

## **LA SELECCION DE MODELOS EN EL CONTEXTO DE HIPOTESIS ANIDADAS. UN EXPERIMENTO DE MONTECARLO**

Daniel Coronado Guerrero  
Manuel Acosta Seró  
DEPARTAMENTO DE ECONOMIA GENERAL  
UNIVERSIDAD DE CADIZ

### **Resumen:**

En muchas ocasiones, cuando pretendemos explicar un determinado fenómeno económico a través de un modelo, nos enfrentamos con la inexistencia de una teoría claramente articulada que impide determinar con ciertas garantías el proceso generador de datos, es decir, el modelo teórico que ha podido generar la muestra de observaciones con la que contamos. En estos casos, en los que existe cierto grado de incertidumbre sobre las variables explicativas a considerar, los procedimientos de selección se revelan como instrumentos de especial trascendencia para determinar el "modelo óptimo". En este trabajo realizamos, en una primera parte, una breve exposición de algunos de los criterios propuestos en la literatura sobre la selección de modelos. En una segunda parte, dentro del marco de análisis que nos proporciona el Modelo Lineal General, llevamos a cabo un experimento tipo Montecarlo con la finalidad de indagar en el comportamiento de distintos criterios de selección ante situaciones diferentes.

**Palabras clave:** selección de modelos, hipótesis anidadas, experimento tipo Montecarlo.

### **Introducción**

Aunque generalmente la teoría económica nos proporciona una base para elegir las variables que deben figurar como explicativas en un modelo econométrico, a menudo sólo es capaz de ofrecer al investigador una orientación general. En consecuencia, suele existir cierto grado de incertidumbre sobre los regresores a incluir o excluir. Para afrontar este tópico, los procedimientos de selección se revelan como instrumentos de especial trascendencia en orden a determinar el "modelo óptimo". En este trabajo consideramos un caso particular del problema general de selección de modelos, consistente en la elección de la especificación correcta entre una secuencia de alternativas anidadas. Analizamos las propiedades de cinco de los criterios más utilizados y estudiamos su comportamiento ante diferentes situaciones a través de un ejercicio de simulación tipo Montecarlo.

### **Criterios para la selección del modelo óptimo.**

Los criterios de selección propuestos en la literatura han sido muy variados y, en su mayoría, están basados en la obtención de un equilibrio entre la reducción de la suma de la variancia residual y la aplicación de un principio de parsimonia.<sup>1</sup> En esta línea, se pueden encontrar varias propuestas. Al popular coeficiente de determinación corregido de Theil (1961), se le han ido añadiendo una larga lista de criterios. Baste recordar, por ejemplo, el Coeficiente Cp de Mallows, (1973), el estadístico AIC de Akaike (1974), BIC de Sawa (1978), SBIC de Schwarz (1978), HQ para modelos autorregresivos de Hannan y Quinn (1979), el "Criterio Predicción" de Amemiya (1980), o las propuestas de Shibata (1981), Breiman y Freedman (1983) y, más recientemente, Aznar (1993). Entre estos, algunos de los más utilizados, y que analizaremos en el siguiente apartado a través de un experimento de simulación, son los que a continuación exponemos:

2.1 *Coeficiente de determinación corregido.* Theil (1961, pág. 213) propuso un "coeficiente de determinación corregido", con la finalidad de eliminar una de las principales debilidades del coeficiente de determinación:

$$\overline{R}^2 = 1 - \frac{(n-1)}{(n-K)}(1 - R^2)$$

Elegiríamos el modelo con mayor coeficiente de determinación corregido. Algunas críticas sobre este estadístico pueden encontrarse en Mayer (1975) y Amemiya (1980).

2.2 *El coeficiente Cp.* Una solución que puede ser satisfactoria para la selección del modelo óptimo consiste en la elección de aquél que presente un error de predicción más pequeño. El Coeficiente Cp (Mallows, 1973) gira en torno a esta idea. Una expresión operativa del Coeficiente Cp para el modelo lineal

---

<sup>1</sup> A estos criterios habría que añadir aquellos procedimientos secuenciales o estrategias de selección (forward, backward, stepwise), que tuvieron una amplia difusión en la década de los setenta. En los trabajos de Hocking (1976) y Thompson (1978) se realiza una amplia revisión sobre la selección de regresores desde esta óptica tradicional. En la misma línea de aplicación de "criterios paramétricos", otro tipo de diagnosis consiste en la utilización de un contraste de hipótesis que compara una versión restringida del modelo con otra alternativa en la que figura el modelo general. El rechazo de la hipótesis sobre la base del test sugiere la incompatibilidad de los datos y el modelo, lo que se toma como una evidencia de subespecificación del modelo en cuestión. En este contexto, los test de la Razón de verosimilitud, Test de Wald y el Contraste de los multiplicadores de Lagrange son los que se vienen utilizando con mayor frecuencia.

general es la siguiente (véase Thompson, 1978, o Myers, 1990, pág. 181, para su derivación):

$$C_p = \frac{SCR_i}{\hat{s}^2} + 2 K_i - n$$

Donde  $\hat{s}^2$  es un estimador de  $s^2$  y se obtiene a partir del modelo con el conjunto total de  $k$  variables. A nivel práctico, Chatterjee y Price (1977, pág. 205) resaltan varias de las principales limitaciones del  $C_p$ , centradas sobre todo en la necesidad de obtener una buena estimación de  $s^2$ .

**2.3 Criterio de "Predicción" Amemiya (PC).** Con la finalidad de incluir una consideración sobre la pérdida asociada con la elección de un modelo incorrecto, Amemiya (1980) considera un criterio basado en el error cuadrático medio de la predicción, obteniendo la siguiente expresión, que impone una penalización más alta por añadir variables que el coeficiente de determinación corregido de Theil:

$$PC = \hat{s}_i \left( 1 + \frac{K_i}{n} \right)$$

**2.4 El estadístico AIC.** Bajo los denominados criterios información, la adecuación de un modelo necesita la cuantificación a través de una medida de la distancia entre la verosimilitud -en sentido probabilístico- de un modelo y la del correcto. Atendiendo a este principio, Akaike (1974) propone la elección de aquella ecuación que minimiza el estadístico AIC, que para el modelo lineal toma la siguiente expresión:

$$AIC = \ln \left( \frac{SCR_i}{n} \right) + \frac{2 K_i}{n}$$

El primer término del estadístico mide la bondad del ajuste del modelo, dado un conjunto de datos. El segundo término se interpreta como una penalización que debe pagarse por incrementar el número de parámetros. En este sentido, como apunta Sawa (1978), el AIC puede considerarse como una formulación explícita del denominado "principio de parsimonia" en la construcción de modelos.

2.5 *Criterio de Schwarz*. El Criterio de Schwarz (1978) trata el problema de la selección de un modelo a partir de otros de diferente dimensión desde una óptica bayesiana. Supone una penalización fija por elegir el modelo equivocado y considera una secuencia infinita de modelos anidados, cada uno de los cuales tiene una probabilidad "a priori" distinta de cero. El criterio propuesto por Schwarz (1978), conocido como "Criterio Información Bayesiano de Schwarz" toma la siguiente expresión para el modelo lineal (Geweke y Meese, 1981):

$$SBIC = \ln \tilde{S}_i^2 + \frac{K_i}{n} \ln(n)$$

Como señala el propio Schwarz (1978), tanto el AIC y como el SBIC proporcionan una formulación matemática del principio de parsimonia en la construcción de modelos. Cuantitativamente, el SBIC difiere del AIC sólo en que la dimensión del modelo se multiplica por  $(\frac{1}{2}) \ln(n)$ , con lo cual el criterio favorece más a los de menor dimensión que el AIC.

### **Una simulación a través de un experimento tipo montecarlo**

Nos proponemos ahora analizar el comportamiento de los cinco criterios anteriores a través de un experimento de simulación tipo Montecarlo y utilizando el marco de trabajo que nos proporciona el Modelo Lineal General. Algunos referentes obligados los constituyen, por un lado, el trabajo de Geweke y Meese (1981). Estos autores utilizan un experimento de Montecarlo basado en tres "modelos verdaderos" de diferentes tamaños, en los que la relación entre una variable dependiente  $Y_t$  y un conjunto de variables explicativas (corrientes  $X_t$ , o retardadas  $X_{t-s}$ ) es fuerte. Todos los coeficientes de regresión son mayores o iguales a 0,5 y la relación entre  $Y_t$  y los valores retardados de las explicativas comienza con el término corriente  $X_t$ . Por otro lado, Holmes y Hutton (1989) estudian las propiedades en pequeñas muestras para varios criterios de selección de modelos en aquellos casos en los que la relación verdadera de la variable endógena y las exógenas es muy débil y/o existen retardos. Ambos trabajos están centrados sobre todo en modelos con variables exógenas desplazadas, considerando como muestras pequeñas las de tamaño 50.

3.1 *Planteamiento.* En esta simulación, como veremos a continuación, hemos reducido intencionadamente el tamaño muestral con respecto a los trabajos antes mencionados, con la finalidad de estudiar el comportamiento de los criterios desde una óptica realista, planteando las siguientes situaciones, que son las que consideramos se les pueden presentar con más frecuencia al investigador:

a) El objetivo de la construcción del modelo es fundamentalmente la predicción. Existe una relación de causalidad fuerte entre la variable endógena y los regresores (coeficientes de determinación superiores a 0,9). En estas condiciones asumimos dos supuestos en los que se cumplen las hipótesis habituales de normalidad, media nula y variancia constante de la perturbación aleatoria:

a.1) El modelo correcto es el que contiene el número máximo de regresores. La ecuación utilizada es la siguiente:

$$Y_t = 10 + 0,9 X_1 + 0,8 X_2 + 0,7 X_3 + 0,6 X_4 + 0,5 X_5 + 0,4 X_6 + e_t$$

Sobre este modelo se han construido una secuencia de otros anidados con 5, 4, 3, 2 y 1 variables.

a.2) El modelo correcto es el que contiene el mínimo número de regresores. La ecuación utilizada es la siguiente:

$$Y_t = 10 + 0,9 X_1 + e_t$$

Sobre este modelo se han construido una secuencia de otros modelos con 2, 3, 4, 5 y 6 regresores.

b) El objetivo de la construcción del modelo es la explicación. Existe una relación de causalidad débil entre la variable endógena y los regresores (coeficientes de determinación inferiores a 0,5). En estas condiciones asumimos los mismos dos supuestos anteriores y el comportamiento de ruido blando de la perturbación aleatoria. El experimento se ha llevado a cabo utilizando muestras pequeñas de tamaño 20, y grandes de 200 observaciones.

3.2 *Resultados.* Con la finalidad de llevar a cabo el seguimiento de los criterios según los supuestos anteriores se han generado cien muestras para cada

uno de los modelos correctos y de la secuencia de anidados. La generación de muestras para la variable endógena se ha realizado a partir de los modelos especificados como correctos, utilizando perturbaciones aleatorias con distribución normal, media nula y variancia constante (lógicamente la variancia especificada en el "supuesto b" es superior al "supuesto a", con la finalidad de obtener relaciones débiles). Una vez generadas las muestras se han estimado tanto el correcto como los derivados, para tamaños muestrales de 20 y 200 observaciones, se han calculado la suma de cuadrados de los residuos y, finalmente, se han obtenido los criterios de selección. El experimento completo ha supuesto la estimación de 4.800 modelos. Los resultados en cada supuesto figuran en las tablas 1 a 4. Para cada criterio se indica, de las cien réplicas, el número de veces que se selecciona el modelo correcto con " $k_0$ " regresores, y las veces que se eligen los modelos mal especificados.

-TABLA 1-

SELECCION DEL MODELO OPTIMO RELACION FUERTE . MODELO CORRECTO: $k'=k_0=6$					
CRITERIO	$\overline{R}^2$	$C_p$	$PC$	$AIC$	$SBIC$
n=200					
$k'=k_0=6$	100	100	100	100	100
$k'=k_0-1$	0	0	0	0	0
$k'=k_0-2$	0	0	0	0	0
$k'=k_0-3$	0	0	0	0	0
$k'=k_0-4$	0	0	0	0	0
$k'=k_0-5$	0	0	0	0	0
n=20					
$k'=k_0=6$	95	93	93	95	93
$k'=k_0-1$	5	7	7	5	7
$k'=k_0-2$	0	0	0	0	0
$k'=k_0-3$	0	0	0	0	0
$k'=k_0-4$	0	0	0	0	0
$k'=k_0-5$	0	0	0	0	0
FUENTE: Elaboración propia.					

-TABLA 2-

SELECCION DEL MODELO OPTIMO RELACION FUERTE. MODELO CORRECTO: $k'=k_0=1$					
CRITERIO	$\overline{R}^2$	$C_p$	$PC$	$AIC$	$SBIC$
n=200					
$k'=k_0=1$	45	80	79	79	97
$k'=k_0+1$	9	5	5	5	2
$k'=k_0+2$	10	6	7	7	1
$k'=k_0+3$	6	5	5	5	0
$k'=k_0+4$	15	2	2	2	0
$k'=k_0+5$	15	2	2	2	0
n=20					
$k'=k_0=1$	31	72	65	64	82
$k'=k_0+1$	10	12	13	11	12
$k'=k_0+2$	13	5	5	5	2
$k'=k_0+3$	13	7	8	9	3
$k'=k_0+4$	9	0	3	4	0
$k'=k_0+5$	24	4	6	7	1
FUENTE: Elaboración propia.					

-TABLA 3-

SELECCION DEL MODELO OPTIMO RELACION DEBIL. MODELO CORRECTO: $k'=k_0=6$					
CRITERIO	$\overline{R}^2$	$C_p$	$PC$	$AIC$	$SBIC$
n=200					
$k'=k_0=6$	94	85	85	85	61
$k'=k_0-1$	6	14	14	14	32
$k'=k_0-2$	0	1	1	1	7
$k'=k_0-3$	0	0	0	0	0
$k'=k_0-4$	0	0	0	0	0
$k'=k_0-5$	0	0	0	0	0
n=20					
$k'=k_0=6$	35	19	28	28	12
$k'=k_0-1$	31	21	24	26	18
$k'=k_0-2$	22	30	22	21	22
$k'=k_0-3$	1	1	2	3	3
$k'=k_0-4$	9	23	18	16	31
$k'=k_0-5$	2	6	6	6	14
FUENTE: Elaboración propia.					

-TABLA 4-

SELECCION DEL MODELO OPTIMO RELACION DEBIL. MODELO CORRECTO: $k'=k_0=1$					
CRITERIO	$\overline{R}^2$	$C_p$	$PC$	$AIC$	$SBIC$
n=200					
$k'=k_0=1$	33	73	73	73	97
$k'=k_0+1$	12	11	11	11	3
$k'=k_0+2$	6	7	7	7	0
$k'=k_0+3$	15	5	4	4	0
$k'=k_0+4$	13	2	3	3	0
$k'=k_0+5$	21	2	2	2	0
n=20					
$k'=k_0=1$	34	69	58	56	86
$k'=k_0+1$	18	13	16	16	7
$k'=k_0+2$	9	6	7	8	3
$k'=k_0+3$	8	6	5	5	3
$k'=k_0+4$	17	3	9	9	1
$k'=k_0+5$	14	3	5	6	0
FUENTE: Elaboración propia.					

En la Tabla 1 figuran los resultados para el Supuesto a.1), es decir, existe una relación de causalidad fuerte entre la variable endógena y los regresores, y el modelo correcto es el que contiene el número máximo de regresores. Como puede apreciarse, todos los criterios utilizados arrojan prácticamente los mismos resultados. Incluso para muestras pequeñas no existen diferencias significativas por el hecho de utilizar uno u otro criterio.

En la Tabla 2 se recogen los resultados del experimento para el Supuesto a.2), es decir, la relación de causalidad es fuerte, pero el modelo correcto es el que contiene el número mínimo de regresores. En este caso, la utilización del Coeficiente de determinación corregido no resulta adecuada, ni siquiera cuando

se dispone de muestras grandes. Para el caso de muestras pequeñas, el coeficiente de determinación corregido tiende a la sobreparametrización, es decir, a elegir al modelo con el máximo número de regresores casi en igual proporción que la elección del modelo correcto con un único regresor. El coeficiente SBIC es el que se aproxima al modelo real en el mayor número de casos, incluso en muestras pequeñas, debido a la alta penalización que impone este criterio por el hecho de añadir variables. El resto de criterios tiene un comportamiento similar entre ellos.

De la consideración conjunta de las Tablas 1 y 2 se desprende que, cuando se desconoce totalmente la estructura de la especificación correcta, pero la relación de causalidad entre variables es fuerte, el criterio SBIC se perfila como el más adecuado para la elección del modelo óptimo, tanto en muestras grandes como pequeñas. En estos casos, el coeficiente de determinación corregido es el menos apropiado. Estos resultados confirman los obtenidos por otros autores que han realizado experimentos de similares características (Geweke y Meese, 1981; Holmes y Hutton, 1989).

En la Tabla 3 se reflejan los resultados cuando la relación de causalidad es débil y el modelo correcto es el que contiene el máximo número de variables. En este caso el coeficiente de determinación corregido es el que tiene un comportamiento mejor, tendiendo a la elección del modelo correcto en más ocasiones que el resto de criterios, incluso en el caso de utilización de muestras pequeñas. El criterio SBIC es el que menos se aproxima a la realidad, tanto en muestras grandes, con errores importantes, como en pequeñas, donde tiende a la subespecificación de forma evidente. El resto de criterios tiene un comportamiento parecido en el caso de muestras grandes, pero para muestras pequeñas el Coeficiente  $C_p$  arroja unos peores resultados.

En la Tabla 4 se recogen los resultados para una relación débil y el modelo correcto es el que contiene el menor número de regresores. El criterio SBIC sigue siendo el que acierta en más ocasiones. El coeficiente de determinación corregido tiende a la sobreparametrización, pero mantiene un comportamiento similar en muestras grandes y pequeñas.

De nuevo, de la consideración conjunta de las Tablas 3 y 4 se desprende que, cuando se desconoce totalmente la estructura de la especificación correcta, pero la relación de causalidad entre variables es débil, el criterio SBIC no sería recomendable porque tiene una tendencia muy elevada a la elección de modelos erróneos subespecificados. Algo similar le ocurre al  $C_p$  (para muestras pequeñas). La valoración conjunta de la elección del modelo correcto con el máximo y mínimo número de variables conduce a pensar que los criterios PC y AIC se perfilan como los más adecuados, porque si el verdadero modelo fuera el que contiene el máximo número de variables, las diferencias con el coeficiente de determinación corregido son muy reducidas (tanto para muestras grandes, como pequeñas); sin embargo, si el modelo correcto fuera el que contiene el mínimo número de variables, el coeficiente de determinación corregido elige de forma correcta muchas menos veces que los criterios PC o AIC. Este resultado contradice el obtenido por Holmes y Hutton (1989), quienes llegan a la conclusión que el criterio de la variancia residual (equivalente al coeficiente de determinación corregido) es el que nos llevaría al modelo óptimo e inferencias correctas.

## **Conclusiones**

Como varios autores han puesto de relieve, la selección del adecuado conjunto de regresores y/o del modelo apropiado son problemas difíciles para los que no existe una solución satisfactoria. Es imposible proporcionar una prescripción que pueda seguirse en todos los casos. Con la finalidad de estudiar el comportamiento de varios de los criterios más utilizados en la práctica, hemos llevado a cabo un experimento tipo Montecarlo que revela un diferente comportamiento de aquéllos ante situaciones diferentes. La simulación evidencia que para relaciones de causalidad fuerte, con un objetivo centrado en la predicción, la aplicación sistemática del coeficiente de determinación corregido no proporciona unos resultados satisfactorios por la excesiva tendencia a la sobreparametrización de este coeficiente. Cualquiera de los otros criterios considerados serían más apropiados. Cuando la relación de causalidad es débil - el objetivo es la explicación y no la predicción-, y se desconoce totalmente la

estructura de la especificación correcta, los coeficientes PC o AIC son los que proporcionan los mejores resultados. El coeficiente de determinación corregido arroja en este caso unos resultados algo inferiores a los anteriores.

## Referencias bibliográficas

- AKAIKE, H. (1974): "A new look at the statistical model identification". IEEE Transactions on Automatic Control, AC-19, pp. 716-723.
- AMEMIYA, T. (1980): "Selection of regressors". International Economic Review, vol. 21, pp. 331-354.
- AZNAR, A. (1993): "La selección de un modelo econométrico como búsqueda de una aproximación razonable al PGD". Cuadernos Económicos de ICE, nº 55. pp. 63-86.
- BREIMAN, L.; FREIDMAN, D. (1983): "How many variables should be entered in a regression equation. JASA, vol. 78, pp. 131-136.
- CHATTERJEE, S.; PRICE, B. (1977): Regression analysis by example. Ed. Wiley and Sons, Nueva York.
- GEWEKE, J.; MEESE, R. (1981): "Estimating regression models of finite but unknown order". International Economic Review, vol. 22, nº 1, pp. 55-70.
- HANNAN, E.J.; QUINN, B.G. (1979): "The determination of the order of an autoregression". Journal of the Royal Stat. Society, Series B, vol 41, pp. 190-095.
- HOCKING, R.R. (1976): "The analysis and selection of variables in linear regression". Biometrics, vol. 32, pp. 1-49.
- HOLMES, J.M.; HUTTON, P.A. (1989): "Optimal model selection when the true relationship is weak and occurs with a delay". Economics Letters, vol. 30, pp. 333-339.
- MALLOWS, C.L. (1973): "Some comments on  $C_p$ ". Technometrics, vol. 15, pp. 661-676.
- MAYER, T. (1975): "Selecting economic hypotheses by goodness of fit". The economic Journal, vol. 85, pp. 877-883.
- MYERS, R.H. (1990): Classical and modern regression with applications. Ed. PWS-KENT, Boston.
- SAWA, T. (1978): "Information criteria for discriminating among alternative regression models". Econometrica, vol. 46, nº 6, pp. 1273-91.
- SCHWARZ, G. (1978): "Estimating the dimension of a model". The Annals of Statistics, vol. 6, nº 2, pp. 461-464.
- SHIBATA, R. (1981): "An optimal selection of regression variables". Biometrika, vol. 68, nº 1, pp. 45-54.
- THEIL, H. (1961): Economics Forecast and Policy. Ed. North Holland, Amsterdam.
- THOMPSON, M.L. (1978): "Selection of variables in multiple regression: a review and evaluation". International Statistical Review, vol. 46, pp. 1-19.