las variables, ya que eligen el PGD aproximadamente el mismo número de veces, independientemente del grado de colinealidad entre las variables de éste. Nuestra idea es que si el PGD presenta una alta multicolinealidad entre el último regresor y el penúltimo, es por que el último regresor no es necesario para la explicación de la endógena, en cuyo caso habría que seleccionar el modelo alternativo más veces que si no hay problemas de correlación entre los regresores del PGD.

Gráfico 5
BEC MULTICOLINEALIDAD (2−1)

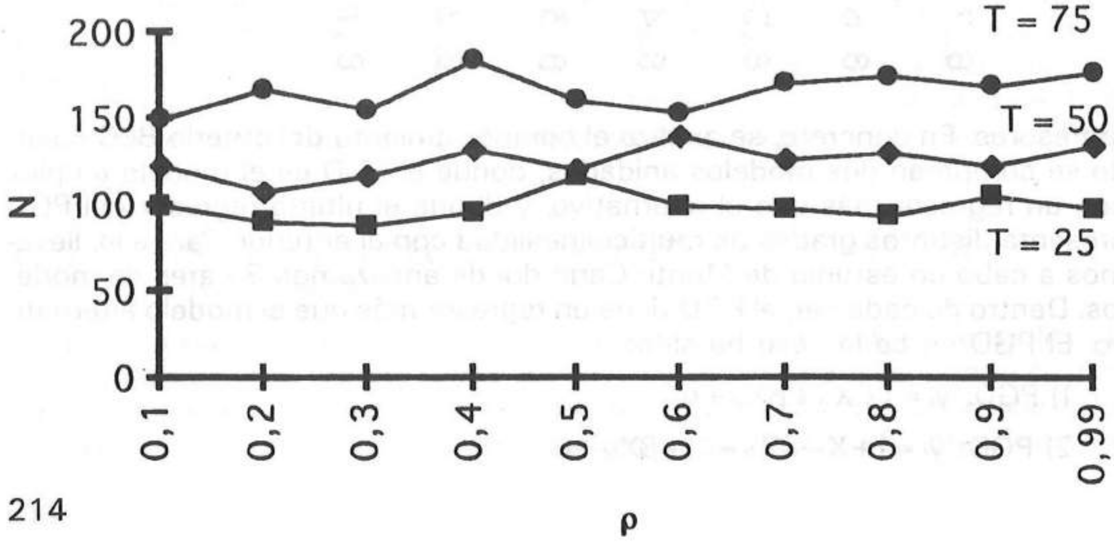
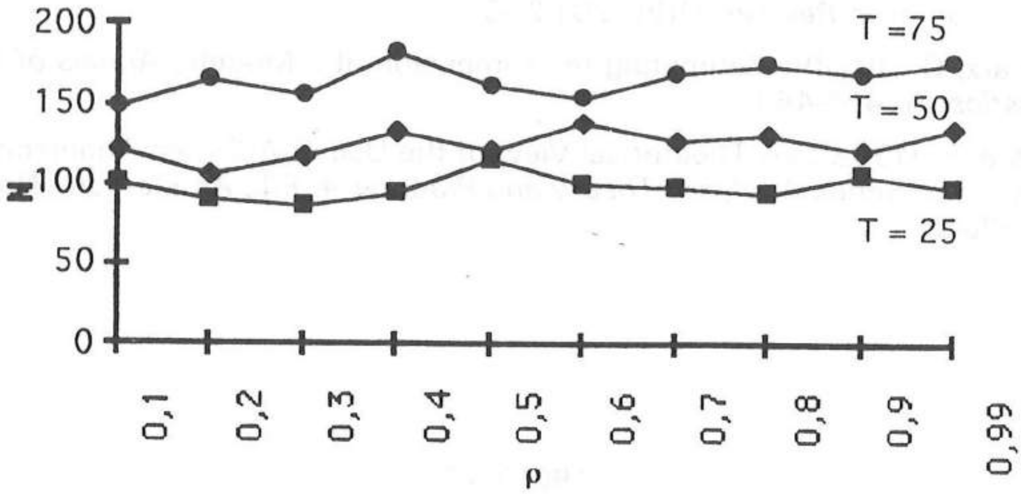


Gráfico 6
BEC MULTICOLINEALIDAD (8−7)



7. CONCLUSIONES

Puede concluirse del desarrollo de este trabajo que aunque el criterio BEC en teoría discrimina entre modelos más o menos próximos entre sí, del estudio de Monte Carlo realizado para tamaños muestrales pequeños, no se desprende la misma conclusión ya que, debido a la consistencia del criterio, éste tiende a seleccionar, incluso en muestras pequeñas el PGD, independientemente de lo próximo que esté a él el modelo alternativo.

Respecto del grado de colinealidad entre variables, el criterio BEC selecciona el modelo con independencia de si existe o no multicolinealidad entre los regresores del mismo; este hecho parece mostrar que la propiedad de consistencia del criterio BEC lleva en ocasiones a seleccionar modelos en los que algunas variables apenas aportan nada a la especificación seleccionada y únicamente son elegidas por formar parte del PGD.

REFERENCIAS BIBLIOGRÁFICAS

Ayuda, M. I. (1994): «El criterio PEC aplicado al contraste de hipótesis no anidadas», Tesis leída en el Departamento de Análisis Económico de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Zaragoza.

Aznar, A. (1989): *Econometric Model Selection. A New Approach,* Dordrecht, Kluwer Academic Publishers.

Aznar, A. y García-Olaverri, M. C. (1995): «El criterio PEC, propiedades asintótica y resultados de un estudio de Monte Carlo», Estadística Española, 139, 217-238.

García-Olaverri, M. C. (1993): «El criterio ACOR: nuevos desarrollos teóricos y resultado de un estudio de Monte Carlo». Tesis leída en el Departamento de Análisis Económico de la Facultad de Ciencias Económicas y Empresariales de la Universidad de Zaragoza.

Geweke, J. y Meese, R. (1981): «Estimating Regression Models of Finite but Unknown Order», *International Economic Review,* 22, 1, 55-70.

Mills, J. A. y Prasad, K. (1992): «A Comparison of Model Selection Criteria», *Econometric Review,* 11(2), 201-233.

Schwarz, C. (1978): «Estimating the Dimension of a Model», *Annals of Statistics,* 6, 461-464.

Shibata, R. (1983): «A Theoretical View of the Use of AIC», en Anderson, O. D., *Time Series Analysis: Theory and Practice,* 4. Ed., Amsterdam, North-Holland.

ABSTRACT

In this article the behaviour of small samples of a consistent selection criterion is analysed, the BEC criterion proposed by Geweke and Meese (1981, p. 55-70). The analysis is carried out in the context of selection between stacked models in which the largest model of those compared is considered to be the data process generator (DPG). In this study the situations under which this criterion tends to select more restricted models than the data process generator (DPG) are determined, concentrating ourselves mainly on the behaviour of the criterion, when faced with the proximity between the compared models, and also faced with the multicollinearity among the DPG regressors. By means of the Monte Carlo study the behaviour of the BEC criterion in small samples is studied, and it is verified how, as a result of being consistent, it tends to select every time (100%) the DPG when this sample size increases, and due to this consistency it generally tends to select the DPG independently of the multicollinearity level among the latter's regressors and the proximity of the models which are under comparison.

Key words: selection, BEC, SBIC, consistence, Monte Carlo.