ICAE

Instituto Complutense de Análisis Económico

UNIVERSIDAD COMPLUTENSE

FACULTAD DE ECONOMICAS

Campus de Somosaguas

28223 MADRID

Teléfono 394 26 11 - FAX 294 26 13

FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES BIBLIOTECA - UCM -

UNIVERSITAS COMPLUTENSIS UNIVERSIDAD DE MADRID

XI/49 (9610)

Documento de trabajo

Un método de inicialización del filtrado para modelosen espacio de los estados con inputs estocásticos

José Casals y Sonia Sotoca

No.9610 Septiembre 1996

ICAE

Instituto Complutense de Análisis Económico

UNIVERSIDAD COMPLUTENSE

\K/
49
(9610)

UN METODO DE INICIALIZACION DEL FILTRADO PARA MODELOS EN ESPACIO DE LOS ESTADOS CON INPUTS ESTOCASTICOS

FACULTAD DE CIENCIAS ECONOMICAS Y EMPRESARIALES
BIBLIOTECA
UCM

José Casals*
Sonia Sotoca*

Universidad Complutense de Madrid
Campus de Somosaguas
28223 Madrid

ABSTRACT

We derive exact expressions for the conditional mean and variance of the initial state of a state space system with stochastic inputs, under stationarity or nonstationarity. These results generalize those of De Jong and Chu-Chun-Lin (1994) and provide a useful initialization method to obtain maximum likelihood estimates of the model parameters. As final estimates are sensitive to initial conditions, the presence of stochastic inputs -a frequent situation in Econometrics- should be considered when computing the mean and variance of the initial state. The results obtained with this method compare favourably with those obtained by standard procedures.

RESUMEN

En este trabajo se derivan las expresiones exactas de la media y varianza condicional del estado inicial de un modelo en espacio de los estados con inputs estocásticos, generalizando los resultados teóricos obtenidos por De Jong y Chu-Chun-Lin (1994). Se muestra que las condiciones iniciales exactas dependen del carácter estacionario o no estacionario del modelo y que las estimaciones finales de los parámetros son sensibles a la presencia de inputs estocásticos, siendo ésta una situación frecuente en Econometría. Los resultados obtenidos con datos simulados muestran el favorable funcionamiento del procedimiento desarrollado frente a otros más clásicos.

* Queremos agredecer los comentarios y sugerencias recibidos de Emilio Domínguez, Rafael Flores y Alfonso Novales.

Key words: initial conditions, stationarity, Kalman filter, exact maximum likelihood, state space models.

JEL classification codes: C32; C40.

Mailing address: Departamento de Economía Cuantitativa. Universidad Complutense. Somosaguas. 28223 Madrid. Spain. E-mail: eccua01@emducms1.sis.ucm.es.

Nº E. 5306520188
Nº C: X-53-231025-9

1. Introducción.

Es conocido que la estimación por máxima verosimilitud exacta (MVE) de un modelo en espacio de los estados (EE) puede abordarse utilizando alguna técnica de filtrado, que habitualmente es el filtro de Kalman (FK), o bien, el filtro de Chandrasekhar (FCH) (ver, entre otros, Shea, 1987; Gardner *et al.*, 1980; Terceiro, 1990; Sotoca, 1994). Puesto que la mayor parte de las especificaciones econométricas dinámicas (como por ejemplo, modelos ARMA, ARMAX, VAR, VARMA, modelos de componentes no observables) tienen una representación en EE, todo resultado válido para un modelo en EE es aplicable a cualquier representación convencional expresado en esta forma.

Se desarrolla, a partir de los trabajos de De Jong (1988) y De Jong y Chu-Chun-Lin (1994), un procedimiento de inicialización del filtro usado en la estimación por MVE de un modelo en EE con inputs estocásticos. En concreto, se obtienen expresiones exactas para inicializar un sistema estacionario o no estacionario, generalizando los resultados teóricos obtenidos por De Jong y Chu-Chun-Lin (1994) que sólo consideran la presencia de inputs deterministas. En la práctica, el uso de esta inicialización requiere conocer o analizar la estructura estocástica de los inputs presentes en el modelo. No obstante, en este trabajo también se deriva una forma aproximada para inicializar el filtrado que no exige llevar a cabo ese análisis previo. Es importante disponer de estos criterios de inicialización, al menos, por dos razones: 1) porque en Econometría los regresores de un modelo suelen ser estocásticos y 2) porque las estimaciones finales de los parámetros son sensibles a las condiciones iniciales del filtro de Kalman, sensibilidad tanto mayor cuanto menor sea el tamaño de la muestra disponible.

Se han usado distintas aproximaciones para tratar las dificultades de inicializar un filtro en la estimación de modelos en EE no estacionarios. Estas incluyen i) suponer que el estado inicial tiene una varianza "suficientemente" grande (ver Burridge y Wallis, 1985); ii) usar el filtro de información en lugar del FK (ver Kitagawa, 1981) o iii) considerar que el vector de estado inicial es fijo pero desconocido, obteniendo el estimador máximo-verosímil de dicho vector (ver Rosenberg, 1973). No obstante, y olvidando los problemas de estos criterios de inicialización, no se ha tratado explícitamente el problema de inicializar el FK aplicado a un modelo en EE con variables exógenas estocásticas. Aunque éste es un caso frecuente en la modelización econométrica de relaciones económicas, autores como De Jong (1988), De Jong y Chu-Chun-Lin (1994) o Danyang y Xuanhuang (1994), sólo consideran la existencia de efectos fijos.

El trabajo se estructura de la siguiente forma.

En el apartado 2 se introduce la notación usada en el resto y se detalla el problema de inicializar el FK para evaluar la FVE de modelos en EE estacionarios y no estacionarios. En el apartado 3 se deriva la expresión de la FVE de un modelo con inputs estocásticos, siguiendo a Marshall (1992) en la prueba del resultado. En el apartado 4 se ofrecen las expresiones exactas de la media y varianza del estado inicial de un modelo con inputs estocásticos, presentándose la prueba del resultado en el Apéndice 1. En el apartado 5 se derivan condiciones iniciales "aproximadas" para el estado de un sistema con inputs estocásticos cuando es desconocida la estructura de autocorrelación de los mismos. En el apartado 6 se presentan resultados usando datos simulados que muestran cómo las estimaciones finales de los parámetros son sensibles a las condiciones iniciales del filtro si no se tiene en cuenta el carácter estocástico de las variables explicativas. Por último, en el apartado 7 se resumen las conclusiones más importantes del trabajo.

2. Modelización en EE y la inicialización del filtrado.

Es sabido que un modelo en forma de EE puede escribirse como:

$$z_t = H^a x_t^a + D^a u_t + C^a v_t^a \tag{1}$$

$$x_{t+1}^a = \Phi^a x_t^a + \Gamma^a u_t + E^a w_t^a \tag{2}$$

en donde x_t^a es un vector $(n \times 1)$ de variables de estado, z_t un vector $(m \times 1)$ de variables observables, u_t un vector $(r \times 1)$ de variables exógenas estocásticas y w_t^a y v_t^a son perturbaciones aleatorias con esperanza nula, matrices de covarianzas Q^a y R^a, respectivamente, y una matriz de covarianzas contemporánea denotada por S^a. También es conocido que la mayor parte de las especificaciones econométricas dinámicas tienen una representación en EE en donde los elementos de H^a, D^a, C^a, Φ^a, Γ^a, E^a, Q^a, R^a y S^a

son funciones de los parámetros desconocidos que incluimos en un vector θ.

A partir de (1)-(2), el propósito es estimar el vector θ por MVE sobre la base de los datos $Z = [z_1, z_2, \ldots, z_N]$ y $U = [u_1, u_2, \ldots, u_N]$, usando el FK o el FCH. La función de densidad conjunta de la muestra puede escribirse como:

$$f(Z/U) = \prod_{t=1}^{N} f(z_t/Z^{t-1}, U) \tag{3}$$

donde Z^{t-1} es el conjunto de información disponible sobre las variables z_t hasta el instante t-1 y $f(z_t/Z^{t-1}, U)$ es la función de densidad condicionada por Z^{t-1} y U. Bajo la hipótesis de normalidad de w_t^a, v_t^a y del estado inicial x_1^a, la distribución de z_t condicionada también será normal y esto permite caracterizar analíticamente $f(z_t/Z^{t-1}, U)$. La evaluación de (3) puede llevarse a cabo utilizando el FK, cuyas condiciones iniciales se corresponden con los dos primeros momentos de $f(z_1/U)$. Es decir, tomando esperanzas condicionadas en (1) cuando $t = 1$, se obtiene que:

$$E(z_1/U) = H^a \bar{x}_1 + D^a u_1 \tag{4}$$

$$cov(z_1/U) = H^a P_1 (H^a)^T + C^a R^a (C^a)^T \tag{5}$$

donde $\bar{x}_1 = E(x_1^a/U)$ y $P_1 = E[(x_1^a - \bar{x}_1/U)(x_1^a - \bar{x}_1/U)^T]$. La utilización del FK para evaluar la FVE exige especificar los valores de $\bar{x}_1$ y P_1. Estas condiciones iniciales dependerán fundamentalmente de si el modelo (1)-(2) es estacionario o no, en el sentido de que las raíces de la matriz Φ^a sean, en módulo, mayores, iguales o menores que la unidad. Un resultado conocido es que si el modelo es estacionario y no existen variables exógenas, la inicialización adecuada del filtro es $\bar{x}_1 = 0$ y P_1 la solución a la ecuación de Lyapunov $P_1 = \Phi^a P_1 (\Phi^a)^T + E^a Q^a (E^a)^T$ (ver, por ej., Anderson y Moore, 1979, cap. 4; Terceiro, 1990, cap. 4).

Bajo estacionariedad y si las variables exógenas son efectos fijos, las expresiones exactas para la media y la varianza condicional del estado inicial del sistema fueron dadas por De Jong y Chu-Chun-Lin (1994). El problema de inicialización cuando aparecen variables exógenas estocásticas en (1)-(2) es resuelto en este trabajo, tanto en el caso estacionario como no estacionario, mostrándose que las soluciones habituales son incorrectas.

En el caso de modelos no estacionarios (es decir, todos los autovalores de la matriz Φ^a son en módulo mayores o iguales a la unidad), es natural suponer una gran incertidumbre asociada al estado inicial, es decir, $P_1 = kI$ con $k \rightarrow +\infty$, que equivale a establecer $P_1^{-1} = 0$. El problema de este tipo de inicialización es que si el FK comienza con una matriz P_1 arbitrariamente grande, puede degradarse numéricamente en pocas propagaciones. Una solución, consistente en usar el filtro de información, tampoco está exenta de problemas, dado que el uso de este filtro requiere que Φ^a tenga rango completo y esto no siempre es cierto. En concreto, en este trabajo se utiliza la formulación en EE propuesta en Terceiro (1990, cap. 2), donde Φ^a, en general, no tiene inversa.

El caso de que algunos autovalores de Φ^a sean en módulo mayores o iguales a la unidad y otros menores que la unidad, también ha sido estudiado por De Jong y Chu-Chun-Lin (1994). La idea fundamental es que la matriz de covarianzas P_1 tiende a infinito, pero P_1^{-1} es una matriz no nula y deficiente en rango. En esta situación, dichos autores derivan la expresión exacta de P_1^{-1}, aunque sólo para modelos con inputs deterministas.

Por último, diversos autores (ver, por ejemplo, Rosenberg, 1973; Terceiro, 1990), consideran que el estado inicial del sistema puede tratarse como un vector de parámetros constantes pero desconocidos, que pueden estimarse por máxima verosimilitud, tanto en modelos estacionarios como no estacionarios. Sin embargo, en el apartado 6 se muestra con datos simulados que estas condiciones iniciales generan sesgos en la estimación por MVE de modelos que incorporan componentes de medias móviles.

3. La función de verosimilitud de un modelo en EE con inputs estocásticos.

Usando la regla de Bayes, es fácil mostrar que la verosimilitud de la muestra en logaritmos puede escribirse como:

$$\ell(Z/U,\theta) = \ell(x_1^a/U) + \ell(Z/U,x_1^a) - \ell(x_1^a/Z,U) \tag{6}$$

donde el primer sumando de (6) representa la verosimilitud del estado inicial condicionada por los inputs, el segundo término es la llamada verosimilitud de la descomposición de las innovaciones y el último término del lado derecho de (6) representa la verosimilitud del estado inicial condicionada por z_t y u_t. Bajo normalidad de x_1^a, w_t^a y v_t^a, el primer sumando

de (6), sin tener en cuenta el término constante, tiene la expresión:

$$\ell(x_1^a/U) = \log|P_1| + (x_1^a - \bar{x}_1)^T P_1^{-1}(x_1^a - \bar{x}_1) \tag{7}$$

donde $\bar{x}_1$ y P_1 denotan la media y varianza condicional de x_1^a, cuyas expresiones exactas se derivan en el apartado 4. El segundo sumando de (6) es:

$$\ell(Z/U,x_1^a) = \sum_{t=1}^{N} \log|B_t| + \sum_{t=1}^{N} \tilde{z}_t^T(x_1^a) B_t^{-1} \tilde{z}_t(x_1^a) \tag{8}$$

donde $\tilde{z}_t$ es un vector de innovaciones definido como $\tilde{z}_t = z_t - E(z_t/Z^{t-1},U)$, $B_t = E(\tilde{z}_t\tilde{z}_t^T)$ y $\tilde{z}_t(x_1^a)$ denota que estas innovaciones están condicionadas por x_1^a, mientras que B_t sólo depende de P_1. La evaluación de (8) exige el cálculo de $\tilde{z}_t$ y B_t, variables que pueden obtenerse mediante la propagación del FK.

Por último, el término $\ell(x_1^a/Z,U)$ estará caracterizado por la $E(x_1^a/Z,U)$ y la $cov(x_1^a/Z,U)$, momentos que a su vez dependen de $\bar{x}_1$ y P_1. Siguiendo a Marshall (1992), es fácil derivar la expresión de esta verosimilitud. Se puede comprobar que las matrices Φ^a, Γ^a, H^a, D^a y K_t (ganancia del FK) no dependen de x_1^a y que (ver Terceiro, 1990):

$$\tilde{z}_t^c = H\bar{\Phi}_{t-1}x_1^a + \tilde{z}_t \tag{9}$$

donde $\tilde{z}_t^c$ son las innovaciones resultantes del FK fijando condiciones iniciales nulas para el estado inicial y su covarianza (a partir de ahora FK(0,0)), $\tilde{z}_t$ las innovaciones obtenidas si el proceso realmente hubiese sido generado con condiciones iniciales nulas y $\bar{\Phi}_t = (\Phi - K_tH)\bar{\Phi}_{t-1}$ comenzando con $\bar{\Phi}_1 = I$. Si denotamos por $\tilde{z}$ al vector $(m \times N) \times 1$ de innovaciones $\tilde{z}_t$, X a la matriz con bloques-fila $H\bar{\Phi}_{t-1}$ y B a una matriz diagonal por bloques que contiene las correspondientes matrices de covarianzas de cada vector de N innovaciones, es posible escribir conjuntamente el sistema:

$$\begin{bmatrix} x_1^a \\ \tilde{z}^c \end{bmatrix} = \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} \begin{bmatrix} x_1^a \\ \tilde{z} \end{bmatrix} \tag{10}$$

y la distribución conjunta de $[(x_1^a)^T \;\; (\tilde{z}^c)^T]^T$ es también normal con media y matriz de covarianzas, respectivamente:

$$\begin{bmatrix} \bar{x}_1 \\ X\bar{x}_1 \end{bmatrix} \; ; \; \Sigma = \begin{bmatrix} P_1 & P_1X^T \\ XP_1 & XP_1X^T + B \end{bmatrix} \tag{11}$$

A partir de (11) se puede obtener la esperanza condicional del estado inicial x_1^a y su matriz de covarianzas como:

$$E(x_1^a/\tilde{z}^c, U) = \bar{x}_1 + P_1 X^T [X P_1 X^T + B]^{-1} (\tilde{z}^c - X\bar{x}_1) \tag{12}$$

$$cov(x_1^a/\tilde{z}^c, U) = P_1 - P_1 X^T [X P_1 X^T + B]^{-1} X P_1 \tag{13}$$

o bien, aplicando el lema de inversión de matrices:

$$E(x_1^a/\tilde{z}^c, U) = [P_1^{-1} + W]^{-1} [P_1^{-1}\bar{x}_1 + w] \tag{14}$$

$$cov(x_1^a/\tilde{z}^c, U) = [P_1^{-1} + W]^{-1} \tag{15}$$

donde $W = X^T B^{-1} X$ y $w = X^T B^{-1} \tilde{z}^c$. Por tanto, la expresión de la función de verosimilitud exacta del modelo (1)-(2) es:

$$\begin{aligned} \ell(Z/U,\theta) = {} & \log|P_1| + \bar{x}_1^T P_1^{-1} \bar{x}_1 + \sum_{t=1}^{N} \log|B_t| + \sum_{t=1}^{N} (\tilde{z}_t^c)^T B_t^{-1} \tilde{z}_t^c + \\ & + \log|P_1^{-1} + W| - [P_1^{-1}\bar{x}_1 + w]^T [P_1^{-1} + W]^{-1} [P_1^{-1}\bar{x}_1 + w] \end{aligned} \tag{16}$$

que coincide con la ofrecida por De Jong (1988) (ver Teorema, pág. 66), aunque la prueba es mucho más inmediata y directa. Por tanto, la FVE puede evaluarse usando un FK(0,0) y después del filtrado, se corrige el efecto de esta inicialización arbitraria. Con esta estrategia se consigue modelizar las condiciones iniciales sin ninguna implicación en el filtrado. Para modelos ARMA vectoriales, es fácil demostrar que (16) coincide con la ofrecida por Mauricio (1995) en un contexto diferente a la modelización en EE.

Derivando (16) con respecto a $\bar{x}_1$, el estimador máximo-verosímil de x_1^a es $W^{-1}w$, independientemente de P_1. Eligiendo $P_1 = 0$ junto con $\bar{x}_1 = W^{-1}w$ (elección que coincide con el algoritmo de Rosenberg (1973)), (16) se convierte en:

$$\ell(Z/U,\theta) = \sum_{t=1}^{N} \log|B_t| + \sum_{t=1}^{N} (\tilde{z}_t^c)^T B_t^{-1} \tilde{z}_t^c - w^T W^{-1} w \tag{17}$$

donde el último término del lado derecho de (17) es el factor de corrección en la verosimilitud de la descomposición de las innovaciones obtenidas con un FK(0,0).

4. Media y varianza del estado inicial de un modelo en EE con inputs estocásticos.

En este apartado se derivan las expresiones exactas de la media y varianza condicional de x_1^a de un modelo en EE con inputs estocásticos. Para ello, se parte de (1)-(2), es decir,

del modelo que establece la relación entre las endógenas y las exógenas, donde sin pérdida de generalidad, se supone que $\epsilon_t = w_t^a = v_t^a$ es un vector de perturbaciones aleatorias con esperanza nula y matriz de covarianzas $\Sigma_\epsilon = Q^a = R^a = S^a$. Además, la estructura estocástica seguida por las variables exógenas puede representarse en EE como:

$$x_{t+1}^b = \Phi^b x_t^b + E^b a_t \tag{18}$$

$$u_t = H^b x_t^b + C^b a_t \tag{19}$$

donde a_t es un vector de perturbaciones con esperanza nula y matriz de covarianzas Σ_a y los elementos de Φ^b, E^b, H^b y C^b son constantes en el tiempo, aunque pueden ser desconocidos. Siguiendo a Terceiro (1990, cap.3) es posible escribir 1) la relación entre las endógenas y las exógenas (ecuaciones (1)-(2)) y 2) el modelo de las exógenas (ecuaciones (18)-(19)) conjuntamente de la forma:

$$x_{t+1} = \Phi x_t + E w_t \tag{20}$$

$$y_t = H x_t + C w_t \tag{21}$$

donde:

$$\Phi = \begin{bmatrix} \Phi^a & \Gamma^a H^b \\ 0 & \Phi^b \end{bmatrix} \; ; \; E = \begin{bmatrix} E^a & \Gamma^a C^b \\ 0 & E^b \end{bmatrix} \; ; \; H = \begin{bmatrix} H^a & D^a H^b \\ 0 & H^b \end{bmatrix} \tag{22}$$

$$C = \begin{bmatrix} C^a & D^a C^b \\ 0 & C^b \end{bmatrix} \; ; \; w_t = \begin{bmatrix} \epsilon_t \\ a_t \end{bmatrix} \; ; \; x_t = \begin{bmatrix} x_t^a \\ x_t^b \end{bmatrix} \; ; \; y_t = \begin{bmatrix} z_t \\ u_t \end{bmatrix} \tag{23}$$

Puesto que Φ es cuadrada, existe una matriz U tal que: $\Phi = U \mathrm{diag}[\Phi^N, \Phi^E] U^{-1}$, o bien:

$$\Phi = \begin{bmatrix} U^N & U^E \end{bmatrix} \begin{bmatrix} \Phi^N & 0 \\ 0 & \Phi^E \end{bmatrix} \begin{bmatrix} T^N \\ T^E \end{bmatrix} \tag{24}$$

donde Φ^N y Φ^E son formas de Jordan que contienen las raíces no estacionarias y estacionarias de Φ, respectivamente (ver Petkov *et al.*, 1991, cap. 1). A partir de (24), es fácil descomponer el modelo (20)-(21) en dos subsistemas:

$$x_{t+1}^N = \Phi^N x_t^N + T^N E w_t \tag{25}$$

$$x_{t+1}^E = \Phi^E x_t^E + T^E E w_t \tag{26}$$

donde (25) es el subsistema no estacionario y (26) el subsistema estacionario. Bajo estas condiciones, se cumple el siguiente resultado.

Teorema: Las expresiones exactas de la media y varianza condicional del estado inicial del modelo (1)-(2) vienen dadas por:

$$E(x_1^a/\tilde{u}^c) = P_{1,2}^*[P_{2,2}^*]^{-1}[S^b + P_{2,2}^{*^{-1}}]^{-1}s^b \tag{27}$$

$$cov(x_1^a/\tilde{u}^c) = P_{1,1}^* - P_{1,2}^* S^b[S^b + P_{2,2}^{*^{-1}}]^{-1}P_{2,2}^{*^{-1}}P_{2,1}^* \tag{28}$$

siendo $P_{i,j}^*$ el bloque *(i,j)* de la matriz de covarianzas P_1^* definida por:

$$P_1^* = \begin{bmatrix} kU^{N,a}(U^{N,a})^T + U^{E,a}P_1^E(U^{E,a})^T & kU^{N,a}(U^{N,b})^T + U^{E,a}P_1^E(U^{E,b})^T \\ kU^{N,b}(U^{N,a})^T + U^{E,b}P_1^E(U^{E,a})^T & kU^{N,b}(U^{N,b})^T + U^{E,b}P_1^E(U^{E,b})^T \end{bmatrix} \tag{29}$$

La matriz P_1^E de (29) es la solución a la ecuación de Lyapunov del subsistema (26) (ver Anderson *et al.*, 1979). Las matrices $U^{N,a}$, $U^{N,b}$, $U^{E,a}$, $U^{E,b}$ son particiones de U^N y U^E. Es decir, $(U^N)^T = \left[(U^{N,a})^T \ (U^{N,b})^T\right]^T$ y $(U^E)^T = \left[(U^{E,a})^T \ (U^{E,b})^T\right]^T$. Además, k es una constante arbitrariamente grande; S^b es una matriz definida por $S^b = X^T(B^b)^{-1}X$ donde X es una matriz con bloques-fila $H^b\bar{\Phi}_{t-1}^b$ y $\bar{\Phi}_t^b = (\Phi^b - K_tH^b)\bar{\Phi}_{t-1}^b$ comenzando con $\bar{\Phi}_1^b = I$ y K_t es la ganancia del FK(0,0). Denotamos por $\tilde{u}^c$ al vector $(r\times N)\times 1$ de innovaciones $\tilde{u}_t^c$ resultantes de aplicar el FK(0,0) al modelo (21)-(22). La matriz B^b es diagonal por bloques y contiene las matrices que devuelve el FK(0,0) como covarianzas de cada vector de r innovaciones, suponiendo que esa inicialización es correcta. Por último, s^b es un vector definido como $s^b = X^T(B^b)^{-1}\tilde{u}^c$.

La prueba del Teorema se ofrece en el Apéndice 1.

Las expresiones (27) y (28) son aplicables al caso estacionario y no estacionario. No obstante, es fácil derivar las expresiones de la media y varianza condicional de x_1^a en cualquiera de estas situaciones:

(i) En el caso estacionario, puesto que todas las raíces de Φ son, en módulo, menores que la unidad, se cumple que $U^N = 0$ y $U^E = I$. Por tanto, $P_1^* = P_1^E$ y (27)-(28) pasan a ser:

$$E(x_1^a/\tilde{u}^c) = P_{12}P_{22}^{-1}[S^b + P_{22}^{-1}]^{-1}s^b \tag{30}$$

$$cov(x_1^a/\tilde{u}^c) = P_{11} - P_{12}\, S^b[S^b + P_{22}^{-1}]^{-1}P_{22}^{-1}P_{21} \tag{31}$$

donde P_{11}, P_{12}, P_{21} y P_{22} son los bloques que forman la matriz P_1^E. Se observa que la media y varianza condicional de x_1^a dependen de la estructura de autocorrelación de los inputs. Si los inputs siguen un proceso de ruido blanco, entonces $P_{12} = 0, E(x_1^a/\tilde{u}^c) = 0$ y $cov(x_1^a/\tilde{u}^c) = P_{11}$ donde $P_{11} = \Phi^a P_{11}(\Phi^a)^T + E^a\Sigma_\varepsilon(E^a)^T + \Gamma^a C^b\Sigma_a(C^b)^T(\Gamma^a)^T$. El último término de P_{11} es habitualmente ignorado si se desconoce que la estructura de los inputs es de ruido blanco.

(ii) En el caso no estacionario, todas las raíces de Φ son, en módulo, mayores o iguales a la unidad y se cumple que $U^E = 0$, $U^N = I$ y $P_1^* = kI$. Por tanto, (27)-(28) se reducen a $E(x_1^a/\tilde{u}^c) = 0$ y $cov(x_1^a/\tilde{u}^c) = kI$ con $k \rightarrow +\infty$.

(iii) Si algunas raíces de Φ son, en módulo, mayores o iguales a la unidad y otras menores que la unidad, se demuestra que la $E(x_1^a/\tilde{u}^c)$ es una expresión finita y la $cov(x_1^a/\tilde{u}^c)$ arbitrariamente grande, pero su inversa es una matriz no nula y finita. Este resultado coincide con el obtenido por De Jong y Chu-Chun-Lin (1994) para modelos con inputs deterministas y la prueba se ofrece en el Apéndice 2.

5. Condiciones iniciales aproximadas del estado de un modelo con inputs estocásticos.

En este apartado se derivan condiciones iniciales "aproximadas" cuando en la práctica se desconozca el modelo seguido por las exógenas y el objetivo último sea estimar por MVE el modelo de relación. Para ello, se parte de la siguiente descomposición de (2) en su parte estacionaria y no estacionaria:

$$x_{t+1}^{a,N} = \Phi^{a,N}x_t^{a,N} + T^{a,N}\Gamma^a u_t + T^{a,N}E^a w_t^a \tag{32}$$

$$x_{t+1}^{a,E} = \Phi^{a,E}x_t^{a,E} + T^{a,E}\Gamma^a u_t + T^{a,E}E^a w_t^a \tag{33}$$

Sabiendo que existe una matriz V tal que: $\Phi^a = V\mathrm{diag}(\Phi^{a,N}, \Phi^{a,E})V^{-1}$, o bien:

$$\Phi^a = \begin{bmatrix} V^{a,N} & V^{a,E} \end{bmatrix} \begin{bmatrix} \Phi^{a,N} & 0 \\ 0 & \Phi^{a,E} \end{bmatrix} \begin{bmatrix} T^{a,N} \\ T^{a,E} \end{bmatrix} \tag{34}$$

y siguiendo a De Jong y Chu-Chun-Lin (1994), la distribución del estado inicial del sistema completo es normal con media $V^{a,N}\bar{x}_1^N + V^{a,E}\bar{x}_1^E$ y matriz de covarianzas $kV^{a,N}(V^{a,N})^T + V^{a,E}P_1^E(V^{a,E})^T$, donde P_1^E es la solución a la ecuación de Lyapunov de (33). A partir de aquí, veremos cómo inicializar el sistema analizando separadamente el caso estacionario y no estacionario.

(i) En el caso estacionario, se cumple que $V^{a,N} = \mathbf{0}$, $V^{a,E} = T^{a,E} = I$ y $\Phi^{a,E} = \Phi^a$. A su vez, la formulación en EE permite descomponer el sistema estacionario en otros dos subsistemas distintos; el que llamamos determinista:

$$x_{t+1}^{a,d} = \Phi^a x_t^{a,d} + \Gamma^a u_t \qquad z_t^d = H^a x_t^{a,d} + D^a u_t \tag{35}$$

y el estocástico:

$$x_{t+1}^{a,s} = \Phi^a x_t^{a,s} + E^a w_t^a \qquad z_t^s = H^a x_t^{a,s} + C^a v_t^a \tag{36}$$

donde es fácil comprobar que $x_{t+1}^a = x_{t+1}^{a,d} + x_{t+1}^{a,s}$ y $z_t = z_t^d + z_t^s$. Suponiendo que no existe correlación entre $x_1^{a,s}$ y los inputs y que $E(x_1^{a,s}, x_1^{a,d}) = \mathbf{0}$, puede obtenerse que:

$$E(x_1^a / U) = E(x_1^{a,d} / U) + E(x_1^{a,s}) \tag{37}$$

$$P_1 = P_1^d + P_1^s \tag{38}$$

y las condiciones iniciales del subsistema estocástico son $E(x_1^{a,s}) = \mathbf{0}$ y $P_1^s = \Phi^a P_1^s (\Phi^a)^T + E^a Q^a (E^a)^T$. En cuanto al subsistema determinista, si las variables u_t son estocásticas, la $E(x_1^{a,d}/U)$ es desconocida y puede estimarse por máxima verosimilitud. Entonces, $E(x_1^{a,d}/U) = W^{-1}w$ junto con $P_1^d = \mathbf{0}$. Por tanto, la inicialización del sistema completo viene dada por:

$$E(x_1^a / U) = W^{-1}w \tag{39}$$

$$P_1 = \Phi^a P_1^s (\Phi^a)^T + E^a Q^a (E^a)^T \tag{40}$$

La inicialización dada por (39)-(40), resultante de la descomposición (35) y (36), tiene otra justificación usando el argumento de tiempo inmemorial (ver De Jong y Chu-Chun-Lin, 1994, pág. 154; Anderson y Moore, 1979, cap. 4), que supone que el proceso se ha iniciado antes de su observación, en un instante lejano en el tiempo y por tanto, la ecuación (2)

se cumple $\forall t = 0, -1, -2, \ldots$. Es decir, las condiciones iniciales del FK se corresponden ahora con los dos primeros momentos de $f(z_1 / U_{1:N}, U_{0:-T})$ donde $U_{1:N} = [u_1, u_2, \ldots, u_N]$, $U_{0:-T} = [u_0, u_{-1}, \ldots, u_{-T}]$ y sólo observamos los inputs desde el instante $t = 1$ hasta N, pero no desde $t = 0$ hasta $t = -T$. Entonces, $E[x_1^a / U_{1:N}, U_{0:-T}] = \Phi^a E[x_0^a / U_{1:N}, U_{0:-T}] + \Gamma^a u_0$, es una expresión desconocida y distinta de cero, aunque $E(x_0^a / U_{1:N}, U_{0:-T}) = \mathbf{0}$ siguiendo el argumento de tiempo inmemorial. Sin embargo, bajo estacionariedad, se sigue cumpliendo que $P_1 = \Phi^a P_1 (\Phi^a)^T + E^a Q^a (E^a)^T$.

(ii) Si el modelo de relación es no estacionario, se tiene que $V^{a,E} = \mathbf{0}$, $V^{a,N} = T^{a,N} = I$ y $\Phi^a = \Phi^{a,N}$ y el estado inicial tiene media $\bar{x}_1^N$ y matriz de covarianzas kI con $k \to +\infty$. Nótese que la FVE dada en (16), depende de la matriz de información inicial y no de la matriz de covarianzas, salvo el primer término de la misma que es $\log|P_1|$. No tiene sentido calcular éste cuando se reduce a $\log|kI|$ con $k \to +\infty$. Sin embargo, el límite de la distancia entre la función de verosimilitud y ese término es:

$$\sum_{t=1}^{N} \log|B_t| + \sum_{t=1}^{N} (\tilde{z}_t^c)^T B_t^{-1} \tilde{z}_t^c + \log|W| - w^T W^{-1} w \tag{41}$$

donde W y w vienen definidas como en (16) y en este caso, no afecta a la evaluación de la función de verosimilitud el que la media del estado inicial sea nula o no.

(iii) La elección de condiciones iniciales si el modelo es parcialmente no estacionario exige trabajar tanto con el subsistema estacionario como con el no estacionario (ver ecuaciones (32) y (33)). La matriz de información inicial es $P_1^{-1} = \lim_{k \to +\infty} [kV^{a,N}(V^{a,N})^T + V^{a,E}P_1^E(V^{a,E})^T]^{-1}$ que es finita e igual a $P_1^{-1} = (T^{a,E})^T (P_1^E)^{-1} T^{a,E}$ donde P_1^E satisface la ecuación de Lyapunov $P_1^E = \Phi^{a,E} P_1^E (\Phi^{a,E})^T + T^{a,E} E^a Q^a E^{a^T} (T^{a,E})^T$. Esta expresión se corresponde con la ofrecida en el Teorema 3 por De Jong y Chu-Chun-Lin (1994), siendo éste uno de los resultados fundamentales del citado trabajo.

En cuanto a la media del estado inicial, si el modelo presenta inputs estocásticos habría que estimarla. No obstante, puede aplicarse el mismo procedimiento desarrollado para el caso estacionario, es decir, descomponer el subsistema estacionario (32) en su parte determinista y en su parte estocástica. El estado inicial del subsistema no estacionario tiene una media condicional distinta de cero, pero como la incertidumbre asociada al mismo es infinita, los términos de la función de verosimilitud no se ven afectados sea cual sea el valor elegido de esta media.

6. Resultados con datos simulados.

En este apartado se ilustra el funcionamiento de los criterios de inicialización propuestos, usando datos simulados. En las Tablas 1 y 2 se resumen los resultados de la estimación por MVE de modelos estacionarios que relacionan una variable endógena con una exógena, considerando distintos tamaños muestrales. Se ha generado una variable exógena con estructura AR(1) regular (Tabla 1) o AR(1) estacional (Tabla 2). Se presenta la media de las estimaciones obtenidas con 1000 realizaciones, así como las desviaciones típicas muestrales, usando tres criterios de inicialización del FK: (1) las condiciones iniciales exactas [ecuaciones (30) y (31)]; (2) la inicialización habitual que supone ignorar el carácter estocástico de la exógena, $\bar{x}_1 = \mathbf{0}$ y $P_1 = \Phi^a P_1 (\Phi^a)^T + E^a Q^a (E^a)^T$ y (3) la inicialización "aproximada" desarrollada en el apartado 4 donde el estado inicial se estima por máxima verosimilitud con $P_1 = \Phi^a P_1 (\Phi^a)^T + E^a Q^a (E^a)^T$. Además, se ofrece el error cuadrático medio (ECM) de la estimación de cada coeficiente del modelo de relación. A la vista de los resultados ofrecidos en las Tablas 1 y 2, se pueden extraer las siguientes conclusiones.

Para tamaños muestrales cortos (30 y 50 observaciones), el uso de las condiciones iniciales exactas genera menos sesgo en la estimación puntual de los parámetros del modelo de relación y la precisión de la estimación es siempre mayor que con los otros dos juegos de condiciones iniciales. A medida que aumenta el tamaño de la muestra, son menos sensibles las estimaciones finales de los parámetros a las condiciones iniciales del filtro. Se observa el buen funcionamiento de las condiciones iniciales aproximadas, independientemente del tamaño muestral, generando resultados muy similares a los obtenidos con las expresiones exactas. Por último, al incorporar estructura estacional al modelo, los efectos de una inicialización incorrecta sobre las estimaciones finales son más importantes que los resultantes de un modelo sin estacionalidad, sobre todo para muestras cortas (ver Tabla 2).

El segundo modelo simulado es un proceso univariante MA(1)xMA(1)$_4$ en el que se han considerado distintos valores de los parámetros y muestras de 100 observaciones. En este

caso, el FK se ha inicializado de dos formas: (1) $\bar{x}_1 = W^{-1}w$ y $P_1 = \mathbf{0}$ y (2) $\bar{x}_1 = \mathbf{0}$ y $P_1^{-1} = (P_1^E)^{-1}$ siendo P_1^E la solución a la ecuación de Lyapunov $P_1^E = E^a Q^a (E^a)^T$. Los resultados de la estimación se presentan en la Tabla 3 y se ofrece la media de la estimación por MVE con 500 realizaciones, la desviación típica muestral y el ECM de la estimación de cada parámetro. Los resultados muestran que, en general, el juego de condiciones iniciales másximo-verosímiles sesga al alza la estimación de los parámetros media móvil, llegando a generar una estimación no invertible si los parámetros están cerca de la no invertibilidad.

El último modelo simulado corresponde a una especificación de componentes no observables utilizada por Young (1988), García-Ferrer et al. (1996), entre otros, en la que se descompone una variable (y_t) en una tendencia (t_t) más un componente irregular (ϵ_t). Además, se supone que la primera diferencia de la tendencia es otro componente no observable (s_t) que sigue un paseo aleatorio. La formulación en EE de este modelo es:

$$\begin{bmatrix} t_{t+1} \\ s_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} t_t \\ s_t \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} a_t \tag{42}$$

$$y_t = t_t + \epsilon_t \tag{43}$$

donde las perturbaciones a_t y ϵ_t tienen esperanza nula, varianzas σ_a^2 y σ_ϵ^2, respectivamente y están incorreladas entre sí. El sistema (42)-(43) es no estacionario, ya que las dos raíces de la matriz Φ son unitarias. Una vez obtenida la estimación por MVE de los dos parámetros desconocidos (σ_a^2 y σ_ϵ^2), se puede extraer eficientemente la tendencia de la serie aplicando un *smoothing* de intervalo fijo (ver Anderson y Moore, 1979). La práctica habitual consiste en fijar *ad hoc* un ratio de varianzas $\sigma_a^2/\sigma_\epsilon^2$ (por ejemplo, igual a 0.1) y posteriormente, extraer el componente tendencial condicionado a ese valor del ratio. Los resultados de la estimación de (42)-(43) usando datos simulados se presentan en la Tabla 4. La clave para la correcta estimación por MVE de este modelo no estacionario, es la elección adecuada de condiciones iniciales del FK que son $\bar{x}_1 = \mathbf{0}$ y $P_1^{-1} = \mathbf{0}$. El otro juego de condiciones iniciales usado es el máximo verosímil, que en este caso funciona de forma parecida a las exactas ya que no aparecen parámetros de medias móviles y tampoco inputs estocásticos. En general, los resultados de la Tabla 4 muestran que la media de la estimación por MVE de los dos parámetros está muy próxima a los valores teóricos de los mismos, para los dos criterios de inicialización usados.

Tabla 1: Resultados de la estimación máximo-verosímil de un modelo con input estocástico. Valores teóricos de los parámetros $\phi = 0.8$, $\omega_0 = 0.6$, $\omega_1 = 0.3$, $\delta_1 = 0.5$ y $\sigma_a^2 = 0.1$. Número de realizaciones: 1000.

$$(1 - \phi B)u_t = \epsilon_t \quad ; \quad z_t = \frac{(\omega_0 + \omega_1 B)}{1 - \delta_1 B} u_t + a_t$$

	Condiciones iniciales exactas (1)				Condiciones iniciales habituales (2)				Condiciones iniciales aproximadas (3)			
N	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}_a^2$	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}_a^2$	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}_a^2$
30	0.6030 (0.0632) 3.9970*	0.3044 (0.0901) 8.7516*	0.4935 (0.0522) 2.7690*	0.0885 (0.0243) 0.7211*	0.6501 (0.0962) 11.763*	0.2146 (0.1526) 30.579*	0.5326 (0.0804) 7.5208*	0.1560 (0.0897) 11.188*	0.6031 (0.0636) 4.0513*	0.3059 (0.0931) 8.7044*	0.4921 (0.0522) 2.7921*	0.0855 (0.0235) 0.7607*
50	0.6004 (0.0481) 2.3155*	0.3042 (0.0714) 5.1100*	0.4967 (0.0407) 1.6672*	0.0921 (0.0184) 0.4015*	0.6310 (0.0690) 5.7209*	0.2497 (0.1081) 14.225*	0.5190 (0.0537) 3.2399*	0.1340 (0.0617) 4.9650*	0.6004 (0.0481) 2.3182*	0.3053 (0.0713) 5.1089*	0.4959 (0.0406) 1.6683*	0.0902 (0.0180) 0.4209*
100	0.6013 (0.0322) 1.0391*	0.3049 (0.0475) 2.2759*	0.4965 (0.0259) 0.6855*	0.0977 (0.0139) 0.1974*	0.6169 (0.0422) 2.0671*	0.2784 (0.0615) 4.2514*	0.5067 (0.0300) 0.9428*	0.1190 (0.0338) 1.5014*	0.6013 (0.0322) 1.0377*	0.3055 (0.0474) 2.2803*	0.4961 (0.0259) 0.6859*	0.0967 (0.0137) 0.1991*
200	0.6004 (0.0226) 0.5106*	0.3021 (0.0337) 1.1420*	0.4983 (0.0186) 0.3484*	0.0985 (0.0102) 0.1070*	0.6086 (0.0260) 0.7489*	0.2882 (0.0391) 1.6651*	0.5036 (0.0201) 0.9165*	0.1097 (0.0192) 0.4611*	0.6004 (0.0226) 0.5115*	0.3024 (0.0337) 1.4211*	0.4981 (0.0186) 0.3488*	0.0980 (0.0102) 0.1077*

N: Tamaño muestral
* Error cuadrático medio de la estimación multiplicado por 1000
() Desviación típica muestral
(1) Ecuaciones (30) y (31); (2) $\bar{x}_1 = \mathbf{0}$ y $P_1 = \Phi^a P_1 (\Phi^a)^T + E^a Q^a (E^a)^T$; (3) $\bar{x}_1 = W^{-1} w$ y $P_1 = \Phi^a P_1 (\Phi^a)^T + E^a Q^a (E^a)^T$

Tabla 2: Resultados de la estimación máximo-verosímil de un modelo estacional con input estocástico. Valores teóricos de los parámetros $\phi = 0.8$, $\omega_0 = 0.6$, $\omega_1 = 0.3$, $\delta_1 = 0.5$ y $\sigma_a^2 = 0.1$. Número de realizaciones: 1000.

$$(1 - \phi B^4)u_t = \epsilon_t \quad ; \quad z_t = \frac{(\omega_0 + \omega_1 B)}{1 - \delta_1 B^4} u_t + a_t$$

	Condiciones iniciales exactas (1)				Condiciones iniciales habituales (2)				Condiciones iniciales aproximadas (3)			
N	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}_a^2$	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}_a^2$	$\hat{\omega}_0$	$\hat{\omega}_1$	$\hat{\delta}_1$	$\hat{\sigma}_a^2$
30	0.6076 (0.0556) 3.1458*	0.3028 (0.0472) 2.2303*	0.4918 (0.0499) 2.5603*	0.0883 (0.0267) 0.8478*	0.6788 (0.0975) 15.725*	0.3335 (0.0692) 5.9167*	0.4651 (0.0799) 7.6002*	0.1830 (0.0839) 13.930*	0.6099 (0.0581) 3.4744*	0.3042 (0.0490) 2.4226*	0.4876 (0.0532) 2.9853*	0.0759 (0.0232) 1.1190*
50	0.6036 (0.0412) 1.7065*	0.2999 (0.0349) 1.2167*	0.4962 (0.0391) 1.5412*	0.0930 (0.0201) 0.4529*	0.6408 (0.0614) 5.4382*	0.3146 (0.0451) 2.2436*	0.4836 (0.0528) 3.0599*	0.1520 (0.0549) 5.7169*	0.6063 (0.0421) 1.8114*	0.3012 (0.0361) 1.3045*	0.4931 (0.0400) 1.6503*	0.0851 (0.0186) 0.5648*
100	0.5997 (0.0279) 0.7802*	0.2999 (0.0225) 0.5057*	0.4994 (0.0249) 0.6198*	0.0979 (0.0141) 0.2030*	0.6171 (0.0349) 1.5101*	0.3069 (0.0255) 0.6985*	0.4938 (0.0292) 0.8923*	0.1315 (0.0345) 2.1796*	0.6006 (0.0283) 0.8003*	0.3005 (0.0228) 0.5212*	0.4980 (0.0249) 0.6271*	0.0939 (0.0136) 0.2221*
200	0.6002 (0.0193) 0.3721*	0.3003 (0.0158) 0.2506*	0.4992 (0.0176) 0.3099*	0.0982 (0.0099) 0.1020*	0.6088 (0.0223) 0.5730*	0.3037 (0.0171) 0.3067*	0.4964 (0.0193) 0.3864*	0.1146 (0.0180) 0.5362*	0.6008 (0.0194) 0.3752*	0.3006 (0.0160) 0.2564*	0.4985 (0.0177) 0.3149*	0.0962 (0.0098) 0.1098*

N: Tamaño muestral
* Error cuadrático medio de la estimación multiplicado por 1000
() Desviación típica muestral
(1) Ecuaciones (30) y (31); (2) $\bar{x}_1 = \mathbf{0}$ y $P_1 = \Phi^a P_1 (\Phi^a)^T + E^a Q^a (E^a)^T$; (3) $\bar{x}_1 = W^{-1} w$ y $P_1 = \Phi^a P_1 (\Phi^a)^T + E^a Q^a (E^a)^T$

Tabla 3: Resultados de la estimación máximo-verosímil del modelo $z_t = (1 - \theta B)(1 - \Theta B^s)a_t$ usando 100 observaciones. Número de realizaciones: 500.

Valores teóricos de los parámetros	Condiciones iniciales $\bar{x}_1 = W^{-1}w$; $P_1 = 0$			Condiciones iniciales $\bar{x}_1 = 0$; $P_1 = E^a Q^a (E^a)^T$		
	$\hat{\theta}$	$\hat{\Theta}$	$\hat{\sigma}_a^2$	$\hat{\theta}$	$\hat{\Theta}$	$\hat{\sigma}_a^2$
$\theta = 0.75$	0.7556	0.7667	0.9170	0.7551	0.7614	0.9700
$\Theta = 0.75$	(0.0859)	(0.0986)	(0.1383)	(0.0823)	(0.0936)	(0.1413)
$\sigma_a^2 = 1.0$	2274.3*	2309.9*	26.016*	2272.1*	2292.9*	20.852*
$\theta = 0.90$	1.0461	1.3325	0.8651	0.9174	0.9143	0.9547
$\Theta = 0.90$	(0.2094)	(0.2899)	(0.1857)	(0.0630)	(0.0748)	(0.1385)
$\sigma_a^2 = 1.0$	65.194*	271.10*	52.683*	4.2718*	5.7990*	21.234*
$\theta = 0.95$	1.1326	1.2576	0.9299	0.9573	0.9531	0.9719
$\Theta = 0.95$	(0.1577)	(0.2196)	(0.1490)	(0.0522)	(0.0663)	(0.1295)
$\sigma_a^2 = 1.0$	58.212*	142.84*	27.115*	2.7781*	4.4053*	17.560*

* Error cuadrático medio de la estimación multiplicado por 1000
() Desviación típica muestral

Tabla 4: Resultados de la estimación máximo-verosímil de un modelo de componentes no observables. Valores teóricos de los parámetros $\sigma_a^2 = 1.0$ y $\sigma_\epsilon^2 = 10$. Número de realizaciones: 500.

$$\begin{bmatrix} t_{t+1} \\ s_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} t_t \\ s_t \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} a_t$$

$$y_t = t_t + \epsilon_t$$

	Condiciones iniciales $\bar{x}_1 = W^{-1}w$; $P_1 = 0$		Condiciones iniciales $\bar{x}_1 = 0$; $P_1^{-1} = 0$	
N	$\hat{\sigma}_a^2$	$\hat{\sigma}_\epsilon^2$	$\hat{\sigma}_a^2$	$\hat{\sigma}_\epsilon^2$
50	0.9623 (0.5771)	9.5648 (2.4254)	1.0663 (0.6071)	9.7685 (2.4237)
	0.3345*	6.0720*	0.3730*	5.9279*
100	0.9571 (0.3812)	10.001 (1.7176)	1.0068 (0.3926)	10.130 (1.7270)
	0.1472*	2.9502*	0.1542*	2.9993*

N Tamaño muestral
* Error cuadrático medio de la estimación
() Desviación típica muestral

7. Conclusiones.

En este trabajo se generalizan los resultados teóricos obtenidos en De Jong y Chu-Chun-Lin (1994), derivándose la inicialización exacta del estado de un modelo que incorpora inputs estocásticos. Dado que las estimaciones finales de los parámetros son sensibles a las condiciones iniciales del filtro, es importante tener en cuenta el carácter estocástico de los inputs, situación frecuente en Econometría, al calcular la media y la varianza condicional del estado inicial. Estas expresiones dependen básicamente de dos aspectos: 1) la estructura de autocorrelación de los inputs que se suponga y 2) el carácter estacionario o no estacionario del modelo de relación entre las endógenas y los inputs, cuestión también tratada en De Jong y Chu-Chun-Lin (1994), si bien estos autores sólo suponen la existencia de efectos fijos.

Además, en el caso de que el objetivo último del análisis sea estimar los parámetros del modelo de relación y no se conozca la estructura estocástica de los inputs, se deriva un juego de condiciones iniciales "aproximadas", que tienen en cuenta la naturaleza aleatoria de dichas variables. La idea clave para derivar estas condiciones iniciales consiste precisamente en la flexibilidad de la representación en espacio de los estados, que permite descomponer fácilmente el sistema en lo que llamamos subsistema determinista y estocástico. Una inicialización adecuada de cada uno de estos dos subsistemas proporciona las condiciones iniciales apropiadas para el sistema completo.

Los métodos de inicialización propuestos han sido aplicados a la estimación de distintos modelos usando datos simulados y distintos tamaños muestrales. Los resultados obtenidos muestran la importancia de la inicialización del filtrado en la estimación final de los parámetros de modelos muy sencillos, tanto estacionarios como no estacionarios. En concreto, los resultados sugieren que: i) ignorar el carácter estocástico de los inputs genera estimaciones menos eficientes (en términos del error cuadrático medio) de los parámetros del modelo de relación; ii) criterios de inicialización clásicos, como el algoritmo de Rosenberg (1973) o el de Terceiro (1990), sesgan al alza la estimación final de parámetros de medias móviles, siendo estas estructuras frecuentes en la modelización univariante de series económicas y iii) la correcta estimación por máxima verosimilitud exacta de modelos no estacionarios, requiere elegir adecuadamente las condiciones iniciales del filtro de Kalman.

8. Referencias.

Anderson, B.D.O. y Moore , J.B. (1979). *Optimal Filtering*. Englewood Cliffs. NJ: Prentice Hall.

Burridge, P. y Wallis, K.F. (1985). Calculating the variance of seasonally adjusted series. *Journal of the American Statistical Association*, 80, 541-552.

Danyang, L. y Xuanhuang, L. (1994). Optimal State Estimation without the Requirement of a Priori Statistics Information of the Initial state. *IEEE Transactions on Automatic Control*, 39, 10, 2087-2091.

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.

García-Ferrer, A., del Hoyo, J., Novales, A. y Young, P.C. (1996). Recursive Identification, Estimation and Forecasting of Nonstationary Economic Time Series with Applications to GNP International Data, en *Bayesian Analysis in Statistics and Econometrics*, Essays in Honor of Arnold Zellner, (eds., D.A. Berry, K.M. Chaloner and J.K. Geweke), John Wiley and Sons, Inc.

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.

Kitagawa, G. (1981). A Nonstationary Time Series Model and its Fitting by Recursive Filter. *Journal of Time Series Analysis*, 2, 103-116.

Marshall, P. (1992). State Space Models with Diffuse Initial Conditions. *Journal of Time Series Analysis*, 13, 5, 411-414.

Mauricio, J.A. (1995). Exact Maximum Likelihodd Estimation of Stationary Vector ARMA Models. *Journal of the American Statistical Association*, 90, 429, 282-291.

Petkov, P.Hr., Christov, N.D. y Konstantinov, M.M. (1991). *Computational Methods for Linear Control Systems*. Prentice-Hall, Englewood Cliffs, N.J.

Rosenberg, B. (1973). The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression. *Annals of Economic and Social Measurement*, 2, 4, 399-428.

Shea, B.L. (1987). Estimation of Multivariate Time Series. *Journal of Time Series Analysis*, 8, 95-109.

Sotoca, S. (1994). Aplicación del Filtro de Chandrasekhar a la Estimación por Máxima Verosimilitud Exacta de Modelos Dinámicos. *Estadística Española*, 36, 136, 259-285.

Terceiro, J. (1990). *Estimation of Dynamic Econometric Models with Errors in Variables*. Springer-Verlag, Heidelberg.

Apéndice 1. A partir de (25) y (26), De Jong y Chu-Chun-Lin (1994) demuestran que x_1^E y x_1^N siguen una distribución conjunta:

$$x_1^* = \begin{bmatrix} x_1^N \\ x_1^E \end{bmatrix} \sim N\left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} kI & 0 \\ 0 & P_1^E \end{bmatrix} \right\} \tag{1.1}$$

y por tanto, el estado inicial del sistema completo $Ux_1^* \sim N(0, P_1^*)$, donde $P_1^* = kU^N(U^N)^T + U^E P_1^E (U^E)^T$. Sabiendo que U^N y U^E se han particionado en $(U^N)^T = \left[(U^{N,a})^T \;\; (U^{N,b})^T\right]^T$ y $(U^E)^T = \left[(U^{E,a})^T \;\; (U^{E,b})^T\right]^T$, P_1^* puede escribirse como en (29). A partir de (21), es fácil demostrar que (ver Terceiro, 1990):

$$\tilde{u}_t^c = \begin{bmatrix} 0 & X \end{bmatrix} \begin{bmatrix} x_1^a \\ x_1^b \end{bmatrix} + \tilde{u}_t \tag{1.2}$$

siendo $\tilde{u}_t^c$ las innovaciones resultantes de un FK(0,0), $\tilde{u}_t$ las innovaciones que se obtendrían si el proceso hubiese sido generado con un estado inicial y matriz de covarianzas nulos. Además, X es una matriz con bloques-fila $H^b \bar{\Phi}_{t-1}^b$ y $\bar{\Phi}_t^b = (\Phi^b - K_t H^b)\bar{\Phi}_{t-1}^b$ con $\bar{\Phi}_1^b = I$ y K_t es la ganancia del FK(0,0). Si denotamos por $\tilde{u}$ al vector $(r \times N) \times 1$ de innovaciones $\tilde{u}_t$ y B^b a una matriz diagonal por bloques que contiene las matrices de covarianzas de cada vector de estas innovaciones, es posible escribir:

$$\begin{bmatrix} x_1^* \\ \tilde{u}^c \end{bmatrix} = \begin{bmatrix} I & 0 \\ X^* & I \end{bmatrix} \begin{bmatrix} x_1^* \\ \tilde{u} \end{bmatrix} \tag{1.3}$$

siendo $X^* = \begin{bmatrix} 0 & X \end{bmatrix}$. Por tanto, la distribución de $\left[(x_1^*)^T \;\; (\tilde{u}^c)^T\right]^T$ es normal con media nula y matriz de covarianzas:

$$\Sigma = \begin{bmatrix} P_1^* & (P_1^*)^T (X^*)^T \\ X^* P_1^* & X^* P_1^* (X^*)^T + B^b \end{bmatrix} \tag{1.4}$$

y las expresiones de la media y varianza condicional de x_1^* son:

$$E[x_1^* / \tilde{u}^c] = (P_1^*)^T (X^*)^T [B^b + X^* P_1^* (X^*)^T]^{-1} \tilde{u}^c \tag{1.5}$$

$$cov[x_1^* / \tilde{u}^c] = P_1^* - P_1^* (X^*)^T [B^b + X^* P_1^* (X^*)^T]^{-1} X^* P_1^* \tag{1.6}$$

En el caso de que sólo se conozca la relación entre las variables endógenas y las exógenas, la esperanza y varianza condicionales relevantes son la $E(x_1^a / \tilde{u}^c)$ y $cov(x_1^a / \tilde{u}^c)$. A partir de (1.5) y (1.6) sus expresiones exactas son:

$$E[x_1^a / \tilde{u}^c] = P_{1,2}^* X^T [B^b + X P_{2,2}^* X^T]^{-1} \tilde{u}^c \tag{1.7}$$

$$cov[x_1^a / \tilde{u}^c] = P_{1,1}^* - P_{1,2}^* X^T [B^b + X P_{2,2}^* X^T]^{-1} X P_{2,1}^* \tag{1.8}$$

y aplicando el lema de inversión de matrices a (1.7)-(1.8), se obtienen (27)-(28), respectivamente. ■

Apéndice 2. Para demostrar que la esperanza y la inversa de la matriz de covarianzas del estado inicial de (20)-(21) son expresiones finitas, es necesario conocer la forma concreta de la matriz U que se aplica a la matriz de transición Φ (ver (24)).

En este apéndice se demuestra que puede obtenerse U como $U = VSI^*W$ donde $\Phi = VSI^*W \mathrm{diag}[\Phi^N, \Phi^E](VSI^*W)^{-1}$. Φ^N sólo contiene autovalores mayores o iguales a la unidad de Φ y Φ^E los autovalores menores a la unidad, siendo posible que Φ^a y Φ^b compartan autovalores comunes. La estructura de las matrices que definen U se ofrece a continuación. Usando la factorización (24) para Φ^a y Φ^b:

$$\Phi^a = \begin{bmatrix} V^{N,a} & V^{E,a} \end{bmatrix} \begin{bmatrix} \Phi^{N,a} & 0 \\ 0 & \Phi^{E,a} \end{bmatrix} \begin{bmatrix} V^{N,a} & V^{E,a} \end{bmatrix}^{-1} ; \; \Phi^b = \begin{bmatrix} V^{N,b} & V^{E,b} \end{bmatrix} \begin{bmatrix} \Phi^{N,b} & 0 \\ 0 & \Phi^{E,b} \end{bmatrix} \begin{bmatrix} V^{N,b} & V^{E,b} \end{bmatrix}^{-1} \tag{2.1}$$

se puede construir V de forma que $\Phi = V\Phi^* V^{-1}$, siendo:

$$V = \begin{bmatrix} V^{N,a} & V^{E,a} & 0 & 0 \\ 0 & 0 & V^{N,b} & V^{E,b} \end{bmatrix} ; \; \Phi^* = \begin{bmatrix} \Phi^{N,a} & 0 & \times & \times \\ 0 & \Phi^{E,a} & \times & \times \\ 0 & 0 & \Phi^{N,b} & 0 \\ 0 & 0 & 0 & \Phi^{E,b} \end{bmatrix} \tag{2.2}$$

donde el símbolo $(\times)$ en (2.2) denota elementos no nulos y Φ^* puede diagonalizarse por bloques mediante dos transformaciones: 1) Haciendo $\Phi^* = S\Phi^{**}S^{-1}$ donde la estructura de S y de Φ^{**} es:

$$S = \begin{bmatrix} I & 0 & 0 & S_1 \\ 0 & I & S_2 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \end{bmatrix} ; \; \Phi^{**} = \begin{bmatrix} \Phi^{N,a} & 0 & \times & 0 \\ 0 & \Phi^{E,a} & 0 & \times \\ 0 & 0 & \Phi^{N,b} & 0 \\ 0 & 0 & 0 & \Phi^{E,b} \end{bmatrix} \tag{2.3}$$

2) Reordenando los términos de Φ^{**} mediante la transformación $\Phi^{**} = I^*\Phi^{***}(I^*)^T$:

$$I^* = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & 0 & I \end{bmatrix} ; \; \Phi^{***} = \begin{bmatrix} \Phi^{N,a} & \times & 0 & 0 \\ 0 & \Phi^{N,b} & 0 & 0 \\ 0 & 0 & \Phi^{E,a} & \times \\ 0 & 0 & 0 & \Phi^{E,b} \end{bmatrix} \tag{2.4}$$

Nótese que existe una solución única para S_1 y S_2 ya que los bloques $[\Phi^{N,a}\,\Phi^{E,b}]$ y $[\Phi^{N,b}\,\Phi^{E,a}]$ no contienen autovalores comunes. Por último, se obtiene la descomposición de Jordan de cada bloque de la matriz Φ^{***}:

$$\Phi^{***} = \begin{bmatrix} W_{11} & 0 \\ W_{12} & 0 \\ 0 & W_{21} \\ 0 & W_{22} \end{bmatrix} \begin{bmatrix} \Phi^N & 0 \\ 0 & \Phi^E \end{bmatrix} \begin{bmatrix} W_{11} & 0 \\ W_{12} & 0 \\ 0 & W_{21} \\ 0 & W_{22} \end{bmatrix}^{-1} = W \mathrm{diag}[\Phi^N, \Phi^E] W^{-1} \tag{2.5}$$

donde Φ^N sólo recoge autovalores mayores o iguales a la unidad de Φ y Φ^E los menores que la unidad, siendo ambas matrices formas de Jordan. Usando (1.1) es fácil mostrar que los bloques (1,2) y (2,2) de la matriz de covarianzas del estado inicial de (20)-(21) tienen la forma:

$$\begin{aligned} P_{1,2}^* &= k V^{N,a} W_{1,1} W_{1,2}^T (V^{N,b})^T + k V^{E,a} S_2 W_{1,2} W_{1,2}^T (V^{N,b})^T + \\ &+ V^{N,a} S_1 W_{2,2} P_1^E W_{2,2}^T (V^{E,b})^T + V^{E,a} W_{2,1} P_1^E W_{2,2}^T (V^{E,b})^T \end{aligned} \tag{2.6}$$

$$P_{2,2}^* = k V^{N,b} W_{1,2} W_{1,2}^T (V^{N,b})^T + V^{E,b} W_{2,2} P_1^E W_{2,2}^T (V^{E,b})^T \tag{2.7}$$

y aplicando el Teorema 3 de De Jong y Chu-Chun-Lin (1994, pp. 155), $P_{2,2}^{*^{-1}}$ es una matriz finita y ortogonal a $V^{N,b}$. Por tanto, $P_{1,2}^* P_{2,2}^{*^{-1}}$ y la $E(x_1^a / \tilde{u}^c)$ son expresiones finitas.

Con respecto a la $cov(x_1^a / \tilde{u}^c)$, que puede tender a infinito, es fácil demostrar que $[cov(x_1^a / \tilde{u}^c)]^{-1} = V_{1,1} - V_{1,2}(V_{2,2} + S^b)^{-1} V_{2,1}$ es una matriz finita y no nula, donde:

$$\begin{bmatrix} V_{1,1} & V_{1,2} \\ V_{2,1} & V_{2,2} \end{bmatrix} = \begin{bmatrix} P_{1,1}^* & P_{1,2}^* \\ P_{2,1}^* & P_{2,2}^* \end{bmatrix}^{-1}$$

■