

MODELIZACIÓN Y PREDICCIONES PARA LA SERIE TEMPORAL DEL IPC, UTILIZANDO EL ESPACIO DE ESTADO BALANCEADO

M^a Cristina Suárez Riestra

Matemática Aplicada. Universidad de Oviedo

Estudios de Economía Aplicada I.S.S.N.: 1133-3197

Nº. 2 de 1994 p.p. 145-167

RESUMEN

En el presente trabajo, se presenta el proceso de modelización de la serie del IPC, bajo una perspectiva de parametrización poco habitual en el área econométrica, en Espacio de Estado Balanceado. En él se aplican técnicas algebraicas, como por ejemplo la descomposición canónica singular, y estadísticas, para llevar a cabo objetivos de modelización y posterior estudio del comportamiento predictivo; para lo que se requiere un soporte de software específico, como es el caso de utilización del MATLAB y programas confeccionados al efecto. Los resultados son muy alentadores, tanto en el estudio que se presenta como con otros realizados, que utilizan las mismas técnicas.

Palabras Clave: Serie de Tiempo, Espacio de Estado Balanceado, Matriz de Hankel, Descomposición Canónica Singular, Predicción.

ABSTRACT

The current work present the IPC modelling process, under a different parametrization, no usual in the econometric area, in Balanced State Space. In that, the most important thing is the application of algebraic techniques as by instance the singular valued decomposition and statistics, in order to get the targets of modelling and forecasting. It is necessary a specific environment of software, that is the case of using the MATLAB and the realization of some appropriates programs. The results are very good, so in the present work as in others ones that use the same techniques.

Key Words: Time Series, Balanced State Space, Hankel Matrix, Singular Valued Decomposition, Forecasting.

1.- INTRODUCCIÓN.

Para realizar el estudio que se presenta a continuación se ha elegido una forma de representación poco habitual en el área de aplicación en que se enclava. El tipo elegido, en el presente trabajo, es la representación en Espacio de Estado Balanceado, proveniente de la teoría de sistemas, útil para tratar este tipo de problemas de carácter temporal, como podrá comprobarse a continuación, y como ya se ha hecho a lo largo de otros estudios realizados,

que pueden encontrarse por ejemplo en Mittnik (1989), Suárez R.(1994a), y Suárez R. y Mayo R. (1994) entre otros.

La característica fundamental, de este tipo de parametrización concreto, es la de encontrar una representación lineal de carácter matricial para la serie en estudio, por medio de la cual se puedan realizar predicciones (Aoki (1990)); teniendo, pues, como objetivos en primer lugar el de modelización y en segundo lugar el de realización de predicciones, de la forma mas fiable posible. Para lo cual, se han realizado estudios de evaluativos de simulación, así como estudios del tipo que se presentan aquí, comparativos, en el sentido de dejar fuera de la muestra valores reales para evaluar la calidad de las mismas (Mittnik (1989), (1990)).

Los procesos, que se llevan a cabo para alcanzar los objetivos planteados, requieren el manejo de un software específico, en este caso, elaborado al efecto por la autora del presente trabajo, con los algoritmos necesarios tanto para la fase de modelización como para la de realización de predicciones. Los programas confeccionados están apoyados en un entorno de software comercial denominado MATLAB, (Marcus, (1992)) por medio del cual, es posible realizar todos los pasos del proceso que se describen en los siguientes epígrafes.

2.- BREVE DESCRIPCIÓN DE LA REPRESENTACIÓN EN ESPACIO DE ESTADO BALANCEADO. FASE DE MODELIZACIÓN.

La representación de un sistema en espacio de estado balanceado, (Moore (1981)), es un tipo de parametrización concreto dentro de la forma general de representar los sistemas en espacio de estado, (Ogata (1967), Furuta and others (1988)), a través de dos ecuaciones. Más que deducir una ecuación m-dimensional en diferencias de alto orden para los datos dados, (modelado ARMAX), siguiendo la metodología de Box-Jenkins, (Box-Jenkins (1976), Alcaide y Alvarez (1992)), aquí, el objetivo es estimar un sistema de primer orden de una determinada dimensión dado por:

$$\begin{aligned} z_{t+1} &= Az_t + Bx_t + K\epsilon_t \\ y_t &= Cz_t + Dx_t + \epsilon_t \end{aligned} \tag{1}$$

donde z_t es el vector de estado n-dimensional, que contiene la información de la historia del proceso y_t , que puede ser necesaria y suficiente para predecir el futuro.

La primera de las ecuaciones en (1), es denominada ecuación de transición y la segunda es denominada ecuación de salida o ecuación de medida o de observación.

La matriz A , representa la dinámica del sistema, es de dimensiones $n \times n$, la matriz B , se denomina de acoplo con las entradas, es de dimensiones $n \times r$, la matriz K , es una matriz de ponderación de las perturbaciones, es de dimensiones $n \times m$, la matriz C , es la matriz de observación, es de dimensiones $m \times s$ y la matriz D , de acoplo con las entradas en la ecuación de observación, es de dimensiones $m \times r$; todas ellas son las que conforman el sistema. El considerar una representación balanceada es con la finalidad de que los modelos encontrados, para los procesos que se traten, tengan las propiedades concretas de controlabilidad, observabilidad, estabilidad y sean representaciones minimales, (Moore (1981)), lo cual, les confiere propiedades no patológicas para representar sus dinámicas. En base a esa elección, es posible encontrar el tipo de representación adecuado para cada proceso, a través de los algoritmos que se describen, el cual, basado en una estimación de parámetros mínimo cuadrática, permite encontrar las matrices A , B , C y D , representativas del sistema.

En primer lugar, es necesario confeccionar unas matrices X e Y , para aplicar el algoritmo mínimo cuadrático, que nos permitirá estimar la matriz más fundamental del proceso: la matriz de bloques de Hankel, (Kung and Lin (1981)). Para ello es necesario elegir quienes van a ser las entradas al sistema y quienes las salidas del mismo. Por supuesto, estas últimas son siempre las series en estudio o una modificación de las mismas, cara a trabajar con procesos estacionarios. Las entradas al sistema, debido al carácter de los procesos temporales, son desconocidas y se suelen elegir a tal efecto o bien la series retardadas, o señales PRBS (pseudo aleatorias), generadas convenientemente en el ordenador en base a procesos normales $(0,1)$, por medio de la utilización de la función signo, se puede ver en Åström and Wittenmark (1984).

Por otro lado, las variables de estado, características de este tipo de representación, han de partir de un valor inicial; en algunos casos, por defecto, este se elige como cero, sin afectar excesivamente el proceso, dado que, el efecto del estado inicial se desvanece en pocos pasos por las características de la representación, en otros, como en el presentado a continuación para el IPC, es posible hacer una estimación del estado inicial por medio de una elección conveniente de las matrices confeccionadas con las entradas: X y las salidas: Y . En la práctica el modelizador desconoce los parámetros M_i , bloques de la matriz de Hankel, fundamental por otro lado en el proceso, que es necesario

estimar y ha de estimarlos a partir de los datos de la muestra que posee, ya que esa es la única información disponible.

Dado que (1) ha de ser un sistema estacionario, los parámetros estimados por mínimos cuadrados para el sistema, a partir de los datos, son los obtenidos al encontrar la matriz:

$$\Theta = [s_1, s_2, \dots, s_k, M_0 \ M_1 \ M_2 \ \dots \ M_k]^T \quad (2)$$

que se calcula a partir de la operación matricial, mínimo cuadrática, dada por:

$$\Theta = (X^T X)^{-1} X^T Y \quad (3)$$

donde, X e Y, matrices con entradas y salidas, están dadas por las matrices confeccionadas al efecto, ec. (4); N representa el tamaño elegido de la muestra y k el orden estimado para el sistema.

$$Y = [y_0 \ y_1 \ \dots \ y_N] \quad X^T = \begin{bmatrix} I_k & 0_{k \times (N-k)} \\ x_0 \ x_1 \ \dots \ x_k \ x_{k+1} \ \dots \ x_N \\ 0 \ x_0 \ \dots \ x_{k-1} \ x_k \ \dots \ x_{N-1} \\ \vdots \ \vdots \ \quad \quad \quad \vdots \ \quad \quad \quad \vdots \\ 0 \ 0 \ \dots \ x_0 \ x_1 \ \dots \ x_{N-k} \end{bmatrix} \quad (4)$$

Para elegir de forma aproximada, con visos de realidad, el entero k se pueden utilizar diferentes criterios, entre los que destacan el criterio de selección estándar del orden AIC, BIC, o FPE, utilizados por Akaike, Shibata, Granger y Newbold, Ljung, Wei.

Para probar si el sistema es determinista, podemos observar si los residuos dados por la ecuación mínimo cuadrática son ruido blanco, usando los test, al efecto, propuestos por Poskitt y Tremayne (1981).

Si la hipótesis de ruido blanco es aceptada, construimos la matriz de Hankel finita, a partir de los bloques estimados M_i y de los valores estimados s_i , correspondientes al estado inicial, si se tratase de un sistema estocástico una de las alternativas es utilizar el filtro de Kalman como estimador predictivo, (Suárez R. (1994a)).

De esta forma, la matriz estimada de Hankel, confeccionada con los bloques estimados anteriormente, es:

$$H_x = \begin{bmatrix} s_1 & M_1 & M_2 & \dots & M_k \\ s_2 & M_2 & M_3 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ s_k & M_k & 0 & \dots & 0 \end{bmatrix} \quad (5)$$

Para obtener un modelo de espacio de estado balanceado a partir de la ecuación anterior, se emplea la técnica de descomposición del valor singular (SVD), (Strang (1980)), para realizar la descomposición de la matriz de bloques de Hankel.

La SDV de H_x conduce a

$$H_x = U Q V^T$$

donde $U^T U = I$ y $V^T V = I$, y Q es una matriz diagonal que contiene los valores singulares de H_x ordenados de mayor a menor, esto es:

$$q_1 \geq q_2 \geq \dots \geq q_{n_{\max}} \geq 0 \quad \text{con} \quad n_{\max} = K \min(m, r)$$

Los factores de H_x que son $H_x = O C$, O es la matriz de observabilidad y C la de controlabilidad, se calculan por medio de las siguientes expresiones:

$$O = U Q^{1/2}$$

$$C = Q^{1/2} V^T$$

donde $Q^{1/2}$ denota la matriz diagonal, que contiene la raíz cuadrada de los valores singulares de la descomposición canónica.

A partir de este instante, se pueden de forma inmediata, deducir las matrices del sistema, ec.(9), (Mittnik (1989), (1990)):

$$\begin{aligned} z_{t+1} &= A z_t + B x_t \\ y_t &= C z_t + D x_t + \epsilon_t \end{aligned}$$

siguiendo los pasos que se detallan:

1ª) La matriz $D = M_0$.

2ª) La matriz C está dada por las m primeras filas de la matriz de observabilidad.

3ª) La matriz B está dada por las r primeras columnas de la matriz de controlabilidad.

4ª) Para obtener la matriz A construimos una matriz, transformada de la matriz de Hankel, del siguiente modo:

$$H_n^* = \begin{bmatrix} M_2 & \dots & M_{n+1} \\ \vdots & & \vdots \\ M_{n+1} & \dots & M_{2n} \end{bmatrix} = O A C \quad (10)$$

A partir de la cual es posible determinar A , despejando:

$$A = O^* H_n^{\dagger} C^*$$

donde el asterisco denota la inversa generalizada de las matrices afectadas.

Dadas A , B , C y D , si no se supone, como en este caso, que el estado inicial sea cero, este se puede calcular del siguiente modo:

$$z_0 = O^* (Y - T X)$$

siendo X , Y y O^* matrices conocidas y la matriz T determinada por la expresión:

$$T = \begin{bmatrix} D & 0 & \dots & \dots & \dots & 0 \\ CB & D & \dots & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ CA^{n-2}B & \dots & \dots & CB & D & \end{bmatrix} \quad (13)$$

O bien, alternativamente, utilizando los valores s_i en la siguiente expresión:

$$z_0 = O^* s \text{ donde } s = (s_1^T, \dots, s_n^T)$$

El usar la SDV para obtener O y C es lo que origina que (9) sea llamada representación en Espacio de Estado Balanceada. Es decir, al modelo (9), se le denomina **representación en espacio de estado balanceado**, si los dos productos dados por la ecuación (15) son diagonales e iguales, (Aoki (1990)).

$$O^T O = Q^{1/2} U^T U Q^{1/2} = Q$$

$$C C^T = Q^{1/2} V^T V Q^{1/2} = Q$$

Los modelos en espacio de estado balanceado conducen a una reducción secuencial del espacio de parámetros. Si comparamos los k primeros valores singulares con los últimos $n_{\max} - k$, estos son muy pequeños en algunos casos, y esto es lo que nos permite separar del modelo en espacio de estado supuestamente sobreparametrizado original, un subsistema "dominante" y otro "débil". Dividiendo el sistema en espacio de estado en subsistemas como se expresa en la ecuación (16), donde los valores singulares próximos a cero corresponderían al subsistema débil, dado por las cajas $(A_{22} \ B_2 \ C_2 \ D)$, que son originados probablemente por errores de muestreo en el cálculo de los M_i estimados o por alguna otra razón en el proceso de cálculo. De este modo quedaría:

$$\begin{bmatrix} z_{1,t+1} \\ z_{2,t+1} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} x_t$$

$$y_t = \begin{bmatrix} C_1 & C_2 \end{bmatrix} \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} + D x_t + \epsilon_t \quad (16)$$

Así, solamente el subsistema "dominante" formado por las cajas (A_{11}, B_1, C_1, D) , que está asociado con los n valores singulares más grandes de la matriz de Hankel estimada, debería ser retenido para predecir los futuros valores de y_t .

Habiendo decidido que k , sea la dimensión del vector de estado, una aproximación de rango n , de la matriz estimada de Hankel, está dada eliminando los últimos valores singulares $n_{\max} - k$ en Q , así como, las correspondientes columnas de las matrices estimadas U y V . Esto nos lleva a manejar matrices de menor dimensión que las originales. Esta es una propiedad, que poseen internamente las representaciones en espacio de estado balanceadas, característica, por otro lado, que es muy atractiva debido a que

representaciones de más baja dimensión $k < n_{\max}$ pueden ser conseguidas simplemente eliminando las $n_{\max} - k$ filas de las matrices estimadas A y B, y lo mismo con las columnas de las matrices estimadas B y C, y están anidadas en las obtenidas de mayores dimensiones, es decir, contenidas estrictamente dentro del modelo original. Las aproximaciones de más bajo orden, por tanto, se obtienen simplemente por eliminación de las filas y columnas en las matrices correspondientes, sin necesidad de realizar de nuevo todo el proceso. Una consecuencia importante de esto es, pues, que las aproximaciones anidadas de dimensión más baja, son fáciles de obtener sin tener que realizar de nuevo la estimación, operación que sí es necesaria de realizar en los modelos convencionales. Por tanto, los modelos de predicción, de dimensión $n_{\max} - i$ ($i = 0, 1, 2, \dots, n_{\max}$), se pueden evaluar con bajo coste y pueden servir como criterio de selección del modelo definitivo; dato, por otro lado, muy interesante, a tener en cuenta a la hora de realizar los estudios prácticos de aplicación de esta teoría y que ha sido utilizado en el presente trabajo, comprobando su utilidad.

3.- ESTUDIO DE LA SERIE IPC.

3.1. Identificación y estimación del modelo.

Se va a estudiar a continuación la serie de datos del IPC español, serie muy analizada por diferentes investigadores desde el punto de vista de la metodología de Box-Jenkins, debido a las influencias que dicho valor puede tener en el mercado socio-económico. La evolución de este índice es un dato de gran interés, de ahí que con esta metodología se intente conseguir el objetivo planteado de predicción.

La serie que se va a analizar ver Fig. 1, serie del IPC, corresponde al período que va desde 1980.01 a 1988.10. 106 datos mensuales, base 1983. Esta ha sido analizada por muchos investigadores del tema, con mayor o menor éxito. La serie citada, muestra una fuerte tendencia, por lo que será preciso trabajar en diferencias y dado que los datos presentan una cierta dispersión, se le va a realizar también una transformación logarítmica, con el fin de concentrar los valores en una banda de amplitud menor.

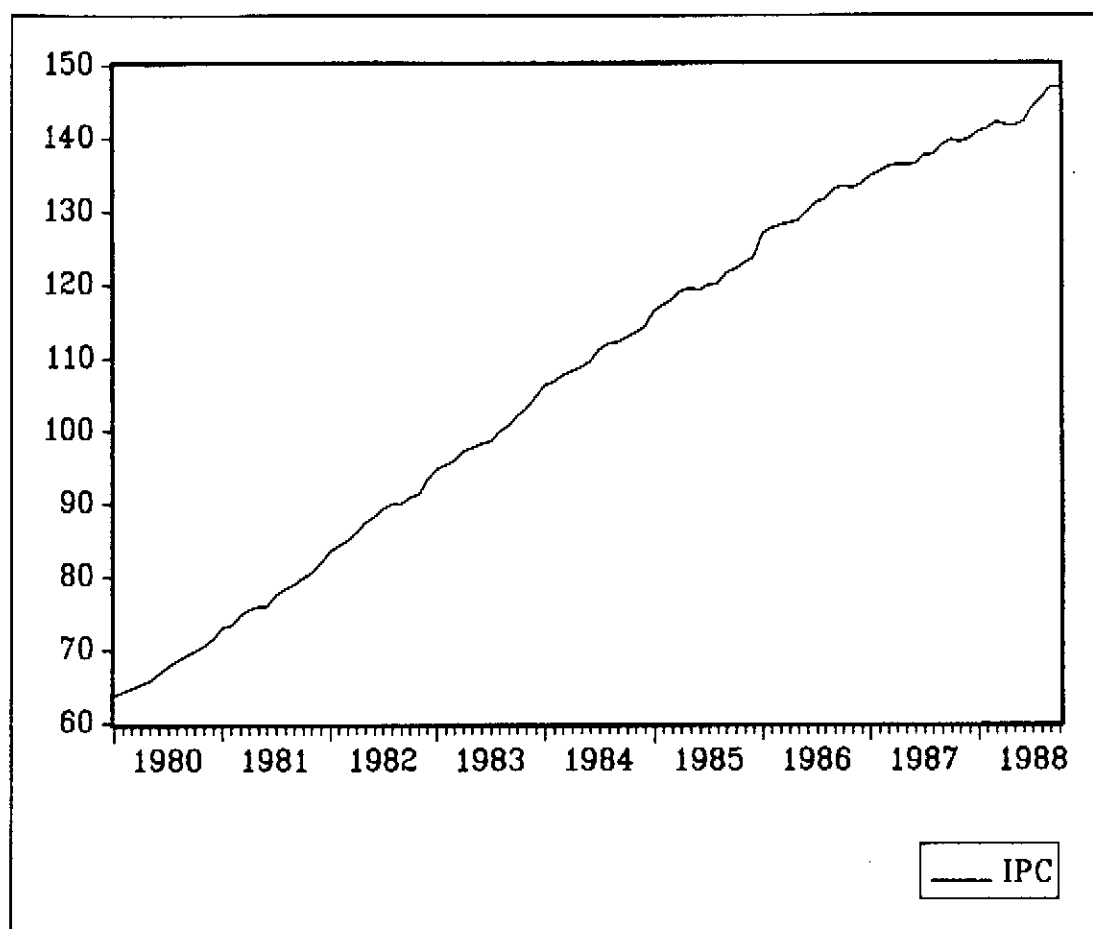


Fig. 1.- Serie del IPC. Rango de valores: 1980.01 a 1989.10.

Dado que, con una diferencia, la serie sigue presentando una cierta tendencia en su último período marcadamente decreciente, indicativa de menores tasas de inflación en los últimos años, se va a realizar una nueva diferencia con lo que se consiguen valores más estacionarios. En un segundo análisis de la serie se determina la presencia de una estacionalidad 12, por ejemplo, analizando los valores a través de una tabla de Buys-Wallot o estudiando las funciones de autocorrelación. Este problema, en este caso, y en otros similares ya realizados, (Suárez R., (1994a)), se ha resuelto aplicando doce diferencias sobre esas dos ya realizadas, con lo que se tendría la serie

transformada de este modo: $(1-B)^{14}LIPC$, que será la que se tratará de identificar con un modelo en espacio de estado balanceado, representándola bajo el nombre de D14IPC, Fig. 2, a la cual, posteriormente, realizando las transformaciones oportunas se reconvertirá en la original estimada. Es necesario tener en cuenta que, se pierden valores de la serie original al realizar las 14 diferencias, pero, este es un problema menor en este tipo de representación, ya que, no se necesitan tantos valores como en otros tipos de representaciones, como por ejemplo si se utilizase la metodología de Box-Jenkins.

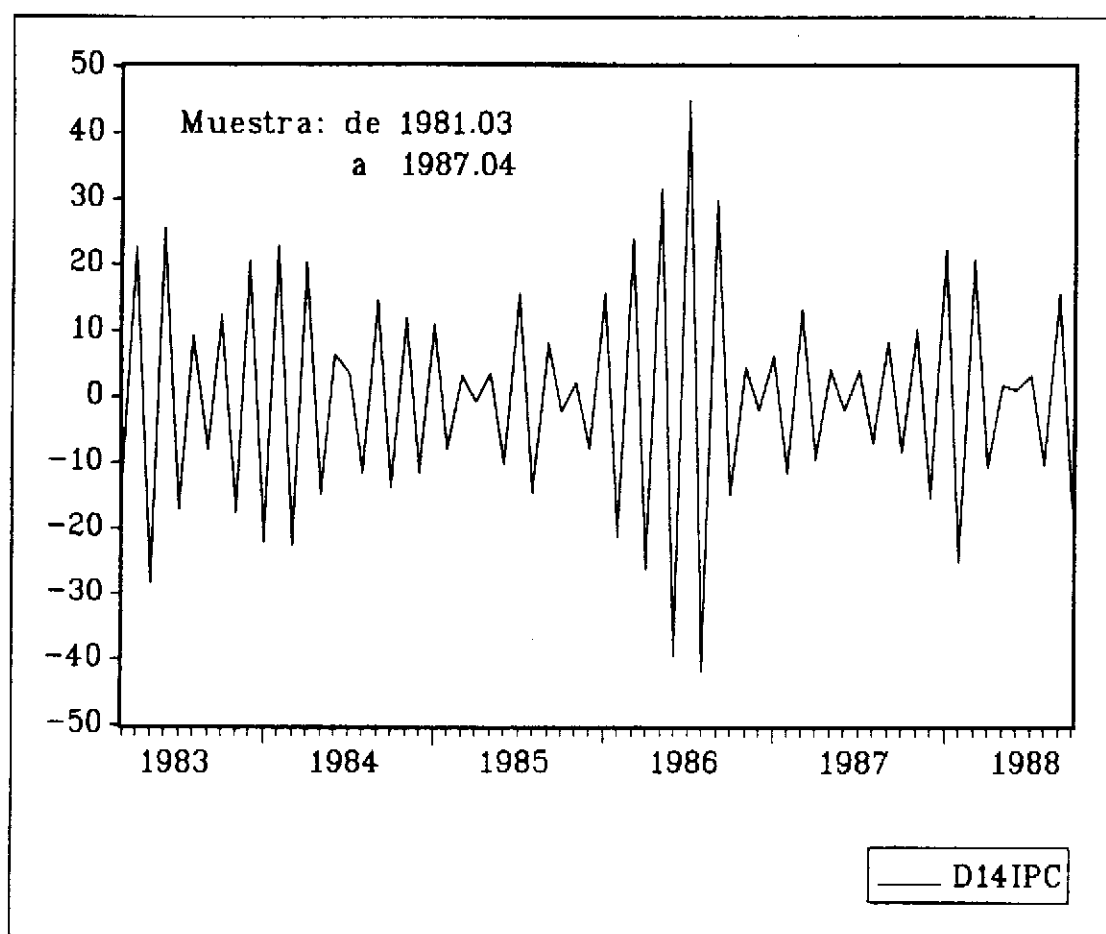


Fig. 2.- Variable transformada definitiva a partir de la serie del IPC.

Una vez seleccionada la variable definitiva, que se va a manejar como salida y elegidas como entradas la primera diferencia del logaritmo de la serie, convenientemente actualizada, se aplica el algoritmo de estimación en espacio de estado balanceado; y se deduce el orden del modelo definitivo, considerando el sistema dominante, procedente de otro de orden mas elevado, teniendo en cuenta que: los valores muy pequeños de la diagonal de la matriz central de la SDV, de la matriz de Hankel, para los datos de la serie en estudio: D14IPC, ver Fig. 2, pueden ser suprimidos; se encuentra, pues, un modelo definitivo, anidado en el original estimado de orden elevado, en concreto, en este caso el elegido es de orden ocho y se expone a continuación:

MATRIZ DIAGONAL DE LA SDV DE LA MATRIZ DE HANKEL:

$$S = \text{diag}(6870.3 \quad 4097.7 \quad 1794.9 \quad 602.8 \quad 160.5 \quad 34.4 \quad 5.9 \quad 0.8)$$

Realizados los pasos convenientes del algoritmo descrito en este trabajo, para identificar la matrices estimadas del sistema, bajo la hipótesis de orden 8 son:

MATRICES DE LA ECUACIÓN DE TRANSICIÓN:

La matriz A de dimensiones 8x8

-0.9281	0.2552	0.0304	0.0440	0.0063	0.0031	0.0002	0.0001
-0.2552	-0.7396	0.3955	0.0398	0.0410	0.0034	0.0017	0.0000
0.0304	-0.3955	-0.5015	0.4592	0.0273	0.0285	0.0006	0.0007
-0.0440	0.0398	-0.4592	-0.2766	0.4663	0.0077	0.0181	-0.0007
0.0063	-0.0410	0.0273	-0.4663	-0.0910	0.4400	-0.0086	0.0116
-0.0031	0.0034	-0.0285	0.0077	-0.4400	0.0547	0.3976	-0.0186
0.0002	-0.0017	0.0006	-0.0181	-0.0086	-0.3976	0.1671	0.3493
-0.0001	0.0000	-0.0007	-0.0007	-0.0116	-0.0186	-0.3493	0.2541

La matriz B de dimensiones 8x1 :

$$[26.12 \quad -33.56 \quad 23.85 \quad -11.11 \quad 3.59 \quad -0.82 \quad 0.13 \quad -0.015]^T$$

MATRICES DE LA ECUACIÓN DE OBSERVACIÓN:

La matriz de salida, denominada C de dimensiones 1x8 es:

$[-26.12 \ -33.57 \ -23.85 \ -11.11 \ -3.59 \ -0.83 \ -0.13 \ -0.015]$

La matriz D se considera nula.

El **estado inicial** estimado en el proceso, a partir de los datos de la serie D14IPC, esta dado como una matriz 8×1 ; cada elemento de la matriz representa un estado inicial estimado, de los ocho que maneja el sistema.

$[-0.29 \ 0.1879 \ 0.0069 \ -0.0779 \ 0.1573 \ -0.0644 \ 0.0393 \ -0.008]^T$

Conocidos los valores de las matrices estimadas para el sistema y el estado inicial estimado, importante a la hora de identificar bien los primeros valores de la serie, aunque no determinante para el sistema en estudio, como ya ha sido mencionado, la salida simulada, correspondiente al sistema representado por esas matrices, se simboliza por SALIDA y está representada junto con la variable a la que modela en la Fig.3. Los dos valores como se puede observar se simultanean prácticamente, debido a que la identificación de la dinámica que representa el proceso ha sido bastante bien conseguida, este hecho se puede observar también en la Fig. 4., gráfico EIPC-IPC, de los valores del IPC estimado, manejando valores reales, corregidos los errores en cada paso, variable EIPC y de los valores reales del IPC.

Un dato a tener en cuenta, a la hora de valorar la calidad de un modelo frente a otros alternativos, es el valor de los errores que, en este caso, han dado:

Varianza residual	Desviación típica
8.6492e-005	0.0093

Se ha trabajado con distintos rangos muestrales, cara a analizar su efecto, no notando ningún cambio sustancial de realizar el estudio con 74 datos a realizarlo con 91 datos, esto ya había sucedido en otros estudios, si la dinámica es captada, el número de datos muestrales mínimo ya es suficiente en cuanto a precisión de los resultados se refiere, por lo que, se eligió la muestra de 74 datos y se dejaron los restantes para evaluar las predicciones.

Como puede verse los valores de la variable estimada van evolucionando del mismo modo que la variable original, por lo que llegado este punto, sólo queda por realizar el estudio fuera del alcance de la muestra para ver el funcionamiento del modelo en su etapa predictiva y analizar los errores correspondientes.

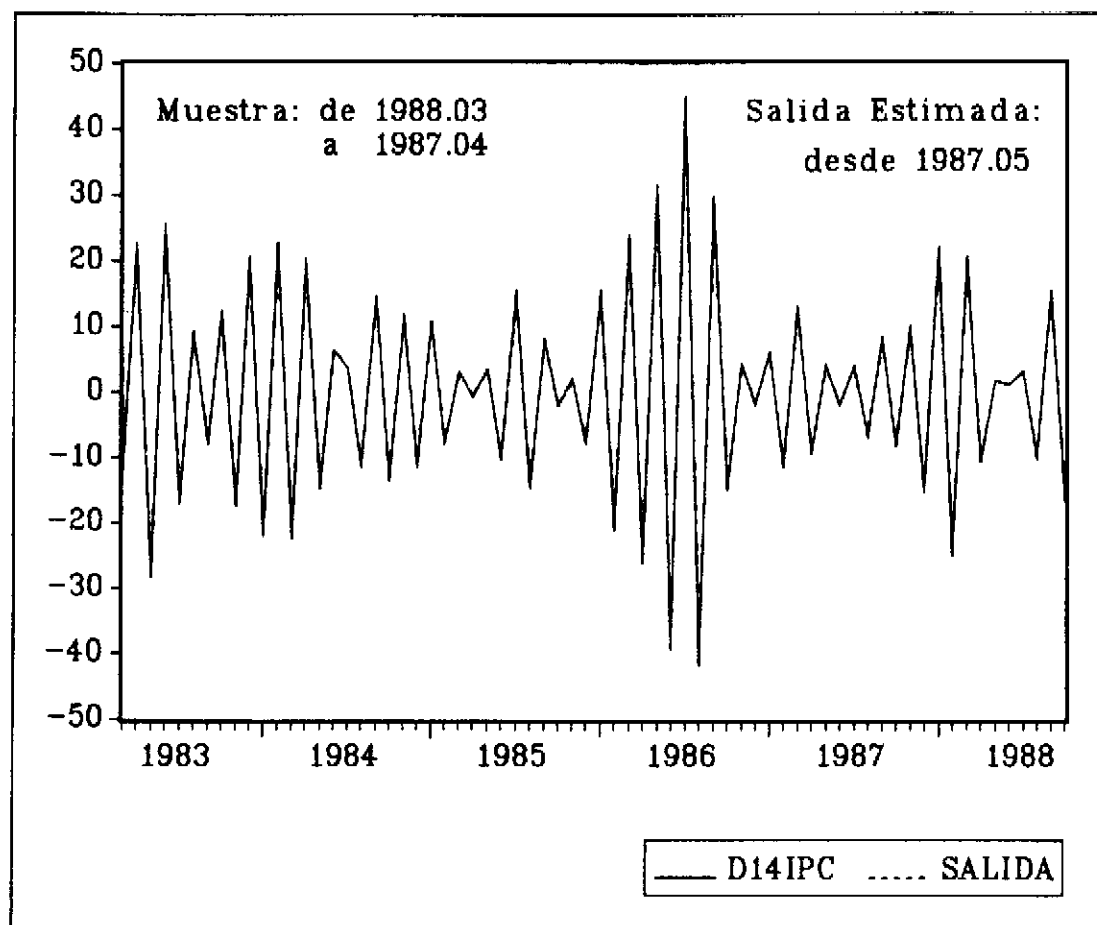


Fig. 3.- Serie IPC transformada y salida estimada, dentro y fuera del rango muestral.

3.2. Elaboración de predicciones, en base al modelo encontrado.

En el proceso, que se realiza a continuación de elaboración de predicciones, es de destacar que, al utilizar representaciones en espacio de estado, se tiene la ventaja de poder realizar predicciones un paso adelante sin demasiada complicación ya que en el instante k se conoce el estado del instante $k+1$, con lo que, por medio de una simple ecuación matricial se puede resolver el valor del siguiente estimado.

Nos encontramos en algunos casos con el problema de la inversión de las transformaciones efectuadas, que con la ayuda del MATLAB y utilizando productos de matrices, basados los coeficientes utilizados en desarrollos binómicos, se reduce notoriamente; aunque no se puede afirmar, que en este punto el trabajo no sea un tanto laborioso. En consecuencia, el proceso a seguir puede ser dinámico o estático; cara a una verdadera predicción no se han de corregir los valores estimados con los reales una vez conocidos estos, no obstante, para que el modelo sea más preciso, es conveniente realizar las actualizaciones necesarias con el fin de que los errores de predicción sean mínimos.

En la tabla 1, que se muestra a continuación, aparecen reflejados los valores reales y los estimados por los dos procedimientos, anteriormente mencionados, utilizando lo descrito, con coeficientes hasta el orden 14, que es la diferencia utilizada. En concreto dichos coeficientes han sido obtenidos a partir de los elementos de la fila correspondiente del triángulo de Tartaglia, implementado y manejado a tal efecto.

PERÍODO	IPC-REAL	IPC- ESTIMADO (DINÁMICA)	IPC- ESTIMADO (ESTÁTICA)
1987.5	136.3000	136.4400	136.4359
1987.6	136.3000	136.4693	136.2567
1987.7	137.7000	136.0197	137.6874
1987.8	137.6000	137.6179	137.6602
1987.9	138.9000	133.7804	139.0075

Tabla 1

A la vista de la tabla adjunta se puede observar para 5 pasos adelante una buena predicción en el caso estático y muy buenas predicciones en el caso dinámico, alimentando cada vez la respuesta con valores reales de la variables diferenciadas conocidas. Los errores que se cometen tienen una desviación típica muy pequeña, de ahí, que sean tan precisos los valores. Aunque se trata de un modelo de orden elevado, es bastante sencillo de identificar y están totalmente automatizadas sus respuestas de cara a la obtención de valores de predicción.

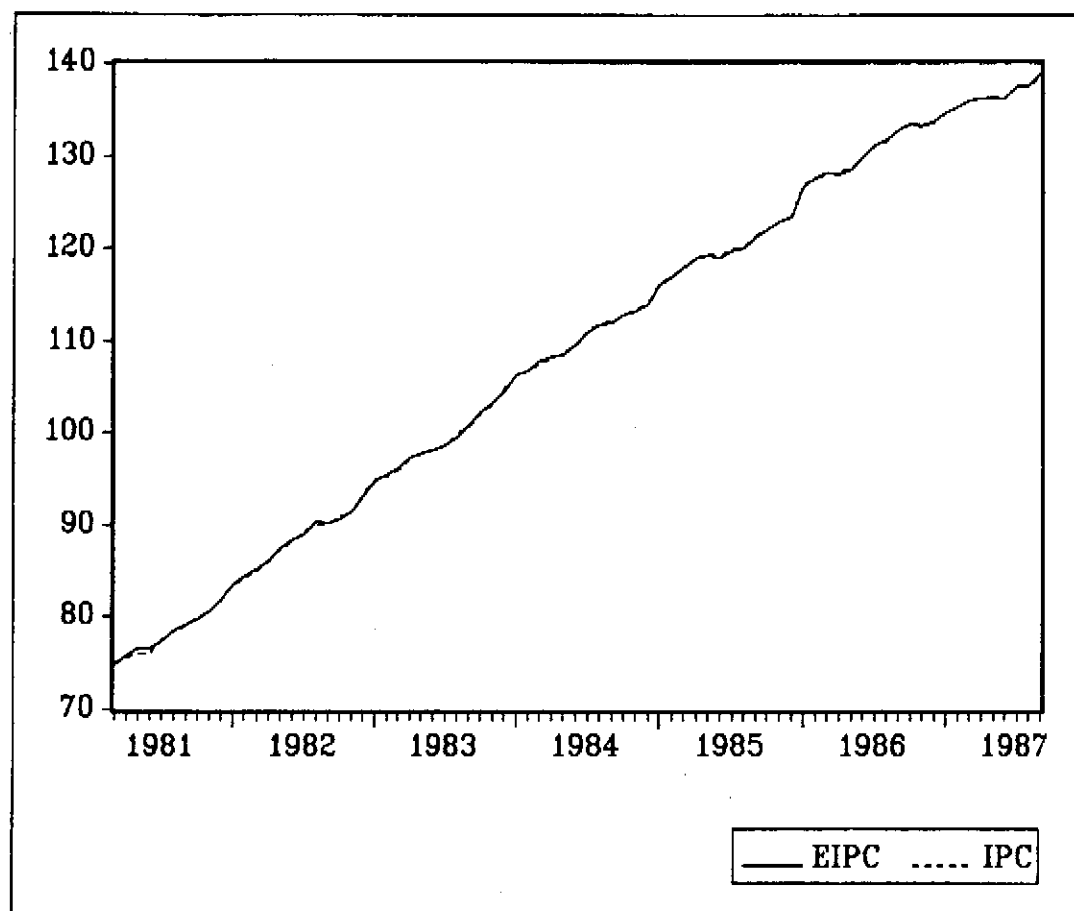


Fig. 4.- Valores reales de la serie IPC y valores estimados, dentro y fuera del rango muestral.

El haber mostrado predicciones sólo 5 pasos adelante, no significa que, el estudio realizado no pueda garantizar predicciones de mayor longitud, todo lo contrario, el estudio predictivo se puede prolongar tanto como se desee, sobre todo en el caso estático. No cabe duda que los errores se irán acumulando y la precisión será cada vez menor, pero, eso ocurre con cualquier tipo de representación en la que se pretenda lograr objetivos a muy largo plazo.

Las dinámicas cambiantes de los procesos difícilmente se pueden reflejar con modelos estáticos y si lo hacen no será de forma general, dependerá mucho de las características del mismo.

En la Fig. 4, se representan predicciones para la serie, los cuales se pueden comparar con los valores conocidos de la serie, fuera del rango muestral; como puede observarse, en dicha gráfica, los valores encontrados, para la serie estimada, se ajustan fielmente a los originales del IPC.

4.- COMPORTAMIENTO PREDICTIVO A LARGO PLAZO.

A continuación, se describen, para la misma serie aunque considerada de otro modo, las estrategias de modelización empleadas para conseguir el tipo de representación elegida para la misma. Los resultados, que se muestran, han sido obtenidos utilizando los procedimientos descritos, siguiendo a Mittnik (1989) y (1990); haciendo gran hincapié en el estudio de predicciones, dado que, estas se hacen para valores de interés económico actual pues son hechas hasta bien entrado el siglo XXI.

Se ha elegido la misma serie que en el epígrafe anterior, pero, se han considerado los índices de forma anual y con mayor amplitud, ya que, se trabaja con datos anuales desde 1961 hasta 1992, realizando predicciones para 14 años fuera del rango, es decir, hasta el año 2006, económicamente interesantes, a pesar de no tener, en mi opinión, un especial significado a demasiado largo alcance, debido a la cantidad de factores que influyen en dicho índice. Una vez encontrado el modelo, los valores de predicción fuera del rango muestral tienen la utilidad de ser predicciones reales y, además, la particularidad de estar basados en un tipo de representación que ha sido poco utilizada para este tipo de aplicaciones, por lo que su contrastación futura parece ser de gran curiosidad. El proceso a seguir es el de encontrar para tal serie, un modelo en espacio de estado balanceado y, a partir de él, se intentarán alcanzar los objetivos planteados de predicción.

Para modelizar el proceso se realizaron diferentes aproximaciones con modelos de distintos órdenes y el aceptado, finalmente como definitivo, ha sido un sencillo modelo de orden dos, que se detalla a continuación. Cabe señalar que, esta misma serie modelada en el apartado anterior, con un modelo de orden elevado, cambia radicalmente con esta forma de tomar la información

de la misma, ya que como se ha dicho, los datos manejados en aquel caso eran mensuales, por lo que el proceso de modelización capta de forma totalmente distinta la dinámica de la serie, debido a que aparecían factores estacionales, fruto del carácter mensual de los mismos, que no se dan en este caso.

A partir de la muestra elegida de 21 datos anuales del IPC, se realizan dos transformaciones, en primer lugar se calculan la segundas diferencias para la serie en estudio, cara a conseguir un proceso estable, siendo esta la serie que se modelizará y en segundo lugar, se considera una primera diferencia de la variable para considerar dichos valores como entradas al sistema.

Determinadas, pues, las señales que van a ser consideradas como entradas y las que van a ser consideradas como salidas, sólo queda por pasar a aplicar el algoritmo de identificación, por mínimos cuadrados, implementado. Las matrices encontradas, en la estimación del modelo del proceso, son las siguientes:

MATRIZ DE PARÁMETROS Y ESTADO INICIAL:

$$\Theta = \begin{bmatrix} -2.1663 \\ 2.9180 \\ -0.7806 \end{bmatrix}$$

$$\text{ESTADO INICIAL} = \begin{bmatrix} 0.0351 \\ -0.1116 \end{bmatrix}$$

MATRIZ DIAGONAL DE LA DESCOMPOSICIÓN SINGULAR:

$$S = \begin{bmatrix} 3.1137 & 0.0 \\ 0.0 & 0.1957 \end{bmatrix}$$

MATRICES DE LA ECUACIÓN DE TRANSICIÓN:

$$A = \begin{bmatrix} -0.2359 & -0.2359 \\ -0.2359 & -0.2359 \end{bmatrix} \quad B = \begin{bmatrix} 1.7116 \\ -0.1076 \end{bmatrix}$$

MATRICES DE LA ECUACIÓN DE OBSERVACIÓN:

$$C = \begin{bmatrix} 1.7116 \\ 0.1076 \end{bmatrix} \quad D = [-2.1663]$$

A partir de esas matrices, estimadas en base a los datos de la muestra elegida de la serie IPC, se puede pasar a analizar el modelo y realizar predicciones, ya que, es el principal objetivo y se conocen todos los instrumentos necesarios para llevarlas a cabo. Las nuevas entradas, en la etapa de predicción, serán consideradas siempre a partir de valores diferenciados de las variables estimadas. Es decir se trata de un proceso real de predicción, ya que se desconocen los valores que tal serie va a tomar en los años venideros, ver Fig. 5.

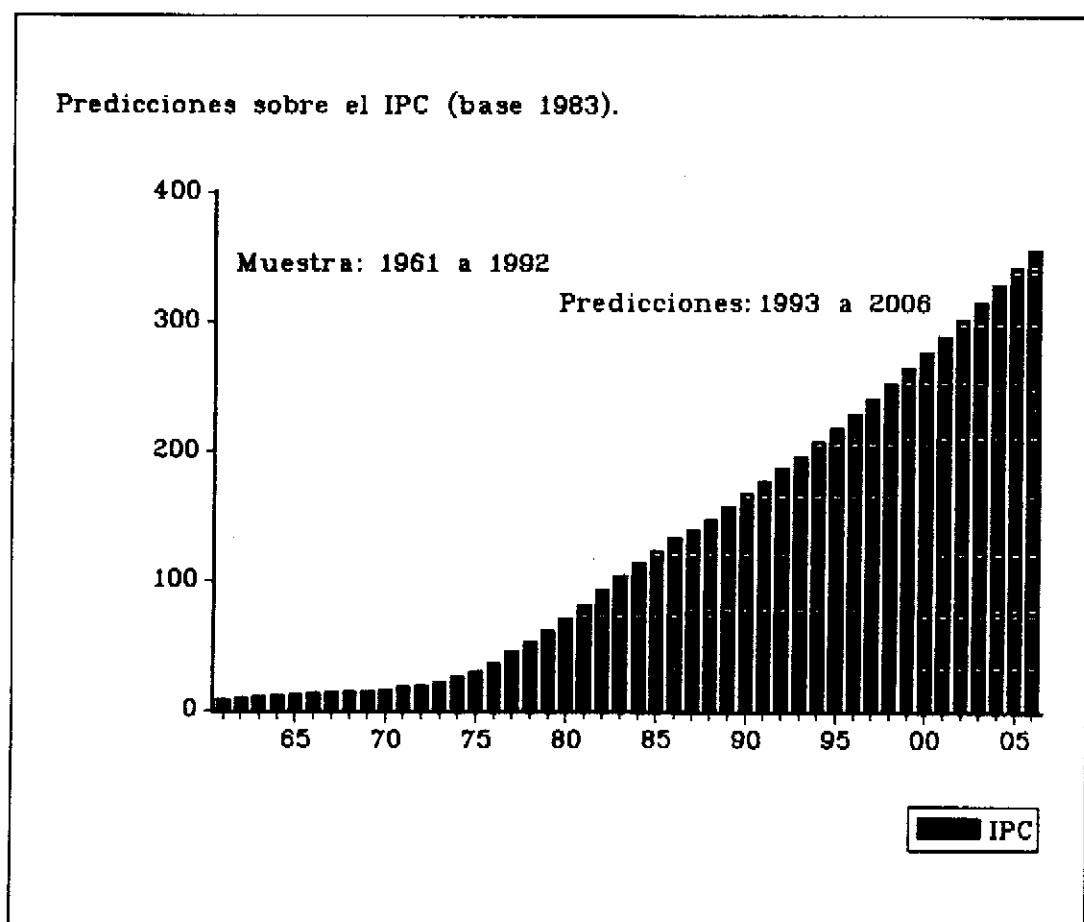


Fig. 5.- Proceso IPC, predicciones 1993 al 2006.

A la vista de los datos encontrados para la serie en estudio del IPC, se predicen unos valores de comportamiento decrecientes décima a décima para el mismo, de manera continuada desde el valor actual.

Partiendo de la realización de predicciones a partir del año 92, y con una predicción de un índice para final del año 93 muy próxima al valor real, considerando la posibilidad de error que nos da el modelo, parece que, tal índice no caería por debajo del 4.6 para final del año, en la situación mas favorable y observando el comportamiento en años sucesivos, podemos afirmar que en base al modelo encontrado, los índices estarían en una banda de decrecimiento en relación al actual. Es de señalar, que este estudio fue realizado con anterioridad al conocimiento del índice del año 93.

Con los modelos en espacio de estado, las predicciones un paso adelante suelen ser bastante sencillas, ya que siempre nos queda un estado predeterminado sin utilizar, que aglutina información suficiente de la historia pasada del proceso, de forma que, con las consiguientes entradas futuras del mismo, señales que podrían ser en algunos casos de tipo totalmente aleatorio, nos dará valores de predicción un paso adelante, en general, bastante fiables.

El que ciertos modelos sean necesarios para alcanzar predicciones a largo plazo, no resta generalidad a lo anteriormente dicho, dado que, aunque las predicciones un paso adelante son bastante fiables, utilizando este tipo de representación, siempre es posible seguir estimando valores para la serie; aunque en algunos casos con pérdida de precisión, como en todos los procesos de este tipo y, por consiguiente, con la correspondiente pérdida de interés.

5.- CONCLUSIONES.

Cabe resumir, de este análisis de procesos temporales univariantes con problemas de estacionalidades o sin ellos, que admiten, sin lugar a dudas, el tipo de parametrización elegida para su representación. Los modelos encontrados, en general, están vinculados con la estacionalidad a la hora de determinar su orden, quizá sea esta una de las conclusiones más importantes en aras a captar la dinámica existente, ya que, por otro lado, el tratamiento final es similar a otros casos sin esta problemática y su utilidad cara al comportamiento predictivo ha quedado mostrada. Los estudios realizados que tratan problemas concretos, para los que se desean realizar predicciones a largo plazo, ver Suárez R. (1994a), han mostrado su efectividad sin ningún problema;

de manera sucinta se puede decir, que todo se reduce a la realización de un cierto número de cálculos matriciales, embebidos en un ciclo for ... end , que con la ayuda del MATLAB se llevan a cabo de forma automática.

El hecho de tener que trabajar con modelos de órdenes elevados, en determinados casos, no influye para nada en el algoritmo; la única posible desventaja radica en el número de datos que son necesarios para trabajar con este tipo de modelos si son de orden elevado, pues, según el orden, se suelen necesitar cuatro veces en número los datos a manejar, pero eso, también es un problema en otros tipos de representaciones. Las ventajas que tiene este tipo de representación elegido son diversas, la longitud de las series no suele ser problema, los modelos encontrados cumplen los objetivos para los que son construidos, la manejabilidad de los modelos es evidente, ya que sólo es necesario conocer las matrices para representarlos, simularlos y calcular predicciones, haciendo uso, por supuesto, del soporte de software citado que ofrece el apoyo para toda la implementación; lo que evita, en cierta medida, la subjetividad del modelizador aunque cuentan con la desventaja, hasta el momento, de no ser de aplicación generalizada, debido a que las aplicaciones tienen que ser realizadas con la elaboración del propio software, para el entorno elegido, el MATLAB por ejemplo en este caso, por otro lado reutilizable una vez elaborado; siendo esta quizá la tarea que dificulta un poco su aplicación y divulgación en la actualidad.

BIBLIOGRAFÍA.

- ALCAIDE, A. y ÁLVAREZ, N. (1992). *Econometría. Modelos Deterministas y Estocásticos. Teoría y práctica*. Editorial CEURA. Madrid.
- ASTRÖM, K. J. and WITTENMARK, B. (1984). *Computer Controlled Systems. Theory and Design*. Prentice-Hall Inc.
- AOKI, M. (1990). *State Space Modelling of Time Series*. Springer-Verlag. Berlin.
- BOX, G. E. P. and JENKINS, G. M. (1976). *Time Series Analysis, Forecasting and Control*. Holden-Day. San Francisco. USA.
- FURUTA, K., SANO, A. and ATHERTON, D. (1988). *State Variable Methods in Automatic Control*. John Wiley & Sons. Great Britain.

- KUNG, L.M. and LIN, D.W. (1981). *Optimal Hankel-norm Reduction: Multivariate Systems*. IEEE Transaction in Automatic Control. AC-26. pp. 832-852.
- MARCUS, M. (1992). *Matrices and MATLAB™*. A tutorial. A Prentice Hall International Student Edition.
- MITTNIK, S. (1989). *Multivariable Time Series Analysis with State Space Models*, in S. Mittnik (ed.). System Theoretic Methods in Economic Modelling I. Oxford: Pergamon Press.
- MITTNIK, S. (1990). *Forecasting with Balanced State Space Representations of Multivariable Distributed Lag Models*. Journal of Forecasting. Wiley Publishers. Vol. 9. pp. 207-218.
- MOORE, B.C. (1981). *Principal Component Analysis in Linear Systems: Controllability, Observability, and Model Reduction*. IEEE Transactions in Automatic Control. AC-26. pp.17-32.
- OGATA, K. (1967). *State Space Analysis of Control Systems..* Prentice-Hall.
- POSKITT, D. S. and TREMAYNE, A.R. (1981). *An approach to testing Linear Time Series Models*. Annals of Statistics, 9, 974-986.
- STRANG, G. (1980). *Linear Algebra*. Academic Press, Inc. New York.
- SUÁREZ R., M. C. y MAYO R., J.L. (1994). *A new Procedure of Modelling. VARMA Processes in Balanced State Space*. 10th IFAC Symposium on System Identification. SYSID'94. Volumen Preprints. Copenhagen. Denmark.
- SUÁREZ R., M. C. (1994a). *Sistemas Automáticos de Diseño en Espacio de Estado para la Modelización de Series Temporales*. Tesis Doctoral. U.N.E.D. Madrid.
- SUÁREZ R., M. C. y MAYOR., J.L. (1994b). *Predicciones usando un Modelo en Espacio de Estado Balanceado*. Un caso de estudio. VII Reunión Anual de ASEPELT. Palma de Mallorca. Junio 1994. España.