CAPÍTULO 8

CRITERIOS PARA ELEGIR EL MEJOR MODELO.

8.1. INTRODUCCIÓN

En el Capítulo 3 hemos hecho referencia a un marco general en el que se derivaban y se evaluaban diferentes procedimientos de contraste. Todo giraba en torno a dos puntos. El primero, hacía referencia a cómo llegar a los procedimientos admisibles que eran los que utilizaban eficientemente la evidencia disponible en los datos. El segundo, consistía en elegir un par (ε_0, δ_0) de las infinitas combinaciones posibles de los tamaños de los dos errores. La solución de la primera cuestión admite un tratamiento estadístico-matemático sobre el cual la Econometría ha proporcionado pautas claras. Sobre el segundo, las orientaciones son más escasas. Hemos comentado dos vías seguidas en la literatura econométrica: Verificacionista y Preferencialista. En la primera, se especificaba a priori el nivel del tamaño del error Tipo 1 fijando un valor muy pequeño; en la segunda no hay ninguna fijación a priori y los tamaños de los dos errores son el resultado de un proceso de toma de decisiones de forma que ambos tamaños dependen de la evidencia muestral.

La evolución de todos los criterios se hacía dentro de un paradigma con dos rasgos distintivos:

- Uno de los modelos que se comparan es el verdadero Proceso Generador de Datos (PGD).

- Todo procedimiento de contraste debe perseguir seleccionar siempre el modelo verdadero.

La evaluación comparada se hace utilizando procedimientos analíticos y métodos de Monte-Carlo. Puede hablarse, incluso, de una cierta "perversidad implícita" en el uso de estos últimos métodos de simulación por la necesidad de asumir de forma "realista" el modelo PGD.

En este Capítulo seguiremos en este marco comentado aunque a lo largo del mismo también consideraremos una situación en la que el PGD no es ninguno de los modelos que se compara y lo que se trata es, no tanto de llegar siempre al modelo verdadero, sino de obtener aproximaciones que sean razonables y aceptables desde algún punto de vista que se especificará.

8.2. MODELOS Y CONTRASTES

Los resultados que vamos a presentar en esta sección se han derivado dentro de un marco con dos modelos lineales anidados lo cual, creemos, no resta generalidad a los mismos.

Los dos modelos lineales anidados pueden escribirse como:

$$M1: y = X_1\beta_1 + u_1 \tag{8.1}$$

$$M2: y = X_2\beta_2 + u_2 \tag{8.2}$$

en donde y es el vector Tx1 de observaciones de la variable dependiente; X_2 es la matriz Txk_2 de observaciones de las k_2 variables que aparecen en M_2 y X_1 es la matriz Txk_1 de observaciones de las k_1 variables que aparecen en M1; por el carácter anidado de ambos modelos se tiene que: $X_2 = (X_1, X^*)$ en donde X* es una matriz Tx $(k_2 - k_1)$ de las observaciones de las (k_2-k_1) variables que aparecen en M2 pero no en M1; u_1 y u_2 son, cada uno de ellos, vectores de T perturbaciones aleatorias. Cuando asumimos que el Proceso Generador de Datos (PGD) es M_1 entonces se cumple que:

$$u_1 \sim N(0,\ \sigma_1^2 I_T)$$

y cuando suponemos que el PGD es M2 entonces se tiene:

$$u_2 \sim N(0,\ \sigma_2^2 I_T)$$

El modelo M2 puede escribirse como:

$$y = X_1\beta_1 + X^*\beta^* + u_2 \tag{8.3}$$

Asumiremos que se dispone de T_1 observaciones extramuestrales y que, para cada una de ellas, podemos escribir:

$$M1: y_p = x'_{p\,1}\beta_1 + u_{p1} \qquad p = 1, ..., T_1 \tag{8.4}$$

$$M2: y_p = x'_{p\,2}\beta_2 + u_{p2} = x'_{p\,1}\beta_1 + x^{*'}_p\beta^* + u_{p2} \qquad p = 1, ..., T_1 \tag{8.5}$$

Si el PGD es M_1 entonces $u_{p1} \sim N(0, \sigma_1^2)$ y se distribuye independientemente del resto de las perturbaciones de los periodos extramuestrales de M1 y de los elementos del vector u_1. Si el PGD es M2 entonces $u_{p2} \sim N(0, \sigma_2^2)$ y se distribuye independientemente del resto de las perturbaciones de los periodos extramuestrales de M2 y de los elementos de u_2.

Para cada modelo se define el vector de residuos MCO como:

$$\hat{u}_i = y - X_i\hat{\beta}_i \tag{8.6}$$

con $$\hat{\beta}_i = (X_i'X_i)^{-1}X_i'y \qquad i = 1, 2 \tag{8.7}$$

Los estimadores MCO de las varianzas de las perturbaciones de los modelos vienen dados por:

$$\hat{\sigma}_i^2 = \frac{\hat{u}_i'\hat{u}_i}{T - k_i} \qquad i = 1, 2 \tag{8.8}$$

y los estimadores Máximo-Verosímiles (MV) por:

$$\tilde{\sigma}_i^2 = \frac{\hat{u}_i'\hat{u}_i}{T} \qquad i = 1, 2 \tag{8.9}$$

Para cada modelo y periodo extramuestral obtenemos el predictor MCO y el correspondiente error de predicción que escribiremos como:

$$e_{pi} = y_p - \hat{y}_{pi} = y_p - x_{pi}'\hat{\beta}_i \qquad i = 1, 2 \tag{8.10}$$

en donde suponemos que para cada período extramuestral que hemos denotado con el subíndice p, cuando se calculan los estimadores MCO de β_i, i = 1, 2, utilizando la expresión escrita en (8.7) en X_i e y se incluyen todas las observaciones previas al correspondiente periodo extramuestral.

Nuestro punto de partida es que los dos modelos son esféricos. Una vez que se garantiza el cumplimiento de esta característica, en la literatura econométrica se ha propuesto el uso de determinados criterios para comparar modelos alternativos.

En el marco de este trabajo sería imposible analizar las características de todos los criterios desarrollados en dicha literatura. Nosotros vamos a limitar nuestro análisis a los criterios que aparecen en el Cuadro 8.1. Otros análisis comparados con diferentes criterios pueden verse en Leamer (1978), Geweke y Meese (1981), Aparicio (1985), Engle y Brown (1985), Aznar (1989), Gourieroux y Monfort (1989), Mills y Prassad (1992) y García-Olaverri (1993).

En el Cuadro 8.1 aparece la denominación del criterio, el estadístico en el que se basa y la regla de decisión que se utiliza. Todos los términos que aparecen en el Cuadro han sido ya definidos; simplemente, indicar que F_ε [(k_2-k_1), T-k_2] y $\chi^2_\varepsilon(k_2 - k_1)$ hacen referencia a los puntos críticos correspondientes al nivel de significación y grados de libertad indicados y que C_{pi} toma la forma siguiente:

$$C_{pi} = x'_{pi}(X_i' X_i)^{-1} x_{pi} \text{ y } C_{i.} = \sum_{p=1}^{T_1} C_{pi} \qquad (8.11)$$

La justificación de la mayor parte de los criterios que aparecen en el Cuadro 8.1 puede encontrarse en Aznar (1989) y Aznar y Trívez (1993). La justificación de los criterios PIC y PICF dentro de un marco bayesiano puede verse en Phillips (1996) y en Phillips y Ploberger (1994, 1996). Estos autores derivan el criterio PIC como una aproximación asintótica a la densidad predictiva de la razón de verosimilitud de las dos hipótesis que se contrastan. El criterio PICF se obtiene condicionando la aproximación anterior a la información correspondiente a un primer periodo muestral.

CUADRO 8.1. Criterios y Regla de Decisión.

CRITERIO	ESTADÍSTICO	REGLA DE DECISIÓN: SE ACEPTA M1 frente a M2 cuando:
Coeficiente de Determinación	$R_i^2 = 1 - \frac{\hat{u}_i'\hat{u}_i}{\sum (y_t - \bar{y})^2}$	$R_1^2 > R_2^2$
Cef.Det.Corregido	$\overline{R}_i^2 = 1 - \frac{T-1}{T-k_i}\left(\frac{\hat{u}_i'\hat{u}_i}{\sum (y_t - \bar{y})^2}\right)$	$\overline{R}_1^2 > \overline{R}_2^2$
Contrastes t y F	$F = \frac{(\hat{u}_1'\hat{u}_1 - \hat{u}_2'\hat{u}_2)/(k_2 - k_1)}{\hat{u}_2'\hat{u}_2/(T-k_2)}$	$F < F_\varepsilon[(k_2 - k_1), T - k_2]$
Wald	$W = T\frac{\tilde{\sigma}_1^2 - \tilde{\sigma}_2^2}{\tilde{\sigma}_2^2}$	$W < \chi_\varepsilon^2(k_2 - k_1)$
Multiplicadores de Lagrange	$LM = T\frac{\tilde{\sigma}_1^2 - \tilde{\sigma}_2^2}{\tilde{\sigma}_1^2}$	$LM < \chi_\varepsilon^2(k_2 - k_1)$
Razón de Verosimilitud	$LR = T\ln\left(\frac{\tilde{\sigma}_1^2}{\tilde{\sigma}_2^2}\right)$	$LR < \chi_\varepsilon^2(k_2 - k_1)$
C_p de Mallows	$C_{pi} = \frac{\tilde{\sigma}_1^2}{\tilde{\sigma}_2^2} + \frac{2k_i}{T-k_2}$	$C_{p1} < C_{p2}$
Akaike	$AIC_i = \ln\tilde{\sigma}_i^2 + \frac{2k_i}{T}$	$AIC_1 < AIC_2$
Densidad Predictiva	$PIC_i = \hat{u}_i'\hat{u}_i + \frac{(k_2 - k_1)\hat{u}_2'\hat{u}_2}{T-k_2}\ln\tilde{\sigma}_2^2 + + \frac{\hat{u}_2'\hat{u}_2}{T-k_2}\ln(X_i'X_i)$	$PIC_1 < PIC_2$
Información	$BEC_i = \tilde{\sigma}_i^2 + \frac{k_i \ln T}{T - K_2}\tilde{\sigma}_2^2$	$BEC_1 < BEc_2$
Schwarz	$SBIC_i = \ln\tilde{\sigma}_i^2 + \frac{k_i \ln T}{T}$	$SBIC_1 < SBIC_2$
Error Cuadrático Medio de Predicción	$ECMP_i = \frac{1}{T_1}\sum_{p=1}^{T_1} e_{pi}^2$	$ECMP_1 < ECMP_2$
Densidad Predictiva Condicional	$PICF_i = \sum_{P=1}^{T_i}\ln(\tilde{\sigma}_{1p}^2(1 + C_{pi})) + + \sum_{p=1}^{T_i}\frac{e_{pi}^2}{\tilde{\sigma}_{pi}^2(1 + C_{pi})}$	$PICF_1 < PICF_2$
Varianza estimada	$AVE_i = \tilde{\sigma}_i^2\left[1 + f(T)\ k_i\right.$	$AVE_1 < AVE_2$

A continuación, vamos a formular, de forma alternativa, la regla de decisión expresando cada criterio como una combinación de un factor de ajuste y de un factor de parsimonia, destacando cómo es este factor de parsimonia el que marca la diferencia entre los criterios.

La expresión genérica que nos va a servir para formular la regla de decisión de todos los criterios es la siguiente: se elige el modelo M1 frente al modelo M2 cuando:

$$\hat{u}_1'\hat{u}_1 < \hat{u}_2'\hat{u}_2 \,.\, h(\,) \tag{8.12}$$

h( ) es lo que hemos llamado factor de parsimonia. Es una función cuyos argumentos dependen del criterio que se considere pero que siempre toma un valor superior a 1 y es creciente de la medida que se toma del tamaño del modelo.

La forma que adopta h( ) para todos los criterios formulados en términos de las sumas de cuadrados de residuos puede verse en la columna 2 del Cuadro 8.2. Obviando los criterios $\overline{R}^2$ y PIC se pueden distinguir tres grupos: el primero está formado por los criterios F, C_p y BEC; el segundo está formado por los criterios W y LM y el tercero abarca a los criterios LR, AIC, SBIC. El criterio AVE tiene una estructura más particular a la que nos referiremos posteriormente.

Derivemos ahora h( ) para el $\overline{R}^2$ y uno de cada uno de los tres grupos.

Utilizando el criterio $\overline{R}^2$ se elige M1 frente a M2 cuando:

$$1-\frac{T-1}{T-k_1}\cdot\frac{\hat{u}_1'\hat{u}_1}{\Sigma(y_t-\bar{y})^2} > 1-\frac{T-1}{T-k_2}\cdot\frac{\hat{u}_2'\hat{u}_2}{\Sigma(y_t-\bar{y})^2}$$

Eliminando términos comunes y agrupando se obtiene:

$$\hat{u}_1'\hat{u}_1 < \hat{u}_2'\hat{u}_2\cdot\frac{T-k_1}{T-k_2}$$

Utilizando el criterio W, se elige M1 frente a M2 cuando:

$$T\frac{\hat{u}_1'\hat{u}_1-\hat{u}_2'\hat{u}_2}{\hat{u}_2'\hat{u}_2} < \chi_\varepsilon^2(k_2-k_1)$$

que es equivalente a:

$$\hat{u}_1'\hat{u}_1 < \hat{u}_2'\hat{u}_2\cdot\left(1+\frac{\chi_\varepsilon^2}{T}\right)$$

Para el criterio F, se elige M1 cuando:

$$\frac{(\hat{u}_1'\hat{u}_1 - \hat{u}_2'\hat{u}_2)/(k_2 - k_1)}{\hat{u}_2'\hat{u}_2/(T - k_2)} < F_\varepsilon[(k_2 - k_1), T - k_2]$$

que es equivalente a:

$$\frac{\hat{u}_1'\hat{u}_1}{\hat{u}_2'\hat{u}_2} < 1 + \frac{k_2 - k_1}{T - k_2} F_\varepsilon$$

o bien

$$\hat{u}_1'\hat{u}_1 < \hat{u}_2'\hat{u}_2 \left(1 + \frac{k_2 - k_1}{T - k_2} F_\varepsilon\right)$$

Por último, con el criterio LR se elige M1 frente a M2 cuando:

$$T \ln\left(\frac{\hat{u}_1'\hat{u}_1}{\hat{u}_2'\hat{u}_2}\right) < \chi_\varepsilon^2 \, (k_2 - k_1)$$

Esta expresión es equivalente a:

$$\left(\frac{\hat{u}_1'\hat{u}_1}{\hat{u}_2'\hat{u}_2}\right) < \exp\left\{\frac{\chi_\varepsilon^2}{T}\right\}$$

de donde se obtiene la expresión para h( ).

Vamos a terminar esta sección reescribiendo todas las expresiones h( ) en la forma que corresponde al criterio F redefiniendo la forma que toma el punto crítico F_ε. Los resultados aparecen en la última columna del Cuadro 8.2.

Escribimos h( ) en la forma siguiente:

$$1 + \frac{(k_2 - k_1)}{T - k_2} F \tag{8.13}$$

Para el criterio $\bar{R}^2$ podemos escribir:

$$\frac{T - k_1}{T - k_2} = 1 + \frac{k_2 - k_1}{T - k_2} F_{\bar{R}^2}$$

con $F_{\bar{R}^2} = 1$

A partir de estos resultados se pueden concretar los valores que toman puntos críticos implícitos correspondientes a los diferentes criterios.

CUADRO 8.2. Formas Alternativas de La Regla de Decisión

CRITERIO	FORMA 1: $\hat{u}_1'\hat{u}_1 < \hat{u}_2'\hat{u}_2$ h( ) h ( )	$h(F) = 1 + \frac{k_2 - k_1}{T - k_2} F$ F
$\bar{R}^2$	$(T - k_1)/(T - k_2)$	$F_{\bar{R}^2} = 1$
F	$1 + \frac{(k_2 - k_1)}{T - k_2} F_\varepsilon$	$F_F = F_\varepsilon$
W	$1 + \frac{\chi^2_\varepsilon}{T}$	$F_W = \frac{T - k_2}{T} \frac{\chi^2_\varepsilon}{k_2 - k_1}$
LM	$\frac{T}{T - \chi^2_\varepsilon}$	$F_{LM} = \frac{T - k_2}{k_2 - k_1} \frac{\chi^2_\varepsilon}{T - \chi^2_\varepsilon}$
LR	$\exp\left(\frac{\chi^2_\varepsilon}{T}\right)$	$F_{LR} = \left(\exp\left(\frac{\chi^2_\varepsilon}{T}\right) - 1\right) \frac{T - k_2}{k_2 - k_1}$
C_p	$1 + \frac{2(k_2 - k_1)}{T - k_2}$	$F_{Cp} = 2$
AIC	$\exp\left(\frac{2(k_2 - k_1)}{T}\right)$	$F_{AIC} = \left(\exp\left(\frac{2(k_2 - k_1)}{T}\right) - 1\right) \frac{T - k_2}{k_2 - k_1}$
PIC	$1 + \frac{k_2 - k_1}{T - k_2} \ln \tilde{\sigma}_2^2 + \frac{1}{T - k_2} \ln \frac{\lvert X_2'X_2 \rvert}{\lvert X_1'X_1 \rvert}$	$F_{PIC} = -\ln \tilde{\sigma}_2^2 + \frac{1}{k_2 - k_1} \ln \frac{\lvert X_2'X_2 \rvert}{\lvert X_1'X_1 \rvert}$
BEC	$1 + \frac{\ln T(k_2 - k_1)}{T - k_2}$	$F_{BEC} = \ln T$
SBIC	$\exp\left((k_2 - k_1)\frac{\ln T}{T}\right)$	$F_{SBIC} = \left(\exp\left(\frac{\ln T\,(k_2 - k_1)}{T}\right) - 1\right)\frac{T - k_2}{k_2 - k_1}$
AVE	$\frac{1 + f(T)k_2}{1 + f(T)k_1}$	$\frac{(T - k_2) f(T)}{1 + f(T)k_1}$

Podemos fijar un nivel de significación, tal como ε = 0,05, y teniendo en cuenta los grados de libertad obtener el correspondiente punto crítico $F_{0,05}$. A partir de este valor, podemos decir que cualquier criterio que tenga un valor de la F –tal como aparecen en la última columna del Cuadro 8.2- superior a $F_{0,05}$ dicho criterio garantizará un nivel del ε inferior al 5%. Por el contrario, para el criterio para el que el valor de la F sea inferior a $F_{0,05}$ podemos decir que la garantía no llegará a ese nivel del 5%.

Una ilustración numérica de estos resultados puede verse en el Cuadro 8.3. En este cuadro se calculan los valores de la F que corresponden a los diferentes criterios, suponiendo 2 tamaños muestrales y dos diferencias de tamaños entre los dos modelos.

CUADRO 8.3. Punto Crítico Implícito T = 45(T = 125) ε = 0,05

PUNTO CRÍTICO	$k_2 = 2,\ k_1 = 1$	$k_2 = 5,\ k_1 = 1$
$F_{\bar{R}^2}$	1	1
$F_{0,05}$	4,08(3,92)	2,61(2,45)
F_W	3,67(3,78)	2,11(2,27)
F_{LM}	4,01(3,89)	2,67(2,46)
F_{LR}	3,83(3,84)	2,34(2,36)
F_{Cp}	2	2
F_{AIC}	1,95(1,98)	1,94(1,98)
F_{BEC}	3,89(4,82)	3,80(4,82)
F_{SBIC}	3,79(4,84)	4,01(5,01)

A partir del contenido del Cuadro 8.3 se observa que los resultados son diferentes según sean los valores que toman $(k_2 - k_1)$ y el tamaño muestral.

En general, podemos decir que los contrastes W, LM y LR se comportan de forma similar a como lo hace el contraste F. Los criterios $\bar{R}^2$, C_p y AIC son menos parsimoniosos que el contraste F aunque la diferencia tiende a decrecer conforme la diferencia $(k_2 - k_1)$ tiende a ser mayor. Por último, los criterios BEC y SBIC tienden a ser más parsimoniosos que el criterio F, haciéndose la distancia mayor conforme la diferencia $(k_2 - k_1)$ se hace mayor.

Cualquiera que sea la aproximación que se siga, la conclusión a la que se llega es que la diferencia entre los diferentes criterios radica en el diferente peso que asignan al factor de parsimonia; según sea la ponderación asignada, la combinación que resulta de los dos tamaños de error difiere. Una ilustración gráfica de este hecho puede verse en la Figura 8.1.

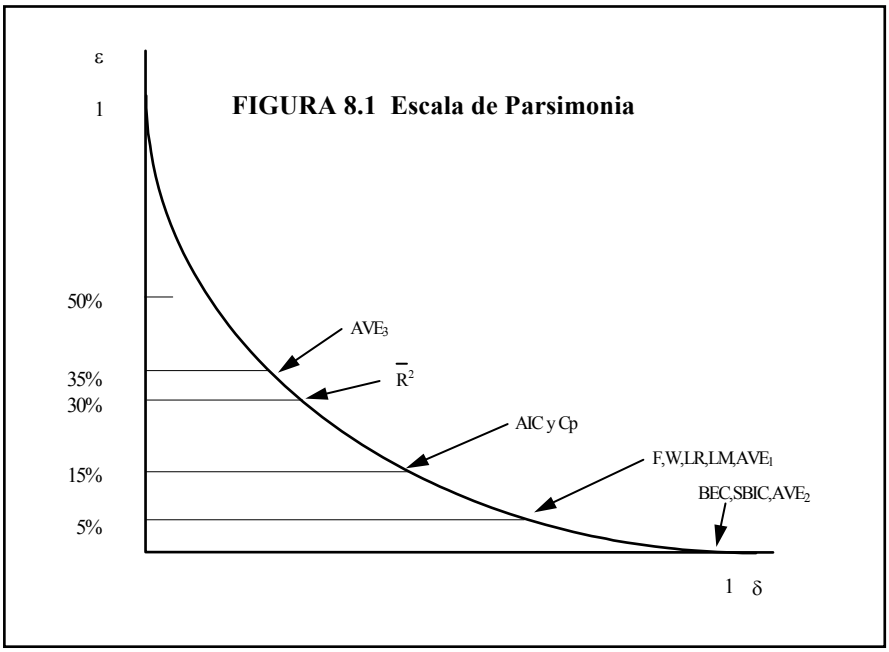

FIGURA 8.1 Escala de Parsimonia

A la vista de estos resultados parece claro que la cuestión relevante a contestar es la siguiente: ¿Qué nivel de parsimonia voy a permitir en el proceso de selección de modelos? o, equivalentemente ¿cómo se determina la función h( )?.

En la literatura econométrica se han seguido, al menos, tres líneas para justificar la forma que debe adoptar la función h( ). Utilizaremos el criterio AVE para ilustrar estas tres alternativas.

Supongamos, en primer lugar, que el investigador está interesado en garantizar que, si se toma el modelo M1 como hipótesis nula, la probabilidad de rechazar esa hipótesis nula sea un valor especificado a priori, ε. En general, este valor se tomará pequeño, digamos el 1% ó el 5%.

Sea F_ε el punto crítico correspondiente al criterio F tal como puede verse en el Cuadro 8.2. Entonces, igualando las funciones h( ) correspondientes a los criterios F y AVE se obtiene:

$$1+\frac{k_2-k_1}{T-k_2}F_\varepsilon=\frac{1+f(T)k_2}{1+f(T)k_1}$$

de donde:

$$f(T) = \frac{F_\varepsilon}{T - k_2 - k_1 F_\varepsilon}$$

El criterio AVE definido de esta manera garantiza que el tamaño del Error Tipo 1 es igual a ε. Esta definición hace al criterio AVE equivalente al criterio F y asociados como el LR, W y LM.

Supongamos ahora que pretendemos definir el criterio AVE de forma tal que se garantice que, asintóticamente, los dos tamaños de error sean cero. Como se verá en la sección siguiente este resultado se garantiza si se cumplen las dos condiciones siguientes conforme $T \to \infty$:

$$h(T) - 1 \to 0$$

$$T(h(T) - 1) \to \infty$$

Una función f(T) que garantiza el cumplimiento de estas dos condiciones es la siguiente:

$$f(T) = \frac{\ln T}{T}$$

Supongamos, por último, que lo que se pretende es que la regla de decisión basada en el AVE sea la consecuencia de la minimización de la comparación de la estimación de las funciones de riesgo correspondientes a ambos modelos después de adoptar una función de pérdida.

Suponiendo que el proceso generador de datos es uno de los dos modelos que se comparan y que la función de pérdida es el cuadrado del error de predicción menos su esperanza matemática, entonces las funciones de riesgo son las respectivas varianzas del error de predicción que podemos escribir como:

$$\sigma_i^2 \left(1 + x_{pj}' (X_j' X_j)^{-1} x_{pj} \right) \qquad i = 1,2, \quad j = 1,2$$

haciendo:

$$x_{pj}' (X_j' X_j)^{-1} x_{pj} = \frac{k_j}{T}$$

se puede escribir:

$$\sigma_i^2 (1 + \frac{k_j}{T})$$

Como la varianza σ_i^2 es desconocida, cada modelo se estima utilizando el estimador máximo-verosímil. La regla de decisión resultante es: se elige M1 frente a M2 cuando se tiene que:

$$\tilde{\sigma}_1^2\left(1+\frac{k_1}{T}\right) < \tilde{\sigma}_2^2\left(1+\frac{k_2}{T}\right)$$

o, equivalentemente, cuando:

$$\hat{u}_1'\hat{u}_1 < \hat{u}_2'\hat{u}_2 \; x \frac{T+k_2}{T+k_1}$$

que es la regla de decisión correspondiente al criterio AVE haciendo:

$$f(T) = \frac{1}{T}$$

La literatura nos brinda, por lo tanto, tres opciones. En la primera, se garantiza un tamaño del Error Tipo 1 fijado a priori. En la segunda, la garantía gira en torno a hacer muy pequeños los tamaños de los dos errores cuando la muestra se hace grande. En la tercera, se minimiza un riesgo estimado. Podiamos considerar tres criterios AVE diferentes: AVE_1, AVE_2 y AVE_3. La diferencia entre ellos radica en la posición que ocupan en la línea de combinaciones de los tamaños de los dos errores de la Figura 8.1. Es decir, se diferencian en la forma de determinar el tamaño de ambos errores, resultando cada uno de ellos con un grado de parsimonia diferente

Utilizando el mismo procedimiento empleado para derivar los valores de la F implícitos que aparecen en la segunda columna del Cuadro 8.2 se obtiene que:

$$F_{AVE_1} = F_\varepsilon$$

$$F_{AVE_2} = \ln T \; x \frac{T-k_2}{T}$$

$$F_{AVE_3} = \frac{T-k_2}{T+k_1}$$

En general, para los valores habituales de T, k_2 y k_1 podemos escribir:

$$F_{AVE2} > F_{AVE1} > F_{AVE3} < 1$$

Esta desigualdad nos indica que el criterio AVE_2 es el más parsimonioso y que el menos parsimonioso es el AVE_3. Como su valor de F implícito es menor que 1 este criterio es incluso menos parsimonioso que el $\overline{R}^2$.

8.3. ANÁLISIS COMPARADO DE LOS CRITERIOS

A la hora de comparar los contrastes analizados en la sección anterior vamos a distinguir dos escenarios. En el primero, se considera que uno de los modelos que se compara es el PGD. En el segundo, se supone que ninguno de los modelos que se compara es el PGD.

Situados en el primer escenario las preguntas relevantes han sido ya formuladas en el Capítulo 3: ¿Son admisibles los procedimietos de contraste estudiados en la sección anterior? ¿Cómo se elige el par de los tamaños de los dos errores?.

Respecto a si son admisibles o no ya se ha visto en el Capítulo 3 que la derivación de un contraste a partir de la Razón de Verosimilitud, Wald o Multiplicadores de Lagrange garantizaba, al menos asintóticamente, su admisibilidad. En la sección anterior se ha demostrado que la expresión general escrita en (8.12) podía identificarse con cualquiera de los tres principios redefiniendo adecuadamente h( ).

Siendo todos admisibles la diferencia entre ellos radica en la diferente forma que toma el par (ε, δ), tal como se ilustra en la Figura 8.1.

Dentro del escenario en el que estamos situados en el que uno de los modelos que se compara es el PGD, entonces la aspiración máxima de todo procedimiento de selección es determinar siempre el modelo verdadero sin ningún tipo de error. Esto, en muestras finitas, en general es imposible de obtener, pero para muestras de tamaño infinito se hace factible como se verá al final de esta sección.

Veamos ahora algunos resultados que nos ayudarán a entender la forma y tamaño de los dos errores.

Si, genéricamente, para un criterio escribimos: $C(M_1) = \hat{u}_1'\hat{u}_1$ y $C(M_2) = \hat{u}_2'\hat{u}_2$ x h(T) entonces teniendo en cuanta la regla escrita en (8.12) el tamaño del error Tipo 1 puede escribirse como:

$$\text{Prob } \{C(M_1) - C(M_2) > 0 \,/\, M_1\} = \varepsilon \tag{8.14}$$

El tamaño del error Tipo 2 viene dado por:

$$\text{Prob } \{C(M_1) - C(M_2) < 0 \,/\, M_2\} = \delta \tag{8.15}$$

A lo largo de esta sección vamos a derivar algunos resultados sobre estos dos tamaños, en primer lugar para cualquier tamaño muestral y, a continuación, derivaremos los resultados en un marco asintótico.

Respecto al tamaño del error Tipo 1 bastaría tener en cuenta lo que hemos dicho en la parte final de la sección anterior. Hemos demostrado que a cada criterio le corresponde un punto crítico implícito, de la distribución F tal como puede verse en el Cuadro 8.3 y que cuanto mayor es este punto crítico implícito menor es el tamaño del error Tipo 1. También se ha destacado que el valor que toma el punto crítico implícito depende del tamaño muestral y de los valores que toman k_1 y k_2.

En lo que respecta al tamaño del error Tipo 2 hay que tener en cuenta que si M2 es el PGD entonces se tiene que:

$$\hat{u}_1 = y - X_1\hat{\beta}_1 = X_2\beta_2 + u_2 - X_1(X_1'X_1)^{-1}X_1'(X_2\beta_2 + u_2) =$$

$$= X_2\beta_2 - X_1B_{12}\beta_2 + u_2 - X_1(X_1'X_1)^{-1}X_1'u_2 \qquad (8.16)$$

en donde:

$$B_{12} = (X_1'X_1)^{-1}X_1'X_2$$

A partir de (8.16) se tiene:

$$E\hat{u}_1 = h_{12}^*$$

y $$E\hat{u}_1'\hat{u}_1 = h_{12}^{*2} + (T - k_1)\sigma_2^2$$

en donde:

$$h_{12}^{*2} = \beta_2'X_2'M_1X_2\beta_2 \quad \text{y} \quad h_{12}^* = M_1X_2\beta_2$$

A continuación, vamos a demostrar que el tamaño del error Tipo 2 escrito en (8.15) depende de σ_2^2 y h_{12}^{*2} e indicaremos la dirección de la dependencia.

Sean μ_c y s_c la media y desviación típica de $C(M_1) - C(M_2)$. Entonces la expresión (8.15) puede escribirse como:

$$\text{Prob } \{C(M_1) - C(M_2) < 0 / M_2\} =$$

$$\text{Prob } \{Z < A_c\}$$

en donde:

$$A_c = \frac{-\mu_c}{s_c} \tag{8.17}$$

y Z es una variable aleatoria con media cero y varianza la unidad. Vamos a demostrar que A_c es una función creciente de σ_2^2 y decreciente de h_{12}^{*2}.

En primer lugar hay que tener en cuenta que:

$$\mu_c = E\,[C(M_1) - C(M_2)] =$$

$$= h_{12}^{*2} + (T - k_1)\sigma_2^2 - (T - k_2)\sigma_2^2\, h(T)$$

que también puede escribirse como:

$$\mu_c = h_{12}^{*2} + (k_2 - k_1)\sigma_2^2 - (k_2 - k_1)\sigma_2^2\ F =$$

$$= \sigma_2^2\, x \left[(k_2 - k_1)(1 - F) + \frac{h_{12}^{*2}}{\sigma_2^2} \right]$$

En lo que respecta a la varianza se tiene que:

$$s_c^2 = Var(\hat{u}_1'\hat{u}_1) + Var(\hat{u}_2'\hat{u}_2)\left(1 + \frac{k_2 - k_1}{T - k_2} F\right)^2 -$$

$$- 2\,Cov(\hat{u}_1'\hat{u}_1, \hat{u}_2'\hat{u}_2)\left(1 + \frac{k_2 - k_1}{T - k_2} F\right) \tag{8.18}$$

Para derivar el primer término, reescribamos (8.16) como:

$$\hat{u}_1 = M_1 X_2 \beta_2 + M_1 u_2$$

Definamos ahora un nuevo vector con $T\text{-}k_1$ elementos de la forma siguiente:

$$V_1 = P\,\hat{u}_1$$

en donde P es una matriz $(T\text{-}\,k_1)$ x T que cumple:

$$P'P = M_1 \quad y \quad PP' = I_{T-k_1}$$

Ahora podemos escribir:

$$V_1 \sim N(PM_1X_2\beta_2,\ \sigma_2^2 I_{T-k_1})$$

Utilizando un resultado standard en estadística se tiene que:

$$\hat{u}_1'\hat{u}_1 = V_1'V_1 \sim \chi^2(h_{12}^{*2},\ T - k_1) \tag{8.19}$$

en donde h_{12}^{*2} es el parámetro de no centralidad.

A partir de (8.19) se tiene que:

$$\text{Var}\ (\hat{u}_1'\hat{u}_1) = 2\sigma_2^4\left(\frac{2h_{12}^{*2}}{\sigma_2^2} + T - k_1\right)$$

Para la suma de cuadrados de los residuos del segundo modelo se tiene que:

$$\text{Var}\ (\hat{u}_2'\hat{u}_2) = 2\sigma_2^4(T - k_2)$$

Utilizando el resultado que sirve para establecer la independencia del numerador y denominador del contraste standard de la F, podemos escribir:

$$\text{Cov}[(\hat{u}_1'\hat{u}_1 - \hat{u}_2'\hat{u}_2)(\hat{u}_2'\hat{u}_2)] = 0$$

de donde:

$$\text{Cov}\ (\hat{u}_1'\hat{u}_1, \hat{u}_2'\hat{u}_2)] = \text{Var}(\hat{u}_2'\hat{u}_2)$$

Utilizando todos estos resultados, (8.18) puede escribirse como:

$$s_c^2 = 2\sigma_2^4\left(\frac{2h_{12}^{*2}}{\sigma_2^2} + T - k_1\right) + 2\sigma_2^4(T - k_2)\left[\left(\frac{k_2 - k_1}{T - k_2}F\right)^2 - 1\right]$$

$$= 2\sigma_2^4\left[\frac{2h_{12}^{*2}}{\sigma_2^2} + k_2 - k_1 + \frac{(k_2 - k_1)^2F^2}{(T - k_2)}\right]$$

Sustituyendo las expresiones obtenidas de μ_c y s_c en (8.17) se llega a:

$$A_c = \frac{-\mu_c}{s_c} = \frac{(k_2 - k_1)(F - 1) - \dfrac{h_{12}^{*2}}{\sigma_2^2}}{\sqrt{2}\left[\dfrac{2h_{12}^{*2}}{\sigma_2^2} + (k_2 - k_1) + \dfrac{(k_2 - k_1)^2F^2}{T - k_2}\right]^{1/2}} \tag{8.20}$$

En esta expresión se ve que el valor de A_c depende de $(k_2 - k_1)$, F, h_{12}^{*2} y σ_2^2.

Utilizando las reglas standard de derivación se obtiene que:

$$\frac{\partial A_c}{\partial \sigma_2^2} > 0 \quad \text{y} \quad \frac{\partial A_c}{\partial h_{12}^{*2}} < 0$$

Un resultado importante a partir de (8.20) es que para poder calcular el tamaño del error Tipo 2 es necesario conocer los valores que toman σ_2^2 y h_{12}^{*2} supuesto conocido F. Por lo tanto, no podemos decir nada con carácter general respecto de este error de Tipo 2 para todo tamaño muestral.

Debido a esta dificultad de determinar con carácter general el tamaño del error Tipo 2, el análisis ha seguido una línea asintótica que permite llegar a resultados más concretos.

Un resultado importante es que el tamaño del error Tipo 1, asintóticamente, tiende a cero cuando el punto crítico implícito (F) que aparece en el Cuadro 8.3 tiende a infinito. Los criterios $\overline{R}^2$, F, W, LM, LR, C_{pi} y AVE_1 claramente no lo cumplen.

Para el criterio AIC podemos escribir:

$$F_{AIC} = \left(\exp\left(\frac{2(k_2 - k_1)}{T} \right) - 1 \right) \text{ x } \frac{T - k_2}{k_{2-k_1}} =$$

$$= \left[1 + \frac{2(k_2 - k_1)}{T - k_2} - 1 \right] \text{ x } \frac{T - k_2}{k_2 - k_1} = 2$$

por lo que el criterio AIC tampoco lo cumple.

El criterio BEC se ve de forma inmediata que cumple la exigencia y en lo que respecta al criterio SBIC se tiene que:

$$F_{SBIC} = \left(\exp\left(\frac{\ln T \ (k_2 - k_1)}{T} \right) - 1 \right) \text{ x } \frac{T - k_2}{k_{2-k_1}} =$$

$$= \left[1 + \frac{\ln T (k_2 - k_1)}{T - k_2} - 1 \right] \text{ x } \frac{T - k_2}{k_2 - k_1} = \ln T \to \infty$$

y, por lo tanto, también cumple el requisito. También es inmediato demostrar que el criterio AVE_2 cumple el requisito.

Hay que hacer notar que, aunque los criterios t, F, LR, W y LM no cumplen el requisito, teniendo en cuenta que fijan el nivel de significación tan pequeño como se quiera podemos decir que, a efectos prácticos, también garantizan un tamaño del error Tipo 1 muy pequeño, próximo a cero. Y esto es así para todo tamaño muestral.

Respecto al error Tipo 2 la condición suficiente es que F/T tienda a cero conforme el tamaño muestral crece.

Teniendo en cuenta que h_{12}^{*2} es $O_p(T)$, dividimos numerador y denominador de (8.20) por T de forma que el numerador de esta expresión tenderá a una cantidad negativa y el denominador tenderá a cero resultando que A_c tenderá a $-\infty$, cumpliéndose el resultado. Puede verse como este requisito lo cumplen todos los criterios recogidos en los Cuadros 8.1 y 8.2.

Teniendo en cuenta (8.13), las condiciones comentadas para garantizar que, asintóticamente, los tamaños de los dos errores se hagan cero se pueden establecer prestando atención a la función h( ).

A partir de (8.13) podemos escribir:

$$F = [h(\ ) - 1] \text{ x } \frac{T - k_2}{k_2 - k_1}$$

Las condiciones de que $\frac{F}{T} \to 0$ y $F \to \infty$ son equivalentes, respectivamente, a:

$$h(\ )\text{-}1 \to 0$$

$$(h(\ )\text{-}1)T \to \infty$$

Como ya hemos indicado, el criterio que cumple estas condiciones garantiza asintóticamente que los tamaños de los dos errores tenderán a cero.

Supongamos ahora que cambiamos de escenario y que nos situamos en un marco en el que ninguno de los modelos que se comparan es el PGD. En este caso, el objetivo de cualquier procedimiento de contraste ya no es seleccionar el PGD, porque este no es ninguno de los que se comparan, sino obtener una aproximación razonable al PGD.

¿Qué es una aproximación razonable?. Para dar respuesta a esta pregunta debemos adoptar una medida de distancia o función de pérdida que nos pondere la calidad de la aproximación a un PGD que es desconocido.

En la literatura se han considerado varias medidas de distancia pero las dos más utilizadas han sido:

– Distancia de Kullback-Liebler.

– Error cuadrático de Predicción (ECP).

La Distancia de Kullback-Liebler ha sido ampliamente utilizada en Econometría para discriminar entre modelos. El criterio AIC de Akaike adopta esta función de pérdida. Recientemente, el libro de Burnham y Anderson (1998) proporciona un tratamiento completo de los temas relacionados con este tipo de distancia.

Nosotros adoptaremos la segunda función de pérdida. Tomamos esta decisión porque nos parece que es una función de pérdida más estrechamente asociada con los usos potenciales de todo modelo econométrico.

Tomaremos T_1 observaciones extramuestrales y siempre consideraremos que la predicción es de un periodo hacia delante.

La función de pérdida la escribiremos como:

$$ECMP_i = \frac{1}{T_1}\Sigma e_{pi}^2 \qquad i = 1,2 \tag{8.21}$$

Los errores de predicción han sido definidos en (7.7) y (7.8).

Suponiendo que el PGD puede ser cualquiera de los tres modelos contemplados en la Sección 7.2., entonces podemos derivar las correspondientes funciones de riesgo aplicando a (8.21) el operador expectativa. Este riesgo será el Error Cuadrático Medio de Predicción (ECMP) que, genéricamente, podemos escribir como:

$$ECMP(M_i)_j = E_j\left[\frac{1}{T_1}\Sigma e_{pi}^2\right] \qquad i= 1,2 \qquad j=1, 2, 3 \tag{8.22}$$

Teniendo en cuenta que, en esta sección, suponemos que M1 está anidado en M2, que ECMP = Sesgo^2 + Varianza y considerando los resultados contenidos en el Cuadro 7.1 se llega a los resultados contenidos en el Cuadro 8.4.

En un marco estacionario puede asumirse la siguiente equivalencia:

$$C_{pi} = x_{pi}^{'}(X_i^{'}X_i)^{-1}x_{pi} \approx \frac{k_i}{T}$$

CUADRO 8.4 Error Cuadrático Medio de Predición

	M1	M2	M3
M1	ECMP(M1)= $=\frac{1}{T_1}\sigma_1^2\sum(1+C_{p1})$	ECMP(M1)= $=h_{12}^2+\frac{\sigma_2^2}{T_1}\Sigma(1+C_{p1})$	ECMP(M1)= $=h_{13}^2+\frac{\sigma_3^2}{T_1}\Sigma(1+C_{p1})$
M2	ECMP(M2)= $=\frac{1}{T_1}\sigma_1^2\Sigma(1+C_{p2})$	ECMP(M2)= $=\frac{1}{T_1}\sigma_2^2\Sigma(1+C_{p2})$	ECMP(M2)= $=h_{23}^2+\frac{\sigma_3^2}{T_1}\Sigma(1+C_{p2})$

C_{pi} ha sido definido en (8.11) y $h_{ij}^2=\frac{1}{T_1}\sum_{p=1}^{T_1}h_{ijp}^2$ $\left(h_{ijp}\right.$ definido en Capítulo 7$\left.\right)$.

Aunque en el Cuadro 8.4 se ha abierto la posibilidad de que cualquiera de los tres modelos puede ser el PGD, para una disciplina como la Economía en la que no cabe casi la experimentación y las relaciones objeto de estudio tienen una gran complejidad de efectos a lo largo del tiempo y en el espacio, la opción más razonable es suponer que el PGD es un modelo muy amplio, desconocido, que abarca a los otros dos modelos que se comparan.

Supongamos entonces que el PGD es el modelo M3. Decimos que el modelo M1 proporciona una aproximación a M3 más razonable que la que proporciona el modelo M2 cuando:

$$\text{ECMP(M1)} < \text{ECMP(M2)}$$

o, equivalentemente, cuando:

$$h_{13}^2+\sigma_3^2\frac{T+k_1}{T}<h_{23}^2+\sigma_3^2\frac{T+k_2}{T}$$

o, también cuando:

$$W=\left(h_{13}^2-h_{23}^2\right)+\sigma_3^2\frac{k_1-k_2}{T}<0 \qquad (8.23)$$

El primer término siempre es positivo mientras que el segundo siempre tiene signo negativo.

Se trata ahora de encontrar un estadístico tal que su media coincida con W y su varianza tiende a cero conforme la evidencia muestral sea mayor.

En las secciones anteriores se ha visto que había dos tipos de criterios. El primer grupo estaba basado en el uso de la expresión (8.12) que realiza una combinación explícita de una medida de ajuste y otra de parsimonia. El segundo grupo se limitaba al Error Cuadrático Medio de Predicción y este combinaba implícitamente las dos medidas mencionadas

A partir de la expresión (8.12) podía pensarse en el siguiente estadístico:

$$\frac{\hat{u}_1'\hat{u}_1}{T} - \frac{\hat{u}_2'\hat{u}_2}{T}\cdot h(T) \tag{8.24}$$

A continuación, hacemos h(T) = 1+g(T) y extendemos el resultado (8.12) en dos direcciones: la primera, al caso en que el PGD sea el modelo M3 y, la segunda, al caso en que tomemos los residuos del modelo M2. Tomando esperanzas se tiene que:

$$E\,\frac{\hat{u}_1'\hat{u}_1}{T} = \frac{h_{13}^{*2}}{T} + \frac{T-k_1}{T}\sigma_3^2$$

$$E\,\frac{\hat{u}_2'\hat{u}_2}{T} = \frac{h_{23}^{*2}}{T} + \frac{T-k_2}{T}\sigma_3^2$$

Introduciendo estos resultados en (8.24) y haciendo: h(T) = 1+g(T) se tiene que:

$$E\left[\frac{\hat{u}_1'\hat{u}_1}{T} - \frac{\hat{u}_2'\hat{u}_2}{T} - \frac{\hat{u}_2'\hat{u}_2}{T}g(T)\right] =$$

$$= \frac{h_{13}^{*2} - h_{23}^{*2}}{T} + \sigma_3^2\,\frac{k_2-k_1}{T} - \left(\frac{h_{23}^{*2}}{T} + \sigma_3^2\,\frac{T-k_2}{T}\right)g(T) \tag{8.25}$$

Se ve claramente que la esperanza difiere de W por lo que ninguno de los estadísticos basados en la expresión (8.12) proporciona un método razonable para elegir la aproximación más útil.

Considerar ahora el siguiente estadístico:

$$ECM = ECMP_1 - ECMP_2 \tag{8.26}$$

en donde $ECMP_1$ y $ECMP_2$ han sido definidos en (8.21). Notar la diferencia entre $ECMP(M_i)$ y $ECMP_i$.

Resultado 8.1: La esperanza matemática de ECM es W; es decir:

$$E(ECM) = W \tag{8.27}$$

Prueba: Basta tener en cuenta los desarrollos utilizados para derivar las casillas de la última columna del Cuadro 8.4 □

Resultado 8.2: Las varianzas de $ECMP_1$ y $ECMP_2$ vienen dadas respectivamente por:

$$Var[ECMP_1] = \frac{2\sigma_3^4}{T_1}\left(1 + \frac{2h_{13}^2}{\sigma_3^2}\right) \tag{8.28}$$

$$Var[ECMP_2] = \frac{2\sigma_3^4}{T_1}\left(1 + \frac{2h_{23}^2}{\sigma_3^2}\right) \tag{8.29}$$

Prueba: Para el primer modelo se tiene:

$$Var(ECMP_1) = \frac{1}{T_1^2}\sum_{p=1}^{T_1} Var(e_{pi}^2) + \frac{1}{T_1^2}\sum_{p,p'} Cov(e_{p1}e_{p'1}) \tag{8.30}$$

Teniendo en cuenta que estos errores de predicción pueden interpretarse como residuos recursivos, su covarianza es cero tal como puede verse en el Capítulo 4 de Aznar (1989). Ver también el Ejercicio 8.3.

Respecto a las varianzas se tiene que, aplicando (7.28) al modelo M1, se puede escribir:

$$e_{p1} \sim N\left[h_{13p}\ ,\ \sigma_3^2\left(1 + \frac{k_1}{T}\right)\right]$$

de donde:

$$\frac{e_{p1}}{\sigma_3\left(1 + \frac{k_1}{T}\right)^{1/2}} \sim N\left(\frac{h_{13p}}{\sigma_3\left(1 + \frac{k_1}{T}\right)^{1/2}}\ ,\ 1\right)$$

Por lo tanto, el cuadrado del término de la izquierda será una variable χ^2 con parámetro de no centralidad igual a $\dfrac{h_{13p}^2}{\sigma_3^2\left(1+\dfrac{k_1}{T}\right)}$ y 1 grado de libertad.

Su varianza toma la forma siguiente:

$$\text{Var}\left[\frac{e_{p1}^2}{\sigma_3^2\left(1+\frac{k_1}{T}\right)}\right]=2\left[1+\frac{2h_{13p}^2}{\sigma_3^2\left(1+\frac{k_1}{T}\right)}\right]$$

A partir de aquí se obtiene que:

$$\text{Var}\left(e_{p1}^2\right)=2\sigma_3^4\left(1+\frac{k_1}{T}\right)^2\left[1+\frac{2h_{13p}^2}{\sigma_3^2\left(1+\frac{k_1}{T}\right)}\right]$$

Si el tamaño muestral es grande, $\dfrac{k_1}{T}$ estará próximo a cero y podemos escribir:

$$\text{Var}\left(e_{p1}^2\right)=2\sigma_3^4\left[1+\frac{2h_{13p}^2}{\sigma_3^2}\right]$$

Sustituyendo en (8.30) se tiene que:

$$\text{Var}\left[\text{ECMP}_1\right]=\frac{2\sigma_3^4}{T_1}\left[1+\frac{2h_{13}^2}{\sigma_3^2}\right]$$

llegándose a (8.28).

La derivación de (8.29) se obtiene aplicando el mismo proceso. □

La conclusión a la que se llega es que las dos varianzas y, por tanto, la covarianza son $O_p(T_1^{-1})$. Eso significa que si T_1 es grande entonces la varianza de ECM tiene a hacerse cero cumpliéndose la exigencia anteriormente formulada.

8.4. RESULTADOS A PARTIR DE EJERCICIOS DE MONTE-CARLO

En esta sección vamos a presentar los resultados obtenidos a partir de dos ejercicios de Monte-Carlo, uno para un marco estacionario y otro para un marco no estacionario.

En el marco estacionario consideramos los dos modelos siguientes:

$$M1:\ Y_t = \beta_0 + \beta_1 X_{1t} + u_{1t}$$

$$M2:\ Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_{2t}$$

Se trata de discriminar entre estos dos modelos suponiendo, primero, que los datos los genera M1 y, después, que los genera M2.

Cuando genera los datos M1, hacemos $\beta_0 = 1$ y $\beta_1 = 0{,}5$; u_{1t} es una variable distribuida normalmente con media cero y para su varianza se suponen tres valores: 0,001, 0,1 y 10.

Si los datos los genera M2 entonces se tiene que: $\beta_0 = 1$ y $\beta_1 = 1$ y para β_2 se permiten dos valores: 0,1 y 0,7. Para u_{2t} se supone el mismo proceso comentado para u_{1t}.

En ambos casos, los valores de las variables X_{1t} y X_{2t} se generan a partir de una variable distribuida normalmente con media cero y varianza la unidad siendo las dos independientes. El ejercicio se lleva a cabo para tres tamaños muestrales: 50, 100 y 500 y el número de simulaciones es igual a 1.000. El programa utilizado para llevar a cabo todos los cálculos ha sido Gauss versión 3.2.12.

Los resultados para el caso en que genera el modelo restringido aparecen en el Cuadro 8.5. En la primera columna aparecen los tres tamaños muestrales, en la segunda, los tres valores supuestos para la varianza de la perturbación del modelo y en las restantes columnas aparecen, para el criterio indicado, el número de veces que se elige el modelo amplio. El nivel de significación adoptado para los criterios F y LR es el 1%. La razón de no incluir otros criterios de los ya estudiados a lo largo de este capítulo es que se comportan de forma similar a alguno de los ya incluídos en el cuadro. Así, por ejemplo, los criterio W y LM reproducen los resultados del criterio LR; el criterio BEC reproduce los resultados del criterio SBIC y el criterio C_p se comporta como el criterio AIC.

Los datos del Cuadro 8.5 nos ponen de manifiesto todas las cuestiones comentadas previamente. Existe una escala de más a menos parsimonia; la distinción es clara entre los cuatro primeros y los restantes. Entre los cuatro primeros, para tamaños muestrales más pequeños se ve que los criterios más parsimoniosos son el F y el LR. Para T=500, los cuatro tienden a igualarse mientras que si hubiéramos considerado tamaños muestrales mayores, los más parsimoniosos hubieran sido los criterios AVE2 y SBIC que es la situación reflejada en la Figura 8.1.

Solo los criterios AVE2 y SBIC son sensibles al tamaño muestral. Conforme este crece el tamaño del error Tipo 1 de ambos se hace más pequeño.

En la última columna aparecen los datos correspondientes al Error Cuadrático Medio de Predicción. No es sensible al tamaño muestral y en la escala de parsimonia se situa entre los criterios AIC y $\overline{R}^2$. Se han distinguido dos casos según sea el número de observaciones extramuestrales. El error se hace menor cuando se pasa a un número mayor de observaciones extramuestrales.

Los resultados para el caso en que los datos los genere el modelo amplio pueden verse en el Cuadro 8.6. La estructura es similar a la del Cuadro 8.5 añadiendo la tercera columna que recoge los dos valores del parámetro de la variable X_2 , β .

En este caso, todos los criterios son sensibles al tamaño muestral; el tamaño del error se hace menor al crecer la evidencia muestral.

Un hecho destacable es la diferencia existente para todos los criterios según sea el valor de β. Esto es lógico porque este valor determina el sesgo en el que incurre el modelo restringido. Cuanto mayor es el valor de este parámetro menor es el tamaño del error.

CUADRO 8.5. Marco Estacionario. PGD: Modelo Restringido $\varepsilon = 0{,}01$

TM	VP	F(=AVE1)	LR	AVE2	SBIC	AIC	$\overline{R}^2$	AVE3	ECM $\left(T_1 = \frac{TM}{2}\right)$	ECM $\left(T_1 = \frac{6}{10}TM\right)$
50	0,001	12	14	68	61	168	310	339	227	187
	0,1	14	16	58	50	167	307	333	236	188
	10	13	16	84	79	177	311	333	220	183
100	0,001	7	8	42	36	155	325	336	211	183
	0,1	10	14	40	38	149	313	322	222	201
	10	6	8	37	34	162	327	335	211	190
500	0,001	7	7	11	11	141	316	319	220	177
	0,1	14	18	22	21	170	322	325	198	183
	10	8	9	13	12	158	321	324	214	183

CUADRO 8.6. Marco Estacionario. PGD: Modelo Amplio.

TM	VP	β	F(=AVE1)	LR	AVE2	SBIC	AIC	$\overline{R}^2$	AVE3	ECM1	ECM2
50	0,001	0,1	31	34	125	113	262	425	447	265	278
		0,7	973	978	998	998	999	1000	1000	976	944
	0,1	0,1	31	31	114	101	251	409	436	266	295
		0,7	967	973	999	998	1000	1000	1000	958	970
	10	0,1	25	29	113	100	255	426	444	262	288
		0,7	969	974	999	998	999	1000	1000	969	946
100	0,001	0,1	42	48	123	116	337	502	512	345	339
		0,7	1000	1000	1000	1000	1000	1000	1000	994	997
	0,1	0,1	46	57	136	123	362	527	535	320	362
		0,7	1000	1000	1000	1000	1000	1000	1000	997	995
	10	0,1	54	59	147	136	355	525	537	341	350
		0,7	1000	1000	1000	1000	1000	1000	1000	994	994
500	0,001	0,1	325	358	393	388	788	892	892	684	683
		0,7	1000	1000	1000	1000	1000	1000	1000	1000	1000
	0,1	0,1	335	354	391	389	791	877	879	709	663
		0,7	1000	1000	1000	1000	1000	1000	1000	1000	1000
	10	0,1	347	378	423	419	822	912	913	689	680
		0,7	1000	1000	1000	1000	1000	1000	1000	1000	1000

CUADRO 8.7. Suma de los Tamaños de los Dos Errores (%)

		F(=AVE1)		LR		AVE2		SBIC		AIC		$\overline{R}^2$		AVE3		ECM$\left(T_1 = \frac{1}{2}TM\right)$		ECM$\left(T_1 = \frac{6}{10}TM\right)$	
TM	VP	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2	1	2
50	0,1	98,3	4,7	98,5	4,3	95,4	5,9	94,9	5,2	91,6	16,7	89,8	30,7	89,4	33,3	94,1	26,6	92,2	23
	10	98,8	4,4	98,7	4,2	98,1	8,5	97,9	8,1	92,5	17,8	88,5	31,1	88,9	33,3	93,2	27,4	92,0	21,3
100	0,1	96,4	1	95,7	1,4	90,4	4	91,5	3,8	78,7	14,9	78,6	31,3	78,7	32,2	86	22,7	80	20,5
	10	95,1	0,6	94,9	0,8	88,0	3,7	89,8	3,4	80,7	16,2	80,5	32,7	79,8	33,5	86,1	21,6	85	19,6
500	0,1	67,9	1,4	66,4	1,8	63,1	2,2	63,2	2,1	47,9	17	44,5	32,2	44,6	32,5	55,7	22	50,2	20,5
	10	66,1	0,8	64,1	0,9	59	1,3	59,3	1,2	33,6	15,8	40,9	32,1	41,1	32,4	53,4	21,4	49,4	18,3

Por último, en el Cuadro 8.7 se recoge la suma de los tamaños de los dos errores para cada criterio y cada combinación tamaño muestral – varianza de la perturbación. Para cada criterio aparecen dos columnas una para cada valor del parámetro β .

En todos los casos un tamaño muestral mayor lleva a que la suma sea menor, lo cual es lógico pues a mayor evidencia en la muestra mayor es la probabilidad de acertar. Un segundo hecho destacable es que la suma es mucho mayor para $\beta = 0{,}1$ (1) que para $\beta = 0{,}7$ (2). Esto también es lógico porque el proceso generador de datos está enviando una señal más clara en el segundo caso. Un último hecho destacable es que conforme nos desplazamos de un criterio más parsimonioso a otro menos parsimonioso el sumatorio para $\beta = 0{,}1$ decrece mientras que el sumatorio para $\beta = 0{,}7$ crece; además, no lo hacen los dos al mismo ritmo pues mientras el decrecimiento del primero es muy ligero, el crecimiento del segundo es más brusco. El resultado es que la diferencia del primer sumatorio entre los criterios F y AVE3 es de 10 puntos siendo mayor el primero, mientras que la diferencia del segundo sumatorio es de casi 30 puntos siendo mayor el segundo.

A la vista de estos resultados y de lo comentado en secciones anteriores cabría pensar en recomendaciones respecto a qué criterios utilizar en el trabajo aplicado. Estas recomendaciones dependen del marco que se adopte en el contraste y el objetivo que se plantee para el mismo.

Respecto al marco caben dos alternativas según que se considere que el proceso generador de datos sea uno de los modelos que se compara u otro modelo diferente de mayor complejidad. En una ciencia social como es la Economía en la que no es posible la experimentación, existen muchas razones para pensar en que lo más normal es el escenario en el que ninguno de los modelos que se compara genere los datos.

Podemos pensar en una primera estrategia en la que o bien por la información a priori o bien por la información que proporciona la evidencia muestral disponible, el investigador está dispuesto a mantener que el PGD es uno de los dos modelos que se comparan. En este caso, si genera los datos el modelo restringido, entonces los criterios menos parsimoniosos siempre van a cometer más error; si los datos los genera un modelo amplio muy distante (midiendo la distancia por el valor que toma el parámetro de la variable X_2), del restringido, entonces todos los criterios van a comportarse de forma similar cometiendo un error muy pequeño; si los datos los genera un modelo amplio muy próximo al restringido entonces los criterios menos parsimoniosos se

comportarán mejor cometiendo un error más pequeño. A la vista de estos resultados la mejor opción sería utilizar un criterio parsimonioso del tipo AVE2 ó asociados, SBIC y BEC. Con esta opción se garantiza que el porcentaje de acierto será elevado en dos situaciones: cuando genera los datos el modelo restringido o cuando los genera un modelo amplio distante del restringido. El porcentaje de fallos será mayor relativamente a otros criterios menos parsimoniosos cuando genere los datos un modelo amplio que esté próximo al modelo restringido. Desde nuestro punto de vista, la conclusión a la que se llega en este caso no tiene mucha relevancia porque no es muy diferente concluir que la variable X_2 no tiene ningún efecto sobre la variable y o que tiene un efecto pero muy pequeño. Sabemos que si el efecto de la variable X_2 sobre y es relevante será detectado con toda seguridad y que solamente puede cometerse algún error si dicho efecto es poco relevante.

Es importante destacar que la única decisión con garantias que puede adoptarse es el rechazo del modelo restringido utilizando un criterio muy parsimonioso (digamos el criterio SBIC o el contraste F con un nivel de significación en torno al 1%). Ese rechazo indicaría que la diferencia entre las verosimilitudes de los dos modelos es tan grande que aún penalizando mucho la del modelo amplio es este el modelo elegido. En el caso en que se acepte la hipótesis nula con un critero muy parsimonioso se está indicando que las verosimilitudes de los dos modelos no son muy diferentes; eso significa que la evidencia contenida en los datos no envian un mensaje claro respecto a que modelo mantener. O bien los datos son malos o bien los dos modelos están próximos de acuerdo con alguna métrica. En este caso, sería útil prestar atención al valor de probabilidad del contraste de la F y si este valor no supera el 10-15%, decidir rechazar la hipótesis nula alertando de que la evidencia de los datos no es muy concluyente. Cuando el valor de probabilidad supere el nivel de significación implícito del criterio menos parsimonioso (digamos el 30% que corresponde al criterio $\overline{R}^2$) entonces no hay evidencia ni para rechazar, ni para aceptar la hipótesis nula. Lo que podría hacerse en este caso es estimar ambos modelos, estudiar su proximidad y llevar a cabo con ellos algún ejercicio de simulación y predicción. Con base en los resultados de estos ejercicios se podría tomar una decisión sobre qué modelo mantener pero siempre, poniendo de manifiesto que la decisión está abierta a todo tipo de dudas porque la evidencia contenida en los datos no permite una discriminación clara.

Si consideramos un último escenario en el que ninguno de los modelos que se compara es el PGD, entonces la recomendación hay que formularla no en base al tamaño que toman los dos errores sino en base a la calidad de la aproximación al PGD desconocido que cada uno de los modelos proporciona. En este marco, los criterios a utilizar son los basados en la estimación de una función de riesgo que coinciden con los

cuatro menos parsimoniosos mencionados. La decisión última dependerá de la función de pérdida que se decida adoptar.

Examinados los resultados para el marco estacionario pasemos ahora a analizar los correspondientes al ejercicio en un marco no estacionario.

Considerar el siguiente PGD:

$$
\begin{aligned}
y_t &= \beta X_t + u_{1t} \\
X_t &= \delta_2 + X_{t-1} + u_{2t} \\
u_{1t} &= \rho_{111} u_{1t-1} + \rho_{121} u_{2t-1} + \varepsilon_{1t} \\
u_{2t} &= \rho_{211} u_{1t-1} + \rho_{221} u_{2t-1} + \varepsilon_{2t}
\end{aligned}
$$

con $$\varepsilon_t \text{ iid} \sim N\left[0, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix}\right]$$

El número de simulaciones es de 1000. Siempre se han considerado tres tamaños muestrales: 50, 100 y 500. Los valores para las varianzas σ_1^2 y σ_2^2 , 0,1 y 2 y dos valores para el coeficiente de la relación de cointegración, 0,1 y 1.

Los modelos que se comparan son:

$$
\begin{aligned}
M1: \ \Delta y_t &= \phi_{11} \Delta y_{t-1} + v_{1t} \\
M2: \ \Delta y_t &= \phi_{11} \Delta y_{t-1} + \phi_{12} \Delta y_{t-2} + v_{2t} \\
M3: \ y_t &= \beta X_t + \phi \hat{u}_{1t-1} + \delta \Delta X_t + v_{3t}
\end{aligned}
$$

CUADRO 8.8. Marco No Estacionario. PGD: Modelo Con Cointegracion

TM	VP	β	(1)	(2)	(3)
50	0,1	0,1	687	767	734
		1	996	996	997
	2	0,1	691	776	731
		1	992	994	993
100	0,1	0,1	891	945	908
		1	1000	1000	1000
	2	0,1	911	949	919
		1	1000	1000	1000
500	0,1	0,1	1000	1000	1000
		1	1000	1000	1000
	2	0,1	1000	1000	1000
		1	1000	1000	1000

en donde:

$$\hat{u}_{1t} = y_t - \hat{\beta} X_t$$

$$\hat{\beta} = \frac{\Sigma X_t y_t}{\Sigma X_t^2}$$

La comparación se lleva a cabo utilizando el error cuadrático medio de predicción (ECMP). Se elige aquel modelo con un menor valor de este estadístico.

En el Cuadro 8.8 pueden verse los resultados para el caso en que $\delta_2 = 0$ y $\rho_{111} = 0{,}5$ y $\rho_{211} = 0{,}1$. La estructura del cuadro es la siguiente. En la primera columna aparecen los tres tamaños muestrales, en la segunda, los valores supuestos para las varianzas, en la tercera, los valores supuestos para el coeficiente de la relación de cointegración y, en las tres restantes, como sigue:

(1) Número de veces que el ECMP del modelo M3 es simultáneamente, menor que los correspondientes a los otros dos modelos.

(2) Número de veces que el ECMP de M3 es menor que el M1.

(3) Número de veces que el ECMP de M3 es menor que el de M2.

Como puede verse, los resultados son sensibles al tamaño muestral y al valor del coeficiente de la relación de cointegración pero no a la varianza de las perturbaciones. Cuanto mayor es el tamaño muestral y cuanto mayor es el coeficiente de la relación de cointegración mayor es el número de veces en que se toma la decisión correcta de elegir el modelo con cointegración. Incluso aunque la cointegración sea pequeña ($\beta = 0{,}1$) si el tamaño muestral es grande se puede llegar con mucha probabilidad a la decisión correcta. Los resultados no varían para valores de δ diferentes de cero pero si que varían conforme lo hacen los coeficientes autorregresivos, ρ_{111} y ρ_{221}; conforme estos coeficientes se aproximan a 1 (digamos 0,8) entonces la probabilidad de tomar una decisión correcta disminuye notablemente.

Por último consideramos un PGD sin cointegración del siguiente tipo:

$$y_t = \delta_1 + y_{t-1} + u_{1t}$$
$$X_t = \delta_2 + X_{t-1} + u_{2t}$$

Para u_{1t} y u_{2t} suponemos el mismo proceso autorregresivo comentado para el modelo con cointegración. Para δ_1 y δ_2 supondremos dos valores, 0 y 1 y el resto de los parámetros toman los mismos valores ya supuestos en el caso anterior.

Los resultados de este ejercicio pueden verse en el Cuadro 8.9. La estructura es la misma que la del Cuadro 8.8 cambiando β por δ_1 y δ_2.

Estos resultados muestran con claridad que el método funciona bien incluso para tamaños muestrales como 50. Cuando el tamaño muestral está en torno a 100 ya se elige siempre el modelo sin cointegración.

CUADRO 8.9. Marco No Estacionario. PGD: Modelo Sin Cointegracion

TM	VP	$\delta_1 = \delta_2$	(1)	(2)	(3)
50	0,1	0	40	40	58
	0,1	1	4	4	7
	2	0	36	38	55
	2	1	13	15	15
100	0,1	0	7	7	10
	0,1	1	0	0	1
	2	0	5	5	11
	2	1	1	1	1
500	0,1	0	0	0	0
	0,1	1	0	0	0
	2	0	0	0	0
	2	1	0	0	0

EJERCICIOS

8.1). Se dispone de T =100 observaciones para una variable dependiente (y) , y seis variables explicativas. Para la variable y se sabe que: $\sum (y_t - \bar{y})^2 = 20$.

Se consideran los modelos anidados M1 , M2, el primero con 3 variables y el segundo con las seis variables, con los siguientes coeficientes de determinación:

$R_1^2 = 0{,}8$, y $R_2^2 = 0{,}9$.

1).Suponiendo que la hipótesis nula es M1 escribir la región crítica que corresponde a los siguientes criterios: $\overline{R}^2$, AIC , LR y SBIC e indicar la decisión que se tomaría utilizando cada uno de los cuatro criterios.

2). Interpretar cada uno de los anteriores contrastes como un contraste de la F derivando el correspondiente punto crítico.

3). Derivar cual sería el tamaño del error tipo 1 que corresponde a cada uno de los cuatro criterios cuando el tamaño muestral tiende a infinito.

8.2). Sea M1 un modelo lineal anidado en otro modelo lineal, M2. Sean $\hat{u}_1$ y $\hat{u}_2$ los respectivos vectores de residuos MCO.

1). Obtener las esperanzas y matrices de varianzas y covarianzas de ambos vectores de residuos generando los datos M2.

2). Demostrar que si genera los datos M2, se cumple que :

$$\hat{u}_2' \hat{u}_2 \leq \hat{u}_1' \hat{u}_1$$

¿Se cumple esta desigualdad si genera los datos M1?.

3). Para discriminar entre M1 y M2 un investigador propone utilizar el criterio AIC de Akaike alegando que dicho criterio garantiza un tamaño del error tipo 1 igual al 5% que,en este caso, corresponde a un punto crítico del contraste de la F igual a 3,5. Evaluar esta propuesta demostrando los resultados que sean necesarios.

4). Otro investigador propone utilizar conjuntamente los criterios $\overline{R}^2$ y el contraste de la F tomando un nivel de significación del 5%. Evaluar la coherencia de esta propuesta.

8.3). Sea M1 un modelo con k_1 regresores que está anidado en otro modelo M2 con k_2 regresores. Tomando toda la información hasta el periodo T ,y utilizando el modelo M1 se hace una predicción MCO para el periodo siguiente; sea e_{T+1} el error de predicción correspondiente. A continuación, usando la información hasta T+1 se

formula la predicción para el periodo T+2; sea e_{T+2} el error de predicción corespondiente. Suponiendo que genera los datos M2 se pide:

1). Derivar $E\left(\frac{e_{T+1}^2 + e_{T+2}^2}{2}\right)$

2). Derivar $Var\left(\frac{e_{T+1}^2 + e_{T+2}^2}{2}\right)$

3). Utilizando los resultados de los dos puntos anteriores comentar la utilidad de utilizar el promedio de los cuadrados de dos errores de predicción sucesivos con un periodo de horizonte para discriminar entre M1 y M2.

8.4). Para un modelo lineal con tres regresores se conoce que:

$\bar{y}=4$; $y'y=200$; $y'X(X'X)^{-1}X'y=190$; $T=10$;

Se pide:

1). Calcular los valores que tomarían los estadísticos $\overline{R}^2$, SBIC y AIC.

2). Suponiendo que el modelo anterior se va a comparar con otro modelo lineal que tiene cinco regresores derivar y calcular el valor de los factores de parsimonia y de los puntos críticos implicitos de la F correspondientes a cada uno de los tres criterios comentados.

3). Determinar qué valor debería tomar la suma de cuadrados de los residuos del modelo con cinco regresores para que dicho modelo fuera aceptado utilizando el criterio AIC y fuera rechazado utilizando el criterio SBIC. Determinar qué valor debería tomar dicha suma para que el modelo con cinco regresores fuera rechazado por los tres criterios.

8.5). Suponer que se están comparando tres modelos, M1, M2 y M3. El primero está anidado en el segundo y éste en el tercero. Con 96 observaciones se han obtenido los siguientes estadísticos:

	k	$\hat{u}'\hat{u}$	$\hat{\sigma}$	AIC	SBIC
M1	5	11.62	?	-2.00	-1.87
M2	7	?	.33	?	-1.95
M3	9	9.58	.33	-2.11	?

en k se incluye como parámetro la varianza de la perturbación del modelo.
Se pide:

1). Calcular las casillas que faltan.

2). Discriminar entre M1 y M2 utilizando los siguientes criterios: $R^2, \overline{R}^2$ y el contraste de la F con un nivel de significación del 5% sabiendo que F (0.05) = 3.09.

3). Derivar y calcular el factor de parsimonia de los criterios $R^2, \overline{R}^2$ y AIC cuando se discrimina entre M1 y M2.

8.6). Sean y_t y x_t dos variables I(1) que están cointegradas. Demostrar que el coeficiente de determinación de la regresión de y_t sobre x_t tiende siempre a ser superior al de la regresión de Δy_t sobre Δx_t. Demostrar que ese resultado no tiene porque cumplirse si las dos variables son I(0).

8.7). El criterio AIC esta basado en la función de pérdida de la distancia de Kullback-Leibler.Utilizando este concepto resolver las dos cuestiones siguientes:

1). Dos expertos predicen que la proporción de consumidores de un determinado producto será respectivamente .7 y .5. Si la verdadera proporción es .6 ¿Qué predicción de las dos está más próxima de la verdadera?.

2). Suponer que la verdadera distribución viene dada por la distribución Normal estandar N(0, 1) ¿ Qué modelo, el N(.5, 1) o el N (0, 1.5) está más próximo de la verdadera distribución?.

8.8). El criterio AIC de Akaike se escribe alternativamente como:

$$AIC = l(\tilde{\theta}) - k$$

o como:

$$AIC = -2l(\tilde{\theta}) + 2k$$

en donde $l(\tilde{\theta})$ es el valor que toma el logaritmo de la función de verosimilitud sustituyendo los parámetros por sus estimadores máximo-verosímiles, y k es el número de parámetros que se estiman.

Utilizando el criterio AIC resolver la siguiente cuestión planteada en Sakamoto, Ishiguro y Kitagawa (1986):

Una máquina produce bolas esféricas. Se conoce que los diámetros de estas bolas, si la máquina funciona normalmente, tienen una media de un centímetro y una desviación típica de .01 cm. . Un dia se eligen aleatoriamente 20 bolas y se miden sus diámetros.

Los resultados son los siguientes.

.999	1.013	.974	.993	.989
1.001	1.008	1.003	.989	1.009
1.001	.977	1.023	.994	.988

1.005 1.006 .995 1.003 1.027

¿ Podemos concluir, a partir de esta información, diciendo que la máquina sigue funcionando normalmente?.
