

# DISCRIMINACIÓN ENTRE MODELOS DE SERIES TEMPORALES MEDIANTE LA UTILIZACIÓN DE UNA DISTANCIA. APLICACIÓN A LA IDENTIFICACIÓN AUTOMÁTICA

MOISÈS VALLS I COLOM Y ALBERT PRAT BARTES  
UNIVERSITAT POLITÈCNICA DE BARCELONA

*Un problema que presenta la metodología Box-Jenkins es que de las tres fases de que consta: identificación, estimación y verificación, únicamente la estimación está automatizada y por tanto la identificación y la verificación han de ser realizadas por el operador. El objetivo de este trabajo es precisamente automatizar eficientemente la fase de identificación de forma que en general el modelo ya sea válido al verificarlo. La automatización se consigue a través de una distancia entre las autocorrelaciones y autocorrelaciones inversas de la serie temporal y las de los posibles modelos a identificar. Esta distancia es fácilmente extensible a otras similares de gran interés: Se ha definido una distancia entre las autocorrelaciones de dos intervalos de una serie temporal y se podría definir también una distancia entre autocorrelaciones que midiera la influencia de cada observación. Además, los conceptos introducidos en el estudio son aplicables a la identificación no automática de series temporales, de forma que su contenido puede ser de gran utilidad tanto para automatizar como para ayudar al operador a identificar modelos.*

DISCRIMINATION BETWEEN TIME SERIES MODEL BY USING DISTANCES,  
APPLICATION TO AUTOMATIC IDENTIFICATION

Keywords: TIME SERIES, ARMA MODELS, AUTOMATIC IDENTIFICATION, DISTANCES,  
MODEL DISCRIMINATION.

## 1. INTRODUCCIÓN.

El procedimiento Box-Jenkins a diferencia de los métodos clásicos de previsión presenta la particularidad de que no está totalmente automatizado. Este hecho representa en muchos casos una ventaja, pues el trabajo de un identificador experimentado operando interactivamente con el ordenador produce generalmente buenos resultados. Sin embargo, hay ocasiones en que sería preferible disponer de un modelador automático de series<sup>1</sup>. Las ventajas que se obtendrían serían múltiples:

a) Posibilidad de modelar gran cantidad de series temporales en poco tiempo: Supóngase, por ejemplo, que una empresa quiere hacer previsiones de las ventas de cada uno de sus 5.000 productos o que un importante banco quiere tratar la serie de recursos ajenos de cada una de sus sucursales.

El esfuerzo humano que se requeriría sería extraordinario si no se dispusiese de un modelador automático de series temporales.

b) Posibilidad de tratar series sin disponer de personal suficientemente instruido.

c) Conseguir un mayor rigor en la elección de los modelos a estimar. En algunas ocasiones, aún poseyendo una gran experiencia, se identifican series que presentan una estructura muy parecida a la del modelo identificado, pero con valores incompatibles con los de las autocorrelaciones o autocorrelaciones parciales del modelo. En cambio al automatizar, el ordenador detecta con facilidad en la fase de identificación si el modelo es o no adecuado para la serie temporal.

- Moisés Valls i Colom - Catedra d'Estadística Teòrica i Aplicada de E.T.S.E.I.B. de la Universitat Politècnica de Barcelona - Av. Diagonal, 647 - Barcelona

- Albert Prat i Bartés - Catedra d'Estadística Teòrica i Aplicada de E.T.S.E.I.B. de la Universitat Politècnica de Barcelona - Av. Diagonal, 647 - Barcelona.

- Article rebut el Octubre del 1983.

d) Más rapidez en el tratamiento de series temporales: Actualmente, modelar una serie temporal supone primeramente, ejecutar el programa de identificación, en segundo lugar analizar los resultados y procesar el programa de estimación, en tercer lugar analizar los nuevos resultados y en caso de no ser satisfactorios volver a empezar el ciclo identificando otro modelo. Finalmente, una vez superada la validación del modelo, ejecutar el programa de previsión. En cambio, con el tratamiento automático de series temporales, sólo haría falta procesar un único programa.

En general los métodos de modelado automático de series temporales aparecidos hasta el momento, desprecian en cierta forma la etapa de identificación e intentan encontrar un modelo aceptable a base de muchas iteraciones que no siempre convergen hacia el modelo óptimo. Este trabajo constituye una respuesta a este inconveniente proporcionando un método de identificación que selecciona modelos con suficientes garantías de buen funcionamiento.

## 2. LA FUNCIÓN DE AUTOCORRELACIÓN INVERSA.

### 2.1. INTRODUCCIÓN.

La identificación de modelos se basa habitualmente en el análisis de las funciones de autocorrelación (ACF) y de autocorrelación parcial (PACF) de la serie temporal. Se ha creído conveniente por múltiples razones, substituir la PACF por la IACF (función de autocorrelación inversa) de la serie.

La función de autocorrelación inversa (IACF) de una serie temporal fué definida por Cleveland /4/ como la función de autocorrelación asociada al recíproco del espectro de la serie. Cleveland demostró la gran importancia que tiene la autocorrelación inversa como instrumento para identificar y estimar modelos.

Chatefield /3/ presentó una definición alternativa a la de Cleveland en el dominio del tiempo y amplió algunos aspectos referentes a la autocorrelación inversa. Posteriormente, Hosking /5/ ha encontrado la distribución asintótica de la autocorrelación inversa mues-

tral.

### 2.2 DEFINICIÓN DE LA AUTOCORRELACIÓN INVERSA EN EL DOMINIO TEMPORAL.

Si  $\Gamma(B) = \sum_{k=-\infty}^{\infty} \gamma_k B^k$  es la función generatriz de las autocovarianzas de un proceso  $\{x_t\}$ , las autocovarianzas inversas  $\gamma_{inv,k}$  de  $\{x_t\}$  serán las de la función generatriz de autocovarianzas inversas  $\Gamma_{inv}(B)$  que se define como

$$\Gamma(B) \cdot \Gamma_{inv}(B) = 1$$

La autocorrelación inversa  $\rho_{inv,k}$  queda pues definida por

$$\rho_{inv,k} = \frac{\gamma_{inv,k}}{\gamma_{inv,0}}$$

La principal consecuencia de la definición de la IACF de un modelo ARMA es la siguiente:

Siendo  $\{x_t\}$  un proceso ARMA (p,q)

$$\phi(B)x_t = \theta(B)a_t$$

y  $\{y_t\}$  el proceso ARMA (q,p)

$$\theta(B)y_t = \phi(B)a_t$$

la autocorrelación inversa de  $\{x_t\}$  será idéntica a la autocorrelación de  $\{y_t\}$  y viceversa.

## 3. DISTANCIA ENTRE FUNCIONES DE AUTOCORRELACIÓN Y AUTOCORRELACIÓN INVERSA.

### 3.1. DISTANCIA ENTRE FUNCIONES DE AUTOCORRELACIÓN.

Sea  $(x_t)$  una serie temporal perteneciente a un modelo

$$\phi(B)x_t = \theta(B)s_t \quad (3.1)$$

Se define  $\xi$  como el vector de dimensión L:

$$\xi = \sqrt{N} \begin{bmatrix} r_1 - \rho_1 \\ r_2 - \rho_2 \\ \vdots \\ r_L - \rho_L \end{bmatrix}$$

donde  $r_j$  y  $\rho_j$  son respectivamente las  $j$ -ésimas autocorrelaciones de la serie  $\{x_t\}$  y del modelo (3.1.) y  $N$  la longitud de la serie.-- T.W. Anderson /1/ demostró que  $\xi$  sigue una - distribución límite de probabilidad  $N(Q, \underline{W})$ , - siendo

$$\underline{O} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{y} \quad \underline{W} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1L} \\ w_{12} & w_{22} & \cdots & w_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ w_{1L} & w_{2L} & \cdots & w_{LL} \end{bmatrix}$$

de término general

$$w_{ij} = \sum_{k=-\infty}^{\infty} (\rho_k \rho_{k+i-j} + \rho_k \rho_{k+i+j} + 2\rho_i \rho_j \rho_k^2 - 2\rho_k \rho_i \rho_{k+j} - 2\rho_k \rho_j \rho_{k+i}) \quad (3.2)$$

Por ser  $\underline{W}$  definida positiva, existe una única matriz triangular  $\underline{V}$  con todos los elementos - de la diagonal positivos, tal que

$$\underline{W} = \underline{V} \cdot \underline{V}'$$

Premultiplicado  $\xi$  por  $\underline{V}^{-1}$ , se obtiene que

$$\underline{F} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_L \end{bmatrix} = \underline{V}^{-1} \xi$$

sigue una distribución asintótica

$$N(\underline{O}, \underline{V}^{-1} \underline{W} (\underline{V}^{-1})')$$

$$\text{pero } \underline{V}^{-1} \underline{W} (\underline{V}^{-1})' = \underline{V}^{-1} \underline{V} \underline{V}' (\underline{V}^{-1})' = \underline{V}^{-1} \underline{V} (\underline{V}^{-1} \underline{V})' = \underline{I}$$

Si se define

$$D_L = \sum_{k=1}^L f_k^2$$

como que  $f_k$  son variables aleatorias que asintóticamente son IIN(0,1),  $D_L$  tendrá una distribución límite de probabilidad de  $\chi^2$  con  $L$  grados de libertad.

$$\text{Por otra parte } D_L = \underline{F}' \underline{F} = (\underline{V}^{-1} \xi)' (\underline{V}^{-1} \xi) = \xi' (\underline{V}^{-1})' \underline{V}^{-1} \xi = \xi' (\underline{V} \underline{V}')^{-1} \xi = \xi' \underline{W}^{-1} \xi$$

ha resultado ser la distancia de Mahalanobis

El hecho de seguir una distribución de probabilidad conocida permite probar con un riesgo  $\alpha$  si una serie temporal  $\{x_t\}$  pertenece a un

determinado modelo

$$\phi(B)X_t = \theta(B)a_t$$

Al plantear la prueba de significación

$$H_0 : D_L(\rho_{-1}, \rho_{-2}) = 0$$

$$H_1 : D_L(\rho_{-1}, \rho_{-2}) > 0$$

siendo  $\rho_{-1}$  y  $\rho_{-2}$  los vectores de autocorrelaciones teóricas de la serie y del modelo respectivamente, si  $H_0$  es cierta, con una confianza  $1-\alpha$  se cumplirá:

$$D_L < \chi_{1-\alpha, L}^2$$

y por tanto dependerá de que  $D_L$  sea mayor o menor que  $\chi_{1-\alpha, L}^2$  para rechazar o no  $H_0$ .

Además como  $D_L$  tiene en cuenta la interdependencia entre las autocorrelaciones discriminará eficazmente entre los diversos modelos.

Aunque en realidad  $D_L$  es el cuadrado de una distancia, se respetará su denominación habitual de distancia de Mahalanobis

### 3.2. DISTANCIA ENTRE FUNCIONES DE AUTOCORRELACIÓN INVERSA.

Análogamente, se puede definir  $\xi_{inv}$  como

$$\xi_{inv} = \sqrt{N} \begin{bmatrix} r_{inv,1} - \rho_{inv,1} \\ r_{inv,2} - \rho_{inv,2} \\ \vdots \\ r_{inv,L} - \rho_{inv,L} \end{bmatrix}$$

donde  $r_{inv,j}$  y  $\rho_{inv,j}$  son respectivamente -- las  $j$  - ésimas autocorrelaciones inversas de la serie  $\{x_t\}$  y del modelo (3.1) y  $N$  la longitud de la serie.

Hosking /5/ demostró que  $\xi_{inv}$  sigue una distribución asintótica normal de media cero y matriz de variancias covariancias  $\underline{W}_{inv}$  de -- término general

$$w_{inv,ij} = \sum_{k=0}^{\infty} (\rho_{inv,k} \rho_{inv,k+i-j} + \rho_{inv,k} \rho_{inv,k+i+j} + 2\rho_{inv,i} \rho_{inv,j} \rho_{inv,k}^2 - 2\rho_{inv,i} \rho_{inv,k} \rho_{inv,k+j} - 2\rho_{inv,j} \rho_{inv,k} \rho_{inv,k+i}) \quad (3.3)$$

Se puede por tanto definir la distancia

$$D_{inv,L} = \xi'_{inv} W^{-1}_{inv} \xi_{inv}$$

que seguirá asintóticamente una ley de  $\chi^2$  -- con L grados de libertad.

Evidentemente  $D_{inv,L}$  gozará de las mismas -- propiedades que tenía  $D_L$ .

### 3.3. DISTANCIA CONJUNTA.

Si un determinado modelo (3.1) cumple simul-  
táneamente

$$D_L < \chi^2_{1-\alpha,L} \quad \text{y} \quad D_{inv,L} < \chi^2_{1-\alpha,L} \quad (3.4)$$

se considerará adecuado para la serie tempo-  
ral  $\{x_t\}$ . En caso de que más de un modelo -  
con el mismo número de parámetros cumpliera  
(3.4) se escogería el que proporcionase dis-  
tancias inferiores. Sin embargo, podría suce-  
der que

$$D_{L_A} < D_{L_B} \quad \text{y} \quad D_{inv,L_A} > D_{inv,L_B}$$

siendo

$$D_{L_A}, D_{inv,L_A} \quad \text{y} \quad D_{L_B}, D_{inv,L_B}$$

las distancias a los modelos A y B, con lo -  
que en este caso no sabríamos que modelo ---  
identificar.

Para estos casos habría que intentar hallar  
una combinación de  $D_L$  y  $D_{inv,L}$  que sirviese  
de criterio a minimizar, para lo cual se ne-  
cesitaría conocer la distribución conjunta -  
de las dos distancias, pues de otro modo se  
desconocería la distribución del nuevo esta-  
dístico y no se podría garantizar su eficien-  
cia.

Si con  $D_L$  y  $D_{inv,L}$  no podemos identificar un  
único modelo, intentaremos descartar el má-  
ximo número de modelos posible con ambas dis-  
tancias y definir el vector  $\xi^{(c)}$  para cada -  
uno de los restantes modelos, donde

$$\xi^{(c)} = \sqrt{N} \begin{bmatrix} r_1 - \rho_1 \\ r_2 - \rho_2 \\ \vdots \\ r_L - \rho_L \\ r'_{inv,1} - \rho_{inv,1} \\ r'_{inv,2} - \rho_{inv,2} \\ \vdots \\ r'_{inv,L} - \rho_{inv,L} \end{bmatrix}$$

Las autocorrelaciones inversas  $r'_{inv,j}$  se ob-  
tendrán de la forma siguiente

Sea

$$\psi(B)x_t = a_t$$

uno de los modelos a comparar, donde

$$\psi(B) = \sum_{K=0}^{\infty} \psi_K B^K$$

Su modelo dual será

$$y_t = \psi^2(B)a_t$$

y por tanto

$$y_t = \psi(B)x_t$$

con lo cual una vez obtenida la serie  $\{y_t\}$  a  
partir de  $\{x_t\}$  podremos calcular sus autocor-  
relaciones que coincidirán con las autocorre-  
laciones inversas de  $\{x_t\}$

El vector  $\xi^{(c)}$  seguirá una distribución lími-  
te de probabilidad  $N(0, \underline{W}^{(c)})$ , donde

$$\underline{W}^{(c)} = \begin{bmatrix} \underline{W} & \underline{M} \\ \underline{M}' & \underline{W}_{inv} \end{bmatrix}$$

$\underline{W}$  y  $\underline{W}_{inv}$  son las matrices de variancias cova-  
riancias de  $\xi$  y  $\xi_{inv}$  respectivamente, y  $\underline{M}$  la  
matriz de covariancias entre  $\xi$  y  $\xi_{inv}$ .

Podemos pues definir

$$D_L^{(c)} = \xi^{(c)'} (\underline{W}^{(c)})^{-1} \xi^{(c)}$$

que seguirá asintóticamente una ley de  $\chi^2$  --  
con 2L grados de libertad.

Como que  $\underline{W}$  y  $\underline{W}_{inv}$  son conocidas, el problema queda reducido a encontrar la matriz de covariancias  $\underline{M}$ .

Valls /7/ demostró que el término general de la matriz  $\underline{M}$  es de la forma

$$M_{i,j} = \frac{\delta_{ij} + 2\rho_i \rho_{inv,j}}{\gamma_o \gamma_{inv,o}} \quad (3.5)$$

donde

$$\delta_{ij} = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

### 3.4. ESTIMACIÓN DE LA MATRIZ DE VARIANCIAS COVARIANCIAS.

Como que las autocorrelaciones  $r_h$  y  $r_{inv,h}$  tienden asintóticamente hacia  $\rho_h$  y  $\rho_{inv,h}$  se podría substituir en las expresiones (3.2) y (3.3) las autocorrelaciones y autocorrelaciones inversas teóricas del modelo por las obtenidas en la serie temporal, con lo que las matrices de variancias covariancias  $\underline{W}$  y  $\underline{W}_{inv}$  serían únicas para todos los modelos.

Las distancias así definidas tendrían pues la forma

$$D_L = \underline{\xi}' \underline{W}^{-1} \underline{\xi} \quad \text{y} \quad D_{inv,L} = \underline{\xi}'_{inv} \underline{W}_{inv}^{-1} \underline{\xi}_{inv}$$

Como  $\underline{W}$  y  $\underline{W}_{inv}$  tienen que ser definidas positivas, convendría utilizar algún procedimiento de estimación que garantizase su invertibilidad /2/.

### 3.5. DISTANCIA ENTRE AUTOCORRELACIONES DE UNA SERIE TEMPORAL.

Sean  $\underline{r}^{(1)}$  y  $\underline{r}^{(2)}$  los vectores de autocorrelaciones de una serie temporal en dos intervalos de tiempo disjuntos y de longitud n. Como que cada vector tiene una distribución límite de probabilidad conocida

$$\underline{r} \sim N \left( \underline{\rho}, \frac{1}{n} \underline{W} \right)$$

la diferencia de ambos vectores tendrá también una distribución asintótica conocida y fácilmente calculable

$$\underline{r}^{(1)} - \underline{r}^{(2)} \sim N \left( \underline{0}, \frac{2}{n} \underline{W} \right)$$

ya que los vectores  $\underline{r}^{(1)}$  y  $\underline{r}^{(2)}$  son independientes entre sí.

Para conocer si las autocorrelaciones de la serie se han mantenido constantes a lo largo del tiempo, se escogerían dos intervalos disjuntos de observaciones de longitud n suficientemente grande y se calcularía la distancia

$$D_{aut} = \frac{n}{2} \left[ \underline{r}^{(1)} - \underline{r}^{(2)} \right]' \underline{W}^{-1} \left[ \underline{r}^{(1)} - \underline{r}^{(2)} \right]$$

que seguirá una ley de  $\chi^2$  con L grados de libertad, siendo L la dimensión de los vectores  $\underline{r}$ .

Como que en general el modelo de la serie será desconocido, la matriz de variancias covariancias se estimará a partir de las autocorrelaciones de la serie.

Análogamente, se podría definir una distancia entre las autocorrelaciones inversas de la serie.

Estas distancias nos permitirán pues comprobar si las autocorrelaciones y autocorrelaciones inversas de la serie han permanecido constantes a lo largo del tiempo. En caso de que hubiesen variado, convendría modelar únicamente las últimas observaciones de la serie una vez comprobada su invariabilidad.

## 4. IDENTIFICACIÓN AUTOMÁTICA DE SERIES ESTACIONARIAS.

### 4.1. INTRODUCCIÓN

Los procesos estacionarios son de interés debido a que son identificables y a que son fácilmente extensibles a procesos no estacionarios ARIMA (p,d,q) y ARIMA (p,d,q) (P,D,Q)S. Los procedimientos que se definirán a continuación intentan detectar mediante la utilización de las distancias definidas en el apartado anterior el modelo que presenta una estructura más parecida a la de la serie previamente estacionarizada y conteniendo el menor número de parámetros posible. Para estacionarizar las series temporales se podrían utilizar los criterios de diferenciación definidos por Valls /7/ basados en la dualidad

existente entre las funciones de autocorrelación y de autocorrelación inversa de una serie temporal.

Si el modelo contiene un término constante  $\theta_0$ , como que no afecta ni a las autocorrelaciones ni a las autocorrelaciones inversas no se tendrá en cuenta en la etapa de identificación, considerándolo únicamente al estimar el modelo.

4.2. SERIES TEMPORALES ESTACIONARIAS NO ESTACIONALES.

4.2.1. PROCESO DE IDENTIFICACIÓN

Sea  $\{x_t\}$  una serie temporal estacionaria y no estacional de modelo

$$\phi(B)x_t = \theta(B)a_t \tag{4.1}$$

Con un cierto riesgo  $\alpha$ , la distancia de la serie  $\{x_t\}$  al modelo (4.1) cumplirá

$$D_L < \chi^2_{1-\alpha, L}$$

y exactamente igual con las autocorrelaciones inversas.

Para probar si la serie  $\{x_t\}$  sigue un determinado tipo de modelo, habrá que estimar los parámetros, comparar las funciones de autocorrelación (ACF) y autocorrelación inversa (IACF) de la serie con las del modelo y en el caso de que las distancias sean suficientemente pequeñas, aceptar el modelo estimado. Las distribuciones de referencia de  $D_L$  y  $D_{inv, L}$  una vez estimado el modelo serán una  $\chi^2$  con  $L-p$  y  $L-q$  grados de libertad siendo  $p$  y  $q$  el número de parámetros estimados con las autocorrelaciones y autocorrelaciones inversas respectivamente. Dado un cierto valor de  $k=p+q$  el modelo de  $k$  parámetros más apropiado para  $\{x_t\}$  será el que minimice simultáneamente  $D_L$  y  $D_{inv, L}$  y si además cumple

$$D_L < \chi^2_{1-\alpha, L} \text{ y } D_{inv, L} < \chi^2_{1-\alpha, L}$$

se aceptará el modelo.

En caso de que los modelos que minimizan  $D_L$  y  $D_{inv, L}$  no coincidan se identificará el que proporcione una distancia conjunta  $D_L^{(c)}$  menor. El proceso de identificación automática de series temporales estacionarias y no esta-

cionales será pues el indicado en la Fig.1.

Sabemos que las ACF y IACF de un modelo cumplen las ecuaciones:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_{p-1} \rho_{k-p+1} + \phi_p \rho_{k-p} \text{ para } k > q \tag{4.2}$$

$$\rho_{inv, k} = \theta_1 \rho_{inv, k-1} + \theta_2 \rho_{inv, k-2} + \dots + \theta_{q-1} \rho_{inv, k-q+1} + \theta_q \rho_{inv, k-q} \text{ para } k > p \tag{4.3}$$

Substituyendo las ACF y IACF del modelo por las de la serie y con  $k$  tomando valores desde  $q+1$  hasta  $q+p$  en la primera ecuación y desde  $p+1$  hasta  $p+q$  en la segunda, se formará un sistema de  $p+q$  ecuaciones con  $p+q$  incógnitas fácilmente resoluble.

4.2.2. SIMULACIÓN

Se han simulado 80 series temporales de 800 observaciones cada una, correspondientes a 6 tipos de modelos diferentes de 1, 2 y 3 parámetros. Para ahorrar tiempo de ordenador se han utilizado las ecuaciones (4.2) y (4.3) para estimar los modelos. Las autocorrelaciones inversas se han estimado con un nuevo procedimiento propuesto por Valls /7/. Las matrices de variancias covariancias  $\underline{W}$  y  $\underline{W}_{inv}$  se han obtenido a partir de las autocorrelaciones y autocorrelaciones inversas teóricas de los modelos estimados en 45 series y a partir de las autocorrelaciones y autocorrelaciones inversas de las series en las 35 restantes. Evidentemente el modelo identificado depende del riesgo de la prueba: Para cada riesgo se identificará un modelo según el proceso 4.2.1.

Los resultados con  $\alpha=0'05$  y  $\alpha=0'001$  han sido los indicados en las Tablas T1 y T2.

Con  $\alpha=0'05$  de las 45 series en que se ha utilizado la matriz de variancias covariancias teórica  $\underline{W}$ , 39 han resultado estar bien identificadas (86'67%), mientras que las 35 restantes en que se ha estimado  $\underline{W}$ , se han identificado 34 correctamente (97'14%).

Para dar idea de la magnitud de las distancias, se han calculado los promedios de las distancias de las series del modelo MA(2) a los diferentes modelos de 1 y 2 parámetros:

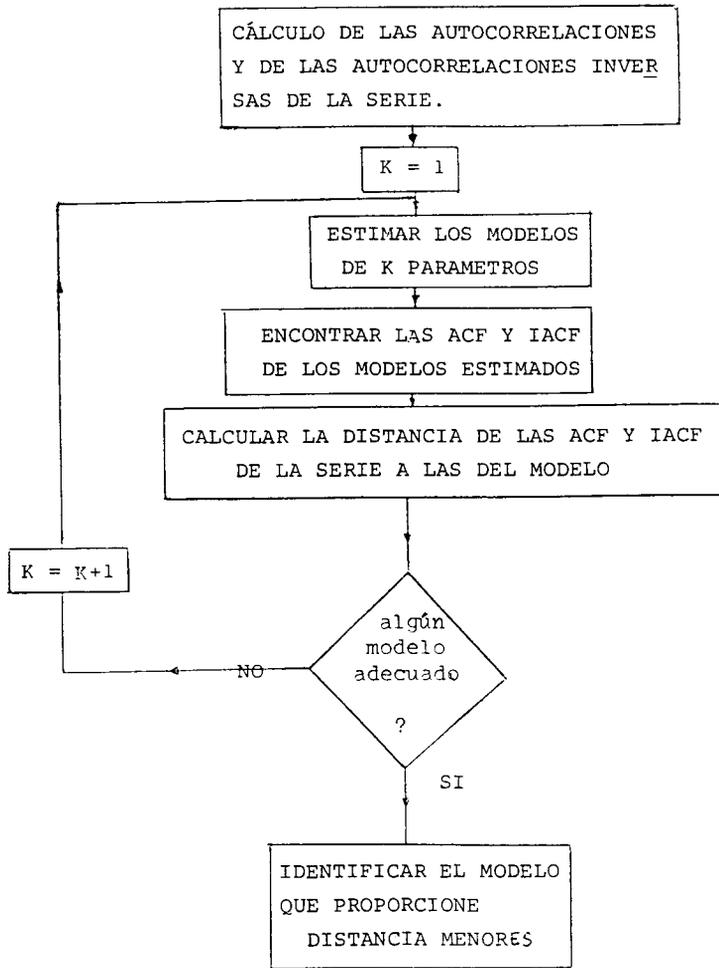


Figura 1. Proceso de identificación de series estacionarias no estacionales.

Tabla T1.  $\alpha = 0'05$  : 91'25% de modelos bien identificados

		MODELO IDENTIFICADO						
		AR(1)	MA(1)	AR(2)	ARMA(1,1)	MA(2)	AR(3)	OTROS
MODELO SIMULADO	AR(1)	17	2					1
	MA(1)		20					
	AR(2)			6			2	1
	ARMA(1,1)				9			1
	MA(2)					11		
	AR(3)						10	

Tabla T2.  $\alpha = 0'001$  : 97'5% de modelos bien identificados

		MODELO IDENTIFICADO						
		AR(1)	MA(1)	AR(2)	ARMA(1,1)	MA(2)	AR(3)	OTROS
MODELO SIMULADO	AR(1)	18	2					
	MA(1)		20					
	AR(2)			9				
	ARMA(1,1)				10			
	MA(2)					11		
	AR(3)						10	

Tabla T3. Promedio de las distancias de las series de modelo MA(2)

	D	D <sub>inv</sub>
MA(1)	102'86	286'45
AR(1)	150'08	215'87
MA(2)	1'72	0'57
ARMA(1,1)	943'33	1910'64
AR(2)	57'18	108'84

En promedio las distancias al modelo MA(2) - son muy inferiores al resto.

Con  $\alpha=0'001$  han aparecido únicamente 2 series mal identificadas que han resultado tener el mismo modelo.

$$(1+0'1B) x_t = a_t$$

habiéndose identificado como MA(1). Sin embargo, con  $N=800$ , los dos modelos son prácticamente equivalentes.

Estos resultados permiten confirmar la eficacia del procedimiento de identificación propuesto en 4.2.1, lo cual comporta las siguientes implicaciones:

- + Las distancias discriminan bien entre modelos y admiten la distribución de referencia externa propuesta.
- + El método de estimación propuesto en 4.2.1 proporciona buenos estimadores de los parámetros del modelo.

+ La aproximación de la matriz de variancias covariancias estimada a partir de las propias autocorrelaciones de la serie parece ser muy productiva, pues el porcentaje de series bien estimadas con  $\alpha=0.05$  ha superado en un 10% el correspondiente a las series en que se ha utilizado la W teórica.

#### 4.3. SERIES TEMPORALES ESTACIONARIAS ESTACIONALES.

##### 4.3.1. PROCESO COMPLETO DE IDENTIFICACIÓN

Dada una cierta serie temporal estacionaria y estacional  $\{x_t\}$ , se quiere identificar el modelo del tipo

$$\phi(B)\phi_s(B)x_t = \theta(B)\theta_s(B)a_t$$

que mejor se le adapte. De este modelo en principio únicamente se conoce la estacionaridad  $s$ .

Un posible proceso de identificación sería - el propuesto en la Fig. 2.

Este método requiere bastante cálculo de ordenador, pues se tendría que ensayar

$$\frac{K_o^2 + 3K_o + 2}{2} \quad \frac{K_s^2 + 3K_s + 2}{2}$$

modelos para cada serie temporal.

##### 4.3.2. PROCESO DE IDENTIFICACIÓN SIMPLIFICADO.

Peña /6/ demostró que la función de autocorrelación de un proceso estacional viene dado por la expresión

$$\rho_j = \frac{T_j + \sum_{i=1}^{\infty} R_{si} (T_{j+si} + T_{si-j})}{1 + 2 \sum_{i=1}^{\infty} R_{si} T_{si}}$$

donde  $T_j$  y  $R_j$  son respectivamente los coeficientes de autocorrelación teóricos de las partes regular y estacional de dicho proceso.

Efectuando las 3 hipótesis:

- 1) Estacionalidad multiplicativa
- 2)  $T_k \approx 0$  para todo  $k > s/2$
- 3)  $T_{inv,k} \approx 0$  para todo  $k > s/2$

se cumplirán 3 importantes propiedades para las autocorrelaciones

- A)  $\rho_j \approx T_j \quad j < s/2$
- B)  $\rho_{hs} \approx R_{hs}$
- C)  $\rho_{hs+k} = \rho_{hs-k} \approx R_{hs} T_k \quad k < s-q$

y análogas propiedades para las autocorrelaciones inversas.

Las propiedades A y B permiten obtener en forma aproximada las primeras autocorrelaciones de la parte regular y todas las correspondientes a la parte estacional.

Por tanto, con las primeras autocorrelaciones y autocorrelaciones inversas se puede identificar la parte no estacional del modelo, mientras que las correspondientes a la parte estacional pueden ser consideradas como las de un modelo no estacional que una vez identificado, puede ser ya formulado como estacional.

De acuerdo con la primera hipótesis, los dos modelos identificados pueden superponerse y formar un único modelo del tipo

$$\phi(B)\phi_s(B)x_t = \theta(B)\theta_s(B)a_t$$

#### 5. CONCLUSIONES.

El procedimiento de identificación automático utilizado pretende identificar el modelo B/J que mejor se adapte a la serie teniendo el menor número de parámetros posible. Para series de longitud suficientemente grande lo consigue al 100%, ya que el procedimiento es asintóticamente óptimo.

Para series de longitud finita, los resultados también son excelentes: 97,5% de series bien identificadas.

El resultado del procedimiento de identificación automática será únicamente el modelo seccionado y tendrá aun de superar las etapas de estimación y verificación antes de hacer previsiones, habiendo sin embargo muchas probabilidades de que el modelo inicial sea ya válido al verificarlo.

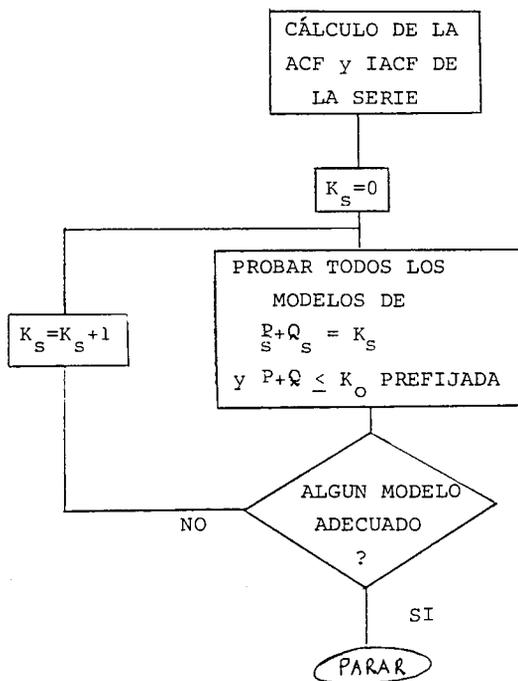


Figura 2. Proceso completo de identificación de series estacionarias estacionales.

En caso de no utilizar el procedimiento automático, las distancias a los diferentes modelos servirán para que el operador identifique con mayores garantías de éxito.

Con el procedimiento propuesto se consigue también más rigor al seleccionar el modelo: En algunas ocasiones se identifican modelos que presentan una estructura similar a la de la serie pero con valores de las autocorrelaciones o autocorrelaciones inversas incompatibles con el modelo.

Sea por ejemplo, el modelo simulado

$$(1-0,5B)^3 X_t = a_t$$

que ha producido las autocorrelaciones y autocorrelaciones inversas indicadas en la Tabla T4.

Analizando las ACF y IACF, el experto segura-

mente identificaría un modelo AR(2). Las distancias a los diferentes modelos con 4 autocorrelaciones comparadas son las de la tabla T5.

$$\text{Como que } D(\text{AR}(2)) = 40,10 > \chi_{2,0,95}^2 = 5,991$$

$$\text{y } D_{\text{inv}}(\text{AR}(2)) = 13,365 > \chi_{4,0,95}^2 = 9,488$$

se rechaza el modelo AR(2). En cambio para el modelo AR(3):

$$1,81 < \chi_{1,0,95}^2 = 3,841 \text{ y } 2,829 < \chi_{4,0,95}^2 = 9,488$$

y por tanto, con el procedimiento propuesto se identificaría el modelo teórico simulado AR(3).

En este caso, el procedimiento de identificación automática seguramente hubiese superado al operador más experimentado. Quizás en algunas ocasiones el procedimiento propuesto será inferior a la identificación clásica,

Tabla T4 . ACF y IACF de la serie simulada

K	1	2	3	4
$r_k$	. 8984	. 6966	. 4875	. 3202
S.E. ( $r_k$ )	. 0058	. 0171	. 0282	. 0364
$r_{inv,k}$	-. 6995	. 2207	-. 0135	-. 0063
S.E. ( $r_{inv,k}$ )	. 0111	. 0270	. 0324	. 0328

Tabla T5. Distancias a los diferentes modelos

		D	D <sub>inv</sub>
MA	(1)	840,71	607,51
AR	(1)	722,63	211,85
MA	(2)	375,48	356,01
ARMA	(1,1)	149,20	110,85
AR	(2)	40,10	13,36
MA	(3)	177,47	164,63
ARMA	(1,2)	75,08	65,40
ARMA	(2,1)	38,89	32,08
AR	(3)	1,81	2,83

pero las diferencias son en general largamente compensadas por la rapidez y automatización de la identificación.

En resumen, las aportaciones introducidas -- han permitido construir un procedimiento de identificación automática de modelos asintóticamente óptimo y han contribuido considerablemente a mejorar el criterio del operador al identificar un modelo. En la práctica, -- los resultados han evidenciado la gran utilidad de las aportaciones introducidas.

## 6. BIBLIOGRAFÍA

/1/ ANDERSON, T.W. "The statistical analysis of time series" -Wiley (1971).  
 /2/ BURG, J.P. "Maximum entropy spectral -- analysis" -Ph D Thesis Geophysics department. Standfor University (1975).

/3/ CHATFIELD, C. "Inverse autocorrelations" J.R.S.S. A 142, 363-377 (1979).  
 /4/ CLEVELAND, W.S. "The inverse autocorrelations of a time series and their applications" Technometrics, 14, 277-298 (1972).  
 /5/ HOSKING, J.R.M. "The asymptotic distribution of the sample inverse autocorrelations of an autoregressive-moving average process" 67, 1, 223-226 (1980).  
 /6/ PEÑA, D. "Interacción en la identificación de modelos ARIMA univariantes" E.O.I. de Madrid. Cuadernos económicos ICE (1979)  
 /7/ VALLS, M. "Identificació automàtica de sèries temporals" - Tesis Doctoral E.T.S.E.I. de Barcelona (1983).

Z. NOTAS.

- (1) En el presente trabajo sólo se trata de la automatización de la fase de identificación.