

Un algoritmo rápido para evaluar la verosimilitud exacta de modelos VARMAX periódicos

por

JOSÉ CASALS, SONIA SOTOCA y MIGUEL JEREZ (1)
Departamento de Fundamentos del Análisis Económico II
Universidad Complutense de Madrid
E-Mail: eccua06@sis.ucm.es

RESUMEN

En este trabajo se deriva un algoritmo rápido y estable para evaluar la función de verosimilitud exacta de procesos VARMAX periódicos. Su estabilidad numérica y eficiencia computacional se consiguen combinando a) una formulación de dimensión mínima en espacio de los estados, en forma *steady-state innovations* con b) un procedimiento computacional que aprovecha las propiedades de esta representación. El algoritmo es aplicable a modelos estacionarios y no estacionarios y facilita el cálculo de los segundos momentos exactos de las estimaciones. Algunas pruebas de simulación y un caso real muestran su buen funcionamiento.

(1) Sonia Sotoca agradece el apoyo financiero de la CICYT, proyecto PB95-0912/95 (nº 5290806) y de Fundación Caja de Madrid. Todos los autores agradecen las sugerencias recibidas en las sesiones sobre series temporales del XXIV Congreso Nacional de Estadística e Investigación Operativa, así como los comentarios de un evaluador anónimo.

Palabras clave: modelos VARMAX periódicos, máxima verosimilitud exacta, filtro de Kalman, modelos en espacio de los estados.

Clasificación AMS: C13; C32; C40.

1. INTRODUCCIÓN

Como es conocido, la estimación por máxima verosimilitud exacta de modelos econométricos de series temporales puede abordarse aplicando el filtro de Kalman a su representación equivalente en espacio de los estados (EE) (ver Gardner *et al.*, 1980 y Terceiro, 1990). En este trabajo se aplica este enfoque a la estimación por máxima verosimilitud exacta de modelos periódicos.

Los modelos periódicos son útiles para modelizar y predecir series estacionales (Noakes *et al.*, 1985). Sin embargo su estimación es más compleja y costosa que la de los modelos ARMA estacionales, debido a su elevado número de parámetros.

Como indica Vecchia (1985), una manera de estimar un proceso ARMA periódico (PARMA) consiste en aplicar un algoritmo estándar de máxima verosimilitud a su representación VARMA equivalente. Esto supone un elevado coste computacional, tanto en términos de almacenamiento como de tiempo de cálculo. Por ello, este autor sugiere un método eficiente que, sin embargo, no proporciona estimaciones exactas cuando el proceso incluye factores autorregresivos. Li y Hui (1988) resuelven esta deficiencia derivando un algoritmo de máxima verosimilitud exacta para modelos PARMA estacionarios, que extiende los resultados de McLeod (1975) al caso periódico.

Otros trabajos recientes proponen algoritmos rápidos para la estimación de modelos PARMA utilizando un enfoque recursivo. Así, Adams y Goodwin (1995) plantean un método que proporciona estimaciones consistentes pero no eficientes. Boshnakov (1996) adapta el procedimiento de Hannan y Rissanen (1982) y Hannan y Kavalieris (1984) al caso de modelos PARMA estacionarios. Estos métodos presentan los siguientes inconvenientes: a) sólo contemplan el caso univariante y estacionario, b) no consideran la posible existencia de variables exógenas en el modelo y c) cuando no todas las estaciones presentan la misma estructura dinámica incurren en ineficiencia, ya que amplían innecesariamente la dinámica del sistema.

En este trabajo se propone una formulación generalizada - proceso VARMAX periódico o PVARMAX - que evita estas limitaciones y desarrolla un algoritmo de máxima verosimilitud exacta. Este método es computacionalmente eficiente y numéricamente estable gracias a la combinación de: a) una representación de dimensión mínima en EE en forma *steady-state innovations* y b) un algoritmo para evaluar la función de verosimilitud que aprovecha las propiedades del filtro de Kalman cuando se aplica a esta representación. Plantear el problema en EE tiene ventajas adicionales, ya que facilita la aplicación de resultados existentes para tratar modelos estacionarios o no estacionarios, con variables explicativas estocásticas y/o deterministas (Casals y Sotoca, 1997), detectar raíces unitarias sin necesidad de cálculos adicionales y obtener los segundos momentos exactos de las estimaciones (Terceiro, 1990).

La estructura del artículo es la siguiente. En la sección 2 se define el proceso generalizado PVARMAX. En la sección 3 se deriva su formulación de dimensión mínima en EE y se caracterizan las situaciones de no estacionariedad y no invertibilidad. La sección 4 desarrolla el algoritmo para evaluar la función de verosimilitud exacta del modelo PVARMAX en EE. En la sección 5 se resume la estructura del algoritmo y sus principales propiedades. La validez del procedimiento se comprueba en la sección 6, simulando modelos análogos a los utilizados por Vecchia (1985) y Li y Hui (1988) en el caso univariante y una estructura multivariante similar a las anteriores. En la sección 7 se muestra un caso real y en la sección 8 se resumen las principales conclusiones.

2. PROCESO PVARMAX

Sean $\{z_t\}$ un vector $(m \times 1)$ de series temporales estacionales y $\{u_t\}$ un vector $(r \times 1)$ de variables exógenas de tamaño N . Se dice que $\{z_t\}$ sigue un proceso PVARMAX $_s(p, q, g)$, (α^i : 1, 2, ..., s), donde s es el número de períodos muestrales que forman una estación completa, si

$$F_t(B)z_t = G_t(B)u_t + \Theta_t(B)a_t ; a_t \sim \text{IID}(0, \Sigma_t) \quad [1]$$

siendo

$$F_t(B) = I - F_{t,1}B - F_{t,2}B^2 - \dots - F_{t,p}B^p \quad [2]$$

$$\Theta_t(B) = I - \Theta_{t,1}B - \Theta_{t,2}B^2 - \dots - \Theta_{t,q}B^q \quad [3]$$

$$G_t(B) = G_{t,0} + G_{t,1}B + G_{t,2}B^2 + \dots + G_{t,g}B^g \quad [4]$$

y las matrices de coeficientes cambian de acuerdo con la siguiente ley de variación determinista de período s

$$F_t(B) = F_{t/s}(B); G_t(B) = G_{t/s}(B); \Theta_t(B) = \Theta_{t/s}(B) \quad [5]$$

$$\Sigma_t = \Sigma_{t/s} \quad t = 1, 2, \dots, N \quad [6]$$

A los parámetros de $F_t(B)$ se les denomina parámetros AR estacionales, los de $\Theta_t(B)$ son los parámetros MA estacionales y los de Σ_t son las varianzas estacionales del error. Por tanto, un proceso PVARMAX puede interpretarse como un modelo VARMAX cuyos coeficientes cambian en el tiempo de forma periódica (ver Lütkepohl, 1993).

El modelo [1]-[6] incluye como casos particulares las formulaciones periódicas convencionales. Por ejemplo, un proceso PARMA $_s(p, q)$ ($\alpha^i, 1, 2, \dots, s$), es un caso particular de [1]-[6] con $m=1$ y $g_i = 0$. Es decir

$$f_t(B)z_t = \Theta_t(B)a_t \quad [7]$$

$$f_t(B) = 1 - f_{t,1}B - f_{t,2}B^2 - \dots - f_{t,p}B^p \quad [8]$$

$$\Theta_t(B) = 1 - \theta_{t,1}B - \theta_{t,2}B^2 - \dots - \theta_{t,q}B^q \quad [9]$$

$$f_t(B) = f_{t/s}(B); \Theta_t(B) = \Theta_{t/s}(B); E(a_t a_t^T) = \sigma_t^2, \sigma_{t/s}^2, \sigma_{t/s}^2, \quad t = 1, 2, \dots, N \quad [10]$$

3. FORMULACIÓN EN ESPACIO DE LOS ESTADOS DE UN PROCESO PVARMAX

Estimar un proceso PVARMAX en la forma [1]-[6] incurriría en la misma ineficiencia computacional que los métodos ya existentes cuando no todas las estaciones presentan la misma estructura. Por ello, en esta sección se propone una representación en EE que garantiza que la dimensión del vector de estado es la mínima imprescindible.

Las expresiones [1]-[6] pueden representarse en forma de EE como

$$x_{t/s} = \Phi_t x_t + \Gamma_t u_t + E_t a_t \quad [11]$$

$$z_t = H_t x_t + D_t u_t + a_t \quad [12]$$

donde el vector de perturbaciones a_t tiene una matriz de covarianzas Q_t y $\Phi_t, \Gamma_t, E_t, H_t$ y D_t son matrices de tamaño $(n_{t+1} \times n_t), (n_{t+1} \times r), (n_{t+1} \times m), (m \times n_t)$ y $(m \times r)$, respectivamente, que cambian en el tiempo de acuerdo con la ley de variación periódica $\Phi_t = \Phi_{t/s}, \Gamma_t = \Gamma_{t/s}, E_t = E_{t/s}, H_t = H_{t/s}, D_t = D_{t/s}$ y $Q_t = Q_{t/s}, t = 1, 2, \dots, N$. Nótese que la dimensión del vector de estado $x_{t/s}$ también cambia con la estación.

Las expresiones [11]-[12] caracterizan un modelo en EE en forma *steady-state innovations*. Como se verá en la sección 4, esta estructura garantiza determinadas propiedades de convergencia del filtro de Kalman que pueden aprovecharse para mejorar su estabilidad numérica y eficiencia computacional.

Para determinar la relación entre el modelo [1]-[6] y las matrices de [11]-[12] conviene distinguir entre el caso en que los s modelos tienen una misma estructura dinámica de aquél en que distintas estaciones tienen estructuras diferentes.

3.1. Formulación cuando todas las estaciones tienen la misma estructura dinámica.

Cuando todas las estaciones tienen la misma estructura dinámica, se cumple que $k_j' = k, (\forall j = 1, 2, \dots, s)$, siendo $k_j' = \max(p_j, q_j, g_j)$. En este caso, la norma de conversión del modelo [1]-[6] a la forma [11]-[12] es

$$\Phi_t' = \begin{bmatrix} F_{1,1} & I & 0 & \ddot{y} & 0 \\ F_{2,2} & 0 & I & \ddot{y} & 0 \\ \vdots & \vdots & \vdots & \ddot{y} & \vdots \\ F_{k&2,k&1} & 0 & 0 & \ddot{y} & I \\ F_{k&1,k} & 0 & 0 & \ddot{y} & 0 \end{bmatrix}; \Gamma_t' = \begin{bmatrix} G_{1,1} & F_{1,1} & G_{t,0} \\ G_{2,2} & F_{2,2} & G_{t,0} \\ \vdots & \vdots & \vdots \\ G_{k&2,k&1} & F_{k&2,k&1} & G_{t,0} \\ G_{k&1,k} & F_{k&1,k} & G_{t,0} \end{bmatrix} \quad [13]$$

$$E_t' = \begin{bmatrix} \Theta_{1,1} & F_{1,1} \\ \Theta_{2,2} & F_{2,2} \\ \vdots & \vdots \\ \Theta_{k&2,k&1} & F_{k&2,k&1} \\ \Theta_{k&1,k} & F_{k&1,k} \end{bmatrix}; H_t' = [I \ 0 \ \ddot{y} \ 0]; D_t' = G_{t,0}; Q_t' = R_t' = S_t' = E(a_t a_t^T) \quad [14]$$

donde $F_{i,j} = F_{i,j,s}$; $G_{i,j} = G_{i,j,s}$; $\Theta_{i,j} = \Theta_{i,j,s}$, $\forall i, j = 0, 1, \dots, k&1$.

Ejemplo. Si $s = 3$ y cada estación sigue un proceso AR(1), la representación convencional del modelo PAR₃(1) resultante es

$$\begin{aligned} z_{t,1} &= f_{1,1} z_{t-1} + a_{t,1} \\ z_{t,2} &= f_{2,1} z_{t-1} + a_{t,2} \\ z_{t,3} &= f_{3,1} z_{t-2} + a_{t,3} \quad t = 0, 3, 6, \dots \end{aligned} \quad [15]$$

y la representación de cada uno de estos modelos en EE es, respectivamente

$$\begin{aligned} z_{t,1} &= x_{t,1} + a_{t,1} \\ x_{t,2} &= f_{2,1} x_{t-1} + f_{2,1} a_{t,1} \end{aligned} \quad [16]$$

$$\begin{aligned} z_{t,2} &= x_{t,2} + a_{t,2} \\ x_{t,3} &= f_{3,1} x_{t-2} + f_{3,1} a_{t,2} \end{aligned} \quad [17]$$

$$\begin{aligned} z_{t,3} &= x_{t,3} + a_{t,3} \\ x_{t,4} &= f_{1,1} x_{t-3} + f_{1,1} a_{t,3} \end{aligned} \quad [18]$$

3.2. Formulación cuando no todas las estaciones tienen la misma estructura dinámica.

Cuando los s modelos no tienen la misma estructura dinámica (es decir, $k_1 \dots k_2 \dots \ddot{y} \dots k_s$), la norma de conversión [13]-[14] sigue siendo válida, pero da lugar a un vector de estado de dimensión no mínima. En la representación de dimensión mínima, el número de filas de Φ_t es $n_{\text{filas}} = m_j \sum_{j=1}^s (k_j \% k_j^{\&})$, donde

$k_j^{\&}$ es la parte entera de $\frac{k_j}{s}$

$$k_j^{\&} = \begin{cases} 1 & \text{si } k_j \% k_j^{\&} = s \& j \& t \text{ cuando } j > t \\ 1 & \text{si } k_j \% k_j^{\&} = s \& j \% s \& t \text{ cuando } j \neq t \\ 0 & \text{en caso contrario} \end{cases} \quad [19]$$

y es fácil comprobar que $n_{t \in \{1, \dots, s\}} = n_{t \in \{1, \dots, s\}}, \alpha_t = 1, 2, \dots, N$. La norma de conversión para obtener una formulación de dimensión mínima consiste en eliminar los estados redundantes, es decir, aquéllos que no estén afectados por ningún input observable (u_t) o no observable (a_t). La forma de construir las matrices Φ_t , Γ_t y E_t es la siguiente. Sea $k = \max\{k_j\}$ ($\alpha_t = 1, 2, \dots, s$). Entonces, si $k_{tj} > j$ para $j = 1, 2, \dots, k$

- 1) La matriz Γ_t se construye añadiendo los bloques de filas $[G_{tj} \quad F_{tj} \quad G_{t,0}]$.
- 2) La matriz E_t se construye añadiendo los bloques de filas $[G_{tj} \quad F_{tj}]$.
- 3) La matriz Φ_t se construye añadiendo los bloques de filas $[F_{tj} \quad 0 \dots 0]$.

Si $k_{tj} > j$ para $j = 1, 2, \dots, k$, se añaden a la matriz Φ_t las columnas $[0 \quad 0 \quad \dots \quad I]^T$.

Ejemplo: Sea un modelo donde $s = 3$, $k_1 = 6$, $k_2 = 1$ y $k_3 = 1$. Siguiendo la norma de conversión [13]-[14], la representación en EE de dimensión no mínima es

$$\Phi_1 = \begin{bmatrix} F_{21} & I & 0 & 0 & 0 & 0 \\ F_{32} & 0 & I & 0 & 0 & 0 \\ F_{13} & 0 & 0 & I & 0 & 0 \\ F_{24} & 0 & 0 & 0 & I & 0 \\ F_{35} & 0 & 0 & 0 & 0 & I \\ F_{16} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \Phi_2 = \begin{bmatrix} F_{31} & I & 0 & 0 & 0 & 0 \\ F_{12} & 0 & I & 0 & 0 & 0 \\ F_{23} & 0 & 0 & I & 0 & 0 \\ F_{34} & 0 & 0 & 0 & I & 0 \\ F_{15} & 0 & 0 & 0 & 0 & I \\ F_{26} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \Phi_3 = \begin{bmatrix} F_{11} & I & 0 & 0 & 0 & 0 \\ F_{22} & 0 & I & 0 & 0 & 0 \\ F_{33} & 0 & 0 & I & 0 & 0 \\ F_{14} & 0 & 0 & 0 & I & 0 \\ F_{25} & 0 & 0 & 0 & 0 & I \\ F_{36} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad [20]$$

y la aplicación de la regla [19] elimina los estados redundantes, proporcionando las siguientes matrices de dimensión mínima

$$\Phi_1 = \begin{bmatrix} F_{21} & 0 \\ F_{13} & I \\ F_{16} & 0 \end{bmatrix}; \Phi_2 = \begin{bmatrix} F_{31} & 0 & 0 \\ F_{12} & I & 0 \\ F_{15} & 0 & I \end{bmatrix}; \Phi_3 = \begin{bmatrix} F_{11} & I & 0 \\ F_{14} & 0 & I \end{bmatrix} \quad [21]$$

donde la dimensión del vector de estado $x_{t \in \{1, \dots, s\}}$ ha pasado de ser 6, igual a la máxima dinámica $k_1 = 6$, a ser cambiante con la estación y nunca mayor que 3.

3.3. Condiciones de estacionariedad e invertibilidad.

Las condiciones de estacionariedad e invertibilidad de un proceso periódico suelen caracterizarse a partir de su representación VARMA equivalente de parámetros fijos (ver, por ejemplo, Franses y Paap, 1994; Boswijk y Franses, 1995). Procediendo de la misma forma, para caracterizar las condiciones de estacionariedad de un proceso PVARMAX es necesario escribir la ecuación de estado [11] en la forma equivalente

$$x_{t \in \{1, \dots, s\}} = (\Phi_{t \in \{1, \dots, s\}} \Phi_{t \in \{1, \dots, s\}} \dots \Phi_t) x_t + \sum_{j=1}^s \sum_{i=1}^{k_j} [\Phi_{t \in \{1, \dots, s\}}] (\Gamma_{t \in \{1, \dots, s\}} u_{t \in \{1, \dots, s\}} + E_{t \in \{1, \dots, s\}} a_{t \in \{1, \dots, s\}}) \quad [22]$$

cuyas matrices no cambian cada s periodos debido a la ley de variación: $\Phi_t = \Phi_{t \in \{1, \dots, s\}}, \Gamma_t = \Gamma_{t \in \{1, \dots, s\}}, E_t = E_{t \in \{1, \dots, s\}}, t = 1, 2, \dots, N$. Una vez descrito el modelo en la forma [12]-[22], se dice que es *ciclo-estacionario* si todos los autovalores de la matriz $\Phi = \sum_{j=1}^s \Phi_{s \in \{1, \dots, s\}}$ son, en módulo, menores que la unidad (ver Anderson y Moore, 1979).

Análogamente, para caracterizar las condiciones de invertibilidad del modelo PVARMAX conviene utilizar su

representación VARX(4). De esta forma, el modelo [11]-[12] puede escribirse como

$$\mathbf{x}_{t\%1} = \bar{\Phi}(\mathbf{x}_{t\%1\&s}) + \bar{\Gamma} \begin{bmatrix} u_t \\ u_{t\&1} \\ \vdots \\ u_{t\%1\&s} \end{bmatrix} + \bar{E} \begin{bmatrix} z_t \\ z_{t\&1} \\ \vdots \\ z_{t\%1\&s} \end{bmatrix} \quad [23]$$

$$\mathbf{z}_{t\%1} = H_{t\%1} \bar{\Phi}(\mathbf{x}_{t\%1\&s}) + H_{t\%1} \bar{\Gamma} \begin{bmatrix} u_t \\ u_{t\&1} \\ \vdots \\ u_{t\%1\&s} \end{bmatrix} + H_{t\%1} \bar{E} \begin{bmatrix} z_t \\ z_{t\&1} \\ \vdots \\ z_{t\%1\&s} \end{bmatrix} + D_t u_t + a_{t\%1} \quad [24]$$

en donde

$$\bar{\Phi} = (\Phi_t + E_t H_t)(\Phi_{t\&1} + E_{t\&1} H_{t\&1}) \ddot{Y} (\Phi_{t\&1\%s} + E_{t\&1\%s} H_{t\&1\%s}) \quad [25]$$

$$\bar{E} = \begin{bmatrix} E_t, (\Phi_t + E_t H_t) E_{t\&1}, (\Phi_t + E_t H_t)(\Phi_{t\&1} + E_{t\&1} H_{t\&1}) E_{t\&2}, \ddot{Y}, \\ s\&2 \\ \vdots \\ K \\ j=0 \end{bmatrix} (\Phi_{t\&i} + E_{t\&i} H_{t\&i}) E_{t\%1\&s} \quad [26]$$

$$\bar{\Gamma} = \begin{bmatrix} \bar{\Gamma}_t, (\Phi_t + E_t H_t) \bar{\Gamma}_{t\&1}, (\Phi_t + E_t H_t)(\Phi_{t\&1} + E_{t\&1} H_{t\&1}) \bar{\Gamma}_{t\&2}, \ddot{Y}, \\ s\&2 \\ \vdots \\ K \\ j=0 \end{bmatrix} (\Phi_{t\&i} + E_{t\&i} H_{t\&i}) \bar{\Gamma}_{t\%1\&s}, \text{ con } \bar{\Gamma}_t = \Gamma_t + E_t D_t \quad [27]$$

Puesto que la ecuación [24] puede interpretarse como un proceso VARX(4) sustituyendo recursivamente $\mathbf{x}_{t\%1\&s}$ por su expresión según [23], resulta inmediato ver que para que los valores pasados de $\langle z_t \rangle$ tengan un efecto decreciente sobre el valor actual (concepto habitual de invertibilidad) es necesario que los autovalores de $\bar{\Phi}$ sean menores en módulo que la unidad. Esta es la misma condición necesaria y suficiente de invertibilidad obtenida por Bentarzi y Hallin (1994).

4. LA FUNCIÓN DE VEROSIMILITUD EXACTA DE UN MODELO PVARMAX

Supongamos que a) la matriz de observaciones $\mathbf{Z} = [z_1, z_2, \dots, z_N]$ ha sido generada por el modelo [11]-[12], b) las perturbaciones \mathbf{a}_t y el estado inicial \mathbf{x}_1 son independientes y siguen una distribución normal y c) \mathbf{x}_1 tiene una esperanza desconocida, $\bar{\mathbf{x}}_1$, y una matriz de covarianzas, \mathbf{P}_1 , también desconocida. Entonces (ver De Jong, 1988) la función soporte de verosimilitud puede expresarse como

$$l(\mathbf{Z}/\mathbf{U}, \theta) = \log^* \mathbf{P}_1^* + \sum_{j=1}^N \bar{\mathbf{x}}_1^T \mathbf{P}_1^{\&1} \bar{\mathbf{x}}_1 + \sum_{j=1}^N \log^* \mathbf{B}_t^* + \sum_{j=1}^N \bar{\mathbf{z}}_t^T \mathbf{B}_t^{\&1} \bar{\mathbf{z}}_t + \log^* \mathbf{P}_1^{\&1} + \log^* \mathbf{W}_N^* + [\mathbf{P}_1^{\&1} \bar{\mathbf{x}}_1 + \mathbf{w}_N]^T [\mathbf{P}_1^{\&1} + \mathbf{W}_N]^{\&1} [\mathbf{P}_1^{\&1} \bar{\mathbf{x}}_1 + \mathbf{w}_N] \quad [28]$$

y la evaluación de [28] puede realizarse en tres fases claramente diferenciadas: a) cálculo de $\bar{\mathbf{z}}_t$ y \mathbf{B}_t , b) cálculo de \mathbf{w}_N y \mathbf{W}_N y c) determinación de las condiciones iniciales $\bar{\mathbf{x}}_1$ y \mathbf{P}_1 .

Los valores de $\bar{\mathbf{z}}_t$ y \mathbf{B}_t pueden obtenerse propagando un filtro de Kalman a partir de un estado inicial y una incertidumbre asociada nulos, proceso que denotaremos a partir de ahora FK(0,0). Esto da lugar al siguiente filtro recursivo

$$\hat{x}_{t|t} = \Phi_t \hat{x}_{t|t-1} + \Gamma_t u_t + K_t \tilde{z}_t \quad [29]$$

$$\tilde{z}_t = z_t + H_t \hat{x}_{t|t-1} + D_t u_t \quad [30]$$

$$B_t = Q_t \quad [31]$$

$$K_t = E_t \quad [32]$$

$$P_{t|t} = 0 \quad [33]$$

en donde [29] propaga la estimación óptima del vector de estado contando con la información disponible hasta el instante anterior. Las ecuaciones [30] y [31] calculan las innovaciones, interpretables como errores de predicción a horizonte un período, y su matriz de covarianzas, respectivamente. La ganancia óptima del filtro viene dada por [32]. La ecuación [33] indica que $\bar{P} = 0$ es la única solución de la ecuación de Riccati del filtro de Kalman, propiedad asociada a los modelos en forma *steady-state innovations* (ver Anderson y Moore, 1979) y que cumplen, por tanto, los modelos PVARMAX expresados en la forma [11]-[12]. Aprovechar esta propiedad supone una ganancia computacional al evitar el cálculo recursivo de $P_{t|t}$ y una importante ganancia en estabilidad numérica, ya que esta propagación es la principal fuente de errores numéricos del Filtro de Kalman.

Por otra parte, el vector w_N y la matriz W_N se calculan mediante las recursiones

$$w_t = w_{t-1} + \bar{\Phi}_{t-1}^T H_t^T B_t^{-1} \tilde{z}_t \quad [34]$$

$$W_t = W_{t-1} + \bar{\Phi}_{t-1}^T H_t^T B_t^{-1} H_t \bar{\Phi}_{t-1} \quad [35]$$

$$\bar{\Phi}_t = (\Phi_t + K_t H_t) \bar{\Phi}_{t-1} \quad [36]$$

donde w_t y W_t se inicializan a cero, $\bar{\Phi}_0 = I$ y K_t es la ganancia de un FK(0,0).

La elección adecuada de condiciones iniciales (\bar{x}_1 y P_1) para la evaluación de [28] dependerá de si a) el modelo es estacionario o no y b) si contiene variables exógenas deterministas y/o estocásticas (ver De Jong y Chu-Chun-Lin, 1994; Casals y Sotoca, 1997).

4.1. Modelos PVARMAX estacionarios.

Si el modelo es estacionario y no tiene variables exógenas, la inicialización adecuada del filtro de Kalman es $\bar{x}_1 = 0$ y P_1 es la solución de la ecuación de Lyapunov correspondiente a [22]. Por tanto, la expresión [28] se convierte en

$$l(Z/U, \theta) = \log \prod_{t=1}^N P_1^{-1} \log \prod_{t=1}^N Q_t^{-1} \tilde{z}_t^T Q_t^{-1} \tilde{z}_t + \log \prod_{t=1}^N P_1^{-1} W_N^{-1} w_N + w_N^T [P_1^{-1} W_N]^{-1} w_N \quad [37]$$

donde w_N y W_N se calculan a partir de las recursiones $w_t = w_{t-1} + \bar{\Phi}_{t-1}^T H_t^T Q_t^{-1} \tilde{z}_t$ y $W_t = W_{t-1} + \bar{\Phi}_{t-1}^T H_t^T Q_t^{-1} H_t \bar{\Phi}_{t-1}$, siendo $\bar{\Phi}_t = (\Phi_t + E_t H_t) \bar{\Phi}_{t-1}$ (con $\bar{\Phi}_0 = I$).

Denotando por M al factor Cholesky de P_1 , sabemos que $P_1 = MM^T$, $P_1^{-1} = M^{-1} M^{-T}$ y $\log P_1^{-1} = \log P_1^{-1} + \log W_N^{-1} = \log I + M^T W_N M^{-1}$. Por tanto, [37] puede escribirse como

$$l(Z/U, \theta) = \log \prod_{t=1}^N M^T W_N M^{-1} \log \prod_{t=1}^N Q_t^{-1} \tilde{z}_t^T Q_t^{-1} \tilde{z}_t + w_N^T M [I + M^T W_N M]^{-1} M^T w_N \quad [38]$$

Si L es el factor Cholesky de $I + M^T W_N M$, tal que $(I + M^T W_N M)^{-1} = (L L^T)^{-1}$, es fácil obtener recursivamente el vector λ que satisface el sistema triangular de ecuaciones $L \lambda = M^T w_N$ y, consecuentemente, [38] puede escribirse de forma simplificada como

$$l(Z/U, \theta) = \log^* L^{-2} + \sum_{t=1}^N \log^* Q_t^* + \sum_{t=1}^N \bar{z}_t^T Q_t^{-1} \bar{z}_t + \lambda^T \lambda \quad [39]$$

Asimismo, si N es el número total de observaciones en la muestra y N_i es el número de observaciones de la estación i -ésima, de forma que $\sum_{i=1}^s N_i = N$, el término $\sum_{t=1}^N \log^* Q_t^*$ de [39] puede calcularse como $\sum_{i=1}^s N_i \log^* Q_i^*$.

4.2. Modelos PVARMAX no estacionarios.

Si el modelo no es estacionario y no tiene variables exógenas, el filtro [29]-[33] debe inicializarse con $\bar{x}_1 = 0$ y $P_1^{-1} = 0$, lo que supone asociar al estado inicial una incertidumbre infinita. En este caso, sólo es posible evaluar una aproximación a la función soporte de verosimilitud, definida como el límite de la distancia $l(Z/U, \theta) + \log^* P_1^*$ cuando P_1 tiende a infinito, que es

$$\sum_{i=1}^s N_i \log^* Q_i^* + \sum_{t=1}^N \bar{z}_t^T Q_t^{-1} \bar{z}_t + \log^* W_N^* + w_N^T W_N^{-1} w_N \quad [40]$$

donde w_N y W_N se obtienen a partir de [34]-[36], teniendo en cuenta que $B_t = Q_t$ y $K_t = E_t \alpha^{-1}$, $\alpha = 1, 2, \dots, N$.

4.3. Modelos PVARMAX parcialmente no estacionarios.

Si el modelo es parcialmente no estacionario y no contiene variables exógenas, algunas raíces de $\Phi^k = \sum_{j=1}^s \Phi_{s,j} \lambda_j^k$ son, en módulo, mayores o iguales a la unidad y otras, menores que la unidad. En esta situación, las condiciones iniciales adecuadas son $\bar{x}_1 = 0$ y P_1^{-1} es una matriz finita y distinta de cero (ver De Jong y Chu-Chun-Lin, 1994).

La posible presencia de variables exógenas en el modelo (estacionario, no estacionario o parcialmente no estacionario) y su carácter determinista o estocástico modifica las condiciones iniciales del filtro de Kalman (ver Casals y Sotoca, 1997).

5. ESTRUCTURA DEL ALGORITMO Y PROPIEDADES

5.1. Caso estacionario.

La evaluación de la función de verosimilitud [39] para un proceso PVARMAX estacionario puede estructurarse en los siguientes pasos:

- 1) Calcular la factorización de Cholesky de las s matrices Q_i , denotada por Q_i^{-1} , su determinante $(|Q_i^{-1}| = |Q_i|^{-1})$ y una matriz R_i tal que $R_i Q_i R_i^T = I$ ($R_i = [Q_i^{-1}]^{1/2}$).
- 2) Calcular la matriz P_1 a partir de la solución de la ecuación de Lyapunov $P_1 = \Phi^c P_1 (\Phi^c)^T + Q^c$ y su factor Cholesky M .
- 3) Calcular las secuencias \bar{z}_t , w_t y W_t , a partir de las expresiones [29]-[33] y [34]-[36], respectivamente.
- 4) Evaluar el vector $M^T w_N$.
- 5) Calcular la matriz $I + M^T W_N M$, su factor Cholesky L y su determinante $|I + M^T W_N M| = |L|^{-2}$.

- 6) Usar un procedimiento de sustitución recursiva para resolver en λ el sistema triangular $L\lambda = M^T w_N$.
- 7) Calcular los términos $\sum_{j=1}^N \bar{z}_t^T Q_t^{\&1} \bar{z}_t$ y $\sum_{j=1}^N \bar{z}_t^T Q_t^{\&1} \bar{z}_t + \lambda \lambda^T$. Para ello no es necesario invertir explícitamente la matriz Q_t , sino que conviene utilizar las inversas de los s factores de Cholesky calculadas en el paso 1) ($R_j, \alpha^j, 1, 2, \dots, s$).

5.2. Caso no estacionario.

En el caso no estacionario, los pasos necesarios para evaluar la función de verosimilitud [40] son los siguientes:

- 1) Idéntico al caso estacionario.
- 2) No es necesario, ya que [40] no depende de P_1 .
- 3) Idéntico al caso estacionario.
- 4), 5) y 6) Se calculan los términos $\sum_{j=1}^s N_j \log^* P_j^* + \log^* W_N^*$ a partir de la factorización de Cholesky de W_N .
- 7) Se evalúa el término $\sum_{t=1}^N \bar{z}_t^T Q_t^{\&1} \bar{z}_t + w_N^T W_N^{\&1} w_N$ aprovechando nuevamente los factores de Cholesky previamente calculados.

Consecuentemente, las diferencias fundamentales entre este caso y el caso estacionario consisten en que no se calcula el factor Cholesky de P_1 (denotado por M) ya que tiende a infinito. En cambio, sí se evalúa el término $\log^* P_1^* + \log^* W_N^*$ y la parte de P_1 asociada a los componentes estacionarios del sistema.

5.3. Propiedades del algoritmo.

Las principales propiedades del algoritmo desarrollado en las secciones 5.1 y 5.2 son las siguientes:

- 1) Evita el cálculo del término $\log^* P_1^* + \log^* W_N^*$ que aparece en [28], usando la descomposición de Cholesky de la matriz P_1 . Esto mejora la estabilidad numérica en la evaluación de la función de verosimilitud.
- 2) Permite detectar situaciones de no estacionariedad sin necesidad de cálculos adicionales. Según la sección 3.3, la estacionariedad del proceso se comprueba calculando el módulo de los autovalores de Φ^L . Para ello, lo más eficiente es resolver la ecuación de Lyapunov $P_1 = \Phi^L (P_1 (\Phi^L)^T + Q^L)$ mediante una descomposición Schur compleja de Φ^L (ver Petkov *et al.*, 1991) lo que proporciona, como subproducto del proceso de cálculo, los autovalores de Φ^L .
- 3) Permite detectar situaciones de no invertibilidad sin necesitar cálculos adicionales. De acuerdo con la sección 3.3, si el modelo es invertible la matriz $\bar{\Phi}_t^L = (\Phi_t^L + E_t H_t^T) \bar{\Phi}_{t-1}^L$ converge a cero cuando $t \ll N$, siendo N el tamaño muestral, y esta secuencia se evalúa en la recursión [36].
- 4) Si a_t y x_1 son variables aleatorias normales independientes, las innovaciones del filtro [29]-[33] se distribuyen como una normal con esperanza nula y, cuando $t \ll N$, convergen en media cuadrática a las perturbaciones a_t (Shea, 1989, pág. 162). Por tanto estas innovaciones son instrumentos adecuados para la diagnosis del modelo.

5. RESULTADOS CON DATOS SIMULADOS

En este apartado se presenta un conjunto de resultados de simulación que ilustran las propiedades del procedimiento desarrollado en el apartado 4. En el primer ejercicio se utiliza la misma estructura $PARMA_2(1,1)$ que en Vecchia (1985) y Li y Hui (1988): $(1\%f_{t,1}B)z_t' (1\%\theta_{t,1}B)a_t$, con $f_{t,1}' f_{t/s,1}'$, $\theta_{t,1}' \theta_{t/s,1}'$ y $\sigma_t^2 \sigma_{t/s}^2$, $s=2$ para $t' 1,2,\dotsc,N$. Por tanto, para la primera estación se tiene

$$(1\%f_{1,1}B)z_t' (1\%\theta_{1,1}B)a_{1,t}, a_{1,t} \sim \text{IID}(0, q_{1,1}), t' 1,3,5,\dotsc [41]$$

mientras que el modelo de la segunda estación es

$$(1\%f_{2,1}B)z_t' (1\%\theta_{2,1}B)a_{2,t}, a_{2,t} \sim \text{IID}(0, q_{2,1}), t' 2,4,6,\dotsc [42]$$

A partir de esta formulación, se simulan muestras de 100 y 200 observaciones, con 150 repeticiones para cada tamaño muestral. En el Cuadro 1 se ofrece, para cada tamaño muestral, la media de las estimaciones, la raíz cuadrada del error cuadrático medio (RMSE), la desviación típica muestral (DTM) y la desviación típica exacta (DTE), calculada sustituyendo los valores teóricos de los parámetros en la matriz de información exacta (ver Terceiro, 1990). Comparando esta cifra con la desviación típica muestral, se obtiene una idea de la eficiencia de la estimación y se puede comprobar si las desviaciones típicas muestrales convergen a las exactas, calculadas a partir de la cota de Cramer-Rao.

Tanto el RMSE como el sesgo de nuestras estimaciones son menores, y en algunos casos mucho menores, que los presentados por Vecchia (1985) y prácticamente iguales a los de Li y Hui (1988). Por ejemplo, según Vecchia, el RMSE del parámetro $f_{1,1}$ con $N=100$ es .825, mientras que con nuestro procedimiento es .200, ver Cuadro 1. También cabe destacar cómo mejora la eficiencia del estimador, medida por el RMSE, al aumentar el tamaño muestral. En cuanto a la eficiencia computacional, el número de flops (operaciones aritméticas de punto flotante) usado en la estimación del $PARMA_2(1,1)$ es similar al usado en la estimación de un $ARMA(1,1)$ no periódico.

Cuadro 1
RESULTADOS DE LA ESTIMACIÓN DEL MODELO [41]-[42]

		$f_{1,1}'$ &.8	$\theta_{1,1}'$ &.3	$q_{1,1}'$ 1.0	$f_{2,1}'$ &.6	$\theta_{2,1}'$ &.5	$q_{2,1}'$ 1.0
N=100	Media	-0.754	-0.239	0.960	-0.649	-0.548	0.968
	RMSE	0.200	0.217	0.213	0.217	0.288	0.218
	DTM	0.195	0.208	0.209	0.211	0.284	0.216
	DTE	0.383	0.394	0.201	0.275	0.304	0.200
N=200	Media	-0.753	-0.257	0.961	-0.636	-0.538	0.990
	RMSE	0.147	0.144	0.144	0.168	0.193	0.136
	DTM	0.139	0.138	0.138	0.164	0.190	0.136
	DTE	0.269	0.277	0.142	0.194	0.214	0.141

Para validar el comportamiento del algoritmo cuando las estructuras estocásticas de cada estación son diferentes, se ha simulado el proceso: $(1\%f_{t,1}B\%f_{t,2}B^2)z_t' (1\%\theta_{t,1}B)a_t$, con $s=2$ y $t' 1,2,\dotsc,N$, de forma que para la primera estación se tiene un modelo $ARMA(1,1)$

$$(1\%f_{1,1}B)z_t' (1\%\theta_{1,1}B)a_{1,t}, a_{1,t} \sim \text{IID}(0, q_{1,1}), t' 1, 3, 5, \dot{y} \quad [43]$$

mientras que el modelo de la segunda estación es un AR(2)

$$(1\%f_{2,1}B\%f_{2,2}B^2)z_t' a_{2,t}, a_{2,t} \sim \text{IID}(0, q_{2,1}), t' 2, 4, 6, \dot{y} \quad [44]$$

En el Cuadro 2 se muestran los resultados correspondientes a la estimación de [43]-[44], para los mismos tamaños muestrales y el mismo número de repeticiones que el ejercicio anterior. Las conclusiones son análogas a las del caso anterior.

Cuadro 2
RESULTADOS DE LA ESTIMACIÓN DEL MODELO [43]-[44].

		$f_{1,1}'$ &.5	$\theta_{1,1}'$ &.9	$q_{1,1}'$ 16.0	$f_{1,2}'$ &.2	$f_{2,2}'$ &.7	$q_{2,1}'$ 64.0
N=100	Media	-0.555	-0.958	14.903	-0.183	-0.663	62.278
	RMSE	0.158	0.174	3.408	0.094	0.096	12.509
	DTM	0.168	0.184	3.580	0.095	0.103	13.036
	DTE	0.138	0.154	3.256	0.088	0.093	12.714
N=200	Media	-0.526	-0.932	15.581	-0.184	-0.685	64.275
	RMSE	0.125	0.136	2.299	0.064	0.075	9.714
	DTM	0.127	0.140	2.336	0.066	0.077	9.718
	DTE	0.098	0.110	2.282	0.062	0.067	9.019

Por último en el Cuadro 3 se muestran los resultados de la estimación de un modelo bivalente definido, siguiendo [1]-[4], como: $(I\%F_{t,1}B)z_t' (I\%\Theta_{t,1}B)a_t$, con $s=2$ y $t' 1, 2, \dot{y}, N$, de forma que para la primera estación el modelo es

$$\begin{bmatrix} 1\%f_{11}^1 B & f_{12}^1 B \\ 0 & 1\%f_{22}^1 B \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}, \begin{bmatrix} 1\%\theta_{11}^1 B & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}, E \left\{ \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \begin{bmatrix} a_{1,t} & a_{2,t} \end{bmatrix} \right\}, \begin{bmatrix} q_{11}^1 & 0 \\ 0 & q_{22}^1 \end{bmatrix}, t' 1, 3, 5, \dot{y} \quad [45]$$

mientras que para la segunda estación se tiene

$$\begin{bmatrix} 1\%f_{11}^2 B & 0 \\ f_{21}^2 B & 1\%f_{22}^2 B \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1\%\theta_{22}^2 B \end{bmatrix} \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix}, E \left\{ \begin{bmatrix} a_{1,t} \\ a_{2,t} \end{bmatrix} \begin{bmatrix} a_{1,t} & a_{2,t} \end{bmatrix} \right\}, \begin{bmatrix} q_{11}^2 & 0 \\ 0 & q_{22}^2 \end{bmatrix}, t' 2, 4, 6, \dot{y} \quad [46]$$

Los resultados del Cuadro 3 muestran nuevamente que las estimaciones son adecuadas, tanto en términos de sesgo como de RMSE.

Cuadro 3
RESULTADOS DE LA ESTIMACIÓN DEL MODELO [45]-[46]

		f_{11}^1	f_{12}^1	f_{22}^1	θ_{11}^1	q_{11}^1	q_{22}^1	f_{11}^2	f_{21}^2	f_{22}^2	θ_{22}^2	q_{11}^2	q_{22}^2
		&.8	&.3	&.5	-0.4	1.0	.5	&.7	&.3	&.5	&.8	.5	1.0
N=100	Media	-0.787	-0.295	-0.498	-0.371	0.923	0.490	-0.688	-0.291	-0.483	-0.791	0.483	0.953
	RMSE	0.177	0.143	0.087	0.248	0.217	0.099	0.072	0.130	0.281	0.359	0.092	0.206
	DTM	0.176	0.143	0.087	0.246	0.203	0.098	0.071	0.130	0.280	0.359	0.091	0.200
	DTE	0.162	0.135	0.083	0.260	0.201	0.100	0.067	0.112	0.305	0.366	0.100	0.200
N=200	Media	-0.805	-0.290	-0.499	-0.415	0.975	0.499	-0.693	-0.305	-0.494	-0.825	0.503	0.968
	RMSE	0.113	0.098	0.061	0.176	0.141	0.071	0.049	0.076	0.234	0.282	0.073	0.140
	DTM	0.113	0.097	0.061	0.176	0.139	0.017	0.048	0.076	0.234	0.281	0.073	0.137
	DTE	0.114	0.095	0.058	0.183	0.142	0.071	0.048	0.079	0.215	0.258	0.071	0.141
N=500	Media	-0.795	-0.294	-0.499	-0.405	0.987	0.496	-0.700	-0.306	-0.497	-0.797	0.496	0.991
	RMSE	0.073	0.062	0.043	0.118	0.071	0.041	0.029	0.052	0.141	0.164	0.046	0.093
	DTM	0.073	0.062	0.043	0.118	0.070	0.041	0.029	0.052	0.141	0.164	0.045	0.093
	DTE	0.072	0.060	0.037	0.115	0.090	0.045	0.030	0.050	0.136	0.163	0.045	0.089

6. PREDICCIÓN DE LA DEMANDA DE EFECTIVO DE LA RED COMERCIAL DE UNA ENTIDAD FINANCIERA

El objetivo de una entidad financiera es cubrir la demanda diaria de efectivo de la red de oficinas del mismo. Para ello, se dispone de una muestra de demanda agregada (D_t) de lunes a viernes desde Enero de 1996 hasta Marzo de 1998 en millones de pts. El primer paso en el análisis es depurar la serie de un conjunto de irregularidades predecibles como son, por ejemplo, los días principio y fin de mes o el viernes anterior a Navidad.

El siguiente paso ha sido modelizar la demanda usando un modelo no periódico. Un análisis univariante estándar sugiere representar esta variable mediante un proceso ARIMA(1,0,0)×(0,1,1)₅. Los resultados de la estimación por máxima verosimilitud exacta de este modelo son

$$\begin{aligned} (1 + 0.63B)L_5 D_t &= (1 + 0.88B^5)\hat{a}_t \\ (0.03) & \quad (0.02) \end{aligned} \quad [47]$$

$$\hat{\sigma}_a = 506.69$$

donde B es el operador retardo, $L_5 = 1 + B^5$ y el período estacional es $s = 5$. Entre paréntesis, se ofrece la desviación típica de cada parámetro y $\hat{\sigma}_a$ es la desviación típica residual en millones de pesetas. Los residuos no muestran estructura dinámica adicional.

La estimación del modelo periódico identificado para esta misma variable proporciona los siguientes resultados

$$\begin{aligned} \text{Lunes: } (1 + 0.15B + 0.79B^2 + 0.24B^3)(1 + 0.08B^5 + 0.41B^{10})D_{1,t} &= \hat{a}_{1,t} \\ (0.08) \quad (0.13) \quad (0.14) \quad (0.09) \quad (0.09) & \\ \hat{\sigma}_1 = 595.99 & \end{aligned} \quad [48]$$

$$\text{Martes: } (1 \pm 0.54B)_{L_5} D_{2,t} \pm (1 \pm 0.85B^5) \hat{a}_{2,t} \quad [49]$$

$$(0.05) \quad (0.05)$$

$$\hat{\sigma}_2 \pm 457.25$$

$$\text{Miércoles: } (1 \pm 0.55B \pm 0.19B^2)(1 \pm 0.33B^5)_{L_5} D_{3,t} \pm (1 \pm 0.93B^5) \hat{a}_{3,t} \quad [50]$$

$$(0.08) \quad (0.06) \quad (0.11) \quad (0.05)$$

$$\hat{\sigma}_3 \pm 405.60$$

$$\text{Jueves: } (1 \pm 0.82B) D_{4,t} \pm \hat{a}_{4,t} \quad [51]$$

$$(0.07)$$

$$\hat{\sigma}_4 \pm 429.13$$

$$\text{Viernes: } (1 \pm 0.51B \pm 0.43B^2)_{L_5} D_{5,t} \pm (1 \pm 0.79B^5) \hat{a}_{5,t} \quad [52]$$

$$(0.10) \quad (0.12) \quad (0.06)$$

$$\hat{\sigma}_5 \pm 463.11$$

Obsérvese que la estructura dinámica es distinta para cada día de la semana y que el único modelo que coincide con la estructura no periódica es el correspondiente a los martes. Asimismo las desviaciones típicas residuales son también muy dispares, siendo muy volátil la demanda de los lunes en comparación con el resto de los días. Los parámetros del modelo periódico son significativos a un nivel de confianza del 95% y los residuos resultantes no rechazan la hipótesis de ruido blanco.

La equivalencia entre los modelos no periódico y periódico se rechaza mediante un test de razón de verosimilitudes, que toma un valor de 89.41 para 17 grados de libertad. Esto es debido a que el valor soporte de la verosimilitud del modelo [47] es 4408.44, mientras que en el modelo [48]-[52] es 4363.73. En el Cuadro 4 se muestran distintos criterios de información que confirman el mayor poder predictivo del modelo periódico frente al univariante.

Cuadro 4

Criterio Información	Modelo univariante [47]	Modelo periódico [48]-[52]
AIC	1532	1522
BIC	1534	1537
HQ	1533	1522

donde AIC denota el criterio de Akaike (1974), BIC denota el criterio bayesiano de Schwartz (1978) y HQ el de Hannan y Quinn (1979). Nótese que sólo el criterio BIC llevaría a rechazar la especificación periódica, pero esto se debe a que es el criterio que más penaliza el incremento de parámetros que supone la representación periódica.

Puesto que el objetivo de la entidad es cubrir la demanda de efectivo, la predicción puntual de la demanda es tan importante como la de la desviación típica del error de predicción. Ambos valores permiten calcular el aprovisionamiento de efectivo que permite cubrir la demanda con un determinado nivel de confianza. Puesto que los lunes son más volátiles que el resto de los días, el modelo no periódico tenderá a generar una infracobertura de la demanda los lunes y una sobrecobertura el resto de los días de la semana. Concretamente, para un nivel de confianza del 95%, la infracobertura supone aproximadamente un 10% de la demanda media de efectivo de los lunes.

7. CONCLUSIONES

La literatura econométrica ha mostrado un interés creciente por los modelos periódicos, debido a su utilidad para predecir series temporales estacionales. Sin embargo, su estimación es compleja, por el elevado número de parámetros que los caracterizan y costosa en términos de tiempo de cálculo. En este trabajo se propone un procedimiento eficiente y estable numéricamente para estimar modelos VARMAX periódicos por máxima verosimilitud exacta.

Los métodos estándar adolecen de falta de flexibilidad ya que: a) tratan de forma diferenciada la estimación de modelos univariantes y multivariantes, b) sólo contemplan el caso estacionario, c) no consideran la posible existencia de variables exógenas en el modelo y d) cuando no todas las estaciones presentan la misma estructura dinámica incurren en ineficiencia, ya que aumentan artificialmente la dinámica del sistema.

En este trabajo se formula un proceso periódico generalizado, se deriva su representación en espacio de los estados de dimensión mínima y se propone un algoritmo rápido y estable para calcular su función de verosimilitud exacta. Estas aportaciones configuran un proceso de modelización y cálculo que, frente a los existentes en la literatura, presenta ventajas en términos de flexibilidad de la representación, eficiencia computacional, estabilidad numérica y capacidad para aplicar resultados conocidos en el contexto de modelos en espacio de los estados.

En cuanto al primer aspecto (flexibilidad de la representación), se parte de la definición de una familia generalizada de procesos estocásticos periódicos (PVARMAX). Esta representación incluye los casos univariante y multivariante, cuando lo habitual en la literatura es tratar ambas situaciones de forma diferenciada.

El segundo aspecto, eficiencia y estabilidad numérica, se consigue combinando a) una representación del proceso PVARMAX como un modelo en espacio de los estados de dimensión mínima en forma *steady-state innovations* y b) un algoritmo estable para calcular la función de verosimilitud exacta, que aprovecha las propiedades de convergencia del filtro de Kalman cuando se aplica a modelos en forma *steady-state innovations*. Así, se consigue que el número de operaciones usado en la estimación de un PARMAX_s(p, q, g), donde p_i p , q_i q , g_i g , (α^i 1,2,...,s), s es el número de estaciones consideradas y g el número de variables exógenas, no difiera sustancialmente del correspondiente a la estimación de una estructura ARMAX_s(p, q, g) univariante.

En cuanto al tercer y último aspecto, capacidad para aplicar resultados conocidos, el uso de una representación en espacio de los estados facilita a) el tratamiento de modelos estacionarios o no estacionarios, con variables explicativas estocásticas y/o deterministas (Casals y Sotoca, 1997), b) la detección de situaciones de no estacionariedad y no invertibilidad sin necesidad de cálculos adicionales y c) el cálculo de la matriz de información analítica y, por tanto, de los segundos momentos de las estimaciones (Terceiro, 1990).

REFERENCIAS

- Adams, G.J. y Goodwin, G.C. (1995): «Parameter Estimation for Periodic ARMA Models». *Journal of Time Series Analysis*, 16, 2, 127-145.
- Anderson, B.D.O. y Moore, J.B. (1979): «Optimal Filtering». *Prentice Hall*, Englewood Cliffs (Nueva Jersey).
- Bentarzi, M. y Hallin, M. (1994): «On the Invertibility of Periodic Moving-Average Models». *Journal of Time Series Analysis*, 15, 3, 263-268.
- Boshnakov, G.N. (1996): «Recursive Computation of the Parameters of Periodic Autoregressive Moving-Average Processes». *Journal of Time Series Analysis*, 17, 4, 333-349.

- Boswijk, P. y Franses, P.H. (1995): «Unit roots in Periodic Autorregressions». *Journal of Time Series Analysis*, 17, 3, 221-245.
- Casals, J. y Sotoca, S. (1997): «Exact Initial Conditions for Maximum Likelihood Estimation of State Space Models with Stochastic Inputs». *Economics Letters*, 57, 3, 261-267.
- De Jong, P. (1988): «The Likelihood for a State Space Model». *Biometrika*, 75, 1, 165-169.
- De Jong, P. y Chu-Chun-Lin, S. (1994): «Stationary and Non-Stationary State Space Models». *Journal of Time Series Analysis*, 15, 2, 151-166.
- Franses, P.H. y Paap, R. (1994): «Model Selection in Periodic Autorregressions». *Oxford Bulletin of Economics and Statistics*, 56, 4, 421-439.
- Gardner, G., Harvey, A.C. y Phillips, G.D.A. (1980): «Algorithm 154. An Algorithm for Exact Maximum-Likelihood Estimation of Autoregressive-Moving Average Models by Means of Kalman Filtering». *Applied Statistics*, 29, 311-317.
- Hannan, E.J. y Kavalieris, L. (1984): «Multivariate Linear Time Series Models». *Advances in Applied Probability*, 16, 492-561.
- Hannan, E.J. y Rissanen, J. (1982): «Recursive Estimation of Mixed Autoregressive-Moving Average Model». *Biometrika*, 69, 81-94.
- Lütkepohl, H. (1993): «Introduction to Multiple Time Series Analysis». *Springer-Verlag*, Berlin.
- Li, W.K. y Hui, Y.V. (1988): «An Algorithm for the Exact Likelihood of Periodic Autoregressive Moving Average Models». *Commun. Statist.-Simula.*, 17, 4, 1483-1494.
- McLeod, A.L. (1975): «Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Time Series». *Applied Statistics*, 24, 255-256.
- Noakes, D.J., McLeod, A.L. y Hipel, K. (1985): «Forecasting Monthly Riverflow Time Series». *International Journal of Forecasting*, 1, 179-190.
- Petkov, P.Hr., Christov, N.D. y Konstantinov, M.M. (1991): «Computational Methods for Linear Control Systems». *Prentice-Hall*, Englewood Cliffs (New Jersey).
- Shea, B.L. (1989): «Algorithm AS 242. The Exact Likelihood of a Vector Autoregressive-Moving Average Model». *Applied Statistics*, 38, 161-204.
- Terceiro, J. (1990): «Estimation of Dynamic Econometric Models with Errors in Variables». *Springer-Verlag*, Heidelberg.
- Vecchia, A.V. (1985): «Maximum Likelihood Estimation for Periodic Autoregressive Moving Average Models». *Technometrics*, 27, 4, 375-384.

A FAST ALGORITHM TO COMPUTE THE EXACT LIKELIHOOD OF PERIODIC VARMAX PROCESSES

SUMMARY

In this work we derive a fast algorithm to compute the exact likelihood of periodic VARMAX processes. Its computational efficiency and numerical stability are achieved by combining a) a minimal dimension state-space model in steady-state innovations form with b) a computational procedure that takes advantage of the properties of this representation. The algorithm can be applied to stationary and nonstationary models and makes easy the calculation of the analytical second-order moments of the estimates. Simulation results and an example with real data illustrate its good behavior.

Key words: periodic VARMAX models, exact maximum likelihood, Kalman filter, state-space models.

AMS Classification: C13; C32; C40.

APÉNDICE: CÁLCULO EFICIENTE DE LA MATRIZ DE INFORMACIÓN.

La función soporte de verosimilitud [25], sin tener en cuenta constantes, puede escribirse como

$$l(\mathbf{Z}/\mathbf{U}, \boldsymbol{\theta}) = \prod_{t=1}^N \log \left(\mathbf{B}_t^* \sum_{j=1}^N \tilde{\mathbf{z}}_t^T \mathbf{B}_t^{*j} \tilde{\mathbf{z}}_t \right) \quad [1.1]$$

donde $\tilde{\mathbf{z}}_t$ y \mathbf{B}_t son las innovaciones y su matriz de covarianzas resultantes del siguiente filtro de Kalman inicializado adecuadamente

$$\tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t-1} + \mathbf{D}_t \mathbf{u}_t \quad [1.2]$$

$$\hat{\mathbf{x}}_{t|t} = \boldsymbol{\Phi}_t \hat{\mathbf{x}}_{t|t-1} + \boldsymbol{\Gamma}_t \mathbf{u}_t + \mathbf{K}_t \tilde{\mathbf{z}}_t \quad [1.3]$$

$$\mathbf{K}_t = (\boldsymbol{\Phi}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T + \mathbf{E}_t \mathbf{Q}_t) \mathbf{B}_t^{*j} \quad [1.4]$$

$$\mathbf{P}_{t|t} = \boldsymbol{\Phi}_t \mathbf{P}_{t|t-1} \boldsymbol{\Phi}_t^T + \mathbf{E}_t \mathbf{Q}_t \mathbf{E}_t^T + \mathbf{K}_t \mathbf{B}_t \mathbf{K}_t^T \quad [1.5]$$

$$\mathbf{B}_t = \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T + \mathbf{Q}_t \quad [1.6]$$

A partir de [1.1] es fácil ver que la expresión analítica del gradiente de la función de verosimilitud es

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_j} = \sum_{t=1}^N \text{tr} \left[\mathbf{B}_t^{*j} \frac{\partial \mathbf{B}_t}{\partial \boldsymbol{\theta}_j} \right] + 2 \sum_{t=1}^N \tilde{\mathbf{z}}_t^T \mathbf{B}_t^{*j} \frac{\partial \tilde{\mathbf{z}}_t}{\partial \boldsymbol{\theta}_j} + \sum_{t=1}^N \tilde{\mathbf{z}}_t^T \mathbf{B}_t^{*j} \frac{\partial \mathbf{B}_t}{\partial \boldsymbol{\theta}_j} \mathbf{B}_t^{*j} \tilde{\mathbf{z}}_t \quad [1.7]$$

donde las derivadas del vector de innovaciones y de su matriz de covarianzas se obtienen teniendo en cuenta [1.2]-[1.6]

$$\frac{\partial \tilde{\mathbf{z}}_t}{\partial \boldsymbol{\theta}_j} = \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\theta}_j} \hat{\mathbf{x}}_{t|t-1} + \frac{\partial \mathbf{D}_t}{\partial \boldsymbol{\theta}_j} \mathbf{u}_t \quad [1.8]$$

$$\frac{\partial \mathbf{B}_t}{\partial \boldsymbol{\theta}_j} = \frac{\partial \mathbf{H}_t}{\partial \boldsymbol{\theta}_j} \mathbf{P}_{t|t-1} \mathbf{H}_t^T + \frac{\partial \mathbf{Q}_t}{\partial \boldsymbol{\theta}_j} + \frac{\partial \boldsymbol{\Phi}_t}{\partial \boldsymbol{\theta}_j} \hat{\mathbf{x}}_{t|t-1} \hat{\mathbf{x}}_{t|t-1}^T + \frac{\partial \boldsymbol{\Gamma}_t}{\partial \boldsymbol{\theta}_j} \mathbf{u}_t + \frac{\partial \mathbf{K}_t}{\partial \boldsymbol{\theta}_j} \tilde{\mathbf{z}}_t \quad [1.9]$$

siendo $\bar{\Phi}_t = \Phi_t$ & $K_t H_t$ y $\bar{\Gamma}_t = \Gamma_t$ & $K_t D_t$. La propagación de [1.9] requiere calcular la derivada de la ganancia del filtro de Kalman. A partir de [1.4] se tiene que

$$\frac{MK_t}{N_j} = \left[\frac{M\Phi_t}{N_j} P_{t|t-1} H_t^T + \frac{M\Phi_t}{N_j} \frac{MP_{t|t-1}}{N_j} H_t^T + \frac{ME_t}{N_j} Q_t + E_t \frac{MQ_t}{N_j} \right] B_t^{*1} + K_t \frac{MB_t}{N_j} B_t^{*1} \quad [1.10]$$

$$\frac{MB_t}{N_j} = H_t \frac{MP_{t|t-1}}{N_j} H_t^T + \frac{MQ_t}{N_j} \quad [1.11]$$

$$\frac{MP_{t|t-1}}{N_j} = \bar{\Phi}_t \frac{MP_{t-1|t-1}}{N_j} \bar{\Phi}_t^T + A_{it} + A_{it}^T \quad [1.12]$$

siendo

$$A_{it} = \frac{M\Phi_t}{N_j} P_{t|t-1} \Phi_t^T + \frac{ME_t}{N_j} Q_t K_t^T + E_t \frac{MQ_t}{N_j} K_t^T + \frac{1}{2} \frac{M(E_t Q_t E_t^T)}{N_j} + \frac{1}{2} K_t \frac{MQ_t}{N_j} K_t^T \quad [1.13]$$

Para validar el modelo estimado es necesario calcular la matriz de covarianzas de las estimaciones. Ésta puede obtenerse a partir de la matriz de información del modelo, ya que su inversa proporciona una cota inferior a la matriz de covarianzas de los estimadores de los parámetros. El término general de la matriz de información, Terceiro (1990), es

$$M(\theta)_{ij} = \sum_{t=1}^N \frac{1}{2} \text{tr} \left[B_t^{*1} \frac{MB_t}{N_j} B_t^{*1} \frac{MB_t}{N_j} \right] + \sum_{t=1}^N \text{tr} \left[B_t^{*1} E \left[\frac{N\bar{z}_t}{N_j} \frac{N\bar{z}_t^T}{N_j} \right] \right] \quad [1.14]$$

donde se ha tenido en cuenta que el proceso de innovaciones tiene esperanza nula y que la derivada de este proceso no depende de las innovaciones. En [1.14] el único término desconocido es la esperanza del producto de las derivadas del vector de innovaciones. A menudo se aproxima la esperanza de ese producto por el producto de las derivadas. Sin embargo, para calcularlo de forma exacta, puede tenerse en cuenta que

$$E \left[\frac{N\bar{z}_t}{N_j} \frac{N\bar{z}_t^T}{N_j} \right] = B_t^{ij} + \bar{z}_{it} \bar{z}_{it}^T \quad [1.15]$$

donde \bar{z}_{it} es la media de $N\bar{z}_t/N_j$ y B_t^{ij} es la matriz de covarianzas de $N\bar{z}_t/N_j$ y $N\bar{z}_t/N_j$. El problema queda reducido, por tanto, a calcular los dos sumandos del lado derecho de [1.15]. Para ello, se define el sistema ampliado

$$x_{t+1}^c = \Phi_t^c x_t^c + \Gamma_t^c u_t + K_t^c z_t \quad [1.16]$$

$$z_t^c = H_t^c x_t^c + D_t^c u_t \quad [1.17]$$

donde

$$\mathbf{x}_t^c \cdot \begin{bmatrix} \bar{x}_{t+1} \\ \frac{\bar{M}_{t+1}}{\bar{M}_i} \\ \frac{\bar{M}_{t+1}}{\bar{M}_j} \end{bmatrix}; \Phi_t^c \cdot \begin{bmatrix} \bar{\Phi}_t & \mathbf{0} & \mathbf{0} \\ \bar{\Phi}_t & \bar{\Phi}_t & \mathbf{0} \\ \bar{\Phi}_t & \mathbf{0} & \bar{\Phi}_t \end{bmatrix}; \Gamma_t^c \cdot \begin{bmatrix} \Gamma_t \\ \frac{\bar{M}_t}{\bar{M}_i} \& \mathbf{K}_t \frac{\bar{M}_t}{\bar{M}_j} \\ \frac{\bar{M}_t}{\bar{M}_j} \& \mathbf{K}_t \frac{\bar{M}_t}{\bar{M}_j} \end{bmatrix}; \mathbf{K}_t^c \cdot \begin{bmatrix} \bar{K}_t \\ \frac{\bar{M}_t}{\bar{M}_i} \\ \frac{\bar{M}_t}{\bar{M}_j} \end{bmatrix} \quad [1.18]$$

$$\mathbf{z}_t^c \cdot \begin{bmatrix} \bar{z}_t \\ \frac{\bar{M}_t}{\bar{M}_i} \\ \frac{\bar{M}_t}{\bar{M}_j} \end{bmatrix}; \mathbf{H}_t^c \cdot \begin{bmatrix} \mathbf{0} & \& \mathbf{H}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \& \mathbf{H}_t \end{bmatrix}; \mathbf{D}_t^c \cdot \begin{bmatrix} \frac{\bar{M}_t}{\bar{M}_i} \\ \& \\ \frac{\bar{M}_t}{\bar{M}_j} \\ \& \end{bmatrix} \quad [1.19]$$

Si se denota por $\bar{\mathbf{x}}_t^c \cdot E(\mathbf{x}_t^c)$, $\bar{\mathbf{z}}_t^c \cdot E(\mathbf{z}_t^c)$, $\mathbf{P}_t^c \cdot E[(\mathbf{x}_t^c \& \bar{\mathbf{x}}_t^c)(\mathbf{x}_t^c \& \bar{\mathbf{x}}_t^c)^T]$ y $\mathbf{B}_t^c \cdot E[(\mathbf{z}_t^c \& \bar{\mathbf{z}}_t^c)(\mathbf{z}_t^c \& \bar{\mathbf{z}}_t^c)^T]$, las ecuaciones que permiten calcular estos momentos son

$$\bar{\mathbf{x}}_t^c \cdot \Phi_t^c \bar{\mathbf{x}}_t^c \& \Gamma_t^c \mathbf{u}_t \quad [1.20]$$

$$\bar{\mathbf{z}}_t^c \cdot \mathbf{H}_t^c \bar{\mathbf{x}}_t^c \& \mathbf{D}_t^c \mathbf{u}_t \quad [1.21]$$

$$\mathbf{P}_t^c \cdot \Phi_t^c \mathbf{P}_t^c (\Phi_t^c)^T \& \mathbf{K}_t^c \mathbf{B}_t^c (\mathbf{K}_t^c)^T \quad [1.22]$$

$$\mathbf{B}_t^c \cdot \mathbf{H}_t^c \mathbf{P}_t^c (\mathbf{H}_t^c)^T \quad [1.23]$$

La propagación de [1.20]-[1.21] permite calcular en cada instante los valores de \bar{z}_{it} y \bar{z}_{jt} , y las ecuaciones [1.22]-[1.23] proporcionan el valor de la matriz \mathbf{B}_t^{ij} , ya que la estructura por bloques de \mathbf{B}_t^c es

$$\mathbf{B}_t^c \cdot \begin{bmatrix} \mathbf{B}_t^{ii} & \mathbf{B}_t^{ij} \\ \mathbf{B}_t^{ji} & \mathbf{B}_t^{jj} \end{bmatrix}$$

Las condiciones iniciales del sistema dado por [1.20]-[1.23] son $\bar{\mathbf{x}}_1^c \cdot [\bar{x}_1^T \ \mathbf{0} \ \mathbf{0}]^T$ y $\mathbf{P}_1^c \cdot \text{diag}[\mathbf{P}_1, \mathbf{0}, \mathbf{0}]$ donde el cálculo correcto de $\bar{\mathbf{x}}_1$ y \mathbf{P}_1 se ha discutido en la Sección 4.

La matriz de información puede calcularse eficientemente descomponiendo el sistema [1.20]-[1.23] en un subsistema de dimensión menor, que permita calcular las variables \bar{z}_{it} , \bar{z}_{jt} y \mathbf{B}_t^{ij} . Así, descomponiendo la matriz \mathbf{P}_t^c en 3×3 bloques de dimensión n_t , de forma que $\mathbf{P}_t^{c ij}$ se corresponde con el bloque (i, j) , se tiene que

$$\mathbf{B}_t^{ij} \cdot \mathbf{H}_t \mathbf{P}_t^{c23} \mathbf{H}_t^T \quad [1.24]$$

y dada la estructura de Φ_t^c , el término \mathbf{P}_t^{c23} de [1.24] se calcula usando las expresiones

$$\mathbf{P}_t^{c23} \cdot \frac{\bar{\Phi}_t}{\bar{M}_i} \mathbf{P}_t^{c11} \frac{\bar{\Phi}_t^T}{\bar{M}_j} \& \frac{\bar{\Phi}_t}{\bar{M}_i} \mathbf{P}_t^{c21} \frac{\bar{\Phi}_t^T}{\bar{M}_j} \& \frac{\bar{\Phi}_t}{\bar{M}_i} \mathbf{P}_t^{c13} \bar{\Phi}_t^T \& \frac{\bar{\Phi}_t}{\bar{M}_i} \mathbf{P}_t^{c23} \bar{\Phi}_t^T \& \frac{\bar{M}_t}{\bar{M}_i} \mathbf{B}_t \frac{\bar{M}_t^T}{\bar{M}_j} \quad [1.25]$$

$$\mathbf{P}_t^{c11} \cdot \Phi_t \mathbf{P}_t^{c11} \Phi_t^T \& \mathbf{K}_t \mathbf{B}_t \mathbf{K}_t^T \quad [1.26]$$

$$P_{t,i}^{c21} = \frac{\bar{\Phi}_t}{N_{\theta_i}} P_t^{c11} \Phi_t^T + \bar{\Phi}_t P_t^{c21} \Phi_t^T + \frac{N K_t}{N_{\theta_i}} B_t K_t^T \quad [1.27]$$

$$P_{t,i}^{c13} = \Phi_t P_t^{c11} \frac{\bar{\Phi}_t^T}{N_{\theta_j}} + \Phi_t P_t^{c13} \bar{\Phi}_t^T + K_t B_t \frac{N K_t^T}{N_{\theta_j}} \quad [1.28]$$

donde puede observarse que no aparecen los bloques (2,2) y (3,3) de P_t^c . Por tanto, la carga computacional que supone la propagación de [1.25]-[1.28] es equivalente a la de un sistema de dimensión $2n_t$, no simétrico. En cuanto a la propagación de [1.20]-[1.21] para obtener \bar{z}_{it} y \bar{z}_{jt} , se calcula la secuencia de medias una sola vez para cada coeficiente estimado de la siguiente forma

$$\bar{x}_{t,i} = \Phi_t \bar{x}_t + \Gamma_t u_t \quad [1.29]$$

$$\bar{x}_{t,i,j} = \bar{\Phi}_t \bar{x}_{t,i} + \frac{N \Phi_t}{N_{\theta_j}} \bar{x}_t + \frac{N \Gamma_t}{N_{\theta_j}} u_t + K_t \frac{N D_t}{N_{\theta_j}} u_t \quad [1.30]$$

$$\bar{z}_{t,i} = H_t \bar{x}_{t,i} + \frac{N D_t}{N_{\theta_j}} u_t \quad [1.31]$$

donde $\bar{x}_{t,i} = E(\bar{x}_{t,i|t})$ y $\bar{x}_{t,i,j} = E(\bar{x}_{t,i,j|t}/N_{\theta_j})$. La descomposición efectuada en [1.15] mejora la eficiencia en los cálculos. La obtención de B_t^{ij} supone una menor carga computacional que la correspondiente a la obtención directa de la esperanza del producto de las derivadas de las innovaciones, dado que la primera de estas expresiones es más simple al cancelarse las variables exógenas por tratarse de un momento centrado en el valor medio. Dado que el número de propagaciones es igual a $h(h+1)/2$ donde h es la dimensión del vector θ , la carga que pueda suponer la propagación h veces de [1.29]-[1.31] queda sobradamente compensada.

Si no existen variables exógenas en el modelo, puede suponerse que la esperanza del producto de las derivadas del proceso de innovaciones coincide con la matriz B_t^{ij} , tomando como condiciones iniciales $P_1^c = \text{diag}(P_1 + \bar{x}_1 \bar{x}_1^T, \mathbf{0}, \mathbf{0})$. Por tanto, en este caso, no sería necesario propagar el sistema [1.29]-[1.31].