

MODELIZACIÓN DE SERIES TEMPORALES ESTACIONARIAS EN ESPACIO DE ESTADOS.

MANUEL VARGAS VARGAS

Área de Estadística. Facultad de Ciencias Sociales de Cuenca.

Universidad de Castilla-La Mancha.

E-mail: mvargas@ecem-ab.uclm.es

MODELIZACIÓN DE SERIES TEMPORALES ESTACIONARIAS EN ESPACIO DE ESTADOS.

Resumen:

Este trabajo aborda el tratamiento de series temporales múltiples a través de la representación en espacio de estados de procesos estocásticos. Aunque mucho menos conocida y utilizada que la representación VARMA y sus variantes, constituye una alternativa equivalente que evita algunas de las dificultades básicas de dicha modelización. Con este enfoque, se desarrollan algoritmos novedosos en el campo de la Economía Cuantitativa, basados en la descomposición en valores singulares, para la determinación de la dimensión del modelo o para la estimación de sus parámetros. Junto a éstos, se utilizan otros algoritmos más conocidos, como el filtro de Kalman, originariamente diseñados para procesos expresados en espacio de estados.

El trabajo se centrará en el análisis de series temporales estacionarias, para las que se exponen los diversos enfoques existentes y se desarrolla una propuesta de modelización. Igualmente, se destacan las ventajas que se presentan sobre los métodos clásicos para procesos autorregresivos de medias móviles y se esbozan las potencialidades de análisis de la representación propuesta.

1.- Introducción.

Con el desarrollo del análisis de series múltiples se han puesto de manifiesto deficiencias en la modelización ARMA, tales como la gran necesidad de parámetros para capturar las relaciones entre las variables o la dificultad de identificar modelos, problemas que aún no se han resuelto de forma satisfactoria. Simultáneamente, y en el campo de la ingeniería, la teoría de sistemas ha desarrollado diversos algoritmos que permiten identificar y modelizar procesos estocásticos de forma distinta a como se viene haciendo en Economía Cuantitativa. Esta formulación de modelos en espacio de estados para series temporales múltiples no es desconocida en Economía; sin embargo su utilización ha estado restringida a objetivos concretos tales como el cálculo “fácil” de la función de verosimilitud o el análisis estructural.

Este paralelismo en el tratamiento de series temporales ha propiciado la aparición de consideraciones y técnicas que, aunque desarrolladas en ámbitos distintos, pueden proporcionar resultados fructíferos en cualquiera de ellos. Este hecho se ha concretado en la aparición de diversos trabajos que pretenden salvar las diferencias metodológicas y terminológicas entre ambos campos, resaltando la equivalencia básica de ambas metodologías. Así, obras como las de Moore (1981), Otter (1985) o Aoki (1987, 1990) han introducido en el campo económico conceptos y algoritmos relacionados con los modelos en espacio de estados que no eran utilizados con anterioridad y que resuelven de forma alternativa problemas básicos en el análisis de series. En algunos casos, estas soluciones son equivalentes a las que se obtienen con el tratamiento clásico mientras que en otros presentan ventajas relativas; son estos últimos casos los que justifican un estudio detallado de esta nueva metodología.

Una de las diferencias fundamentales reside en la fase de identificación⁽¹⁾ donde se han desarrollado algoritmos propios y relativamente desconocidos para la determinación del modelo generador de datos subyacente.

Como puede verse en, por ejemplo, Hannan y Deistler (1988), todo proceso estocástico débilmente estacionario de rango completo y con función de densidad espectral $\Phi_Y(e^{i\cdot})$ sin ceros en la circunferencia unidad, tiene una realización innovacional en espacio de estados:

$$\begin{aligned} X_{t+1} &= FX_t + G\varepsilon_t \\ Y_t &= \mu + HX_t + \varepsilon_t \end{aligned} \quad t \in Z$$

donde μ es el vector de medias de la serie y ε es el proceso innovacional de Y, cuya secuencia de matrices de impulso-respuesta es idéntica a la de coeficientes de la descomposición de Wold de Y. La estructura básica del modelo supone la existencia de un vector de estado, X_t , que actúa en cada instante como estadístico suficiente para la dinámica del sistema. De esta forma, la mejor predicción del vector de observaciones para el siguiente instante del tiempo reviste la forma $\hat{Y}_{t+1|t} = \mu + HX_t$.

La ecuación de transición del modelo anterior representa a un proceso markoviano de primer orden. La renovación del vector de estado consta de dos componentes, la evolución propia de X_t resumida en la matriz F, y la corrección producida por la innovación del proceso a través de la matriz G. De esta forma, se resume el comportamiento dinámico en una ecuación en diferencias de primer orden⁽²⁾.

Esta representación, sin embargo, no es única. Si el determinante de la función de densidad espectral no posee ceros en el intervalo $[0, 2\pi]$, cualquier representación

¹ En la terminología ARMA este término hace referencia a la determinación de los órdenes del proceso, mientras que en teoría de sistemas tiene una significación más amplia, englobando también a la estimación de las matrices del sistema.

² Esta singularidad permite obviar la determinación del número de retardos de la serie que pueden presentar una correlación significativa con el valor actual: en cada instante, el vector de estado está recogiendo toda la información relevante, por atrasada que ésta sea.

innovacional invariante de dimensión mínima vendrá dada por:

$$\begin{aligned} X_{t+1} &= F(S)X_t + G(S)\varepsilon_t \\ Y_t &= \mu + H(S)X_t + \varepsilon_t \end{aligned} \quad t \in Z$$

con $F(S) = SFS^{-1}$, $H(S) = HS^{-1}$, $G(S) = SG$ y $S \in Gl(n)$. Una implicación fundamental de este resultado es que todos los algoritmos que aseguren una representación minimal en espacio de estados para un proceso estocástico están dando soluciones equivalentes⁽³⁾. A pesar de esto, no todas las soluciones son igualmente fáciles de obtener, lo que justifica la existencia de varios métodos alternativos. Entre ellos, se ha popularizado el basado en una representación “canónica” del proceso en la metodología VARMA. En este trabajo se pretenden desarrollar otros algoritmos originarios de la teoría de sistemas y que buscan el mismo fin. Así, para la representación de procesos en espacio de estados se considerará de un modelo innovacional de dimensión p de la forma⁽⁴⁾:

$$\begin{aligned} X_{t+1} &= FX_t + G\varepsilon_t \\ Y_t &= \mu + HX_t + \varepsilon_t \end{aligned} \quad t \in Z \tag{1}$$

llamada representación innovacional ya que ε_t se interpreta como el error de predicción a un período de Y_t dados los valores Y_s para $s < t$, por lo que son variables incorreladas. Además y por construcción, el vector de innovaciones ε_t no presenta autocorrelación⁽⁵⁾.

Para analizar y estimar el modelo dado en (1) se han propuesto diversos algoritmos que, aunque basados en una idea común, desarrollan de forma distinta la identificación de las matrices del sistema, y que pueden consultarse en Vargas (1999). Para su exposición, el resto del trabajo se estructurará como sigue: en el epígrafe segundo se abordará la

³ Esta equivalencia consiste en que es posible encontrar una matriz no singular que relaciona unas matrices con otras, por lo que las únicas diferencias consisten en la utilización de bases distintas en el espacio vectorial de estados utilizadas para expresar el sistema.

⁴ Se supone, sin pérdida de generalidad, que la serie Y_t está centrada, por lo que no aparecerá dicha media en la formulación del proceso. Además, como se destacará más adelante, es usual reescalar la serie Y_t dividiendo cada componente por su desviación típica.

⁵ Esta forma particular de la representación en espacio de estados no supone ninguna restricción, ya que cualquier otra expresión puede convertirse fácilmente a ésta (Hannan y Deistler, 1988).

determinación de la dimensión del sistema; el tercero recogerá los algoritmos de estimación de las matrices del modelo; la estimación del vector de estado será el tema del cuarto epígrafe y en el quinto se estudiará la determinación de las condiciones iniciales del filtrado y la predicción de valores futuros. Por último, el epígrafe seis resumirá las principales conclusiones.

2.- Determinación de la dimensión del sistema.

Una de los aspectos fundamentales en la modelización de series temporales consiste en la estimación de la dimensión del modelo, decisión que condiciona todo el desarrollo posterior. Esta etapa es exclusiva de los algoritmos desarrollados en teoría de sistemas, ya que los modelos en espacio de estados comunes en Economía Cuantitativa son estructurales, es decir, suponen conocida la estructura dinámica del proceso generador de los datos, limitándose a la estimación de los parámetros que rigen dicha estructura, como ocurre en los trabajos de Harrison y Stevens (1976), Harvey (1981) o West y Harrison (1989).

En el campo de los modelos estáticos, trabajos como el de Bartlett (1939), Lawley (1959) o Rao (1965, 1979), utilizan los coeficientes de correlación canónica y su distribución estadística para determinar el número de factores comunes en modelos multidimensionales. La generalización a modelos dinámicos se realiza en los trabajos de Akaike (1974a, 1974b, 1975, 1976), donde introduce variables canónicas entre los valores pasados y futuros de las series para construir modelos en espacio de estados, lo que permite seguir utilizando las propiedades estadísticas de los coeficientes de correlación canónica. Posteriormente, se han desarrollado otros algoritmos que, bien matizando el anterior, bien basados en criterios de información, han ampliado el número de herramientas disponibles para la identificación del orden de un sistema. A pesar del gran trabajo, teórico y de simulación, que se ha llevado a cabo en los últimos años, no se ha presentado una clara supremacía de ninguno de los métodos, debiendo recurrir frecuentemente a la utilización conjunta de varios de ellos para minimizar los posibles errores de especificación.

Para analizar más detalladamente estos algoritmos, considérense los vectores de valores futuros y pasados de la serie:

$$\begin{aligned} Y_t^+ &= (Y_t', Y_{t+1}', Y_{t+2}' \dots)' \\ Y_{t-1}^- &= (Y_{t-1}', Y_{t-2}', Y_{t-3}' \dots)' \end{aligned} \quad (2)$$

y sean R_+ y R_- las correspondientes matrices de correlación⁽⁶⁾. Si por $\Gamma_l = E[Y_{t+l} Y_t']$ para $l = 0, 1, 2, \dots$ se denotan a las matrices de autocorrelación, se puede expresar la igualdad:

$$\mathcal{H} = E[Y_t^+ Y_{t-1}^{-'}] = \begin{pmatrix} \Gamma_1 & \Gamma_2 & \Gamma_3 & \dots \\ \Gamma_2 & \Gamma_3 & \Gamma_4 & \dots \\ \Gamma_3 & \Gamma_4 & \Gamma_5 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \quad (3)$$

donde la matriz de autocorrelaciones \mathcal{H} es de tipo Hankel por bloques, es decir, con los elementos de las contradiagonales iguales.

Por el teorema de Kronecker (Golub y Van Loan (1983)) la dimensión mínima para el vector de estado coincide con el rango de la matriz \mathcal{H} ⁽⁷⁾. Para determinar de forma empírica el rango de esta matriz se puede utilizar la descomposición en valores singulares (DVS a partir de este momento) propuesta por primera vez en el trabajo de Eckart y Young (1936): dada la matriz \mathcal{H} es posible encontrar una descomposición de la forma $\mathcal{H} = U \Sigma V$ donde $U'U = VV' = \text{Id}$ y Σ es una matriz diagonal con elementos (llamados valores singulares) no negativos y ordenados en forma decreciente. Por la estructura de estas matrices, la dimensión de \mathcal{H} coincide con el número de valores singulares no nulos. Para evitar la utilización de matrices de orden infinito, se suelen truncar los vectores de datos,

⁶ Hay que recordar que, al estar tipificadas todas las series, es equivalente hablar de correlación y de covarianza.

⁷ Al mismo resultado se puede llegar considerando la estructura canónica implicada por los índices de Kronecker, ya que su suma coincide con el grado de McMillan y, por tanto, con la dimensión mínima del vector de estado.

tanto futuros como pasados en la forma:

$$Y_t^+ = (Y_t', Y_{t+1}', \dots, Y_{t+N_f-1}')'$$

$$Y_{t-1}^- = (Y_{t-1}', Y_{t-2}', \dots, Y_{t-N_p}')'$$

que definen la matriz⁽⁸⁾:

$$\mathcal{H}_{N_f}^{N_p} = \begin{pmatrix} \Gamma_1 & \Gamma_2 & \dots & \Gamma_{N_p} \\ \Gamma_2 & \Gamma_3 & \dots & \Gamma_{N_p+1} \\ \vdots & \vdots & \vdots & \vdots \\ \Gamma_{N_f} & \Gamma_{N_f+1} & \dots & \Gamma_{N_p+N_f-1} \end{pmatrix} \quad (4)$$

de dimensión ($pN_f \times pN_p$), con las que se puede enunciar el siguiente teorema (Hannan y Deistler, 1988):

Teorema: Un sistema en espacio de estados (F, H, G) tiene dimensión mínima (n) si y sólo si $\text{rango}(\mathcal{H}) = n$.

Fijando valores de N_p y N_f suficientemente grandes⁽⁹⁾ y estimando la matriz $\mathcal{H}_{N_f}^{N_p}$ mediante las matrices de autocorrelaciones muestrales $\hat{\Gamma}_l = T^{-1} \sum_{t=1}^T Y_{t+l} Y_t'$ se puede

descomponer la matriz $\hat{\mathcal{H}}_{N_f}^{N_p} = \hat{U} \hat{\Sigma} \hat{V}'$ y utilizar como estimación de su rango el número de

valores singulares de la matriz $\hat{\Sigma}$ significativamente positivos.

⁸ Con esta estructura se estaría analizando la relación lineal entre N_f valores futuros ($Y_t, Y_{t+1}, \dots, Y_{t+N_f-1}$) y N_p valores pasados ($Y_{t-1}, Y_{t-2}, \dots, Y_{t-N_p}$). Generalmente, se consideran valores iguales para N_f y N_p .

⁹ Si fuesen inferiores a la verdadera dimensión del sistema, sería imposible determinarla correctamente. Sin embargo, suele ser habitual el planteamiento contrario, es decir, imponer que el rango “estimado” no sea superior a los valores elegidos (Aoki, 1987).

Para esto, considerando los vectores normalizados $d_t^+ = R_+^{-1/2} Y_t^+$ y $d_{t-1}^- = R_-^{-1/2} Y_{t-1}^-$, la matriz de covarianza⁽¹⁰⁾ de estos nuevos vectores se puede expresar como una versión reescalada de la matriz tipo Hankel $\hat{\mathcal{H}}_{N_f}^{N_p}$ mediante $E[d_t^+, d_{t-1}^-] = R_+^{-1/2} \hat{\mathcal{H}}_{N_f}^{N_p} R_-^{-1/2}$. Calculando su descomposición en valores singulares $E[d_t^+, d_{t-1}^-] = P\Gamma Z'$, en la diagonal de la matriz Γ aparecen los coeficientes de correlación canónica ordenados de forma decreciente. Ahora, se pueden definir las variables canónicas rotando los vectores de datos mediante las matrices de la descomposición en valores singulares, obteniendo:

$$u_t^+ = P' d_t^+, \quad y \quad u_{t-1}^- = Z' d_{t-1}^- \quad (5)$$

que son utilizados por Akaike como vectores de estado en su propuesta de modelización en espacio de estados.

Dada la estructura de las matrices P y Z, la dimensión de las variables canónicas coincide con el rango de la matriz Γ , es decir, con el número de coeficientes de correlación canónica no nulos, n. Así pues, la dimensión del modelo será n y no la dimensión de los vectores de datos d_t . En la práctica, la determinación de cuántos coeficientes son positivos puede ser una tarea muy difícil debido a la variabilidad muestral. Por ello, se han propuesto varios enfoques para determinar la dimensión del vector de estado. El primero consiste en utilizar una aproximación a la distribución estadística de los coeficientes de correlación canónica propuesta en Bartlett (1939), donde, bajo la hipótesis nula de que sólo hay n coeficientes estrictamente positivos, se tiene que:

$$B_n = \left\{ T - \frac{1}{2} [p(N_p + N_f) + 1] \right\} \ln \prod_{j>n+1} (1 - \gamma_j^2) \rightarrow \chi_{(pN_f - n)(pN_p - n)}^2 \quad (6)$$

donde N_p es el número de retardos considerados, N_f es el horizonte de predicción y p es la dimensión del proceso. De esta forma, podemos plantear contrastes de hipótesis

¹⁰ Debido a la tipificación hecha, esta matriz es, en realidad, la de autocorrelaciones.

secuencialmente para $n = 1, 2, \dots$ hasta que no se rechace la hipótesis de nulidad del estadístico en un valor n^* que se considerará como la dimensión del modelo en espacio de estados.

Posteriormente, en Lawley (1959) o en Glynn y Muirhead (1978) se sugiere una modificación de (6) para mejorar la aproximación, proponiendo la expresión conocida como estadístico de Bartlett-Lawley:

$$L_n \equiv - \left[T - n - \frac{1}{2} ((N_p + N_f) p + 1) + \sum_{j=1}^n \gamma_j^{-2} \right] \ln \prod_{j>n} (1 - \gamma_j^2) \rightarrow \chi_{(N_f p - n)(p N_p - n)} \quad (7)$$

que, en estudios de Montecarlo, se ha revelado más preciso que el originario de Bartlett⁽¹¹⁾.

Otro enfoque para la determinación de la dimensión del modelo en espacio de estados se basa en criterios de información. Particularmente, en el trabajo de Gel'fand y Yaglom (1959) se puede comprobar que la información mutua⁽¹²⁾ entre las variables canónicas viene dada por la ecuación:

$$I(u_t^+, u_{t-1}^-) = -\ln \det(Id - \Gamma^2) = -\sum_{i=1}^n \ln(1 - \gamma_i^2) \quad (8)$$

Así, basándose en esta igualdad, Desai y Pal (1983) proponen el cociente:

$$\frac{\sum_{i=1}^n \ln(1 - \gamma_i^2)}{\sum_{i=1}^p \ln(1 - \gamma_i^2)} \quad (9)$$

¹¹ El uso de estas distribuciones asintóticas se ve limitado en la práctica por su sensibilidad al supuesto de normalidad en el proceso. En Muirhead y Waternaux (1980) puede encontrarse una extensión al caso de distribuciones no normales.

¹² La definición de información mutua puede verse en Ibragimov y Rozanov (1978) o en Jewell y Bloomfield (1983).

como criterio para elegir la dimensión del modelo en espacio de estados⁽¹³⁾.

También basado en la función de información mutua, Li y Xie (1996) desarrollan un algoritmo, conocido como LIC y que, dado un conjunto de matrices de autocorrelación determina el modelo (originariamente autorregresivo) que minimiza la información entre pasado y futuro⁽¹⁴⁾.

Otros criterios basados en la función de información de Kullback-Leibler son más conocidos y utilizados, sobre todo en la estimación del orden de un modelo autorregresivo. Básicamente centrados en la obtención de una aproximación a un estimador insesgado para la esperanza de esta función de información, se desarrollan a partir del trabajo pionero de Akaike (1973, 1974a, 1974b), donde se propone el conocido criterio AIC⁽¹⁵⁾ basado en la aproximación de la anterior esperanza por el desarrollo de Taylor de primer orden (Akaike, 1973; Linhart y Zucchini, 1986). Posteriormente se han sugerido modificaciones para mejorar dicha aproximación, como pueden ser criterios propuestos por Brockwell y Davis (1991) o por Hurvich y Tsai (1989, 1993), o bien propuestas que basadas en ideas semejantes han aportado criterios tales como el de Schwarz (1978) o el de Hannan y Quinn (1979).

3. Estimación de las matrices del sistema.

Tras la determinación de la dimensión del vector de estado es posible abordar la estimación de las matrices del sistema. En primer lugar, se va a exponer un algoritmo basado en la teoría de sistemas y desarrollado originariamente en Aoki (1987, 1990). Para

¹³ Se trata de elegir los estados suficientes como para preservar un porcentaje suficientemente alto de la información estadística existente en los datos.

¹⁴ Esta función de información es un caso particular de la de información mutua.

¹⁵ Akaike's Information Criterion.

ello, se parte del modelo innovacional dado por la ecuación (1)⁽¹⁶⁾:

$$\begin{aligned} X_{t+1} &= FX_t + G\varepsilon_t \\ Y_t &= HX_t + \varepsilon_t \end{aligned} \quad t \in Z$$

a partir del cual se pueden expresar las matrices de autocorrelación como $\Gamma_l = HF^{l-1}\Omega$

donde $\Omega = E[X_{t+1}Y_t']$ sin más que utilizar el modelo de forma recursiva. Si ahora se define

la matriz de observación mediante $O' = (H', F'H', \dots, (F')^{m-1}H')$ y la matriz⁽¹⁷⁾

$\Upsilon = (\Omega, F\Omega, \dots, F^{m-1}\Omega)$, se obtiene la igualdad $\mathfrak{H}_m^m = O\Upsilon$ ⁽¹⁸⁾.

Calculando la esperanza del valor futuro de la serie dados los valores pasados, se obtiene que $E[Y_t^+ | Y_{t-1}^-] = \hat{\mathfrak{H}}_m^m R^{-1} Y_{t-1}^-$ donde $R = E[Y_{t-1}^- Y_{t-1}^{-'}]$. De esta forma, la proyección ortogonal del vector de valores futuros en el espacio vectorial generado por Y_{t-1}^-

se puede expresar mediante la matriz de observación como $E[Y_t^+ | Y_{t-1}^-] = OX_t$. De aquí

igualando términos, descomponiendo la matriz $\hat{\mathfrak{H}}_m^m$ y simplificando, se obtiene la ecuación de estimación de los estados:

$$\Upsilon R^{-1} Y_{t-1}^- = X_t \quad (10)$$

Para estimar las matrices del sistema, definiendo la matriz $J' = (I_{(p \times p)}, 0, \dots, 0)$, se pueden obtener los resultados:

$$\begin{aligned} 1- \hat{E}[Y_t | Y_{t-1}^-] &= J' \hat{O} \hat{\Upsilon} \hat{R}^{-1} Y_{t-1}^- = \hat{H} \hat{\Upsilon} \hat{R}^{-1} Y_{t-1}^- \text{ , de donde se deduce que} \\ \hat{E}[\hat{Y}_t | Y_{t-1}^-] &= (\hat{\Gamma}_1, \hat{\Gamma}_2, \dots, \hat{\Gamma}_m) = J' \hat{\mathfrak{H}}_m^m = \hat{H} \hat{\Upsilon}. \end{aligned} \quad (11)$$

¹⁶ Ya reexpresado sin la media.

¹⁷ En el desarrollo del algoritmo esta matriz jugará un papel parecido al de la matriz de control, semejanza que justifica todo el algoritmo y que, como se comentará más adelante, no está exenta de problemas.

¹⁸ A partir de este momento se va a asumir que el número de filas y columnas en la matriz de tipo Hankel es igual, restricción no necesaria pero que simplifica tanto la exposición como los cálculos posteriores.

$$2.- \hat{E}[Y_t^+ Y_{t-1}^{-'}] = (\hat{\Gamma}_1, \hat{\Gamma}_2, \dots, \hat{\Gamma}_m)' = \hat{\mathcal{H}}_m^m J = \hat{O} \hat{\Omega}. \quad (12)$$

$$3.- \hat{E}[Y_{t+1}^+ | Y_{t-1}^-] = \hat{O} \hat{F} X_t = \hat{O} \hat{F} \hat{\Upsilon} \hat{R}^{-1} Y_{t-1}^-.$$

En esta expresión, si se define la matriz $\mathcal{H}^l =$
$$\begin{pmatrix} \hat{\Gamma}_2 & \hat{\Gamma}_3 & \dots & \hat{\Gamma}_{m+1} \\ \hat{\Gamma}_3 & \hat{\Gamma}_4 & \dots & \hat{\Gamma}_{m+2} \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\Gamma}_{m+1} & \hat{\Gamma}_{m+2} & \dots & \hat{\Gamma}_{2m} \end{pmatrix}$$
 se obtiene

la expresión:

$$\mathcal{H}^l = \hat{O} \hat{F} \hat{\Upsilon} \quad (13)$$

De estas tres ecuaciones se pueden extraer las estimaciones de las matrices F, H y Ω de forma implícita. Para despejarlas se necesita invertir las matrices \hat{O} y $\hat{\Upsilon}$. Sin embargo, éstas pueden no ser invertibles, por lo que se han de utilizar matrices inversas generalizadas (Golub y Van Loan, 1983). Como se sabe, la inversas generalizadas no son únicas, por lo que se ha de hacer alguna elección. Aquí entra en juego la descomposición en valores singulares de la matriz \mathcal{H}_m^m . Como $\hat{\mathcal{H}}_m^m = \hat{O} \hat{\Upsilon} = \hat{U} \hat{\Sigma} \hat{V}'$, se pueden considerar las matrices $\hat{\Upsilon}^- = \hat{V} \hat{\Sigma}^{-1/2}$ y $\hat{O}^- = \hat{\Sigma}^{-1/2} \hat{U}'$ como inversas generalizadas, lo que supone una parametrización del vector de estado (Aoki y Havenner, 1992). Con ello, las estimaciones para las matrices del sistema adoptan la forma:

$$\hat{F} = \hat{O}^- \mathcal{H}^l \hat{\Upsilon}^- \quad \hat{H} = \hat{\mathcal{H}}_1^l \hat{\Upsilon}^- = J' \hat{O} \quad \hat{\Omega} = \hat{O}^- \hat{\mathcal{H}}_1^l = \hat{\Upsilon} J \quad (14)$$

donde $\hat{\mathcal{H}}_1^l$ y $\hat{\mathcal{H}}_1^l$ representan al primer bloque de p filas y al primer bloque de p columnas de la matriz $\hat{\mathcal{H}}_m^m$ respectivamente.

Con estos resultados, sólo queda por estimar la matriz G en el modelo (1), matriz de filtrado que incorpora las innovaciones al vector de estado. Para ello, basándose en los

trabajos de Vaughan (1970) y Laub (1983), se definen las matrices $\Theta = E[X_t X_t']$ y $\Psi = E[\varepsilon_t \varepsilon_t']$ y utilizando las ecuaciones del modelo podemos expresar las igualdades:

$$\begin{aligned} - \Theta &= F\Theta F' + G\Psi G' \\ - \Psi &= \Gamma_0 - H\Theta H' \\ - G\Psi &= \Omega - F\Theta H' \end{aligned}$$

Uniendo todas las ecuaciones y sustituyendo las dos últimas en la primera se obtiene la ecuación de tipo Riccati dada por:

$$\Theta = F\Theta F' + (\Omega - F\Theta H')(\Gamma_0 - H\Theta H')^{-1}(\Omega - F\Theta H')' \quad (15)$$

que permite calcular estimaciones $\hat{\Theta}$, $\hat{\Psi}$ y \hat{G} condicionadas a las estimaciones anteriores \hat{F} , $\hat{\Gamma}_0$, $\hat{\Omega}$ y \hat{H} .

Para la resolución de la ecuación (15), siguiendo a Aoki (1987), se define la matriz $D = F' - H'\Gamma_0^{-1}\Omega'$ y con ella la nueva matriz:

$$S^* = \begin{pmatrix} D - H'\Gamma_0^{-1}HD^{-1}\Omega\Gamma_0^{-1}\Omega' & H'\Gamma_0^{-1}HD^{-1} \\ -D^{-1}\Omega\Gamma_0^{-1}\Omega' & D^{-1} \end{pmatrix} \quad (16)$$

Expresando esta matriz en su descomposición de Schur $Q'S^*Q = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix}$

donde cada submatriz S_{ij} es cuasitriangular superior y particionando la matriz Q de vectores de Schur en forma similar a S^* , se puede probar (Vaughan, 1970) que la matriz dada por $\hat{\Theta} = Q_{21}Q_{11}^{-1}$ es solución de la ecuación de Riccati (15). Así, con la estimación de esta matriz, se pueden obtener sucesivamente las estimaciones $\hat{\Psi}$ y \hat{G} , completando la fase de estimación del modelo.

Este algoritmo, tal y como se presenta en Aoki (1987) o Aoki y Havenner (1992), está inspirado en un método de representación y simplificación de modelos en espacio de estados para sistemas input-output lineales y determinísticos, desarrollado en teoría de control en Popov (1969) o Rosenbrock (1970) y que puede encontrarse en trabajos como los de Moore (1981) o Pernebo y Silverman (1982). Las ventajas fundamentales de este algoritmo se centran en las propiedades de anidamiento y estabilidad. La primera de ellas establece una relación entre las estimaciones para modelos equilibrados⁽¹⁹⁾ de diferentes órdenes, pieza fundamental cuando se pretende una reducción de la dimensión del vector de estado que conserve “en la medida de lo posible” las características de la función de transferencia del modelo. De hecho, si se particiona la matriz $\mathcal{H} = O\Upsilon$ en la forma

$O = [O_1 \ O_2]$, $\Upsilon = \begin{bmatrix} \Upsilon_1 \\ \Upsilon_2 \end{bmatrix}$ y denotamos por U_1 , Σ_1 y V_1 a las matrices correspondientes a los

n_1 valores singulares mayores, es decir, $O_1 = U_1 \Sigma_1^{-\frac{1}{2}}$ y $\Upsilon_1 = \Sigma_1^{-\frac{1}{2}} V_1'$, por la ortogonalidad

de las matrices U y V se obtiene (Aoki, 1987):

$$O_1^{-1} \mathcal{H}^F \Upsilon_1^{-1} = \Sigma_1^{-\frac{1}{2}} U_1' [U_1 \Sigma_1^{-\frac{1}{2}} \ U_2 \Sigma_2^{-\frac{1}{2}}] F \begin{bmatrix} \Sigma_1^{-\frac{1}{2}} V_1' \\ \Sigma_2^{-\frac{1}{2}} V_2' \end{bmatrix} V_1 \Sigma_1^{-\frac{1}{2}} = [I_{n_1} \ 0] F \begin{bmatrix} I_{n_2} \\ 0 \end{bmatrix} = F_{11} \quad (17)$$

donde $\mathcal{H}^F = OF\Upsilon$ es la matriz obtenida calculando la covarianza entre Y_{t+1}^+ e Y_t^- ; la matriz

F_{11} es el cuadrante superior izquierdo de la matriz F y permanece invariante si n_1 aumenta.

También $H = \mathcal{H}^H \Upsilon = \mathcal{H}^H [V_1 \Sigma_1^{-\frac{1}{2}} \ V_2 \Sigma_2^{-\frac{1}{2}}] = [H_1 \ H_2]$.

Esta propiedad de anidamiento garantiza que si se especifica un tamaño para el vector de estado inferior al correcto, las estimaciones de las matrices F , H y Ω serían una aproximación de menor orden a las verdaderas matrices del sistema, proporcionando un

¹⁹ El equilibrio del modelo viene garantizado por la elección de matrices inversas hecha en (9).

modelo “aproximado” pero con menor dimensión que el verdadero. Por contra, si la dimensión del modelo estimado es superior a la correcta, se sabe que una parte de las matrices del sistema es redundante, manteniéndose el resto válida, por lo que no es necesario recalcularlas⁽²⁰⁾.

Otra ventaja importante reside en la estabilidad de los submodelos. En Pernebo y Silverman (1982) se puede ver que todos los submodelos equilibrados de uno asintóticamente estable son, así mismo, asintóticamente estables. Este hecho, junto a la propiedad de anidamiento, facilita la obtención y análisis de submodelos, situación de interés cuando se busca la simplicidad en la interpretación de los datos.

Por otro lado, el desarrollo original utiliza la matriz de Hankel expresándola en términos de la tripleta de matrices $\{F, H, G\}$, que describe a un sistema input-output. Sin embargo el presente algoritmo utiliza la tripleta $\{F, H, \Omega\}$ que no siempre describe un proceso estocástico. Para que así sea se ha de cumplir una condición adicional (Faure et al. 1979) que asegura que una tripleta como la utilizada describe a un proceso estocástico si y sólo si la inecuación matricial:

$$\begin{pmatrix} \Pi - F\Pi F' & \Omega - F\Pi H' \\ \Omega' - H\Pi F' & \Gamma_0 - H\Pi H' \end{pmatrix} \geq 0 \quad (18)$$

se cumple al menos para una matriz semidefinida positiva Π . Se puede probar que esta condición es equivalente a imponer la inecuación $(G' Id)\Psi(G' Id) \geq 0$ (Heij, Kloek y Lucas, 1992). La no negatividad de la expresión anterior puede ser expresada también

utilizando la función de densidad espectral de un proceso como:

²⁰ Esta propiedad se debe a la introducción de la condición de ortogonalidad en la estimación de las matrices y no en el vector de estado, en el que puede existir correlación entre los componentes. Este hecho diferencia el algoritmo del propuesto por Akaike (1975), donde las componentes del vector de estado sí son ortogonales pero pierde la propiedad de anidamiento en la estimación de las matrices.

$$S(z) = \sum_{k=-\infty}^{\infty} \Gamma_k z^k = \Gamma_0 + H(zId - F)^{-1} \Omega + \Omega (z^{-1}Id - F')^{-1} H'$$

A su vez, ésta puede ser factorizada como $S(z) = G(z)G'(z^{-1})$ donde $G(z) = \sum_{k=0}^{\infty} G_k z^{-k}$ (Faurre et al. 1979). Esto significa en particular que se cumple la inecuación $S(e^{i\theta}) \geq 0$ para $\theta \in [0, 2\pi)$.

El cumplimiento de esta condición de no negatividad no está garantizado por el algoritmo de estimación propuesto, por que se pueden producir varios problemas (ver Heij, Kloek y Lucas (1992) para un análisis detallado):

- En primer lugar, el ya comentado de la posible no correspondencia entre la tripleta $\{F, H, \Omega\}$ y un proceso estocástico. En particular, se podría encontrar una estimación \hat{F} cuyos autovalores no fueran todos menores que la unidad, con la consiguiente pérdida de la estabilidad del modelo. Por otro lado, en el desarrollo del algoritmo de estimación se ha utilizado implícitamente el hecho de que la matriz de observación es de rango máximo, requisito que no se garantiza en la estimación \hat{O} .

- En segundo lugar, en la resolución algebraica de la ecuación de Ricatti la matriz propuesta ($\hat{\Theta} = Q_{21} Q_{11}^{-1}$) es solución si y sólo si los autovalores de S^* son estrictamente menores que la unidad. Este hecho estaría garantizado si se cumpliera la condición de no negatividad tratada anteriormente; en concreto, el algoritmo de Vaughan (1970) falla sólo si la matriz S^* tiene un autovalor de módulo la unidad⁽²¹⁾. En Hannan y Poskitt (1988) puede verse que estas deficiencias de rango del espectro en el círculo unidad están

²¹ En Heij, Kloek y Lucas (1992) se demuestra que si S^* tiene un autovalor en $z = e^{i\theta}$ entonces el espectro es singular en dicho punto.

relacionadas con la existencia de componentes determinísticas en el proceso estocástico⁽²²⁾.

Todo el proceso de estimación consiste en la aproximación de la matriz de autocorrelaciones, \mathcal{H} , por un modelo que recoja la mayor parte del comportamiento de la serie. Así, se propone aproximar dicha matriz por otra resultante de eliminar en la descomposición en valores singulares aquéllos que no son significativamente distintos de cero, de forma que la norma de la diferencia entre las dos matrices sea mínima. Se sabe (Golub y Van Loan, 1983) que para matrices estructuradas, el error mínimo es el primer valor singular excluido. Sin embargo, esta aproximación no tiene por qué mantener la estructura de Hankel, por lo que su interpretación como matriz de covarianzas entre valores pasados y futuros no está del todo clara (Heij, Kloek y Lucas, 1992)⁽²³⁾.

Por estos problemas, es necesaria la existencia de una fase de diagnóstico del modelo estimado que asegure la correcta especificación de éste. En caso de deficiencias, las soluciones que se han propuesto son meramente *ad hoc*. En este sentido, Vaccaro y Vukina (1993) proponen una solución basada en aproximar la secuencia de autocorrelaciones $\{\Gamma_k\}$ de un modelo con deficiencias por otra $\{\bar{\Gamma}_k\}$ que produzca una estimación válida del proceso y tal que la matriz de inicial $\bar{\Gamma}_0 = \Gamma_0$ ⁽²⁴⁾. Otra alternativa propuesta en la literatura consiste en la no consideración de modelos con singularidades, pasando a estimar otros modelos con distinto número de estados o de retardos (Dorfman y Havenner, 1992).

²² La contrapartida en la formulación ARMA sería que singularidades en el círculo unidad son equivalentes a raíces unitarias en la parte de medias móviles, traduciéndose en la no invertibilidad del proceso.

²³ La existencia de una aproximación óptima en el sentido de tener un error mínimo está demostrada en Adamjan et al. (1971). Así mismo, un algoritmo que produzca modelos en espacio de estados eficientes está desarrollado en Glover (1984), aunque tampoco preserva la estructura de Hankel.

²⁴ Esta aproximación está basada en la detección de singularidades en la función de densidad espectral original mediante un rastreo en todas las posibles frecuencias. Una vez detectadas las frecuencias conflictivas, se propone una pequeña alteración que haga posible la densidad espectral en esa frecuencia. Por ello, la labor computacional puede ser inmensa, ya que supone una evaluación del espectro en todo el intervalo $[0, \pi)$.

Un desarrollo algo diferente que no precisa de la resolución de la ecuación de Ricatti puede encontrarse en trabajos como Östermark y Aoki (1992) u Östermark (1997). En ellos, partiendo de la ecuación (10) y de la estimación de la matriz H en (14), es posible obtener una estimación del vector de estado y de las innovaciones mediante las expresiones:

$$\begin{aligned}\hat{X}_t &= \Upsilon R^{-1} Y_{t-1}^- \\ \hat{e}_t &= Y_t - \hat{H} \hat{X}_t\end{aligned}\tag{14-1}$$

Sustituyendo la estimación del vector de estado en la ecuación de transición, multiplicando por la derecha por \hat{X}_t' y tomando esperanzas, se obtiene la expresión:

$$E[\hat{X}_{t+1} \hat{X}_t'] = \dots = FE[\hat{X}_t \hat{X}_t']\tag{14-2}$$

que, despejando, proporciona la estimación de la matriz F.

Por último, despejando en la ecuación de transición, postmultiplicando por \hat{e}_t' y tomando esperanzas se llega a la ecuación:

$$E[(\hat{X}_{t+1} - \hat{F} \hat{X}_t) \hat{e}_t'] = \dots = GE[\hat{e}_t \hat{e}_t']\tag{14-3}$$

de donde se obtiene la estimación de la matriz G que completa la fase de estimación de las matrices del modelo.

Otros autores, basándose en el algoritmo expuesto, han introducido modificaciones para facilitar la labor de estimación conservando las propiedades de anidamiento y estabilidad. Así, los trabajos de Mittnik (1989) y de Otter (1989) han desarrollado un planteamiento de estimación que, conservando la esencia del de Aoki, presenta ventajas computacionales. Básicamente, dada una muestra de N realizaciones de un proceso estocástico de orden p débilmente estacionario, se busca construir un modelo innovacional

en espacio de estados lineal e invariante de la forma dada por (1):

$$\begin{aligned} X_{t+1} &= FX_t + G\varepsilon_t \\ Y_t &= HX_t + \varepsilon_t \end{aligned} \quad t \in Z$$

que aproxime la secuencia muestral observada. Para encontrar estimaciones de las matrices F, H y G se considera que el proceso Y_t está generado por un proceso autorregresivo de la forma⁽²⁵⁾:

$$Y_t = \sum_{i=1}^m M_i Y_{t-i} + \varepsilon_t \quad (19)$$

donde ε_t es un ruido blanco de media nula y varianza $E(\varepsilon_s, \varepsilon_t') = \delta_{st} \Sigma$. Dada la muestra se pueden estimar las matrices de coeficientes por mínimos cuadrados a través de la ecuación

$$\hat{\Theta} = (X'X)^{-1} X'Y \quad \text{donde } \Theta = [M_1 \ M_2 \ \dots \ M_k]', \ Y = [Y_{k+1} \ Y_{k+2} \ \dots \ Y_N]'$$
 y

$$X' = \begin{bmatrix} Y_k & Y_{k+1} & \dots & Y_{N-1} \\ Y_{k-1} & Y_k & \dots & Y_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ Y_1 & Y_2 & \dots & Y_{N-k} \end{bmatrix}, \text{ con } k \geq p \text{ para permitir la correcta identificación del modelo.}$$

Para obtener la representación dada en (1), se plantea un sistema intermedio dado por:

$$\begin{aligned} X_{t+1} &= AX_t + GY_t \\ Y_t &= HX_t + \varepsilon_t \end{aligned} \quad (20)$$

donde la relación entre las matrices dinámicas F y A viene dada por $F = A + GH$. La ventaja de considerar valores retrasados de Y como inputs en la ecuación de transición permite interpretar los coeficientes autorregresivos M_i como parámetros de impulso respuesta de (20), dados por $M_i = HA^{i-1}G$ para $i = 1, 2, \dots$ con lo que se puede construir la matriz

²⁵ En la práctica, basta con que se pueda aproximar razonablemente bien por dicho proceso autorregresivo.

estimada:

$$\hat{\mathcal{H}}_k = \begin{bmatrix} \hat{M}_1 & \hat{M}_2 & \dots & \hat{M}_k \\ \hat{M}_2 & \hat{M}_3 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{M}_k & 0 & \dots & 0 \end{bmatrix} \quad (21)$$

que es de tipo Hankel por bloques. A partir de aquí, este algoritmo sigue un proceso análogo al anterior. Específicamente, esta matriz puede ser factorizada a través de:

$$\hat{\mathcal{H}}_k = \hat{O}_k \hat{R}_k \quad (22)$$

donde $O_k' = [C' \ (CF)' \ \dots \ (CF^{k-1})']$ y $R_k = [G \ FG \ \dots \ F^{k-1}G]$; utilizando la descomposición en valores singulares para estimar la factorización (22) se puede expresar la igualdad $\hat{\mathcal{H}}_k = \hat{O}_k \hat{R}_k = \hat{U} \hat{Q} \hat{V}'$ donde $\hat{U}' \hat{U} = \hat{V}' \hat{V} = Id$ y la matriz \hat{Q} contiene en su diagonal los valores singulares de la matriz $\hat{\mathcal{H}}_k$ ordenados de forma decreciente. A partir de esta igualdad se pueden obtener las estimaciones⁽²⁶⁾:

$$\hat{O}_k = \hat{U} \hat{Q}^{1/2} \quad ; \quad \hat{R}_k = \hat{Q}^{1/2} \hat{V}' \quad (23)$$

que proporciona un modelo internamente equilibrado. En el caso no estocástico, la matriz Q adoptaría la forma $Q = diag(q_1, q_2, \dots, q_n, 0, \dots, 0)$ siendo n la dimensión del sistema; sin embargo, ante la presencia de perturbaciones aleatorias ocurre que $q_i > 0$ en toda la diagonal de la matriz Q, dificultando la identificación de n. En este caso, las técnicas de reducción de dimensión ya comentadas o las expuestas en Holt y Antill (1977) u Otter (1985) permiten

²⁶ Esta idea, además de en Aoki (1987, 1990) puede encontrarse en los trabajos de Kung (1978), Moore (1981), Pernebo y Silverman (1982) o Zeiger y McEwen (1974).

separar el modelo en dos subsistemas, uno llamado “dominante” y otro “débil” de la forma:

$$\begin{aligned} \begin{bmatrix} X_{1,t+1} \\ X_{2,t+1} \end{bmatrix} &= \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} + \begin{bmatrix} G_1 \\ G_2 \end{bmatrix} \varepsilon_t \\ Y_t &= \begin{bmatrix} H_1 & H_2 \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} + \varepsilon_t \end{aligned} \quad (24)$$

donde el subsistema débil (A_{22} , G_2 , H_2) contribuye poco a la respuesta del sistema y se puede atribuir a la presencia de perturbaciones aleatorias. Por contra, el subsistema dominante (A_{11} , G_1 , H_1) asociado con los n valores singulares mayores puede considerarse como una aproximación válida al proceso generador de datos. La matriz de Hankel correspondiente al subsistema dominante, que denotaremos por $\tilde{\mathcal{H}}$, se obtiene eliminando los $p_k - n$ valores singulares inferiores de la matriz \hat{Q} así como las correspondientes columnas de las matrices \hat{U} y \hat{V} . Así, denotando por \tilde{Q} , \tilde{U} y \tilde{V} a las matrices modificadas, se puede expresar la igualdad⁽²⁷⁾ $\tilde{H} = \tilde{U} \tilde{Q} \tilde{V}'$ que proporcionaría las estimaciones:

$$\tilde{R} = \tilde{U} \tilde{Q}^{-1/2}, \quad \tilde{O} = \tilde{Q}^{-1/2} \tilde{V}' \quad (25)$$

De esta forma, utilizando las definiciones de las matrices \tilde{O} y \tilde{R} , una estimación de la matriz H de (20) viene dada por las primeras p filas de \tilde{O} y la estimación de la matriz G a través de las primeras p columnas de \tilde{R} . Ahora, una estimación de la matriz A viene dada a través de la ecuación $\hat{A} = \tilde{Q}^{-1/2} \tilde{U}' \mathcal{H}_k^{\dagger} \tilde{V} \tilde{Q}^{-1/2}$ donde $\mathcal{H}_k^{\dagger} = L^{-1} \mathcal{H}_k$. Con este resultado, la estimación de la matriz dinámica del sistema original (1) es $\hat{F} = \hat{A} + \hat{G} \hat{H}$.

²⁷ Esta matriz $\tilde{\mathcal{H}}$ no es una matriz tipo Hankel, por lo que su utilización como matriz de autocorrelaciones no está clara. Sin embargo, las discrepancias son mínimas, por lo que en la práctica se suele usar para estimar las matrices del sistema.

Como se puede observar, este método de estimación proporciona directamente una estimación de la matriz de filtrado G , sin necesidad de resolver la ecuación de Riccati que aparece en el método de estimación de Aoki. Este hecho, unido a la propiedad de anidamiento de las estimaciones facilita la labor de aproximación por modelos de orden inferior. Específicamente, si tenemos una estimación del modelo (20) dada por $(\hat{A}, \hat{G}, \hat{H})$ y se considera un sistema de dimensión $n_1 < n$, la estimación de tal subsistema se obtiene eliminando las correspondientes filas y columnas de las matrices $(\hat{A}, \hat{G}, \hat{H})$ sin más necesidad de cálculos adicionales. Para el sistema original (1) como se tiene:

$$\begin{bmatrix} \hat{F}_{11} & \hat{F}_{12} \\ \hat{F}_{21} & \hat{F}_{22} \end{bmatrix} = \begin{bmatrix} \hat{A}_{11} + \hat{G}_1 \hat{H}_1 & \hat{A}_{12} + \hat{G}_1 \hat{H}_2 \\ \hat{A}_{21} + \hat{G}_2 \hat{H}_1 & \hat{A}_{22} + \hat{G}_2 \hat{H}_2 \end{bmatrix}$$

se cumple que una aproximación de dimensión inferior al sistema se obtiene también eliminando las filas y columnas adecuadas de las estimaciones $(\hat{F}, \hat{G}, \hat{H})$, por lo que la propiedad de anidamiento se traspasa del sistema (20) al (1) original⁽²⁸⁾.

Un planteamiento semejante al expuesto pero partiendo de una representación de (1) como convolución del input y output de la forma $Y_t = \Psi(L)\varepsilon_t + \delta_t = \sum_{i=1}^{\infty} \Psi_i \varepsilon_{t-i} + \delta_t$ puede encontrarse en el trabajo de Otter y Van Dal (1989).

También basado en la representación (1) para procesos estocásticos, en los trabajos de Young *et al.* (1989) se desarrolla un algoritmo de estimación de matrices denominado descomposición espectral secuencial⁽²⁹⁾ y donde se explotan las propiedades espectrales del algoritmo de Kalman junto a una primera aproximación a estimaciones máximo verosímiles.

²⁸ A cambio de esta facilidad se pierde la propiedad de equilibrio para el sistema (1), ya que este método de estimación parte de una realización internamente equilibrada para la representación (19), propiedad que no se trasmite al modelo innovacional.

²⁹ También conocido como SSD, del inglés Sequential Spectral Decomposition.

Por último, dado el modelo de partida y desde un punto de vista teórico, quizás el enfoque más obvio para la estimación de las matrices sea el de máxima verosimilitud. Si las perturbaciones en el modelo (1) se consideran normales, la función de verosimilitud de las observaciones puede ser obtenida a partir de la descomposición en error de predicción del filtrado de Kalman (Schweppe, 1965). Así, en teoría, es posible maximizar esta verosimilitud respecto a los parámetros desconocidos utilizando algún algoritmo numérico de optimización. Sin embargo, como puede verse en el trabajo de Harvey y Peters (1984), los resultados de los trabajos realizados sobre este tema indican que se trata de un método singularmente complejo, incluso en los modelos más simples⁽³⁰⁾.

Hasta este momento, se está considerando un sistema donde todas las series son endógenas. Sin embargo, la existencia de información exógena puede ayudar a comprender la evolución de estas variables endógenas, por lo que resulta interesante su inclusión en el modelo en espacio de estados. Así, se puede reformular el sistema (1) introduciendo el proceso exógeno Z_t , de dimensión k , quedando la expresión:

$$\begin{aligned} X_{t+1} &= FX_t + DZ_t + G\varepsilon_t \\ Y_t &= HX_t + CZ_t + \varepsilon_t \end{aligned} \quad t \in Z \quad (26)$$

que constituye el modelo básico en espacio de estados para un proceso con variables exógenas. Como la determinación del número de estados necesarios para describir el proceso es equivalente al modelo sin información exógena, la única diferencia se produce en la estimación de las matrices del modelo (26). Aunque de naturaleza similar, el algoritmo de estimación ha de recoger, por un lado, la estimación de las nuevas matrices C y D , y, por otro, la interrelación entre los tres procesos. Por su similitud, se va a exponer de forma esquemática el planteamiento de la fase de estimación siguiendo el esquema de Östermark (1997).

³⁰ Esta circunstancia ha originado que en la inmensa mayoría de los trabajos teóricos sobre la modelización en espacio de estados el tema de la estimación se asuma resuelto por máxima verosimilitud, sin entrar en la complejidad numérica que esta solución conlleva, limitándose a la exposición del filtrado de Kalman. Como consecuencia, tampoco se ha prestado suficiente atención a los algoritmos de estimación que, desarrollados en teoría de control, suponen alternativas para la estimación por máxima verosimilitud, manteniéndolos casi desconocidos en el campo de la Economía Cuantitativa.

En primer lugar, para construir las matrices de tipo Hankel, es preciso ordenar los vectores de datos conforme a:

$$Y_t^+ = (Y_t', Y_{t+1}', \dots, Y_{t+N_f-1}')'$$

$$Y_{t-1}^- = (Y_{t-1}', Y_{t-2}', \dots, Y_{t-N_p}')'$$

$$Z_{t-1}^- = (Z_{t-1}', Z_{t-2}', \dots, Z_{t-N_p}^*)'$$

y construir las matrices:

$$\Omega_1 = E[Z_t Y_{t-1}^{-\prime}] \quad \mathcal{H} = E[Y_t^+, Y_{t-1}^{-\prime}]$$

$$\mathcal{H}_Z = E[Y_t^+, Z_{t-1}^{-\prime}] \quad R_- = E[Y_t^-, Y_t^{-\prime}]$$

$$Y = E[Y_{t-1}^-, Z_{t-1}^{-\prime}] \quad Z = E[Z_t, Z_{t-1}^{-\prime}]$$

$$Z^+ = Z'(ZZ')^{-1}$$

Con éstas ya se puede estimar la relación entre el output y los vectores de estado y exógeno, estimando las matrices C y H. Para ello, se construye la matriz auxiliar:

$$O\Upsilon = (\mathcal{H} - \mathcal{H}_Z Z^+ \Omega_1) (I - R_-^{-1} Y Z^+ \Omega_1)^{-1}$$

donde O y Υ son las matrices de observabilidad y controlabilidad respectivamente. Así, a través de su descomposición en valores singulares $O\Upsilon = U\Sigma V'$ se puede asignar:

$$\begin{aligned} O &= U\Sigma^{\frac{1}{2}} \\ \Upsilon &= \Sigma^{\frac{1}{2}} V' \end{aligned} \tag{27}$$

Construyendo ahora la matriz $Q_1 = (\mathcal{H}_Z - O\Upsilon R_-^{-1} Y) Z'(ZZ')^{-1}$ es posible estimar las matrices C y H a través de las expresiones:

$$\begin{aligned} \hat{C} &= [I_{(p \times p)}, 0 \dots 0] Q_1 \\ \hat{H} &= [I_{(n \times n)}, 0 \dots 0] O \end{aligned} \tag{28}$$

Para la obtención de las tres matrices de la ecuación de transición, recordando la expresión de los estados dada en (10), se puede construir la variable instrumental⁽³¹⁾

$\hat{X}_t = \Upsilon R^{-1} Y_{t-1}^-$ y, utilizando la ecuación de observación ya estimada, obtener una estimación del proceso residual $\hat{E}_t = Y_t - \hat{H}\hat{X}_t - \hat{C}Z_t$. Con esta expresión del vector instrumental de estado, la ecuación de transición adoptará la forma $\hat{X}_{t+1} = F\hat{X}_t + DZ_t + G\varepsilon_t$.

Esta ecuación puede ser utilizada para obtener un sistema de ecuaciones lineales postmultiplicando por \hat{X}_t' y por Z_t' y tomando esperanzas:

$$\begin{aligned} E[\hat{X}_{t+1} \hat{X}_t'] &= \dots = FE[\hat{X}_t \hat{X}_t'] + DE[Z_t \hat{X}_t'] \\ E[\hat{X}_{t+1} Z_t'] &= \dots = FE[\hat{X}_t Z_t'] + DE[Z_t Z_t'] \end{aligned} \quad (29)$$

de donde se puede extraer la estimación:

$$[\hat{F}, \hat{D}] = \left[E[\hat{X}_{t+1} \hat{X}_t'], E[\hat{X}_{t+1} Z_t'] \right] \begin{bmatrix} E[\hat{X}_t \hat{X}_t'] & E[\hat{X}_t Z_t'] \\ E[Z_t \hat{X}_t'] & E[Z_t Z_t'] \end{bmatrix}^{-1} \quad (30)$$

Por último, la matriz G puede ser hallada a través de la expresión⁽³²⁾:

$$\hat{G} = E\left[(\hat{X}_{t+1} - \hat{F}\hat{X}_t - \hat{D}Z_t) \hat{E}_t' \right] E\left[\hat{E}_t \hat{E}_t' \right]^{-1} \quad (31)$$

concluyendo la fase de estimación de las matrices del sistema.

4. Estimación del vector de estado: filtrado y alisado del sistema.

³¹ Ya que esta variable se utilizará sólo para la identificación de las matrices. Como se verá más adelante, se pueden hallar los estados del sistema de forma más eficiente.

³² Ver, por ejemplo, Östermark y Aoki (1992) u Östermark (1997).

Tras la determinación del orden y de las matrices del sistema (1) sólo queda la generación de la serie de valores del vector de estado, necesarios para la simulación y predicción con el modelo estimado.

En el caso del primer algoritmo propuesto (Aoki, 1987, 1990), la ecuación (10) dada por $\Upsilon R^{-1} Y_{t-1}^- = X_t$ proporciona estimaciones para el vector de estado dependiendo de las matrices $\Upsilon = (\Omega, F\Omega, \dots, F^{t-1}\Omega)$, $R = E[Y_{t-1}^- Y_{t-1}^{-'}]$ y $\Omega = E[X_{t+1}, Y_t']$, ya estimadas en la fase anterior. Como se puede observar, las dos primeras matrices tienen dimensiones que dependen de t , por lo que a medida que avanzamos temporalmente la necesidad de cálculos aumenta enormemente, pudiendo llegar a hacer inviable el uso de esta ecuación. Sin embargo dada la primera observación, Y_0 , es posible utilizar (10) para estimar el primer estado, X_1 , sin grandes complejidades computacionales y, a partir de éste, utilizar el modelo estimado para generar el resto de los valores del vector de estado. Específicamente, disponiendo del estado X_t y de la observación Y_t , se puede utilizar la ecuación de medida $Y_t = HX_t + \varepsilon_t$ para estimar la innovación en el instante t , valor que será utilizado posteriormente en la ecuación de transición $X_{t+1} = FX_t + G\varepsilon_t$ para generar el valor del estado en el instante siguiente. Se tiene así un algoritmo recursivo que permite la identificación de vector X_t junto a la serie de innovaciones ε_t a partir de la especificación del estado inicial. Sin embargo, el carácter de estimación que tiene este último se transmite a toda la serie de valores de estado y no existen herramientas que aseguren las buenas propiedades estadísticas de la serie así estimada.

Este mismo problema se presenta en los otros algoritmos de estimación, ya que se siguen utilizando las ecuaciones (1) para construir la serie de valores de estado a partir de uno inicial, diferenciándose exclusivamente en la determinación de este último.

Así pues, el problema consiste en encontrar un método para la construcción de la mejor estimación del estado X_t dadas las observaciones Y_s , $1 \leq s \leq t+n$. El algoritmo que

resuelve este problema es el conocido filtro de Kalman (Kalman, 1960; Kalman y Bucy, 1961).

El filtro de Kalman es una herramienta muy poderosa en el análisis de datos y suscitó desde su aparición un gran interés, enfocando la atención de los estadísticos y econométricos en la representación en espacio de estados de procesos estocásticos⁽³³⁾. Éste puede concebirse como un algoritmo de ortogonalización de Gram-Schmidt de la serie del output en el espacio generado por el vector de estado, generando así las innovaciones del modelo. Como método de estimación, dada la estructura en espacio de estados que se adopta para modelizar el proceso, en un instante del tiempo se tienen dos posibles estimaciones para el vector de estado: por una parte, la suministrada a través de la ecuación de transición del modelo, que indica cómo se genera el valor del estado para el instante inmediatamente siguiente en función de la información actual disponible (tanto en el estado como en la innovación); por otra, la suministrada a través de la ecuación de observación, donde se introduce la innovación contemporánea al vector que se estima. El estimador de Kalman, se obtiene como una combinación lineal de ambas estimaciones imponiendo que sea insesgado y de varianza mínima, condiciones que determinan las dos funciones de peso de la combinación lineal. Además, si se impone que el vector de estado inicial sea gaussiano, se puede probar que el estimador de Kalman coincide con el de mínimos cuadrados lineales. Entre las ventajas que presenta el filtrado de Kalman pueden destacarse dos fundamentales: por un lado, el algoritmo no se restringe a procesos estacionarios sino que permite la no estacionariedad siempre que el proceso sea estable; por otro lado, se presenta un método recursivo que permite el cálculo en tiempo real de las estimaciones. Por contra, este algoritmo está restringido a procesos con función de transferencia racional⁽³⁴⁾.

Dada la gran bibliografía sobre el tema, en este trabajo no se analizará

³³ De hecho, la importancia de este algoritmo ha ocasionado que dicha representación sea estudiada exclusivamente para la aplicación de este algoritmo.

³⁴ Esta restricción supone sólo una pérdida de eficiencia de las estimaciones ya que se están considerando modelos en espacio de estados de dimensión finita aunque sólo sean como aproximación al proceso generador de datos, por lo que la función de transferencia será siempre racional.

exhaustivamente el algoritmo, limitándose a una exposición somera que, sin embargo, ponga de manifiesto las implicaciones para la estimación de modelos en espacio de estados. Para un análisis detallado pueden consultarse las obras de Kalman (1960) y Kalman y Bucy (1961) junto a las de Anderson y Moore (1979), Hannan y Deistler (1988), Meinhold y Singpurwalla (1984) o Reinsel (1993).

El algoritmo parte del modelo (1) dado por:

$$\begin{aligned} X(t+1) &= FX(t) + G\varepsilon(t) \\ Y(t) &= HX(t) + \varepsilon(t) \end{aligned} \quad t \in Z$$

donde $\Psi = E[\varepsilon_t, \varepsilon_t']$ ⁽³⁵⁾ y la referencia temporal se introduce entre paréntesis por comodidad. Se denotará por $\hat{X}(t|s)$ la estimación del vector de estado en el instante t dada la información recogida hasta el instante s. En estas condiciones, el filtrado de Kalman se puede analizar en tres etapas:

a) Condiciones iniciales:

Vienen dadas por el valor del estado en un instante inicial $\hat{X}(1|0)$ y su correspondiente varianza $P(1) = E[X(1), X(1)']$. Estos valores tienen que ser determinados con anterioridad a la utilización del algoritmo, lo que ha originado una amplia literatura sobre el problema de inicialización del filtrado, en particular, sobre la elección del estado inicial, ya que la matriz P(1) sólo se interpreta como una medida de incertidumbre sobre éste ⁽³⁶⁾.

Para un instante t cualquiera, se supone conocida una estimación inicial del vector

³⁵ Esta matriz está estimada, bien a partir de la resolución de la ecuación de Ricatti (15), bien a partir de la secuencia de innovaciones auxiliares dada en (14-3).

³⁶ Para no romper el hilo argumental de la exposición, el análisis del problema de inicialización del filtro de Kalman se abordará en el epígrafe siguiente.

de estado $\hat{X}(t|t-1)$ junto a una primera estimación de su varianza $P(t) = E[X(t), X(t)']$. Con estas condiciones iniciales y la ecuación de observación del modelo se puede obtener la matriz de covarianza del output del sistema en el instante actual a través de:

$$\Sigma(t) = HP(t)H' + \Psi \quad (32)$$

con lo que se resume toda la información disponible antes de la obtención del output $Y(t)$.

b) Observación y actualización de la información:

En el instante en que se tiene la observación de la serie $Y(t)$, utilizando la ecuación de observación del sistema se puede obtener la innovación a través de la diferencia entre el output y el valor previsto para éste como:

$$E(t) = Y(t) - H\hat{X}(t|t-1) \quad (33)$$

Esta innovación es utilizada junto a la estimación original del vector de estado para actualizar ésta a través de la ecuación:

$$\hat{X}(t|t) = \hat{X}(t|t-1) + P(t)H' \Sigma(t)^{-1} E(t) \quad (34)$$

con lo que se obtiene el estimador de Kalman para el estado en el instante t .

c) Predicción del vector de estado:

Una vez analizada la información disponible en el instante actual, se predice el valor del vector de estado para el instante siguiente a través de la ecuación:

$$\hat{X}(t+1|t) = F\hat{X}(t|t-1) + K(t)E(t) \quad (35)$$

donde la matriz $K(t)$ se conoce como matriz de ganancia de Kalman y responde a la ecuación:

$$K(t) = [FP(t)H' + G\Psi]\Sigma(t)^{-1} \quad (36)$$

Esta matriz permite la consideración de la innovación $E(t)$ para la predicción del vector de estado. Hay que destacar que en la determinación automática del vector de estado expuesta al principio del epígrafe se utilizaba la ecuación de transición del modelo (1), que sólo se diferencia de la actual (35) en la matriz G . La rigidez introducida al ser esta matriz constante en el modelo hace que se gane eficiencia con el filtrado de Kalman, ya que la variabilidad de la matriz de ganancia permite un mejor ajuste a las observaciones.

Una vez obtenido el estimador de Kalman se puede actualizar la información sobre su varianza a través de la ecuación:

$$P(t+1) = FP(t)F' + G\Psi G' - K(t)\Sigma(t)K(t)' \quad (37)$$

con lo que se está en condiciones de reiniciar el ciclo utilizando los resultados de las ecuaciones (35) y (37) como condiciones iniciales para el instante $t+1$.

El conjunto de ecuaciones (32) - (37) forman el algoritmo de estimación de Kalman y su obtención detallada puede consultarse en cualquiera de las referencias bibliográficas citadas anteriormente.

Este algoritmo permite estimar el valor del vector de estado $X(t)$ dada la información disponible hasta ese momento en tiempo real y de forma recursiva. Si por cualquier circunstancia no se dispone de la observación $Y(t)$ se puede utilizar igualmente el algoritmo pero considerando que la innovación $E(t)$ es cero en la ecuación (33). Esto origina que la actualización del estimador de Kalman dada en (34) adopte la nueva forma:

$$\hat{X}(t|t) = \hat{X}(t|t-1) \quad (34 \text{ bis})$$

y que la predicción al instante siguiente venga dada por:

$$\hat{X}(t+1|t) = F\hat{X}(t|t-1) \quad (35 \text{ bis})$$

permaneciendo idénticas el resto de las ecuaciones del filtrado de Kalman. Como era de esperar, a falta de la información del output, el estimador de Kalman es el estimador suministrado por la ecuación de transición del modelo en espacio de estados⁽³⁷⁾.

Durante el desarrollo anterior, se ha obtenido la estimación $\hat{X}(t|t)$ a partir de la información disponible hasta ese momento, sin utilizar la información de $\mathbf{Y}(s)$ para valores $s > t$. En el caso de utilizar toda la información disponible para estimar los vectores de estado el algoritmo de Kalman recibe el nombre de alisado. Así, considerando que se tiene información para $0 \leq t \leq s$ se obtiene la estimación $\hat{X}(t|s)$ correspondiente al estado en el instante t a través de la ecuación (Hannan y Deistler, 1988):

$$\hat{X}(t|s) = \hat{X}(t|t) + A(t) \{ \hat{X}(t+1|s) - \hat{X}(t+1|t) \} \quad (38)$$

donde la matriz $A(t)$ pondera la ganancia de información y responde a la ecuación

$$A(t) = \{ P(t)F' - P(t)H'K(t)' \} P(t+1)' \quad (39)$$

para $t = s-1, s-2, \dots$ Como se puede observar, las estimaciones se van obteniendo en sentido contrario a la evolución temporal, empezando por la última observación y acabando por la primera. De esta forma, son utilizadas todas las observaciones disponibles para estimar los vectores de estado en todos los instantes muestrales.

Otro aspecto fundamental del algoritmo de filtrado es su estabilidad numérica, sobre todo en la resolución de la ecuación de Riccati (37) que actualiza la varianza del estimador

³⁷ Esta simple modificación del algoritmo permite el tratamiento de series con algunas observaciones omitidas. Si se supone normalidad en la innovación, como el estimador de Kalman coincide con el de mínimos cuadrados ordinarios, el algoritmo actúa como si la observación $\mathbf{Y}(t)$ fuera sustituida por su estimación máximo verosímil, permitiendo una primera estimación de valores perdidos.

de Kalman⁽³⁸⁾. Por este motivo se han sugerido otras alternativas más robustas computacionalmente para resolver esta ecuación. Entre ellas cabe destacar varias áreas de trabajo: por un lado, la utilización del lema de inversión de matrices para actualizar $P(t)^{-1}$ en lugar de la varianza, dando lugar al llamado filtro de información. Otra alternativa consiste en la utilización de algoritmos que actualizan la raíz cuadrada de la matriz $P(t|t)$, que se han demostrado numéricamente más estables que los basados en la resolución de la ecuación de Riccati. Por último, se han propuesto algoritmos basados en la actualización del incremento de varianza $\Delta(t) = P(t+1) - P(t)$, denominados algoritmos de tipo Chandrasekhar, que presentan menos necesidad de cálculo cuando la dimensión del modelo es grande y ciertas ventajas numéricas⁽³⁹⁾.

Otro fenómeno que afecta seriamente a la aplicabilidad del filtrado es la conocida como divergencia del filtro. En Sage y Melsa (1971) u Otter (1985) se puede ver que, cuando el modelo está mal especificado, aunque la covarianza del error sea pequeña puede ocurrir que el error “actual” de estimación no esté acotado. Este fenómeno, producido por la rápida tendencia de la matriz de ganancia a cero, se manifiesta en un distanciamiento (a veces cíclico) entre las observaciones y las predicciones (en Sage y Melsa puede encontrarse un interesante práctico de este problema⁽⁴⁰⁾)

5. Determinación de las condiciones iniciales del filtrado y predicción.

Un aspecto importante del algoritmo de filtrado de Kalman es la determinación del estado inicial y su varianza junto con la influencia que ésta pueda tener sobre aspectos

³⁸ La posible inestabilidad numérica en esta estimación no está relacionada con la divergencia del filtro de Kalman, fenómeno que se debe a una mala especificación del modelo.

³⁹ Para un análisis de todos estos tipos de algoritmos puede consultarse la obra de Anderson y Moore (1979).

⁴⁰ El ejemplo se centra en la especificación de un paseo aleatorio sin deriva cuando el verdadero proceso generador de datos posee deriva. Ante un ejemplo de especificación concreto puede ocurrir que la deriva no sea lo suficientemente importante para detectarla con los contrastes habituales, pero de bastante magnitud como para producir la divergencia del filtro.

claves como la estabilidad o convergencia de las iteraciones. Sin embargo, cuando se trabaja con procesos estacionarios, esta cuestión queda en segundo término en la mayoría de las aplicaciones empíricas ya que suele ser habitual considerar que $\hat{X}(1|0) = 0$ y obtener su varianza particularizando la ecuación (37) a la expresión:

$$P(1) = FP(1)F' + G\Psi G'$$

Esta práctica está avalada por las propiedades de convergencia y estabilidad del filtrado de Kalman. Específicamente, si los procesos involucrados son estacionarios se puede probar que cuando t converge a infinito $\Sigma(t) \rightarrow \Sigma$, $K(t) \rightarrow K$, $P(t) \rightarrow 0$ y $E(t)$ converge en media cuadrática a $\varepsilon(t)$ (ver, por ejemplo, Hannan y Deistler, 1988). En particular, sustituyendo la expresión de $E(t)$ dada por la ecuación (33) en la de predicción del estimador de Kalman (35), se obtiene que:

$$\hat{X}(t+1|t) = (F - K(t)H)\hat{X}(t|t-1) + K(t)Y(t) \quad (40)$$

donde por las propiedades de convergencia comentadas se tiene que:

$$(F - K(t)H) \rightarrow (F - KH) \quad (41)$$

por lo que todos los autovalores de $(F - K(t)H)$ pasarán a estar dentro del círculo unidad si y sólo si es cierto para los autovalores de $(F - KH)$. Si ocurre esto se dice que la ecuación (40) es uniformemente asintóticamente estable (ver Jazwinski, 1970, para una definición formal) y, entre otras consecuencias, se deduce que cualquier error en la iniciación del algoritmo así como en las observaciones $Y(t)$ presenta un efecto que decrece geométricamente a cero, por lo que la condición inicial no es excesivamente relevante en el filtrado. De hecho, asumiendo una representación innovacional como la dada por (1) de dimensión mínima⁽⁴¹⁾ se puede demostrar (ver Hannan y Deistler, 1988) que $(F - KH)$ tiene todos sus autovalores inferiores en módulo a la unidad si y sólo si $\det\{k(z)\} \neq 0$ para $|z| \leq 1$ donde $k(z) = Id + \sum_{j=1}^{\infty} HF^{j-1}Gz^j$ es la función de transferencia

⁴¹ Es decir, con dimensión igual al grado de McMillan o, equivalentemente, controlable y observable.

del modelo (1) ⁽⁴²⁾.

Sin embargo, cuando la amplitud muestral no es grande, puede ocurrir que la consideración de nulidad para el valor inicial del estado provoque un mal ajuste del algoritmo de Kalman mientras que una buena especificación mejore sensiblemente las estimaciones sucesivas. Es en tales casos donde tiene interés el estudio de fórmulas que determinen el “mejor” valor inicial para el estado ⁽⁴³⁾. Igualmente, puede ser fructífero no especificar un único valor para iniciar el filtrado; en muchas aplicaciones, entre las que cabe destacar los modelos económicos no estacionarios o los modelos estructurales, es más realista asumir que las condiciones iniciales son parcialmente difusas, por lo que se introduce una perturbación aleatoria en el estado inicial. Se origina así el llamado filtrado difuso de Kalman (ver, por ejemplo, Kohn y Ansley (1986, 1987, 1989) o Kitagawa y Gersch (1984) para un análisis detallado de este algoritmo). En la misma línea, se puede considerar toda una distribución para el estado inicial (usualmente normal) que, unida a la normalidad asumida para la innovación, puede convertir el filtrado de Kalman en un algoritmo de actualización de los momentos de la distribución, planteamiento adoptado, por ejemplo, en Fahrmeir y Tutz (1991) o desde un punto de vista bayesiano en Harrison y Stevens (1976), West y Harrison (1989) o Vargas y Gámez (1995).

Otra alternativa, adoptada en este trabajo, utiliza la fase de estimación de matrices descrita por los algoritmos anteriores para extraer ecuaciones que relacionan el estado inicial con las primeras observaciones del sistema. Así, en el algoritmo de Aoki, la ecuación (10) particularizada a $t = 1$ proporcionaría la relación $\Upsilon R^{-1} Y(0)^- = X(1|0)$, que se puede utilizar como inicialización del filtrado de Kalman. Si se utiliza el algoritmo de Mittnik, en el vector de estado inicial se recoge el efecto de las observaciones premuestrales $\mathbf{Y}(0)$,

⁴² Esta condición se puede relajar ya que es posible que $\left\| \prod_{j=1}^t (F - K(j)H) \right\| \rightarrow 0$ aunque no lo haga geoméricamente incluso si $\det\{k(z)\} = 0$ para algún z de módulo unidad.

⁴³ Obviamente, el criterio para elegir el punto inicial del vector de estado es el de minimizar la norma de la matriz de varianza $P(t)$ para los sucesivos valores de t .

$\mathbf{Y}(-1), \dots, \mathbf{Y}(1-m)$ ⁽⁴⁴⁾. Si ahora se definen los vectores s_k para $k=1, 2, \dots, m$ que recojan la influencia del estado inicial en $\mathbf{Y}(k)$, se puede expresar la relación:

$$s_k = \sum_{j=k}^m M_k Y(k-j) = C F^{k-1} X(1|0) \quad (42)$$

Con estas variables se puede alterar el algoritmo expuesto construyendo la matriz de Hankel modificada (Mittnik, 1989):

$$\mathcal{H}^e = \begin{bmatrix} s_1 & M_1 & M_2 & \dots & M_m \\ s_2 & M_2 & M_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_m & M_m & 0 & \dots & 0 \end{bmatrix}$$

que puede ser factorizada a través de la expresión $\mathcal{H}^e = O_m R_m^e$ donde O_m es la matriz de observabilidad usual y $R_m^e = [X(1|0) \quad G \quad FG \dots F^{m-1}G]$. La estimación de esta nueva matriz a través de la descomposición en valores singulares de \mathcal{H}^e proporciona, en su primera columna una estimación del vector inicial de estado.

Por último, una vez estimadas las matrices del modelo original (1) junto a la serie de vectores de estado y la de innovaciones, se puede abordar la predicción óptima ⁽⁴⁵⁾ de valores futuros para la serie. En este caso, al considerar como nulas las innovaciones futuras, la predicción que se obtiene utilizando el filtrado de Kalman y la deducida del modelo original son iguales; por ello, la predicción del vector de estado a un período dada la información disponible hasta T responde a la relación:

⁴⁴ Recuérdese que en este algoritmo se está aproximando el modelo en espacio de estados por uno autorregresivo de orden m.

⁴⁵ La optimalidad de la predicción depende en gran medida de los supuestos iniciales que se hagan. Particularmente si asumimos la normalidad en el vector de estado inicial y en las innovaciones, se obtienen las predicciones máximo verosímiles.

$$\hat{X}(T+1|T) = F \hat{X}(T|T) \quad (43)$$

con error cuadrático medio dado por la ecuación:

$$P(T+1) = F P(T) F' + G \Psi G' \quad (44)$$

Así mismo, el valor del output se puede predecir utilizando la ecuación de observación del modelo original, resultando:

$$\hat{Y}(T+1|T) = H \hat{X}(T+1|T) \quad (45)$$

y error cuadrático medio asociado dado por la expresión:

$$E[(Y(T+1) - \hat{Y}(T+1|T))(Y(T+1) - \hat{Y}(T+1|T))'] = H' P(t+1) H + \Psi \quad (46)$$

ecuaciones que pueden ser utilizadas secuencialmente para la determinación de predicciones a un horizonte de n períodos.

6.- Conclusiones.

Este trabajo aborda la teoría subyacente en la representación en espacio de estados de procesos estocásticos, adaptando los avances que se han conseguido en otros campos científicos (teoría de realización estocástica, teoría de control o ingeniería de sistemas). En esta metodología, las propiedades de controlabilidad y observabilidad garantizan, bajo supuestos poco restrictivos, la minimalidad de la representación (Vargas, 1999), mitigando el problema de la gran cantidad de parámetros necesarios en la metodología VARMA. Por otro lado, los algoritmos detallados están basados en la descomposición en valores singulares, método mucho más robusto computacionalmente y que solventa, en gran medida, los problemas de la optimización de la función de verosimilitud y de la estabilidad numérica de las estimaciones así obtenidas.

La estructura particular de los modelos en espacio de estados, descompone la serie observada en dos sumandos: una combinación lineal de las componentes del vector de

estado, que resume la evolución dinámica del sistema; y un proceso de innovación. Esta particularidad hace que, conocido el valor del vector de estado en un instante dado, el pasado de la serie sea irrelevante; es decir, el estado es un estadístico suficiente (y minimal en la representación adoptada) para la evolución del proceso. Esta característica, exclusiva de la modelización en espacio de estados, la asemeja a un análisis factorial dinámico, donde las variables de estado desempeñan el papel de componentes principales. Con esto, se evita la especificación de un número, constante, de retardos para captar las peculiaridades dinámicas de la serie y se mantiene, en cada instante, toda la información relevante. Además, esta estructura permite eliminar de la serie observada la componente innovacional, pudiendo interpretar el resultado de esta diferencia como la señal implícita de la serie. Este proceso supone, realmente, una labor de alisado, ya que elimina la componente no predecible, por lo que el algoritmo de estimación puede ser interpretado como un filtrado de series. Aunque no es un planteamiento inédito en la literatura de series temporales, sí es novedoso en cuanto que no asume ninguna estructura en las matrices del sistema. En la casi totalidad de las aplicaciones empíricas del filtrado en espacio de estados se supone conocida, en primer lugar, la dimensión del vector de estado así como las matrices del modelo o, a lo sumo, se introduce algún parámetro en éstas para su estimación por máxima verosimilitud. Así, la eficiencia de las estimaciones del método están condicionadas a la estructura que se asuma, normalmente determinada por consideraciones extra-muestrales. Por contra, el planteamiento defendido en este trabajo determina el número de estados y la estimación de las matrices en función de la información suministrada por la muestra. Se consigue con ello una mayor versatilidad en la especificación y, por tanto, una mayor adecuación a los datos disponibles sin renunciar a las propiedades del filtrado de Kalman.

Por último, el comportamiento dinámico de un sistema múltiple viene caracterizado por las funciones de impulso-respuesta. En la metodología estudiada, su cálculo se ve facilitado por la propiedad markoviana de la ecuación de transición, que permite expresarlas como el producto de tres matrices. Se consigue con ello una forma simple e intuitiva de analizar las interacciones dinámicas entre las componentes del sistema, determinando fácilmente qué componentes presentan mayor poder de generar respuestas

dinámicas y cuáles son más sensibles a alteraciones del sistema.

Por todo ello, la modelización de series temporales en espacio de estados puede constituir una alternativa válida a la clásica VARMA que, bajo un enfoque analítico algo distinto, proporciona herramientas estadísticas útiles para el análisis de series.

7.- Bibliografía:

ADAMJAN, V.M, AROV, D.Z. & KREIN, M.G. (1971): "Analytic properties of Schmidt-pairs for a Hankel operator and the generalized Schur-Takagi problem". *Mathematics USSR Sbornik* 15, 31-73.

AKAIKE, H. (1973): "Information theory and an extension of the maximum likelihood principle". En *2nd International Symposium on Information Theory* (B.N. Petrov & F. Csaki, eds.). Budapest: Akademia Kiado.

AKAIKE, H. (1974a): "Markovian representation of stochastic processes and its applications to the analysis of autoregressive moving average processes". *Ann. Inst. statist. Math.* 20, 363-388.

AKAIKE, H. (1974b): "Stochastic theory of minimal realization". *IEEE Trans. Autom. Control* **AC** 19, 667-674

AKAIKE, H. (1974c): "A new look at the statistical model identification". *IEEE Trans. Autom. Control* **AC** 19, 716-723.

AKAIKE, H. (1975): *Markovian representation of stochastic processes by canonical variables*. *SIAM J. Control Optim.* 13, 162-173.

AKAIKE, H. (1976): "Canonical Correlation analysis of time series and the use of an information criterion". En *System Identification and Case Studies* (Eds. R. Mehra and D. Lainiotis). Academic Press, New York.

ANDERSON, B.D.O. & MOORE, J.B. (1975): *Optimal Filtering*. Prentice-Hall, New Jersey.

AOKI, M. (1987): *State Space Modeling of Time Series*. New York, Springer-Verlag.

AOKI, M. (1990): *State Space Modeling of Time Series. Second, Revised and Enlarged Edition.* New York, Springer-Verlag.

AOKI, M. & HAVENNER, A.M. (1992): "State Space modeling of multiple time series". *Econometric Reviews* 10, 1-99.

BARTLETT, M.S. (1939): "A Note on Tests of Significance in Multivariate Analysis". *Proc. Cambridge Phil. Soc.* 35, 180-185.

BROCKWELL, P.J. & DAVIS, R.A. (1991): *Time Series: Theory and Methods*, 2^a ed. New York: Springer Verlag

DESAI, U.B. & PAL, D. (1983): "A Realization Approach to Stochastic Model Reduction and Balanced Stochastic Realization". Preprint IEEE Conf. On Decision and Control, 1105-1112.

DORFMAN, J.H. & HAVENNER, A.M. (1992): "A Bayesian approach to state space multivariate time series modeling". *Journal of Econometrics* 52, 315-346.

ECKART, C. y YOUNG, G. (1936): "The approximation of one matrix by another of lower rank". *Psychometrika*, 1, 211-218.

FAHRMEIR, L. & TUTZ, G. (1991): *Multivariate Statistical Modelling Based on Generalized Linear Models.* New York: Springer Verlag.

FAURRE, P., CLERGET, M. & GERMAIN, F. (1979): *Opérateurs rationnels positifs.* Dunod, Paris.

GEL'FAND, I.M. & YAGLOM, A.M. (1959): "Calculation of the Amount of Information about a Random Function Contained in Another Such Function". *Am. Math. Soc. Transl. Series 2*, 12, 191-198.

GLOVER, K. (1984): "All optimal Hankel norm approximations of linear multivariable systems and their L^∞ error bounds". *International Journal of Control* 39, 1115-1193.

GLYNN, W.J. & MUIRHEAD, R.J. (1978): "Inference in canonical correlation analysis". *Journal of Multivariate Analysis*, 8, 468-478.

GOLUB, G.H. & VAN LOAN, C.F. (1983): *Matrix Computations.* Johns Hopkins University Press, Baltimore.

- HANNAN, E.J. & DEISTLER, M.** (1988): *Statistical Theory of Linear Systems*. John Wiley, New York.
- HANNAN, E.J. & POSKITT, D.S.** (1988): "Unit canonical correlations between future and past". *The Annals of Statistics* 16, 784-790.
- HANNAN, E.J. & QUINN, B.G.** (1979): "The determination of the order of an autoregression" *Journal of the Royal Statistical Society, B* 41, 190-195.
- HARRISON, P.J. & STEVENS, C.F.** (1976): "Bayesian Forecasting". *Journal of The Royal Statistical Society, serie B* 38, 205-228.
- HARVEY, A.C.** (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.
- HARVEY, A.C. & PETERS, S.** (1984): "Estimation procedures for structural time-series models". London School of Economics, Discussion Paper n° A28.
- HEIJ, C., KLOEC, T. & LUCAS, A.** (1992): "Positivity conditions for stochastic state space modelling of time series". *Econometric Reviews* 11(3), 379-396.
- HOLT, J.N. & ANTILL, R.J.** (1977): "Determining the number of terms in a Prony algorithm exponential fit". *Math. Biosci.* 36, 319-332.
- HURVICH, C.M. & TSAI, C.L.** (1989): "Regression and time series model selection in small samples". *Biometrika*, 76, 297-307.
- HURVICH, C.M. & TSAI, C.L.** (1993): "A corrected Akaike information criterion for vector autoregressive model selection". *Journal of Time Series Analysis*, 14, 271-279.
- IBRAGIMOV, I.A. & ROZANOV, Y.A.** (1978): *Gaussian Random Process*. New York: Springer Verlag.
- JAZWINSKI, A.H.** (1970): *Stochastic Processes and Filtering Theory*. New York: Academic Press.
- JEWELL, N.P. & BLOOMFIELD, P.** (1983): "Canonical correlations of past and future for time series: definitions and theory". *Annals of Statistics*, 11, 837-847.
- KALMAN, R.E.** (1960): "A new approach to linear filtering and prediction problems". *Journal of Basic Engineering*, 82, 35-45.
- KALMAN, R.E. & BUCY, R.S.** (1961): "New results in linear filtering and prediction theory". *Journal of Basic Engineering*, 83, 95-108.

- KITAGAWA, G. & GERSCH, W.** (1984): "A smoothness priors-state space modeling of time series with trend and seasonality". *Journal of American Statistical Association*, 79, 378-389.
- KOHN, R. & ANSLEY, C.F.** (1986): "Estimation, Prediction and interpolation for ARIMA models with missing data". *Journal of American Statistical Association* 81, 751-761.
- KOHN, R. & ANSLEY, C.F.** (1987): "Signal extraction for finite nonstationary time series". *Biometrika* 74, 411-421.
- KOHN, R. & ANSLEY, C.F.** (1989): "Filtering and Smoothing Algorithms for state space models" *Computers Math. Applic.* 18, n° 6/7, 515-528.
- KUNG, S.Y.** (1978): "A new identification and model reduction algorithm via singular value decomposition". *Proc. 12th Ann. Asilomar Conf. Circuits, Systems and Computer*, 705-714.
- LAUB, A.J.** (1983): "Numerical Aspects of Solving Algebraic Riccati Equations". *Proc. IEEE Conf. Decision and Control*, 184-186.
- LAWLEY, D.N.** (1959): "Test of Significance in Canonical Analysis". *Biometrika*, 41, 59-66.
- LI, L. & XIE, Z.** (1996): "Model selection and order determination for time series by information between the past and the future". *Journal of time series analysis*, 17, 65-84.
- LINHART, H. & ZUCCHINI, W.** (1986): *Model Selection*. New York: Wiley.
- MEINHOLD, R.J. & SINGPURWALLA, N.D.** (1983): "Understanding the Kalman Filter". *American Statistician* 37, 123-127.
- MITTNIK, S.** (1984): "Time Series analysis via approximate realization theory". 1984 Winter Meeting Econometric Society, Dallas, Texas.
- MITTNIK, S.** (1989): "Multivariate Time Series analysis with state space models". *Computers Math. Applic.* 17, n° 8/9, 1189-1201.
- MOORE, C.B.** (1981): "Principal component analysis in linear systems: controllability, observability and model reduction". *IEEE Automatic Control* 26, 17-32.
- MUIRHEAD, R.J. & WATERNAUX, C.M.** (1980): "Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal

populations”. *Biometrika*, 67, 31-43.

ÖSTERMARK, R. (1997): “Modeling Cointegrated Processes by a Vector-Valued State Space Algorithm. Evidence on the Impact of Japanese Stock Prices on the Finnish Derivatives Market”. En M. Aoki & A.M. Havenner (Eds.) *Applications of Computer Aided Time Series Modeling*. Lecture Notes in Statistics, 119. New-York, Springer-Verlag.

ÖSTERMARK, R. & AOKI, M. (1992): “Time Series Evidence of Impacts of the US Economy on the Scandinavian Economy”. IFAC Workshop on Economic Time Series Analysis and System Identification, Vienna.

OTTER, P.W. (1985): *Dynamic Feature Space Modelling, Filtering and Self-Tuning Control of Stochastic Systems*. Springer, Berlin.

OTTER, P.W. & VANDAL, R. (1989): “State-space approximation of multi-input multi-output systems with stochastic exogenous inputs”. *Computers Math. Applic.* 18, n° 6/7, 529-538.

PERNEBO, L. & SILVERMAN, L.M. (1982): "Model reduction via balanced state space representations". *IEEE Automatic Control* 27, 382-387.

POPOV, V.M. (1969): “Some Properties of Control Systems with Matrix Transfer Functions” *Lecture Notes in Mathematics*, 144, 169-180. Berlin, Springer-Verlag.

RAO, C.R. (1965): *Linear Statistical Inference and Its Applications*. New York: John Wiley.

RAO, C.R. (1979): “Separation Theorems for Singular Values of Matrices and Their Applications in Multivariate Analysis”. *Journal of Multivariate Analysis*, 9, 362-377.

REINSEL, G.C. (1993): *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York.

ROSENBROCK, H.H. (1970): *State Space and Multivariable Theory*. New York, Wiley.

SAGE, A.P. & MELSA, J.L. (1971): *Estimation Theory with Applications to Communications and Control*. McGraw-Hill.

SCHWARZ, G. (1978): “Estimating the Dimension of a Model”. *Annals of Statistics*, 6, 461-464

SCHWEPPE, F. (1965): “Evaluation of likelihood function for Gaussian signals”. *IEEE*

Trans. Inform. Theory 11, 61-70.

VACCARO, R.J. & VUKINA, T. (1993): "A solution to the positivity problem in the state-space approach to modeling vector-valued time series". *Journal of Economic Dynamics and Control* 17, 401-421.

VARGAS, M. (1999): *Modelización de series temporales múltiples en espacio de estados. Análisis de procesos no estacionarios y cointegración*. Tesis doctoral. Facultad de Ciencias Sociales de Cuenca. U.C.L.M.

VARGAS, M. Y GÁMEZ, M. (1995): "Modelización dinámico-bayesiana de series temporales univariantes". IX Reunión Asepelt-España, Santiago de Compostela.

VAUGHAN, D.R. (1970): "A nonrecursive algebraic solution for the discrete Riccati equation". *IEEE Automatic Control* 15, 597-599.

WEST, M. & HARRISON, J. (1989): *Bayesian Forecasting and dynamic models*. Springer Verlag, New York.

YOUNG, P., NG, C. & ARMITAGE, P. (1989): "A Systems approach to Recursive Economic Forecasting and Seasonal Adjustment". *Computers Math. Applic.* 18, nº 6/7, 481-501.

ZEIGER, H.P. & McEWEN, A.J. (1974): "Approximate linear realization of given dimension via Ho's algorithm". *IEEE Trans. Autom. Control*, AC19, 153.