

CRITERIOS DE SELECCION DE MODELOS ARIMA

Daniel Peña Sánchez de Rivera
y
Gonzalo Arnáiz Tovar
Escuela de Organización Industrial

RESUMEN

Este trabajo compara la eficacia de algunos contrastes diagnósticos para discriminar entre modelos ARIMA, y discute la utilización, en este contexto, del criterio de Información de Akaike (MAIC). Hemos demostrado que el MAIC equivale a un contraste F para probar si la reducción de varianza aportada por la introducción de nuevos parámetros es significativa. El nivel crítico de este contraste es variable en función del número incremental de parámetros introducidos.

Los resultados teóricos se ilustran con un experimento de simulación. La conclusión principal respecto a los contrastes diagnósticos evaluados es que la eficacia del estadístico Q de Box-Pierce aumenta al utilizarlo conjuntamente con un contraste de rachas sobre los signos de los coeficientes de autocorrelación simple. El contraste para la normalidad de Kolmogorov-Smirnov, aplicado a la distribución de los residuos, parece ser de escasa utilidad para detectar errores en la estructura de los modelos. Finalmente, discutimos los problemas de utilizar un criterio automático de selección y presentamos, mediante casos publicados en la literatura especializada, los riesgos asociados a procedimientos automáticos de construcción de modelos.

1. INTRODUCCION

La clase de modelos ARIMA con estacionalidad multiplicativa, ocupa un lugar central en la modelización de series temporales propuestas por Box-Jenkins (1970). Esta clase admite la representación general siguiente:

$$\Phi_p(B^S) \phi_p(B) (\nabla^d \nabla_S^D Z_t^{(\lambda)} - \mu) = \theta_q(B) \theta_Q(B^S) a_t \quad (1.1)$$

Donde B es el operador de retardo definido por $B Z_t \equiv Z_{t-1}$; $\Phi_p(B^S)$ y $\phi_p(B)$ son operadores autorregresivos estacionarios (con ceros fuera del círculo unidad); ∇ es el operador diferencia regular, definido por $\nabla = 1 - B$; ∇_S el operador diferencia estacional, definido por $\nabla_S = 1 - B^S$, $\theta_q(B)$ y $\theta_Q(B^S)$ son operadores media móvil invertibles; $Z_t^{(\lambda)}$ es la serie transformada mediante la transformación instantánea Box-Cox; μ es la media de la serie estacionaria y a_t es un proceso de variables aleatorias normales homocedásticas e independientes que llamaremos en el futuro para simplificar un proceso de ruido blanco (1).

Un problema que ha suscitado considerable atención en la literatura especializada es el hecho de que, dada la falta de un criterio objetivo último de ajuste, distintos analistas pueden, ante una misma serie, obtener modelos no idénticos (véase, por ejemplo, Newbold y Granger (1974)). El problema es importante, porque dos modelos pueden ajustarse prácticamente igual a un conjunto de datos y conducir, sin embargo, a funciones de previsión esencialmente distintas (2).

Recientemente, Akaike (1972) ha presentado un estadístico para la selección automática de modelos, el MAIC, que constituye una generalización del procedimiento FPE (error final de predicción), introducido

para seleccionar, automáticamente, el orden de un proceso autorregresivo. Uno de los objetivos principales de este trabajo, es evaluar teórica y empíricamente este criterio.

El trabajo está estructurado como sigue. En la sección 2 describimos los contrastes diagnósticos que estudiaremos, la relevancia de la varianza residual como criterio de ajuste y los contrastes disponibles para elegir entre modelos, y presentamos los fundamentos teóricos del criterio MAIC de Akaike. La última parte de esta sección 2.4, incluye una comparación teórica entre un contraste de la F de Snedecor basado en los residuos y el criterio MAIC.

La sección 3, resume los resultados de un experimento de simulación. La sección 4, comenta los resultados de Ozaki y Akaike, que han utilizado el MAIC para la selección automática de modelos ARIMA. Finalmente, la sección 5, resume las conclusiones más importantes.

2. CRITERIOS DE SELECCION

2.1 Contrastes diagnósticos

Se llaman contrastes diagnósticos a las pruebas que se aplican a los residuos estimados de un modelo para buscar evidencia que permita rechazar las hipótesis en las que nos hemos basado para estimarlo. Los contrastes diagnósticos para un modelo ARIMA van encaminados a analizar las siguientes propiedades de los residuos estimados: a) Si su media muestral no es significativamente distinta de cero; b) si su varianza es constante; c) si se distribuyen normalmente; d) si son independientes.

Un objetivo importante es diseñar contrastes que, además de ser potentes para revelar las deficiencias del modelo, indiquen vías alternativas para reformularlo.

Puede ocurrir en la práctica que los resultados de las pruebas diagnósticas sean aceptables para más de un modelo. El criterio generalmente seguido en estos casos es ponderar subjetivamente, conjuntamente con los resultados de los contrastes diagnóstico: a) la flexibilidad del modelo; lo que conduce a preferir en igualdad de condiciones, un modelo con un número de diferencias suficientes para que la media de la serie estacionaria sea cero, a otro con una media muestral distinta de cero, (véase Box-Jenkins (1970), pág. 192); b) los estadísticos obtenidos en la etapa de estimación. A igualdad de condiciones, escogeremos aquél modelo que tenga una varianza residual σ_a^2 menor y parámetros más significativos individual y conjuntamente (3).

No conocemos estudios sistemáticos sobre el peso que debemos atribuir a los contrastes diagnósticos con vistas a seleccionar entre modelos. La discusión de éstos contrastes se ha centrado en el estadístico de Box-Pierce (1970), que ha sido criticado por Davies y otros (1977), y modificado recientemente por Ljung y Box (1978). En este trabajo, hemos

intentando una evaluación –desde el punto de vista de selección entre modelos– de los siguientes estadísticos: a) el contraste Q de Box-Pierce; b) un contraste de rachas de independencia, del tipo Wald-Wolfowitz sobre la estructura de cambios de signos en la f_{as} ; c) un contraste de independencia basado en rachas, aplicado a los residuos estimados; d) el contraste de Kolmogorov - Smirnov sobre la normalidad de los residuos.

La razón de considerar estos contrastes es la siguiente: una de las posibles razones para explicar la baja potencia del contraste Q es que este contraste no tiene en cuenta los signos de los coeficientes de autocorrelación simple (cas).

Es, sin embargo, claro que si en el correlograma observásemos que la gran mayoría de los cas son negativos (o positivos), esto nos haría sospechar falta de aleatoriedad. No solamente el número absoluto de cas negativos y positivos es importante, también lo es su secuencia de ordenación en el correlograma. Si observamos una alternancia continua de signos, sospecharemos falta de aleatoriedad, de la misma forma que si no observamos casi ninguna. Esto nos conduce a un test de rachas del tipo Wald-Wolfowitz. (Conover (1971), págs. 351-356).

Hemos decidido aplicar también este contraste a los residuos estimados por razones análogas. Finalmente, hemos considerado interesante introducir un contraste de normalidad para observar empíricamente si la utilización de un modelo inadecuado destruye su configuración normal. Si se estableciesen las implicaciones de distintos errores de formulación en la distribución de los residuos estimados, se dispondría de un nuevo contraste diagnóstico que, al utilizar información distinta de la de los contrastes de independencia, sería presumiblemente muy útil.

2.2 La varianza residual

De los estadísticos obtenidos en la etapa de estimación, el más importante es la varianza residual estimada $\hat{\sigma}_a^2$. Este estadístico nos permite construir una medida de la predictibilidad del modelo mediante:

$$R^2 = 1 - \frac{\hat{\sigma}_a^2}{\hat{\sigma}_w^2}$$

donde $\hat{\sigma}_w^2$ es la varianza de la serie estacionaria y $\hat{\sigma}_a^2$ representa el error de previsión con horizonte 1.

Parece razonable admitir que si la estimación del modelo se basa en minimizar $\hat{\sigma}_a^2$ (4), el modelo con varianza mínima debe de ser el elegido. Sin embargo, la dualidad existente entre la parte autorregresiva y la media móvil, hace el problema complejo.

No existe todavía una teoría general del contraste de hipótesis para modelos ARMA (p, q) y en concreto, carecemos de contrastes rigurosos que nos indiquen cuando la reducción de varianza obtenida mediante una reformulación del modelo es significativa. Cuando el proceso es AR puro, sí podemos aplicar los resultados de la teoría de modelos lineales y efectuar un contraste mediante:

$$F_{(K,N)} = \frac{\Delta VE(K)}{K \hat{\sigma}_a^2(p+K)} \quad (2.1)$$

Donde $\Delta VE(K)$ representa el incremento de variación explicada al pasar del modelo AR (p) al modelo AR ($p+K$) y $\hat{\sigma}_a^2(p+K)$ representa la varianza residual del modelo con $p+K$ parámetros. Este contraste es asintóticamente equivalente a los obtenidos a partir de la razón de verosimilitudes (véase Anderson (1971) y Bartlett (1978)). El más utilizado, (Bartlett (1978; pág. 307)) se basa en el estadístico:

$$\chi_K^2 \sim N \ln \frac{\hat{\sigma}_a^2(p)}{\hat{\sigma}_a^2(p+K)} \quad (2.2)$$

que es asintóticamente una χ^2 con K grados de libertad, siendo N el número total de datos. Desconocemos la extensión rigurosa de estos contrastes para procesos ARMA (p, q). Un argumento a favor de la utiliza-

ción del contraste F en modelos ARMA, es el resultado obtenido por Box-Pierce (1970) que demuestran que si tenemos dos series generadas por el mismo proceso de ruido blanco $\{a_t\}$, la primera ARMA (p, q) y la segunda AR $(p + q)$ y se verifica la condición:

$$\phi_{p+q}(B) = \phi_p(B) \theta_q(B)$$

Donde $\phi_{p+q}(B)$ es el filtro autorregresivo del proceso AR $(p + q)$ y $\phi_p(B)$ y $\theta_q(B)$ los filtros autorregresivos y media móvil del proceso ARMA (p, q) los residuos estimados de ambos modelos serán aproximadamente iguales. Este teorema sugiere que un contraste, a partir de los residuos del modelo ARMA $(p + 1, q + 1)$, de que simultáneamente los dos coeficientes ϕ_{p+1} y θ_{q+1} son cero, es equivalente al efectuado a partir de los residuos de un modelo AR de orden $p + q + 2$. Dado que, en esta segunda formulación, el contraste F —o equivalentemente el de χ^2 — sí es aplicable, al ser los residuos sobre los que se apoya el test iguales, parece razonable admitir una generalización del contraste para modelos generales ARMA (p, q) .

2.3. El Criterio de Akaike

Akaike (1972), ha recomendado la utilización de un estadístico, basado en la teoría de la información y las propiedades del método de máxima verosimilitud, el AIC, para elegir entre modelos (7). Vamos a exponer brevemente, sus fundamentos teóricos.

Dada una distribución de probabilidad discreta $\{x_i, p_i\}$, donde x_i representa los valores particulares de la variable aleatoria, y p_i las probabilidades respectivas, Shannon y Wiener introducen el concepto de entropía mediante:

$$H = - \sum p_i \ln p_i = - E \{ \ln p_i \}$$

La función H es siempre positiva y alcanza su máximo cuando $p_i = p_j$, ($\forall i, j$), siendo cero cuando para algún valor x , $p_i = 1$. (Jerez (1972)). Por tanto, la entropía puede considerarse como una medida na-

tural de la incertidumbre global asociada a una distribución de probabilidad discreta.

La extensión de este concepto al caso continuo mediante:

$$H = -E \{ \text{Ln } f(x) \} = - \int \text{Ln } f(x) f(x) dx$$

presenta problemas importantes. Puede demostrarse –Reza (1961)– que la entropía así definida es, para algunas distribuciones, negativa y, para otras, infinita. Una dificultad adicional, es que la entropía de una distribución continua no es invariante respecto a transformaciones monótonas de la variable, como en el caso discreto. Kullback y Leibler (1951) parten de la noción de información como idea central, definiendo la variación de información al pasar de la distribución inicial $\{y_i, q_i\}$ a la final $\{x_i, p_i\}$ por:

$$I(x; y) = \sum p_i \text{Ln } p_i - \sum p_i \text{Ln } q_i = \sum p_i \text{Ln } \frac{p_i}{q_i}$$

La relación de esta medida con la de Shannon puede verse de la forma siguiente: Si suponemos que el estado inicial es de incertidumbre máxima con $q_i = 1/n$, tendremos:

$$I(x; y) = \sum p_i \text{Ln } p_i + \text{Ln } n = H_{\text{max}} - H(x)$$

y la variación de información al pasar de q a p es simplemente la reducción de entropía en el sentido de Shannon.

Por otra parte, la información aportada por un experimento sobre $\{x_i, p_i\}$ cuando observo y_j viene dada por: (Reza (1961))

$$I(x/y_j) = \sum p(x_i/y_j) \text{Ln } \frac{p(x_i/y_j)}{p(x_i)}$$

que es, precisamente, la variación de información –en el sentido de Kullback– de pasar de $p(x_i)$ a $p(x_i/y_j)$.

Esta medida admite una generalización directa para distribuciones continuas. Definimos variación de información al pasar de f a g mediante:

$$I(g; f) = \int g(x) \ln \frac{g(x)}{f(x)} dx \quad (2.3)$$

que es conocida como cantidad de información –en el sentido de Kullback– al pasar de la distribución f a la g y es positiva a no ser que ambas distribuciones coincidan en casi todo punto.

El criterio de selección de modelos propugnado por Akaike (1974) se basa en la expresión (2.3). Supongamos que $g(x)$ representa la distribución verdadera de una variable y sea $f(x/\theta)$ una familia paramétrica de modelos, donde θ representa un vector de parámetros. La idea de Akaike es escoger aquel modelo que minimiza la cantidad de información requerida para pasar de $f(x/\theta)$ a $g(x)$, es decir, escoger aquel modelo más próximo –desde esta perspectiva– al modelo teórico. Si tenemos en cuenta que:

$$I(g(\cdot); f(\cdot/\theta)) = \int g(x) \ln g(x) dx - \int g(x) \ln f(x/\theta) dx,$$

minimizar $I(g; f)$ es equivalente a maximizar la segunda integral, que es la esperanza de $\ln f(x/\theta)$. Un estimador consistente de esta cantidad es su media muestral dada por:

$$\frac{1}{N} \sum \ln f(x/\theta) \quad (2.4)$$

y, maximizar esta expresión –que es en definitiva el objetivo del método de la máxima verosimilitud– conduce a minimizar $I(g; f)$.

Este resultado muestra como el método de máxima verosimilitud puede considerarse como un criterio natural de ajuste de un modelo.

Si tenemos en cuenta el sesgo del estimador (2.4), la minimización de la integral conduce a escoger el modelo con un valor menor de:

$$AIC = -2 \left(-\sum_{j=1}^N \ln f(x_j/\theta) \right) + 2K$$

Donde K es el número de parámetros estimados —dimensión del vector θ — y $\hat{\theta}$ es el estimador de máxima verosimilitud. Este criterio se conoce con la abreviatura *MAIC*.

La aplicación de este criterio a un modelo ARMA es inmediato. Su función de verosimilitud es (Box y Jenkins (1970), p 213):

$$Ln(\theta, \sigma_a^2) = f(\theta) - \frac{N}{2} \ln \sigma_a^2 - \frac{\sum a_t^2}{2 \sigma_a^2}$$

donde θ representa el vector de parámetros del modelo y $f(\theta)$ es despreciable cuando el tamaño muestral es grande. Entonces:

$$AIC = -2 \left(-\frac{N}{2} \ln \hat{\sigma}_a^2 - \frac{\sum \hat{a}_t^2}{2 \hat{\sigma}_a^2} \right) + 2K$$

$$AIC = N \ln \hat{\sigma}_a^2 + N + 2K \quad (2.5)$$

Donde N es el número total de datos utilizados en la estimación de los parámetros del modelo ARMA, y K representa el número de parámetros estimados que será, para un modelo estacional, igual a la suma de los órdenes de los operadores más uno (debido a $\hat{\sigma}_a^2$), más uno (únicamente cuando la media de la serie estacionaria sea distinta de cero).

El enfoque de Akaike es atractivo porque amplía la teoría de estimación clásica de Fisher-Neyman y Pearson, donde la forma del modelo se supone conocida a priori, a situaciones más generales donde la identificación de la estructura del modelo puede resolverse como un problema de estimación. Akaike (1976), ha utilizado este criterio para investigar representaciones canónicas multivariantes para series temporales y ha presentado sus implicaciones en el análisis bayesiano, en especial en el establecimiento de distribuciones a priori (Akaike (1978b)), y en la selección bayesiana del orden de un proceso autorregresivo (Akaike (1978a)).

Dado que el establecimiento de este criterio se basa en un procedimiento heurístico su “racionalidad” para la selección de modelos depende de su eficacia empírica. Ozaki (1977) ha aplicado este criterio a las series utilizadas en el libro de Box-Jenkins (1970) obteniendo modelos con un número de parámetros muy superior a los estimados por estos autores. Akaike (1974) (1978a) ha presentado ejemplos de utilización con series simuladas. Los resultados de estos autores serán comentados en la sección 4.

2.4. Comparación teórica entre criterios

Si comparamos, utilizando la expresión (2.5), los procedimientos: a) escoger el modelo σ_a^2 con mínima; b) escoger el modelo MAIC; observamos que, para modelos con igual número de parámetros, coinciden, por lo que solo pueden aparecer diferencias si comparamos modelos con distinto número de parámetros. Para ver las implicaciones del MAIC supongamos dos modelos con $p + K$ y p parámetros respectivamente. Escogeremos con el criterio MAIC el primero si:

$$AIC (p + K) < AIC (p)$$

$$N \text{Ln} \hat{\sigma}_a^2 (p + K) + 2 (p + K + 1) < N \text{Ln} \hat{\sigma}_a^2 (p) + 2 (p + 1)$$

$$N \text{Ln} \frac{\hat{\sigma}_a^2 (p + K)}{\hat{\sigma}_a^2 (p)} + 2 K < 0$$

$$N \text{Ln} \frac{\hat{\sigma}_a^2 (p)}{\hat{\sigma}_a^2 (p + K)} > 2 K \quad (2.6)$$

Si contrastásemos la reducción de varianzas residual mediante un Test F , construiríamos: (8)

$$F_{(K,N)} = \frac{N \hat{\sigma}_a^2 (p) - N \hat{\sigma}_a^2 (p + K)}{K \hat{\sigma}_a^2 (p + K)}$$

y poniendo:

$$N \hat{\sigma}_a^2(p) = \hat{\sigma}_a^2(p + K) N + K F_{(K, N)} \quad (2.7)$$

y sustituyendo (2.7) en (2.6):

$$N \operatorname{Ln} \frac{\hat{\sigma}_a^2(p)}{\hat{\sigma}_a^2(p + K)} = N \operatorname{Ln} \left(1 + K F_{(K, N)} \frac{(K, N)}{N} \right) \approx K F_{(K, N)} > 2 K$$

$$F_{K, N} > 2$$

donde hemos utilizado que, cuando Z es pequeño, $\operatorname{Ln}(1 + Z) \approx Z$. El criterio de Akaike equivale a realizar un contraste de la F con nivel crítico igual a 2. Si comparamos (2.6) con (2.2) vemos que es también equivalente asintóticamente a realizar un contraste χ_K^2 con nivel crítico $2 K$.

Puede argumentarse que el criterio de Akaike está encaminado a seleccionar el modelo óptimo entre un conjunto de modelos teóricos que establecemos a priori –por ejemplo fijando los órdenes máximos de p y q para modelos ARMA– y no puede compararse con un problema distinto como es el de seleccionar entre modelos. En la sección 4 comentaremos los problemas que aparecen al proceder mecánicamente poniendo todo el peso en un criterio único. Lo que queremos resaltar es que si procedemos, siguiendo la metodología de Box-Jenkins, a través de las etapas de identificación –estimación– contrastes diagnósticos, el problema de elección se presenta únicamente respecto a un pequeño número de modelos (no más de 2 ó 3).

Si utilizamos el criterio MAIC para decidir, estamos, de hecho, efectuando contrastes F para comprobar si la reducción de variabilidad por incrementar el número de parámetros es significativa. Hemos mostrado que estos contrastes se realizan con un F crítico igual a 2, sea cual sea el valor de K . Si suponemos que N es grande y utilizamos las tablas de la F , (K, ∞) –o también alternativamente, las de χ_K^2 – obtenemos que la probabilidad de esta región crítica varía del 0,22 aproximadamente para $K = 1$ a 0,1 para $K = 4$ y 0,05 para $K = 8$. Por lo tanto existe un riesgo apreciable de escoger modelos con un número mayor

de parámetros que los requeridos. Shibata (1976) ha obtenido teóricamente para procesos AR la distribución del orden seleccionado en función del orden verdadero y sus resultados, obtenidos con un enfoque muy diferente, concuerdan con los nuestros. Podemos resumir que el criterio de Akaike conducirá, en promedio, alrededor del 25% de los casos a seleccionar modelos sobreparametrizados.

En las simulaciones del cuadro 1 el MAIC escoge el modelo correctamente para las tres simulaciones.

3. RESULTADOS EXPERIMENTALES

3.1. Planteamiento

Para comparar los estadísticos a los que nos hemos referido en la sección anterior hemos efectuado un experimento de simulación. El programa que hemos utilizado consta de dos partes: Un generador de ruido blanco a_t , seguido de un filtro lineal que genera Z_t .

El generador de ruido blanco está construido a partir del algoritmo de Marsaglia y Mc Laren (1965), utilizando el generador multiplicativo congruencial:

$$x_n \equiv 16807 x_{n-1} \pmod{2^{31} - 1}$$

Este generador se ha contrastado respecto: a) a la uniformidad mediante el contraste de Kolmogorov-Smirnov; b) a la independencia con el contraste de rachas recomendado por Knuth (1969). En ambos casos no hemos encontrado ninguna evidencia de posibles deficiencias.

A continuación utilizamos el algoritmo de Box y Müller (1965) para generar variables normales. La secuencia obtenida $\{a_t\}$, ha sido sometida a un análisis Box-Jenkins para contrastar su carácter de ruido blanco.

Una vez comprobada la adecuación de las $\{a_t\}$ esta secuencia se filtra a través de:

$$Z_t = \frac{\theta_a(B) \Theta_Q(B^S)}{\phi_p(B) \Phi_P(B^S) \nabla^d \nabla_s^D} a_t$$

para obtener el modelo ARIMA estacional. En todos los casos hemos desechado los cien primeros valores iniciales de Z_t para evitar la influencia de las condiciones iniciales. Hemos trabajado con series de 168 datos

El experimento ha consistido en obtener realizaciones simuladas de 17 secuencias $\{a_t\}$ a las que hemos sometido a distintos filtros para generar 17 procesos ARIMA multiplicativos. A la vista de las fas y fap muestrales hemos estimado para cada muestra de 5 a 10 modelos distin-

tos, obteniendo al final datos de estimación de 111 distintos modelos. Los resultados que presentamos se basan en estas 111 modelizaciones.

3.2. Resultados

La potencia del contraste Q para revelar errores en la modelización se ha comprobado que es baja, resultado que coincide con los datos ya publicados. Como ejemplo, citaremos que dada una realización del proceso:

$$(1 - 0,8 B) Z_t = (1 - 0,4 B) a_t$$

El contraste Q no rechazó la hipótesis de aleatoriedad para ninguno de los residuos de los modelos siguientes: AR (1), AR (2), AR (3), ARIMA (1, 1, 1), IMA (1, 1), IMA (1, 2).

Hemos calculado el coeficiente de correlación ordinal de Spearman (Lehman (1975), pp. 297-307) entre la ordenación mediante σ_a^2 y Q de los modelos estimados. El coeficiente es, en todos los casos, entre 0,7 y 0,85 lo que indica un grado apreciable de covariación como era de esperar.

La tabla 2 presenta una comparación entre el test Q y el test de rachas sobre las *fas*. En ambos casos el test de la hipótesis de aleatoriedad se ha construido con un nivel de significación de 0,05. Los datos incluidos se refieren a resultados con los 94 modelos erróneos estimados.

La tabla muestra como, aunque la potencia de ambos es baja, su utilización conjunta es mejor que la de uno solo por separado. El porcentaje de casos detectados por el test Q es del 29 por 100, el test de Rachas detecta el 23 por 100, mientras que utilizándolos conjuntamente la falta de adecuación sería detectada en el 40 por 100 de los casos.

TABLA 2

		Q		
		Aceptado	Rechazado	
R A C H A S	Aceptado	56	16	72
	Rechazado	10	12	22
		66	28	94

El contraste directo sobre los residuos ha resultado ser de utilidad mucho menor, como preveíamos. Únicamente en 13 casos (14 por 100) este contraste ha rechazado la aleatoriedad y de ellos en 11 la falta de adecuación era señalada por los contrastes de aleatoriedad.

El contraste de normalidad ha resultado ser de una utilidad muy baja. Únicamente 2 veces en las 111 simulaciones el contraste Kolmogorov Smirnov sobre la distribución de los residuos ha conducido a rechazar la hipótesis de normalidad.

En resumen, los contrastes diagnósticos estudiados (9) raramente conducen a la selección de un modelo único y en todos los casos hemos obtenido entre 4 y 8 modelos aceptable. No parece que el valor de los estadísticos evaluados pueda utilizarse razonablemente como criterio de selección.

Pasando a la varianza residual σ_a^2 , señalaremos que, de los 17 modelos, en 12 el modelo correcto fue también de el varianza mínima. En la tabla 3 presentamos un resumen de los cinco casos en que esto no ocurrió.

TABLA 3

Modelo correcto	$\hat{\sigma}_a^2$	AIC	Modelo estimado con $\hat{\sigma}_a^2$ mín.	σ_a^2	AIC
ARMA (1,1)	1.00	+6	AR (3)	.97	+2,88
IMA (1,1) \times IMA (1,1) _{1 2}	.925	-7,1	IMA (1,1) \times ARIMA (1,1,1) _{1 2}	.913	-7,29
IMA (1,1)	.941	-6,22	ARIMA (5,1,0)	.919	-2,19
IMA (1,1)	.832	-26,9	ARIMA (3,1,3)	.810	-23,4
IMA (1,1)	.741	-46,36	ARIMA (2,1,3)	.728	-41,33

La primera columna de la tabla da la estructura verdadera del modelo, la varianza residual estimada en un modelo con estructura correcta y el valor de AIC. A continuación damos la representación con varianza estimada mínima y el valor del AIC para ese modelo.

La relación que hemos demostrado entre el contraste F y el criterio de Akaike explica los resultados obtenidos. En el primer caso, los parámetros teóricos del modelo generado eran $\phi = 0,6$ y $\theta = 0,2$. Al estimar un ARMA (1,1) obtuvimos:

$$(1 - 0,34 B) Z_t = (1 - 0,01 B) a_t$$

(0,22) (0,23)

con una correlación entre estimaciones de 0,95. El AR (3) estimado fue:

$$(1 - .34B - .06B^2 + .19B^3) Z_t = a_t$$

(.08) (.08) (.08)

Los resultados son explicables si tenemos en cuenta que:

$$(1 - .6 B) Z_t = (1 - .2 B) a_t \rightarrow (1 - .4 B - .08 B^2 - 0.1 B^3) Z_t = a_t$$

En este caso el contraste de la F conduciría, análogamente al MAIC, al AR(3). En el segundo caso, donde el MAIC conduce al modelo más complejo, el test de la F hubiera escogido el modelo correcto. En los tres casos siguientes es claro bajo cualquier perspectiva que los modelos simples son superiores.

4. SELECCION AUTOMATICA DEL MODELO OPTIMO

Los resultados anteriores se han obtenido con la importante restricción de que los modelos estimados surgen de un proceso simple pero efectivo de identificación en el sentido de Box-Jenkins. Únicamente en los modelos IMA(1,1) hemos acudido a parametrizaciones excesivas para estudiar el comportamiento de las estimaciones en estos casos.

En nuestra opinión, un excelente ejemplo de los riesgos de eludir la etapa de identificación y confiar ciegamente en un criterio único de ajuste es el trabajo de Ozaki (1977). Este autor utiliza el MAIC para seleccionar de forma automática un modelo óptimo para las series utilizadas como ejemplo en Box-Jenkins (1970). El autor utiliza un programa que va probando todos los modelos posibles con (p, d, q) menores que unos órdenes prefijados. Los resultados de Ozaki están resumidos en la tabla 4, donde presentamos factorizados los polinomios AR y MA.

Para la serie A el MAIC de Ozaki es un ARMA(3,3).

Como vemos, existe redundancia de parámetros en los operadores AR y MA, lo que conduce, al simplificar, al modelo ARMA(1,1) de Box y Jenkins (1970). En situaciones como estas –cancelación de operadores– la función de verosimilitud no puede aproximarse cerca de su máximo por una función cuadrática y, en consecuencia, las varianzas y covarianzas de los estimadores están calculadas muy aproximadamente mediante la inversión de la matriz de información (hessiano).

TABLA 4

SERIE	MODELO B/J	$\hat{\sigma}_a^2$	AIC	MODELO OZAKI	$\hat{\sigma}_a^2$	AIC
A	$(1 - .92B)Z_t = (1 - .58B)a_t$	0,0977	-450,26	$(1 - .95B)(1 - 1.1B + .98B^2)Z_t = (1 - .7B)(1 - 1B + .98B^2)a_t$	0,0910	-456,16
A				$(1 - .35B - .19B^2 - .02B^3 - .03B^4 + .02B^5 - .08B^6 - .2B^7)Z_t = a_t$	0,0925	-450,89
B	$\bar{V}Z_t = a_t$	55,22	1463,5	$(1 - .9B + .98B^2)\nabla Z_t = (1 - .87B + .99B^2)a_t$	50	1453,5
C	$(1 - .82B)\nabla Z_t = a_t$	0,0181	-902,98	$(1 - .71B)(1 - 1.99B + .99B^2)Z_t = a_t$	0,0165	-915,17
D	$(1 - .87B)Z_t = a_t$	0,0903	-739,6	$(1 - 1.82B + .84B^2)\nabla Z_t = (1 - 1.99B + .99B^2)a_t$	0,0871	-746,59
E	$(1 - .36B)(1 - 1.2B + .76B^2)Z_t = a_t$	218,2	548,56	$(1 - 1.99B + .99B^2)(1 - 1.58B + .95B^2)(1 - .26B + .64B^2)Z_t = (1 - 1.86 + .93B^2)(1 - .52B)a_t$	179,2	540,87

Esta situación se agrava además cuando las raíces de los polinomios están muy próximas a la unidad, como ocurre en este caso. Esto explica el bajo valor de $\hat{\sigma}_a^2$. El modelo AR(7) estimado por Ozaki para esta misma serie, es un claro ejemplo de sobreparametrización. Los parámetros ϕ_3 , ϕ_4 , ϕ_5 y ϕ_6 no son estadísticamente significativos, y un contraste F comparando este modelo con el ARMA(1,1) no rechaza el modelo simple con $\alpha = 0,01$.

La tabla ilustra como las formulaciones del resto de las series caen en estos mismos problemas: Redundancia de parámetros (B, E) y raíces unitarias en los operadores (C, D, E). En estas condiciones, los algoritmos de estimación son muy aproximados y obtendremos una mala estimación de los coeficientes que se traducirá en mayores errores de previsión al salirnos del período muestral. Akaike (1974) ha estudiado la serie E ajustando modelos autorregresivos hasta llegar al orden 20. El criterio MAIC conduce a escoger un AR(2), lo que está en contradicción con los resultados de Ozaki. Nos preguntamos si estas diferencias en la estimación de un modelo con un mismo paquete de programas —los del Institute of Statistical Mathematics de Tokio— no son debidas a las peculiaridades del algoritmo de estimación, cuando la situación de estimación no está bien definida por las razones ya comentadas.

Además, en el trabajo al que nos referimos, Akaike analiza la misma serie, pero utilizando 176 observaciones en lugar de las 100 que contiene la serie E . El modelo escogido en este caso sí es AR(8). Posteriormente Akaike (1978 a) escoge un modelo ARMA(7,2) para esta misma serie con 176 observaciones.

En todos estos casos, el ajuste se efectuó por un procedimiento automático. Los resultados demuestran, en nuestra opinión, las ventajas del proceso de identificación recomendadas por Box y Jenkins e ilustran la importancia de una parametrización escueta como procedimiento para obtener modelos fiables y robustos. Aplicar un criterio mecánico para seleccionar el modelo óptimo para una serie temporal es un procedimiento que difícilmente puede conducir, en conjunto, a buenos resultados.

Es cierto que la concepción “artesanal” implícita en la metodología de Box-Jenkins requiere un grado apreciable de experiencia y que la aparición de nuevas herramientas teóricas que faciliten la construc-

ción de modelos razonables puede contribuir a extender las aplicaciones de los modelos de series temporales; pero, en nuestra opinión, en toda construcción de un modelo la capacidad de pensar críticamente sobre ese problema en concreto es insustituible.

5. CONCLUSIONES

La conclusión principal de este trabajo, es que no existe un estadístico único en el que podamos confiar para la selección del modelo. La varianza residual es, en nuestra opinión, y en el supuesto de que no aparecen indicios de mala especificación del modelo, el estadístico más fiable. Este criterio puede completarse con un contraste de la F , en situaciones en las que debemos elegir entre un modelo simple y otro más elaborado. El criterio de Akaike combina el valor de la varianza y la significatividad del contraste. En nuestra opinión, el criterio es atractivo al unir el problema de la identificación con la teoría de la información y el método de máxima verosimilitud. Sin embargo, desde un punto de vista práctico, su utilidad es, creemos, limitada. El objetivo central para el que se utiliza, la selección automática de modelos, nos parece poco relevante, y para decidir entre un pequeño número de modelos razonables se reduce a un contraste F con nivel de significación especificado encubiertamente.

En segundo lugar, creemos importante la búsqueda de contrastes diagnósticos más potentes. El contraste de rachas que hemos introducido en la *fás* se ha revelado como razonablemente útil por lo que recomendamos su uso. El interés principal de un contraste de este tipo es que utiliza información complementaria al contraste Q , por lo que conviene utilizarlos conjuntamente.

NOTAS

- (1) El lector puede encontrar en castellano una descripción de la modelización mediante este enfoque en Treadway (1978), o Peña (1978).
- (2) Por ejemplo el modelo $(1 - \phi B) Z_t = a_t$ resulta, cuando ϕ es positivo y próximo a 1, difícil de distinguir del modelo $\nabla Z_t = (1 - \theta B) a_t$. Sin embargo, sus funciones de previsión son sensiblemente distintas. El primero conduce a una previsión que decrece geométricamente hasta cero, mientras que el se-

gundo origina, a partir del horizonte 2, una función de previsión constante para cualquier período.

$$\hat{Z}_t(l) = \phi^l Z_t \text{ para el AR}(1)$$

$$\hat{Z}_t(l) = Z_t - \theta a_t \text{ para el IMA}(1,1)$$

- (3) Estos resultados no son independientes entre si ya que, por ejemplo, la varianza $\hat{\sigma}_a^2$ está relacionada con la significatividad de los parámetros, etc.
- (4) La estimación de los modelos ARMA mediante mínimos cuadrados es asintóticamente equivalente a la maximización de la función de verosimilitud.
- (5) Los errores estandar de los parámetros ϕ estimados están, en todos los casos, entre 0,04 y 0,1. Las correlaciones entre parámetros estimados son de orden 0,7.
- (6) Las series contienen 168 datos y el valor F obtenido es 1,0054, que debe compararse con una $F_{(2,168)}$.
- (7) El criterio de Akaike esta ya implementado en algunos programas comerciales, como en el paquete SIFT, distribuido en Inglaterra por Uniliver.
- (8) Suponemos que N es grande y, también, que la estimación incluye Backforecasting con lo que las sumas cuadráticas del numerador tienen el mismo número de términos. La estimación de la varianza está definida por:

$$\hat{\sigma}_a^2(p) = \frac{\sum a_t^2(p)}{N}$$

Siendo N el número total de datos utilizados en la estimación. Cuando N es grande y p pequeño la división por el número de grados de libertad en lugar de N conduce, prácticamente, al mismo resultado.

- (9) En este trabajo no hemos evaluado el contraste sobre el Periodograma. Su potencia, (Vease Ozaki (1977)) parece ser baja.

BIBLIOGRAFIA

- Akaike, H. (1972): *Information theory and an Extension of the Maximum likelihood principle*. In proc. and int. Symp. Information theory, supp. to Problems of Control and Information Theory. pp. 267-281.
- Akaike, H. (1974): *A New Look at the Statistical Model Identification*. IEEE Transactions on Automatic Control, Vol. AC-19, núm. 6, December, pp. 716-722.
- Akaike, H. (1976): *Canonical Correlation Analysis of time series and the use of an Information criterion*. En System Identification: Advances and case studies. Ed. R.K. Mehra y D.G. Lainiotis. Academic Press.

- Akaike, H. (1978a): *On the likelihood of a time series model*. Ponencia invitada en la 1978 Conference on Time Series Analysis and Forecasting, Cambridge University.
- Akaike, H. (1978b): *A new look at the Bayes procedure*. *Biometrika*, 65, 1, pp. 53-59.
- Anderson, T.W. (1971). *The Statistical Analysis of Time Series*. Wiley.
- Bartlett, M.S. (1978): *An Introduction to Stochastic Processes*. (Third Edition). Cambridge University Press.
- Box, G.E.P. y Muller, M. (1955): *A note on the generation of normal deviates*. *Annals of Mathematical Statistics*. XXIX. pp. 610-611.
- Box, G.E.P. y Jenkins, G.M. (1970): *Time Series Analysis, Forecasting and Control*. Holden-Day.
- Box, G.E.P. y Pierce, D.A. (1970): *Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models*. *J.A. S.A.* Vol. 55, 332, pp. 1509-1526.
- Canover, W.J. (1971): *Practical Nonparametric Statistics*. Wiley.
- Jerez, M. (1972): *Fundamentos Matemáticos de la Estadística y de la Información*. ETSII. Madrid.
- Knuth, D.E. (1969): *Seminumerical algorithms. (The art of Computer Programming, Vol. 2)*. Addison - Wesley.
- Kullback, S. y Leibler, R.A. (1951): *On Information and Sufficiency*. *Ann. Math. Statist.*, 22, 79-86.
- Kullback, S. (1959): *Information theory and Statistics*. Wiley. New York.
- Lehmann, E.L. (1975): *Nonparametric Statistics Methods based on Ranks*. Holden-Day.
- Ljung, G. y Box, G.E.P. (1978): *On a measure of lack of fit in Time Series Models*. *Biometrika*, 65, 2, pp. 297-303.
- McLaren, D. y Marsaglia, G. (1965): *Uniform Random Number Generators*. *JACM* 12, pp. 83-89.
- Newbold, P. y Granger, C.W. (1974): *Experience with forecasting univariate time series and the combination of forecasts*. *J.R. Statist. Soc. A.* 137, pp. 131-146.
- Ozaki, T. (1977): *On the order determination of ARIMA Models*. *Appl. Statist.* 26, Núm. 3, pp. 290-301.
- Peña, D. (1978): *La metodología de Box-Jenkins. Una aplicación al consumo de Gasolina*. *Información Comercial Española*, Vol. 542, pp. 135-152.

- Reza, F.M. (1961): *An Introduction to Information Theory*. McGraw-Hill. New York.
- Shibata, R. (1976): *Selection of the order of an autoregressive model by Akaike's information criterion*". *Biometrika*, 63, 1, pp. 117-26.
- Treadway, A.B. (1978): *Sobre la modelización estadística de la Balanza de Pagos española*. *Información Comercial Española*. Núm. 24, Abril. pp. 24-46.