

MODELADO DE SERIES TEMPORALES CON MÉTODOS EN BLOQUE Y RECURSIVOS. DESARROLLO DE ESTIMADORES Y PREDICTORES ADAPTATIVOS

D. DE LA FUENTE Y D.F. GARCÍA

Universidad de Oviedo

En este artículo se presenta un análisis comparativo entre los algoritmos más interesantes para la estimación de parámetros de series temporales, tanto en bloque como recursivos. Se propone que los modelos autorregresivos largos constituyen una parametrización general para modelizar series inestables, cuyos parámetros pueden estimarse adecuadamente con algoritmos recursivos, tales como los filtros celosía.

**Modeling of time-series with batch and recursive methods.
Development of adaptative estimators and predictors.**

Keywords: Time series; parameter-estimation methods; autorregressive vectors; adaptative estimation.

1. INTRODUCCIÓN

Previsión es la proyección formalizada del futuro de un modelo de hechos pasados reales. Del amplio campo de los métodos de series temporales en el contexto de la previsión, en este trabajo se analizará una interesante alternativa a los métodos Box-Jenkins que se conoce como previsión adaptativa.

Los objetivos en el análisis de series temporales en cuanto a la previsión adaptativa, podemos describirlos de la siguiente forma, a saber: Previsión y Control.

—David De La Fuente García - E.T.S.I.I. - Universitat d'Oviedo - Dept. d'Administració d'Empreses i Comptabilitat.

—Daniel Fernando García Martínez - E.T.S.I.I. - Universitat d'Oviedo - Dept. d'Ingeniería Elèctrica, Electrónica de Computadors i Sistemes - Carretera de Castiello, s/n. - 33204-Gijón.

—Article rebut el desembre de 1988.

El análisis clásico, aunque de interés permanente, en series temporales ha tenido como estudio básico, la descomposición de la serie en sus cuatro componentes habituales: tendencia, variaciones cíclicas, variaciones estacionales y variaciones accidentales o erráticas, estudiando cada componente por separado e integrando los resultados, para conocer la evolución global de la serie.

Además de la previsión, un objetivo importante en el análisis de series temporales, es el control. Para alcanzar este objetivo, es preciso disponer de un modelo paramétrico dinámico del tipo entrada-salida.

La aproximación clásica intenta cubrir uno de los fines del análisis que es la previsión, pero por su propia naturaleza, no permite abordar el otro aspecto de la teoría de sistemas que es la supervisión o control.

La aproximación frecuencial al análisis de series temporales presenta problemas conceptuales y de estimación eficiente del espectro, así como la utilización de ventanas ad-hoc. Precisamente el análisis temporal que aquí abordamos, con base en modelos autorregresivos largos, permite resolver estas dificultades.

Por tanto, nuestra elección del dominio temporal, resulta ser además un enfoque eficiente para un análisis correcto de un sistema, en el que el dominio frecuencial de indudable interés en muchas áreas de conocimiento, como procesamiento de señales, descomposición de series, cancelación de ruido, etc. La contribución de Box-Jenkins, hizo posible la aplicación práctica de las ideas iniciales de Yule (1927), Slutsky (1937), Mann y Wall (1943), Quenoville (1937) y otros, a la solución de problemas en previsión y control haciendo obsoletas técnicas que venían siendo habituales en previsión tales como alisado o promedios móviles, etc.

Posteriormente a esa época se han producido varias alternativas para mejorar esta metodología, que es restringida al exigir que las series sean estables e invertibles y no permitir análisis en tiempo real. Estas alternativas son el punto de comienzo de nuestro trabajo.

Existen otros tipos de predicciones no cuantitativas como el método Delphy, en cuentas de mercado, analogía con los ciclos de vida, etc., que resultan bastante subjetivas.

Otra gran familia de técnicas predictivas la constituyen los modelos econométricos, cuya estructura viene dada "a priori" supuestamente derivada de la teoría económica, mientras que un modelo de series temporales se infiere directamente desde los datos, dentro de una clase general de modelos que permite reproducir relaciones dinámicas como las que empíricamente se observan en series.

Es obvio que ambas propiedades (ajustarse a la evidencia empírica y ser consistente con la teoría económica) no deben ser antagónicas, sino simultáneas. Ambas son deseables. Sin embargo, la econometría clásica desde el punto de vista de la predicción, ha sido víctima de dos problemas. Primero, al orientarse

a modelos de ecuaciones simultáneas con gran número de variables (a veces miles), la especificación de la dinámica ha sido relegada a segundo término. En modelos grandes, sólo es posible hoy por hoy incorporar los esquemas dinámicos más simples y rudimentarios. Segundo, si el modelo económico viene, en principio, determinado por la teoría económica, no es de extrañar que su estructura dinámica sea débil; la teoría económica nada nos dice sobre la dinámica de la economía. Ver Maravall (1981).

Como consecuencia de esta insuficiencia en la especificación dinámica, los grandes modelos econométricos son de una pobreza predictiva considerable. En definitiva, la predicción es una extrapolación de la relación entre el presente y el pasado hacia la relación entre el futuro y el presente. Si la primera relación no está capturada, la predicción será poco fiable.

Es en este contexto, en nuestra opinión, donde se encuentra una de las razones básicas de la popularidad alcanzada por los modelos de series temporales. De la estructura dinámica de los datos (reflejada por ejemplo, en la función de autocorrelación de la serie), se infiere un modelo que genera movimientos semejantes a los observados. En la medida en que son modelos específicamente diseñados para capturar relaciones dinámicas, resultan muy adecuados para la predicción. Una virtud complementaria que presentan es que son modelos generalmente pequeños, de gran eficiencia computacional y facilidad de uso. Se han convertido, pues, en un instrumento fundamental en la elaboración de predicciones económicas. Su limitación, de todos modos, surge de dos terrenos: Por una parte, el tamaño de los modelos no permite el tratamiento conjunto de muchas variables (no es posible, por mucho que se simplifique su parametrización, utilizar modelos con, por ejemplo, 100 variables). Por otra parte, a veces no se puede a partir del modelo de series temporales recuperar los parámetros que son de interés para el economista (por ejemplo, separar parámetros de una función de oferta de los de una función de demanda), lo cual puede limitar su uso en problemas de política económica.

En relación con este último punto, tiene interés un último comentario general. Es importante integrar el análisis de series temporales con los modelos econométricos, y algunos pasos lentos se están dando en esta dirección (ver, por ejemplo, Sims (1980) y Litterman (1982)). Es sabido que, un modelo econométrico dinámico estructural tiene una forma reducida que puede verse como un modelo multivariante de series temporales del tipo ARIMA (Zellner-Pal, 1974). Se conoce también que un modelo multivariante ARIMA implica modelos ARIMA univariantes para las distintas series (ver Maravall, 1981). Lo que resulta más difícil es inferir la forma estructural del modelo econométrico a partir de los modelos ARIMA. En general, una infinidad de formas estructurales serían compatibles con un modelo ARIMA determinado. Estos pueden verse, pues, como formas reducidas, robustas frente a la especificación de la estructura, particularmente aptas para su uso en la predicción. Es de esperar

que llegará un día en que modelos grandes, de fácil manejo, bien especificados y directamente interpretables en términos de teoría económica, y que reproduzcan bien la evolución de los datos en el tiempo, estarán disponibles. Hasta que llegue ese "lejano" día, los modelos de series temporales seguirán cumpliendo una función importante e insustituible.

2. TIPOS DE MODELOS Y PARAMETRIZACIONES

Para el modelado y la simulación con suficiente generalidad de un vector estocástico, resulta imprescindible especificar una parametrización que sea suficientemente amplia así como la utilización de un generador de ruido adecuado. Esta sección se dedica a establecer la relación entre la parametrización AR y otras al uso.

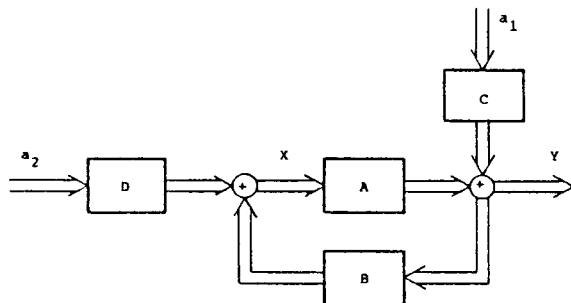


FIGURA 1. Transformaciones de modelos de ESTADO, ARMA o MA en AR

Si el vector temporal $Y(t)$ admite un modelo lineal, una forma de parametrización canónica debe recoger los modelos de los ruidos a la entrada y a la salida, la función de transferencia de planta y la realimentación tal y como esquematizamos en la figura 1 donde:

Y : Vector de series de salida.

X : Vector de series de entrada.

A : Matriz de polinomios de transformada en z ó L de la planta.

B : Matriz de polinomios de transformada en z ó L de la realimentación.

C : Matriz de polinomios de transformada en z ó L del ruido de medida.

D : Matriz de polinomios de transformada en z ó L del polinomio de observación.

Bajo este esquema y considerando X , Y , a_1 y a_2 vectores aleatorios multi-variables, podemos expresar las siguientes relaciones:

$$(1) \quad Y = A \cdot X + C \cdot a_1$$

$$(2) \quad X = B \cdot Y + D \cdot a_2$$

que se pueden escribir como

$$(3) \quad Y = A \cdot (B \cdot Y + D \cdot a_2) + C \cdot a_1$$

o bien:

$$(4) \quad Y \cdot (I - A \cdot B) = A \cdot D \cdot a_2 + C \cdot a_1$$

Análogamente:

$$(5) \quad X = B \cdot (A \cdot X + C \cdot a_1) + D \cdot a_2$$

$$(6) \quad X \cdot (I - B \cdot A) = B \cdot C \cdot a_1 + D \cdot a_2$$

Y así las ecuaciones matriciales se escriben del siguiente modo:

$$(7) \quad \begin{bmatrix} I - B \cdot A & 0 \\ 0 & I - A \cdot B \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} B \cdot C & D \\ C & A \cdot D \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

La fórmula (7) es la conocida representación ARMA. De (7) se pueden obtener parametrizaciones alternativas. El modelo Autorregresivo se obtiene premultiplicando (1) por C^{-1} y (2) por D^{-1} , resultando:

$$(8) \quad C^{-1} \cdot Y - C^{-1} \cdot A \cdot X = a_1$$

$$(9) \quad D^{-1} \cdot X - D^{-1} \cdot B \cdot Y = a_2$$

Que es representado en (10) en forma matricial.

$$(10) \quad \begin{bmatrix} -C^{-1} \cdot A & C^{-1} \\ D^{-1} & -D^{-1} \cdot B \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

Procediendo de forma similar se pueden obtener las parametrizaciones en Media Móvil y de Estados.

Los modelos autorregresivos largos, también llamados VAR (de vectores autorregresivos) utilizan ecuaciones en diferencias estocásticas lineales, en las que

cada elemento de un vector de variables es regresado sobre sus propios valores pasados y sobre los valores pasados de cada uno de las otras variables del sistema. A este tipo de modelado se le puede incorporar información a priori.

La principal debilidad de esta especificación y la razón por la que en el pasado no ha sido muy utilizada, es que el número de parámetros libres crece cuadráticamente con el número de variables del sistema. Para un sistema de orden moderado, el modelo se hace altamente sobreparametrizado, pero este inconveniente puede considerarse soslayable, pues para algunos casos de series no estables puede ser la única forma de tratarlas gracias a los llamados filtros celosía (en inglés Lattice o Ladder), siendo esta una de las ideas que motivaron el estudio.

En el vector autorregresivo, el modelo estadístico subyacente, VAR, es un sistema lineal dinámico, con un vector Y de $(n \times 1)$ salidas, generadas por ecuaciones de diferencias estocásticas. El vector Y puede incluir tanto variables endógenas como exógenas.

Si no se considera realimentación, caso habitual en series temporales, al proponer un modelo en forma diferencial para una sola serie temporal, existen tres posibles parametrizaciones:

$$(11) \quad Y(t) = a(t) + \Theta(1)a(t-1) + \dots \quad \text{MA}(\infty)$$

$$(12) \quad y(t) = a(t) + \varphi(1)Y(t-1) + \dots \quad \text{AR}(\infty)$$

$$(13) \quad \varphi(B) \cdot Y(t) = \Theta(B) \cdot a(t) \quad \text{ARMA}(p, q)$$

La primera de ellas es la conocida descomposición de Wold (1954), quien la desarrolló para el caso multivariable, se conoce por modelo en medias móviles.

La segunda representación es semejante a un modelo de regresión sobre valores pasados, con la diferencia de que el modelo considerado aquí es dinámico, y el modelo de regresión habitual tiende a ser estático. Si las variables $Y(t-1)$, $Y(t-2)$, ..., se combinasen con otras variables explicativas $X(1)$, $X(2)$, ..., tendríamos un modelo de función de transferencia causal.

Tal y como se entendía en la literatura convencional, los modelos de regresión estáticos para estimar relaciones han sido duramente criticados en los últimos tiempos (Akaike (1976), Box y Neubold (1971) y Hernández (1975)) entre otros, debido a la mala especificación de la autocorrelación entre series y la falta de detección de la realimentación.

3. MÉTODOS RECURSIVOS Y MÉTODOS EN BLOQUE

Antes de tratar este punto vamos a justificar por qué el trabajo se realiza sobre series temporales discretas, y no continuas. En primer lugar, porque se evi-

tan las dificultades que aparecen en el análisis de series estocásticas continuas; estas dificultades son debidas al problema que surge al aplicar la teoría de integración clásica a variables aleatorias. En términos generales, la diferenciación y la integración implican continuidad o predictibilidad sobre intervalos de tiempo corto, mientras que la aleatoriedad implica lo contrario. Así, en el análisis discreto las integrales son reemplazadas por sumas discretas y las diferenciales por diferencias. En segundo lugar porque cualquier análisis serio en series temporales, implica una gran cantidad de cálculos, lo que únicamente se puede resolver con un ordenador que es una máquina digital. Finalmente porque la mayor parte de los datos económicos son discretos.

Por otra parte debemos distinguir dos tipos de procesamiento de la información; uno sería el llamado en tiempo real (on-line) y otro sería en tiempo no real (off-line). Cada uno de ellos tiene su justificación según la aplicación que se realice.

El análisis de series temporales trata de conseguir un modelo de la realidad de un proceso (caja negra) que depende de varias variables que se pueden medir a su salida y a su entrada o sólo a su salida y que denominaremos identificación MIMO (multi input/multi output). Por supuesto existen casos más sencillos de una entrada y una salida SISO (simple input/simple output) y otros casos mixtos.

En los casos reales, tenemos que hablar de fenómenos estocásticos donde existe una componente que es el ruido o parte no detectada que intentamos determinar o separar de la señal verdadera. El método más habitual que se utiliza para conseguir esto, es el método de mínimos cuadrados, en el que se minimiza una función cuadrática del error, definido como diferencia de los valores que nosotros estimamos con nuestro modelo y los datos reales. Cuanto más pequeño sea ese error, más nos acercamos a la realidad.

El método de predicción mínimo cuadrático se determina por el estadístico de segundo orden del proceso observado, o sea, la varianza. En la mayor parte de los casos este estadístico no es conocido a priori y tiene que ser estimado desde los datos. Si el proceso es estable y tenemos muchos datos, es posible estimar la secuencia de autocorrelación y calcular el predictor mínimo cuadrático para todos los casos.

En la práctica, las señales son a menudo no estables, teniendo valores estadísticos variables en el tiempo. En tales situaciones, el proceso de estimar el estadístico de segundo orden para determinar el error mínimo cuadrático y calcular los parámetros del modelo del predictor, necesita ser realizado muy a menudo, para seguir los cambios de estos estadísticos. Este problema de predicción de una serie temporal, sin conocimiento de sus estadísticos, lo denominaremos predicción adaptativa.

Las técnicas de procesamiento adaptativo pueden ser clasificadas en dos tipos. En el primero, se hace el procesamiento de los datos en bloque. En el segundo, previsión recursiva, los parámetros u otra información, se van estimando y revisando conforme se dispone de cada nuevo dato.

4. MÉTODOS EN BLOQUE

A continuación se describen varios algoritmos (que trabajan con todos los datos a la vez) para estimar los parámetros de los modelos. Se analizan desde los más sencillos a los más complejos, presentando los diversos resultados obtenidos.

4.1. MÉTODO DE LOS MÍNIMOS CUADRADOS

Seguidamente se realiza un análisis de la Estimación Mínimo Cuadrática Ordinaria. De las especificaciones AR, MA, ARMA y espacios de estados, partiremos de la primera, ya que como veremos, las demás se reducen a ella, vamos a considerar el caso en que no tengamos realimentación.

El modelo del que partimos, es un modelo lineal AR, o sea:

$$(14) \quad Y(k) = \sum_{i=1}^{\infty} a(i) Y(k-i) + W(k)$$

$W(k)$ es el ruido del sistema que se podría considerar de media cero y varianza 1, y que llamaremos ruido blanco, cuando su espectro es plano, o las funciones de autocorrelación y autocorrelación parcial estén perfectamente dentro de los límites de las bandas de confianza y no tengan estructura. Realmente como no podemos tener un orden infinito en i para expresar un modelo, hay que truncarlo a un orden finito, llegando a una representación:

$$(15) \quad Y(k) = \sum_{i=1}^n a(i) Y(k-i) + W(k)$$

siendo $W(k)$ el ruido blanco inaccesible y $a(i)$ los parámetros del modelo. K indica el instante considerado.

El estimador mínimo cuadrático ordinario para los parámetros $a(i)$ de este modelo es:

$$(16) \quad \hat{a}(r, i) = (\underline{Y}' \cdot \underline{Y})^{-1} \cdot \underline{Y}'$$

donde $\hat{a}(r, i)$ es el vector estimado de $a(i)$ por medio de r observaciones, y la matriz \underline{Y} , tiene la siguiente expresión:

$$(17) \quad y = \begin{bmatrix} Y(k-1) & \dots & Y(k-n) \\ Y(k-2) & \dots & Y(k-n-1) \\ \vdots & & \vdots \\ Y(k-r) & \dots & Y(k-n-r+1) \end{bmatrix}$$

A veces los estimadores anteriores se suelen ponderar con pesos para dar más importancia a los últimos datos, y se llaman mínimos cuadrados ponderados.

Estos estimadores son bastante utilizados en econometría; no obstante al ser métodos en bloque, si se tuviese que calcular la inversa de una matriz cada vez que se amplía la matriz de datos, resultaría un método impracticable. Análogamente para órdenes de retardo superior, o cuatro, por ejemplo, habría que calcular inversas de 4×4 , lo que resultaría bastante ineficiente, solamente hacemos mención a ellos debido a que son los precursores de los algoritmos que veremos a continuación.

4.2. MÉTODO DE DURBIN-LEVINSON.

En 1946, Levinson presentó un procedimiento iterativo para determinar la función ponderatriz de un filtro lineal. Este procedimiento es ahora conocido como el algoritmo de Durbin-Levinson.

Según Akaike (1973): "Este procedimiento es probablemente la contribución más significativa en el campo del análisis con computador digital de series temporales".

Este algoritmo fue utilizado por Durbin (1960) para la solución de las ecuaciones de Yule-Walker en el caso escalar. Posteriormente Whittle (1963), lo extendió para el caso multivariable. Los problemas conectados con este algoritmo como inversión de matrices de Toeplitz fueron estudiadas por Akaike (1973 b). La estimación espectral autorregresiva por Burg (1967,1975) y la estimación del orden de modelos ARMA por Akaike (1970), Shibata (1976) y Hannan (1980).

Ya que el algoritmo de Levinson se desarrolló en el contexto del diseño de un filtro y predicción del mismo, y éste era un marco muy restringido, se olvidó rápidamente y por eso es mucho más conocido como el algoritmo de Durbin aunque los dos métodos son esencialmente el mismo.

Queremos estimar una señal $Y(t)$ en función de sus valores precedentes $Y(t-1)$, $Y(t-2)$, ..., $Y(t-M)$. Que lo podemos expresar:

$$(18) \quad Y(t)' = \sum_{m=1}^M a(m) \cdot Y(t-m)$$

donde $a(m)$ son los coeficientes de un predictor lineal, de aquí podemos derivar las ecuaciones de Yule-Walker que minimizan el error cuadrático bien por diferenciación o por el principio de ortogonalidad llegando a:

$$(19) \quad \sum_{m=1}^M a(M, m) \cdot r(m, 1) = r(1) \quad 1 = 1 \dots M$$

Los coeficientes de reflexión o coeficientes de correlación parcial son $a(M, M)$, donde el primer subíndice indica el máximo retardo. Se pueden calcular por la siguiente expresión:

$$(20) \quad a(M, M) = \begin{cases} r(1) & M = 1 \\ \frac{r(M) - \sum_{j=1}^{M-1} a(M-1, j) \cdot r(M-j)}{1 - \sum_{j=1}^{M-1} a(M-1, j) \cdot r(M)} & M = 2 \end{cases}$$

$$(21) \quad a(M, k) = a(M-1, k) - a(M, M) \cdot a(M-1, M-k)' \quad k = 1, \dots, M-1$$

Démonos cuenta que aquí trabajamos con covarianzas estimadas. La primera ecuación (20) calcula los últimos elementos de la regresión de la serie sobre los sucesivos retardos obtenidos, por ejemplo:

(22)

$$\begin{aligned} p = 1 \quad Y(t)' &= \boxed{\hat{a}(1, 1)} \cdot Y(t-1) \\ p = 2 \quad Y(t)' &= \hat{a}(2, 1) \cdot Y(t-1) + \boxed{\hat{a}(2, 2)} \cdot Y(t-2) \\ &\dots\dots\dots \\ p = k \quad Y(t)' &= \hat{a}(k, 1) \cdot Y(t-1) + \hat{a}(k, 2) \cdot Y(t-2) + \dots + \boxed{\hat{a}(k, k)} \cdot Y(t-k) \end{aligned}$$

donde básicamente lo que se ha hecho ha sido calcular los llamados coeficientes de autocorrelación parcial $\hat{a}(k, k)$ (en los filtros de celosía se llaman coeficientes

de reflexión), que puede contemplarse como la correlación parcial existente entre $Y(y - k)$ y $Y(k)$ cuando se han eliminado de ambas variables el efecto de $Y(t - 1), Y(t - 2), \dots, Y(t - k + 1)$. La segunda ecuación calcula los parámetros intermedios.

Aunque posteriormente describiremos el caso multivariable, que hemos programado, podemos decir que tiene las siguientes características:

- Sólo sirve para modelos AR.
- Es poco sensible a valores iniciales diferentes de la covarianza inicial de la serie.
- Muy sensible al valor medio de la serie.
- Si se calcula el orden de la serie por el criterio de Akaike, los valores de los parámetros de órdenes superiores al marcado por este criterio son nulos o casi nulos, dando ruido blanco en los residuos.
- No estima modelos MA ó ARMA.
- No estima adecuadamente cuando las series utilizadas son no estables y no invertibles. Sí hace previsión (si tiene el orden adecuado para los no estables) relativamente aceptable.

MODELO 1

$$(23) \quad Y(t) = 1.9 \cdot Y(t - 1) - 0.9 \cdot Y(t - 2) + \varepsilon(t) \quad \varepsilon(t) \approx N(0, 1)$$

Este modelo tiene la raíz 1, en el círculo unidad. El modelo siguiente, modelo 2 es una serie no estable con una raíz en el valor 1.05 y otra en 0.7.

MODELO 2

$$(24) \quad Y(t) = 1.75 \cdot Y(t - 1) - 0.7 \cdot Y(t - 2) + \varepsilon(t) \quad \varepsilon(t) \approx N(0, 1)$$

TABLA I

Parámetros estimados para diferentes órdenes de los modelos 1 y 2 con el algoritmo de Durbin-Levinson

MODELO	ORDEN	PARÁMETROS ESTIMADOS			
		a1	a2	a3	a4
1	2	1,91	-0,92		
	3	1,91	-0,90	0,01	
	4	1,90	-0,96	-0,11	0,06
2	2	0,05	-1,05		
	3	0,95	-0,51	10,1	

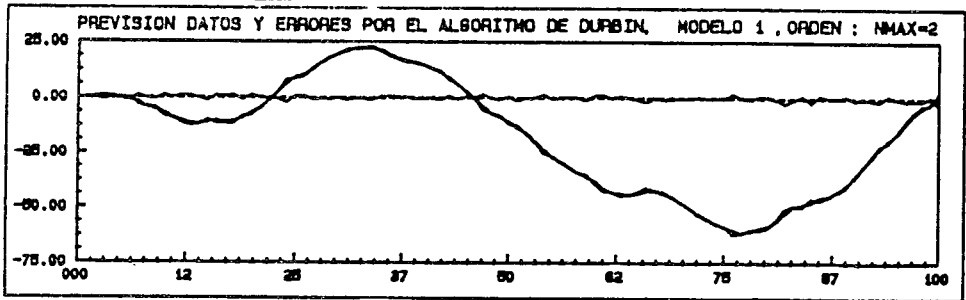


FIGURA 2. Predicción del modelo 1 con el algoritmo de Durbin-Levinson

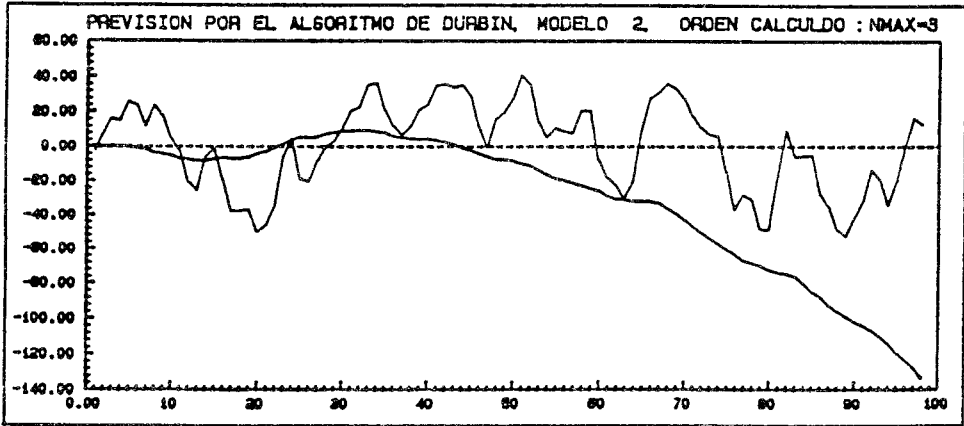


FIGURA 3. Predicción del modelo 2 con el algoritmo de Durbin-Levinson

4.3. MÉTODO DE BURG

Definiendo los residuos desde las predicciones basadas en un modelo autor-regresivo de orden M hacia adelante por:

$$(25) \quad e^f(i, M) = x(i) - \sum_{m=1}^M a(M, m) \cdot x(i - m) \quad i = M + 1 \dots N$$

donde el "*" indica complejo conjugado, y los errores de predicción hacia atrás por:

$$(26) \quad e^b(i, M) = x(i) - \sum_{m=1}^M a^*(M, m) \cdot x(i + m) \quad i = 1, \dots, N - M$$

donde los coeficientes del filtro hacia atrás son complejos conjugados inversos de los coeficientes del filtro hacia adelante, se obtiene facilmente las siguientes relaciones:

$$(27) \quad e^f(i, M) = e^f(i, M - 1) + a(M, M) \cdot e^b(i - 1, M - 1)$$

$$(28) \quad e^b(i, M) = e^b(i - 1, M - 1) + a^*(M, M) \cdot e^f(i, M - 1)$$

Burg planteó que la estimación de $a(M, M)$ se podría calcular como la suma de los errores al cuadrado hacia delante y hacia atrás, así si llamamos:

$$(29) \quad P^f(M) = E[|e^f(i, M)|^2] \quad \text{y} \quad P^b(M) = E[|e^b(M+i, M)|^2]$$

las covarianzas hacia adelante y hacia atrás, la maximización de la entropía será:

$$(30) \quad P(M, M) = 1/2|P^f(M, N) + P^b(M, N)| \geq \sum_{i=M+1}^N |e^f(i, M)|^2 + \sum_{i=1}^{N-M} |e^b(i, M)|^2 = \text{MIN}$$

Por diferenciación de (30) se llega a la siguiente expresión:

$$(31) \quad a(M, M) = \frac{-2 \cdot \sum_{i=M+1}^N e^{b^*}(i-1, M-1) \cdot e^f(i, M-1)}{\sum_{i=M+1}^N [|e^b(i-1, M-1)|^2 + |e^f(i, M-1)|^2]}$$

Definiendo el denominador de la expresión (31) en el paso $M-1$ por $D(M-1)$, se tiene la siguiente fórmula recursiva:

$$(32) \quad D(M-1) = -|e^b(N, M-1)|^2 - |e^f(M, M-1)|^2 + (1 - |a(M-1, M-1)|^2) \cdot D(M-2)$$

Hemos hecho el programa para este algoritmo, ya que es la base con el algoritmo de Durbin-Levinson del algoritmo celosía, que es la herramienta más avanzada que planteamos y nos ha dado los siguientes resultados:

- Calcula parámetros AR.
- Para pocos datos estima mucho mejor que el algoritmo de Durbin-Levinson.
- Si la sobreparametrización no es excesiva, hace nulos los parámetros sobre-dimensionados, dando ruido blanco los residuos.

Las predicciones han sido buenas para modelos AR. Para casos no estables o no invertibles estima mal; estima bien para órdenes más elevados que el suyo propio, dando valores nulos los parámetros en exceso y hace previsiones aceptables. Tampoco calcula los parámetros de series MA ó ARMA.

El algoritmo de Burg no es generalizable directamente a un modelo multi-variable con los datos aunque, podría serlo por medio de los residuos (Jones, 1978).

TABLA II

Parámetros estimados para diferentes órdenes de los modelos 1 y 2 con el algoritmo de Burg

MODELO	ORDEN	PARÁMETROS ESTIMADOS			
		a1	a2	a3	a4
1	2	-1.89	0.91		
	3	-1.7	0.54	0.19	
	4	-1.67	0.63	0.008	0.16
2	2	-1.7	0.75		
	3	-1.57	0.41	0.19	

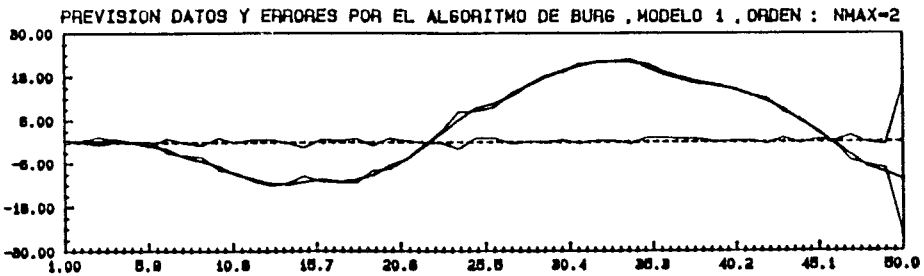


Figura 4. Previsiones y errores del modelo 1 con el algoritmo de Burg.

MÉTODO DE WITTLE-ROBINSON

Este algoritmo es el de Durbin-Levinson generalizado al caso multivariable. En este caso, el algoritmo de Burg no es directamente generalizable ya que las matrices autorregresivas hacia adelante y hacia atrás no son iguales, lo mismo que sus predicciones, aunque Akaike demostró que tienen el mismo determinante. Sean $Y(1), Y(2)\dots Y(N)$ un vector de series temporales multivariadas y la matriz de covarianzas estimadas en el retardo j viene dada por:

$$(33) \quad R(j) = 1/N \sum_{t=1}^{N-j} Y(t+j) \cdot Y(t)'$$

las predicciones autorregresivas hacia adelante de orden M son:

$$(34) \quad Y^f(t) = \sum_{k=1}^M A(k, M) \cdot Y(t-k)$$

y hacia atrás

$$(35) \quad Y^b(t) = \sum_{k=1}^M B(k, M) \cdot Y(t+k)$$

El algoritmo de Durbin-Levinson generalizado permite el cálculo de las matrices de los parámetros $A(k, M)$ y $B(k, M)$. Si se comienza el procedimiento dando:

$$(36) \quad P^f(0) = P^b(0) = R(0)$$

se puede llegar a las siguientes recursiones:

$$(37) \quad A(k, M) = A(k, M-1) - A(M, M) \cdot B(M-k, M-1)$$

$$(38) \quad B(k, M) = B(k, M-1) - B(M, M) \cdot A(M-k, M-1)$$

y las matrices de covarianza del error de predicción son:

$$(39) \quad P^f(M) = [I - A(M, M) \cdot B(M, M)] \cdot P^f(M-1)$$

$$(40) \quad P^b(M) = [I - B(M, M) \cdot A(M, M)] \cdot P^b(M-1)$$

que expresan el cambio de las matrices de predicción del error de covarianza, en términos de autorregresión o matrices de reflexión, los residuos hacia delante o hacia atrás están relacionados por:

$$(41) \quad e^f(i, M) = e^f(i, M + 1) - A(M, M) \cdot e^b(i - M, M - 1) \quad i = M + 1 \dots N$$

$$(42) \quad e^b(i, m) = e^b(i, M - 1) - B(M, M) \cdot e^f(i + M, M - 1) \quad i = 1 \dots N - M$$

Para desarrollar estas fórmulas en el ordenador se ha utilizado la descomposición de Cholesky. Como conclusión podemos decir que tiene las siguientes características:

- Permite estimar modelos AR multivariantes.
- Es muy sensible a valores iniciales de las series.
- Es poco sensible a diferentes valores de $R(0)$.
- En el modelo AR multivariable, a partir del orden estimado por A.I.C., da valores nulos o casi nulos en los parámetros y ruido blanco en los residuos.
- No permite estimar series no estables no invertibles.

Vamos a describir a continuación el efecto de estimación y predicción sobre una serie multivariable AR(1). Sea:

$$(43) \quad \begin{bmatrix} Y_1(t) \\ Y_2(t) \end{bmatrix} + \begin{bmatrix} 0.5 & 0 \\ 0.25 & 0.7 \end{bmatrix} \cdot \begin{bmatrix} Y_1(t-1) \\ Y_2(t-2) \end{bmatrix} = \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \end{bmatrix}$$

donde $\varepsilon_1(t) \approx N(0, 1)$ y $\varepsilon_2(t) \approx N(0, 1)$.

Las series se representan en las figura 5 y 6. Cuando se da orden 2 al modelo, los parámetros que corresponden a este orden son nulos. Las predicciones hacia atrás se observan en las figuras 7 y 8, y en los gráficos 8 y 9 se puede observar que las funciones de autocorrelación y autocorrelación parcial de los residuos son ruido blanco. La función de correlación cruzada no es significativa para ningún retardo como se puede ver en la figura 10.

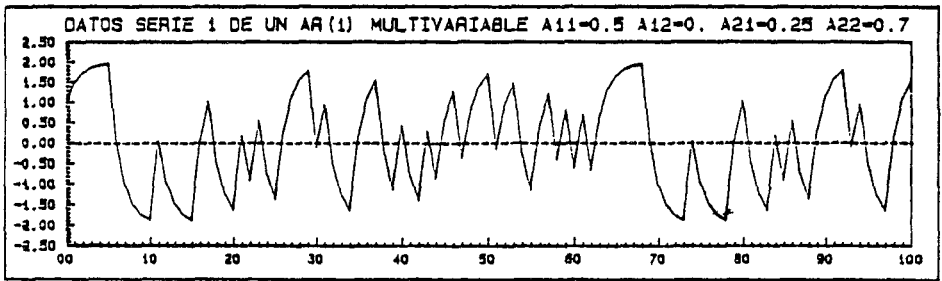


FIGURA 5. Datos de la serie $Y_1(t)$

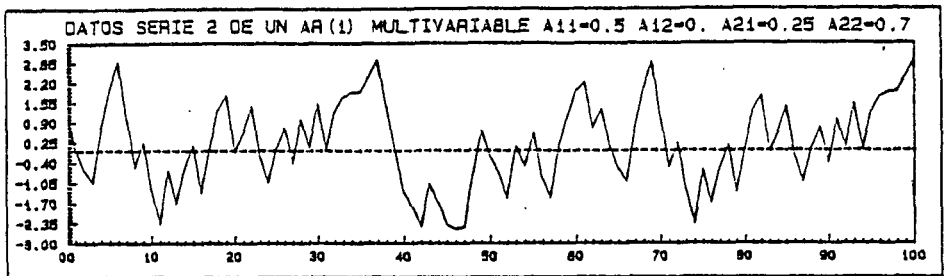


FIGURA 6. Datos de la serie $Y_2(t)$

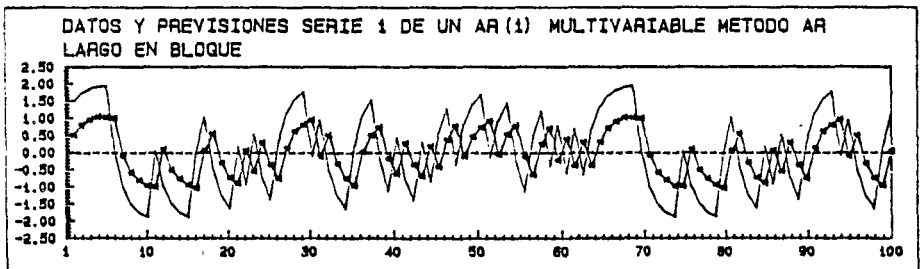


FIGURA 7. Previsiones para la serie Y_1

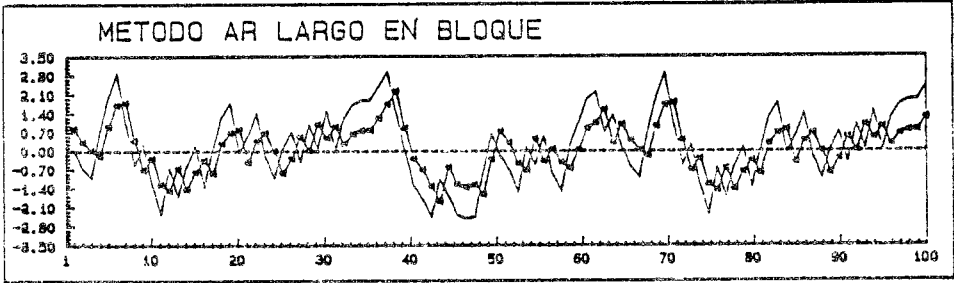


FIGURA 8. Previsiones para la serie Y_2

ACF VARIABLE IS A.MAXLAG IS 15./

NUMBER OF OBSERVATIONS > 97
 MEAN OF THE (DIFFERENCED) SERIES > -0.0263
 STANDARD ERROR OF THE MEAN > 0.1016
 T-VALUE OF MEAN (AGAINST ZERO) > -0.2592

PLOT OF SERIAL CORRELATION

LAG	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.008						I					
2	0.011						I					
3	-0.046						XI					
4	-0.052						XI					
5	0.007						I					
6	0.017						I					
7	-0.062						XXI					
8	-0.048						XI					
9	-0.088						XXI					
10	0.034						IX					
11	-0.046						XI					
12	0.049						IX					
13	-0.047						XI					
14	-0.038						XI					
15	0.001						I					

PACF VARIABLE IS A.MAXLAG IS 15./

NUMBER OF OBSERVATIONS > 97
 MEAN OF THE (DIFFERENCED) SERIES > -0.0263
 STANDARD ERROR OF THE MEAN > 0.1016
 T-VALUE OF MEAN (AGAINST ZERO) > -0.2592

PLOT OF SERIAL CORRELATION

LAG	CORR.	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
1	0.008						I					
2	0.011						I					
3	-0.046						XI					
4	-0.051						XI					
5	0.009						I					
6	0.017						I					
7	-0.067						XXI					
8	-0.050						XI					
9	-0.084						XXI					
10	0.033						IX					
11	-0.057						XI					
12	0.037						IX					
13	-0.052						XI					
14	-0.041						XI					
15	-0.004						I					

FIGURA 9. ACFs y PACFs de los residuos de la serie Y_1

ACT VARIABLE IS B.MAQLAG IS 15./

NUMBER OF OBSERVATIONS * 37
 MEAN OF THE (DIFFERENCED) SERIES * 0.0612
 STANDARD ERROR OF THE MEAN * 0.1012
 T-VALUE OF MEAN (AGAINST ZERO) * 0.6043

PLOT OF SERIAL CORRELATION

LAG	CORR.			
1	-0.005		I	*
2	-0.020		I	*
3	0.013		I	*
4	0.013		I	*
5	-0.044		XI	*
6	-0.037		XI	*
7	0.011		I	*
8	0.002		I	*
9	-0.009		I	*
10	-0.039		XI	*
11	-0.007		I	*
12	-0.024		XI	*
13	0.053		IX	*
14	-0.034		XI	*
15	-0.025		XI	*

PACT VARIABLE IS B.MAQLAG IS 15./

NUMBER OF OBSERVATIONS * 37
 MEAN OF THE (DIFFERENCED) SERIES * 0.0612
 STANDARD ERROR OF THE MEAN * 0.1012
 T-VALUE OF MEAN (AGAINST ZERO) * 0.4043

PLOT OF SERIAL CORRELATION

LAG	CORR.			
1	-0.005		I	*
2	-0.020		XI	*
3	0.013		I	*
4	0.013		I	*
5	-0.043		XI	*
6	-0.037		XI	*
7	0.009		I	*
8	0.002		I	*
9	-0.006		I	*
10	-0.040		XI	*
11	-0.011		I	*
12	-0.036		XI	*
13	0.055		IX	*
14	-0.034		XI	*
15	-0.026		XI	*

FIGURA 10. ACFs y PACFs de los residuos de la serie Y_2

PLOT OF SERIAL CORRELATION

LAG	CORR.			
-15	-0.090		I	*
-14	-0.002		XXI	*
-13	-0.057		I	*
-12	0.062		XI	*
-11	0.038		IX	*
-10	-0.012		IX	*
-9	0.021		I	*
-8	-0.030		IX	*
-7	0.016		XI	*
-6	0.011		I	*
-5	0.026		I	*
-4	-0.040		IX	*
-3	-0.054		XI	*
-2	0.004		XI	*
-1	-0.012		I	*
0	-0.122		I	*
1	0.037		XXI	*
2	0.009		IX	*
3	0.000		I	*
4	0.026		I	*
5	0.039		IX	*
6	-0.059		IX	*
7	-0.013		XI	*
8	-0.093		I	*
9	-0.027		XXI	*
10	-0.036		XI	*
11	-0.004		XI	*
12	-0.056		I	*
13	-0.029		XI	*
14	0.003		XI	*
15	-0.026		I	*

FIGURA 11. Función de correlación cruzada de los residuos de las dos series

5. MÉTODOS RECURSIVOS

Las técnicas recursivas que son apropiadas para aplicaciones en tiempo real, utilizan los datos nuevos tan pronto se hacen disponibles, mientras que en el procesamiento en bloque, la estimación de los parámetros se realiza solamente una vez por período.

Los algoritmos recursivos más interesantes sirven para el cálculo de los coeficientes del predictor mínimo cuadrático y han sido desarrollados hace bastante tiempo. Entre los más conocidos tenemos los llamados mínimos cuadrados recursivos (M.C.R.) que son muy utilizados en identificación de sistemas. Este algoritmo, está muy relacionado con el filtro de Kalman (1961), que estima los parámetros de un modelo cuando en éstos, la varianza de las estimaciones y las medidas, cambian continuamente.

Estos algoritmos son conceptualmente bastante simples pero requieren gran cantidad de cálculos. Así, para adaptar los parámetros de un predictor de orden N utilizando M.C.R. requiere del orden de N^2 operaciones (multiplicaciones y sumas). Varias técnicas subóptimas fueron desarrolladas para reducir los cálculos necesarios. Quizás, la más conocida de éstas sea la técnica del gradiente, para resolver problemas de optimización mínimo cuadrática, llamado por Widrow-Hoff, LMS (Least Mean Squares) (1975). Este algoritmo requiere sólo del orden de N operaciones por paso temporal.

Por la simplicidad y fácil realización el algoritmo de gradiente, fue muy utilizado en la última década. Mientras la técnica del gradiente ha producido interesantes soluciones para muchos problemas de procesamiento adaptativo, su comportamiento es inferior a las técnicas de mínimos cuadrados.

Una importante manifestación de este comportamiento subóptimo es el relativo a su tasa de convergencia (por tasa de convergencia entendemos el tiempo que tarda el algoritmo en responder a cambios repentinos en el proceso estocástico).

En los últimos 10 años se desarrollaron, en el área de estimación mínimo cuadrática, nuevas técnicas que han tenido gran importancia en el procesamiento adaptativo. De éstas, las más importantes, son los llamados algoritmos de mínimos cuadrados "rápidos", que fueron introducidos por Morf (1975) y otros autores contemporáneos como Lee (1981).

Estos algoritmos son capaces de adaptar recursivamente los coeficientes del predictor de orden N , utilizando del orden de N operaciones por paso temporal. En otras palabras, son capaces de realizar la predicción mínimo cuadrática, a un precio de cálculo comparable con las técnicas subóptimas de gradiente. Estos algoritmos tienen propiedades de convergencia superior a los del gradiente, ya que son mínimos cuadrados reales. Estos métodos que reciben el nombre genérico de algoritmos celosía (Lattice), tienen muchas variedades que surgen

según se tomen los datos para calcular los parámetros del modelo. Debido a esta variedad, todavía queda mucho campo de investigación, pero hoy en día pueden considerarse una alternativa a los mínimos cuadrados tradicionales.

La estructura celosía tiene unas características significativas que brevemente expondremos pero que particularizaremos a lo largo del trabajo, según nuestros descubrimientos. En términos generales cabe destacar:

- 1) Eficiencia de cálculo. La estructura de celosía pertenece a los llamados algoritmos rápidos de estimación. Calculando el predictor óptimo relacionado con una matriz de covarianza de orden $N \times N$ con $O(N^2)$ operaciones, comparado con $O(N^3)$ operaciones requeridas para los métodos de inversión de una matriz.
- 2) Recursividad en el Orden y en el Tiempo. Los métodos celosía son recursivos en el tiempo, o sea, según los datos y en el orden, porque mantienen en su memoria los órdenes inferiores al que se ha considerado.
- 3) Buenas propiedades numéricas. Estos métodos pertenecen a la clase de algoritmos raíz cuadrada (que resultan de la factorización de Cholesky) y que son reconocidos como robustos y relativamente insensibles a errores de redondeo.
- 4) Facilidad para su realización en hardware de propósito especial. Debido a la estructura modular de los filtros celosía, es posible utilizar una distribución en paralelo. El filtro celosía normalizado genera una escala de valores menores que la unidad, en todas las cantidades que utiliza, por lo que es fácil de realizar en un ordenador en aritmética de punto fijo. Además de estas ventajas, los métodos celosía permiten, como hemos dicho, el modelado de series no estables. Ello nos ha motivado para realizar un estudio profundo del tema, por entender que estos métodos son unos excelentes candidatos para la estimación paramétrica.

Por otra parte la metodología de Box-Jenkins desarrollada ultimamente por Box Tiao (1983-87) utiliza modelos en ecuaciones en diferencias estocásticas múltiples: (ARMA) vector autorregresivo-media móvil. Tiene grandes virtudes y es capaz de captar distribuciones con grandes retardos y con una parametrización escueta (el menor número de parámetros posibles). Pero presenta varios problemas, ya que al añadir la media móvil, la relación entre los parámetros es no lineal, obligando a utilizar un algoritmo de optimación no lineal, en general, tipo Gauss-Marquardt (1965). Esta pérdida de linealidad, hace que la estimación, inferencia estadística y la interpretación sea difícil en el caso univariante y muy difícil en los modelos multivariantes (Litterman, R.B. (1984)), a la vez que exige trabajar en bloque con los datos. Se ha desarrollado el paquete S.C.A. (1985) para este tipo de modelado multivariable de Box-Jenkins.

5.1. MÉTODO DE LOS MÍNIMOS CUADRADOS RECURSIVOS

Aunque el criterio de minimización es el mismo que en el caso en lote, se pretende actualizar el vector de parámetros, según los nuevos datos se van haciendo disponibles. Esto se consigue de tal forma que, en vez de invertir la matriz (17), se utilice el lema de la inversión matricial (de Woodbury). esto es:

$$(44) \quad (A + B \cdot C \cdot D)^{-1} = A^{-1} - A^{-1} \cdot B (C^{-1} + D \cdot A^{-1} \cdot B)^{-1} \cdot D \cdot A^{-1}$$

donde A , C y $(A + BCD)$ son matrices no singulares y cuadradas, llegándose a las siguientes expresiones para la estimación de los parámetros, que pueden desarrollarse en las dos fórmulas siguientes:

$$(45) \quad \hat{a}(r) = \hat{a}(r - 1) + P(r) \cdot \underline{Y}(r) \cdot (Y(r) - \underline{Y}(r)) \cdot \hat{a}(r - 1)$$

donde $P(r)$ es la matriz de covarianzas para la observación r y $P(1)$ es un valor elevado que se adapta según la ecuación:

$$(46) \quad P(r) = P(r - 1) - \frac{P(r - 1) \cdot \underline{Y}(r) \cdot \underline{Y}(r)' \cdot P(r - 1)}{1 + \underline{Y}(r)' \cdot P(r - 1) \cdot \underline{Y}(r)}$$

el denominador de esta última expresión es un escalar, cuando sólo tenemos una serie. Para el caso de varias series, las expresiones son análogas, pero debemos interpretar qué representa cada parámetro, y con quien está relacionado; pues sería lo mismo tener una salida multivariable de n variables, que n medidas de una salida univariante. En este caso $\underline{Y}(r) = (-Y(t - 1), \dots, -Y(t - r))$.

También se podría hacer estimación recursiva, por incremento del número de parámetros en vez de datos, pero se llega a situaciones semejantes. Otros aspectos del algoritmo se pueden encontrar entre otros en Graupe (1984).

Como hemos visto, este algoritmo se ha presentado para un modelo AR, y no calcula los parámetros MA de un modelo ARMA. Lo hemos programado y las características más significativas son:

- Sirve para modelos AR univariante y multivariante.
- Es poco sensible a los valores iniciales de $P(0)$, [$0 < P(0) \leq 10000$].
- Muy sensible a los valores medios de las series, o sea, $Y \neq 0$; bajo esta condición calcula mal los parámetros.
- Poco sensible a la sobreparametrización (o sea, simular por ejemplo AR(2) y estimar AR(4)).
- Poco sensible a valores iniciales de los parámetros distintos de cero, si el número de simulaciones es grande.
- Poco sensible a la varianza del ruido.

Como conclusión podemos decir, que con respecto a la estimación de los parámetros de un modelo, incluye todo excepto los valores iniciales de los parámetros y la varianza del ruido, y por supuesto estiman mal o no estiman cuando las series son no estables o no invertibles. Las figuras 12 y 13 ilustran estas conclusiones.

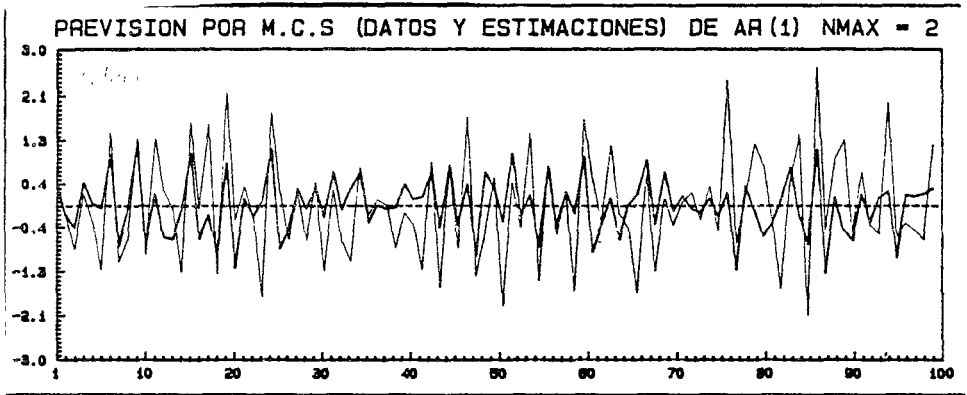


FIGURA 12. Previsión y datos por el método de M.C.S. de un modelo AR(1)

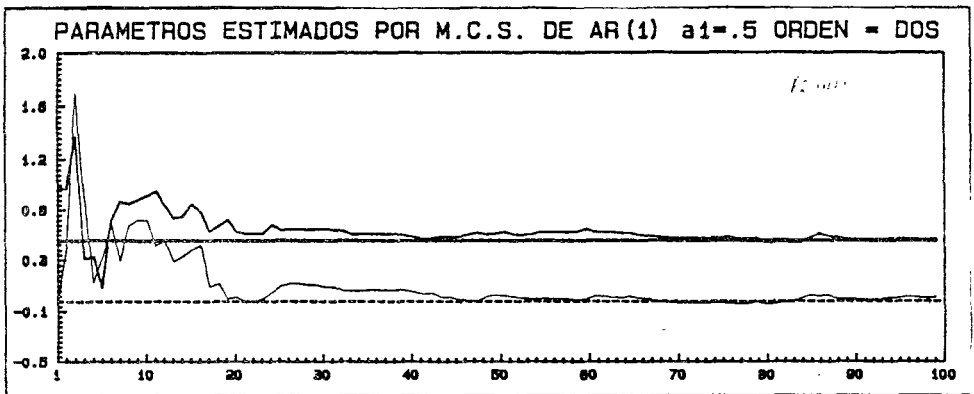


FIGURA 13. Estimación de los parámetros

También podemos decir, que la función cuadrática de los errores que minimizamos puede tener varios mínimos, sobre todo cuando la función de transferencia es complicada y por lo tanto para encontrar un mínimo absoluto, es interesante el buscar unos valores iniciales (aunque no dependa mucho) adecuados. Por eso se puede partir de la estimación de los Mínimos Cuadrados Ordinarios y, después utilizar Mínimos Cuadrados Secuenciales.

Ya que este algoritmo es muy apropiado para la estimación de modelos AR, nos dará pie para formular el planteamiento crucial de este apartado, basándonos en el desarrollo de la sección 2. Se puede reducir cualquier modelo de series temporales a modelos AR. Y para ello, desarrollaremos en el siguiente apartado uno de los más interesantes métodos de estimación que son los mínimos cuadrados extendidos, donde la idea es estimar los parámetros de un modelo ARMA. En esencia se trata de una estimación AR aumentada con la parte MA, lo que ratifica nuestra idea de que los modelos AR son una generalización más que un caso particular.

5.2. MÉTODO DE LOS MÍNIMOS CUADRADOS EXTENDIDOS RECURSIVOS

Sea un modelo ARMA(p, q), donde (p, q) son los órdenes del sistema. En este método se pretende hacer una regresión pseudolineal que representaremos por:

$$(47) \quad Y(t) = A' \cdot X(t) + W(t)$$

donde en el vector de regresión $X(t)$ incluiremos las variables inobservables de la parte MA que estimaremos desde los datos $Y(t)$. La idea de estimar los parámetros de A' y reconstruir la parte no observada de $X(t)$, fue introducida por Solo (1978).

La diferencia con los mínimos cuadrados secuenciales, reside en que aquí se hace una estimación de $w(t)'$ mediante el modelo ARMA(p, q). Sea el modelo:

$$(48) \quad \begin{aligned} y(t) + a(1) y(t-1) + \dots + a(p) y(t-p) = w(t) + \\ + b(1) w(t-1) + \dots + b(q) w(t-q) \end{aligned}$$

$$(49) \quad \begin{aligned} w(t) = y(t) + \hat{a}(1) y(t-1) + \dots + \hat{a}(p) y(t-p) - \\ - c(1)' w(t-1)' - \dots - c(q)' w(t-q)' \end{aligned}$$

y el vector $X(t)$ es:

$$(50) \quad X(t) = (-Y(t-1), \dots, -Y(t-p), w(t-1)', \dots, w(t-q)')$$

utilizando las fórmulas mínimo cuadráticas tenemos:

$$(51) \quad W(t)' = Y(t) - \hat{a}(t-1)' \cdot X(t)$$

$$(52) \quad \hat{a}(t) = \hat{a}(t-1) + P(t) \cdot X(t) \cdot [Y(t) - \hat{a}(t-1)' \cdot X(t)]$$

$$(53) \quad P(t) = P(t-1) - \frac{P(t-1) \cdot X(t) \cdot X(t)' \cdot P(t-1)}{1 - X(t)' \cdot P(t-1) \cdot X(t)}$$

Este método que lo hemos programado, permite estimar la parte MA de una serie temporal, y resulta semejante a los mínimos cuadrados secuenciales. En los gráficos que siguen, presentamos algunos resultados ilustrativos sobre este algoritmo.

Queremos puntualizar de nuevo que lo que se está estimando es un modelo AR ($p+q$), lo que ocurre es que en este modelo AR ($p+q$) los elementos que corresponden al orden q son un submodelo MA (q). O sea, hemos pasado de un modelo ARMA (p, q) \rightarrow AR ($p+q$), con esta estimación evitamos los métodos no lineales de estimación que utiliza Box-Jenkins.

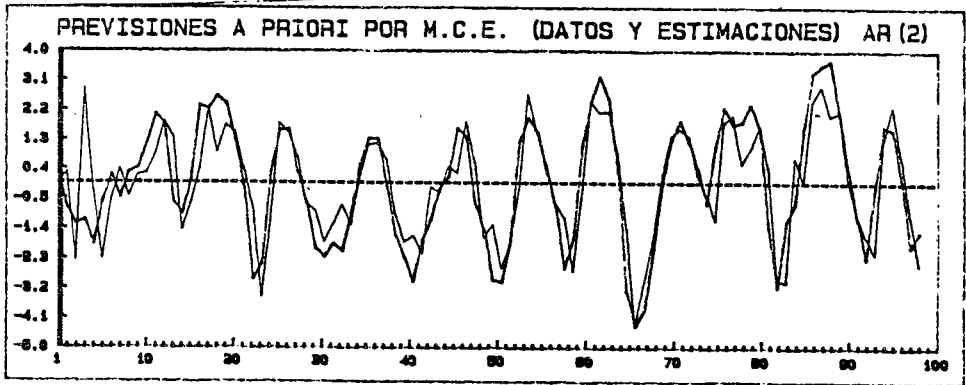


FIGURA 14. Previsión y datos con estimación mínimo cuadrática extendida

5.3. MÉTODO DE LA MÁXIMA VEROSIMILITUD RECURSIVA

Se presenta a continuación una aproximación alternativa que puede interpretarse como un prefiltrado del vector $X(t)$ en (54) con el modelo del ruido. Como este modelo puede ser inestable es necesario hacer, en su caso, una regularización de sus raíces. Este método se expone para subrayar que la estimación recursiva mínima cuadrática se realiza sobre modelos AR.

Siguiendo a Aström (1972) y Söderstrom (1973), este algoritmo surge al querer minimizar el error de predicción al cuadrado. Si se hace el desarrollo de Taylor del funcional del error cuadrático, llegamos a una serie de condiciones entre los errores de previsión y las derivadas de dicho desarrollo, que nos permite definir el vector:

$$(54) \quad \Gamma(t) = \begin{bmatrix} -Y(t-1) \\ \varepsilon(t-1) \end{bmatrix}$$

obteniendo otro vector llamado $\varphi(t)$, que se obtiene al filtrar el polinomio de la parte MA de $\Gamma(t)$, dándonos la relación:

$$(55) \quad \varphi(t) = -b(t-1)' \cdot \varphi(t-1) + \Gamma(t)$$

y el error de predicción:

$$(56) \quad w(t) = Y(t) - A(t-1)' \cdot \Gamma(t)$$

llegando a las ecuaciones siguientes:

$$(57) \quad \hat{a}(t) = \hat{a}(t-1) + 1/\beta P(t) \cdot \varphi(t) \cdot \varepsilon(t)$$

$$(58) \quad P(t) = P(t-1) - \frac{P(t-1) \cdot \varphi(t) \cdot \varphi(t)' \cdot P(t-1)}{1 + \varphi(t)' \cdot P(t-1) \cdot \varphi(t)}$$

donde β es el peso que se le da a los datos para olvidar los más viejos (factor de olvido.)

Este método está programado en la referencia Ljung (1983), por eso no entraremos en más detalles. Sin embargo, hemos hecho aplicaciones con él y resulta que es semejante a los mínimos cuadrados, excepto cuando la parte MA no es invertible. Debido a este segundo filtrado que se hace para obtener el vector φ , debemos de hacer una supervisión permanente del polinomio B , para que sus raíces pertenezcan al circuito unidad. El algoritmo desarrollado por Ljung lleva incorporado este mecanismo y en series estables e invertibles funciona muy bien; es más, bajo ciertas condiciones favorables trabaja mejor que los mínimos cuadrados extendidos.

Muñoz (1986) comprobó que este algoritmo trabaja mucho mejor que el paquete BMPD (P2T), donde se trata la metodología Box-Jenkins para los casos donde los polos y los ceros están muy cerca en modelos estables e invertibles.

Por otra parte, otro método con el que se puede trabajar es el de las variables instrumentales, Mayne (1967), Wong y Polak (1967), y Young (1965,1968). En este algoritmo se emplea un vector $\pi(t)$, que se elige de un modelo determinista de entrada y salida conocida, o sea, el caso típico de control con señal u de referencia, como aquí estamos considerando que no disponemos de ella no lo trataremos. Este método supone una clasificación a priori de las variables que no es necesaria en nuestra aproximación.

Existen diversas variantes de la estimación y previsión adaptativa descrita que obtienen mejores resultados en situaciones particulares. Por ejemplo, los mínimos cuadrados generalizados que se utilizan para evitar el sesgo en presencia de autocorrelación. Métodos de factorización a la Cholesky para facilitar el cálculo recursivo, tales como LDU y otros. Sería tan trivial como prolijo reformular estos métodos, como lo hemos hecho en las páginas que preceden, en términos de predictores adaptativos AR.

5.4. MÉTODO CELOSÍA (LATTICE O LADDER)

El arranque de estos métodos está ligado a la contribución de Levinson y Durbin. En 1946, Levinson introdujo un procedimiento iterativo para determinar la función ponderatriz de un filtro lineal. Este procedimiento se conoce actualmente como algoritmo de Durbin-Levinson. El procedimiento utilizado por Durbin en 1960 para la solución de las ecuaciones de Yule-Walker en el caso de una serie, fue extendido por Wittle (1963) y Robinson (1963) al caso multivariable.

Akaike mostró, a su vez, la relación entre el algoritmo de Durbin-Levinson y el problema de inversión de bloques de Matrices de Toeplitz. La parametrización AR resultó ser muy útil en la estimación eficiente del espectro (Burg, 1967, 1975), así como en la estimación del orden de modelo ARMA, (Akaike, 1970), (Shibata, 1976), (Hannan, 1980). Cerrando este ciclo de ideas, Itakura y Saito en Japón y Morf, Kailath y otros en U.S.A., utilizaron modelos celosía en aplicaciones de procesamiento de palabras y en la búsqueda de algoritmos rápidos de factorización, demostrando la relación entre los coeficientes de reflexión de los algoritmos celosía y los coeficientes de autocorrelación parcial que se obtienen de algoritmos a la Durbin-Levinson.

Una vez mostrada la relación entre los coeficientes de reflexión y la función de correlación parcial, ha sido posible acoplar los algoritmos celosía a la resolución eficiente de problemas de estimación paramétrica multiecuacional desde ecuaciones tipo Yule-Walker (Porat, Friedlander, etc.)

De acuerdo con el teorema de Wold, cualquier proceso estocástico, puede descomponerse en una componente determinista y una no determinista, representable como una media móvil. Si la componente en media móvil es invertible, la componente no determinista puede representarse a su vez como un modelo

AR de orden finito, que puede a su vez aproximarse por un VAR de orden finito.

Para alcanzar la estabilidad necesaria en el uso de métodos tipo Box-Jenkins, es necesario aplicar transformaciones ad-hoc (Box-Cox), no siempre de fácil interpretación económica. Los métodos celosía, desde la parametrización AR, permiten estimar modelos para series no estables, desde los datos originales directamente. El algoritmo tiene la siguiente formulación:

Inicialización

$$(59) \quad e^f(t, 0) = e^b(t, 0) = y(t)$$

$$(60) \quad R^f(t, 0) = R^b(t, 0) = \beta(t) \cdot R^f(t-1, 0) + y(t) \cdot y(t)^T$$

$$(61) \quad R^f(-1, 0) = R$$

PARA $n:=1$ HASTA MIN(n_{\max}, t)

$$(62) \quad \Delta(t, n+1) = \beta(t) \cdot \Delta(t-1, n+1) + r^b(t-1, n) \cdot e^f(t, n) / \Theta(t, n)$$

donde:

$$(63) \quad \Delta(-1, n+1) = 0$$

$$(64) \quad e^f(-1)^T = 0$$

$$(65) \quad \Theta(t, n+1) = \Theta(t, n) - e^b(t-1, n)^T \cdot [R^f(t-1, n)]^{-1} \cdot e^b(t-1, n)$$

$$(66) \quad e^f(t, n+1) = e^f(t-1, n) - \Delta(t, n+1) \cdot [R^b(t-1, n)]^{-1} \cdot e^b(t-1, n)$$

$$(67) \quad e^b(t, n+1) = e^b(t-1, n) - \Delta(t, n+1)^T \cdot [R^f(t, n)]^{-1} \cdot e^f(t-1, n)$$

$$(68) \quad R^f(t, n+1) = R^f(t, n) - \Delta(t, n+1) \cdot [R^b(t-1, n)]^{-1} \cdot \Delta(t, n+1)^T$$

$$(69) \quad R^b(t, n+1) = R^b(t-1, n) - \Delta(t, n+1)^T \cdot [R^f(t, n)]^{-1} \cdot \Delta(t, n+1)$$

$$(70) \quad K^f(t, n+1) = \frac{\Delta(t, n+1)}{R^b(t-1, n)}$$

$$(71) \quad K^b(t, n+1) = \frac{\Delta(t, n+1)}{R^f(t, n)}$$

Una vez que se han obtenido los coeficientes de reflexión K^r y K^b , se pueden calcular los parámetros de un modelo autorregresivo de la siguiente forma:

$$(72) \quad B(p, -1) = 0 \quad p = 0..n_{\max} - 1$$

PARA $i = 0$ HASTA n_{\max} DO

$$(73) \quad A(0, i) = B(0, i) = \begin{cases} i & \text{if } i = 0 \\ 0 & \text{if } i > 0 \end{cases}$$

PARA $p = 0$ HASTA $n_{\max}-1$ DO

$$(74) \quad A(p+1, i) = A(p, i) - K^b(p+1) \cdot B(p, i-1)$$

$$(75) \quad B(p+1, i) = B(p, i) - K^f(p+1) \cdot A(p, i)$$

$$(76) \quad A(i) = a(n_{\max}, i)$$

Seguidamente se expone un ejemplo del funcionamiento de estos métodos, usando el modelo multivariable, estable y con realimentación, siguiente:

$$(77) \quad \begin{aligned} Y_1(t) &= 0.8 Y_1(t-1) + 0.1 Y_2(t-1) + a_1(t) \\ Y_2(t) &= 0.2 Y_1(t-1) + 0.7 Y_2(t-1) + a_2(t) \end{aligned}$$

donde $a_1(t) \approx N(0, 1)$ y $a_2(t) \approx N(0, 1)$. Los parámetros estimados se pueden observar en la figura 15.

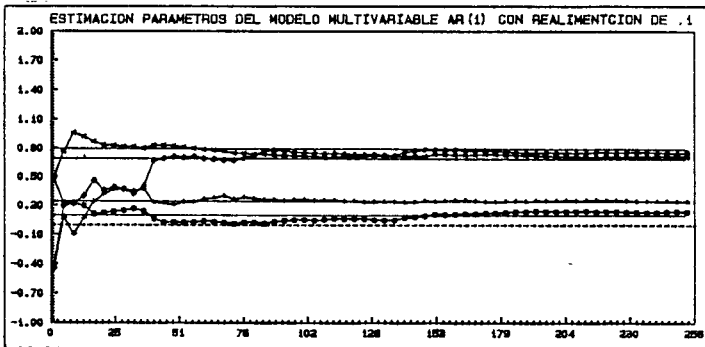


FIGURA 15. Parámetros estimados del modelo estable (77)

Seguidamente se estudia un modelo multivariable, inestable, y sin realimentación, presentándose resultados hasta la saturación del ordenador. El modelo es el siguiente:

$$(78) \quad \begin{aligned} Y_1(t) &= 1.1 Y_1(t-1) + a_1(t) \\ Y_2(t) &= 0.25 Y_1(t-1) + 0.7 Y_2(t-1) + a_2(t) \end{aligned}$$

Los parámetros estimados se pueden observar en la figura 16, calculados con orden 2, cuando en realidad es 1. Los parámetros A_2 convergen a cero.

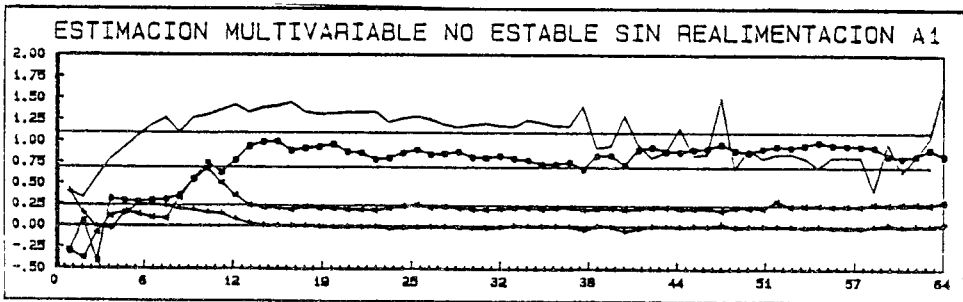


FIGURA 16. Parámetros estimados del modelo inestable (78)

6. CONCLUSIONES

Se ha realizado un estudio comparativo de los algoritmos de estimación de parámetros para series temporales. Se han planteado transformaciones de modelos ARMA en VAR (vectores autorregresivos) comprobando que para éstos últimos las técnicas de estimación adaptativa ofrecen resultados notablemente mejores que los algoritmos en bloque. Incluso para modelos multivariantes no estables, uno de estos algoritmos adaptativos, el celosía, estima los parámetros de la serie.

7. BIBLIOGRAFÍA

- [1] **Akaike, H.** (1973). "Information theory and an extension of the maximum likelihood principle", 2nd International Symposium on Information Theory, Budapest, pp. 267-281.
- [2] **Box, G.E.P. y G.H. Jenkins** (1970). "Time series analysis: Forecasting and control", Holden-Day, San Francisco.
- [3] **Box, G.E.P. y G.D. Tiao** (1975). "Intervention analysis with applications to economic and environmental problems", JASA, 70, pp. 70-79.
- [4] **Box, G.E.P. y G.C. Tiao** (1977). "A canonical analysis of multiple time series", BIOMETRIKA, 64, pp. 355-365.
- [5] **Burg, J.P.** (1967). "Maximum entropy spectral analysis", PhD Dissertation, Dpt. of Geophysics, Stanford University.

- [6] **Durbin, J.** (1960a). "The fitting of time series models", *Rev. Inst. Int. Statistics*, Vol. 28, pp. 233-244.
- [7] **Durbin, J.** (1960b). "Estimation of parameters in time series regression models", *Journal of the Royal Statistical Society*, Vol. 22, pp. 139-153.
- [8] **Fuente, D.** (1987). "Previsión adaptativa por el método celosía en organización de empresas", Tesis doctoral, Universidad de Oviedo.
- [9] **Hannan, E.J.** (1980). "The estimation of the order of an ARMA process". *Ann. Statistics*, Vol. 8, pp. 1071-1081.
- [10] **Hernández Iglesias, C. y Hernández Iglesias F.** (1976). "Causality and Independence Phenomenon. The care of the Demand for Money". *Journal of Econometrics* 15, pp. 247-263.
- [11] **Jones, R.H.** (1978). "Multivariate autoregression estimation usign residuals in applied time series", *Academic Press, New York*, pp. 139-162.
- [12] **Jones, W.B. y A.O. Steinharat** (1985). "Finding the poles of the lattice filter", *IEEE Transactions on Acoustic, Speech and Signal Processing*, Vol. ASSP-33, N° 4, pp. 1323-1331.
- [13] **Kalman, R.E.** (1960). "A new approach to lineal filtering and prediction problems", *Journal of Basic Engineering*, Vol. 82, pp.33-44.
- [14] **Lee, D.T.** (1980). "Canonical Ladder Form Realizations and fast estimation algorithms", Thesis, Standford University.
- [15] **Litterman, R.C.** "Techniques of forecasting using vector autoregressions", Thesis, Minnesota University.
- [16] **Ljung, L. y T. Soderstrom** (1983). "Theory and practice of recursive identification", MIT Press.
- [17] **Maravall, A.** (1985). "Predicción con modelos de Series Temporales". Banco de España. Servicio de Estudios. Documento de Trabajo n° 8501.
- [18] **Morf, M** (1977). "Efficient solutions of covariances equations for linear prediction", *IEEE Transactions on Acoustic Speech and Signal Processing*, Vol. ASSP-25, N° 5, pp. 429-435.
- [19] **Muñoz, P** (1986). "Comparación entre algoritmos de estimación RLS y el paquete BMDP (P2T)", Congreso de Estadística e Investigación Operativa, Málaga, Noviembre 86.
- [20] **Quenouille, M.H.** (1949). "Approximate test of correlation in time series", *Journal of the Royal Statistical Society*, Vol. 11, pp. 68-84.
- [21] **Sargen, T.J.** (1978). "Estimation of dynamic labor demand schedules under rational expectations", *Journal of Political Economy*, Vol. 86, N° 6, pp. 1009-1044.

- [22] **Shibata, R.** (1976). "Selection of the order of an autoregressive model by Akaike's information criterion", *Biometrika*, Vol. 63, pp. 117-126.
- [23] **Sims, C.A.** (1974). "Distributed lags", Chapter 5 of *Frontier of Quantitative Economics*, Vol. II, pp.289-232.
- [24] **Slutzky, E.** (1937). "The summation of random causes as the source cyclic processes", *Econometrika*, Vol. 5, pp. 105-146.
- [25] **Tiao, G.C. y G.E.P. Box** (1983). "An introduction to applied multiple time series analysis", S.C.A. working paper N° 101.
- [26] **Whittle, P** (1963). "On the fitting of multivariate autoregressions and the approximate canonical factorization of a spectral density matrix", *Biometrika*, Vol. 50, pp. 129-134.
- [27] **Widrow, B.** (1975). "Adaptative noise cancelling: Principles and applications", *Proceedings of the IEEE*, Vol. 63, N° 12, pp. 1692-1716.
- [28] **Yule, G.U.** (1972). "On a method of investigation periodicities in disturbed series with special reference to Wolfer's sunspot numbers", *Phis. Transactions of The Royal Society*, Vol. 226, pp. 267-298.

