

UN ALGORITMO ITERATIVO PARA LA ESTIMACION DE MODELOS ARMA CON AUSENCIA DE OBSERVACIONES

A. CRISTÓBAL CRISTÓBAL
Dpto. de Métodos Estadísticos
Universidad de Zaragoza

RESUMEN

En el presente artículo se muestra un algoritmo iterativo para la estimación de parámetros de modelos ARMA en series temporales que tengan alguna observación ausente. Posteriormente, se efectúa la demostración de la convergencia de dicho algoritmo. Se presenta un ejemplo de estimación basado en la simulación de series temporales con un ordenador y se exponen las conclusiones llevadas a cabo por el autor.

Palabras clave: serie temporal; estimación; algoritmo numérico.

Clasificación A.M.S.: 62M10,65U05.

ABSTRACT

In this paper, an iterative algorithm to estimate parameters of ARMA models with missing observations is given. Furthermore, the convergence of the algorithm is proved. An example based in simulation of time series with a computer and the conclusions of this technique are also presented.

keywords: time series; estimation; numerical algorithm.

Classification A.M.S.: 62M10,65U05.

1. VISION GENERAL DEL PROBLEMA

Un problema que se presenta comúnmente en el análisis de series de tiempo es aquel en el que hay determinados valores de alguna observación que están ausentes. Esta ausencia puede deberse a varios motivos, tales como que la observación no haya podido realizarse, se haya perdido su valor o bien la ausencia sea deliberada para propósitos de calibración. Existen además varias series económicas en las que las observaciones se realizan primero diariamente, luego semanalmente, mensualmente, etc., perdiéndose las observaciones intermedias.

Estas situaciones son las que conducen a hallar métodos específicos para la estimación de parámetros y el ajuste de modelos estándar a los datos.

Se han seguido varios caminos para intentar solucionar el problema.

Así, Jones (1962), Parzen (1963) y Dunsmuir y Robinson (1981) han considerado el problema en el dominio de la frecuencia, intentando estimar la función de densidad espectral.

Tan (1979), Marshall (1980) y Shaman y Tan (1982) lo tratan de resolver en el dominio del tiempo, así como Ljung (1982), que expone un planteamiento de maximización de la función de verosimilitud.

Hay también otros autores que proponen métodos especiales para su resolución. Así, Harvey y Phillips (1979) hacen un desarrollo sobre teoría de filtraje de Kalman aplicado a modelos de regresión con residuos ARMA parcialmente observados.

Orchard y Woodbury (1972) desarrollan un método basado en la resolución de sistemas no lineales equivalentes a la maximización de la verosimilitud del modelo.

En esta última línea, Dempster, Laird y Rubin (1977) aplican un algoritmo (EM) para resolver iterativamente las ecuaciones no lineales planteadas por Orchard y Woodbury. Se estudia la convergencia del algoritmo, pero Barham (1981) detecta un fallo en la demostración, e incluso Boyles (1983) elabora un contraejemplo aclaratorio.

No obstante, la convergencia del algoritmo EM es estudiada por Wu (1983), aunque se necesitan ciertas propiedades sobre la serie para que al aplicarle el algoritmo EM, éste resulte convergente. El problema radica en que las ecuaciones de maximización de la verosimilitud del modelo son altamente no lineales en los parámetros.

Damsleth (1980) presenta un algoritmo de interpolación de los

valores ausentes de una serie temporal, mediante el empleo de predicciones pasadas y futuras. Los resultados obtenidos son más pobres que con los anteriores y además la cantidad de información necesaria para llevarlo a cabo es también mayor que en aquéllos.

No obstante, se han hecho pocos intentos en evaluar y comparar los métodos propuestos para determinar cuál debe ser preferible en unas determinadas circunstancias.

La mayoría de los resultados obtenidos en los trabajos expuestos son de naturaleza asintótica, y parece necesario emplear técnicas de simulación para comparar dichos resultados.

2. UN METODO DE ESTIMACION DE LOS PARAMETROS DE MODELOS ARMA EN SERIES TEMPORALES CON VALORES AUSENTES

Sea $\{X_t\}$ una serie temporal que supondremos sigue un modelo ARMA(p, q) de media μ :

$$X_t - \mu + \sum_{i=1}^p a_i(X_{t-i} - \mu) = \varepsilon_t + \sum_{j=1}^q b_j \varepsilon_{t-j}$$

donde $\{\varepsilon_t\}$ es un proceso aleatorio puro. Para asegurar la estacionariedad de la serie, supondremos que:

1. Las raíces del polinomio:

$$\alpha(z) = 1 + \sum_i a_i z^i \quad i: 1, \dots, p$$

y las del polinomio:

$$\beta(z) = 1 + \sum_i b_i z^i \quad i: 1, \dots, q$$

están fuera de la bola unidad.

2. Los polinomios $\alpha(z)$ y $\beta(z)$ no tienen ninguna raíz común.
3. a_p, b_q son distintos de 0.

Definiendo el operador «marcha atrás», B :

$$BX_t = X_{t-1}$$

podemos representar la serie:

$$\alpha(B)(X_t - \mu) = \beta(B)\varepsilon_t$$

Usando una técnica debida a Box y Jenkins (1970) se pueden determinar los errores ε_t mediante la siguiente recurrencia:

Partiendo de las condiciones iniciales:

$$\varepsilon_1 = \varepsilon_2 = \dots = \varepsilon_p = 0$$

se calculan:

$$\begin{aligned} \varepsilon_{p+1} &= X_{p+1} - \mu + \sum_i a_i(X_{p+1-i} - \mu) \\ \varepsilon_{p+2} &= X_{p+2} - \mu + \sum_i a_i(X_{p+2-i} - \mu) - b_1\varepsilon_{p+1} \\ \varepsilon_{p+q} &= X_{p+q} - \mu + \sum_i a_i(X_{p+q-i} - \mu) - \sum_{j=1}^{q-1} b_j\varepsilon_{p+q-j} \\ \varepsilon_{p+q+1} &= \dots \\ & \quad i: 1, \dots, p \end{aligned}$$

Para simplificar la notación, llamaremos ϕ al vector de parámetros:

$$\phi^T = (a^T, b^T, \mu)$$

con

$$\begin{aligned} a^T &= (a_1, \dots, a_p) \\ b^T &= (b_1, \dots, b_q) \end{aligned}$$

Sean X_1, \dots, X_N , N observaciones de la serie anterior, entre las que distinguiremos r accesibles y $N - r$ no accesibles o perdidas.

Llamaremos:

$$T = \{t \in \{1, \dots, N\}; X_t \text{ es accesible}\}$$

Realmente sólo se dispone de r observaciones X_t con $t \in T$. A éstas les llamaremos Y_t , $t = 1, \dots, r$.

2.1. Algoritmo

Para el cálculo de ϕ se va a establecer el siguiente algoritmo iterativo:

Etapla 0: Determinación de la estimación inicial $\phi^{(0)}$. Se ha de minimizar la suma de los cuadrados de los residuos que pueden evaluarse:

$$\min H_1(\phi|Y_1, \dots, Y_r) = \min \sum_{t \in A} \varepsilon_t^2$$

con $A = \{t, \varepsilon_t \text{ no contiene ningún } X_s \text{ con } s \notin T\}$.

Es decir, se minimiza una suma de cuadrados de errores que no contienen observaciones perdidas. Estos pueden evaluarse, y la suma de sus cuadrados se minimiza siguiendo un criterio de mínimos cuadrados, resolviendo el sistema de ecuaciones normales correspondientes.

Etapla 1: Consta de dos pasos y se ha de iterar con $m \geq 1$.

Paso 1: Estimación de los valores perdidos, $X_t^{(m)}$, con $t \notin T$, minimizando la forma cuadrática:

$$H_2(X_t(t \notin T)|Y_1, \dots, Y_r, \phi^{(m-1)}) = \sum_{t \notin A} \varepsilon_t^2$$

Es decir, se toman los errores ε_t que tienen alguna observación perdida, y suponiendo conocidos los valores de los parámetros ϕ como los obtenidos en la iteración anterior, $\phi^{(m-1)}$, se evalúa la suma de cuadrados de éstos como función de las observaciones perdidas. Se aplica un criterio de mínimos cuadrados para minimizar H_2 y así se consiguen las estimaciones $X_t^{(m)}$ para las observaciones perdidas.

Paso 2: Se calculan los nuevos valores de los parámetros $\phi^{(m)}$ minimizando la forma cuadrática:

$$Q(\phi|Y_1, \dots, Y_r, X_t^{(m)} t \notin T) = \sum_{t \geq p+1} \varepsilon_t^2$$

donde $X_t^{(m)} t \notin T$ son las estimaciones de los valores perdidos, calculadas en el paso 1. Al igual que en las anteriores, se sigue un criterio de mínimos cuadrados para la minimización de Q .

Los valores obtenidos $\phi^{(m)}$ entrarán en la siguiente iteración, a no ser que:

$$\|\phi^{(m)} - \phi^{(m-1)}\| < \delta$$

para una tolerancia δ dada y una norma de R^{p+q+1} dada equivalente a la del máximo, en cuyo caso el algoritmo se detendrá, y con ello:

- $\phi^{(m)}$ será la estimación de ϕ aportada por el algoritmo.
- $X_t^{(m)}$ $t \notin T$ será la estimación de las observaciones perdidas aportada por el algoritmo.

2.2. Convergencia del algoritmo

Sea $\Omega \subseteq R^{p+q+1}$ el dominio de los posibles valores de ϕ . Vamos a construir una sucesión $\phi^{(1)}, \phi^{(2)}, \dots, \phi^{(m)}, \dots$ de valores de ϕ . Como primer paso demostraremos que el error cuadrático medio en cada iteración es menor que en la iteración anterior.

Construimos la aplicación:

$$\begin{aligned} M: \Omega \subseteq R^{p+q+1} &\longrightarrow \Omega \\ \phi &\rightarrow M(\phi) \in \Omega \end{aligned}$$

donde $M(\phi)$ es la transformación de ϕ por medio del algoritmo.

Para los valores de la sucesión:

$$\phi^{(m+1)} = M(\phi)$$

Teorema 1: Si $\phi^{(m)} \in \Omega$ y Q está definida por:

$$Q(\phi | Y_1, \dots, Y_r; X_t, t \notin T) = \sum_{t \geq p+1} \varepsilon_t^2$$

entonces:

$$\begin{aligned} Q(M(\phi^{(m)}) | Y_1, \dots, Y_r; X_t^{(m+1)}, t \notin T) &\geq \\ &\geq Q(\phi^{(m)} | Y_1, \dots, Y_r; X_t^{(m)}, t \notin T) \end{aligned}$$

es decir, en cada iteración se va haciendo más pequeño el error cuadrático.

Demostración. Por simplificar, llamaremos:

$$Q(\phi^{(m)}|Y_1, \dots, Y_r; X_t^{(s)} t \notin T) = Q(\phi^{(m)}|X_t^{(s)})$$

Vamos a demostrar que:

$$Q(\phi^{(m+1)}|X_t^{(m+1)}) \geq Q(\phi^{(m)}|X_t^{(m+1)}) \geq Q(\phi^{(m)}|X_t^{(m)})$$

y con ello tendremos demostrado el teorema.

$$\text{i) } Q(\phi^{(m+1)}|X_t^{(m+1)}) \geq Q(\phi^{(m)}|X_t^{(m+1)})$$

Es evidente, ya que $\phi^{(m+1)}$ es el valor de ϕ que minimiza $Q(\phi|X_t^{(m+1)})$.

$$\text{ii) } Q(\phi^{(m)}|X_t^{(m+1)}) \geq Q(\phi^{(m)}|X_t^{(m)})$$

Las dos formas cuadráticas se diferencian únicamente en los errores ε_t que contienen observaciones perdidas. Se podría poner:

$$Q = \sum_{t > p+1} \varepsilon_t^2 = \sum_{t \notin T} \varepsilon_t^2 + \sum_{t \in T} \varepsilon_t^2 = Q_1 + Q_2$$

Así:

$$\begin{aligned} Q(\phi^{(m)}|X_t^{(m+1)}) &= Q_1^1 + Q_2^1 = Q^1 \\ Q(\phi^{(m)}|X_t^{(m)}) &= Q_1^2 + Q_2^2 = Q^2 \end{aligned}$$

Es claro que $Q_2^1 = Q_2^2$, ya que Q_2 sólo depende de ϕ e Y_1, \dots, Y_r , observaciones accesibles. Pero $X_t^{(m+1)}$ es el valor que minimiza Q_1 (etapa 1, paso 1 del algoritmo), luego trivialmente $Q_1^1 \geq Q_1^2$ y se tiene (ii). De (i) y (ii) se concluye el teorema.

Así, según demuestra el teorema 1, la sucesión de errores cuadráticos descende con cada iteración del algoritmo.

Veamos ahora que esta sucesión tiene un mínimo:

Corolario 1: Supongamos que para algún $\phi^* \in \Omega$, y $\{X_t^*\} \in R^{N-r}$ con $t \notin T$:

$$Q(\phi^*|Y_1, \dots, Y_r; X_t^* t \notin T) \geq Q(\phi|Y_1, \dots, Y_r; X_t t \notin T)$$

$\forall \phi \in \Omega$, $\phi \neq \phi^*$ y $\forall \{X_t\} \in R^{N-r}$ con $t \notin T$ y $X_t \neq X_t^*$, $\forall t \notin T$.

Entonces:

$$Q(M(\phi^*)|Y_1, \dots, Y_r; X_t^* t \notin T) = Q(\phi^*|Y_1, \dots, Y_r; X_t^* t \notin T)$$

Demostración. Es obvia si se tiene en cuenta el teorema 1, ya que:

$$Q(M(\phi^*)|X_t^*) \geq Q(\phi^*|X_t^*)$$

Por la hipótesis del corolario 1, si $\phi = M(\phi^*)$, se tendrá:

$$Q(\phi^*|X_t^*) \geq Q(M(\phi^*)|X_t^*)$$

Luego de ambas desigualdades:

$$Q(\phi^*|X_t^*) = Q(M(\phi^*)|X_t^*)$$

con lo que se concluye la demostración.

Corolario 2: Supongamos que para algún $\phi^* \in \Omega$, y $\{X_t^*\} \in R^{N-r}$ con $t \notin T$.

$$Q(\phi^*|X_t^*) < Q(\phi|X_t)$$

$\forall \phi \in \Omega$, $\phi \neq \phi^*$ y $\forall \{X_t\} \in R^{N-r}$ con $t \notin T$ y $X_t \neq X_t^*$, $\forall t \notin T$.

Entonces: $M(\phi^*) = \phi^*$.

Demostración. Se sigue del corolario 1, pues:

$$Q(M(\phi^*)|X_t^*) = Q(\phi^*|X_t^*) < Q(\phi|X_t)$$

Con los corolarios 1 y 2 se ha visto que el valor de ϕ que minimiza la forma cuadrática, si existe, es el punto fijo de M , y límite de la sucesión $\{\phi^{(m)}\}$ con $m \geq 1$. Nótese que Q es una suma de cuadrados, luego $Q > 0$ está acotada inferiormente, y la función $Q(\phi^{(m+1)}|\dots) \leq Q(\phi^{(m)}|\dots)$ es monótona no creciente, luego tiene un límite q :

$$q = \lim Q(\phi^{(m)}|\dots) \text{ con } m \rightarrow \infty$$

No hay pérdida de generalidad al suponer que el valor de los parámetros es finito. Con ello. $\|\phi^{(m)}\| \leq k$ para una cota k apropiada. La sucesión $\{\phi^{(m)}\}$ por lo tanto tendrá una subsucesión convergente:

$$\phi^{(m_k)} \rightarrow \bar{\phi} \text{ con } k \rightarrow \infty$$

Al ser Q continua (suma de cuadrados de funciones polinómicas de ϕ):

$$\lim Q(\phi^{(m_k)}) = Q(\bar{\phi}) \text{ con } k \rightarrow \infty$$

pero, $Q(\phi^{(m)}) \rightarrow q$ y $Q(\phi^{(m_k)})$ es una subsucesión de $Q(\phi^{(m)})$, luego por la unicidad de límites:

$$\Omega(\phi^{(m)}) \rightarrow q = \Omega(\bar{\phi})$$

Con todo ello:

Teorema 2: Sea $M: \Omega \subseteq R^{p+q+1} \rightarrow \Omega$, y consideremos una norma con la que R^{p+q+1} sea completo. Sea $\phi^* \in \Omega$ tal que $M(\phi^*) = \phi^*$, punto fijo de M . Entonces, existe una bola $B(\phi^*, r) \subseteq \Omega$, tal que:

$$\begin{aligned} \forall \phi^{(m)} \in B(\phi^*, r), \\ \|M(\phi^{(m)}) - M(\phi^*)\| \leq L \|\phi^{(m)} - \phi^*\| \end{aligned}$$

con $L < 1$.

Demostración. Por lo que anteriormente se ha dicho, $\phi^* = M(\phi^*)$ es tal que $Q(\phi^*)$ es el límite de la sucesión $\{Q(\phi^{(m)})\}$ con $m \geq 1$. Dado que Q es continua y no lineal, y que posee un mínimo local, ϕ^* , $\exists r > 0$ tal que:

$$\frac{\partial^2 Q}{\partial \phi^2} > 0 \quad \forall \phi \in B(\phi^*, r)$$

o sea que la función es convexa en una $B(\phi^*, r)$. Así se deduce fácilmente que la tesis. Si tenemos un $\phi^{(m)} \in B(\phi^*, r)$, por el teorema 1, $Q(\phi^{(m+1)}) \leq Q(\phi^{(m)})$ y por la convexidad de Q en $B(\phi^*, r)$:

$$\|\phi^{(m+1)} - \phi^*\| < \|\phi^{(m)} - \phi^*\|$$

Al ser $\phi^* = M(\phi^*)$ y $\phi^{(m+1)} = M(\phi^{(m)})$ se deduce la tesis:

$$\|M(\phi^{(m)}) - M(\phi^*)\| \ll L \|\phi^{(m)} - \phi^*\|$$

Luego, con todo ello:

Corolario 3: La sucesión de aproximaciones sucesivas $\{\phi^{(m)}\}$ construida mediante el algoritmo expuesto, es convergente y tiene por límite ϕ^* , que verifica:

$$Q(\phi^{(m)}) \rightarrow Q(\phi^*)$$

y además $\phi^* = M(\phi^*)$.

Demostración. Es consecuencia inmediata a los Teoremas 1 y 2 de los Corolarios 1 y 2.

3. EJEMPLO

Mediante un ordenador se ha simulado un ejemplo de estimación comparada del parámetro de una serie temporal $AR(1)$ de media 0:

$$X(t) + aX(t - 1) = \varepsilon_t$$

empleando el algoritmo aquí expuesto y el algoritmo EM . Los residuos $\{\varepsilon_t\}$ se han tomado normales tipificados e independientes. El proceso de simulación ha sido el siguiente: Para tres valores del parámetro a : 0.5, 0.9 y 0.99 se han generado 100 series temporales $AR(1)$ de media 0 y extensión $N = 52$. Con cada valor del parámetro y para cada serie en particular se han simulado tres situaciones: la primera, donde se ha supuesto un 30 % de observaciones perdidas, la segunda, donde se ha supuesto un 20 % y la última donde se han supuesto perdidas el 10 % de las observaciones. Las pérdidas se han simulado mediante generación de números aleatorios enteros uniformes entre 1 y 52. Los números generados coincidirán con la situación de la observación perdida dentro del orden de la serie.

Una vez determinada la serie, el valor del parámetro y las ausencias, se ha aplicado, por una parte, el algoritmo aquí expuesto para estimar el valor de a , AEST, en cada iteración, y por otra parte, el algoritmo EM , utilizando como estimación inicial del parámetro la aportada en la iteración inicial por nuestro algoritmo. De esta manera, las condiciones iniciales de los dos algoritmos son equivalentes y se salva el sesgo que se puede introducir con una estimación inicial a voluntad. También con el algoritmo EM se calcula la estimación del valor de a , AESTEM, en cada iteración.

Con las dos estimaciones, y para cada iteración, calculamos el error:

$$ERR1 = (AEST - a) ** 2$$

$$ERR2 = (AESTEM - a) ** 2$$

Repetiendo el proceso para las 100 series y para los tres valores del parámetro a , se expresan a continuación las tablas de los errores medios por iteración en las tres situaciones de pérdidas expresadas anteriormente:

La tolerancia fijada para la convergencia de los errores dados es de 1.E-4.

Como puede verse en las tablas, independientemente del número de ausencias, siempre que el valor del parámetro es superior a 0.9, nuestro

algoritmo es preferible al algoritmo *EM*, ya que la precisión es de un orden mayor, 1.E-3, frente a 1.E-2 del algoritmo *EM*.

En estos casos, el parámetro está cerca de la frontera de estacionariedad, lo que hace que los algoritmos convencionales tengan un error mayor que en casos menos conflictivos para el parámetro. Sin embargo, en nuestro algoritmo el estimador de los valores ausentes es una combinación de los valores observados, que además, en el caso que nos ocupa, un modelo autorregresivo, dicha combinación es lineal, lo que provoca un efecto de media móvil en la estimación de las ausencias, suavizando la serie en cada iteración y así, obteniendo unas estimaciones más precisas, en casos que, a priori, puede parecer que vaya a haber problemas con la convergencia del algoritmo.

Cuando el valor del parámetro es $a = 0.5$, los dos algoritmos tienen comportamientos similares, comportándose ligeramente mejor el algoritmo *EM*, cuando el número de ausencias es mayor (Tabla 1), y al contrario cuando el número de ausencias desciende (Tablas 2 y 3). Los

TABLA 1
1.ª situación, con 30 % de ausencias

E R R O R E S						
IT	PAR = 0.5		PAR = 0.9		PAR = 0.99	
	ALG.	EM.	ALG.	EM.	ALG.	EM.
0	0.0821	0.0821	0.0187	0.0187	0.0115	0.0115
1	0.0557	0.0485	0.0067	0.0219	0.0067	0.0166
2	0.0465	0.0386	0.0062	0.0169	0.0064	0.0166
3	0.0426	0.0351	0.0061	0.0138	0.0063	0.0167
4	0.0407	0.0335		0.0122	0.0062	0.0167
5	0.0397	0.0322		0.0115		0.0167
6	0.0391	0.0322		0.0113		
7	0.0387	0.0320				
8	0.0384	0.0319				
9	0.0382					
10	0.0381					

TABLA 2
2.^a situación, con 20 % de ausencias

E R R O R E S						
IT	PAR = 0.5		PAR = 0.9		PAR = 0.99	
	ALG.	EM.	ALG.	EM.	ALG.	EM.
0	0.0314	0.0314	0.0146	0.0146	0.0075	0.0075
1	0.0226	0.0212	0.0073	0.0111	0.0043	0.0144
2	0.0211	0.0221	0.0070	0.0111	0.0042	0.0147
3	0.0206	0.0220	0.0070		0.0042	0.0147
4	0.0203	0.0219	0.0069			
5	0.0201	0.0218	0.0068			
6	0.0200	0.0217				
7	0.0199	0.0217				
8	0.0198	0.0217				
9	0.0198	0.0216				
10	0.0197					

TABLA 3
3.^a situación, con 10 % de ausencias

E R R O R E S						
IT	PAR = 0.5		PAR = 0.9		PAR = 0.99	
	ALG.	EM.	ALG.	EM.	ALG.	EM.
0	0.0189	0.0189	0.0104	0.0104	0.0068	0.0068
1	0.0177	0.0217	0.0074	0.0129	0.0058	0.0152
2	0.0176	0.0217	0.0072	0.0128	0.0053	0.0152
3	0.0176	0.0216	0.0072	0.0128	0.0053	0.0152
4		0.0216			0.0053	

dos algoritmos presentan un orden de convergencia de los errores de 1.E-2.

Por lo hasta aquí expuesto, puede resumirse que ambos algoritmos pueden elegirse indistintamente cuando el valor del parámetro no vaya a plantear problemas de aproximación a la no estacionariedad, pudiéndose utilizar incluso el algoritmo *EM* cuando el número de ausencias sea notable (superior al 25 %). Tienen ambos un orden de convergencia de 1.E-2. Pero en casos en que el valor del parámetro esté cerca de la frontera de estacionariedad será preferible utilizar nuestro algoritmo, pues tiene un orden de convergencia 1.E-3 un grado mayor que el del algoritmo *EM*, que sigue siendo de 1.E-2. Esta elección deberá hacerse independientemente del número de ausencias.

4. CONCLUSIONES

El algoritmo presentado es muy simple y fácilmente implementable en un ordenador, tal y como puede verse en Cristóbal (1985).

La eficacia de la estimación inicial depende en distinta medida en los modelos AR que en los MA y ARMA. En los primeros, la bondad de la estimación inicial depende del número de ausencias. Conforme aumenta éste, la estimación inicial es menos buena (puede comprobarse en el ejemplo). En los otros, debido al método recurrente explicado en el punto 2, para el cálculo de los residuos ε_t , si X_k es una ausencia, no se pueden evaluar los ε_t desde $t = k$ hasta $t = N$, por lo que la estimación inicial será mejor conforme más tarde aparezcan las ausencias. Si la primera ausencia aparece muy pronto, se deberá cortar la serie y hacerla comenzar más tarde, para que exista un número de residuos tal que la estimación inicial de los parámetros no quede muy desvirtuada. Posteriormente se incorporarán los valores iniciales a la serie y se estimarán las ausencias intermedias aplicando el algoritmo.

El algoritmo presentado tiene algunas similitudes con el algoritmo *EM*. La filosofía de este último se basa en la maximización de la verosimilitud, mientras que la de aquél persigue minimizar una función de errores residuales. Si estos errores son gaussianos, los dos algoritmos deben proporcionar estimaciones parecidas en casos no conflictivos, es decir, siempre que no haya peligro por la estacionariedad y siempre que el número de ausencias no sea muy grande.

Pero la aplicación del algoritmo *EM* a las series de medias móviles

MA, o mixtas *ARMA*, incluye dos problemas: uno, la dificultad de evaluación de la función de verosimilitud, problema que se agrava conforme mayor sea el número de ausencias y que también hace difícilmente implementable la idea de Ljung (1982), y otro, que, aunque sea evaluable la función de verosimilitud, es posible que no se pueda realizar el paso de maximización del algoritmo (paso *M*). Estos dos problemas condujeron a Dempster, Laird y Rubin (1977) a definir un nuevo algoritmo, al que llamaron GEM (*EM* generalizado), como una generalización del anterior, de manera que se aumentara la verosimilitud esperada en cada iteración, pero que como Barham (1981) y posteriormente Boyles (1983) expusieron, este esquema no es siempre convergente, aunque en la mayoría de los casos en donde se ha empleado, los resultados han sido satisfactorios.

Las propiedades de convergencia del algoritmo *EM* y del GEM han sido estudiadas por Wu (1983), que ha demostrado teoremas de convergencia del algoritmo para determinadas formas de la función de verosimilitud, basados en condiciones necesarias, que el propio Wu reconoce que son muy difíciles de comprobar para una serie en particular.

A estos dos problemas ha de unirse la dificultad de la implementación del algoritmo en un ordenador. De los métodos resumidos, el algoritmo *EM* es el menos eficaz, computacionalmente hablando (Anderson (1981)).

Además, existe el problema de la estimación inicial. Para aplicar el algoritmo *EM*, debe partirse de una estimación inicial dada por el operador.

Estos cuatro problemas quedan reducidos en la aplicación del algoritmo aquí presentado.

Primero, la evaluación de la función de error mediante los residuos ε_t es muy sencilla en cualquier caso. Segundo, la minimización de dicha función de error no plantea problemas de convergencia al mínimo. Tercero, el algoritmo es fácilmente implementable y es eficaz (en el ejemplo del punto 3, el tiempo de ejecución del algoritmo *EM*, en media, y es de esperar que es el caso *ARMA* general, esa proporción quede reducida por las dificultades antes expresadas en la implementación del algoritmo *EM*). Cuarto, el algoritmo aquí presentado aporta una estimación inicial, que además es muy eficiente en una gran parte de casos, cuando el número de ausencias no es grande o bien los parámetros están cerca de la frontera de estacionariedad.

Por todas estas razones, parece preferible emplear este algoritmo como alternativa al *EM*, y más en los casos en los que se ha dicho que hay una preferencia mayor, por alguna característica de la serie original.

Una variación del algoritmo *EM* puede ser la empleada en el ejemplo del punto 3, es decir, para evitar el sesgo de la introducción de una estimación inicial por parte del operador, puede aprovecharse la etapa 0 del algoritmo presentado para aportar una estimación inicial coherente, y hacer que funcione el algoritmo *EM* con dicha estimación inicial. Como se ha podido ver, dicha variación es eficaz cuando el número de ausencias es grande y no hay problemas de estacionariedad con los parámetros.

BIBLIOGRAFIA

- ANDERSON, O. D. (1982): *Time Series Analysis*, North Holland.
- BARHAM, S. Y. (1981): «The EM algorithm», *M. Sc. Diss.*, University of Wales.
- BARHAM, S. Y., y DUNSTAN, F. D. J. (1982): «Missing values in Time Series», *Time Series Analysis: Theory and Practice 2*, Ed. O. Anderson, North Holland.
- BOX, G. E. P., y JENKINS, G. M. (1970): *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco.
- BOYLES, R. A. (1983): «On the convergence of the EM algorithm», *J. Roy Statist. Soc.*, B-45 No. 1, 47-50.
- CRISTOBAL, A. (1985): *Análisis de Series Temporales en ausencia de observaciones. Estimación de parámetros*. Tesina de Licenciatura, Universidad de Zaragoza.
- DAMSLETH, E. (1980): «Interpolating missing values in a Time Series», *Schand. J. Statist.*, 7, 33-39.
- DEMPSTER, A. P.; LAIRD, N. M., y RUBIN, D. B. (1977): «Maximum likelihood estimation for incomplete data, via the EM algorithm», *J. Roy. Statist. Soc.*, B-39, 1-38.
- DUNSMUIR W., y ROBINSON, P. M. (1981): «Parametric estimators for Time Series with missing observations», *Adv. Appl. Prob.*, 13, 129-146.
- HARVEY A. C., y PHILLIPS G. D. A. (1979): «Maximum likelihood estimation of regression models with autoregressive-moving average disturbances», *Biometrika*, 66, 49-58.

- JONES R. H. (1962): «Spectral analysis with regularly missed observations», *Ann. Math. Stat.*, 33, 455-461.
- LJUNG (1982): *Biometrika*, 69, 265-268.
- MARSHALL, R. J. (1980): «Autocorrelación estimation of Time Series with randomly missing observations», *Biometrika*, 67, 567-571.
- ORCHARD T., y WOODBURY, M. A. (1972): «A missing information principle: Theory and applications», *Proceeding of the 6th. Berkeley Symposium on Mathematical Statistics and Probability*, 1, 697-715.
- PARZEN, E. (1963): «On Spectral Analysis with missing observations and amplitude modulation», *Sankhya*, A-25, 383-392.
- PRIESTLEY M. B. (1981): *Spectral Analysis and Time Series*, Academic Press, London.
- ROBINSON, P. M. (1980): «Estimation and Forecasting for Time Series containing censored or missing observations». *Time Series: Proceedings of the International Conference at Nottingham University*, in March 1979. Ed. O. Anderson, North Holland, 167-171.
- SAKAI, H.; SOEDA, T., y TOKUMARU, H. (1979): «On the relations between fitting autoregressions and periodograms with applications», *Ann. Statist*, 7, 96-107.
- SHAMAN, P., y TAN, S. B. (1982): «The mean square error of one step ahead predictions for an autoregressive process when the coefficients are estimated with missing data», *Time Series Analysis: Theory and Practice*, 1, 53-63. Ed. O. Anderson, North Holland.
- TAN, S. B. (1979): «Maximum likelihood estimation in autoregressive processes with missing observations», *Ph. D. Thesis*, University of Pittsburgh.
- WU, C. F. J. (1983): «On the convergence properties of the EM algorithm», *Ann. Statist*, 11, N. 1, 95-103.