



Departamento de Economía Cuantitativa
Facultad de CC. Económicas y Empresariales
Universidad Complutense de Madrid

Tesis Doctoral

**APLICACIONES ECONOMETRICAS DEL FILTRO DE KALMAN Y ALGUNAS
VARIACIONES NUMERICAS: EL FILTRO DE CHANDRASEKHAR**

Autor: Sonia Sotoca López
Director: Jaime Terceiro Lomba

A mis padres

Agradecimientos

Deseo expresar, en primer lugar, mi más profundo agradecimiento al director de la tesis, el profesor Jaime Terceiro. Su continua labor de dirección e investigación, ha sido imprescindible para que el trabajo llegara a buen fin.

Mi más sincero agradecimiento a Ramón Cancio y especialmente, a Miguel Jerez, por sus valiosos consejos y ayuda de todo tipo.

Fuera del terreno profesional, quiero agradecer el apoyo moral que, en todo momento, he recibido de mis padres y de mis hermanos.

A todos ellos, y a todos los compañeros de Departamento que han mostrado algún interés por este trabajo, muchas gracias.

INDICE

Introducción	2
<u>CAPITULO 1:</u> Teoría de filtros	7
1.1. Modelos en forma de espacio de los estados	7
1.2. El filtro de Kalman en tiempo discreto	12
1.3. Filtros derivados de la factorización de la matriz P	15
1.3.1. Problemas numéricos del filtro de Kalman ..	15
1.3.2. Algoritmos de la <u>raíz cuadrada</u> de $P_{t/t}$	18
1.3.3. Algoritmos de factorización de los incre- mentos de $P_{t+1/t}$	25
1.4. El filtro de Chandrasekhar en tiempo discreto .	26
1.4.1. Derivación de las ecuaciones de Chandra- sekhar	26
1.4.2. Propiedades del filtro de Chandrasekhar ...	37
1.5. Comparación de la factorización de Chandrasekhar con otras factorizaciones alternativas	38
<u>CAPITULO 2:</u> Estimación máximo-verosímil de modelos economé- tricos en forma de espacio de los estados me- diante los filtros de Chandrasekhar	41
2.1. Modelos econométricos en forma de espacio de los estados	41
2.2. Estimación máximo-verosímil de un modelo eco- nómico lineal en espacio de los estados mediante un filtro de Chandrasekhar	44
2.2.1. La función de verosimilitud del modelo	44

2.2.2. Planteamiento del problema de optimización	46
2.2.3. Expresión analítica exacta del gradiente de la función de verosimilitud	49
2.2.4. Expresión analítica exacta del hessiano de la función de verosimilitud	54
2.2.5. Expresión analítica exacta de la matriz de información	69
2.2.6. Propiedades asintóticas del estimador máximo-verosímil	73

CAPITULO 3: Selección de algoritmos y estructura del software 76

3.1. Estructura general de un algoritmo de optimización	76
3.2. Criterios de convergencia	78
3.3. Criterios para elegir la longitud del paso	79
3.3.1. Elección de la longitud de paso óptima	79
3.3.2. Elección de la longitud de paso por <u>backtracking</u>	80
3.4. Direcciones de búsqueda del óptimo	81
3.4.1. Dirección de Cauchy	81
3.4.2. Dirección de Newton	83
3.5. Método de Newton modificado para asegurar el buen condicionamiento de H_k	84
3.6. Descripción de los algoritmos de optimización utilizados	86
3.6.1. Método de <u>line-search</u>	87
3.6.2. Método de <u>model trust-region</u>	91
3.7. Criterios de convergencia utilizados	95
3.8. Escalamiento de variables	97

3.9. Aproximación del gradiente y del hessiano por diferencias finitas	98
3.10. Uso del paquete UNCMIN	100
3.10.1. Elección del método de optimización	100
3.10.2. Cálculo del gradiente y del hessiano	101
3.10.3. Entrada en el programa	102
3.10.4. Control de los resultados	103
3.10.5. Organización del programa de estimación de modelos econométricos en forma de espacio de los estados mediante un filtro de Chandrasekhar	103
<u>CAPITULO 4: Aplicaciones</u>	108
4.1. Modelos simulados	110
4.2. Modelos del Indice Dow Jones	129
4.2.1. Modelo AR(1) para el Indice Dow-Jones	131
4.2.2. Modelo MA(1) para el Indice Dow-Jones	139
4.3. Modelos del Desempleo Femenino del Reino Unido	150
4.4. Modelos de relación entre el Consumo energético y el PIB de Dinamarca	158
4.5. Modelos multivariantes	165
<u>CAPITULO 5: Conclusiones</u>	179
<u>APENDICE A: El filtro de Kalman en tiempo discreto</u>	183
A.1. Derivación del filtro de Kalman	184
A.2. Propiedades del filtro de Kalman	189
A.3. Interpretación de los resultados del filtro de Kalman para modelos gaussianos	191
A.4. Inicialización del filtro de Kalman	192

<u>APENDICE B:</u> Estimación de las condiciones iniciales óptimas del vector de estado	195
B.1. Estimador máximo-verosímil del vector de estado inicial	196
<u>APENDICE C:</u> Ecuaciones de Riccati	200
C.1. Métodos de cálculo directos	202
C.2. Métodos de cálculo iterativos	204
<u>APENDICE D:</u> Datos utilizados	206
Bibliografía	210
Indice de autores	221
Indice de palabras	223

Símbolos empleados (por orden de aparición).

\mathbf{x}_t : vector de estado ($n \times 1$) del sistema en el instante t .

\mathbf{w}_t : vector ($p \times 1$) de perturbaciones de la dinámica del sistema en el instante t .

\mathbf{u}_t : vector ($r \times 1$) de variables de control.

\mathbf{z}_t : vector ($m \times 1$) de variables observadas.

\mathbf{v}_t : vector ($c \times 1$) de errores de observación.

$\bar{\mathbf{x}}_t, \Gamma_t, \mathbf{E}_t, \mathbf{H}_t, \mathbf{D}_t$ y \mathbf{C}_t : matrices características del sistema.

n : dimensión del vector de estado.

m : dimensión del vector de observación.

r : dimensión del vector de variables de control.

p : dimensión del vector de perturbaciones de la dinámica del sistema.

c : dimensión del vector de errores de observación.

\mathbf{x}_0 : vector de estado inicial.

$E(\mathbf{x})$: esperanza matemática de la variable \mathbf{x} .

$E(\mathbf{x}/\mathbf{y})$: esperanza matemática de \mathbf{x} condicionada por \mathbf{y} .

$\text{COV}(\mathbf{x})$: matriz de covarianzas de \mathbf{x} .

$\text{COV}(\mathbf{x}/\mathbf{y})$: matriz de covarianzas de \mathbf{x} condicionada por \mathbf{y} .

δ_{ts} : valor delta de Kronecker.

$$\bar{\mathbf{x}}_0 = E(\mathbf{x}_0)$$

$$\mathbf{P}_0 = \text{COV}(\mathbf{x}_0)$$

\mathbf{I} : matriz identidad.

\mathbf{z}^{t-1} : conjunto de información disponible hasta el instante $t-1$.

$$\mathbf{P}_{t-1/t-1} = \text{COV}(\mathbf{x}_{t-1} / \mathbf{z}^{t-1})$$

$$\mathbf{P}_{t/t-1} = \text{COV}(\mathbf{x}_t / \mathbf{z}^{t-1})$$

$$\hat{\mathbf{z}}_{t/t-1} = E(\mathbf{z}_t / \mathbf{z}^{t-1})$$

$\tilde{\mathbf{z}}_t = \mathbf{z}_t - \hat{\mathbf{z}}_{t/t-1}$: proceso de innovaciones.

$$\mathbf{B}_t = \text{COV}(\mathbf{z}_t / \mathbf{z}^{t-1})$$

$$P_{t/t} = \text{COV}(\mathbf{x}_t / \mathbf{z}_t)$$

K_t : matriz ganancia del filtro de Kalman.

$$\bar{\mathbf{x}}_t = \mathbf{x} - K_t H.$$

$$\bar{\Gamma}_t = \Gamma - K_t D.$$

$\text{tr}(\mathbf{A})$: traza de la matriz \mathbf{A} .

$\text{det}(\mathbf{A})$: determinante de la matriz \mathbf{A} .

$\text{diag} \{a_{11}, a_{22}, \dots, a_{nn}\}$: matriz diagonal con elementos a_{ii} ,

$\forall i = 1, \dots, n$

\bar{K}_t : matriz ganancia normalizada.

$V_{t/t-1} = P_{t+1/t} - P_{t/t-1}$: matriz de incrementos de $P_{t/t-1}$.

$\mathbf{A} > \mathbf{0}$: matriz \mathbf{A} definida positiva.

$\mathbf{A} < \mathbf{0}$: matriz \mathbf{A} definida negativa.

$\text{rang}(\mathbf{A})$: rango de la matriz \mathbf{A} .

\rightarrow : "tiende a".

$\min(a, b, \dots, z)$: el mínimo de los valores a, b, \dots, z .

$\max(a, b, \dots, z)$: el máximo de los valores a, b, \dots, z .

$$\hat{\mathbf{a}}_{t/t-1} = P_{t/t-1}^{-1} \hat{\mathbf{x}}_{t/t-1}$$

$$\hat{\mathbf{a}}_{t/t} = P_{t/t-1}^{-1} \hat{\mathbf{x}}_{t/t}$$

$L(\theta)$: función de verosimilitud del modelo.

$p(z_t / \mathbf{z}^{t-1}, \theta)$: distribución de z_t condicionada al conjunto de información \mathbf{z}^{t-1} .

$N(\mu, \Sigma)$: distribución normal de media μ y matriz de covarianzas Σ .

$\ln x$: logaritmo neperiano de x .

$$\ell(\theta) = -\ln L(\theta)$$

\sim : "se distribuye como".

MIN: minimizar.

G_{ij} : término general del hessiano de la función de verosimilitud.

$\frac{\partial \mathbf{A}}{\partial x}$: derivada parcial de la matriz \mathbf{A} con respecto a x .

$M(\theta)$: matriz de información del modelo.
 \mathbf{p}_k : dirección de búsqueda del óptimo seguida en la iteración k .
 \mathbf{g}_k : vector gradiente de un campo escalar.
 \mathbf{G}_k : hessiano de un campo escalar.
 α_k : longitud de paso tomada en la iteración k .
 θ_* : vector de estimaciones óptimas del vector de parámetros θ .
 $\|\mathbf{x}\|$: norma del vector \mathbf{x} .
 $|x|$: valor absoluto de x .
 $O(h)$: error cometido de orden de magnitud h .
 θ : vector de parámetros del modelo.
 θ^0 : preestimación del vector θ .
 θ^t : valor teórico del vector θ .
 B : operador retardo.
 $\nabla = 1 - B$: operador diferencia.
 $\nabla_s = 1 - B^s$: operador diferencia estacional de período estacional s .
 a_t : valor (no observable) de un proceso ruido blanco en t .
 σ_a^2 : varianza de la distribución de a_t .
 DW_t : valor del Índice Dow-Jones en t .
 acf : función de autocorrelación simple.
 pacf : función de autocorrelación parcial.
 DT : desviación típica.
 DF_t : valor de la serie de Desempleo Femenino en t .
 CE_t : valor de la serie de Consumo Energético en t .
 PIB_t : valor de la serie Producto Interior Bruto en t .
 g : ganancia en estado estacionario de una función de transferencia.
 ML : test del multiplicador de Lagrange.
 $M.C.O$: Mínimos Cuadrados Ordinarios.
 EAR : Ecuación Algebraica de Riccati.

q_t : Cantidad de patata producida en miles de Tm.

p_t : Precios de la patata percibidos por los agricultores en
Pts/Kg.

INTRODUCCION

En este trabajo se presenta un algoritmo de estimación de modelos econométricos en forma de espacio de los estados basado en un filtro de Chandrasekhar. Este filtro es superior al filtro de Kalman por su estabilidad numérica. Además, resulta especialmente ventajoso, desde el punto de vista computacional, cuando es aplicado a los casos más usuales en Econometría. El algoritmo de estimación se ha completado aplicando un método para estimar de forma óptima las condiciones iniciales del filtro. Esto evita la arbitrariedad habitual en estos casos, así como los posibles problemas numéricos derivados de una mala elección de condiciones iniciales.

El procedimiento de estimación utilizado, tanto para los parámetros del modelo como para las condiciones iniciales del filtro, es el de máxima verosimilitud. Para ello, se ha desarrollado la expresión analítica de la función de verosimilitud, por lo que puede hablarse de máxima verosimilitud exacta. También se han derivado las expresiones analíticas exactas del gradiente y del hessiano de la función de verosimilitud, así como de la matriz de información, a partir de la cual se obtiene la matriz de covarianzas de los parámetros.

El problema de estimación por máxima verosimilitud de modelos econométricos en forma de espacio de los estados no es nuevo. Este tema ha sido tratado por Gardner et al. (1980), Cancio (1989) y Terceiro (1990), utilizando el filtro de Kalman. En este último trabajo se desarrolla un mecanismo para el tratamiento (formulación y estimación) de modelos econométricos con errores en las variables endógenas y exógenas.

Desde un punto de vista teórico, la aplicación del filtro de Kalman en la estimación máximo-verosímil de modelos en espacio de los estados es perfecta. Sin embargo, es bien sabido por la experiencia práctica [vid., entre otros, Schlee et al. (1967), Bellantoni y Dodge (1967), Leondes (1970), Fitzgerald (1971) y Thornton y Bierman (1980)], que el filtro de Kalman es muy sensible a errores de redondeo. Esta falta de robustez numérica

da lugar, entre otras cosas, a la obtención de matrices de covarianzas no definidas positivas, lo cual invalida sus resultados.

La literatura de sistemas ha ofrecido diversas soluciones a este problema inherente al filtro de Kalman. Todas ellas se basan en factorizar las matrices de covarianzas mediante transformaciones numéricamente estables. Estas factorizaciones permiten eliminar la ecuación de Riccati cuya propagación genera los problemas numéricos. Así surgió, como primera solución el algoritmo de Potter (1963), aplicable a sistemas en tiempo discreto. Andrews (1968) extendió el trabajo de Potter para tiempo continuo, eliminando además algunas restricciones del mismo. La generalización de Andrews se aplicó a problemas en tiempo discreto por Leondes (1970) y Kaminski et al (1971). Todos estos algoritmos tienen unos requerimientos de cálculo superiores a los del filtro de Kalman. Esta ineficiencia fue resuelta por la factorización UDU^T de Bierman (1974), que es una instrumentación estable de la factorización de Agee y Turner (1972). Por último, se ha desarrollado una nueva familia de algoritmos debidos a Kailath (1973) y Morf et al (1974). Estos filtros están basados en la idea de factorizar los incrementos de las matrices de covarianzas, en lugar de las propias matrices. Esta nueva aproximación culmina con las llamadas ecuaciones de Chandrasekhar, sólo aplicables a sistemas con matrices estructurales constantes en el tiempo, aunque se han extendido al caso en que las matrices de covarianzas del ruido del sistema varían con el tiempo. Recientemente, Verhaegen y Van Dooren (1986, 1988) han estudiado los aspectos computacionales y de robustez numérica de todos estos algoritmos, mostrando las ventajas de la aproximación de Chandrasekhar. Una aplicación de este algoritmo para la estimación de algunos modelos simulados y una especificación bivariante con datos físicos puede verse en Shea (1987, 1989).

En este trabajo, se ha aplicado el filtro de Chandrasekhar a la estimación de modelos econométricos lineales y dinámicos, tanto con datos simulados como con datos económicos reales. Las

especificaciones consideradas presentan distintas estructuras dinámicas así como distintas situaciones de condicionamiento. Por otra parte, los algoritmos de Chandrasekhar son especialmente ventajosos, desde un punto de vista computacional, cuando los modelos presentan alguna estructura estacional, siendo éste un caso frecuente en la práctica econométrica con variables económicas. Los resultados muestran que este criterio proporciona estimaciones comparables a las obtenidas mediante otros métodos, además de poseer las ventajas de eficiencia computacional y de robustez numérica ya citadas. En un contexto econométrico, es conveniente utilizar algoritmos numéricamente robustos, puesto que las variables propagadas recursivamente por un filtro son utilizadas en la evaluación, para cada vector de parámetros, de la función de verosimilitud exacta, de la expresión analítica de su gradiente, de su hessiano y de la matriz de información.

Los algoritmos alternativos al de Kalman no fueron bien acogidos en la práctica por considerarse muy complejos los procedimientos de factorización de matrices. En contra de lo que se suele pensar sobre los algoritmos de factorización, las ecuaciones de Chandrasekhar son fácil y directamente aplicables a la estimación de cualquier modelo econométrico, cuando éste es expresado en una forma de espacio de los estados de dimensión mínima [vid. Terceiro (1990)].

Deseamos destacar que las cuestiones numéricas de un mecanismo de estimación deben preocupar a cualquier econométra que realiza un trabajo empírico. La fiabilidad del software empleado habitualmente en la práctica econométrica, es un aspecto crucial, si después se quieren utilizar los resultados obtenidos para explicar determinados hechos o relaciones entre variables económicas. Por ejemplo, la Econometría se ha preocupado de estudiar problemas numéricos como la multicolinealidad, ofreciendo métodos para detectarla y resolverla. Una parte sustancial del trabajo se encuentra en esta línea, ya que nos ocupamos de la estabilidad numérica de un algoritmo de estimación.

La estructura del trabajo es la siguiente:

En el Capítulo 1 se resume la teoría de filtros lineales en tiempo discreto, derivándose el filtro de Chandrasekhar para modelos estacionarios y no estacionarios. También se revisan los distintos filtros aplicables a nuestro problema y se realiza un análisis comparativo entre ellos.

En el Capítulo 2 se desarrollan los elementos básicos del mecanismo de estimación máximo-verosímil de modelos econométricos (lineales y dinámicos) en forma de espacio de los estados utilizando un filtro de Chandrasekhar. Para ello se obtienen las expresiones analíticas exactas de la función de verosimilitud de la muestra, del gradiente, del hessiano y de la matriz de información.

El Capítulo 3 describe los algoritmos de optimización numérica que se utilizan para maximizar la función de verosimilitud del modelo. Además, se explica la estructura del programa desarrollado en FORTRAN para instrumentar el mecanismo de estimación descrito en el capítulo anterior.

En el Capítulo 4 se muestran los resultados obtenidos al aplicar el mecanismo de estimación, tanto a modelos simulados como a modelos con datos económicos. Todos los resultados se comparan con los obtenidos por otros procedimientos alternativos.

En el Capítulo 5 se resumen las principales conclusiones del trabajo y se sugieren algunas extensiones del mismo.

Por último, se presenta una serie de apéndices cuyo contenido es el siguiente:

En el Apéndice A se deriva el filtro de Kalman en tiempo discreto, describiéndose las propiedades del mismo, su inicialización y la interpretación de sus resultados para el caso de modelos gaussianos. El Apéndice B describe el algoritmo de

estimación máximo-verosímil de las condiciones iniciales del filtro de Chandrasekhar. En el Apéndice C se presenta una visión general sobre la ecuación de Riccati asociada al filtro de Kalman y los distintos procedimientos existentes para encontrar soluciones a la misma. Y por último, en el Apéndice D se ofrecen los datos utilizados en la estimación de los modelos con datos reales presentados en el Capítulo 4.

CAPITULO 1

TEORIA DE FILTROS

1.1. MODELOS EN FORMA DE ESPACIO DE LOS ESTADOS.

Sea un sistema dinámico lineal cuyo estado queda definido en cada instante t por un vector de estado \mathbf{x}_t . La evolución de dicho estado en tiempo discreto puede representarse a través de la siguiente ecuación:

$$\mathbf{x}_t = \Phi_t \mathbf{x}_{t-1} + \Gamma_t \mathbf{u}_t + \mathbf{E}_t \mathbf{w}_t \quad [1.1.1]$$

donde se supone que el estado del sistema en el instante t , viene determinado por:

- 1) el estado del sistema en el instante anterior,
- 2) un vector de variables exógenas (\mathbf{u}_t) y
- 3) un vector de ruido (\mathbf{w}_t).

Además, se dispone de un vector de variables observadas o "medidas" \mathbf{z}_t , que se relacionan con el estado del sistema a través de la ecuación:

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{D}_t \mathbf{u}_t + \mathbf{C}_t \mathbf{v}_t \quad [1.1.2]$$

Un modelo en forma de espacio de los estados queda definido por las dos ecuaciones anteriores, siendo [1.1.1] la ecuación de estado o de transición y [1.1.2] la ecuación de observación.

En esta representación Φ_t , Γ_t , \mathbf{E}_t , \mathbf{H}_t , \mathbf{D}_t y \mathbf{C}_t son las matrices características del sistema. En principio, se supone que son matrices no estocásticas y el subíndice temporal indica que sus elementos pueden cambiar en el tiempo. Φ_t es la matriz ($n \times n$) de transición del sistema; Γ_t es la matriz ($n \times r$) de distribución del control; \mathbf{E}_t es la matriz ($n \times p$) de distribución del ruido de estado; \mathbf{H}_t es la matriz ($m \times n$) de observación; \mathbf{D}_t es la matriz ($m \times r$) de distribución del control en la ecuación de observación; \mathbf{C}_t es la matriz ($m \times c$) de distribución del error de observación; \mathbf{x}_t es el vector ($n \times 1$) de variables de estado; \mathbf{u}_t es el vector

(rx_1) de variables de control; z_t es el vector (mx_1) de variables observadas; w_t es el vector (px_1) de perturbaciones de la dinámica del sistema y v_t es el vector (cx_1) de errores de observación.

Con las dos ecuaciones anteriores se describe el comportamiento de un sistema dinámico en tiempo discreto, siendo conveniente señalar que:

- El que la dinámica del sistema sea de orden uno no supone ninguna restricción, ya que la forma de espacio de los estados permite expresar cualquier relación dinámica lineal. Como ya se verá, la dinámica de un modelo se introduce en esta representación regulando la dimensión del vector de estado.
- Las variables de estado del sistema no son, en general, directamente observables. A priori, lo único que se conoce de ellas es que siguen el proceso dado por la ecuación [1.1.1]. Desde el punto de vista muestral, se observa un vector de valores z_t que son combinación lineal de los componentes de x_t , u_t y de un vector de ruido v_t (vid. ecuación [1.1.2]).
- La linealidad del sistema [1.1.1]-[1.1.2] permite que, en cualquier instante de tiempo t , z_t pueda expresarse en función de los valores pasados y presentes de los ruidos w_t , v_t y del vector de estado inicial x_0 .

Una forma alternativa a la ecuación [1.1.1] es la siguiente:

$$x_{t+1} = \Phi_t x_t + \Gamma_t u_t + E_t w_t \quad [1.1.1']$$

junto con la ecuación de observación dada por [1.1.2]. La única diferencia entre [1.1.1] y [1.1.1'], es que en la primera se establece una relación contemporánea entre el vector de estado

(\mathbf{x}_t) y el vector de perturbaciones del sistema (\mathbf{w}_t) . En principio, ambas formulaciones son equivalentes. Sin embargo, es importante distinguir entre ambas en el caso de existir correlación contemporánea entre los ruidos \mathbf{v}_t y \mathbf{w}_t .

Es fácil demostrar que un modelo no tiene una única representación en forma de espacio de los estados. Para ello, sea \mathbf{L} una matriz cualquiera de orden $(n \times n)$ no singular. Entonces, puede definirse el vector:

$$\mathbf{x}_t^* = \mathbf{L}\mathbf{x}_t$$

Premultiplicando la ecuación [1.1.1] por la matriz \mathbf{L} , se obtiene la nueva ecuación de estado:

$$\mathbf{x}_t^* = \bar{\Phi}_t^* \mathbf{x}_{t-1}^* + \Gamma_t^* \mathbf{u}_t + \mathbf{E}_t^* \mathbf{w}_t$$

y su correspondiente ecuación de observación es ahora:

$$\mathbf{z}_t = \mathbf{H}_t^* \mathbf{x}_t^* + \mathbf{D}_t \mathbf{u}_t + \mathbf{C}_t \mathbf{v}_t$$

en donde $\bar{\Phi}_t^* = \mathbf{L}\bar{\Phi}_t\mathbf{L}^{-1}$; $\Gamma_t^* = \mathbf{L}\Gamma_t$; $\mathbf{E}_t^* = \mathbf{L}\mathbf{E}_t$; $\mathbf{H}_t^* = \mathbf{H}_t\mathbf{L}^{-1}$

Ejemplo: Dado un proceso MA(1) definido como $y_t = a_t + \theta a_{t-1}$, sus posibles formulaciones en forma de espacio de los estados se pueden resumir de la forma:

$$\begin{aligned} x_{t+1} &= \lambda a_t \\ z_t &= (\theta/\lambda)x_t + a_t \end{aligned}$$

siendo λ cualquier escalar distinto de cero y habiéndose hecho el siguiente cambio de variable: $x_{t+1} = a_t$; $z_t = y_t$.

Esto indica que a un mismo vector de variables endógenas (\mathbf{z}_t) y variables de control (\mathbf{u}_t) , le corresponden infinitos vectores de variables de estado (\mathbf{x}_t) . Sin embargo, más adelante

se verá cómo imponiendo ciertas restricciones en las matrices estructurales del sistema y en las matrices de distribución de los ruidos, se puede identificar un único vector de estado para cada modelo. Cuando estas restricciones han sido impuestas, se obtienen las llamadas formulaciones canónicas en espacio de los estados [vid. Wertz (1982)].

La formulación dada por las ecuaciones [1.1.1] y [1.1.2] ó [1.1.1'] y [1.1.2] es la llamada formulación propia en espacio de los estados. Esta forma tiene una ventaja computacional evidente, ya que la dimensión del vector de estado es la mínima posible¹.

Un caso particular de especial interés en muchas aplicaciones, surge cuando las matrices Φ_t , Γ_t , E_t , H_t , D_t y C_t no dependen del tiempo. Su formulación en forma de espacio de los estados es análoga a la ya descrita:

$$x_{t+1} = \Phi x_t + \Gamma u_t + E w_t \quad [1.1.1'']$$

$$z_t = H x_t + D u_t + C v_t \quad [1.1.2'']$$

¹ La formulación impropia en espacio de los estados tiene las siguientes características:

- Establece una relación contemporánea entre las variables de estado y el término de error en [1.1.1].
- Suprime de la ecuación de observación el término de error.

Por tanto, las ecuaciones [1.1.1] y [1.1.2], expresadas en forma impropia, serían ahora:

$$x_{t+1} = \Phi_t x_t + \Gamma_t u_t + E_t w_{t+1}$$

$$z_t = H_t x_t + D_t u_t$$

Esta formulación es menos aconsejable, porque en ella la dimensión del vector de estado aumenta con respecto a la formulación propia y, por tanto, aumenta la carga computacional. Otro problema derivado de esta formulación, es de tipo numérico y proviene de no incluir un término de error en la ecuación de observación [vid. Terceiro (1990)].

Una vez que el modelo que se está analizando se ha expresado en forma de espacio de los estados, existen tres técnicas para extraer alguna información estadística acerca de las variables no observables del sistema. Estas técnicas se basan en calcular:

$$\hat{\mathbf{x}}_{t/t'} = E(\mathbf{x}_t / \mathbf{z}^{t'})$$

donde $\mathbf{z}^{t'}$ es el conjunto de información disponible hasta el instante t' .

Si $t > t'$, se predice el valor futuro del vector de estado dada la información disponible hasta el instante t' . Esta técnica para obtener información acerca de \mathbf{x}_t es la de predicción.

Si $t = t'$, se estima el valor presente del vector de estado usando para ello las medidas disponibles hasta ese momento. Esta técnica es la de filtrado.

Si $t < t'$ y se dispone, en total, de t' observaciones, se intenta calcular el estado \mathbf{x}_t utilizando toda la muestra disponible. Esta técnica se conoce en la literatura como smoothing.

En cada uno de estos casos se obtienen, además, las correspondientes matrices de varianzas-covarianzas del error de estimación del vector de estado:

$$\mathbf{P}_{t/t'} = E[(\mathbf{x}_t - \hat{\mathbf{x}}_{t/t'}) (\mathbf{x}_t - \hat{\mathbf{x}}_{t/t'})' / \mathbf{z}^{t'}]$$

Este trabajo se centra en las técnicas de filtrado y en particular, en el filtro de Kalman y algunas de sus variaciones numéricas como el filtro de Chandrasekhar. Una presentación detallada de las técnicas de predicción y smoothing puede verse en Jazwinsky (1970), Bierman (1973), Anderson y Moore (1979) y Ansley y Kohn (1983, 1985).

1.2. EL FILTRO DE KALMAN EN TIEMPO DISCRETO.

Una vez expresado un modelo en forma de espacio de los estados, pueden utilizarse diversos algoritmos para estimar el vector de estado. El algoritmo más conocido y utilizado para este fin es el llamado filtro de Kalman [vid. Kalman (1960)].

El filtro de Kalman es un procedimiento que permite obtener una estimación recursiva óptima de los elementos del vector de estado \mathbf{x}_t , contando para ello con la información disponible hasta el instante de tiempo t^2 .

Bajo el supuesto de independencia de los ruidos de estado y de observación, las ecuaciones que definen el filtro de Kalman son:

$$\hat{\mathbf{x}}_{t/t-1} = \Phi_t \hat{\mathbf{x}}_{t-1/t-1} + \Gamma_t \mathbf{u}_t \quad [1.2.1]$$

$$\mathbf{P}_{t/t-1} = \Phi_t \mathbf{P}_{t-1/t-1} \Phi_t' + \mathbf{E}_t \mathbf{Q}_t \mathbf{E}_t' \quad [1.2.2]$$

$$\hat{\mathbf{x}}_{t/t} = \hat{\mathbf{x}}_{t/t-1} + \mathbf{K}_t \tilde{\mathbf{z}}_t \quad [1.2.3]$$

$$\tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{H}_t \hat{\mathbf{x}}_{t/t-1} - \mathbf{D}_t \mathbf{u}_t \quad [1.2.4]$$

$$\mathbf{P}_{t/t} = [\mathbf{I} - \mathbf{K}_t \mathbf{H}_t] \mathbf{P}_{t/t-1} \quad [1.2.5]$$

$$\mathbf{K}_t = \mathbf{P}_{t/t-1} \mathbf{H}_t' \mathbf{B}_t^{-1} \quad [1.2.6]$$

$$\mathbf{B}_t = \mathbf{H}_t \mathbf{P}_{t/t-1} \mathbf{H}_t' + \mathbf{C}_t \mathbf{R}_t \mathbf{C}_t' \quad [1.2.7]$$

donde [1.2.1] y [1.2.2] son las llamadas ecuaciones de predicción y [1.2.3]-[1.2.7] definen el ciclo de actualización o filtrado. Dada una estimación inicial del vector de estado, el filtro

² Una derivación del filtro de Kalman puede verse en el Apéndice A, en donde se distingue el caso en que los ruidos de estado y de observación sean independientes entre sí o estén correlacionados contemporáneamente. Además, se describen las propiedades del filtro, la interpretación de los resultados del mismo para modelos gaussianos y sobre todo, se plantea el origen de sus problemas que son siempre de tipo numérico. En este apartado, sólo se resume brevemente la parte del Apéndice que será utilizada en el resto del Capítulo.

proporciona una estimación óptima del vector de estado en el instante t , contando con la información hasta el instante anterior. Cuando está disponible un nuevo dato sobre las variables observadas z_t , el filtro actualiza la estimación obtenida en la fase anterior ponderando el error cometido al predecir z_t mediante su esperanza condicionada por la información hasta $t-1$, por la matriz ganancia K_t . Ese vector de errores de predicción \tilde{z}_t , conocido como proceso de innovaciones, tiene una matriz de covarianzas denotada por B_t . Además, en las dos fases, se calculan las matrices de covarianzas asociadas al error de estimación del vector de estado ($P_{t/t-1}$ y $P_{t/t}$). El filtro de Kalman puede expresarse de una forma alternativa uniendo las fases de predicción y actualización:

$$\hat{x}_{t+1/t} = (\Phi_{t+1} - K_t H_t) \hat{x}_{t/t-1} + K_t (z_t - D_t u_t) + \Gamma_{t+1} u_{t+1} \quad [1.2.8]$$

$$K_t = \Phi_{t+1} P_{t/t-1} H_t' B_t^{-1} \quad [1.2.9]$$

$$P_{t+1/t} = \Phi_{t+1} P_{t/t-1} \Phi_{t+1}' + E_{t+1} Q_{t+1} E_{t+1}' - K_t B_t K_t' \quad [1.2.10]$$

donde [1.2.10] es la llamada ecuación de Riccati del filtro de Kalman.

Si los ruidos de estado y de observación están correlacionados contemporáneamente, las ecuaciones de predicción del filtro siguen siendo [1.2.1] y [1.2.2], mientras que las de actualización pasan a ser las siguientes:

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + K_t \tilde{z}_t \quad [1.2.3']$$

$$\tilde{z}_t = z_t - H_t \hat{x}_{t/t-1} - D_t u_t \quad [1.2.4']$$

$$K_t = (P_{t/t-1} H_t' + E_t S_t C_t') B_t^{-1} \quad [1.2.5']$$

$$B_t = H_t P_{t/t-1} H_t' + C_t R_t C_t' + H_t E_t S_t C_t' + C_t S_t E_t' H_t' \quad [1.2.6']$$

$$P_{t/t} = P_{t/t-1} - (P_{t/t-1} H_t' + E_t S_t C_t') B_t^{-1} (H_t P_{t/t-1} + C_t S_t E_t') \quad [1.2.7']$$

Por último, si se utiliza la formulación [1.1.1']-[1.1.2], pueden unirse las ecuaciones de predicción y de actualización resultando las expresiones:

$$\hat{\mathbf{x}}_{t+1/t} = \Phi_t \hat{\mathbf{x}}_{t/t-1} + \Gamma_t \mathbf{u}_t + \mathbf{K}_t \tilde{\mathbf{z}}_t \quad [1.2.3'']$$

$$\tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{H}_t \hat{\mathbf{x}}_{t/t-1} - \mathbf{D}_t \mathbf{u}_t \quad [1.2.4'']$$

$$\mathbf{K}_t = (\Phi_t \mathbf{P}_{t/t-1} \mathbf{H}_t' + \mathbf{E}_t \mathbf{S}_t \mathbf{C}_t') \mathbf{B}_t^{-1} \quad [1.2.5'']$$

$$\mathbf{P}_{t+1/t} = \Phi_t \mathbf{P}_{t/t-1} \Phi_t' + \mathbf{E}_t \mathbf{Q}_t \mathbf{E}_t' - \mathbf{K}_t \mathbf{B}_t \mathbf{K}_t' \quad [1.2.6'']$$

$$\mathbf{B}_t = \mathbf{H}_t \mathbf{P}_{t/t-1} \mathbf{H}_t' + \mathbf{C}_t \mathbf{R}_t \mathbf{C}_t' \quad [1.2.7'']$$

Esta última forma del filtro de Kalman es la más utilizada en el trabajo.

1.3. FILTROS DERIVADOS DE LA FACTORIZACION DE LA MATRIZ P.

Para resolver los problemas numéricos que presenta el filtro convencional de Kalman en su aplicación, se utilizan algoritmos numéricamente estables derivados de distintas factorizaciones de la matriz $P_{t/t}$.

1.3.1. Problemas numéricos del filtro de Kalman.

El filtro de Kalman es muy sensible a los errores de redondeo y su robustez numérica puede degradarse hasta el punto de invalidar sus resultados. También es sabido que la mayor parte de los problemas numéricos del filtro de Kalman se reflejan en las matrices de covarianzas, que dejan de ser semidefinidas positivas en algún momento de su propagación (vid. Capítulo 4). Por tanto, parece interesante utilizar expresiones alternativas del filtro de Kalman que sean numéricamente estables.

Los problemas numéricos del filtro de Kalman pueden deberse a diferentes causas. Las más importantes son:

- (a) El empleo de condiciones iniciales del filtro arbitrarias.
- (b) La existencia de no linealidades en el problema.
- (c) Los errores de redondeo.

El primer problema puede resolverse, en determinados contextos, aplicando métodos de estimación óptima de las condiciones iniciales del filtro (vid. Apéndice B). Para resolver el segundo problema se ha desarrollado el llamado filtro de Kalman extendido³. Por último, como solución al problema (c) se han desarrollado los algoritmos basados en factorizaciones de la matriz $P_{t/t}$. Usualmente, los efectos de los errores numéricos se manifiestan en el cálculo de la matriz de covarianzas del vector

³ Vid. Jazwinsky (1970) y Anderson y Moore (1979).

de estado, que deja de ser semidefinida positiva. Esto se debe a que, en la versión convencional del filtro de Kalman, esa matriz se propaga como la diferencia de dos matrices semidefinidas positivas, siendo posible su degradación que se traduce en:

- 1) Pérdida de simetría de la matriz.
- 2) Pérdida de condicionamiento de la misma.

La robustez numérica de un filtro de Kalman mejorará, si se toman algunas precauciones en el cálculo. Estas pueden consistir en:

- Calcular sólo la parte triangular inferior o superior de la matriz $P_{t+1/t}$ y posteriormente, forzar su simetría.
- Fijar una matriz de covarianzas inicial, denotada como $P_{0/-1}$, con una traza no demasiado grande, ya que una incertidumbre inicial excesiva tiende a aumentar los errores de redondeo.
- Escalar las unidades del problema para reducir su rango de variación.
- Realizar los cálculos con precisión aritmética doble.

Estas precauciones reducen, pero no resuelven, la pérdida de condicionamiento de P_t/t . Existen en la literatura diversas soluciones a este último problema:

Métodos ad hoc.

- (a) Aproximar la ecuación de Riccati [1.2.10], suprimiendo el último sumando. Este es el término que puede llegar a convertir en definida negativa a la matriz de varianzas-covarianzas. Esta aproximación puede resolver el problema numérico, pero no existe ninguna

razón teórica que la justifique. Aún más, al ser la matriz $\Sigma_t P_{t/t-1} \Sigma_t' > 0$, $E_t Q_t E_t' > 0$ y $K_t B_t K_t' > 0$, la omisión de este último sumando tiende a aumentar artificialmente la varianza del estimador del vector de estado. Por la misma razón, al ser $H_t P_{t/t-1} H_t' > 0$ y aumentar la varianza del vector de estado, aumentaría la varianza del vector de observación y por tanto, el error de predicción. Además, el filtro perdería la propiedad de optimalidad (vid. Apéndice A).

- (b) En el caso de usar las ecuaciones [1.2.1]-[1.2.7], se sustituye $P_{t/t} = (I - K_t H_t) P_{t/t-1}$ por su expresión equivalente [A.5'], pero más estable (vid. Apéndice A). Si se usan las ecuaciones [1.2.8]-[1.2.10], se puede sustituir la ecuación de Riccati [1.2.10] por una versión más estable dada por [A.11] (vid. Apéndice A). Sin embargo, al contrario de lo que suele creerse, estas versiones más estables no garantizan la robustez numérica del filtro de Kalman. Esto ha sido demostrado con diversos ejemplos por Bierman y Thornton (1977, 1980).
- (c) Verificar en cada iteración la matriz $P_{t+1/t}$ manipulando lo mínimo posible los elementos de su diagonal (para mantener su carácter de definida positiva) y los de fuera de su diagonal (para garantizar que los coeficientes de correlación de la matriz sean, en valor absoluto, menores que la unidad). La forma de alterar los elementos de fuera de la diagonal se basa en la siguiente propiedad de las matrices definidas positivas: Si P es definida positiva y se definen los coeficientes de correlación de dicha matriz de la forma $\rho(i,j) = P(i,j) / [P(i,i)P(j,j)]^{1/2}$, $\forall i,j = 1,2 \dots, n$ y siendo $i \neq j$, entonces se cumple que $|\rho(i,j)| < 1$. De acuerdo con esta propiedad, se alteran los elementos de fuera de la diagonal para que los coeficientes de correlación

sean menores que la unidad. Sin embargo, puede mostrarse que, en general, esto no basta para mantener su condición de definida positiva⁴.

1.3.2. Algoritmos de la raíz cuadrada de P_t/t .

La alternativa a todos estos métodos ad-hoc es usar un algoritmo que esté mejor condicionado numéricamente que el filtro de Kalman. Una solución son los llamados filtros de la raíz cuadrada. Estos algoritmos se basan en propagar una matriz análoga a la raíz cuadrada de la matriz de covarianzas del vector de estado, lo que da lugar, por construcción, a una mayor estabilidad numérica que el filtro de Kalman.

Este mejor comportamiento se debe, en gran parte, a que se reduce el rango de variación numérica de las variables que se propagan, lo cual genera menores errores de redondeo. De esta forma, cálculos que incluyen números comprendidos en el rango $[10^{-n}, 10^n]$ se reducen a números comprendidos entre $[10^{-n/2}, 10^{n/2}]$. Además, en estos algoritmos, el número de condición de la raíz cuadrada de P_t/t es la raíz cuadrada del número de condición de P_t/t . Normalmente no se ha puesto demasiado énfasis en estos algoritmos porque, erróneamente, se ha creído que los métodos de factorización de matrices son demasiado complejos para ponerlos en práctica.

⁴ Por ejemplo, sea una matriz de covarianzas P normalizada:

$$P = \begin{bmatrix} 1 & 0.5-\delta & 0.5-\delta \\ 0.5-\delta & 1 & -0.5+\delta \\ 0.5-\delta & -0.5+\delta & 1 \end{bmatrix}$$

Imponer que los coeficientes de correlación sean en valor absoluto menor que la unidad, implica suponer que $-0.5 < \delta < 1.5$. Sin embargo, el $\det(P) = 2\delta(1.5 - \delta)^2$ y para que sea definida positiva debemos exigir además que $\delta > 0$. Por tanto, la condición sobre los coeficientes de correlación no es suficiente.

Todos los algoritmos de la raíz cuadrada se basan en que siempre se puede descomponer una matriz \mathbf{A} simétrica y definida positiva de la forma $\mathbf{A} = \mathbf{B}\mathbf{B}'$, donde \mathbf{B} es llamada la raíz cuadrada de \mathbf{A} . La noción de raíz cuadrada de una matriz es similar a la noción de raíz cuadrada de un número. Sin embargo, las raíces cuadradas de una matriz, si existen, no son únicas. Así, si \mathbf{B} es una raíz cuadrada de \mathbf{A} y \mathbf{T} es una matriz ortogonal, entonces $\mathbf{B}\mathbf{T}$ también será una raíz cuadrada de \mathbf{A} , ya que $\mathbf{A} = \mathbf{B}\mathbf{T}\mathbf{T}'\mathbf{B}' = \mathbf{B}\mathbf{B}'$, al cumplirse que $\mathbf{T}\mathbf{T}' = \mathbf{I}$.

Se han desarrollado en la literatura dos tipos de filtros de la raíz cuadrada:

- (a) El filtro de la raíz cuadrada de la matriz de covarianzas del error de estimación del vector de estado, que se obtiene factorizando la matriz $\mathbf{P}_{t/t}$.
- (b) El filtro de la raíz cuadrada de la matriz de información, que se obtiene factorizando la inversa de $\mathbf{P}_{t/t}$.

Ninguno de estos dos tipos de filtro es superior al otro en todos los casos. Así, los algoritmos del tipo (a) son más aconsejables para problemas donde el vector de variables observadas es escalar (o la dimensión de este vector es considerablemente menor a la del vector de estado) y donde se tiene un buen conocimiento de los parámetros del modelo. Los algoritmos relacionados con la matriz de información, son más convenientes cuando se dispone de una gran cantidad de datos y no se sabe apenas nada de los parámetros del modelo.

Factorización de Potter.

El primero que introdujo un algoritmo de factorización de la matriz $\mathbf{P}_{t/t}$ fue Potter (1963). La aproximación de Potter consiste en encontrar una matriz \mathbf{S} (llamada raíz cuadrada de \mathbf{P}),

tal que $P = SS'$. No es necesario que S tenga ninguna estructura especial. El algoritmo original de Potter parte de la ecuación de observación:

$$z_t = Hx_t + v_t$$

donde se supone que $E(v_t v_t') = I$. La factorización de Potter necesita un vector de estado inicial, \tilde{x} y la correspondiente raíz cuadrada de la matriz de covarianzas de ese vector, denotada por \tilde{S} . De este modo, la matriz de covarianzas del vector de estado inicial se calcularía como $\tilde{P} = \tilde{S}\tilde{S}'$. Dada la ecuación [1.2.5] del filtro de Kalman, la manera de actualizar los factores de la matriz P sería la siguiente:

$$\begin{aligned} \hat{P} &= (I - K_t H) \tilde{P} = \tilde{S}\tilde{S}' - K_t H \tilde{S}\tilde{S}' = \tilde{S}\tilde{S}' - \tilde{S}\tilde{S}' H' B_{t-1}^{-1} H \tilde{S}\tilde{S}' = \\ &= \tilde{S} [I - \tilde{S}' H' B_{t-1}^{-1} H \tilde{S}] \tilde{S}' = \tilde{S} [I - \alpha B_t^{-1} \alpha'] \tilde{S}' \end{aligned}$$

y denotando:

$$\begin{aligned} \alpha &= \tilde{S}' H' \\ B_t &= H \tilde{S}\tilde{S}' H' + I = \alpha' \alpha + I \end{aligned}$$

es fácil definir:

$$\hat{S} = \tilde{S} [I - \alpha B_{t-1}^{-1} \alpha']^{\frac{1}{2}} \quad [1.3.2.1]$$

Cuando el vector de variables observadas es un escalar, la matriz $[I - \alpha B_{t-1}^{-1} \alpha']$ se puede factorizar como $[I - \beta \alpha \alpha']^2$ donde $\beta = 1/[b_t^{\frac{1}{2}}(b_t^{\frac{1}{2}} + 1)]$. Es decir:

$$\begin{aligned} [I - \beta \alpha \alpha'] [I - \beta \alpha \alpha'] &= [I - 2\beta \alpha \alpha' + \beta^2 \alpha \alpha' \alpha \alpha'] = \\ &= [I - (2\beta - \beta^2 \alpha' \alpha) \alpha \alpha'] = [I - b_t^{-1} \alpha \alpha'] \end{aligned}$$

Por tanto: $b_t^{-1} = 2\beta - \beta^2 \alpha' \alpha$

Y hallando las raíces de la ecuación cuadrática:

$$\beta^2 \alpha' \alpha - 2\beta + b_t^{-1} = 0$$

se obtiene que $\beta = 1/[b_t^{\frac{1}{2}}(b_t^{\frac{1}{2}} \pm 1)]$. Para evitar problemas de cancelación, se escoge la raíz $\beta = 1/[b_t^{\frac{1}{2}}(b_t^{\frac{1}{2}} + 1)]$. Y sustituyendo esta factorización en [1.3.2.1], se obtiene:

$$\hat{\mathbf{s}} = \tilde{\mathbf{s}} - [b_t^{\frac{1}{2}}/(b_t^{\frac{1}{2}} + 1)]\mathbf{K}_t\alpha' \quad [1.3.2.2]$$

en donde:

$$\mathbf{K}_t = \tilde{\mathbf{s}}\alpha b_t^{-1} \quad [1.3.2.3]$$

Las expresiones [1.3.2.2] y [1.3.2.3] son las ecuaciones clave de la factorización de Potter. Las ecuaciones de este algoritmo por orden de propagación son las siguientes:

$$\forall j = 1, \dots, n$$

$$\alpha_j' = H_j \mathbf{s}_j \quad [1.3.2.4]$$

$$\sigma_j = 1/(\alpha_j' \alpha_j + 1) \quad [1.3.2.5]$$

$$\bar{\mathbf{K}}_j = \mathbf{s}_j \alpha_j \quad (\text{o bien, } \bar{\mathbf{K}}_j = \mathbf{K}_j/\sigma_j) \quad [1.3.2.6]$$

$$\delta_j = z_j - H_j \hat{\mathbf{x}}_{j/j-1} \quad [1.3.2.7]$$

$$\hat{\mathbf{x}}_{j/j} = \hat{\mathbf{x}}_{j/j-1} + \bar{\mathbf{K}}_j \delta_j \sigma_j \quad [1.3.2.8]$$

$$\gamma_j = \sigma_j / (1 + \sigma_j^{\frac{1}{2}}) \quad [1.3.2.9]$$

$$\mathbf{s}_{j+1} = \mathbf{s}_j - \bar{\mathbf{K}}_j \gamma_j \alpha_j' \quad [1.3.2.10]$$

La ecuación [1.3.2.5] propaga la inversa de la matriz de covarianzas del vector de medidas descrito en la ecuación de observación. La ecuación [1.3.2.6] es la matriz ganancia normalizada, que no es más que la ganancia del filtro de Kalman afectada por un factor de escala. La ecuación [1.3.2.7] expresa la desviación entre el valor real de la medida z_t en el instante t y su esperanza condicional, es decir, es el vector de innovaciones. La ecuación [1.3.2.8] actualiza la estimación del vector de estado del sistema. La expresión [1.3.2.10] actualiza la estimación de la raíz cuadrada de la matriz de covarianzas del

vector de estado. Y por último, las expresiones [1.3.2.4] y [1.3.2.9] son simplemente definiciones⁵.

Un problema de la factorización de Potter es que sólo puede aplicarse cuando el vector de medidas es escalar y cuando éstas son independientes entre sí. Sin embargo, esto último no es realmente una limitación del algoritmo. Así, Andrews (1968) propone su uso preblanqueando antes el vector de medidas si éstas están correlacionadas⁶. El algoritmo de Potter tiene mayores requerimientos de cálculo que la versión convencional del filtro de Kalman. Esto es evidente, pues en la factorización de Potter hay que actualizar n^2 elementos de la matriz S , mientras que en el algoritmo de Kalman sólo se han de actualizar $n(n+1)/2$ elementos distintos de P_t/t . Surgen, por este motivo, modificaciones del mismo como el algoritmo UDU' de Bierman.

Factorización de Bierman.

Bierman (1974) propone el algoritmo UDU', que consiste en factorizar P_t/t de la forma $P = UDU'$ donde U es una matriz triangular superior con unos en la diagonal principal y D una matriz diagonal. Esta es una instrumentación estable de la

⁵ Las condiciones iniciales de este filtro son una estimación inicial del vector de estado, x , y una raíz cuadrada inicial de la matriz de covarianzas del error de esa estimación, llamada S .

La elección de la condición inicial \tilde{S} depende de la forma concreta que tenga la matriz P :

- Si \tilde{P} es una matriz diagonal, su raíz cuadrada $\tilde{S} = \text{diag} \{P(1,1)^{\frac{1}{2}}, P(2,2)^{\frac{1}{2}}, \dots, P(n,n)^{\frac{1}{2}}\}$.
- Si \tilde{P} no es diagonal, una posibilidad para calcular \tilde{S} sería hallar la descomposición de Cholesky de P , es decir, $P = LL'$ y elegir $S = L$.

⁶ Esta solución se basa en que una aplicación de la factorización de matrices definidas positivas, consiste en poder convertir un vector de medidas correlacionado en uno independiente.

descomposición propuesta por Agee y Turner (1972). Esta factorización puede considerarse como un algoritmo del tipo raíz cuadrada al ser $UD^{\frac{1}{2}}$ una posible raíz de $P_{t/t}$. Las ecuaciones que forman el algoritmo de Bierman se ofrecen a continuación:

Supongamos una ecuación de observación donde el vector de medidas es escalar $z_t = hx_t + v_t$ siendo $E(v_t) = 0$ y $E(v_tv_t') = r$. La estimación inicial del vector de estado es \tilde{x} y su matriz de covarianzas $\tilde{P} = \tilde{U}\tilde{D}\tilde{U}'$, donde \tilde{U} es una matriz triangular superior con unos en la diagonal y \tilde{D} una matriz diagonal con elementos positivos. Con estos supuestos, la ganancia del filtro de Kalman y los factores U y D se obtienen de la siguiente forma:

$$f' = h\tilde{U} \quad \text{siendo } f' = [f_1 \dots f_n]$$

$$g = \tilde{D}f' \quad (g_i = \tilde{D}_i f_i)$$

$$\forall j = 1, \dots, n$$

$$\alpha_0 = r$$

$$\alpha_j = \alpha_{j-1} + f_j g_j \quad [1.3.2.7]$$

$$\hat{D}_j = (\alpha_{j-1}/\alpha_j) D_j \quad (\hat{D}_j = \tilde{D}_j \text{ si } \alpha_j = 0) \quad [1.3.2.8]$$

$$\nu_j = g_j \quad [1.3.2.9]$$

$$\lambda = -f_j/\alpha_{j-1} \quad (\lambda = 0 \text{ si } \alpha_{j-1} = 0) \quad [1.3.2.10]$$

$$\forall i = 1, \dots, j-1$$

$$\hat{u}_{ij} = \tilde{u}_{ij} + \nu_i \lambda \quad [1.3.2.11]$$

$$\nu_i = \nu_i + \tilde{u}_{ij} \nu_j \quad [1.3.2.12]$$

La ecuación [1.3.2.7] actualiza la varianza del proceso de innovaciones. [1.3.2.8] actualiza los elementos de la matriz diagonal D . [1.3.2.11] actualiza la columna j -ésima de la matriz factor U . Las expresiones [1.3.2.9] y [1.3.2.10] son definiciones. Por último, [1.3.2.12] es la ganancia normalizada del filtro de Kalman. Como puede verse, no es necesario calcular

\hat{P} , que, sin embargo, puede recuperarse en cada instante de tiempo de la forma $\hat{P} = \hat{U}\hat{D}\hat{U}'$.

La estabilidad numérica de esta factorización se debe a que los términos α_j en [1.3.2.7] se calculan como la suma de dos elementos positivos, evitándose los errores numéricos por cancelación de términos. La derivación de este algoritmo puede verse en Bierman (1977) o Thornton y Bierman (1980). Estos autores demuestran que, en realidad, la robustez numérica del algoritmo se debe al uso de transformaciones ortogonales numéricamente estables en cada iteración del proceso recursivo.

Otra característica de los algoritmos de factorización con matrices triangulares (como es el de Bierman), es que la singularidad o casi singularidad de la matriz P puede detectarse simplemente examinando los elementos de la matriz factor D . Esto se debe a que con esta factorización, $\det(P) = \det(UDU') = d_{11} \dots d_{nn}$, ya que el determinante de una matriz triangular unitaria es uno. Por tanto, la singularidad o casi singularidad de P requerirá que al menos algún elemento de D sea muy pequeño.

Por último, cabe comentar que el hecho de que un algoritmo sea numéricamente más robusto significa que es menos susceptible a los efectos de posibles errores de redondeo. Y el que un algoritmo sea numéricamente estable implica que los errores de redondeo no dan lugar a la divergencia de los resultados del filtro. Por tanto, para problemas razonablemente bien condicionados, los resultados de las estimaciones serán más fiables que los del filtro de Kalman. Esto no quiere decir que dichos errores desaparezcan completamente, sobre todo si el modelo en consideración está especialmente mal condicionado. Por supuesto, si pudiésemos trabajar en precisión aritmética absoluta, los resultados de todos los algoritmos vistos hasta ahora serían idénticos.

1.3.3. Algoritmos de factorización de los incrementos de $P_{t+1/t}$.

Estos filtros han sido desarrollados por Kailath (1973) y Morf, Sidhu y Kailath (1974) para el caso de sistemas con matrices constantes en el tiempo. Están basados en las ecuaciones de Chandrasekhar, derivadas inicialmente para sistemas lineales en tiempo continuo y después extendidas a sistemas en tiempo discreto. Su principal característica consiste en sustituir la ecuación de Riccati del filtro de Kalman, por un conjunto de ecuaciones en diferencias llamadas ecuaciones de Chandrasekhar.

Las ecuaciones de Chandrasekhar se derivan de la factorización de los incrementos de $P_{t+1/t}$ en lugar de la factorización de la propia matriz. Por tanto, es claro que existe una relación entre los algoritmos de Kailath y Morf y los de Potter, Bierman, Thornton et al. Sin embargo, la aproximación de Kailath y Morf tiene importantes ventajas computacionales cuando la dimensión del vector de estado es considerablemente mayor que la del vector de observación del sistema, caso frecuente en Econometría. Además, estos algoritmos disfrutan de las propiedades de robustez numérica de los métodos de factorización. Posteriormente, se verá cómo estos filtros pueden tomar formas alternativas, dependiendo las ventajas numéricas de cada una de ellas de las características concretas del modelo a estudiar.

1.4. EL FILTRO DE CHANDRASEKHAR EN TIEMPO DISCRETO.

Los algoritmos de Chandrasekhar en tiempo discreto se derivan para sistemas con matrices constantes. Como se verá en este apartado, estos filtros tienen una estructura diferente a la de los filtros de Kalman, así como importantes ventajas computacionales y de robustez numérica.

A continuación, y siguiendo a Morf, Sidhu y Kailath (1974), se derivan las ecuaciones que forman un filtro de Chandrasekhar para el caso general en que la matriz C es distinta a la identidad y los términos de error v_t y w_t están correlacionados contemporáneamente ($S \neq 0$).

1.4.1. Derivación de las ecuaciones de Chandrasekhar.

Los algoritmos de Chandrasekhar se derivan de la factorización de la matriz de incrementos de $P_{t/t-1}$, definida como:

$$V_{t/t-1} = P_{t+1/t} - P_{t/t-1} \quad [1.4.1.1]$$

Sea la siguiente factorización de $V_{t/t-1}$:

$$V_{t/t-1} = Y_t M_t Y_t' \quad [1.4.1.2]$$

en donde:

M_t es una matriz cuadrada y simétrica de orden α ,
 Y_t es una matriz $(n \times \alpha)$ y
 α se determina a partir de las condiciones iniciales del sistema.

Los filtros de Chandrasekhar se caracterizan por propagar los términos de la factorización dada en [1.4.1.2], en lugar de la propia matriz $P_{t/t-1}$, como ocurre en los filtros de Kalman.

Como las expresiones [1.2.3''] y [1.2.4''] del filtro de Kalman, que propagan el estimador del vector de estado y el proceso de innovaciones, son independientes de $P_{t/t-1}$, pueden incorporarse directamente a un algoritmo de Chandrasekhar.

La primera ecuación del filtro de Kalman que depende de $P_{t/t-1}$, es la que calcula la matriz de covarianzas del vector de observación. Esta matriz viene dada por la expresión [1.2.7'']. Sin embargo, manipulando adecuadamente [1.2.7''] y teniendo en cuenta [1.4.1.1]-[1.4.1.2], se obtiene:

$$\begin{aligned}
 B_t &= HP_{t/t-1}H' + CRC' = HP_{t-1/t-2}H' + CRC' + \\
 &+ H(P_{t/t-1} - P_{t-1/t-2})H' = \\
 &= HP_{t-1/t-2}H' + CRC' + HV_{t-1/t-2}H' = \\
 &= B_{t-1} + HY_{t-1}M_{t-1}Y_{t-1}'H' \qquad [1.4.1.3]
 \end{aligned}$$

ecuación que ya no depende de $P_{t/t-1}$ y que, por tanto, puede incorporarse a un algoritmo de tipo Chandrasekhar.

Para obtener la segunda ecuación del algoritmo, se parte de la expresión de la ganancia del filtro de Kalman [1.2.5'']. Operando en ella y teniendo en cuenta [1.4.1.1]-[1.4.1.2], se obtiene una nueva expresión para la matriz K_t :

$$\begin{aligned}
 K_{t+1} &= [\Phi P_{t+1/t}H' + ESC'] [HP_{t+1/t}H' + CRC']^{-1} = \\
 &= [\Phi P_{t/t-1}H' + ESC'] B_t B_t^{-1} B_{t+1}^{-1} + \\
 &+ [\Phi (P_{t+1/t} - P_{t/t-1})H'] B_{t+1}^{-1} = \\
 &= [K_t B_t + \Phi V_{t/t-1}H'] B_{t+1}^{-1} = \\
 &= [K_t B_t + \Phi Y_t M_t Y_t' H'] B_{t+1}^{-1} \qquad [1.4.1.4]
 \end{aligned}$$

A partir de [1.4.1.4], se obtiene otra ecuación equivalente, que puede utilizarse alternativamente en el filtro de Chandrasekhar:

$$\begin{aligned}
K_{t+1} &= K_t [H(P_{t/t-1} - P_{t+1/t})H' + CR C' + HP_{t+1/t}H'] B_{t+1}^{-1} + \\
&+ \Sigma V_{t/t-1}H' B_{t+1}^{-1} = \\
&= K_t [-HV_{t/t-1}H' + B_{t+1}] B_{t+1}^{-1} + \Sigma V_{t/t-1}H' B_{t+1}^{-1} = \\
&= K_t - [K_t H V_{t/t-1}H' + \Sigma V_{t/t-1}H'] B_{t+1}^{-1} = \\
&= K_t + [\Sigma - K_t H] V_{t/t-1}H' B_{t+1}^{-1} = \\
&= K_t + [\Sigma - K_t H] Y_t M_t Y_t' H' B_{t+1}^{-1} \quad [1.4.1.4']
\end{aligned}$$

La ecuación [1.4.1.4'] tiene una importante ventaja computacional con respecto a [1.4.1.4]: no exige almacenar en cada iteración la matriz de covarianzas del vector de observación del instante anterior, lo cual permite reducir la memoria requerida por el software.

Las restantes ecuaciones de este algoritmo propagan los términos de la factorización de $V_{t/t-1}$: Y_t y M_t . Para deducir dichas ecuaciones, se parte de la expresión del filtro de Kalman que propaga la matriz de varianzas-covarianzas del vector de estado.

Sabiendo que $V_{t+1/t} = P_{t+2/t+1} - P_{t+1/t}$ y teniendo en cuenta [1.2.6'''], se llega a:

$$\begin{aligned}
V_{t+1/t} &= \Sigma P_{t+1/t} \Sigma' + EQE' - K_{t+1} B_{t+1} K_{t+1}' - \Sigma P_{t/t-1} \Sigma' - EQE' + \\
&+ K_t B_t K_t' = \Sigma V_{t/t-1} \Sigma' - K_{t+1} B_{t+1} K_{t+1}' + K_t B_t K_t'
\end{aligned}$$

Sustituyendo [1.4.1.4'] en la expresión anterior, se obtiene:

$$\begin{aligned}
V_{t+1/t} &= \Sigma V_{t/t-1} \Sigma' + K_t B_t K_t' - K_t B_{t+1} K_t' + K_t H V_{t/t-1} H' K_t' - \\
&- K_t H V_{t/t-1} \Sigma' + K_t H V_{t/t-1} H' K_t' + K_t H V_{t/t-1} H' B_{t+1}^{-1} H V_{t/t-1} \Sigma' - \\
&- \Sigma V_{t/t-1} H' B_{t+1}^{-1} H V_{t/t-1} \Sigma' + \Sigma V_{t/t-1} H' B_{t+1}^{-1} H V_{t/t-1} H' K_t' - \\
&- K_t H V_{t/t-1} H' B_{t+1}^{-1} H V_{t/t-1} H' K_t' - \Sigma V_{t/t-1} H' K_t'
\end{aligned}$$

El segundo, tercer y cuarto sumando de esta expresión se anulan. Y sacando factor común $\bar{\Sigma}_t$ por la izquierda y por la derecha, se llega al resultado:

$$V_{t+1/t} = \bar{\Sigma}_t [V_{t/t-1} - V_{t/t-1}H'B_{t+1}^{-1}HV_{t/t-1}] \bar{\Sigma}_t' \quad [1.4.1.5]$$

donde $\bar{\Sigma}_t = [\Sigma - K_t H]$.

Sustituyendo la matriz $V_{t+1/t}$ por su factorización, se obtiene la expresión:

$$\begin{aligned} Y_{t+1}M_{t+1}Y_{t+1}' &= \bar{\Sigma}_t [Y_t M_t Y_t' - Y_t M_t Y_t' H' B_{t+1}^{-1} H Y_t M_t Y_t'] \bar{\Sigma}_t' = \\ &= \bar{\Sigma}_t Y_t [M_t - M_t Y_t' H' B_{t+1}^{-1} H Y_t M_t] Y_t' \bar{\Sigma}_t' \end{aligned}$$

Por tanto:

$$Y_{t+1} = [\Sigma - K_t H] Y_t \quad [1.4.1.6]$$

$$M_{t+1} = M_t - M_t Y_t' H' B_{t+1}^{-1} H Y_t M_t \quad [1.4.1.7]$$

Las expresiones [1.4.1.6] y [1.4.1.7] son las ecuaciones de Chandrasekhar que propagan las matrices factores Y_t y M_t , respectivamente. Es importante señalar que la matriz M_{t+1} es también definida mediante la ecuación de Riccati [1.4.1.7], pero la diferencia entre la matriz $P_{t+1/t}$ del filtro de Kalman y la matriz M_{t+1} , es que la dimensión de ésta última es generalmente menor a la de $P_{t+1/t}$ y, lo que es más importante, M_{t+1} no necesita ser semidefinida positiva.

La factorización de la matriz $V_{t+1/t}$ no es única. Así, se demuestra que dicha matriz es también equivalente a la expresión:

$$V_{t+1/t} = \bar{\Sigma}_{t+1} [V_{t/t-1} + V_{t/t-1}H'B_t^{-1}HV_{t/t-1}] \bar{\Sigma}_{t+1}' \quad [1.4.1.8]$$

donde $\bar{\Sigma}_{t+1} = [\Sigma - K_{t+1} H]$.

La equivalencia entre [1.4.1.5] y [1.4.1.8] se demuestra en el lema de Morf et al (1974). Y en este caso, la correspondiente factorización de [1.4.1.8] da lugar a las ecuaciones de Chandrasekhar [1.4.1.6'] y [1.4.1.7'], alternativas a [1.4.1.6] y [1.4.1.7], respectivamente.

$$Y_t = [\Phi - K_t H] Y_{t-1} \quad [1.4.1.6']$$

$$M_t = M_{t-1} + M_{t-1} Y_{t-1}' H' B_{t-1}^{-1} H Y_{t-1} M_{t-1} \quad [1.4.1.7']$$

Resumiendo, el filtro queda definido por las ecuaciones:

$$\begin{aligned} \tilde{z}_t &= z_t - H \hat{x}_{t/t-1} - D u_t \\ \hat{x}_{t+1/t} &= \Phi \hat{x}_{t/t-1} + \Gamma u_t + K_t \tilde{z}_t \\ B_t &= B_{t-1} + H Y_{t-1} M_{t-1} Y_{t-1}' H' \\ K_t &= K_{t-1} + [\Phi - K_{t-1} H] Y_{t-1} M_{t-1} Y_{t-1}' H' B_{t-1}^{-1} \\ Y_t &= [\Phi - K_{t-1} H] Y_{t-1} \\ M_t &= M_{t-1} - M_{t-1} Y_{t-1}' H' B_{t-1}^{-1} H Y_{t-1} M_{t-1} \end{aligned}$$

o bien:

$$\begin{aligned} \tilde{z}_t &= z_t - H \hat{x}_{t/t-1} - D u_t \\ \hat{x}_{t+1/t} &= \Phi \hat{x}_{t/t-1} + \Gamma u_t + K_t \tilde{z}_t \\ B_t &= B_{t-1} + H Y_{t-1} M_{t-1} Y_{t-1}' H' \\ K_t &= K_{t-1} + [\Phi - K_{t-1} H] Y_{t-1} M_{t-1} Y_{t-1}' H' B_t^{-1} \\ Y_t &= [\Phi - K_t H] Y_{t-1} \\ M_t &= M_{t-1} + M_{t-1} Y_{t-1}' H' B_{t-1}^{-1} H Y_{t-1} M_{t-1} \end{aligned}$$

y necesita las condiciones iniciales B_0 , K_0 , Y_0 y M_0 . Estas varían dependiendo de si trabajamos con un modelo estacionario o no.



Condiciones iniciales para modelos estacionarios.

En este caso, las condiciones iniciales B_0 y K_0 vienen dadas por:

$$B_0 = HP_0H' + CRC' \quad [1.4.1.9]$$

$$K_0 = [\Phi P_0 H' + ESC'] B_0^{-1} \quad [1.4.1.10]$$

donde [1.4.1.9] es la ecuación que propaga la matriz de covarianzas del vector de observación y [1.4.1.10] es la expresión de la ganancia del filtro de Kalman en el instante inicial.

Para obtener las condiciones iniciales Y_0 y M_0 , se ha de factorizar la matriz V_0 en la forma:

$$V_0 = \bar{Y} \bar{M} \bar{Y}' \quad [1.4.1.11]$$

donde \bar{Y} es una matriz de orden $(n \times \alpha)$ y \bar{M} una matriz cuadrada de orden α con la siguiente estructura:

$$\bar{M} = \begin{bmatrix} M_+ & 0 \\ 0 & M_- \end{bmatrix} \quad \text{siendo } M_+ > 0 \text{ y } M_- < 0.$$

El valor de α viene dado por las condiciones iniciales, ya que $\alpha = \text{rang}(V_0)$.

Además, la ecuación de Riccati del filtro de Kalman expresa que la matriz $P_{t+1/t}$ se propaga de acuerdo con:

$$P_{t+1/t} = \Phi P_{t/t-1} \Phi' + EQE' - K_t B_t K_t' \quad [1.4.1.12]$$

y sustituyendo [1.4.1.12] en [1.4.1.11], se llega a:

$$V_0 = P_{1/0} - P_0 = \Phi P_0 \Phi' + EQE' - K_0 B_0 K_0' - P_0 \quad [1.4.1.13]$$

El valor de P_0 en el caso de modelos estacionarios viene dado por la solución única de la ecuación de Lyapunov⁷:

$$P_0 = \Phi P_0 \Phi' + EQE'$$

Sabiendo esto, la expresión [1.4.1.13] pasa a ser:

$$V_0 = -K_0 B_0 K_0' = -[\Phi P_0 H' + ESC'] [HP_0 H' + CRC']^{-1} [\Phi P_0 H' + ESC']'$$

Por tanto, $\alpha = \text{rang}(V_0) \leq \min(n, m)$ y puede tomarse la siguiente factorización de V_0 :

$$Y_0 = \Phi P_0 H' + ESC' \approx K_0 B_0 \quad [1.4.1.14]$$

$$M_0 = -[HP_0 H' + CRC']^{-1} = -B_0^{-1} \quad [1.4.1.15]$$

Como se observa, la dimensión de M_t viene determinada por la condición inicial de esa misma matriz, de modo que, cuando se trabaja con modelos estacionarios, α es igual a la dimensión del vector de observación del sistema. Además, dado que $M_0 < 0$, si se utiliza la ecuación [1.4.1.7] en su propagación, se conservaría dicha propiedad en todo instante de tiempo. Esto sería conveniente en el caso de procesos estacionarios⁸.

⁷ Si el modelo en espacio de los estados es estacionario, es decir, si los autovalores de Φ son menores que la unidad en valor absoluto, entonces cuando $t \rightarrow \infty$, x_t tenderá a un valor estacionario. En estas condiciones, P_t puede aproximarse por una matriz constante P que es la solución única de esa ecuación de Lyapunov.

⁸ Por supuesto, otra posible factorización de V_0 sería:

$$Y_0 = -K_0$$

$$M_0 = -B_0$$

donde también se comienza con una matriz M_t definida negativa.

Condiciones iniciales para modelos no estacionarios.

En el caso de modelos no estacionarios, podemos considerar \mathbf{x}_0 como un vector de parámetros desconocidos pero constantes. Por tanto, es razonable suponer que la matriz inicial $\mathbf{P}_0 = \mathbf{0}$. Las condiciones iniciales [1.4.1.9] y [1.4.1.10] quedan simplificadas a las siguientes expresiones:

$$\mathbf{B}_0 = \mathbf{C}\mathbf{R}\mathbf{C}' \quad [1.4.1.16]$$

$$\mathbf{K}_0 = \mathbf{E}\mathbf{S}\mathbf{C}'\mathbf{B}_0^{-1} \quad [1.4.1.17]$$

y [1.4.1.13] se convierte ahora en la ecuación:

$$\begin{aligned} \mathbf{V}_0 &= \mathbf{E}\mathbf{Q}\mathbf{E}' - \mathbf{K}_0\mathbf{B}_0\mathbf{K}_0' = \mathbf{E}\mathbf{Q}\mathbf{E}' - (\mathbf{E}\mathbf{S}\mathbf{C}')\mathbf{B}_0^{-1}(\mathbf{E}\mathbf{S}\mathbf{C}')' = \\ &= \mathbf{E}\mathbf{Q}\mathbf{E}' - \mathbf{E}\mathbf{S}\mathbf{C}'(\mathbf{C}\mathbf{R}\mathbf{C}')^{-1}\mathbf{C}\mathbf{S}'\mathbf{E}' = \\ &= \mathbf{E}[\mathbf{Q} - \mathbf{S}\mathbf{C}'(\mathbf{C}\mathbf{R}\mathbf{C}')^{-1}\mathbf{C}\mathbf{S}']\mathbf{E}' \end{aligned}$$

En este caso, se observa que $\alpha = \text{rang}(\mathbf{V}_0) \leq \min(n,p)$ si la matriz \mathbf{E} tiene rango completo. Una factorización de esta última expresión proporciona las condiciones iniciales:

$$\mathbf{Y}_0 = \mathbf{E} \quad [1.4.1.18]$$

$$\mathbf{M}_0 = \mathbf{Q} - \mathbf{S}\mathbf{C}'(\mathbf{C}\mathbf{R}\mathbf{C}')^{-1}\mathbf{C}\mathbf{S}' \quad [1.4.1.19]$$

donde ahora α es igual a la dimensión de la matriz de covarianzas del ruido de estado del sistema. Puesto que $\mathbf{M}_0 > \mathbf{0}$, no se aseguraría el carácter de esta matriz si se utilizase para su propagación [1.4.1.7], siendo por ello conveniente utilizar en su lugar la ecuación [1.4.1.7']⁹.

Por tanto, si se obtienen las condiciones iniciales de las matrices \mathbf{M}_t e \mathbf{Y}_t para los casos de modelos estacionarios y no

⁹ En el caso particular en que no exista correlación entre los ruidos de estado y de observación ($\mathbf{S} = \mathbf{0}$), la matriz $\mathbf{M}_0 = \mathbf{Q}$ si \mathbf{Q} tiene rango completo. Y para conservar el carácter de definida positiva de \mathbf{M}_t , también sería aconsejable utilizar la versión [1.4.1.7'].

estacionarios, no es necesario calcular el rango de la matriz V_0 , rango que define el valor teórico de la dimensión α .

Cálculo de la matriz de covarianzas del vector de estado.

Como ya se ha señalado anteriormente, en este tipo de algoritmos no es necesario calcular en ningún momento la matriz $P_{t+1/t}$. Sin embargo, puede obtenerse fácilmente a partir de [1.4.1.1] y [1.4.1.2]:

$$P_{t+1/t} = P_{t/t-1} + Y_t M_t Y_t'$$

y por recurrencia:

$$P_{t+1/t} = P_0 + \sum_{j=0}^t Y_j M_j Y_j' \quad [1.4.1.20]$$

Por último, se ofrece un resumen del algoritmo obtenido en los dos casos particulares estudiados.

El filtro a propagar es:

$$\tilde{z}_t = z_t - H\hat{x}_{t/t-1} - Du_t$$

$$B_t = B_{t-1} + HY_{t-1}M_{t-1}Y_{t-1}'H'$$

$$K_t = K_{t-1} + [\Phi - K_{t-1}H]Y_{t-1}M_{t-1}Y_{t-1}'H'B_t^{-1}$$

$$\hat{x}_{t+1/t} = \Phi\hat{x}_{t/t-1} + \Gamma u_t + K_t\tilde{z}_t$$

$$M_t = M_{t-1} - M_{t-1}Y_{t-1}'H'B_t^{-1}HY_{t-1}M_{t-1}$$

$$Y_t = [\Phi - K_{t-1}H]Y_{t-1}$$

sujeito a las condiciones iniciales:

$$B_0 = HP_0H' + CRC'$$

$$K_0 = [\Phi P_0H' + ESC']B_0^{-1}$$

$$Y_0 = \Phi P_0H' + ESC'$$

$$M_0 = -B_0^{-1}$$

Filtro de Chandrasekhar para modelos estacionarios.

El filtro a propagar es:

$$\tilde{z}_t = z_t - H\hat{x}_{t/t-1} - Du_t$$

$$B_t = B_{t-1} + HY_{t-1}M_{t-1}Y_{t-1}'H'$$

$$K_t = K_{t-1} + [\Phi - K_{t-1}H]Y_{t-1}M_{t-1}Y_{t-1}'H'B_t^{-1}$$

$$\hat{x}_{t+1/t} = \Phi\hat{x}_{t/t-1} + \Gamma u_t + K_t\tilde{z}_t$$

$$M_t = M_{t-1} + M_{t-1}Y_{t-1}'H'B_{t-1}^{-1}HY_{t-1}M_{t-1}$$

$$Y_t = [\Phi - K_tH]Y_{t-1}$$

sujeito a las condiciones iniciales:

$$B_0 = CRC'$$

$$K_0 = ESC'B_0^{-1}$$

$$Y_0 = E$$

$$M_0 = Q - SC'(CRC')CS'$$

Filtro de Chandrasekhar para modelos no estacionarios.

Otras formas alternativas del algoritmo de Chandrasekhar.

Es posible obtener distintas formas del filtro sustituyendo las ecuaciones [1.4.1.3] y [1.4.1.7'] por ecuaciones que propagan B_t^{-1} y M_t^{-1} . Estas se obtienen fácilmente aplicando el Lema de Inversión de Matrices¹⁰ a la parte derecha de las expresiones [1.4.1.3] y [1.4.1.7'], respectivamente. Las ecuaciones resultantes serían:

$$M_t^{-1} = M_{t-1}^{-1} - Y_{t-1}'H'B_t^{-1}HY_{t-1} \quad [1.4.1.21]$$

$$B_t^{-1} = B_{t-1}^{-1} - B_{t-1}^{-1}HY_{t-1}M_{t-1}^{-1}M_{t-1}^{-1}Y_{t-1}'H'B_{t-1}^{-1} \quad [1.4.1.22]$$

y llamando $M_{t-1}^{-1} = J_{t-1}$, se obtiene la llamada forma simétrica del algoritmo de Chandrasekhar:

$$J_t = J_{t-1} - Y_{t-1}'H'B_t^{-1}HY_{t-1} \quad [1.4.1.23]$$

$$B_t = B_{t-1} + HY_{t-1}J_{t-1}^{-1}Y_{t-1}'H' \quad [1.4.1.24]$$

Otras dos versiones alternativas consisten en utilizar [1.4.1.7'] y [1.4.1.22] ó [1.4.1.21] y [1.4.1.22]. Esta última forma no es operativa pues no se pueden obtener B_t^{-1} y M_t^{-1} , dado el carácter simultáneo de [1.4.1.21] y [1.4.1.22]. Este problema se resuelve sustituyendo en [1.4.1.21] M_t por la expresión:

$$M_t = M_{t-1} + M_{t-1}Y_{t-1}'H'B_{t-1}^{-1}HY_{t-1}M_{t-1}$$

Pero esto da la forma original [1.4.1.3] y [1.4.1.7'].

La formulación simétrica puede ser más ventajosa computacionalmente cuando $\alpha < m$. Sin embargo, en el caso en que $\alpha > m$ resulta más económico utilizar las expresiones [1.4.1.3] y [1.4.1.7'].

¹⁰ $(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)DA^{-1}$ siendo A y C matrices no singulares.

1.4.2. Propiedades del filtro de Chandrasekhar.

- El filtro de Chandrasekhar proporciona una estimación recursiva óptima del vector de estado de un sistema. La optimalidad se debe al empleo de la ganancia del filtro de Kalman, expresada en términos de las matrices factores Y_t y M_t .
- Este filtro es un tipo de algoritmo de la raíz cuadrada de la matriz de incrementos de $P_{t+1/t}$, siendo $Y_t M_t^{\frac{1}{2}}$ una posible raíz cuadrada de dicha matriz.
- Los filtros de Chandrasekhar no necesitan en sí mismos una ecuación de propagación de la matriz de covarianzas $P_{t+1/t}$, aunque puede recuperarse su valor cada vez que se procesa una nueva observación a partir de [1.4.1.20]. Por tanto, es posible calcular la ganancia del filtro sin calcular la matriz de covarianzas del error de estimación de x_t .
- La principal característica de estos algoritmos es su considerable ventaja computacional cuando la dimensión del vector de estado es mayor que la del vector de observación. Esto es evidente, ya que en la ecuación de Riccati hay que actualizar del orden de n^2 elementos, mientras que las ecuaciones de Chandrasekhar sólo exigen actualizar del orden de n elementos distintos.
- A pesar de haber derivado los filtros de Chandrasekhar para un sistema invariante, pueden obtenerse también las ecuaciones correspondientes a un modelo con matrices de covarianzas Q , R y S variantes en el tiempo. Asimismo, también puede derivarse el correspondiente algoritmo de la matriz de información (P_t^{-1}) a partir de la factorización de Chandrasekhar.

- Si el modelo es gaussiano, el filtro de Chandrasekhar es, al igual que el filtro de Kalman, una vía abierta para calcular el valor exacto de la función de verosimilitud de la muestra.

1.5. COMPARACION DE LA FACTORIZACION DE CHANDRASEKHAR CON OTRAS FACTORIZACIONES ALTERNATIVAS.

Como se ha visto, existe una relación entre los filtros de Potter, Bierman y los de Chandrasekhar. Sin embargo, las diferencias entre ambas aproximaciones son la causa de que, en este trabajo, se haya elegido la última:

- (a) Los algoritmos de Potter y Bierman factorizan la propia matriz de covarianzas \mathbf{P} , mientras que el filtro de Chandrasekhar utiliza una factorización de la matriz de incrementos (o de variaciones) de \mathbf{P} .
- (b) El algoritmo de Potter sólo puede aplicarse cuando el vector de variables observadas es un escalar, mientras que la factorización de Chandrasekhar se obtiene partiendo de un vector de orden m de variables observadas.
- (c) La factorización de Potter supone actualizar una matriz \mathbf{S} que contiene n^2 elementos distintos. Para reducir cálculos se puede transformar \mathbf{S} en una matriz triangular y se tendrían que actualizar sólo $n(n+1)/2$ elementos. Sin embargo, el problema en este caso es la necesidad de calcular n raíces cuadradas en lugar de una, pero el consumo de tiempo para calcular una raíz cuadrada es mucho mayor que el utilizado en cualquier otra operación aritmética. La factorización UDU' evita el cálculo de n raíces cuadradas siendo este algoritmo casi tan eficiente como el filtro de Kalman. Los algoritmos de Chandrasekhar suponen el cálculo de

raíces cuadradas de $V_{t+1/t}$ de orden $(n \times \alpha)$ y como frecuentemente en Econometría α es mucho menor que n , parecen ser los más ventajosos computacionalmente cuando se aplican a este campo.

Resumiendo, el algoritmo de Potter resuelve los problemas numéricos inherentes al filtro de Kalman convencional, pero tiene el inconveniente de ser más costoso desde el punto de vista computacional. El algoritmo de Bierman es también numéricamente estable y casi tan eficiente como el de Kalman. Y por último, los filtros de Chandrasekhar heredan las propiedades de estabilidad numérica de los anteriores, con la ventaja de ser los más económicos en cuanto a requerimientos de cálculo cuando son aplicados a los casos más usuales en Econometría. En concreto, en modelos estacionarios, su uso será muy ventajoso cuando la dimensión del vector de estado del sistema sea considerablemente mayor a la del vector de observación.

Los aspectos computacionales y de robustez numérica de todos estos algoritmos han sido estudiados en detalle por Verhaegen y Van Dooren (1986, 1988). En concreto, concluyen que en el caso en que la matriz $P_{0/-1}$ sea nula, los filtros de Chandrasekhar tienen la máxima precisión numérica. Esta robustez puede degradarse si dicha condición inicial no es nula, pero aún en este caso, pueden ser preferibles dichos algoritmos a otros más estables por su eficiencia computacional. Esta ventaja en cuanto a cómputo sólo se mantiene para el caso de sistemas invariantes en el tiempo. Las modificaciones existentes que permiten matrices de covarianzas de los ruidos no constantes suponen, en la práctica, una mayor carga computacional. Otro aspecto crucial en la propagación de errores numéricos en estos filtros es el número de condición de la matriz B_t . Esto puede resolverse, en muchos casos, aplicando un cambio de escala a los datos de las variables observadas.

CAPITULO 2

**ESTIMACION MAXIMO-VEROSIMIL DE MODELOS ECONOMETRICOS EN FORMA DE
ESPACIO DE LOS ESTADOS MEDIANTE LOS FILTROS DE CHANDRASEKHAR**

2.1. MODELOS ECONOMETRICOS EN FORMA DE ESPACIO DE LOS ESTADOS.

Sea la forma estructural de cualquier modelo econométrico lineal dinámico:

$$\bar{F}(B)Y_t = \bar{G}(B)u_t + \bar{\Lambda}(B)\epsilon_t \quad [2.1.1]$$

en donde Y_t es un vector de m variables endógenas,
 u_t es un vector de r variables exógenas,
 ϵ_t es un vector de m variables ruido blanco, es decir:

$$E(\epsilon_t) = 0 \quad \forall t = 1, 2, \dots, n$$

$$E(\epsilon_{t_1}\epsilon_{t_2}') = \Sigma_\epsilon \delta_{t_1, t_2} \quad \delta_{t_1, t_2} = \begin{cases} 1 & \text{si } t_1 = t_2 \\ 0 & \text{en otro caso} \end{cases}$$

Los elementos de las matrices \bar{F} , \bar{G} y $\bar{\Lambda}$ de dimensiones $(m \times m)$, $(m \times r)$ y $(m \times m)$, respectivamente, son polinomios en el operador retardo B y se definen como:

$$\bar{F}(B) = \bar{F}_0 + \bar{F}_1 B + \dots + \bar{F}_p B^p$$

$$\bar{G}(B) = \bar{G}_0 + \bar{G}_1 B + \dots + \bar{G}_q B^q$$

$$\bar{\Lambda}(B) = \bar{\Lambda}_0 + \bar{\Lambda}_1 B + \dots + \bar{\Lambda}_s B^s$$

La forma reducida correspondiente a [2.1.1] es:

$$F(B)Y_t = G(B)u_t + \Lambda(B)\epsilon_t \quad [2.1.2]$$

donde $F(B) = \bar{F}_{0-1} \bar{F}(B)$

$$G(B) = \bar{F}_{0-1} \bar{G}(B)$$

$$\Lambda(B) = \bar{F}_{0-1} \bar{\Lambda}(B)$$

suponiendo que existe la inversa de la matriz \bar{F}_0 . El modelo dado por [2.1.2], puede expresarse en forma de espacio de los estados mediante las ecuaciones:

$$x_{t+1} = \Phi x_t + \Gamma u_t + E w_t$$

$$z_t = H x_t + D u_t + C v_t$$

en donde:

$$\Phi = \begin{bmatrix} -F_1 & I & \dots & 0 \\ -F_2 & 0 & I & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ -F_{k-1} & 0 & \dots & \dots & I \\ -F_k & 0 & \dots & \dots & 0 \end{bmatrix} \quad \Gamma = \begin{bmatrix} G_1 - F_1 G_0 \\ G_2 - F_2 G_0 \\ \dots & \dots & \dots \\ G_k - F_k G_0 \end{bmatrix} \quad E = \begin{bmatrix} \Lambda_1 - F_1 \Lambda_0 \\ \Lambda_2 - F_2 \Lambda_0 \\ \dots & \dots & \dots \\ \Lambda_k - F_k \Lambda_0 \end{bmatrix}$$

[2.1.4]

$$H = [I \ 0 \ \dots \ \dots \ 0] \quad D = G_0 \quad C = \Lambda_0 \quad v_t = w_t = \epsilon_t$$

[2.1.5]

$$Q = R = S = \Sigma_\epsilon \quad z_t = y_t \quad [2.1.6]$$

El vector de estado tendrá, por tanto, k componentes siendo $k = \max(p, q, s)$. Por ejemplo, la representación en espacio de los estados, según la norma de conversión anterior, de un proceso ARMA(2,1) con término constante es:

$$(1 - \phi_1 B - \phi_2 B^2) y_t = \mu + (1 - \theta B) a_t$$

$$\begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{bmatrix} \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} + \begin{bmatrix} \phi_1 \mu \\ \phi_2 \mu \end{bmatrix} + \begin{bmatrix} -\theta + \phi_1 \\ \phi_2 \end{bmatrix} a_t$$

$$z_t = [1 \ 0] \begin{bmatrix} x_{1,t} \\ x_{2,t} \end{bmatrix} + \mu + a_t$$

Existen múltiples representaciones en forma de espacio de los estados para un mismo modelo [vid. Aoki (1987)]. Sin embargo, en este trabajo se ha elegido la de dimensión mínima del vector de estado por diversas razones:

- Esta representación supone una ventaja computacional con respecto a otras de dimensión no mínima.
- Si el modelo econométrico es expresado en esta forma, son directa y fácilmente aplicables las ecuaciones de Chandrasekhar derivadas en el Capítulo 1.
- Más adelante, se plantea un algoritmo de estimación óptima del vector de estado inicial (vid. Apéndice B). Esta estimación será más eficiente si la dimensión de dicho vector es mínima y además, permite evitar casos en los que no está identificado un subconjunto de condiciones iniciales.

Rosenbrock (1970) muestra que una representación en forma de espacio de los estados es mínima si es observable y alcanzable¹. Puede demostrarse que la representación dada por [2.1.4]-[2.1.6] cumple ambas propiedades.

¹ Un modelo en espacio de los estados es observable si, dada una muestra de tamaño suficiente n , el vector de estado puede determinarse de forma exacta. Esta propiedad equivale a imponer que:

$$\text{rang}[\mathbf{H}' \mid (\mathbf{F}')\mathbf{H}' \mid (\mathbf{F}')^2\mathbf{H}' \mid \dots \mid (\mathbf{F}')^{n-1}\mathbf{H}'] = n$$

Un modelo en espacio de los estados es controlable si, en cualquier instante de tiempo, existe una secuencia de vectores de control que permite alcanzar un vector de estado arbitrariamente escogido. Esta propiedad equivale a la siguiente condición:

$$\text{rang}[\mathbf{E} \mid \mathbf{F}\mathbf{E} \mid \mathbf{F}^2\mathbf{E} \mid \dots \mid \mathbf{F}^{n-1}\mathbf{E}] = n$$

que se cumple siempre que el $\text{rang}(\mathbf{E}) = n$.

Por último, hay que señalar que las ecuaciones [2.1.4]-[2.1.6] constituyen una regla convencional para convertir cualquier modelo econométrico lineal en forma de espacio de los estados. Además, se han desarrollado procedimientos para el tratamiento de modelos con errores de observación en las variables output y/o inputs, variables observadas con distinta periodicidad, etc [vid. Terceiro (1990)]. También existen reglas para convertir modelos en forma de espacio de los estados a modelos econométricos en forma final [vid. Aoki (1987)].

2.2. ESTIMACION MAXIMO-VEROSIMIL DE UN MODELO ECONOMETRICO LINEAL EN ESPACIO DE LOS ESTADOS MEDIANTE UN FILTRO DE CHANDRA-SEKHAR.

En este apartado se describe la forma de estimar, por máxima verosimilitud exacta, los parámetros desconocidos de un modelo econométrico lineal expresado en forma de espacio de los estados. Para ello se utiliza un algoritmo de Chandrasekhar. El problema que nos ocupa es estimar un vector de parámetros θ , en el que se incluyen todos los elementos desconocidos de las matrices Φ , Γ , E , H , D , C , Q , R y S .

2.2.1. La función de verosimilitud del modelo.

El método de estimación por máxima verosimilitud está basado en el supuesto de que las observaciones $\{z_1, z_2, \dots, z_t\}$ se distribuyen de forma idéntica e independiente. En este caso, la función de densidad conjunta se define como el producto de las funciones de densidad de cada observación. Es decir:

$$L(\mathbf{z}, \theta) = \prod_{t=1}^n p(z_t)$$

Una vez que las observaciones se han realizado, la función $L(\mathbf{z}, \theta)$ se interpreta como la función de verosimilitud y el

estimador máximo-verosímil se obtiene maximizando esa función con respecto al vector de parámetros desconocidos θ .

Sin embargo, cuando se trabaja con modelos de series temporales no es creíble suponer que las observaciones son independientes entre sí. Por esta razón, no puede usarse la función de densidad anterior y la idea central para resolver este problema consiste en aplicar sucesivamente la definición de función de densidad condicional [vid. Schweppe (1965) y Harvey (1981)]:

$$L(\mathbf{z}, \theta) = \prod_{t=1}^n p(\mathbf{z}_t / \mathbf{z}^{t-1}, \theta) \quad [2.2.1.1]$$

donde $p(\mathbf{z}_t / \mathbf{z}^{t-1}, \theta)$ es la distribución de \mathbf{z}_t condicionada por el conjunto de información disponible hasta el instante $t-1$. Es decir, $\mathbf{z}^{t-1} = (z_1, \dots, z_{t-1})$. Si los términos de error \mathbf{v}_t y \mathbf{w}_t y el vector de estado inicial tienen distribuciones normales, la distribución de \mathbf{z}_t condicionada por \mathbf{z}^{t-1} es también normal, lo cual permite caracterizar analíticamente la expresión [2.2.1.1].

Por tanto, para este tipo de modelos, $p(\mathbf{z}_t / \mathbf{z}^{t-1}, \theta)$ estará completamente caracterizada a través de las expresiones de un filtro de Chandrasekhar. Es decir, si se denota:

$$E(\mathbf{z}_t / \mathbf{z}^{t-1}, \theta) = \hat{\mathbf{z}}_{t/t-1}$$

$$\text{COV}(\mathbf{z}_t / \mathbf{z}^{t-1}, \theta) = E[(\mathbf{z}_t - \hat{\mathbf{z}}_{t/t-1})(\mathbf{z}_t - \hat{\mathbf{z}}_{t/t-1})' / \mathbf{z}^{t-1}, \theta] = \mathbf{B}_t$$

se tiene que:

$$p(\mathbf{z}_t / \mathbf{z}^{t-1}, \theta) = (2\pi)^{-m/2} |\mathbf{B}_t|^{-1/2} e^{(-1/2) \tilde{\mathbf{z}}_t' \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t}$$

siendo:

$$\tilde{\mathbf{z}}_t = \mathbf{z}_t - \hat{\mathbf{z}}_{t/t-1}$$

Se trata de maximizar la función [2.2.1.1] con respecto al vector θ . Y maximizar esa función equivale a minimizar la expresión:

$$\ell(\theta) = \frac{1}{2} \sum_{t=1}^n (\ln 2\pi + \ln |B_t| + \tilde{z}_t' B_t^{-1} \tilde{z}_t) \quad [2.2.1.2]$$

Esta función de verosimilitud se conoce en la literatura como la función de verosimilitud de la descomposición del error de predicción. Esto es así, porque \tilde{z}_t se interpreta como un vector de errores de predicción. Bajo la hipótesis de normalidad, la distribución de dicho vector es:

$$\tilde{z}_t \sim \text{NID}(0, B_t)$$

Es importante señalar que \tilde{z}_t se distribuye $\text{NID}(0, B_t)$ si las matrices del sistema son fijas y conocidas. Esto no es cierto si algún(os) elemento(s) de dichas matrices son parámetros desconocidos y son reemplazados por sus estimadores². La evaluación de la función de verosimilitud dada por [2.2.1.2] exige el cálculo en cada instante de tiempo de \tilde{z}_t y B_t . Ambas expresiones pueden obtenerse a través de un filtro de Kalman o de un filtro de Chandrasekhar.

2.2.2. Planteamiento del problema de optimización.

El problema de optimización consiste en maximizar la función de verosimilitud del modelo con respecto al vector de parámetros a estimar, sujeto a las restricciones que impone un filtro de Chandrasekhar. Dado que las ecuaciones de este filtro cambian, dependiendo de si el modelo es estacionario o no, se ha de plantear un problema distinto para cada tipo de modelo.

² Vid. Watanabe (1985).

Modelos estacionarios.

El problema planteado se reduce a minimizar la función [2.2.1.2] sujeto a las restricciones que impone un filtro de Chandrasekhar para modelos estacionarios. Es decir:

$$\text{MIN} \quad \ell(\theta) = \frac{1}{2} \sum_{t=1}^n (\ln 2\pi + \ln |B_t| + \tilde{z}_t' B_t^{-1} \tilde{z}_t)$$

s.a

$$\tilde{z}_t = z_t - H\hat{x}_{t/t-1} - Du_t \quad [2.2.2.1]$$

$$B_t = B_{t-1} + HY_{t-1}M_{t-1}Y_{t-1}'H' \quad [2.2.2.2]$$

$$K_t = K_{t-1} + [\Phi - K_{t-1}H]Y_{t-1}M_{t-1}Y_{t-1}'H'B_t^{-1} \quad [2.2.2.3]$$

$$\hat{x}_{t+1/t} = \Phi\hat{x}_{t/t-1} + \Gamma u_t + K_t \tilde{z}_t \quad [2.2.2.4]$$

$$M_t = M_{t-1} - M_{t-1}Y_{t-1}'H'B_t^{-1}HY_{t-1}M_{t-1} \quad [2.2.2.5]$$

$$Y_t = [\Phi - K_{t-1}H]Y_{t-1} \quad [2.2.2.6]$$

donde todas las restricciones son de tipo igualdad y se pueden sustituir en la expresión de la función a minimizar.

Modelos no estacionarios.

Cuando se trabaje con modelos no estacionarios, el problema planteado consiste en minimizar la misma función objetivo, teniendo en cuenta las ecuaciones que forman un filtro de Chandrasekhar para modelos no estacionarios.

$$\text{MIN} \quad \ell(\theta) = \frac{1}{2} \sum_{t=1}^n (\ln 2\pi + \ln |B_t| + \tilde{z}_t' B_t^{-1} \tilde{z}_t)$$

s.a

$$\tilde{z}_t = z_t - H\hat{x}_{t/t-1} - Du_t \quad [2.2.2.7]$$

$$B_t = B_{t-1} + HY_{t-1}M_{t-1}Y_{t-1}'H' \quad [2.2.2.8]$$

$$K_t = K_{t-1} + [\Phi - K_t H]Y_{t-1}M_{t-1}Y_{t-1}'H'B_{t-1} \quad [2.2.2.9]$$

$$\hat{x}_{t+1/t} = \Phi \hat{x}_{t/t-1} + \Gamma u_t + K_t \tilde{z}_t \quad [2.2.2.10]$$

$$M_t = M_{t-1} + M_{t-1} Y_{t-1}' H' B_{t-1}^{-1} H Y_{t-1} M_{t-1} \quad [2.2.2.11]$$

$$Y_t = [\Phi - K_t H] Y_{t-1} \quad [2.2.2.12]$$

Como se observa, lo único que cambia con respecto al primer problema planteado son las dos últimas restricciones, que propagan las matrices factores M_t e Y_t .

Con respecto a los dos problemas anteriores, cabe señalar dos aspectos importantes:

- (a) Las restricciones del problema no se imponen directamente sobre el vector θ , sino sobre variables que son funciones implícitas de los componentes de ese vector.
- (b) La estimación de los parámetros del modelo requiere la estimación del vector de estado dada por el filtro de Chandrasekhar. Como no se conoce el verdadero valor de θ , sino una estimación $\hat{\theta}$, las estimaciones del vector de estado derivadas del filtro de Chandrasekhar serán subóptimas con respecto al verdadero modelo. Sin embargo, si se utiliza un procedimiento de optimización tal que $\hat{\theta} \rightarrow \theta$, entonces tanto el proceso de innovaciones como su matriz de covarianzas tenderán a su verdadero valor.

La minimización de [2.2.1.2] sujeta a cualquiera de los dos filtros de Chandrasekhar considerados, exige utilizar algún procedimiento de optimización numérica. Existen diversos algoritmos que pueden usarse, pero todos necesitan las primeras derivadas de la función objetivo y algunos, sus segundas derivadas. Por ello, en los siguientes apartados se obtiene la expresión analítica exacta del gradiente y del hessiano de la función a minimizar, tanto para modelos estacionarios como no estacionarios.

2.2.3. Expresión analítica exacta del gradiente de la función de verosimilitud.

Derivando la función objetivo [2.2.1.2] con respecto al vector de parámetros a estimar, se obtiene la expresión del término general del gradiente³:

$$\frac{\partial \ell(\theta)}{\partial \theta_i} = \sum_{t=1}^n \left\{ \frac{1}{2} \operatorname{tr} \left(\mathbf{B}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_i} \right) + \tilde{\mathbf{z}}_t' \mathbf{B}_t^{-1} \frac{\partial \tilde{\mathbf{z}}_t}{\partial \theta_i} - \frac{1}{2} \tilde{\mathbf{z}}_t' \mathbf{B}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_i} \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t \right\} \quad [2.2.3.1]$$

En [2.2.3.1] se desconocen las derivadas parciales del proceso de innovaciones y de su matriz de varianzas-covarianzas, con respecto a cada componente del vector θ . Para obtenerlas, se utilizan las ecuaciones del filtro de Chandrasekhar descrito en el Capítulo 1.

Gradiente para modelos estacionarios.

Para evaluar el gradiente de la función de verosimilitud en cada vector de parámetros, hay que derivar las expresiones del algoritmo de Chandrasekhar para procesos estacionarios:

³ Es necesario recordar los siguientes resultados sobre derivadas matriciales:

$$\frac{\partial \ln |\mathbf{A}(\beta)|}{\partial \beta} = \operatorname{tr} \left\{ \mathbf{A}(\beta)^{-1} \frac{\partial \mathbf{A}(\beta)}{\partial \beta} \right\}$$

$$\frac{\partial \mathbf{A}(\beta)^{-1}}{\partial \beta} = - \mathbf{A}(\beta)^{-1} \frac{\partial \mathbf{A}(\beta)}{\partial \beta} \mathbf{A}(\beta)^{-1}$$

$$\frac{\partial \bar{z}_t}{\partial \theta_i} = - \frac{\partial H}{\partial \theta_i} \hat{x}_{t/t-1} - H \frac{\partial \hat{x}_{t/t-1}}{\partial \theta_i} - \frac{\partial D}{\partial \theta_i} u_t \quad [2.2.3.2]$$

$$\frac{\partial \hat{x}_{t/t-1}}{\partial \theta_i} = \frac{\partial \bar{\bar{x}}_t}{\partial \theta_i} \hat{x}_{t/t-1} + \bar{\bar{x}}_t \frac{\partial \hat{x}_{t/t-1}}{\partial \theta_i} + \frac{\partial \bar{\Gamma}_t}{\partial \theta_i} u_t + \frac{\partial K_t}{\partial \theta_i} z_t \quad [2.2.3.3]$$

donde se define:

$$\bar{\bar{x}}_t = \bar{x} - K_t H \quad \bar{\Gamma}_t = \Gamma - K_t D$$

$$\begin{aligned} \frac{\partial K_t}{\partial \theta_i} = & \frac{\partial K_{t-1}}{\partial \theta_i} + \frac{\partial \bar{\bar{x}}_{t-1}}{\partial \theta_i} Y_{t-1} M_{t-1} Y_{t-1}' H' B_t^{-1} + \bar{\bar{x}}_{t-1} \left\{ \frac{\partial Y_{t-1} M_{t-1} Y_{t-1}' H' +}{\partial \theta_i} \right. \\ & + Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' H' + Y_{t-1} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' + Y_{t-1} M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} - \\ & \left. - Y_{t-1} M_{t-1} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} \right\} B_t^{-1} \end{aligned} \quad [2.2.3.4]$$

$$\begin{aligned} \frac{\partial B_t}{\partial \theta_i} = & \frac{\partial B_{t-1}}{\partial \theta_i} + \left\{ \frac{\partial H}{\partial \theta_i} Y_{t-1} + H \frac{\partial Y_{t-1}}{\partial \theta_i} \right\} M_{t-1} Y_{t-1}' H' + \\ & + H Y_{t-1} \left\{ \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' H' + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' + M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} \right\} \end{aligned} \quad [2.2.3.5]$$

$$\begin{aligned}
\frac{\partial M_t}{\partial \theta_i} &= \frac{\partial M_{t-1}}{\partial \theta_i} - \left\{ \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} \right\} H' B_t^{-1} H Y_{t-1} M_{t-1} - \\
&- M_{t-1} Y_{t-1}' H' B_t^{-1} \left\{ \frac{\partial H}{\partial \theta_i} Y_{t-1} M_{t-1} + H \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} + H Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} \right\} - \\
&- M_{t-1} Y_{t-1}' \left\{ \frac{\partial H'}{\partial \theta_i} - H' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} \right\} B_t^{-1} H Y_{t-1} M_{t-1}
\end{aligned} \tag{2.2.3.6}$$

$$\frac{\partial Y_t}{\partial \theta_i} = \frac{\partial \bar{y}_{t-1}}{\partial \theta_i} Y_{t-1} + \bar{y}_{t-1} \frac{\partial Y_{t-1}}{\partial \theta_i} \tag{2.2.3.7}$$

Para poder propagar las expresiones [2.2.3.2]-[2.2.3.7], se necesitan las primeras derivadas de las condiciones iniciales de B_t , K_t , Y_t y M_t . En el caso de modelos estacionarios, estas derivadas vienen dadas por las siguientes ecuaciones:

$$\frac{\partial B_0}{\partial \theta_i} = \frac{\partial H}{\partial \theta_i} P H' + H \frac{\partial P}{\partial \theta_i} H' + H P \frac{\partial H'}{\partial \theta_i} + \frac{\partial C}{\partial \theta_i} R C' + C \frac{\partial R}{\partial \theta_i} C' + C R \frac{\partial C'}{\partial \theta_i} \tag{2.2.3.8}$$

$$\begin{aligned}
\frac{\partial K_0}{\partial \theta_i} &= \left\{ \frac{\partial \bar{z}}{\partial \theta_i} P H' + \bar{z} \frac{\partial P}{\partial \theta_i} H' + \bar{z} P \frac{\partial H'}{\partial \theta_i} + \frac{\partial E}{\partial \theta_i} S C' + E \frac{\partial S}{\partial \theta_i} C' + E S \frac{\partial C'}{\partial \theta_i} \right\} B_0^{-1} - \\
&- K_0 \frac{\partial B_0}{\partial \theta_i} B_0^{-1}
\end{aligned} \tag{2.2.3.9}$$

$$\frac{\partial Y_0}{\partial \theta_i} = \frac{\partial \bar{z}}{\partial \theta_i} P H' + \bar{z} \frac{\partial P}{\partial \theta_i} H' + \bar{z} P \frac{\partial H'}{\partial \theta_i} + \frac{\partial E}{\partial \theta_i} S C' + E \frac{\partial S}{\partial \theta_i} C' + E S \frac{\partial C'}{\partial \theta_i} \tag{2.2.3.10}$$

$$\frac{\partial \mathbf{M}_0}{\partial \theta_i} = \mathbf{B}_0^{-1} \frac{\partial \mathbf{B}_0}{\partial \theta_i} \mathbf{B}_0^{-1} \quad [2.2.3.11]$$

siendo \mathbf{P} la solución de la ecuación de Lyapunov:

$$\mathbf{P} = \Phi \mathbf{P} \Phi' + \mathbf{E} \mathbf{Q} \mathbf{E}'$$

y la derivada primera de \mathbf{P} con respecto al vector de parámetros θ , tiene la expresión:

$$\frac{\partial \mathbf{P}}{\partial \theta_i} = \frac{\partial \Phi}{\partial \theta_i} \mathbf{P} \Phi' + \Phi \frac{\partial \mathbf{P}}{\partial \theta_i} \Phi' + \Phi \mathbf{P} \frac{\partial \Phi'}{\partial \theta_i} + \frac{\partial \mathbf{E}}{\partial \theta_i} \mathbf{Q} \mathbf{E}' + \mathbf{E} \frac{\partial \mathbf{Q}}{\partial \theta_i} \mathbf{E}' + \mathbf{E} \mathbf{Q} \frac{\partial \mathbf{E}'}{\partial \theta_i} \quad [2.2.3.12]$$

Gradiente para modelos no estacionarios.

En el caso de modelos no estacionarios, la expresión del gradiente diferirá en las ecuaciones [2.2.3.6] y [2.2.3.7], así como en las condiciones iniciales de \mathbf{K}_t , \mathbf{B}_t , \mathbf{Y}_t y \mathbf{M}_t .

$$\begin{aligned} \frac{\partial \mathbf{M}_t}{\partial \theta_i} &= \frac{\partial \mathbf{M}_{t-1}}{\partial \theta_i} + \left\{ \frac{\partial \mathbf{M}_{t-1}}{\partial \theta_i} \mathbf{Y}_{t-1}' + \mathbf{M}_{t-1} \frac{\partial \mathbf{Y}_{t-1}'}{\partial \theta_i} \right\} \mathbf{H}' \mathbf{B}_{t-1}^{-1} \mathbf{H} \mathbf{Y}_{t-1} \mathbf{M}_{t-1} + \\ &+ \mathbf{M}_{t-1} \mathbf{Y}_{t-1}' \left\{ \frac{\partial \mathbf{H}'}{\partial \theta_i} - \mathbf{H}' \mathbf{B}_{t-1}^{-1} \frac{\partial \mathbf{B}_{t-1}}{\partial \theta_i} \right\} \mathbf{B}_{t-1}^{-1} \mathbf{H} \mathbf{Y}_{t-1} \mathbf{M}_{t-1} + \\ &+ \mathbf{M}_{t-1} \mathbf{Y}_{t-1}' \mathbf{H}' \mathbf{B}_{t-1}^{-1} \left\{ \frac{\partial \mathbf{H}}{\partial \theta_i} \mathbf{Y}_{t-1} \mathbf{M}_{t-1} + \mathbf{H} \frac{\partial \mathbf{Y}_{t-1}}{\partial \theta_i} \mathbf{M}_{t-1} \right\} + \\ &+ \mathbf{M}_{t-1} \mathbf{Y}_{t-1}' \mathbf{H}' \mathbf{B}_{t-1}^{-1} \mathbf{H} \mathbf{Y}_{t-1} \frac{\partial \mathbf{M}_{t-1}}{\partial \theta_i} \end{aligned} \quad [2.2.3.13]$$

$$\frac{\partial \mathbf{Y}_t}{\partial \theta_i} = \frac{\partial \bar{\mathbf{z}}_{t-1}}{\partial \theta_i} \mathbf{Y}_{t-1} + \bar{\mathbf{z}}_{t-1} \frac{\partial \mathbf{Y}_{t-1}}{\partial \theta_i} \quad [2.2.3.14]$$

Las condiciones iniciales de las ecuaciones [2.2.3.13] y [2.2.3.14] vienen dadas por las expresiones:

$$\frac{\partial \mathbf{B}_0}{\partial \theta_i} = \frac{\partial \mathbf{C}}{\partial \theta_i} \mathbf{R} \mathbf{C}' + \mathbf{C} \frac{\partial \mathbf{R}}{\partial \theta_i} \mathbf{C}' + \mathbf{C} \mathbf{R} \frac{\partial \mathbf{C}'}{\partial \theta_i} \quad [2.2.3.15]$$

$$\frac{\partial \mathbf{K}_0}{\partial \theta_i} = \left\{ \frac{\partial \mathbf{E}}{\partial \theta_i} \mathbf{S} \mathbf{C}' + \mathbf{E} \frac{\partial \mathbf{S}}{\partial \theta_i} \mathbf{C}' + \mathbf{E} \mathbf{S} \frac{\partial \mathbf{C}'}{\partial \theta_i} \right\} \mathbf{B}_0^{-1} - \mathbf{E} \mathbf{S} \mathbf{C}' \mathbf{B}_0^{-1} \frac{\partial \mathbf{B}_0}{\partial \theta_i} \mathbf{B}_0^{-1} \quad [2.2.3.16]$$

$$\frac{\partial \mathbf{Y}_0}{\partial \theta_i} = \frac{\partial \mathbf{E}}{\partial \theta_i} \quad [2.2.3.17]$$

$$\begin{aligned} \frac{\partial \mathbf{M}_0}{\partial \theta_i} = & \frac{\partial \mathbf{Q}}{\partial \theta_i} - \left\{ \frac{\partial \mathbf{S}}{\partial \theta_i} \mathbf{C}' + \mathbf{S} \frac{\partial \mathbf{C}'}{\partial \theta_i} \right\} \mathbf{B}_0^{-1} \mathbf{C} \mathbf{S}' + \mathbf{S} \mathbf{C}' \mathbf{B}_0^{-1} \left\{ \frac{\partial \mathbf{B}_0}{\partial \theta_i} \mathbf{B}_0^{-1} \mathbf{C} \mathbf{S}' - \right. \\ & \left. - \frac{\partial \mathbf{C}}{\partial \theta_i} \mathbf{S}' - \mathbf{C} \frac{\partial \mathbf{S}'}{\partial \theta_i} \right\} \end{aligned} \quad [2.2.3.18]$$

2.2.4. Expresión analítica exacta del hessiano de la función de verosimilitud.

El término general del hessiano de la función de verosimilitud tiene la expresión:

$$\begin{aligned}
 G_{ij} = \frac{\partial^2 \ell(\theta)}{\partial \theta_j \partial \theta_i} = \sum_{t=1}^n \left\{ -\frac{1}{2} \operatorname{tr} \left[B_t^{-1} \frac{\partial B_t}{\partial \theta_j} B_t^{-1} \frac{\partial B_t}{\partial \theta_i} \right] + \right. \\
 + \frac{1}{2} \operatorname{tr} \left[B_t^{-1} \frac{\partial^2 B_t}{\partial \theta_i \partial \theta_j} \right] + \frac{\partial z_t'}{\partial \theta_j} B_t^{-1} \frac{\partial z_t}{\partial \theta_i} - \\
 - z_t' B_t^{-1} \frac{\partial B_t}{\partial \theta_j} B_t^{-1} \frac{\partial z_t}{\partial \theta_i} + z_t' B_t^{-1} \frac{\partial^2 z_t}{\partial \theta_j \partial \theta_i} - \\
 - z_t' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} B_t^{-1} \frac{\partial z_t}{\partial \theta_j} + z_t' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} B_t^{-1} \frac{\partial B_t}{\partial \theta_j} B_t^{-1} z_t - \\
 \left. - \frac{1}{2} z_t' B_t^{-1} \frac{\partial^2 B_t}{\partial \theta_j \partial \theta_i} B_t^{-1} z_t \right\} \quad [2.2.4.1]
 \end{aligned}$$

donde se observa que son necesarias las expresiones de las derivadas segundas del proceso de innovaciones y de su matriz de varianzas-covarianzas. Estas se consiguen derivando por segunda vez las ecuaciones del filtro de Chandrasekhar. Como ocurría con el gradiente, las expresiones del hessiano diferirán en modelos estacionarios y no estacionarios.

Hessiano para modelos estacionarios.

Las expresiones necesarias para evaluar, en cada vector de parámetros, el hessiano de la función de verosimilitud son:

$$\begin{aligned} \frac{\partial^2 \bar{z}_t}{\partial \theta_j \partial \theta_i} = & - \frac{\partial^2 H}{\partial \theta_j \partial \theta_i} \hat{x}_{t/t-1} - \frac{\partial H}{\partial \theta_i} \frac{\partial \hat{x}_{t/t-1}}{\partial \theta_j} - \frac{\partial H}{\partial \theta_j} \frac{\partial \hat{x}_{t/t-1}}{\partial \theta_i} - \\ & - H \frac{\partial^2 \hat{x}_{t/t-1}}{\partial \theta_j \partial \theta_i} - \frac{\partial^2 D}{\partial \theta_j \partial \theta_i} u_t \end{aligned} \quad [2.2.4.2]$$

$$\begin{aligned} \frac{\partial^2 \hat{x}_{t/t-1}}{\partial \theta_i \partial \theta_j} = & \frac{\partial^2 \bar{s}_t}{\partial \theta_i \partial \theta_j} \hat{x}_{t/t-1} + \frac{\partial \bar{s}_t}{\partial \theta_i} \frac{\partial \hat{x}_{t/t-1}}{\partial \theta_i \partial \theta_j} + \frac{\partial \bar{s}_t}{\partial \theta_j} \frac{\partial \hat{x}_{t/t-1}}{\partial \theta_i} + \\ & + \bar{s}_t \frac{\partial^2 \hat{x}_{t/t-1}}{\partial \theta_j \partial \theta_i} + \frac{\partial^2 \bar{\Gamma}_t}{\partial \theta_j \partial \theta_i} u_t + \frac{\partial^2 K_t}{\partial \theta_j \partial \theta_i} z_t \end{aligned} \quad [2.2.4.3]$$

Se requiere ahora la expresión de la derivada segunda de la matriz K_t :

$$\begin{aligned} \frac{\partial^2 K_t}{\partial \theta_j \partial \theta_i} = & \frac{\partial^2 K_{t-1}}{\partial \theta_j \partial \theta_i} + \frac{\partial^2 \bar{s}_{t-1}}{\partial \theta_j \partial \theta_i} Y_{t-1} M_{t-1} Y_{t-1}' H' B_t^{-1} + \frac{\partial \bar{s}_{t-1}}{\partial \theta_i} [A_{jt} + L_{jt}] + \\ & + \frac{\partial \bar{s}_{t-1}}{\partial \theta_j} [A_{it} + L_{it}] + \bar{s}_{t-1} \left\{ \frac{\partial A_{it}}{\partial \theta_j} + \frac{\partial L_{it}}{\partial \theta_j} \right\} \end{aligned} \quad [2.2.4.4]$$

siendo:

$$A_{jt} = \frac{\partial Y_{t-1}}{\partial \theta_j} M_{t-1} Y_{t-1}' H' B_t^{-1} \quad [2.2.4.5]$$

$$A_{it} = \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} Y_{t-1}' H' B_t^{-1} \quad [2.2.4.6]$$

$$\begin{aligned}
L_{jt} &= Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' H' B_t^{-1} + Y_{t-1} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} H' B_t^{-1} + \\
&+ Y_{t-1} M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} B_t^{-1} - Y_{t-1} M_{t-1} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_j} B_t^{-1}
\end{aligned}$$

[2.2.4.7]

$$\begin{aligned}
L_{it} &= Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' H' B_t^{-1} + Y_{t-1} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' B_t^{-1} + \\
&+ Y_{t-1} M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} B_t^{-1} - Y_{t-1} M_{t-1} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} B_t^{-1}
\end{aligned}$$

[2.2.4.8]

$$\begin{aligned}
\frac{\partial A_{jt}}{\partial \theta_j} &= \frac{\partial^2 Y_{t-1}}{\partial \theta_j \partial \theta_i} M_{t-1} Y_{t-1}' H' B_t^{-1} + \frac{\partial Y_{t-1}}{\partial \theta_i} \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' H' B_t^{-1} + \\
&+ \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} H' B_t^{-1} + \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} B_t^{-1} - \\
&- \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_j} B_t^{-1}
\end{aligned}$$

[2.2.4.9]

$$\begin{aligned}
\frac{\partial L_{it}}{\partial \theta_j} = & \frac{\partial Y_{t-1}}{\partial \theta_j} \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' H' B_t^{-1} + Y_{t-1} \frac{\partial^2 M_{t-1}}{\partial \theta_j \partial \theta_i} Y_{t-1}' H' B_t^{-1} - \\
& - Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_j} B_t^{-1} + \frac{\partial Y_{t-1}}{\partial \theta_j} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' B_t^{-1} + \\
& + Y_{t-1} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} \frac{\partial H'}{\partial \theta_j} B_t^{-1} - Y_{t-1} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' B_t^{-1} \frac{\partial B_t}{\partial \theta_j} B_t^{-1} + \\
& + \frac{\partial Y_{t-1}}{\partial \theta_j} M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} B_t^{-1} + Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} B_t^{-1} + \\
& + Y_{t-1} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} \frac{\partial H'}{\partial \theta_i} B_t^{-1} + Y_{t-1} M_{t-1} Y_{t-1}' \frac{\partial^2 H'}{\partial \theta_j \partial \theta_i} B_t^{-1} + \\
& + Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} \frac{\partial Y_{t-1}'}{\partial \theta_j} H' B_t^{-1} + Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} B_t^{-1} + \\
& + Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_j} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' B_t^{-1} + Y_{t-1} M_{t-1} \frac{\partial^2 Y_{t-1}'}{\partial \theta_j \partial \theta_i} H' B_t^{-1} - \\
& - Y_{t-1} M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} B_t^{-1} \frac{\partial B_t}{\partial \theta_j} B_t^{-1} - \frac{\partial M_{it}}{\partial \theta_j}
\end{aligned} \tag{2.2.4.10}$$

donde:

$$M_{it} = Y_{t-1} M_{t-1} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} B_t^{-1} \tag{2.2.4.11}$$

$$\begin{aligned}
\frac{\partial M_{jt}}{\partial \theta_j} = & \left\{ \frac{\partial Y_{t-1}}{\partial \theta_j} M_{t-1} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} + Y_{t-1} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} H' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} \right\} B_t^{-1} + \\
& + \left\{ Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} + Y_{t-1} M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} B_t^{-1} \frac{\partial B_t}{\partial \theta_i} \right\} B_t^{-1} - \\
& - Y_{t-1} M_{t-1} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_j} B_t^{-1} \frac{\partial B_t}{\partial \theta_i} B_t^{-1} - \\
& - Y_{t-1} M_{t-1} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} B_t^{-1} \frac{\partial B_t}{\partial \theta_j} B_t^{-1} + \\
& + Y_{t-1} M_{t-1} Y_{t-1}' H' B_t^{-1} \frac{\partial^2 B_t}{\partial \theta_j \partial \theta_i} B_t^{-1} \quad [2.2.4.12]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 B_t}{\partial \theta_j \partial \theta_i} = & \frac{\partial^2 B_{t-1}}{\partial \theta_j \partial \theta_i} + \left\{ \frac{\partial^2 H}{\partial \theta_j \partial \theta_i} Y_{t-1} + \frac{\partial H}{\partial \theta_i} \frac{\partial Y_{t-1}}{\partial \theta_j} \right\} M_{t-1} Y_{t-1}' H' + \\
& + \frac{\partial H}{\partial \theta_i} Y_{t-1} \left\{ \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' H' + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} H' + M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} \right\} + \\
& + \frac{\partial H}{\partial \theta_j} \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} Y_{t-1}' H' + H \frac{\partial^2 T_{jt}}{\partial \theta_j \partial \theta_i} H' + \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} + \\
& + \frac{\partial H}{\partial \theta_j} Y_{t-1} M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} + H T_{jt} \frac{\partial H'}{\partial \theta_i} + H Y_{t-1} M_{t-1} Y_{t-1}' \frac{\partial^2 H'}{\partial \theta_j \partial \theta_i} + \\
& + \frac{\partial H}{\partial \theta_j} Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' H' + H Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} + \\
& + \frac{\partial H}{\partial \theta_j} Y_{t-1} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' + H Y_{t-1} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} \frac{\partial H'}{\partial \theta_j} \quad [2.2.4.13]
\end{aligned}$$

en donde:

$$T_{jt} = \frac{\partial Y_{t-1}}{\partial \theta_j} M_{t-1} Y_{t-1}' + Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' + Y_{t-1} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} \quad [2.2.4.14]$$

$$\begin{aligned}
\frac{\partial^2 T_{it}}{\partial \theta_j \partial \theta_i} &= \frac{\partial Y_{t-1}}{\partial \theta_i} \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' + \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} + \frac{\partial Y_{t-1}}{\partial \theta_j} \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' + \\
&+ Y_{t-1} \frac{\partial^2 M_{t-1}}{\partial \theta_j \partial \theta_i} Y_{t-1}' + Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} \frac{\partial Y_{t-1}'}{\partial \theta_j} + \frac{\partial Y_{t-1}}{\partial \theta_j} M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} \\
&+ Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_j} \frac{\partial Y_{t-1}'}{\partial \theta_i} + \frac{\partial^2 Y_{t-1}}{\partial \theta_j \partial \theta_i} M_{t-1} Y_{t-1}' \quad [2.2.4.15]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 M_t}{\partial \theta_j \partial \theta_i} &= \frac{\partial^2 M_{t-1}}{\partial \theta_j \partial \theta_i} - \left\{ \frac{\partial^2 M_{t-1}}{\partial \theta_j \partial \theta_i} Y_{t-1}' - \frac{\partial M_{t-1}}{\partial \theta_i} \frac{\partial Y_{t-1}'}{\partial \theta_j} \right\} H' B_t^{-1} H Y_{t-1} M_{t-1} - \\
&- \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' H' B_t^{-1} C_{it} - [D_{it} + F_{it} - G_{it}] B_t^{-1} H Y_{t-1} M_{t-1} - \\
&- M_{t-1} Y_{t-1}' H' B_t^{-1} I_{it} - J_{it} B_t^{-1} H \left\{ \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} + Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} \right\} + \\
&+ M_{t-1} Y_{t-1}' H' B_t^{-1} \frac{\partial B_t}{\partial \theta_i} B_t^{-1} C_{it} - H_{it} B_t^{-1} \frac{\partial H}{\partial \theta_i} Y_{t-1} M_{t-1} - \\
&- \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' \left\{ \frac{\partial H'}{\partial \theta_j} - H' B_t^{-1} \frac{\partial B_t}{\partial \theta_j} \right\} B_t^{-1} H Y_{t-1} M_{t-1} - \\
&- M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' B_t^{-1} C_{it} - M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} B_t^{-1} C_{it} + \\
&+ M_{t-1} Y_{t-1}' H' B_t^{-1} \left\{ O_{it} - U_{it} + \frac{\partial B_t}{\partial \theta_j} B_t^{-1} H Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} \right\} \quad [2.2.4.16]
\end{aligned}$$

siendo:

$$C_{it} = \frac{\partial H}{\partial \theta_j} Y_{t-1} M_{t-1} + H \frac{\partial Y_{t-1}}{\partial \theta_j} M_{t-1} + H Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_j} \quad [2.2.4.17]$$

$$\begin{aligned}
D_{it} &= \frac{\partial M_{t-1}}{\partial \theta_j} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' + M_{t-1} \frac{\partial^2 Y_{t-1}'}{\partial \theta_j \partial \theta_i} H' + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} \frac{\partial H'}{\partial \theta_j} + \\
&+ M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' B_{t-1} \frac{\partial B_t}{\partial \theta_j}
\end{aligned} \tag{2.2.4.18}$$

$$\begin{aligned}
F_{it} &= \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} \frac{\partial H'}{\partial \theta_i} + M_{t-1} Y_{t-1}' \frac{\partial^2 H'}{\partial \theta_j \partial \theta_i} - \\
&- M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} B_{t-1} \frac{\partial B_t}{\partial \theta_j}
\end{aligned} \tag{2.2.4.19}$$

$$\begin{aligned}
G_{it} &= \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' H' B_{t-1} \frac{\partial B_t}{\partial \theta_i} + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} H' B_{t-1} \frac{\partial B_t}{\partial \theta_i} + \\
&+ M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} B_{t-1} \frac{\partial B_t}{\partial \theta_i} - M_{t-1} Y_{t-1}' H' B_{t-1} \frac{\partial B_t}{\partial \theta_j} B_{t-1} \frac{\partial B_t}{\partial \theta_i} + \\
&+ M_{t-1} Y_{t-1}' H' B_{t-1} \frac{\partial^2 B_t}{\partial \theta_j \partial \theta_i} - M_{t-1} Y_{t-1}' H' B_{t-1} \frac{\partial B_t}{\partial \theta_i} B_{t-1} \frac{\partial B_t}{\partial \theta_j}
\end{aligned} \tag{2.2.4.20}$$

$$\begin{aligned}
H_{it} &= \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' H' + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} H' + M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} - \\
&- M_{t-1} Y_{t-1}' H' B_{t-1} \frac{\partial B_t}{\partial \theta_j}
\end{aligned} \tag{2.2.4.21}$$

$$I_{it} = \frac{\partial^2 H}{\partial \theta_j \partial \theta_i} Y_{t-1} M_{t-1} + \frac{\partial H}{\partial \theta_i} \frac{\partial Y_{t-1}'}{\partial \theta_j} M_{t-1} + \frac{\partial H}{\partial \theta_i} Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_j} \tag{2.2.4.22}$$

$$J_{it} = \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' H' + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} H' + M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} \tag{2.2.4.23}$$

$$\begin{aligned}
O_{it} = & \frac{\partial B_t}{\partial \theta_j} B_{t-1} H \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} - \frac{\partial H}{\partial \theta_j} \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} - H \frac{\partial^2 Y_{t-1}}{\partial \theta_j \partial \theta_i} M_{t-1} - \\
& - H \frac{\partial Y_{t-1}}{\partial \theta_i} \frac{\partial M_{t-1}}{\partial \theta_j}
\end{aligned} \quad [2.2.4.24]$$

$$U_{it} = \frac{\partial H}{\partial \theta_j} Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} + H \frac{\partial Y_{t-1}}{\partial \theta_j} \frac{\partial M_{t-1}}{\partial \theta_i} + H Y_{t-1} \frac{\partial^2 M_{t-1}}{\partial \theta_j \partial \theta_i} \quad [2.2.4.25]$$

$$\begin{aligned}
\frac{\partial^2 Y_t}{\partial \theta_j \partial \theta_i} = & \frac{\partial^2 \bar{y}_{t-1}}{\partial \theta_j \partial \theta_i} Y_{t-1} + \frac{\partial \bar{y}_{t-1}}{\partial \theta_i} \frac{\partial Y_{t-1}}{\partial \theta_j} + \frac{\partial \bar{y}_{t-1}}{\partial \theta_j} \frac{\partial Y_{t-1}}{\partial \theta_i} + \\
& + \bar{y}_{t-1} \frac{\partial^2 Y_{t-1}}{\partial \theta_j \partial \theta_i}
\end{aligned} \quad [2.2.4.26]$$

La propagación de las expresiones [2.2.4.4]-[2.2.4.26] exige calcular las derivadas segundas de las condiciones iniciales de K_t , B_t , Y_t y M_t . Para modelos estacionarios, estas derivadas son:

$$\begin{aligned}
\frac{\partial^2 B_0}{\partial \theta_j \partial \theta_i} = & \frac{\partial^2 H}{\partial \theta_j \partial \theta_i} P H' + \frac{\partial H}{\partial \theta_i} \frac{\partial P}{\partial \theta_j} H' + \frac{\partial H}{\partial \theta_i} P \frac{\partial H'}{\partial \theta_j} + \frac{\partial H}{\partial \theta_j} \frac{\partial P}{\partial \theta_i} H' + \\
& + H \frac{\partial^2 P}{\partial \theta_j \partial \theta_i} H' + H \frac{\partial P}{\partial \theta_i} \frac{\partial H'}{\partial \theta_j} + \frac{\partial H}{\partial \theta_j} P \frac{\partial H'}{\partial \theta_i} + H \frac{\partial P}{\partial \theta_j} \frac{\partial H'}{\partial \theta_i} + \\
& + H P \frac{\partial^2 H'}{\partial \theta_j \partial \theta_i} + \frac{\partial^2 C}{\partial \theta_j \partial \theta_i} R C' + \frac{\partial C}{\partial \theta_i} \frac{\partial R}{\partial \theta_j} C' + \frac{\partial C}{\partial \theta_i} R \frac{\partial C'}{\partial \theta_j} + \\
& + \frac{\partial^2 C}{\partial \theta_j \partial \theta_i} R C' + \frac{\partial C}{\partial \theta_i} \frac{\partial R}{\partial \theta_j} C' + \frac{\partial C}{\partial \theta_i} R \frac{\partial C'}{\partial \theta_j} + \frac{\partial C}{\partial \theta_j} \frac{\partial R}{\partial \theta_i} C' + \\
& + C \frac{\partial^2 R}{\partial \theta_j \partial \theta_i} C' + C \frac{\partial R}{\partial \theta_i} \frac{\partial C'}{\partial \theta_j} + \frac{\partial C}{\partial \theta_j} R \frac{\partial C'}{\partial \theta_i} + C \frac{\partial R}{\partial \theta_j} \frac{\partial C'}{\partial \theta_i} + \\
& + C R \frac{\partial^2 C'}{\partial \theta_j \partial \theta_i}
\end{aligned} \quad [2.2.4.27]$$

donde la derivada segunda de la matriz P tiene la expresión:

$$\begin{aligned}
 \frac{\partial^2 P}{\partial \theta_j \partial \theta_i} &= \frac{\partial^2 \Phi}{\partial \theta_j \partial \theta_i} P \Phi' + \frac{\partial \Phi}{\partial \theta_i} \frac{\partial P}{\partial \theta_j} \Phi' + \frac{\partial \Phi}{\partial \theta_i} P \frac{\partial \Phi'}{\partial \theta_j} + \frac{\partial \Phi}{\partial \theta_j} \frac{\partial P}{\partial \theta_i} \Phi' + \\
 &+ \Phi \frac{\partial^2 P}{\partial \theta_j \partial \theta_i} \Phi' + \Phi \frac{\partial P}{\partial \theta_i} \frac{\partial \Phi'}{\partial \theta_j} + \frac{\partial \Phi}{\partial \theta_j} P \frac{\partial \Phi'}{\partial \theta_i} + \Phi \frac{\partial P}{\partial \theta_j} \frac{\partial \Phi'}{\partial \theta_i} + \\
 &+ \Phi P \frac{\partial^2 \Phi'}{\partial \theta_j \partial \theta_i} + \frac{\partial^2 E}{\partial \theta_j \partial \theta_i} Q E' + \frac{\partial E}{\partial \theta_i} \frac{\partial Q}{\partial \theta_j} E' + \frac{\partial E}{\partial \theta_i} Q \frac{\partial E'}{\partial \theta_j} + \\
 &+ \frac{\partial E}{\partial \theta_j} \frac{\partial Q}{\partial \theta_i} E' + E \frac{\partial^2 Q}{\partial \theta_j \partial \theta_i} E' + E \frac{\partial Q}{\partial \theta_i} \frac{\partial E'}{\partial \theta_j} + \frac{\partial E}{\partial \theta_j} Q \frac{\partial E'}{\partial \theta_i} + \\
 &+ E \frac{\partial Q}{\partial \theta_j} \frac{\partial E'}{\partial \theta_i} + E Q \frac{\partial^2 E'}{\partial \theta_j \partial \theta_i}
 \end{aligned} \tag{2.2.4.28}$$

$$\begin{aligned}
 \frac{\partial^2 K_0}{\partial \theta_j \partial \theta_i} &= \left\{ \frac{\partial^2 \Phi}{\partial \theta_j \partial \theta_i} P H' + \frac{\partial \Phi}{\partial \theta_i} \frac{\partial P}{\partial \theta_j} H' + \frac{\partial \Phi}{\partial \theta_i} P \frac{\partial H'}{\partial \theta_j} + \frac{\partial \Phi}{\partial \theta_j} \frac{\partial P}{\partial \theta_i} H' + \right. \\
 &+ \Phi \frac{\partial^2 P}{\partial \theta_j \partial \theta_i} H' + \Phi \frac{\partial P}{\partial \theta_i} \frac{\partial H'}{\partial \theta_j} + \frac{\partial \Phi}{\partial \theta_j} P \frac{\partial H'}{\partial \theta_i} + \Phi \frac{\partial P}{\partial \theta_j} \frac{\partial H'}{\partial \theta_i} + \\
 &+ \Phi P \frac{\partial^2 H'}{\partial \theta_j \partial \theta_i} + \frac{\partial^2 E}{\partial \theta_j \partial \theta_i} S C' + \frac{\partial E}{\partial \theta_i} \frac{\partial S}{\partial \theta_j} C' + \frac{\partial E}{\partial \theta_i} S \frac{\partial C'}{\partial \theta_j} + \\
 &+ \frac{\partial E}{\partial \theta_j} \frac{\partial S}{\partial \theta_i} C' + E \frac{\partial^2 S}{\partial \theta_j \partial \theta_i} C' + E \frac{\partial S}{\partial \theta_i} \frac{\partial C'}{\partial \theta_j} + \frac{\partial E}{\partial \theta_j} S \frac{\partial C'}{\partial \theta_i} + \\
 &+ \left. E \frac{\partial S}{\partial \theta_j} \frac{\partial C'}{\partial \theta_i} + E S \frac{\partial^2 C'}{\partial \theta_j \partial \theta_i} \right\} B_0^{-1} - \frac{\partial Z}{\partial \theta_i} B_0^{-1} \frac{\partial B_0}{\partial \theta_j} B_0^{-1} - \\
 &- \frac{\partial Z}{\partial \theta_j} B_0^{-1} \frac{\partial B_0}{\partial \theta_i} B_0^{-1} - Z B_0^{-1} \frac{\partial^2 B_0}{\partial \theta_j \partial \theta_i} B_0^{-1} + \\
 &+ Z B_0^{-1} \left\{ \frac{\partial B_0}{\partial \theta_j} B_0^{-1} \frac{\partial B_0}{\partial \theta_i} + \frac{\partial B_0}{\partial \theta_i} B_0^{-1} \frac{\partial B_0}{\partial \theta_j} \right\} B_0^{-1}
 \end{aligned}$$

[2.2.4.29]

siendo:

$$Z = \Phi PH' + ESC' \quad [2.2.4.30]$$

$$\frac{\partial Z}{\partial \theta_i} = \frac{\partial \Phi}{\partial \theta_i} PH' + \Phi \frac{\partial P}{\partial \theta_i} H' + \Phi P \frac{\partial H'}{\partial \theta_i} + \frac{\partial E}{\partial \theta_i} SC' + E \frac{\partial S}{\partial \theta_i} C' + ES \frac{\partial C'}{\partial \theta_i} \quad [2.2.4.31]$$

$$\frac{\partial Z}{\partial \theta_j} = \frac{\partial \Phi}{\partial \theta_j} PH' + \Phi \frac{\partial P}{\partial \theta_j} H' + \Phi P \frac{\partial H'}{\partial \theta_j} + \frac{\partial E}{\partial \theta_j} SC' + E \frac{\partial S}{\partial \theta_j} C' + ES \frac{\partial C'}{\partial \theta_j} \quad [2.2.4.32]$$

$$\begin{aligned} \frac{\partial^2 Y_0}{\partial \theta_j \partial \theta_i} &= \frac{\partial^2 \Phi}{\partial \theta_j \partial \theta_i} PH' + \frac{\partial \Phi}{\partial \theta_i} \frac{\partial P}{\partial \theta_j} H' + \frac{\partial \Phi}{\partial \theta_i} P \frac{\partial H'}{\partial \theta_j} + \frac{\partial \Phi}{\partial \theta_j} \frac{\partial P}{\partial \theta_i} H' + \\ &+ \Phi \frac{\partial^2 P}{\partial \theta_j \partial \theta_i} H' + \Phi \frac{\partial P}{\partial \theta_i} \frac{\partial H'}{\partial \theta_j} + \frac{\partial \Phi}{\partial \theta_j} P \frac{\partial H'}{\partial \theta_i} + \Phi \frac{\partial P}{\partial \theta_j} \frac{\partial H'}{\partial \theta_i} + \\ &+ \Phi P \frac{\partial^2 H'}{\partial \theta_j \partial \theta_i} + \frac{\partial^2 E}{\partial \theta_j \partial \theta_i} SC' + \frac{\partial E}{\partial \theta_i} \frac{\partial S}{\partial \theta_j} C' + \frac{\partial E}{\partial \theta_i} S \frac{\partial C'}{\partial \theta_j} + \\ &+ \frac{\partial E}{\partial \theta_j} \frac{\partial S}{\partial \theta_i} C' + E \frac{\partial^2 S}{\partial \theta_j \partial \theta_i} C' + E \frac{\partial S}{\partial \theta_i} \frac{\partial C'}{\partial \theta_j} + \frac{\partial E}{\partial \theta_j} S \frac{\partial C'}{\partial \theta_i} + \\ &+ E \frac{\partial S}{\partial \theta_j} \frac{\partial C'}{\partial \theta_i} + ES \frac{\partial^2 C'}{\partial \theta_j \partial \theta_i} \end{aligned} \quad [2.2.4.33]$$

$$\frac{\partial^2 M_0}{\partial \theta_j \partial \theta_i} = - B_0^{-1} \left\{ \frac{\partial B_0}{\partial \theta_j} B_0^{-1} \frac{\partial B_0}{\partial \theta_i} - \frac{\partial^2 B_0}{\partial \theta_j \partial \theta_i} + \frac{\partial B_0}{\partial \theta_i} B_0^{-1} \frac{\partial B_0}{\partial \theta_j} \right\} B_0^{-1} \quad [2.2.4.34]$$

Hessiano para modelos no estacionarios.

En el caso de modelos no estacionarios, las expresiones del hessiano pasan a ser las siguientes:

$$\begin{aligned}
\frac{\partial^2 M_t}{\partial \theta_j \partial \theta_i} &= \frac{\partial^2 M_{t-1}}{\partial \theta_j \partial \theta_i} + \left\{ \frac{\partial^2 M_{t-1} Y_{t-1}'}{\partial \theta_j \partial \theta_i} + \frac{\partial M_{t-1}}{\partial \theta_i} \frac{\partial Y_{t-1}'}{\partial \theta_j} \right\} H' B_{t-1}^{-1} H Y_{t-1} M_{t-1} + \\
&+ \frac{\partial M_{t-1} Y_{t-1}' H' B_{t-1}^{-1} C_{it}}{\partial \theta_i} + [\bar{D}_{it} + \bar{F}_{it} - \bar{G}_{it}] B_{t-1}^{-1} H Y_{t-1} M_{t-1} + \\
&+ M_{t-1} Y_{t-1}' H' B_{t-1}^{-1} I_{it} + J_{it} B_{t-1}^{-1} H \left\{ \frac{\partial Y_{t-1} M_{t-1}}{\partial \theta_i} + Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} \right\} - \\
&- M_{t-1} Y_{t-1}' H' B_{t-1}^{-1} \frac{\partial B_{t-1}}{\partial \theta_i} B_{t-1}^{-1} C_{it} + \bar{H}_{it} B_{t-1}^{-1} \frac{\partial H}{\partial \theta_i} Y_{t-1} M_{t-1} + \\
&+ \frac{\partial M_{t-1}}{\partial \theta_i} Y_{t-1}' \left\{ \frac{\partial H'}{\partial \theta_j} - H' B_{t-1}^{-1} \frac{\partial B_{t-1}}{\partial \theta_j} \right\} B_{t-1}^{-1} H Y_{t-1} M_{t-1} + \\
&+ M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' B_{t-1}^{-1} C_{it} + M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} B_{t-1}^{-1} C_{it} - \\
&- M_{t-1} Y_{t-1}' H' B_{t-1}^{-1} \left\{ \bar{O}_{it} - U_{it} + \frac{\partial B_{t-1}}{\partial \theta_j} B_{t-1}^{-1} H Y_{t-1} \frac{\partial M_{t-1}}{\partial \theta_i} \right\}
\end{aligned}$$

[2.2.4.35]

siendo:

$$\begin{aligned}
\bar{D}_{it} &= \frac{\partial M_{t-1}}{\partial \theta_j} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' + M_{t-1} \frac{\partial^2 Y_{t-1}}{\partial \theta_j \partial \theta_i} H' + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} \frac{\partial H'}{\partial \theta_j} - \\
&- M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_i} H' B_{t-1}^{-1} \frac{\partial B_{t-1}}{\partial \theta_j}
\end{aligned}$$

[2.2.4.36]

$$\begin{aligned}
\bar{F}_{it} &= \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} \frac{\partial H'}{\partial \theta_i} + M_{t-1} Y_{t-1}' \frac{\partial^2 H'}{\partial \theta_j \partial \theta_i} - \\
&- M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_i} B_{t-1}^{-1} \frac{\partial B_{t-1}}{\partial \theta_j}
\end{aligned}$$

[2.2.4.37]

$$\begin{aligned}
\bar{G}_{it} = & \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' H' B_{t-1-1} \frac{\partial B_{t-1}}{\partial \theta_i} + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} H' B_{t-1-1} \frac{\partial B_{t-1}}{\partial \theta_i} + \\
& + M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} B_{t-1-1} \frac{\partial B_{t-1}}{\partial \theta_i} - M_{t-1} Y_{t-1}' H' B_{t-1-1} \frac{\partial^2 B_{t-1}}{\partial \theta_j \partial \theta_i} + \\
& + M_{t-1} Y_{t-1}' H' B_{t-1-1} \frac{\partial B_{t-1}}{\partial \theta_i} B_{t-1-1} \frac{\partial B_{t-1}}{\partial \theta_j} - \\
& - M_{t-1} Y_{t-1}' H' B_{t-1-1} \frac{\partial B_{t-1}}{\partial \theta_j} B_{t-1-1} \frac{\partial B_{t-1}}{\partial \theta_i} \quad [2.2.4.38]
\end{aligned}$$

$$\begin{aligned}
\bar{H}_{it} = & \frac{\partial M_{t-1}}{\partial \theta_j} Y_{t-1}' H' + M_{t-1} \frac{\partial Y_{t-1}'}{\partial \theta_j} H' + M_{t-1} Y_{t-1}' \frac{\partial H'}{\partial \theta_j} - \\
& - M_{t-1} Y_{t-1}' H' B_{t-1-1} \frac{\partial B_{t-1}}{\partial \theta_j} \quad [2.2.4.39]
\end{aligned}$$

$$\begin{aligned}
\bar{O}_{it} = & \frac{\partial B_{t-1}}{\partial \theta_j} B_{t-1-1} H' \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} - \frac{\partial H}{\partial \theta_j} \frac{\partial Y_{t-1}}{\partial \theta_i} M_{t-1} - H' \frac{\partial^2 Y_{t-1}}{\partial \theta_j \partial \theta_i} M_{t-1} - \\
& - H' \frac{\partial Y_{t-1}}{\partial \theta_i} \frac{\partial M_{t-1}}{\partial \theta_j} \quad [2.2.4.40]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 Y_t}{\partial \theta_j \partial \theta_i} = & \frac{\partial^2 \bar{s}_t}{\partial \theta_j \partial \theta_i} Y_{t-1} + \frac{\partial \bar{s}_t}{\partial \theta_i} \frac{\partial Y_{t-1}}{\partial \theta_j} + \frac{\partial \bar{s}_t}{\partial \theta_j} \frac{\partial Y_{t-1}}{\partial \theta_i} + \bar{s}_t \frac{\partial^2 Y_{t-1}}{\partial \theta_j \partial \theta_i} \\
& [2.2.4.41]
\end{aligned}$$

Las expresiones de las segundas derivadas de las condiciones iniciales de K_t , B_t , Y_t y M_t también cambian y pasan a ser las siguientes:

$$\begin{aligned}
\frac{\partial^2 B_0}{\partial \theta_j \partial \theta_i} &= \frac{\partial^2 C}{\partial \theta_j \partial \theta_i} R C' + \frac{\partial C}{\partial \theta_i} \frac{\partial R}{\partial \theta_j} C' + \frac{\partial C}{\partial \theta_i} R \frac{\partial C'}{\partial \theta_j} + \frac{\partial C}{\partial \theta_j} \frac{\partial R}{\partial \theta_i} C' + \\
&+ C \frac{\partial R}{\partial \theta_i} \frac{\partial C'}{\partial \theta_j} + \frac{\partial C}{\partial \theta_j} R \frac{\partial C'}{\partial \theta_i} + C \frac{\partial R}{\partial \theta_j} \frac{\partial C'}{\partial \theta_i} + C R \frac{\partial^2 C'}{\partial \theta_j \partial \theta_i} + \\
&+ C \frac{\partial^2 C'}{\partial \theta_j \partial \theta_i}
\end{aligned} \tag{2.2.4.42}$$

$$\begin{aligned}
\frac{\partial^2 K_0}{\partial \theta_j \partial \theta_i} &= \left\{ \frac{\partial^2 E}{\partial \theta_j \partial \theta_i} S C' + \frac{\partial E}{\partial \theta_i} \frac{\partial S}{\partial \theta_j} C' + \frac{\partial E}{\partial \theta_i} S \frac{\partial C'}{\partial \theta_j} + \frac{\partial E}{\partial \theta_j} \frac{\partial S}{\partial \theta_i} C' + \right. \\
&+ E \frac{\partial S}{\partial \theta_i} \frac{\partial C'}{\partial \theta_j} + \frac{\partial E}{\partial \theta_j} S \frac{\partial C'}{\partial \theta_i} + E \frac{\partial S}{\partial \theta_j} \frac{\partial C'}{\partial \theta_i} + E S \frac{\partial^2 C'}{\partial \theta_j \partial \theta_i} \left. \right\} B_0^{-1} - \\
&- \bar{E} B_0^{-1} \frac{\partial^2 B_0}{\partial \theta_j \partial \theta_i} + \bar{E} B_0^{-1} \frac{\partial B_0}{\partial \theta_i} B_0^{-1} \frac{\partial B_0}{\partial \theta_j} + E \frac{\partial^2 S}{\partial \theta_j \partial \theta_i} C' B_0^{-1} - \\
&- \frac{\partial \bar{E}}{\partial \theta_i} B_0^{-1} \frac{\partial B_0}{\partial \theta_j} B_0^{-1} - \frac{\partial \bar{E}}{\partial \theta_j} B_0^{-1} \frac{\partial B_0}{\partial \theta_i} B_0^{-1} + \\
&+ \bar{E} B_0^{-1} \frac{\partial B_0}{\partial \theta_j} B_0^{-1} \frac{\partial B_0}{\partial \theta_i} B_0^{-1}
\end{aligned} \tag{2.2.4.43}$$

donde:

$$\bar{E} = E S C' \tag{2.2.4.44}$$

$$\frac{\partial \bar{E}}{\partial \theta_i} = \frac{\partial E}{\partial \theta_i} S C' + E \frac{\partial S}{\partial \theta_i} C' + E S \frac{\partial C'}{\partial \theta_i} \tag{2.2.4.45}$$

$$\frac{\partial \bar{E}}{\partial \theta_j} = \frac{\partial E}{\partial \theta_j} S C' + E \frac{\partial S}{\partial \theta_j} C' + E S \frac{\partial C'}{\partial \theta_j} \tag{2.2.4.46}$$

$$\frac{\partial^2 Y_0}{\partial \theta_j \partial \theta_i} = \frac{\partial^2 E}{\partial \theta_j \partial \theta_i} \tag{2.2.4.47}$$

$$\begin{aligned}
\frac{\partial^2 \mathbf{M}_0}{\partial \theta_j \partial \theta_i} &= \frac{\partial^2 \mathbf{Q}}{\partial \theta_j \partial \theta_i} - \frac{\partial^2 \mathbf{S}}{\partial \theta_j \partial \theta_i} \mathbf{C}' \mathbf{B}_0^{-1} \bar{\mathbf{S}}' + \frac{\partial \mathbf{S}}{\partial \theta_i} \frac{\partial \mathbf{C}'}{\partial \theta_j} \mathbf{B}_0^{-1} \bar{\mathbf{S}}' - \\
&- \bar{\mathbf{S}} \mathbf{B}_0^{-1} \frac{\partial \mathbf{B}_0}{\partial \theta_j} \mathbf{B}_0^{-1} \frac{\partial \mathbf{B}_0}{\partial \theta_i} + \bar{\mathbf{S}} \mathbf{B}_0^{-1} \frac{\partial^2 \mathbf{B}_0}{\partial \theta_j \partial \theta_i} \mathbf{B}_0^{-1} \mathbf{C} \mathbf{S}' + \\
&+ \bar{\mathbf{S}} \mathbf{B}_0^{-1} \left\{ \frac{\partial \mathbf{B}_0}{\partial \theta_j} \mathbf{B}_0^{-1} \frac{\partial \mathbf{C}}{\partial \theta_i} \mathbf{S}' - \frac{\partial^2 \mathbf{C}}{\partial \theta_j \partial \theta_i} \mathbf{S}' - \frac{\partial \mathbf{C}}{\partial \theta_i} \frac{\partial \mathbf{S}'}{\partial \theta_j} \right\} - \\
&- \left\{ \frac{\partial \mathbf{S}}{\partial \theta_j} \frac{\partial \mathbf{C}'}{\partial \theta_i} + \mathbf{S} \frac{\partial^2 \mathbf{C}'}{\partial \theta_j \partial \theta_i} - \mathbf{S} \frac{\partial \mathbf{C}'}{\partial \theta_i} \mathbf{B}_0^{-1} \frac{\partial \mathbf{B}_0}{\partial \theta_j} \right\} \mathbf{B}_0^{-1} \bar{\mathbf{S}}' - \\
&- \mathbf{S} \frac{\partial \mathbf{C}'}{\partial \theta_i} \mathbf{B}_0^{-1} \frac{\partial \bar{\mathbf{S}}'}{\partial \theta_j} + \frac{\partial \bar{\mathbf{S}}}{\partial \theta_j} \mathbf{B}_0^{-1} \frac{\partial \mathbf{B}_0}{\partial \theta_i} \mathbf{B}_0^{-1} \bar{\mathbf{S}}' + \\
&+ \bar{\mathbf{S}} \mathbf{B}_0^{-1} \frac{\partial \mathbf{B}_0}{\partial \theta_i} \mathbf{B}_0^{-1} \frac{\partial \bar{\mathbf{S}}'}{\partial \theta_j} - \frac{\partial \bar{\mathbf{S}}}{\partial \theta_j} \mathbf{B}_0^{-1} \frac{\partial \mathbf{C}}{\partial \theta_i} \mathbf{S}' - \\
&- \frac{\partial \bar{\mathbf{S}}}{\partial \theta_j} \mathbf{B}_0^{-1} \mathbf{C} \frac{\partial \mathbf{S}'}{\partial \theta_i} + \bar{\mathbf{S}} \mathbf{B}_0^{-1} \frac{\partial \mathbf{B}_0}{\partial \theta_j} \mathbf{B}_0^{-1} \mathbf{C} \frac{\partial \mathbf{S}'}{\partial \theta_i} + \\
&+ \frac{\partial \mathbf{S}}{\partial \theta_i} \mathbf{C}' \mathbf{B}_0^{-1} \left\{ \frac{\partial \mathbf{B}_0}{\partial \theta_j} \mathbf{B}_0^{-1} \bar{\mathbf{S}}' - \frac{\partial \bar{\mathbf{S}}'}{\partial \theta_j} \right\} - \\
&- \bar{\mathbf{S}} \mathbf{B}_0^{-1} \frac{\partial \mathbf{C}}{\partial \theta_j} \frac{\partial \mathbf{S}'}{\partial \theta_i} - \bar{\mathbf{S}} \mathbf{B}_0^{-1} \frac{\partial^2 \mathbf{S}'}{\partial \theta_j \partial \theta_i} - \\
&- \bar{\mathbf{S}} \mathbf{B}_0^{-1} \frac{\partial \mathbf{B}_0}{\partial \theta_i} \mathbf{B}_0^{-1} \frac{\partial \mathbf{B}_0}{\partial \theta_j} \mathbf{B}_0^{-1} \bar{\mathbf{S}}'
\end{aligned} \tag{2.2.4.48}$$

siendo:

$$\bar{\mathbf{S}} = \mathbf{S} \mathbf{C}' \tag{2.2.4.49}$$

$$\frac{\partial \bar{\mathbf{S}}}{\partial \theta_i} = \frac{\partial \mathbf{S}}{\partial \theta_i} \mathbf{C}' + \mathbf{S} \frac{\partial \mathbf{C}'}{\partial \theta_i} \tag{2.2.4.50}$$

$$\frac{\partial \bar{\mathbf{S}}}{\partial \theta_j} = \frac{\partial \mathbf{S}}{\partial \theta_j} \mathbf{C}' + \mathbf{S} \frac{\partial \mathbf{C}'}{\partial \theta_j} \tag{2.2.4.51}$$

$$\frac{\partial \bar{\mathbf{s}}'}{\partial \theta_i} = \frac{\partial \mathbf{c}}{\partial \theta_i} \mathbf{s}' + \mathbf{c} \frac{\partial \mathbf{s}'}{\partial \theta_i} \quad [2.2.4.52]$$

$$\frac{\partial \bar{\mathbf{s}}'}{\partial \theta_j} = \frac{\partial \mathbf{c}}{\partial \theta_j} \mathbf{s}' + \mathbf{c} \frac{\partial \mathbf{s}'}{\partial \theta_j} \quad [2.2.4.53]$$

2.2.5. Expresión analítica exacta de la matriz de información.

Un instrumento necesario para completar el proceso de estimación de un modelo econométrico, así como para realizar las pruebas de validación del mismo, es la matriz de covarianzas de las estimaciones de sus parámetros. Esta matriz 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. En este apartado, se sigue fundamentalmente el trabajo de Terceiro (1990). Definiendo la matriz de información como la esperanza matemática del hessiano:

$$M(\theta) = E \left\{ \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} \middle| \theta \right\}$$

una forma alternativa de escribir el hessiano de la función de verosimilitud es:

$$\begin{aligned} G_{ij} = & \sum_{t=1}^n \left\{ -\frac{1}{2} \operatorname{tr} \left[\mathbf{B}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_i} \mathbf{B}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_j} \right] + \frac{1}{2} \operatorname{tr} \left[\mathbf{B}_t^{-1} \frac{\partial^2 \mathbf{B}_t}{\partial \theta_i \partial \theta_j} \right] + \right. \\ & + \operatorname{tr} \left[\mathbf{B}_t^{-1} \frac{\partial \tilde{\mathbf{z}}_t}{\partial \theta_i} \frac{\partial \tilde{\mathbf{z}}_t'}{\partial \theta_j} \right] - \operatorname{tr} \left[\mathbf{B}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_j} \mathbf{B}_t^{-1} \frac{\partial \tilde{\mathbf{z}}_t}{\partial \theta_i} \tilde{\mathbf{z}}_t' \right] + \\ & + \operatorname{tr} \left[\mathbf{B}_t^{-1} \frac{\partial^2 \tilde{\mathbf{z}}_t}{\partial \theta_i \partial \theta_j} \tilde{\mathbf{z}}_t' \right] - \operatorname{tr} \left[\mathbf{B}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_i} \mathbf{B}_t^{-1} \frac{\partial \tilde{\mathbf{z}}_t}{\partial \theta_j} \tilde{\mathbf{z}}_t' \right] + \\ & + \operatorname{tr} \left[\mathbf{B}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_i} \mathbf{B}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_j} \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t' \right] - \\ & \left. - \frac{1}{2} \operatorname{tr} \left[\mathbf{B}_t^{-1} \frac{\partial^2 \mathbf{B}_t}{\partial \theta_i \partial \theta_j} \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t' \right] \right\} \end{aligned} \quad [2.2.5.1]$$

donde la expresión [2.2.5.1] se obtiene sin más que aplicar las propiedades del operador traza a [2.2.4.1]. El término general de la matriz de información se obtiene tomando esperanzas en [2.2.5.1] y teniendo en cuenta dos cosas: a) $E(\tilde{\mathbf{z}}_t) = \mathbf{0}$ y b) la derivada del

proceso de innovaciones no depende del propio proceso de innovaciones (vid. ecuación [2.2.3.2]):

$$[M(\theta)]_{ij} = \sum_{t=1}^n \left\{ \frac{1}{2} \operatorname{tr} \left[\mathbf{B}_{t-1} \frac{\partial \mathbf{B}_t}{\partial \theta_i} \mathbf{B}_{t-1} \frac{\partial \mathbf{B}_t}{\partial \theta_j} \right] + \operatorname{tr} \left[\mathbf{B}_{t-1} E \left[\frac{\partial \tilde{\mathbf{z}}_t}{\partial \theta_i} \frac{\partial \tilde{\mathbf{z}}_t'}{\partial \theta_j} \right] \right] \right\} \quad [2.2.5.2]$$

En la expresión [2.2.5.2] el único término desconocido es la esperanza del producto de las derivadas del proceso de innovaciones. El cálculo exacto de esa esperanza requiere propagar el siguiente sistema lineal de dimensión ampliada con matrices variantes en el tiempo:

$$\bar{\mathbf{x}}_{t+1}^a = \Phi_t^a \bar{\mathbf{x}}_t^a + \Gamma_t^a u_t \quad [2.2.5.3]$$

$$\bar{\mathbf{z}}_t^a = \mathbf{H}^a \bar{\mathbf{x}}_t^a + \mathbf{D}^a u_t \quad [2.2.5.4]$$

$$\mathbf{P}_{t+1}^a = \Phi_t^a \mathbf{P}_t^a (\Phi_t^a)' + \mathbf{K}_t^a \mathbf{B}_t (\mathbf{K}_t^a)' \quad [2.2.5.5]$$

$$\mathbf{B}_t^a = \mathbf{H}^a \mathbf{P}_t^a (\mathbf{H}^a)' \quad [2.2.5.6]$$

siendo:

$$\bar{\mathbf{x}}_t^a = E(\mathbf{x}_t^a)$$

$$\bar{\mathbf{z}}_t^a = E(\mathbf{z}_t^a)$$

$$\mathbf{P}_t^a = E[(\mathbf{x}_t^a - \bar{\mathbf{x}}_t^a)(\mathbf{x}_t^a - \bar{\mathbf{x}}_t^a)']$$

$$\mathbf{B}_t^a = E[(\mathbf{z}_t^a - \bar{\mathbf{z}}_t^a)(\mathbf{z}_t^a - \bar{\mathbf{z}}_t^a)']$$

Es decir, el sistema dado por las ecuaciones [2.2.5.3]-[2.2.5.6] propaga la media y la matriz de covarianzas de un vector de estado ampliado y de un vector de innovaciones también ampliado.

Los componentes de dichos vectores de dimensiones $3n$ y $2m$, respectivamente, son los siguientes:

$$\mathbf{x}_t^a = \begin{bmatrix} \hat{\mathbf{x}}_{t/t-1} \\ \frac{\partial \hat{\mathbf{x}}_{t/t-1}}{\partial \theta_i} \\ \frac{\partial \hat{\mathbf{x}}_{t/t-1}}{\partial \theta_j} \end{bmatrix} \quad \mathbf{z}_t^a = \begin{bmatrix} \bar{z}_t \\ \frac{\partial \bar{z}_t}{\partial \theta_i} \\ \frac{\partial \bar{z}_t}{\partial \theta_j} \end{bmatrix}$$

$$\mathbf{s}_t^a = \begin{bmatrix} \bar{s}_t & 0 & 0 \\ \frac{\partial \bar{s}_t}{\partial \theta_i} - K_t \frac{\partial H}{\partial \theta_i} & -K_t & 0 \\ \frac{\partial \bar{s}_t}{\partial \theta_j} - K_t \frac{\partial H}{\partial \theta_j} & 0 & -K_t \end{bmatrix} \quad \mathbf{K}_t^a = \begin{bmatrix} K_t \\ \frac{\partial K_t}{\partial \theta_i} \\ \frac{\partial K_t}{\partial \theta_j} \end{bmatrix}$$

$$\mathbf{\Gamma}_t^a = \begin{bmatrix} \Gamma & & \\ \frac{\partial \Gamma}{\partial \theta_i} - K_t \frac{\partial D}{\partial \theta_i} & & \\ \frac{\partial \Gamma}{\partial \theta_j} - K_t \frac{\partial D}{\partial \theta_j} & & \end{bmatrix} \quad \mathbf{H}^a = \begin{bmatrix} -\frac{\partial H}{\partial \theta_i} & -H & 0 \\ -\frac{\partial H}{\partial \theta_j} & -H & 0 \end{bmatrix} \quad \mathbf{D}^a = \begin{bmatrix} -\frac{\partial D}{\partial \theta_i} \\ -\frac{\partial D}{\partial \theta_j} \end{bmatrix}$$

y las condiciones iniciales del sistema:

$$\mathbf{x}_0^a = \begin{bmatrix} \bar{\mathbf{x}}_0 \\ 0 \\ 0 \end{bmatrix} \quad \mathbf{P}_0^a = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

La condición inicial $\bar{\mathbf{x}}_0$ será obtenida de forma óptima utilizando el algoritmo desarrollado en el Apéndice B. Por tanto, el cálculo de la matriz de información exacta sólo exige propagar simultáneamente las ecuaciones de un filtro de Chandrasekhar junto con las primeras derivadas de dichas expresiones.

No es posible obtener una factorización de los incrementos de la matriz P_t^a , al variar con el tiempo las matrices del sistema Φ_t^a y Γ_t^a . En este caso, serían más apropiados los algoritmos de factorización de la propia matriz de covarianzas⁴. Sin embargo, existe una razón práctica que justifica la propagación de P_t^a de acuerdo con [2.2.5.5]: la matriz de información es evaluada sólo en un punto (el vector de parámetros que maximiza la función de verosimilitud de la muestra). Por tanto, es menos probable que los posibles errores de redondeo en el cálculo de dicha matriz lleguen a invalidar los resultados. Esto no ocurre en el proceso de estimación de los parámetros, donde la matriz P_t es propagada para toda la muestra y evaluada en cada vector de parámetros hasta llegar al óptimo.

Deben señalarse, por último, dos aspectos importantes en el cálculo de la matriz de información:

- No es necesario calcular las derivadas segundas de la función de verosimilitud y de las ecuaciones del filtro de Chandrasekhar. Sólo se necesita información sobre sus primeras derivadas.
- Se calcula la expresión analítica exacta de la matriz de información. Normalmente, se utiliza una aproximación de la misma que casi siempre consiste en sustituir en [2.2.5.2] la esperanza matemática del producto de las derivadas del proceso de innovaciones por el producto de dichas derivadas⁵. El hecho de trabajar con una expresión exacta de dicha matriz es importante tanto en el proceso iterativo de optimización como en la etapa de validación del modelo econométrico que se ha estimado. Así, la contrastación de cualquier hipótesis sobre los parámetros estimados se puede llevar a cabo mediante el test del multiplicador de Lagrange, que utiliza tanto el

⁴ Vid. Bierman y Thornton (1980).

⁵ Vid. Watson y Engle (1983).

valor del gradiente como el de la matriz de información del modelo bajo la hipótesis nula.

2.2.6. Propiedades asintóticas del estimador máximo-verosímil.

Los estudios sobre la consistencia del estimador máximo-verosímil para observaciones dependientes se han desarrollado a partir del trabajo de Cramer (1946) o el de Wald (1949). Siguiendo al primero cabe destacar, entre otros, el trabajo de Crowder (1976) y en la línea de Wald, los trabajos de Caines y Rissanen (1974) y Heijmans y Magnus (1986a, 1986b). Todos estos estudios se diferencian básicamente en usar más o menos supuestos restrictivos y en el grado de dificultad existente a la hora de comprobar su cumplimiento.

En el marco de modelos en espacio de los estados, Pagan (1980) demostró las propiedades de consistencia y normalidad asintótica del estimador máximo-verosímil de θ , aplicando las condiciones de Crowder (1976) a un modelo sin variables exógenas y donde el vector \mathbf{z}_t es escalar. Sin embargo, es inmediato generalizar este resultado al caso en que \mathbf{z}_t sea un vector de m variables observables [vid. Pagan (1980)].

Las propiedades de consistencia y normalidad asintótica son más difíciles de demostrar para un modelo en espacio de los estados que contenga variables exógenas. Caines (1988, Capítulo 7) estudió las propiedades asintóticas del estimador máximo-verosímil de θ , considerando variables exógenas determinísticas y estocásticas. Bajo condiciones muy generales, demostró las propiedades de consistencia y normalidad asintótica del estimador por máxima verosimilitud.

Resumiendo, si θ^* es el verdadero valor del vector de parámetros θ y $\hat{\theta}_N$ el estimador máximo-verosímil de θ dada una muestra de tamaño N , las propiedades asintóticas de dicho estimador son:

(a) Consistencia:

$$\lim_{N \rightarrow \infty} \hat{\theta}_N = \theta^*$$

(b) Insensatez asintótica:

$$\lim_{N \rightarrow \infty} E[\hat{\theta}_N] = \theta^*$$

(c) Normalidad asintótica:

$$\hat{\theta}_N \underset{\text{asy}}{\sim} N(\theta^*, [M(\theta^*)]^{-1})$$

(d) Eficiencia asintótica:

$$\lim_{N \rightarrow \infty} \text{COV}(\hat{\theta}_N) = [M(\theta^*)]^{-1}$$



CAPITULO 3

SELECCION DE ALGORITMOS Y ESTRUCTURA DEL SOFTWARE

3.1. Estructura general de un algoritmo de optimización.

El problema de optimización que nos interesa resolver consiste en:

$$\underset{\theta}{\text{MIN}} \quad \ell(\theta) = \frac{1}{2} \sum_{t=1}^n (\ln 2\pi + \ln |B_t| + \tilde{z}_t' B_t^{-1} \tilde{z}_t) \quad [3.1.1]$$

sujeta a un conjunto de restricciones de tipo igualdad, dadas por las expresiones [2.2.2.1]-[2.2.2.6] o bien por las ecuaciones [2.2.2.7]-[2.2.2.12].

El primer término de esta función es una constante que, como tal, no se minimiza. El tercer sumando es una forma cuadrática que representa la suma de cuadrados de los errores de predicción (\tilde{z}_t), ponderada por la inversa de su matriz de covarianzas (B_t). Si el segundo término fuese despreciable, esta función sería cuadrática y el criterio de máxima verosimilitud sería equivalente al de mínimos cuadrados ponderados.

De la función a minimizar dada en [3.1.1], se conocen sus primeras y segundas derivadas analíticas, desarrolladas en el Capítulo 2.

Notación utilizada.

Antes de comenzar a describir los algoritmos de optimización utilizados para resolver este problema, es necesario establecer la notación usada en este capítulo:

- θ_* : vector ($n \times 1$) de estimaciones óptimas del vector de parámetros θ .
- $\ell(\theta_k)$: valor de la función objetivo en el vector de parámetros θ_k .
- θ_k : estimación del óptimo θ_* en la iteración k -ésima.

- g_k : vector gradiente (nx1) de la función objetivo evaluado en el vector de parámetros θ_k .
- G_k : hessiano de la función objetivo evaluado en el vector de parámetros θ_k .
- p_k : dirección de búsqueda del óptimo seguida en la iteración k-ésima.
- α_k : longitud del paso tomada en la iteración k-ésima.

Adoptando la notación anterior, la estructura general de la mayor parte de los algoritmos que se emplean para optimizar numéricamente [3.1.1] es la siguiente:

- 1) Fijar una estimación inicial θ_0 del óptimo θ_* y el máximo número de iteraciones permitidas, K. Ir a 2).
- 2) Comprobar si en ese punto inicial se cumple la condición de convergencia del algoritmo, previamente establecida:
 - 2.1) Si $\theta_0 \approx \theta_*$, el algoritmo ha finalizado.
 - 2.2) Si no se cumple, pasar a 3).
- 3) Calcular una dirección de búsqueda del óptimo p_k .
- 4) Calcular una longitud de paso α_k .
- 5) Calcular $\theta_{k+1} = \theta_k + \alpha_k p_k$.
- 6) Hacer $k = k+1$. Si $k \leq K$, volver al paso 2). En caso contrario, detener el proceso.

3.2. Criterios de convergencia.

Cualquier algoritmo de optimización numérica necesita un criterio de convergencia de forma que, si dicha condición se cumple, el algoritmo se detiene por considerar suficiente la aproximación al óptimo.

Los criterios de convergencia pueden hacer referencia a:

- 1) El valor de la función objetivo en la última iteración, es decir, comprobar si la mejora en esta iteración ha sido "pequeña".
- 2) El cumplimiento de las condiciones de primer orden en la última iteración.
- 3) El tamaño del paso dado en la última iteración.

Sea $\theta_{\mathbf{k}+1}$ la estimación de θ_* en la iteración $\mathbf{k}+1$ obtenida mediante cualquier algoritmo. Las condiciones correspondientes a cada uno de los criterios anteriores de finalización del algoritmo podrían ser:

- 1) $|\ell(\theta_{\mathbf{k}+1}) - \ell(\theta_{\mathbf{k}})| \leq \epsilon$, donde ϵ es un escalar positivo arbitrariamente pequeño.
- 2) $\|\mathbf{g}_{\mathbf{k}+1}\| \leq \epsilon$, o bien, $|g_{i,\mathbf{k}+1}| \leq \epsilon \quad \forall i = 1, \dots, n$.
- 3) $|\alpha_{\mathbf{k}}| \leq \epsilon$.

El problema de estos tests de convergencia es que son sensibles a la escala de $\ell(\bullet)$ y del vector de parámetros θ . Por ello, suelen emplearse versiones adimensionales, independientes de la métrica del problema en concreto.

3.3. Criterios para elegir la longitud del paso.

En un problema de minimización, los métodos para determinar la longitud de paso intentan garantizar que una vez que se ha llevado a cabo una iteración, el valor de la función objetivo decrezca. Es decir:

$$\ell(\theta_{\mathbf{k}+1}) < \ell(\theta_{\mathbf{k}})$$

Los criterios para elegir la longitud de paso son métodos de line-search o búsqueda lineal de la longitud del paso a partir de una dirección dada. Estos métodos pueden ser de dos tipos:

3.3.1. Elección de la longitud de paso óptima.

Este método consiste en resolver un problema de optimización unidimensional en la variable $\alpha_{\mathbf{k}}$. Este criterio está basado en la idea de que, una vez calculada la dirección de búsqueda $\mathbf{p}_{\mathbf{k}}$, el valor de la función objetivo alcanzado después de dar el paso es:

$$\ell(\theta_{\mathbf{k}+1}) = \ell(\theta_{\mathbf{k}} + \alpha_{\mathbf{k}}\mathbf{p}_{\mathbf{k}})$$

y esta expresión sólo depende de la variable $\alpha_{\mathbf{k}}$. Por tanto, se puede plantear el problema:

$$\min_{\alpha_{\mathbf{k}}} \ell(\theta_{\mathbf{k}+1})$$

después de imponer alguna normalización sobre $\mathbf{p}_{\mathbf{k}}$. Resolviendo este programa se obtiene una solución óptima α_* , que da lugar a la máxima reducción en el valor de la función objetivo cuando nos movemos desde $\theta_{\mathbf{k}-1}$ en la dirección $\mathbf{p}_{\mathbf{k}}$.

3.3.2. Elección de la longitud de paso por backtracking.

Este criterio consiste en comenzar, en cada iteración, con una longitud de paso $\alpha_k = 1$. Si el punto $\theta_k + p_k$ no es aceptable (de acuerdo con algún criterio), se vuelve hacia atrás (reduciendo α_k) hasta que se encuentre un valor $\theta_k + \alpha_k p_k$ aceptable.

En teoría, parece lógico requerir que:

$$\ell(\theta_{k+1}) < \ell(\theta_k)$$

Sin embargo, esta condición no garantiza que la secuencia $\{\theta_k\}$ converga al mínimo. Un primer problema es que el valor de la función puede decrecer muy poco en relación al tamaño del paso. En este sentido, se puede exigir que la tasa media de decrecimiento desde $\ell(\theta_k)$ a $\ell(\theta_{k+1})$ sea, al menos, una fracción de la tasa inicial de decrecimiento del valor de la función en esa dirección. Es decir, dado $\lambda \in (0,1)$, se ha de cumplir que:

$$\ell(\theta_k + \alpha_k p_k) \leq \ell(\theta_k) + \lambda \alpha_k g(\theta_k)' p_k \quad [3.3.2.1]$$

El segundo problema puede ser que el tamaño del paso sea demasiado pequeño en relación con la tasa inicial de decrecimiento del valor de la función. Sin embargo, se puede exigir que la tasa de decrecimiento de $\ell(\cdot)$ desde el punto θ_{k+1} en la dirección p_k sea mayor que una cierta fracción de la tasa de decrecimiento de $\ell(\cdot)$ desde θ_k en la dirección p_k . Es decir, dado $\beta \in (0,1)$, se ha de cumplir que:

$$g(\theta_k + \alpha_k p_k)' p_k \geq \beta g(\theta_k)' p_k \quad [3.3.2.2]$$

En la práctica, esta última condición no es necesaria porque la estrategia de backtracking evita tomar pasos excesivamente pequeños.

Por tanto, se ha de encontrar, en cada iteración, un valor de α_k que cumpla las condiciones [3.3.2.1] y [3.3.2.2].

3.4. Direcciones de búsqueda del óptimo.

Como ya se ha dicho, una estrategia natural para resolver un problema de minimización es asegurar que en cada iteración, decrezca el valor de la función objetivo. Por tanto, la idea básica es elegir una dirección \mathbf{p}_k desde un punto θ_k , tal que si $\theta_k \neq \theta_*$, exista algún desplazamiento a lo largo de esa dirección que haga decrecer el valor de la función objetivo.

Esta dirección es una dirección de descenso¹. Matemáticamente, \mathbf{p}_k es una dirección de descenso desde $\theta_k \neq \theta_*$, si existe algún $\alpha_k > 0$ tal que:

$$\ell(\theta_k + \alpha_k \mathbf{p}_k) < \ell(\theta_k)$$

Una condición necesaria y suficiente para que \mathbf{p}_k sea una dirección de descenso es que:

$$g(\theta_k)' \mathbf{p}_k < 0$$

Es fácil obtener esta condición a partir de un desarrollo en serie de Taylor de la función $\ell(\cdot)$ alrededor del entorno del punto θ_k .

3.4.1. Dirección de Cauchy.

A partir del concepto de dirección de descenso, parece bueno intentar conseguir una gran reducción en el valor de la función objetivo eligiendo \mathbf{p}_k de forma que $g(\theta_k)' \mathbf{p}_k$ sea "grande" y negativo. Sin embargo, debe imponerse alguna normalización sobre \mathbf{p}_k , porque de lo contrario, para cualquier \mathbf{p}_+ tal que $g(\theta_k)' \mathbf{p}_+ < 0$, siempre es posible elegir una dirección \mathbf{p}_k como un múltiplo arbitrariamente grande de \mathbf{p}_+ .

¹ Análogamente, para un problema de maximización se define el concepto de dirección de ascenso.

Por tanto, el objetivo es encontrar \mathbf{p}_k de modo que $g(\theta_k)' \mathbf{p}_k$ sea mínimo sujeto a alguna normalización de \mathbf{p}_k . Es decir:

$$\begin{aligned} \min_{\mathbf{p}_k} & g(\theta_k)' \mathbf{p}_k \\ \text{s.a.} & \|\mathbf{p}_k\| = 1 \end{aligned}$$

La solución a este problema depende del tipo de norma que se utilice. Si se elige la usual $\|\mathbf{p}_k\| = (\mathbf{p}_k' \mathbf{p}_k)^{1/2}$, el mínimo es:

$$\mathbf{p}_k = -g(\theta_k) / \|g(\theta_k)\|$$

y ésta es la dirección de Cauchy. Con respecto a la misma cabe señalar que:

- (a) Aunque es una dirección de descenso, los algoritmos basados en ella sólo tienen una tasa de convergencia lineal².
- (b) Los algoritmos basados en esta dirección se acercan rápidamente al óptimo en las primeras iteraciones, independientemente de las condiciones iniciales de las que se parta. Sin embargo, su comportamiento empeora en la cercanía del óptimo.
- (c) Esta dirección sólo usa información sobre las primeras derivadas analíticas de la función a optimizar.
- (d) Es una dirección muy sensible a los cambios de escala del vector θ .

² Un algoritmo es linealmente convergente si, aplicado a una función lineal, converge en una iteración.

3.4.2. Dirección de Newton.

La dirección de Newton se basa en una aproximación de Taylor de segundo orden a la función objetivo en un entorno de θ_k . Así, sea $\phi(\theta)$ la aproximación de segundo orden de $\ell(\theta)$. Entonces:

$$\phi(\theta_{k+1}) = \ell(\theta_k) + g_k'(\theta_{k+1} - \theta_k) + \frac{1}{2} (\theta_{k+1} - \theta_k)' G_k (\theta_{k+1} - \theta_k)$$

Las condiciones de primer orden del problema:

$$\min_{\theta_{k+1}} [\phi(\theta_{k+1}) - \ell(\theta_k)]$$

son:

$$g_k + G_k(\theta_{k+1} - \theta_k) = 0$$

y por tanto, se obtiene la siguiente regla de actualización:

$$\theta_{k+1} = \theta_k - G_k^{-1} g_k$$

El vector de parámetros actualizado θ_{k+1} será un mínimo global si la matriz G_k es estrictamente definida positiva. El buen condicionamiento de la matriz hessiana es una condición necesaria y suficiente para que:

- (a) La dirección de Newton sea de descenso.
- (b) El algoritmo sea cuadráticamente convergente³.
- (c) La ventaja de este algoritmo es que si θ_0 está lo suficientemente cerca del óptimo θ_* y G es definida positiva, la secuencia $\{\theta_k\}$ generada por el algoritmo converge muy rápidamente a θ_* .

³ Un algoritmo es cuadráticamente convergente si, aplicado a una función cuadrática, converge en una sola iteración.

- (d) Esta dirección no sólo usa información sobre las primeras derivadas de la función objetivo, sino también sobre las segundas derivadas. Es decir, en la actualización del vector de parámetros tiene en cuenta la curvatura de la función objetivo.
- (e) La dirección de Newton es independiente de la escala de θ .

Es importante notar que habría que disponer de alguna modificación del método de Newton cuando el hessiano, evaluado en algún punto, no es definido positivo, que es el principal problema de dicho algoritmo. Se tienen, entonces, los métodos de cuasi-Newton o de Newton modificados.

3.5. Método de Newton modificado para asegurar el buen condicionamiento de H_k .

Este algoritmo consiste en aplicar el método anterior a G_k para encontrar la descomposición de Cholesky de la matriz:

$$H_k = G_k + \mu_k I$$

donde:

- $\mu_k = 0$ si G_k es una matriz definida positiva.
- $\mu_k > 0$ si G_k es una matriz no definida positiva.

En cualquier caso, se resuelve:

$$\theta_{k+1} = \theta_k - H_{k-1}^{-1} g_k$$

y si θ_0 está muy cerca del mínimo θ_* y G_* es definido positivo y está bien condicionado, este algoritmo es equivalente al anterior.

Sin embargo, este último algoritmo tiene la ventaja de que la dirección de Newton modificada $-H_k^{-1} g_k$ es una dirección tal

que, cualquier función $\ell(\cdot)$ dos veces diferenciable, decrece a partir de cualquier punto θ_k , distinto del óptimo.

Dennis y Schnabel (1983) demuestran que es posible calcular a la vez el escalar $\mu_k \geq 0$ y la factorización de Cholesky de H_k . Cuando G_k no es definido positivo, el escalar μ_k más pequeño que hay que añadir al hessiano ha de ser ligeramente mayor que el menor autovalor de dicho hessiano. En primer lugar, se aplica un algoritmo modificado de factorización de Cholesky⁴ a G_k . Este consiste en calcular $G_k + D = LL'$ siendo D una matriz diagonal con elementos no negativos que se convierten en cero si G_k es definido positivo. Por tanto, si $D = 0$, $\mu_k = 0$. Si $D \neq 0$, se calcula una cota superior de μ_k (denotada por b_1), usando para ello el teorema de Gerschgorin⁵. Como existe otra cota superior a μ_k definida como:

$$b_2 = \max_{1 \leq i \leq n} (d_{ii})$$

se elige $\mu_k = \min(b_1, b_2)$. De esta forma, se perturba lo mínimo posible el hessiano para asegurar su buen condicionamiento. Por último, se calcula la factorización de Cholesky del hessiano modificado $H_k = G_k + \mu_k I$.

Por tanto, habría que comenzar en cada iteración con el método de Newton. Este método se usaría si con él se da un paso razonable (por ejemplo, si el valor de la función decrece). Si no ocurre esto, habría que retroceder y tomar un paso dictado por un método global. Con esta estrategia, siempre se acabará usando el método de Newton cerca del óptimo, obteniendo una rápida convergencia local.

En cada iteración k , este algoritmo se puede resumir en los siguientes pasos:

⁴ Vid. Gill y Murrey (1972).

⁵ Vid. Dennis y Schnabel (1983).

- (1) Calcular \mathbf{g}_k y decidir si parar o continuar de acuerdo con algún criterio de convergencia previamente fijado.
- (2) Calcular \mathbf{G}_k .
- (3) Aplicar una técnica de factorización a \mathbf{G}_k y calcular su número de condición. Si dicho hessiano está mal condicionado, se perturba de la forma descrita anteriormente.
- (4) Calcular $\mathbf{s}_k = -\mathbf{H}_k^{-1}\mathbf{g}_k$, siendo $\mathbf{H}_k = \mathbf{G}_k + \mu_k\mathbf{I}$.
- (5) Hacer $\theta_{k+1} = \theta_k + \mathbf{s}_k$ o bien, elegir θ_{k+1} por una estrategia global. Volver a (1).

3.6. Descripción de los algoritmos de optimización utilizados.

Los algoritmos de optimización escogidos para minimizar la función de verosimilitud dada en [3.1.1], utilizan la dirección de Newton por las siguientes razones:

- Dicha función es muy similar a una función cuadrática, sobre todo, si el segundo término de [3.1.1] es despreciable con respecto al tercero.
- Se dispone de la expresión analítica exacta del gradiente y del hessiano de dicha función, que es la información que utiliza la dirección de Newton para actualizar el vector de parámetros en cada iteración.
- Usualmente, puede disponerse de una buena condición inicial sobre el valor óptimo de los parámetros. Esto proviene del conocimiento que se tiene a priori del modelo econométrico objeto de estudio o de la estimación de dichos coeficientes por otros procedimientos distintos al desarrollado en este trabajo.

Con respecto al criterio de elección de la longitud de paso, se ha utilizado un método de backtracking. Se comienza siempre que sea posible con $\alpha_k = 1$, con el objeto de mantener la ventaja del método de Newton cerca de la solución.

Básicamente, se han utilizado tres algoritmos de optimización alternativos:

- un método de line-search.
- dos métodos de model-trust region.

incorporados en UNCMIN (UNConstrained MINimization) y que, a continuación, se describen. UNCMIN es un sistema modular de subrutinas escritas en FORTRAN desarrollado en la Universidad de Colorado y basado en los métodos descritos por Dennis y Schnabel (1983). El paquete está diseñado para minimizar, sin restricciones, campos escalares no lineales y derivables continuamente dos veces.

3.6.1. Método de line-search.

La primera aproximación para iterar desde una solución fuera de la región de convergencia del método de Newton, es el método de line-search (o de búsqueda lineal). La idea de este método es la siguiente:

- (1) Se calcula una dirección de descenso \mathbf{p}_k .
- (2) Se calcula $\theta_{k+1} = \theta_k + \alpha_k \mathbf{p}_k$ para algún $\alpha_k \geq 0$, que haga θ_{k+1} un punto aceptable. En este método, se usa una dirección de cuasi-Newton $\mathbf{p}_k = -\mathbf{H}_k^{-1} \mathbf{g}_k$ donde la matriz hessiano se calcula como $\mathbf{H}_k = \mathbf{G}_k + \mu_k \mathbf{I}$, siendo $\mu_k = 0$ si \mathbf{G}_k es estrictamente definida positiva. Esto permite que el algoritmo mantenga una rápida convergencia local.

Definidas las condiciones [3.3.2.1] y [3.3.2.2], pueden demostrarse los tres resultados siguientes:

(1) **Teorema de Wolfe** (1969,71):

(a) Dada cualquier dirección \mathbf{p}_k , tal que se cumpla que $g(\theta_k)' \mathbf{p}_k < 0$, existe un $\alpha_k > 0$ que satisface las condiciones [3.3.2.1] y [3.3.2.2].

(b) Cualquier método que genere una secuencia de valores $\{\theta_k\}$ tal que $g(\theta_k)'(\theta_{k+1} - \theta_k) < 0$ y satisfaga las condiciones [3.3.2.1]-[3.3.2.2], es un método globalmente convergente.

(2) **Teorema de Dennis y Moré** (1974): Cerca de un mínimo de $\ell(\cdot)$, donde el hessiano es definido positivo, el paso de Newton satisface las condiciones [3.3.2.1]-[3.3.2.2].

Lo importante de estos resultados es que muestran que pueden crearse algoritmos globalmente convergentes intentando siempre primero el paso de Newton.

La estructura del algoritmo es la siguiente:

Algoritmo A.1: Dado un valor $\lambda \in (0, 1/2)$ y $0 < \zeta < 1$, tomar primero $\alpha_k = 1$. Mientras se cumpla la condición:

$$\ell(\theta_k + \alpha_k \mathbf{p}_k) > \ell(\theta_k) + \lambda \alpha_k g(\theta_k)' \mathbf{p}_k$$

se toman pasos $\alpha_k = \rho \alpha_k$, siendo ρ algún valor $\rho \in [\zeta, u]$, que en cada instante es elegido por line-search. Por último, se calcula:

$$\theta_{k+1} = \theta_k + \alpha_k \mathbf{p}_k$$

Para encontrar la longitud de paso α_k en cada iteración, se utiliza el Algoritmo A.1. En el paquete UNCMIN se fija un valor $\lambda = 10^{-4}$. La forma de obtener el valor de α_k que satisface

las condiciones [3.3.2.1] y [3.3.2.2] se describe a continuación. Si se define:

$$\ell(\alpha_k = \alpha) = \ell(\theta_k + \alpha p_k)$$

y es necesario hacer backtracking, se usará la información más reciente acerca del valor de $\ell(\bullet)$. Se dispone de dos fuentes de información acerca de $\ell(\alpha_k = \alpha)$. Por un lado:

$$\ell(\alpha=0) = \ell(\theta_k) \quad \ell'(\alpha=0) = g_k' p_k \quad [3.6.1.1]$$

y por otro lado, al calcular $\ell(\theta_k + p_k)$, se conoce:

$$\ell(\alpha=1) = \ell(\theta_k + p_k) \quad [3.6.1.2]$$

Si $\ell(\theta_k + p_k)$ no satisface las condiciones [3.3.2.1]-[3.3.2.2] (es decir, si $\ell(\alpha=1) > \ell(\alpha=0) + \lambda \ell'(\alpha=0)$), se ajusta $\ell(\alpha_k = \alpha)$ por un modelo cuadrático unidimensional (denotado por $m(\alpha)$) que satisface [3.6.1.1] y [3.6.1.2]. Es decir:

$$m(\alpha) = [\ell(\alpha=1) - \ell'(\alpha=0)]\alpha^2 + \ell'(\alpha=0)\alpha + \ell(\alpha=0)$$

Minimizando la función $m(\alpha)$ con respecto a α , se obtiene la solución:

$$\hat{\alpha} = \frac{-\ell'(\alpha=0)}{2[\ell(\alpha=1) - \ell(\alpha=0) - \ell'(\alpha=0)]}$$

que es un mínimo, ya que:

$$m'(\hat{\alpha}) = 0$$

y también cumple la condición de segundo orden:

$$m''(\hat{\alpha}) = 2[\ell(\alpha=1) - \ell(\alpha=0) - \ell'(\alpha=0)] > 0$$

al poder demostrarse que:

$$\ell(\alpha=1) > \ell(\alpha=0) + \lambda \ell'(\alpha=0) > \ell(\alpha=0) + \ell'(\alpha=0)$$

Es fácil ver que $\hat{\alpha} > 0$ al ser $\ell'(\alpha=0) < 0$. Por tanto, se tomará $\hat{\alpha}$ como el nuevo valor de α_k en el Algoritmo A.1. Además, como $\ell(\alpha=1) > \ell(\alpha=0) + \lambda \ell'(\alpha=0)$, se tiene que:

$$\hat{\alpha} < \frac{1}{2(1 - \lambda)}$$

Y de hecho, si $\ell(\alpha=1) > \ell(\alpha=0)$, se tiene que:

$$\hat{\alpha} < \frac{1}{2}$$

Por tanto, el valor óptimo $\hat{\alpha}$ proporciona una cota superior al valor de $u \approx 1/2$ para obtener el primer valor de ρ en el Algoritmo A.1.

En el paquete UNCMIN, este algoritmo tiene dos características:

- (1) Se fija un paso de longitud mínima. Esto sirve para contrastar si se ha llegado al óptimo. Si la condición [3.3.2.1] no se cumple, pero $\|\alpha_k \mathbf{p}_k\|$ es menor que la longitud de paso mínima, se ha llegado al óptimo.
- (2) El usuario fija una longitud de paso máxima, ya que se pueden dar pasos excesivamente grandes cuando se calcula $\mathbf{p}_k = -\mathbf{H}_k^{-1} \mathbf{g}_k$ y \mathbf{H}_k es una matriz casi singular.

3.6.2. Métodos de model trust-region.

Estos algoritmos no suponen que los pasos de longitud corta deben estar en la dirección de cuasi-Newton. Como se ha visto, si el paso de cuasi-Newton no es satisfactorio, en los métodos de line-search se elige una longitud de paso menor y se mantiene la misma dirección del paso. La nueva longitud de paso se consigue construyendo un nuevo modelo cuadrático unidimensional, basado solamente en la información sobre la función y el gradiente en la dirección de cuasi-Newton.

En estos nuevos métodos, cuando se necesita encontrar un paso más corto, primero se elige la longitud de paso y luego, se usa el modelo cuadrático n-dimensional para elegir la dirección del mismo.

Una vez realizada la primera iteración, parece razonable pensar que se tiene una idea sobre la longitud de paso a tomar a partir de ese punto. Por ejemplo, de la longitud de paso tomada en la iteración k se puede deducir una cota superior de la mejor longitud de paso a adoptar en la iteración $k+1$. Es decir, en estos métodos se usa información sobre la función, obtenida en las iteraciones anteriores, para estimar la mejor longitud de paso en la siguiente iteración.

Supongamos que estamos en un punto $\theta_{\mathbf{k}}$ y se dispone de alguna estimación ($\delta_{\mathbf{k}}$) de la máxima longitud de un "buen" paso. Surge entonces la pregunta de ¿cómo seleccionar un paso de longitud máxima $\delta_{\mathbf{k}}$ desde el punto $\theta_{\mathbf{k}}$? Existe una respuesta natural si volvemos a un modelo cuadrático. Añadiendo la idea de acotar la longitud máxima del paso por $\delta_{\mathbf{k}} > 0$, la respuesta es conseguir un paso $\delta_{\mathbf{k}}$ que resuelva el problema:

$$\begin{aligned} \min \quad & \ell(\theta_{\mathbf{k}}) + g(\theta_{\mathbf{k}})'s + \frac{1}{2} s'H_{\mathbf{k}}s \\ \text{s.a} \quad & \|s\| \leq \delta_{\mathbf{k}} \end{aligned}$$

La solución a este problema viene dada por el siguiente lema:

Sea $\ell: \mathbb{R}^n \rightarrow \mathbb{R}$ dos veces continuamente diferenciable y \mathbf{H}_k una matriz simétrica y definida positiva. Entonces:

$$\mathbf{s}(\mu) = -(\mathbf{H}_k + \mu\mathbf{I})^{-1}g(\theta_k)$$

para un único $\mu \geq 0$ tal que $\|\mathbf{s}(\mu)\| = \delta_k$, salvo que $\|\mathbf{s}(0)\| \leq \delta_k$. En este caso, $\mathbf{s}(0)$ es la solución de Newton. Es evidente que para cualquier $\mu \geq 0$, $\mathbf{s}(\mu)$ define una dirección de descenso desde el punto θ_k .

No obstante, no hay ninguna garantía de que el punto θ_{k+1} que resuelve exacta o aproximadamente el problema anterior sea el siguiente paso aceptable, aunque lo será si δ_k es un "buen" paso acotado. Por tanto, un paso completo en un algoritmo de trust-region tiene la siguiente forma:

Algoritmo A.2: Dada $\ell: \mathbb{R}^n \rightarrow \mathbb{R}$, $\delta_k > 0$, $\theta_k \in \mathbb{R}^n$ y $\mathbf{H}_k \in \mathbb{R}^{n \times n}$ simétrica y definida positiva, repetir en cada iteración:

- (1) $\theta_{k+1} = \theta_k + \mathbf{s}_k$, siendo \mathbf{s}_k la solución del problema anterior.
- (2) Decidir si θ_{k+1} es un punto aceptable de acuerdo con algún criterio establecido y calcular un nuevo valor de δ_k hasta que θ_{k+1} sea un punto aceptable.
- (3) Hacer $\delta_{k+1} = \delta_k$.

Existen dos aproximaciones para encontrar el paso en este tipo de algoritmos:

(1) **El paso óptimo localmente restringido o hook step:** Consiste en encontrar un paso $\mathbf{s}(\mu) = -(\mathbf{H}_k + \mu\mathbf{I})^{-1}g(\theta_k)$ tal que se cumpla

que $\|\mathbf{s}(\mu)\| \approx \delta_k$ y calcular $\theta_{k+1} = \theta_k + \mathbf{s}(\mu)$. Este método encuentra una solución aproximada μ_k a la ecuación escalar:

$$\mathbf{s}(\mu) = \|\mathbf{s}(\mu)\| - \delta_k = 0$$

(2) El paso **double dogleg**: Este método también encuentra una solución aproximada al problema anterior. Sin embargo, más que encontrar un punto $\theta_{k+1} = \theta_k + \mathbf{s}(\mu_k)$ sobre la curva $\mathbf{s}(\mu)$ tal que se cumpla que $\|\theta_{k+1} - \theta_k\| \approx \delta_k$, aproxima esta curva por una función lineal conectando el punto de Cauchy (el óptimo del modelo cuadrático en la dirección de máximo descenso) con la dirección de Newton para ese modelo. Esta estrategia se basa en moverse en la dirección de Cauchy cuando δ_k es pequeño y en la dirección de cuasi-Newton cuando δ_k se incrementa.

Actualización de la región de confianza.

Para completar el paso global dado en el Algoritmo A.2., es necesario decidir si el punto θ_{k+1} encontrado por hook step o por double dogleg step es una iteración aceptable.

- (a) Si θ_{k+1} no es aceptable, se reduce el tamaño de la región de confianza y se minimiza el mismo modelo cuadrático sobre una región de confianza más pequeña.
- (b) Si θ_{k+1} es un punto satisfactorio, hay que decidir si el tamaño de la región de confianza debe aumentar, disminuir o mantenerse para dar el siguiente paso.

La condición que hace de θ_{k+1} un punto aceptable es:

$$\ell(\theta_{k+1}) \leq \ell(\theta_k) + \lambda g(\theta_k)'(\theta_{k+1} - \theta_k) \quad [3.6.2.1]$$

donde $\lambda \in (0, 1/2)$. En UNCMIN, de nuevo, se elige $\lambda = 10^{-4}$ ya que la condición [3.6.2.1] es más fuerte que $\ell(\theta_{k+1}) < \ell(\theta_k)$.

Si θ_{k+1} no cumple la condición [3.6.2.1], se reduce el tamaño de la región de confianza y se vuelve a encontrar la solución aproximada del problema de minimización restringida por hook step o por double dogleg step. El factor de reducción del tamaño de la región de confianza se obtiene aplicando la misma estrategia de backtracking del algoritmo line-search. Aplicando esta estrategia, se obtiene:

$$\alpha_* = \frac{-g(\theta_k)'(\theta_{k+1} - \theta_k)}{2[\ell(\theta_{k+1}) - \ell(\theta_k) - g(\theta_k)'(\theta_{k+1} - \theta_k)]}$$

y el nuevo radio de la región de confianza es:

$$\delta_{k+1} = \alpha_* \|\theta_{k+1} - \theta_k\|$$

Si θ_{k+1} cumple la condición [3.6.2.1], hay que distinguir dos casos:

- (1) Si θ_{k+1} está en el paso de Newton, se toma este paso, se actualiza δ , se forma el nuevo modelo y se pasa a la siguiente iteración.
- (2) Si $(\theta_{k+1} - \theta_k)$ no está en el paso de Newton, primero se intenta dar un paso más largo desde θ_k usando el modelo actual. Si θ_{k+1} es ya un punto aceptable, hay que actualizar el valor δ_{k+1} . En este aspecto, hay tres alternativas: que se doble su tamaño, que se reduzca a la mitad o que se mantenga. Así, si el modelo cuadrático con el que se está trabajando predice bien la función, se aumenta el tamaño de la región. Si por el contrario, se está prediciendo mal, se reduce su tamaño.

En el paquete UNCMIN, en concreto, se define la reducción real en el valor de la función como $g = \ell(\theta_{k+1}) - \ell(\theta_k)$ y la reducción prevista en el valor de la función a través del modelo

cuadrático con que se trabaja como $g_{\text{pred}} = m_{\mathbf{k}}(\theta_{\mathbf{k}+1}) - \ell(\theta_{\mathbf{k}})$. Y definidos así estos incrementos, las estrategias a tomar son:

- Si $g \leq 0.75g_{\text{pred}}$ se toma $\delta_{\mathbf{k}+1} = 2\delta_{\mathbf{k}}$.
- Si $g > 0.10g_{\text{pred}}$ se toma $\delta_{\mathbf{k}+1} = 1/2\delta_{\mathbf{k}}$.
- En cualquier otro caso, se toma $\delta_{\mathbf{k}+1} = \delta_{\mathbf{k}}$.

3.7. Criterios de convergencia utilizados.

El criterio de convergencia utilizado para decidir la finalización de cualquiera de los algoritmos descritos, está referido al cumplimiento de las condiciones de primer orden.

La justificación de usar dicho criterio es que, aunque la condición $\mathbf{g}_{\mathbf{k}} = \mathbf{0}$ también puede darse en un máximo o en un punto de silla, la utilización de métodos globalmente convergentes y la estrategia de perturbar el hessiano siempre que no sea estrictamente definido positivo, hace prácticamente imposible que el algoritmo converga a un máximo o a un punto de silla. Por tanto, en nuestro contexto, la condición $\mathbf{g}_{\mathbf{k}} = \mathbf{0}$ es una condición necesaria y "casi" suficiente para que $\theta_{\mathbf{k}}$ sea un mínimo de la función $\ell(\bullet)$.

En precisión infinita, una condición necesaria para que θ sea un mínimo es que:

$$\mathbf{g}_{\mathbf{k}} = \mathbf{0}$$

Sin embargo, en precisión finita, es necesario relajar la condición anterior y sustituirla por:

$$\|\mathbf{g}_{\mathbf{k}}\| \leq \epsilon$$

siendo ϵ una tolerancia arbitrariamente pequeña. Sin embargo, este test no es adecuado porque depende de la escala de $\theta_{\mathbf{k}}$ y de $\ell(\bullet)$. Por ejemplo, si se fija $\epsilon = 10^{-3}$ y los valores de $\ell(\bullet)$

están en el rango $[10^{-7}, 10^{-5}]$, es muy probable que cualquier valor de $\theta_{\mathbf{k}}$ satisfaga esa condición de convergencia. Por el contrario, si los valores de $\ell(\bullet)$ están en el rango $[10^5, 10^7]$, esa condición puede ser demasiado exigente.

Para resolver este problema, se puede usar el siguiente criterio:

$$|\mathbf{g}_{\mathbf{k}}' \mathbf{G}_{\mathbf{k}}^{-1} \mathbf{g}_{\mathbf{k}}| \leq \epsilon$$

Esta condición es independiente de la escala de $\theta_{\mathbf{k}}$, pero sigue dependiendo de la escala de $\ell(\bullet)$.

Una mejora del test anterior consiste en definir el gradiente relativo (GR) de la función $\ell(\bullet)$ en el punto $\theta_{\mathbf{k}}$, de la siguiente forma:

$$\begin{aligned} \text{GR} &= \frac{\text{tasa de cambio relativa de } \ell}{\text{tasa de cambio relativa de } \theta_{\mathbf{k}}} = \lim_{\delta \rightarrow 0} \frac{\frac{\ell(\theta_{\mathbf{k}} + \delta \mathbf{e}_i) - \ell(\theta_{\mathbf{k}})}{\ell(\theta_{\mathbf{k}})}}{\delta / \theta_{i,\mathbf{k}}} = \\ &= \frac{\mathbf{g}_{\mathbf{k}}^{\theta_{i,\mathbf{k}}}}{\ell(\theta_{\mathbf{k}})} \end{aligned}$$

donde \mathbf{e}_i es un vector unitario y $\theta_{i,\mathbf{k}}$ es la componente i -ésima del vector θ en la iteración \mathbf{k} . Con esta definición, se exige el cumplimiento de la condición:

$$\|\text{GR}(\theta_{\mathbf{k}})\| \leq \epsilon$$

donde ϵ es la tolerancia del gradiente. Este criterio es independiente de cualquier cambio en la métrica de $\ell(\bullet)$ o de $\theta_{\mathbf{k}}$. Tiene el inconveniente de que la idea de cambio relativo en $\ell(\bullet)$ o en $\theta_{\mathbf{k}}$ se rompe si el valor de $\theta_{i,\mathbf{k}}$ o $\ell(\theta_{\mathbf{k}})$ es muy cercano a cero. Sin embargo, este problema se resuelve sustituyendo $\theta_{i,\mathbf{k}}$ y $\ell(\bullet)$ por $\max\{|\theta_{i,\mathbf{k}}|, t(\theta_{i,\mathbf{k}})\}$ y $\max\{|\ell(\theta_{\mathbf{k}})|, t(\ell)\}$, respectivamente,

donde $t(\theta_{i,k})$ y $t(\ell)$ es una estimación de la magnitud típica de $\theta_{i,k}$ y $\ell(\bullet)$, respectivamente. Con estas modificaciones, el test resultante y más utilizado en UNCMIN es:

$$\max_{1 \leq i \leq n} \left| \frac{g_{i,k} \max\{|\theta_{i,k}|, t(\theta_{i,k})\}}{\max\{|\ell(\theta_{\mathbf{k}})|, t(\ell)\}} \right| \leq \epsilon$$

Otro criterio de convergencia disponible en UNCMIN es el siguiente:

$$\|\text{rel}(\theta_{\mathbf{k}})\| \leq \eta$$

donde η es la tolerancia del paso⁶ y $\text{rel}(\theta_{\mathbf{k}})$ es el cambio relativo en los valores de $\theta_{i,k}$ definido como:

$$\text{rel}(\theta_{i,\mathbf{k}}) = \frac{|\theta_{i,k+1} - \theta_{i,k}|}{\max\{|\theta_{i,k+1}|, t(\theta_{i,k})\}} \leq \eta$$

3.8. Escalamiento de variables.

El paquete UNCMIN trabaja con variables escaladas. La razón es que, en la práctica, algunas variables pueden diferir mucho en cuanto a magnitud. Por ejemplo, supongamos que los valores de una variable θ_1 están comprendidos en el rango $[10^2, 10^3]$ y los de una segunda variable θ_2 están comprendidos en el rango $[10^{-7}, 10^{-6}]$. Una métrica tan dispar en estas variables da lugar, frecuentemente, a errores de redondeo importantes. Por ejemplo, al calcular términos como $\|\theta_{\mathbf{k}+1} - \theta_{\mathbf{k}}\|$, en el caso anterior la segunda variable será prácticamente ignorada. Este problema se soluciona reescalando las variables, es decir, cambiando sus unidades de medida. Esto es equivalente a efectuar

⁶ La tolerancia del paso se fija dependiendo del número de dígitos significativos que se desee en la solución. Así, si se quieren p dígitos, se elegiría $\eta = 10^{-p}$.

el cambio de variable $\hat{\theta} = T\theta$, siendo T una matriz diagonal de factores de escala. En el ejemplo anterior, esta matriz puede ser $T = \text{diag}\{10^{-3}, 10^6\}$.

Por tanto, se trabaja en un nuevo espacio de variables, se obtiene el paso global y después, se deshace la transformación aplicada a las variables. El resultado fundamental es que el paso de Newton no cambia ante esta transformación en las variables, pero la dirección de Cauchy sí. Esto quiere decir que un paso de búsqueda lineal en la dirección de Newton no está afectado por el cambio de unidades, pero un paso en la región de confianza sí puede cambiar.

Por estas razones, en el uso del paquete UNCMIN el usuario debe proporcionar una matriz de escala T correspondiente al cambio de unidades deseado en cada problema. En realidad, sólo se ha de ofrecer un vector de tipificaciones de todas las variables del problema y el algoritmo correspondiente construye la matriz $T = \text{diag}\{1/t(\theta_1), \dots, 1/t(\theta_n)\}$, de forma que la magnitud de cada variable transformada es del orden de la unidad. Lo que no permite UNCMIN es un escalamiento dinámico, que podría ser conveniente cuando en el proceso de optimización alguna variable cambie mucho su orden de magnitud.

3.9. Aproximación del gradiente y del hessiano por diferencias finitas.

A veces, las expresiones analíticas exactas del gradiente y del hessiano de una función a optimizar no están disponibles. Esto no ocurre en nuestro contexto, ya que dichas expresiones están dadas por las ecuaciones [2.2.3.2]-[2.2.3.12]. Sin embargo, estas derivadas pueden aproximarse por diferencias finitas, aunque habrá un error asociado a dichas aproximaciones.

La aproximación por diferencias finitas hacia adelante del componente (i,j) de $g(\theta)$ será:

$$a_{ij} = \frac{\ell_i(\theta + h\mathbf{e}_j) - \ell_i(\theta)}{h}$$

donde \mathbf{e}_j es un vector unitario. Esto es equivalente a aproximar la columna j -ésima de $g(\theta)$ por la expresión:

$$A_{.j} = \frac{\ell(\theta + h\mathbf{e}_j) - \ell(\theta)}{h}$$

donde $\|A_{.j} - g(\theta)\| = O(h)$ para un valor de h suficientemente pequeño. Es decir, el error cometido al realizar esta aproximación es de orden h .

Es más precisa la aproximación por diferencias finitas centrales, dada por la expresión:

$$a_i = \frac{\ell(\theta + h\mathbf{e}_i) - \ell(\theta - h\mathbf{e}_i)}{2h}$$

que como se observa, requiere evaluar la función el doble de veces que la aproximación anterior, pero el error cometido es menor, ya que $|a_i - g(\theta)_i| = O(h^2)$.

En Dennis y Schnabel (1983) se muestra que, si se elige adecuadamente el tamaño del paso para la aproximación por diferencias finitas, el método de Newton sigue teniendo una tasa de convergencia cuadrática.

De hecho, proponen perturbar cada componente de θ por su propio tamaño de paso $h_j = \eta^{\frac{1}{2}}\theta_j$ donde η es el épsilon de la máquina⁷. Es decir, no hay razón para que haya un único tamaño del paso, pudiendo ser no conveniente si la magnitud de los componentes θ_j difieren mucho entre sí. Como este tamaño del paso

⁷ El épsilon de la máquina se define como el menor número cuya suma con 1 difiere de 1.

puede ser muy pequeño si los valores de θ_j están muy cerca de cero, se puede modificar a:

$$h_j = \eta^{\frac{1}{2}} \max\{|\theta_j|, t(\theta_j)\}$$

donde $t(\theta_j)$ es la magnitud típica del componente θ_j .

3.10. Uso del paquete UNCMIN.

Los datos requeridos por el programa UNCMIN son la dimensión del problema, una subrutina para evaluar la función $\ell(\theta)$ y una estimación inicial θ_0 del óptimo θ_* .

El usuario puede controlar los métodos para evaluar las primeras y segundas derivadas de la función a optimizar. El uso del programa es ineficiente en el caso de optimización de una función de una variable y sólo tiene una limitación: no puede haber más de 100 parámetros a estimar. Sin embargo, en la práctica, es claro que ésta no es ninguna restricción, pues es raro encontrar modelos econométricos con más de 100 parámetros.

3.10.1. Elección del método de optimización.

El programa dispone de tres estrategias globales de minimización:

- (1) El método line-search.
- (2) Dos métodos de trust-region: el paso óptimo localmente restringido (hook step) y el paso double dogleg.

A menos que el usuario especifique un método concreto, el programa empleará por defecto el método de line-search. El mejor funcionamiento de los tres métodos de optimización alternativos dependerá del problema concreto bajo estudio. Es decir, a priori, no se sabe qué método será más adecuado en una situación particu-

lar. Parece que los métodos de trust-region funcionarán mejor que el de line-search en modelos donde el condicionamiento de la función objetivo sea malo o cuando los parámetros a estimar no puedan superar una zona de invertibilidad.

3.10.2. Cálculo del gradiente y del hessiano.

Este programa permite elegir entre proporcionar el gradiente y el hessiano analíticos de la función a minimizar o calcular una aproximación a dichas expresiones por diferencias finitas. Es un paquete muy flexible en este aspecto, permitiendo elegir entre una amplia gama de combinaciones entre derivadas analíticas y/o calculadas por diferencias finitas.

Además, a menos que el usuario no lo desee, UNCMIN realiza siempre una comparación entre el valor del gradiente y/o el hessiano analítico (proporcionado por el usuario) y su valor aproximado por diferencias finitas. Si ambas magnitudes difieren en más de un 1%, el programa finaliza con un mensaje de error. Esta comparación permite comprobar, en cada caso, si ha habido algún error en el cálculo por parte del usuario de las expresiones analíticas de las primeras y segundas derivadas de la función.

En este trabajo, siempre se proporcionan las expresiones analíticas exactas del gradiente y del hessiano y se pide que el programa compare estas magnitudes con las calculadas por diferencias finitas.

El usuario también puede especificar si la función objetivo es o no "costosa" de evaluar en cada vector de parámetros. Cuando es así, puede ser ventajoso utilizar los métodos de la secante para aproximar el hessiano (que no requieren evaluaciones de la función objetivo), más que el método de diferencias finitas, que sí requiere esas evaluaciones. El método de la

secante utilizado es el de Broyden, Fletcher, Goldfarb y Shanno⁸.

3.10.3. Entrada en el programa.

Dispone de dos subrutinas de entrada.

En la primera, el usuario debe proporcionar los siguientes datos:

- el número de parámetros a estimar.
- un valor aproximado del valor de la función objetivo cerca del óptimo.
- un vector de condiciones iniciales de los parámetros.

En la segunda, además de lo anterior, el usuario puede fijar:

- el tamaño máximo y mínimo permisible del paso.
- el radio de la región de confianza inicial si es elegido ese método de optimización.
- el grado de tolerancia del valor de la función, del gradiente y del paso para los criterios de convergencia.
- las tipificaciones de los parámetros para calcular una matriz de escala de las variables⁹.
- el máximo número de iteraciones permitido, etc.

Como se observa, la entrada del programa permite tener un control absoluto del proceso de optimización.

⁸ Vid. Dennis y Schnabel (1983).

⁹ No es necesario ofrecer este vector de tipificaciones de los parámetros, pero es muy conveniente si se sabe que los órdenes de magnitud de los componentes de θ difieren mucho entre sí.

3.10.4. Control de los resultados.

Cuando el algoritmo de optimización elegido finaliza, el programa devuelve el control ofreciendo una aproximación $\hat{\theta}$ al mínimo local θ_* , el valor de la función en ese punto $\ell(\hat{\theta})$, el valor del gradiente $g(\hat{\theta})$ y un mensaje que indica por qué ha finalizado el algoritmo. Por ejemplo:

- Se ha convergido.
- Se ha excedido el número de iteraciones permitido.
- La función a minimizar es demasiado no lineal.
- La función a minimizar no está acotada inferiormente o el tamaño máximo del paso fijado es demasiado pequeño, etc.

Por último, conviene señalar que el paquete UNCMIN originalmente programado en precisión simple, ha sido convertido a precisión doble para evitar problemas numéricos en el proceso de cálculo.

3.10.5. Organización del programa de estimación de modelos econométricos en forma de espacio de los estados mediante un filtro de Chandrasekhar.

El programa de estimación consta de las siguientes partes:

- Un programa principal.
- El paquete de subrutinas UNCMIN.
- La subrutina de cálculo del valor de la función a optimizar.
- La subrutina de cálculo del valor analítico del gradiente de la función objetivo.
- La subrutina de cálculo del valor analítico del hessiano de la función objetivo.
- La subrutina de cálculo del valor de la matriz de información.

- La subrutina de cálculo de las condiciones iniciales óptimas del vector de estado.
- Las subrutinas del modelo concreto a estimar.

Todas las subrutinas están programadas en FORTRAN y precisión doble.

La función del programa principal es leer las dimensiones del problema y distribuir la memoria disponible entre las matrices necesarias, dando a cada una de ellas la dimensión exacta requerida por el problema. La memoria total disponible puede cambiarse modificando un parámetro del programa.

El programa principal llama a UNCMIN para optimizar la función objetivo. Una vez que el algoritmo elegido finaliza, UNCMIN devuelve el control y el principal llama a la subrutina de cálculo de la matriz de información. Esta rutina calcula dicha matriz, las desviaciones típicas de los parámetros estimados y la matriz de correlación entre ellos.

Cuando UNCMIN necesita el cálculo del valor de la función, del gradiente y del hessiano, llama a las subrutinas correspondientes. Y éstas llaman (opcionalmente) a la rutina de cálculo de las condiciones iniciales óptimas del vector de estado.

El cálculo del valor de la función objetivo, del gradiente y del hessiano, ha sido programado de acuerdo con las expresiones analíticas exactas desarrolladas en el Capítulo 2. La subrutina de cálculo de las condiciones iniciales del vector de estado se corresponde con el algoritmo máximo-verosímil de estimación de dichas condiciones, descrito en el Apéndice A.

A su vez, las subrutinas de cálculo de la función, del gradiente y del hessiano necesitan la especificación concreta del modelo que se quiere estimar. Por ello, estas subrutinas llaman a las subrutinas propias de cada modelo. Estas son tres y contienen:

- Las expresiones de los elementos de las matrices estructurales y de ruido del modelo expresado en espacio de los estados.
- Las expresiones de las primeras derivadas de los elementos de dichas matrices con respecto a cada parámetro a estimar.
- Las expresiones de las segundas derivadas de los elementos de dichas matrices con respecto a cada parámetro a estimar.

Si se desea calcular el gradiente o el hessiano a través de las aproximaciones a dichos valores que permite UNCMIN, se puede suprimir la programación de estas subrutinas.

Estas tres últimas subrutinas cambian con cada modelo. Esto supone reprogramar cada vez que se quiere estimar una especificación distinta. Sin embargo, la programación de dichas rutinas es siempre fácil. Por ejemplo, si la especificación a estimar es un modelo MA(1) doble:

$$y_t = (1 - \theta B)(1 - \theta B^{12})a_t$$

la primera rutina del modelo consistiría en programar los elementos de las siguientes matrices:

$$\Phi = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} \quad \Gamma = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad E = \begin{bmatrix} -\theta \\ 0 \\ 0 \\ \vdots \\ -\theta \\ \theta\theta \end{bmatrix}$$

13×13 13×1 13×1

$$\mathbf{H} = [1 \ 0 \ \dots \ 0] \quad ; \quad D = 0 \quad ; \quad C = 1 \quad ; \quad Q = R = S = \sigma^2$$

1×13

Denotando $E(i,j)$, $Q(i,j)$, $R(i,j)$ y $S(i,j)$ al elemento (i,j) de dichas matrices, la rutina de las primeras derivadas sería:

$$\frac{\partial E(1,1)}{\partial \theta} = -1 ; \quad \frac{\partial E(12,1)}{\partial \theta} = -1 ; \quad \frac{\partial E(13,1)}{\partial \theta} = \theta ; \quad \frac{\partial E(13,1)}{\partial \theta} = \theta$$

$$\frac{\partial Q(1,1)}{\partial \sigma^2} = \frac{\partial R(1,1)}{\partial \sigma^2} = \frac{\partial S(1,1)}{\partial \sigma^2} = 1$$

y el resto de las primeras derivadas serían nulas al no depender ningún otro elemento de Ψ , Γ , E , D , y C de los parámetros θ , θ y σ^2 .

Y por último, la rutina de las segundas derivadas sería:

$$\frac{\partial^2 E(13,1)}{\partial \theta \partial \theta} = 1 \quad ; \quad \frac{\partial^2 E(13,1)}{\partial \theta \partial \theta} = 1$$

siendo también el resto de las segundas derivadas nulas.

CAPITULO 4

APLICACIONES

En este capítulo se presenta un conjunto de resultados empíricos obtenidos al aplicar el mecanismo de estimación descrito en los capítulos anteriores. Este procedimiento se ha aplicado tanto a modelos simulados como a modelos con datos reales.

Todos los modelos se han estimado utilizando tres procedimientos distintos:

- Mínimos cuadrados con retrovisión.
- Máxima verosimilitud exacta con el filtro de Kalman.
- Máxima verosimilitud exacta con el algoritmo de Chandrasekhar.

La aplicación de los dos últimos métodos requiere que los modelos sean previamente expresados en forma de espacio de los estados, siguiendo la norma de conversión expuesta en el Capítulo 2.

La razón de presentar los resultados obtenidos con distintos procedimientos de estimación es doble. Por un lado, se quiere mostrar que el mecanismo desarrollado ofrece resultados comparables a los de cualquier otro método, además de las ventajas computacionales y de robustez numérica ya citadas. Por otro lado, es importante analizar las posibles diferencias que se produzcan en la estimación de una misma especificación por este método y por otros alternativos.

Los modelos estimados son modelos ARMA univariantes, modelos de función de transferencia y modelos multivariantes. Todos ellos han sido especificados de acuerdo con la metodología Box-Jenkins (1976) de series temporales. De acuerdo con este enfoque, las series han sido transformadas adecuadamente para conseguir estacionariedad en media y en varianza, de modo que sólo se ha aplicado para su estimación el algoritmo de Chandrasekhar para modelos estacionarios.

Es importante señalar que este trabajo se centra en la etapa de estimación de un modelo, sin preocuparnos su especificación. Además, no ha sido un objetivo desarrollar un programa para la estimación por máxima verosimilitud con el filtro de Kalman y el de Chandrasekhar lo más eficiente posible, desde un punto de vista informático. Por ello, no son demasiado informativos los valores absolutos de tiempo de cálculo, pero sí los valores relativos con respecto al uso del filtro de Chandrasekhar en lugar del filtro de Kalman. La estructura del programa desarrollado en FORTRAN para un entorno de ordenadores personales es la misma cuando se utiliza un filtro u otro, de forma que la introducción de mejoras en las rutinas de cálculo del mismo reducirían los tiempos de cálculo en términos absolutos, pero no en términos relativos. Por otra parte, el programa empleado para estimar las especificaciones consideradas por mínimos cuadrados con retroprevisión es MICRANAL 1.3 (1984).

A continuación, se ofrecen primero los resultados de la estimación de un conjunto de modelos simulados. Con estos ejemplos se comprueba el buen funcionamiento de nuestro algoritmo de estimación, tanto en situaciones bien como mal condicionadas. Después, se presenta la estimación de una serie de modelos con datos reales como ejemplos ilustrativos de la mayor precisión numérica y del ahorro en tiempo de cálculo de la aproximación de Chandrasekhar con respecto a la de Kalman.

4.1. MODELOS SIMULADOS.

En este apartado, se presentan los resultados obtenidos con modelos simulados al aplicar el procedimiento de estimación por máxima verosimilitud descrito en los capítulos anteriores.

Se han simulado tres tipos de modelos: a) modelos univariantes autorregresivos, b) modelos univariantes de medias móviles y c) modelos multivariantes. Todos los modelos se han estimado aplicando los tres procedimientos citados anteriormente.

Usando los valores teóricos de los parámetros, se han generado 20 realizaciones de 100 observaciones para cada situación estudiada. Además, con objeto de analizar el comportamiento de nuestro método de estimación ante modelos bien y mal condicionados, se ha simulado cada modelo para distintos valores de los parámetros.

Los dos procedimientos de máxima verosimilitud considerados (con el filtro de Kalman y con el de Chandrasekhar), utilizan un método de optimización numérica de tipo line-search (vid. Capítulo 3). Las preestimaciones de los parámetros, necesarias para inicializar el algoritmo de optimización, se han escogido en todos los casos dentro del intervalo $\theta^t \pm 0.2\theta^t$ siendo θ^t el valor teórico de los parámetros. Las mismas condiciones iniciales han sido usadas para el filtro de Kalman y el de Chandrasekhar. Por tanto, son perfectamente comparables entre sí las estimaciones obtenidas con los dos procedimientos de máxima verosimilitud.

Los resultados de las simulaciones se presentan en las Tablas 4.1.1 a 4.1.13. En general, cabe destacar que las estimaciones obtenidas con el método de máxima verosimilitud utilizando el filtro de Chandrasekhar son aceptables. En todos los casos, los valores medios están muy cerca de los verdaderos valores y sus intervalos de confianza siempre contienen los valores teóricos. Además, es importante señalar que las desviaciones típicas calculadas coinciden básicamente con las obtenidas a

partir de la expresión exacta de la matriz de información (vid. Capítulo 2).

Todas las pruebas se han llevado a cabo en un ordenador personal Olivetti con procesador INTEL 80386 (20Mhz). En cada tabla, se ofrece la ganancia (o pérdida) porcentual media en tiempo de cálculo al usar el filtro de Chandrasekhar en lugar del filtro de Kalman. El gráfico 4.1.14 (pág 57), resume cómo crece exponencialmente esta ventaja computacional cuando aumenta la dimensión del vector de estado (n) con respecto a la dimensión del vector de observación (m).

A continuación, se comentan brevemente los resultados de las estimaciones obtenidas para cada situación considerada.

En la Tabla 4.1.1 se presentan los resultados de la simulación de un proceso AR(1) para distintos valores de los parámetros. Las estimaciones resultantes son muy similares con los tres criterios utilizados. Sólo cabe decir que, con máxima verosimilitud y el filtro de Chandrasekhar, los valores medios están ligeramente más cerca de los verdaderos valores de los parámetros. También se observa una pequeña ventaja en cómputo con la aproximación de Chandrasekhar versus el filtro de Kalman (nótese que en este modelo $n=m=1$).

En la Tabla 4.1.2 se muestran los resultados de la simulación de un proceso MA(1). En el caso bien condicionado ($\theta=0.5$), se observa que el procedimiento de mínimos cuadrados con retrovisión sobrestima el parámetro media móvil más que los otros dos métodos. En el caso mal condicionado, se observa que, tanto con mínimos cuadrados con retrovisión como con máxima verosimilitud con el filtro de Kalman, el parámetro está sobrestimado, mientras que esto no ocurre con la aproximación de Chandrasekhar. Esto es importante, pues indica el peligro de algunos procedimientos al estimar parámetros que no pueden superar una región de invertibilidad. Aunque en este caso también

$n=m=1$, se obtiene una pequeña ganancia porcentual en tiempo de cálculo con las ecuaciones de Chandrasekhar.

Con objeto de mostrar cómo crece la ganancia media porcentual en tiempo de cálculo cuando n aumenta con respecto a m , se ha simulado un proceso MA(1) estacional de período trimestral donde $n/m = 4/1$ y un MA(1) estacional de período anual donde $n/m = 12/1$. Los resultados de estas simulaciones se ofrecen en las Tablas 4.1.3 y 4.1.4, respectivamente. Mientras que la ganancia media porcentual en tiempo de estimación es, en algunos casos, de un 43% cuando $n = 4$, esta ganancia llega a ser de más de un 100% cuando $n = 12$. Es decir, el tiempo empleado (medido en segundos) en la estimación de este tipo de modelos se duplica cuando se utiliza el filtro de Kalman en lugar de las ecuaciones de Chandrasekhar. En cuanto a las estimaciones medias, señalar que, de nuevo, el método de máxima verosimilitud con el filtro de Kalman sobrestima más el valor del parámetro media móvil que el algoritmo de Chandrasekhar.

En las Tablas 4.1.5 y 4.1.6 se muestran los resultados de la simulación de un proceso AR(2) con distintos valores de los parámetros. Con esta especificación, los valores medios de las estimaciones son muy similares con los tres procedimientos, aunque se observa que el método de máxima verosimilitud con el filtro de Chandrasekhar llega a resultados más parecidos a los verdaderos valores que los otros dos procedimientos. En este modelo, al ser el ratio de dimensiones $n/m = 2/1$, también se obtiene un ahorro computacional con el filtro de Chandrasekhar, ahorro que varía dependiendo del modelo concreto estimado.

En la Tabla 4.1.7 se ofrecen los resultados de la simulación de un modelo ARMA(1,1). La estimación del parámetro autorregresivo está mucho más cerca de su verdadero valor cuando se aplica cualquier método de máxima verosimilitud que cuando se utiliza mínimos cuadrados con retrovisión. En cuanto al parámetro media móvil, sigue siendo la aproximación de Chandrasekhar la que mejor estima este coeficiente. Como siempre, se

obtiene una pequeña ventaja computacional de aproximadamente un 6% con las ecuaciones de Chandrasekhar versus el filtro de Kalman.

La Tabla 4.1.8 ofrece los resultados de la simulación de un proceso AR(1) con término constante. En ella, se observa un importante error en la estimación del coeficiente asociado al término constante mediante mínimos cuadrados con retrovisión. Sin embargo, esto no ocurre con cualquiera de los dos procedimientos de máxima verosimilitud, cuyos resultados coinciden exactamente en este caso.

En las Tablas 4.1.9 y 4.1.10 se muestran los resultados de la simulación de un proceso AR(1) con errores de observación. Debido a la falta de identificación de este modelo, se ha fijado la varianza residual del mismo y se ha dejado variar la varianza del error de observación, con el fin de obtener distintos ratios señal/ruido. Este proceso se ha generado para una muestra de 100 observaciones y considerando un valor del parámetro autorregresivo de 0.5 (vid. Tabla 4.1.9) y de 0.9 (vid. Tabla 4.1.10). Con cualquiera de los tres criterios utilizados, se observa que los parámetros son estimados con mayor precisión cuanto más alto es el ratio señal/ruido. Sin embargo, se aprecia una ligera mejora en las estimaciones conseguidas con máxima verosimilitud con el filtro de Chandrasekhar con respecto al filtro de Kalman. El ahorro porcentual en tiempo de cómputo es, en este caso, del orden de un 4% en todas las estimaciones. Cabe señalar, además, que la formulación de este proceso con error de observación es distinta a la convencional y se ajusta a la formulación dada por Terceiro (1990) para este tipo de modelos.

En cuanto a los modelos multivariantes, la Tabla 4.1.11 ofrece los resultados de la simulación de la primera especificación considerada. Es un modelo bivariante con una matriz autorregresiva diagonal y distinta de la identidad. En este caso, se observa que los valores medios de los parámetros también coinciden exactamente con los dos procedimientos de máxima verosimili-

tud considerados. Se aprecia una ligera infraestimación de dos parámetros en este modelo: la varianza residual de la segunda serie generada y la covarianza entre los ruidos. A pesar de que en este modelo el ratio de dimensiones $n/m = 1/2$, se consigue un pequeño ahorro computacional cuando se utilizan las ecuaciones de Chandrasekhar en lugar del filtro de Kalman.

En la Tabla 4.1.12 se ofrecen los resultados de la simulación de un modelo bivariante con una matriz media móvil diagonal y distinta de la identidad. La estimación del parámetro media móvil bien condicionado está infraestimado por los dos métodos, aunque ligeramente más por la aproximación de Kalman. Sin embargo, como casi siempre, el parámetro media móvil peor condicionado está sobrestimado por los dos procedimientos, aunque también más con el filtro de Kalman. De nuevo, un caso en que la dimensión n es menor que m y la ganancia en tiempo de cálculo empleado es de aproximadamente un 6% con el algoritmo de Chandrasekhar.

Por último, en la Tabla 4.1.13 se presentan los resultados de la simulación de un modelo bivariante con parte autorregresiva y media móvil. Además, se ha perturbado, en un instante de tiempo determinado, el valor de las dos series generadas. Este efecto ha sido corregido después mediante dos variables de intervención de tipo impulso, estimándose también los coeficientes asociados a dichas variables deterministas. En esta especificación, las estimaciones de todos los parámetros prácticamente coinciden con los dos procedimientos de máxima verosimilitud considerados.

Procedimiento de estimación				
Valor teórico de los parámetros	Mín. cuadrados con retrovisión	Máx. verosimilit. con el Filtro de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	Ganancia en tiempo de cálculo (1)/(2)
$\phi = 0.5$ $\sigma_a^2 = 1.0$	$\hat{\phi} = 0.47467$ (0.07839) $\hat{\sigma}_a^2 = 0.99086$ a (0.13854)	$\hat{\phi} = 0.48715$ (0.07860) $\hat{\sigma}_a^2 = 0.99939$ a (0.14643)	$\hat{\phi} = 0.48747$ (0.07909) $\hat{\sigma}_a^2 = 1.00063$ a (0.14585)	+ 2.42230%
$\phi = 0.9$ $\sigma_a^2 = 1.0$	$\hat{\phi} = 0.87498$ (0.04789) $\hat{\sigma}_a^2 = 0.99103$ a (0.13424)	$\hat{\phi} = 0.88859$ (0.04346) $\hat{\sigma}_a^2 = 0.99626$ a (0.14541)	$\hat{\phi} = 0.89185$ (0.04449) $\hat{\sigma}_a^2 = 1.00044$ a (0.14362)	+ 1.53951%

Tabla 4.1.1: Resultados de la simulación del modelo $(1 - \phi B)y_t = a_t$ con 100 observaciones y 20 realizaciones.

Valor teórico de los parámetros	Procedimiento de estimación			Ganancia en tiempo de cálculo (1)/(2)
	Min. cuadrados con retrovisión	Máx. verosimilit. con el Filtro de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
$\theta = 0.5$ $\sigma^2 = 1.0$ a	$\hat{\theta} = 0.53072$ (0.09872) $\hat{\sigma}^2 = 0.98614$ a (0.14217)	$\hat{\theta} = 0.50298$ (0.08177) $\hat{\sigma}^2 = 0.99442$ a (0.13991)	$\hat{\theta} = 0.50327$ (0.08097) $\hat{\sigma}^2 = 0.99451$ a (0.14002)	+ 0.09420%
$\theta = 0.9$ $\sigma^2 = 1.0$ a	$\hat{\theta} = 0.90214$ (0.04837) $\hat{\sigma}^2 = 1.00469$ a (0.14068)	$\hat{\theta} = 0.91587$ (0.05190) $\hat{\sigma}^2 = 0.99123$ a (0.13676)	$\hat{\theta} = 0.88812$ (0.04829) $\hat{\sigma}^2 = 1.00179$ a (0.13754)	+ 2.35170%

Tabla 4.1.2: Resultados de la simulación del modelo $y_t = (1 - \theta B)a_t$ con 100 observaciones y 20 realizaciones.

Valor teórico de los parámetros	Procedimiento de estimación			Ganancia en tiempo de cálculo (1)/(2)
	Min. cuadrados con retrovisión	Máx. verosimilit. con el Filtro de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
$\theta = 0.5$	$\hat{\theta} = 0.54975$ (0.11096)	$\hat{\theta} = 0.50848$ (0.10603)	$\hat{\theta} = 0.50743$ (0.10640)	+ 20.8804%
$\sigma_a^2 = 1.0$	$\hat{\sigma}_a^2 = 0.98338$ a (0.14655)	$\hat{\sigma}_a^2 = 0.99432$ a (0.14798)	$\hat{\sigma}_a^2 = 0.99524$ a (0.14754)	
$\theta = 0.9$	$\hat{\theta} = 0.90997$ (0.05542)	$\hat{\theta} = 0.93296$ (0.05990)	$\hat{\theta} = 0.91920$ (0.05185)	+ 43.1295%
$\sigma_a^2 = 1.0$	$\hat{\sigma}_a^2 = 1.01943$ a (0.16271)	$\hat{\sigma}_a^2 = 0.98750$ a (0.15265)	$\hat{\sigma}_a^2 = 1.00681$ a (0.15149)	

Tabla 4.1.3: Resultados de la simulación del modelo $y_t = (1 - \theta B^4)a_t$ con 100 observaciones y 20 realizaciones.

Valor teórico de los parámetros	Procedimiento de estimación			Ganancia en tiempo de cálculo (1)/(2)
	Min. cuadrados con retrovisión	Máx. verosimilit. con el Filtro de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
$\theta = 0.5$ $\sigma^2 = 1.0$ a	$\hat{\theta} = 0.72469$ (0.14219) $\hat{\sigma}^2 = 0.95223$ a (0.16580)	$\hat{\theta} = 0.50987$ (0.12378) $\hat{\sigma}^2 = 0.98513$ a (0.19470)	$\hat{\theta} = 0.50814$ (0.12263) $\hat{\sigma}^2 = 0.98630$ a (0.19439)	+ 111.213%
$\theta = 0.9$ $\sigma^2 = 1.0$ a	$\hat{\theta} = 0.86557$ (0.02066) $\hat{\sigma}^2 = 1.06843$ a (0.16348)	$\hat{\theta} = 0.90933$ (0.13217) $\hat{\sigma}^2 = 0.97763$ a (0.20917)	$\hat{\theta} = 0.88603$ (0.13936) $\hat{\sigma}^2 = 1.00027$ a (0.21955)	+ 104.335%

Tabla 4.1.4: Resultados de la simulación del modelo $y_t = (1 - \theta B^{12})a_t$ con 100 observaciones y 20 realizaciones.

Valor teórico de los parámetros	Procedimiento de estimación			Ganancia en tiempo de cálculo (1)/(2)
	Min. cuadrados con retrovisión	Máx. verosimilit. con el Filtro de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
$\phi_1 = 0.3$	$\hat{\phi}_1 = 0.26378$ (0.08079)	$\hat{\phi}_1 = 0.28065$ (0.08083)	$\hat{\phi}_1 = 0.28663$ (0.08220)	+ 4.45594%
$\phi_2 = 0.5$	$\hat{\phi}_2 = 0.50048$ (0.07359)	$\hat{\phi}_2 = 0.50340$ (0.06791)	$\hat{\phi}_2 = 0.50427$ (0.06578)	
$\sigma_a^2 = 1.0$	$\hat{\sigma}_a^2 = 0.99038$ (0.14604)	$\hat{\sigma}_a^2 = 0.98518$ (0.14420)	$\hat{\sigma}_a^2 = 0.99052$ (0.14469)	
$\phi_1 = 0.5$	$\hat{\phi}_1 = 0.47572$ (0.06936)	$\hat{\phi}_1 = 0.47819$ (0.06941)	$\hat{\phi}_1 = 0.47840$ (0.06912)	+ 4.52516%
$\phi_2 = -0.7$	$\hat{\phi}_2 = -0.68571$ (0.09069)	$\hat{\phi}_2 = -0.67258$ (0.08629)	$\hat{\phi}_2 = -0.67439$ (0.08716)	
$\sigma_a^2 = 1.0$	$\hat{\sigma}_a^2 = 0.98211$ (0.14352)	$\hat{\sigma}_a^2 = 0.97895$ (0.14291)	$\hat{\sigma}_a^2 = 0.98594$ (0.14614)	

Tabla 4.1.5: Resultados de la simulación del modelo $(1 - \phi_1 B - \phi_2 B^2)Y_t = a_t$ con 100 observaciones y 20 realizaciones.

Valor teórico de los parámetros	Procedimiento de estimación			Ganancia en tiempo de cálculo (1)/(2)
	Min. cuadrados con retrovisión	Máx. verosimilit. con el Filtro de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
$\phi_1 = 0.1$	$\hat{\phi}_1 = 0.07320$ (0.05378)	$\hat{\phi}_1 = 0.08883$ (0.05370)	$\hat{\phi}_1 = 0.09401$ (0.05623)	+ 11.3059%
$\phi_2 = 0.8$	$\hat{\phi}_2 = 0.77801$ (0.04842)	$\hat{\phi}_2 = 0.77892$ (0.04805)	$\hat{\phi}_2 = 0.78239$ (0.04625)	
$\sigma_a^2 = 1.0$	$\hat{\sigma}_a^2 = 0.97376$ (0.14160)	$\hat{\sigma}_a^2 = 0.98085$ (0.14206)	$\hat{\sigma}_a^2 = 0.99010$ (0.14413)	

Tabla 4.1.6: Resultados de la simulación del modelo $(1 - \phi_1 B - \phi_2 B^2)y_t = a_t$ con 100 observaciones y 20 realizaciones.

Valor teórico de los parámetros	Procedimiento de estimación			Ganancia en tiempo de cálculo (1)/(2)
	Mín. cuadrados con retrovisión	Máx. verosimilit. con el Filtro de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
$\phi = 0.5$	$\hat{\phi} = 0.38882$ (0.30802)	$\hat{\phi} = 0.45739$ (0.15261)	$\hat{\phi} = 0.47176$ (0.16183)	+ 6.02732%
$\theta = 0.7$	$\hat{\theta} = 0.63336$ (0.28032)	$\hat{\theta} = 0.64435$ (0.13789)	$\hat{\theta} = 0.64707$ (0.13979)	
$\sigma_a^2 = 1.0$	$\hat{\sigma}_a^2 = 0.98036$ (0.14955)	$\hat{\sigma}_a^2 = 1.04581$ (0.15408)	$\hat{\sigma}_a^2 = 1.03297$ (0.11480)	

Tabla 4.1.7: Resultados de la simulación del modelo $(1 - \phi B)y_t = (1 - \theta B)a_t$ con 100 observaciones y 20 realizaciones.

Valor teórico de los parámetros	Procedimiento de estimación			Ganancia en tiempo de cálculo (1)/(2)
	Mín. cuadrados con retrovisión	Máx. verosimilit. con el Filtro de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
$\gamma = 0.5$	$\hat{\gamma} = -0.00138$ (0.02239)	$\hat{\gamma} = 0.40936$ (0.10632)	$\hat{\gamma} = 0.40936$ (0.10632)	+ 0.29859%
$\phi = 0.5$	$\hat{\phi} = 0.47698$ (0.07782)	$\hat{\phi} = 0.47233$ (0.07849)	$\hat{\phi} = 0.47233$ (0.07849)	
$\sigma_a^2 = 1.0$	$\hat{\sigma}_a^2 = 0.98885$ (0.14289)	$\hat{\sigma}_a^2 = 0.98907$ (0.14392)	$\hat{\sigma}_a^2 = 0.98907$ (0.14392)	
$\gamma = 0.1$	$\hat{\gamma} = 0.05448$ (0.24905)	$\hat{\gamma} = 0.08004$ (0.12949)	$\hat{\gamma} = 0.08004$ (0.12949)	+ 0.16970%
$\phi = 0.9$	$\hat{\phi} = 0.87599$ (0.04726)	$\hat{\phi} = 0.86901$ (0.05325)	$\hat{\phi} = 0.86901$ (0.05325)	
$\sigma_a^2 = 1.0$	$\hat{\sigma}_a^2 = 0.98403$ (0.14115)	$\hat{\sigma}_a^2 = 0.98668$ (0.14313)	$\hat{\sigma}_a^2 = 0.98668$ (0.14313)	

Tabla 4.1.8: Resultados de la simulación del modelo $(1 - \phi B)y_t = \gamma + a_t$ con 100 observaciones y 20 realizaciones.



	Valor teórico de los parámetros	Procedimiento de estimación		Ganancia en tiempo de cálculo (1)/(2)
		Máx. verosimilit. con el F. de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
Ratio Señal/Ruido 1	$\phi = 0.5$ $\sigma_y^2 = 1.34$	$\hat{\phi} = 0.41279$ (0.16564) $\hat{\sigma}_y^2 = 1.85037$ Y (0.27214)	$\hat{\phi} = 0.41356$ (0.16611) $\hat{\sigma}_y^2 = 1.84974$ Y (0.27216)	+ 5.42543%
Ratio Señal/Ruido 2	$\phi = 0.5$ $\sigma_y^2 = 0.66$	$\hat{\phi} = 0.45348$ (0.06367) $\hat{\sigma}_y^2 = 0.54222$ Y (0.10944)	$\hat{\phi} = 0.45410$ (0.06402) $\hat{\sigma}_y^2 = 0.54236$ Y (0.10936)	+ 4.33996%
Ratio Señal/Ruido 5	$\phi = 0.5$ $\sigma_y^2 = 0.266$	$\hat{\phi} = 0.46463$ (0.02122) $\hat{\sigma}_y^2 = 0.18987$ Y (0.04198)	$\hat{\phi} = 0.46497$ (0.02135) $\hat{\sigma}_y^2 = 0.19014$ Y (0.04188)	+ 4.26222%
Ratio Señal/Ruido 10	$\phi = 0.5$ $\sigma_y^2 = 0.134$	$\hat{\phi} = 0.46640$ (0.01028) $\hat{\sigma}_y^2 = 0.14418$ Y (0.02140)	$\hat{\phi} = 0.46667$ (0.01034) $\hat{\sigma}_y^2 = 0.14444$ Y (0.02134)	+ 4.24205%

Tabla 4.1.9: Resultados de la simulación del modelo $y_t = \phi y_{t-1} + a_t$ con $\sigma_a^2 = 1.0$ e $y_t^* = y_t + v_{yt}$ con 100 observaciones y 20 realizaciones.

	Valor teórico de los parámetros	Procedimiento de estimación		Ganancia en tiempo de cálculo (1)/(2)
		Máx. verosimilit. con el F. de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
Ratio Señal/Ruido 1	$\phi = 0.9$ $\sigma_y^2 = 1.34$	$\hat{\phi} = 0.83260$ (0.02065) $\hat{\sigma}_y^2 = 1.82626$ Y (0.30348)	$\hat{\phi} = 0.83483$ (0.02108) $\hat{\sigma}_y^2 = 1.82414$ Y (0.30325)	+ 4.39990%
Ratio Señal/Ruido 2	$\phi = 0.9$ $\sigma_y^2 = 0.66$	$\hat{\phi} = 0.84026$ (0.01014) $\hat{\sigma}_y^2 = 0.49561$ Y (0.11459)	$\hat{\phi} = 0.84139$ (0.01040) $\hat{\sigma}_y^2 = 0.49678$ Y (0.11460)	+ 4.47277%
Ratio Señal/Ruido 5	$\phi = 0.9$ $\sigma_y^2 = 0.266$	$\hat{\phi} = 0.84492$ (0.00431) $\hat{\sigma}_y^2 = 0.13807$ Y (0.04010)	$\hat{\phi} = 0.84540$ (0.00444) $\hat{\sigma}_y^2 = 0.13899$ Y (0.03991)	+ 4.36077%
Ratio Señal/Ruido 10	$\phi = 0.9$ $\sigma_y^2 = 0.134$	$\hat{\phi} = 0.84588$ (0.00225) $\hat{\sigma}_y^2 = 0.09125$ Y (0.01991)	$\hat{\phi} = 0.84624$ (0.00232) $\hat{\sigma}_y^2 = 0.09199$ Y (0.01979)	+ 4.29979%

Tabla 4.1.10: Resultados de la simulación del modelo $y_t = \phi y_{t-1} + a_t$ con $\sigma_a^2 = 1.0$ e $y_t^* = y_t + v_{yt}$ con 100 observaciones y 20 realizaciones.

Valor teórico de los parámetros	Procedimiento de estimación		Ganancia en tiempo de cálculo (1)/(2)
	Máx. verosimilit. con el F. de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
$\phi = 0.50$	$\hat{\phi} = 0.48933$ (0.08043)	$\hat{\phi} = 0.48933$ (0.08043)	+ 2.35014%
$\sigma_1^2 = 1.00$	$\hat{\sigma}_1^2 = 1.00650$ (0.14031)	$\hat{\sigma}_1^2 = 1.00650$ (0.14031)	
$\sigma_2^2 = 1.25$	$\hat{\sigma}_2^2 = 1.08009$ (0.06580)	$\hat{\sigma}_2^2 = 1.08009$ (0.06580)	
$\sigma_{12}^2 = 0.56$	$\hat{\sigma}_{12}^2 = 0.50460$ (0.07928)	$\hat{\sigma}_{12}^2 = 0.50460$ (0.07928)	

Tabla 4.1.11: Resultados de la simulación del modelo bivalente

$(1 - \phi_1 B)z_{1t} = v_{1t}$; $z_{2t} = v_{2t}$ siendo $v_{2t} = \rho v_{1t} + a_t$
con 100 observaciones y 20 realizaciones.

Valor teórico de los parámetros	Procedimiento de estimación		Ganancia en tiempo de cálculo (1)/(2)
	Máx. verosimilit. con el F. de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
$\theta_1 = 0.70$	$\hat{\theta}_1 = 0.70642$ (0.12027)	$\hat{\theta}_1 = 0.70595$ (0.11997)	+ 6.94051%
$\theta_2 = 0.50$	$\hat{\theta}_2 = 0.46942$ (0.04123)	$\hat{\theta}_2 = 0.47383$ (0.04236)	
$\sigma_1^2 = 1.00$	$\hat{\sigma}_1^2 = 0.99710$ (0.16042)	$\hat{\sigma}_1^2 = 1.00175$ (0.16110)	
$\sigma_2^2 = 1.25$	$\hat{\sigma}_2^2 = 1.08522$ (0.06687)	$\hat{\sigma}_2^2 = 1.08621$ (0.06681)	
$\sigma_{12}^2 = 0.56$	$\hat{\sigma}_{12}^2 = 0.49138$ (0.08894)	$\hat{\sigma}_{12}^2 = 0.49288$ (0.08898)	

Tabla 4.1.12: Resultados de la simulación del modelo bivariante
 $z_{1t} = (1 - \theta_1 B)v_{1t}$; $z_{2t} = (1 - \theta_2 B)v_{2t}$ siendo
 $v_{2t} = \rho v_{1t} + a_t$ con 100 observaciones y 20 realizaciones.

Valor teórico de los parámetros	Procedimiento de estimación		Ganancia en tiempo de cálculo (1)/(2)
	Máx. verosimilit. con el F. de Kalman (1)	Máx. verosimilit. con el F. Chandrasekhar (2)	
$\phi = 0.50$	$\hat{\phi} = 0.46655$ (0.08351)	$\hat{\phi} = 0.46653$ (0.08358)	+ 1.48418%
$\theta = 0.70$	$\hat{\theta} = 0.71767$ (0.14706)	$\hat{\theta} = 0.73588$ (0.18034)	
$\omega_0 = 2.00$	$\hat{\omega}_0 = 2.01038$ (1.12137)	$\hat{\omega}_0 = 2.01023$ (1.12119)	
$\omega_1 = 2.00$	$\hat{\omega}_1 = 2.32448$ (0.45076)	$\hat{\omega}_1 = 2.32348$ (0.44749)	
$\sigma_1^2 = 1.00$	$\hat{\sigma}_1^2 = 1.00625$ (0.13054)	$\hat{\sigma}_1^2 = 1.00626$ (0.13054)	
$\sigma_2^2 = 1.25$	$\hat{\sigma}_2^2 = 1.03237$ (0.10345)	$\hat{\sigma}_2^2 = 1.01663$ (0.12549)	
$\sigma_{12}^2 = -0.03$	$\hat{\sigma}_{12}^2 = -0.03066$ (0.00893)	$\hat{\sigma}_{12}^2 = -0.02852$ (0.00912)	

Tabla 4.1.13: Resultados de la simulación del modelo bivariante

$(1 - \phi B)z_{1t} = \omega_0 \xi_t^{1/t^*} + v_{1t}$; $z_{2t} = \omega_1 \xi_t^{1/t^*} + (1 - \theta B)v_{2t}$
siendo $v_{2t} = \rho v_{1t} + a_t$ con 100 observaciones y 20 realizaciones.

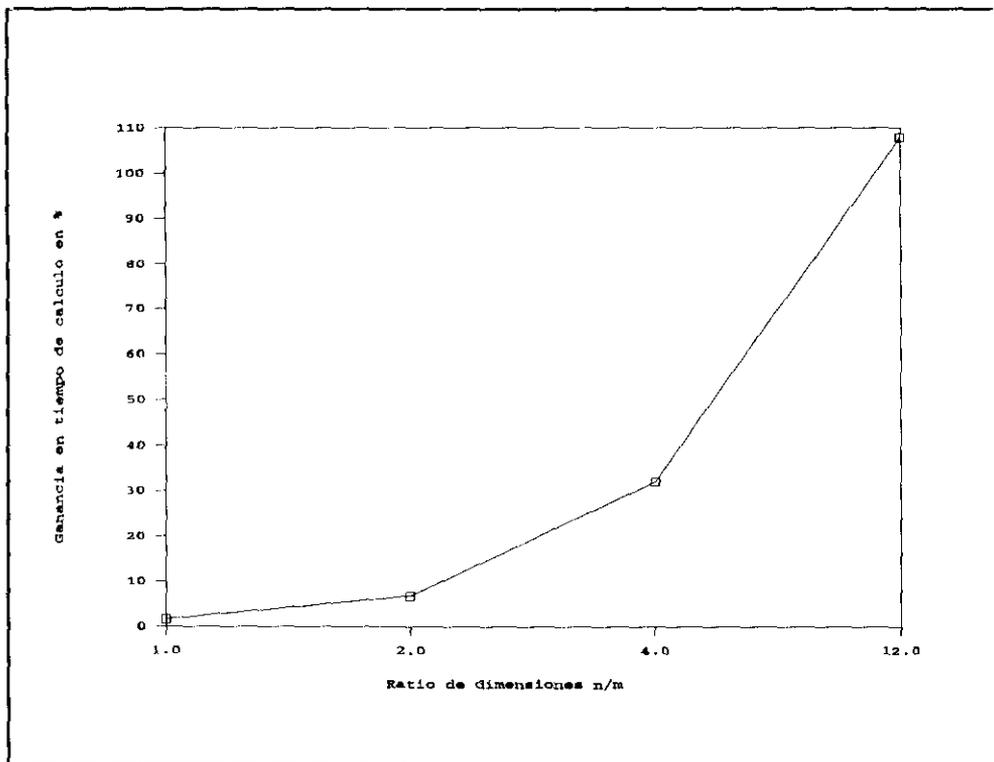


Fig. 4.2.14: Ganancia media en tiempo de cálculo (en %) para distintos ratios n/m .

4.2. Modelos del Índice Dow-Jones.

Esta serie fue analizada por Anderson (1976) de acuerdo con la metodología Box-Jenkins (1976). La muestra utilizada abarca el periodo del 28 de agosto al 18 de diciembre de 1972, disponiéndose, por tanto, de 78 observaciones.

Anderson considera que la serie del Índice Dow-Jones (a partir de ahora, denotada como DW_t) es estacionaria en media diferenciándola una vez. Y con este orden de diferenciación, propone como mejor representación univariante del Índice un proceso AR(1).

Sin embargo, el análisis detallado de esta serie muestra que parece más apropiado tomar dos diferencias regulares (vid. Figuras 4.2.1, 4.2.2 y 4.2.3). En la Figura 4.2.2 se muestra la serie ∇DW_t estandarizada. En ella, se observa la necesidad de tomar una diferencia regular adicional al deambular la serie alrededor de su media. La Figura 4.2.3 sugiere que la serie $\nabla^2 DW_t$ puede ser considerada estacionaria. Aunque no se ofrecen aquí, el examen de las funciones de autocorrelación simple (acf) y parcial (pacf) de la serie con dos diferencias regulares, muestra claramente la presencia de estructura media móvil de orden uno.

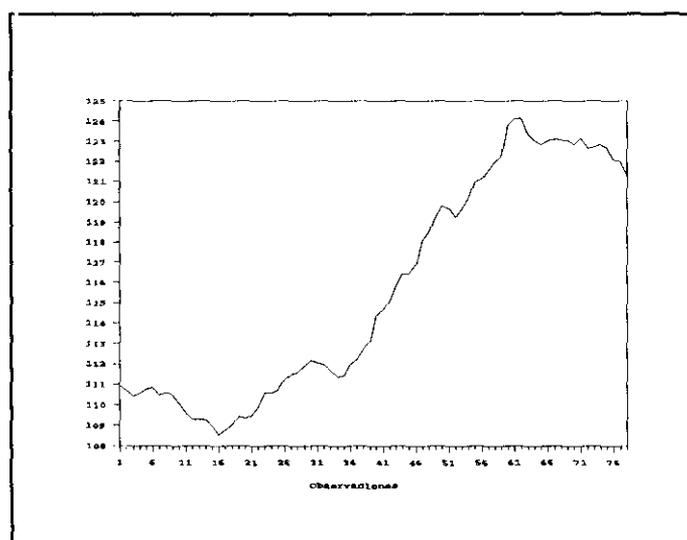


Fig 4.2.1: Serie original del Índice Dow-Jones.

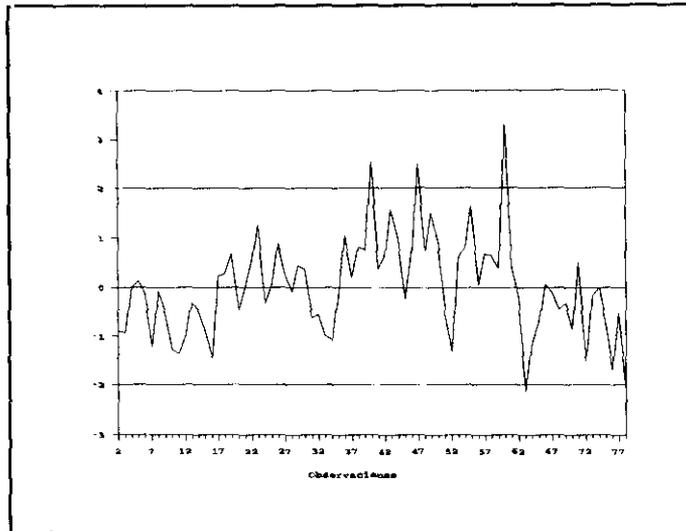


Fig. 4.2.2: Serie ∇DW_t estandarizada.

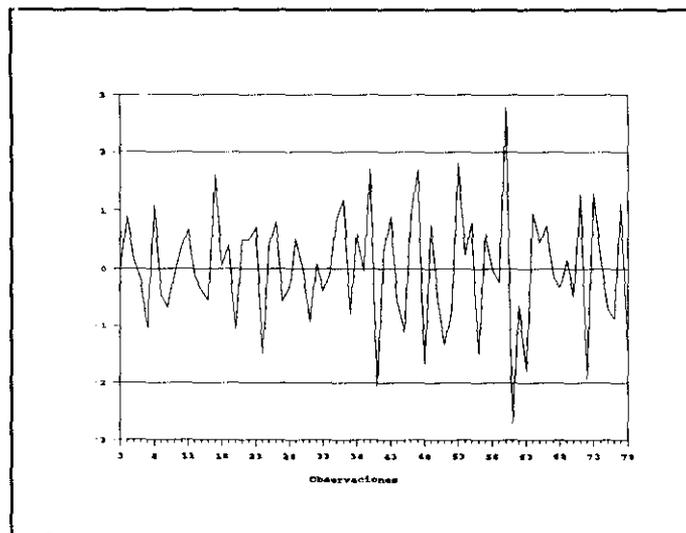


Fig. 4.2.3: Serie $\nabla^2 DW_t$ estandarizada.

Se ha trabajado con dos modelos univariantes para este Índice:

$$(1 - \phi B) \nabla DW_t = a_t \quad [4.2.1]$$

$$\nabla^2 DW_t = (1 - \theta B) a_t \quad [4.2.2]$$

A continuación, se estudia por separado cada una de estas representaciones univariantes del Índice.

4.2.1. Modelo AR(1) para el Índice Dow-Jones.

El resultado de la estimación de este modelo por los tres procedimientos de estimación considerados es:

Método de Estim.	Min. cuadrados con retroprevisión	Máx. verosimil. con el F. Kalman	Máx. verosimil. con el F. Chandrasekhar
Modelo [4.2.1]	$\hat{\phi} = 0.506$ (0.101) $\hat{\sigma}_a^2 = 0.149$	$\hat{\phi} = 0.504$ (0.099) $\hat{\sigma}_a^2 = 0.151$ a (0.024)	$\hat{\phi} = 0.504$ (0.099) $\hat{\sigma}_a^2 = 0.151$ a (0.024)

Tabla 4.2.1.1: Resultados de la estimación del modelo [4.2.1].

donde $\hat{\sigma}_a^2$ es la varianza residual resultante del ajuste y entre paréntesis, se ofrecen las desviaciones típicas de los parámetros estimados.

La Tabla 4.2.1.1 sugiere los siguientes comentarios:

- La estimación del proceso AR(1) es robusta frente a distintos procedimientos de estimación. Se observa que tanto la estimación del parámetro autorregresivo como la de la varianza residual, prácticamente coinciden con los tres métodos. Y en los dos mecanismos de máxima verosimilitud utilizados, las estimaciones coinciden exactamente. Hay que señalar que este hecho se debe a que la función de verosimilitud de este modelo está muy bien condicionada. En las Figuras 4.2.1.2 y 4.2.1.3 se presenta el perfil de la función de verosimilitud de la muestra en un entorno del óptimo, permitiendo variar los valores de los parámetros ϕ y σ_a^2 , respectivamente. En ambos casos, los valores de la función han sido multiplicados por un factor de escala de 10^{16} .
- Con carácter general, cabe indicar que en el procedimiento de mínimos cuadrados con retroprevisión no se ofrecen las desviaciones típicas de las varianzas

residuales estimadas, ya que éstas no son consideradas como parámetros propios del modelo. Es decir, se calculan a posteriori sobre los residuos resultantes del ajuste.

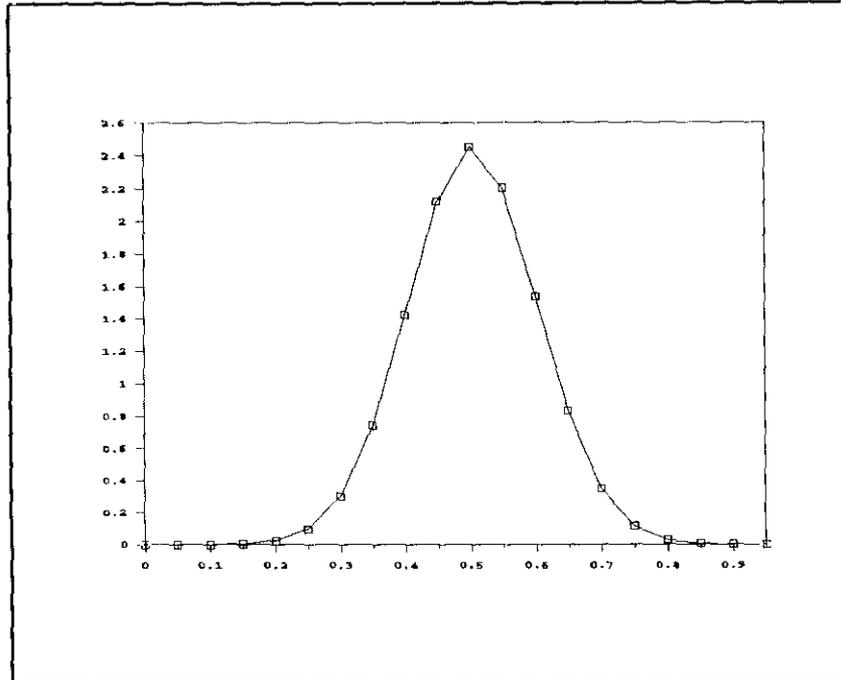


Fig. 4.2.1.2: Perfil de la función de verosimilitud con respecto a ϕ .

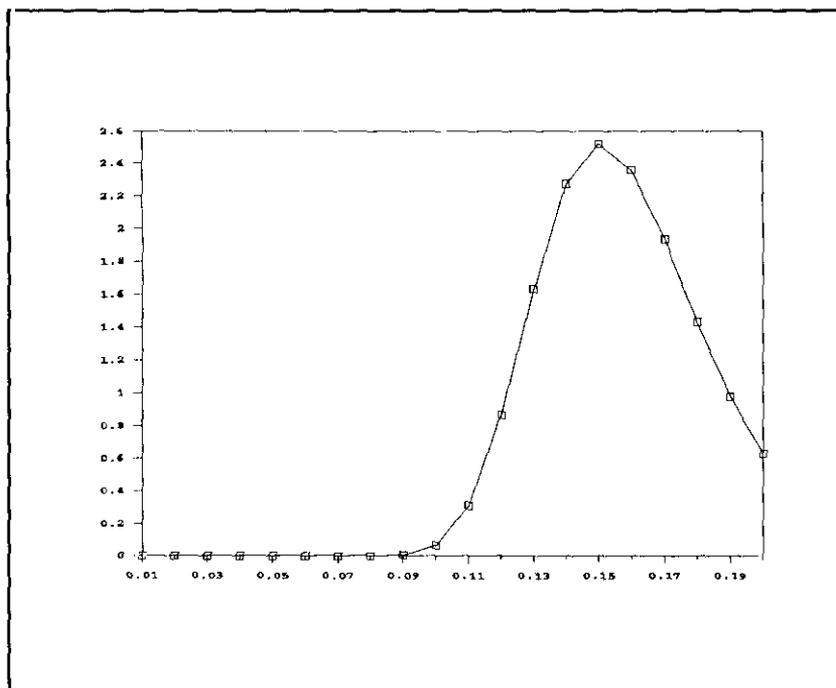


Fig. 4.2.1.3: Perfil de la función de verosimilitud con respecto a σ_B^2 .

CPU	Máx.verosim. con el F. Kalman	Máx.verosim. con el F.Chandrasek.	Ganancia (%)
Proc.INTEL 286 (10Mhz)	104.69	99.63	+ 5.08
Proc.INTEL 386 (20Mhz)	36.04	35.76	+ 0.78

Tabla 4.2.1.4: Tiempos de cálculo en segundos en la estimación del modelo [4.2.1].

La Tabla 4.2.1.4 muestra una mínima ganancia porcentual en tiempo de estimación cuando se utiliza la aproximación de Chandrasekhar, ganancia que varía dependiendo del procesador utilizado.

Por último, se estudia con este ejemplo la seguridad y estabilidad numéricas de los dos filtros utilizados para su estimación. Dado que en este caso, $n=m=1$, todas las matrices que aparecen en las ecuaciones de los filtros son escalares. Por ello, es fácil escribir las expresiones teóricas de los filtros en función de los parámetros a estimar.

Expresiones teóricas del filtro de Chandrasekhar para un modelo AR(1).

Dada la formulación en espacio de los estados de un proceso AR(1), se obtienen las siguientes condiciones iniciales del filtro de Chandrasekhar para modelos estacionarios:

$$P_0 = \frac{\phi^2 \sigma_a^2}{1 - \phi^2} \quad ; \quad B_0 = \frac{\sigma_a^2}{1 - \phi^2}$$

$$K_0 = \phi \quad ; \quad Y_0 = \frac{\phi \sigma_a^2}{1 - \phi^2} \quad ; \quad M_0 = \frac{\phi^2 - 1}{\sigma_a^2}$$

Como puede verse, la varianza inicial del vector de estado y del vector de observación son positivas si $|\phi| < 1$. Pero esta condición se cumple al trabajar con un modelo estacionario. La

matriz factor M_0 (en este caso, escalar) es definida negativa si el proceso es estacionario.

Dadas estas condiciones iniciales, las variables del filtro $\forall t \geq 1$ son:

$$\begin{aligned} B_t &= \sigma_a^2 \\ K_t &= \phi \\ M_t &= -(1/\sigma_a^2) \\ Y_t &= 0 \end{aligned}$$

donde se comprueba que, al utilizar la versión del filtro de Chandrasekhar para modelos estacionarios (vid. Capítulo 1), se mantiene el carácter de definida negativa de M_t , a partir de su condición inicial.

Aunque en esta aproximación es posible calcular, en cada instante de tiempo, la ganancia del filtro sin propagar la matriz de covarianzas del vector de estado, ésta última puede obtenerse de forma recursiva a partir de la expresión [1.4.1.20]. En este ejemplo, se tiene que:

$$P_t = 0 \qquad \forall t \geq 1$$

Por tanto, en un proceso AR(1), teóricamente la varianza del vector de estado ha de ser nula en todo instante de tiempo.

Expresiones teóricas del filtro de Kalman para un modelo AR(1).

La expresión de la varianza del vector de estado inicial es la misma que en el filtro de Chandrasekhar. El resto de las variables del filtro de Kalman para este modelo son:

$$B_t = \begin{cases} \frac{\sigma_a^2}{1 - \phi^2} & \text{si } t = 1 \\ \sigma_a^2 & \forall t \geq 2 \end{cases}$$

$$K_t = \phi \quad \forall t \geq 1$$

$$P_t = 0 \quad \forall t \geq 1$$

Se observa, por tanto, que la única diferencia existente en las expresiones de los dos filtros es que la matriz de covarianzas del vector de observación es, en el instante $t=1$, mayor en el filtro de Kalman que en el de Chandrasekhar. Además, teóricamente la varianza del vector de estado ha de ser nula en todo instante de tiempo. Esto significa que el filtro debería ser asintóticamente estable dado que se cumple que:

$$\lim_{t \rightarrow \infty} P_{t+1/t} = \bar{P} = 0$$

En la Figura 4.2.1.5 se representa la evolución de la varianza estimada del vector de estado una vez que el algoritmo de optimización ha llegado a un vector de parámetros óptimo. En ella, se aprecia cómo los errores numéricos en la propagación, hacen que esta varianza sea estrictamente negativa cuando se calcula con el filtro de Kalman. Además, se observa que dicha varianza oscila constantemente entre dos valores, con lo que se muestra la inestabilidad, aunque sólo numérica, del filtro de Kalman. Por el contrario, la varianza del vector de estado obtenida a partir del filtro de Chandrasekhar es un número estrictamente positivo y constante en todo instante de tiempo, como cabía esperar de su expresión teórica.

Dadas las ecuaciones del filtro de Kalman, los errores numéricos cometidos en la propagación de P_t se transmiten al cálculo de B_t . Esto no se puede apreciar en la Figura 4.2.1.6,

(que muestra la evolución de B_t), al ser los errores numéricos cometidos muy pequeños. Lo que se observa en dicho gráfico, coincide con lo que se dedujo antes teóricamente: en un primer instante, la varianza del vector de observación es mayor para el filtro de Kalman, coincidiendo después para ambos filtros.

La evolución del vector de estado es muy similar cuando se estima con los dos filtros, como lo muestra la Figura 4.2.1.7. La Figura 4.2.1.8 presenta la evolución de la ganancia calculada por ambos filtros, viendo que es estacionaria en la última iteración del algoritmo de optimización y que su valor coincide con la estimación de ϕ en ese momento (es decir, en el óptimo).

Como resumen de este ejemplo cabe señalar que, a pesar de ser el modelo más sencillo concebible y del buen condicionamiento de la función de verosimilitud, se aprecia una ligera ganancia tanto en tiempo de cálculo como en precisión numérica utilizando las ecuaciones de Chandrasekhar en lugar del filtro de Kalman.

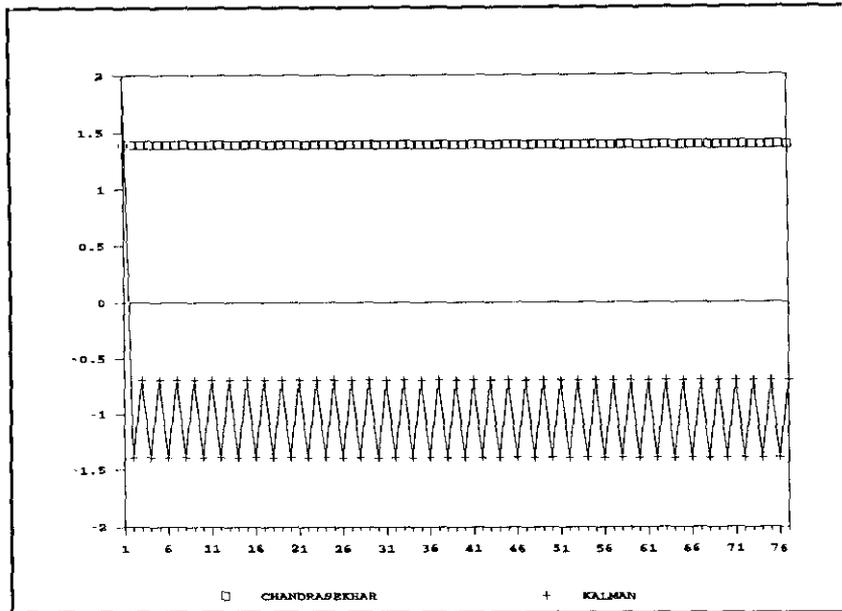


Fig. 4.2.1.5: Varianza estimada del vector de estado del modelo [4.2.1].

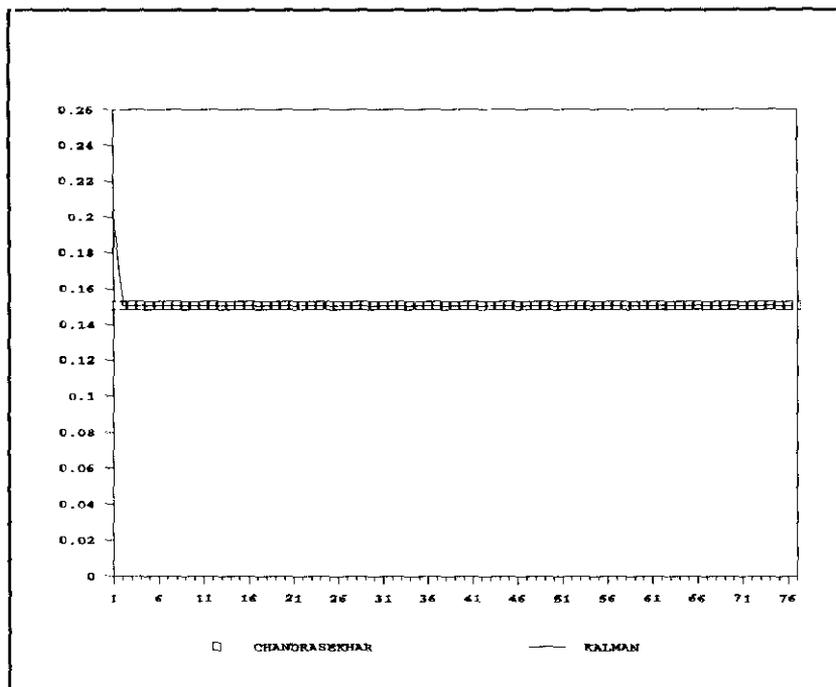


Fig. 4.2.1.6: Varianza estimada del vector de observación del modelo [4.2.1].

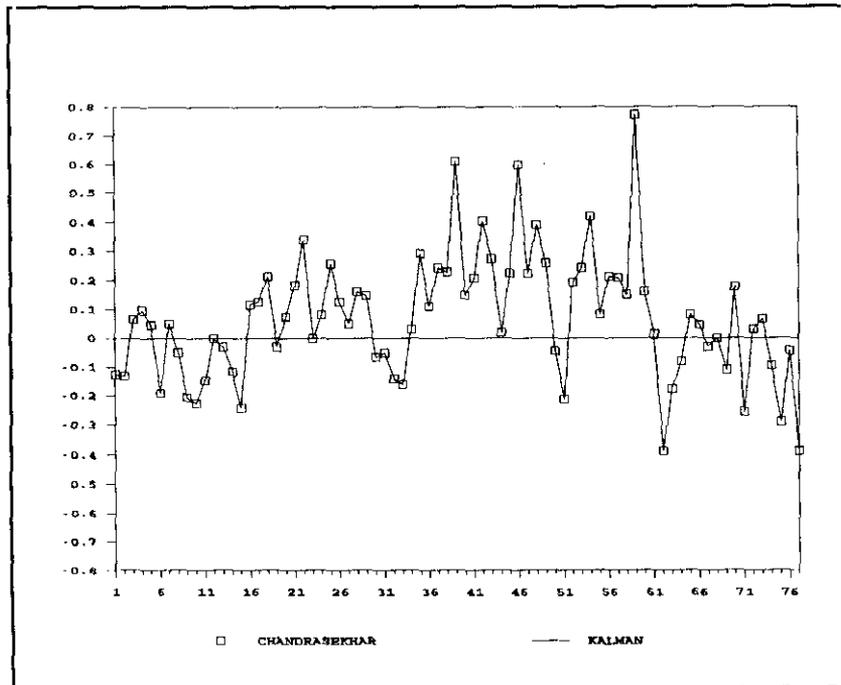


Fig. 4.2.1.7: Estimación del vector de estado del modelo [4.2.1].

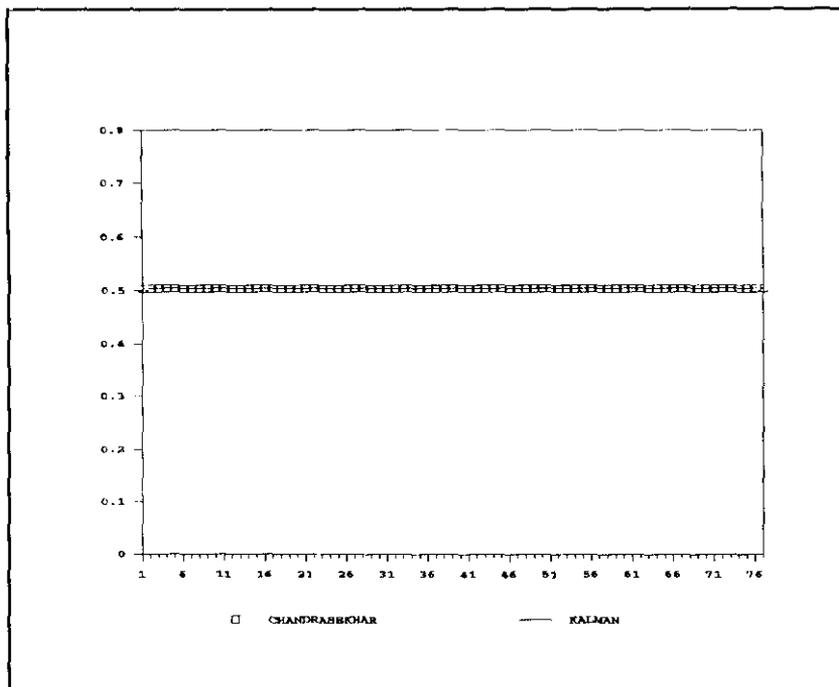


Fig. 4.2.1.8: Evolución de la ganancia calculada del modelo [4.2.1].

4.2.2. Modelo MA(1) para el Índice Dow-Jones.

Una representación univariante que, en nuestra opinión, resulta más adecuada que la propuesta por Anderson, consiste en describir mediante un proceso MA(1) la segunda diferencia regular del Índice.

La estimación de los parámetros de este modelo con los tres métodos utilizados se ofrece en la siguiente tabla:

Método de Estim.	Mín. cuadrados con retrovisión	Máx. verosimil. con el F. Kalman	Máx. verosimil. con el F.Chandrasekhar
Modelo [4.2.2]	$\hat{\theta} = 0.735$ (0.081) $\hat{\sigma}^2 = 0.150$ a	$\hat{\theta} = 0.722$ (0.083) $\hat{\sigma}^2 = 0.152$ a (0.025)	$\hat{\theta} = 0.724$ (0.082) $\hat{\sigma}^2 = 0.153$ a (0.025)

Tabla 4.2.2.1: Resultados de la estimación del modelo [4.2.2].

A la vista de los resultados de la Tabla 4.2.2.1, pueden hacerse los siguientes comentarios:

- La estimación del parámetro media móvil es ligeramente más alta en el caso del procedimiento de mínimos cuadrados con retrovisión que en los otros dos métodos de máxima verosimilitud utilizados. De nuevo, esto puede ser importante a la hora de reformular un modelo cuando este parámetro esté cerca de la no invertibilidad.
- A su vez, la estimación de θ es ligeramente mayor cuando este modelo se estima utilizando la aproximación de Chandrasekhar que cuando se aplica el filtro de Kalman. Esto no es de extrañar, ya que como se observa en las Figuras 4.2.2.2 y 4.2.2.3, el perfil de la función de verosimilitud con respecto a cada parámetro es casi plano en un entorno de puntos que contiene las estimaciones de los parámetros obtenidas con los tres métodos.

Como en el modelo [1], en dichas figuras, los valores de la función de verosimilitud han sido multiplicados por un factor de escala de 10^{16} .

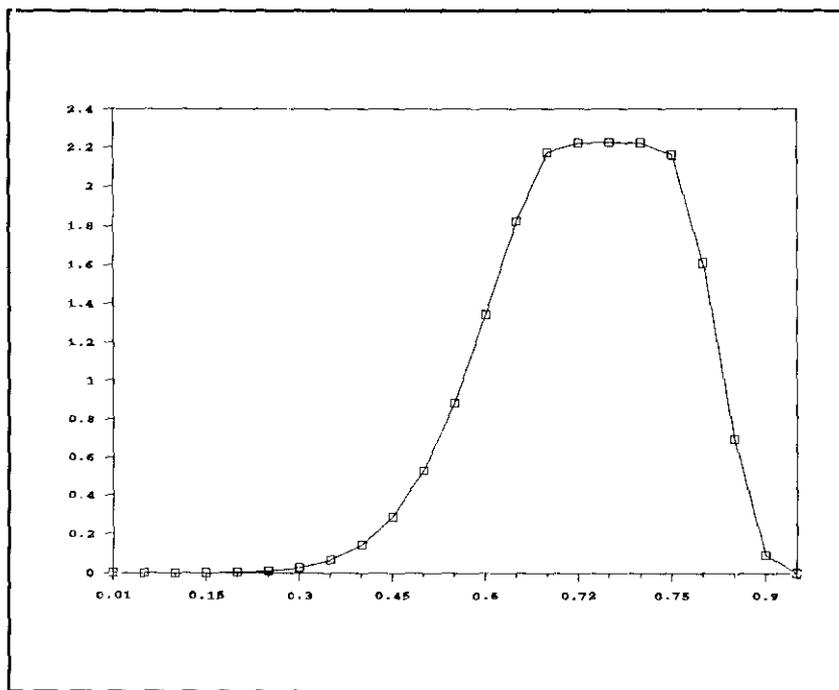


Fig 4.2.2.2: Perfil de la función de verosimilitud con respecto a θ .

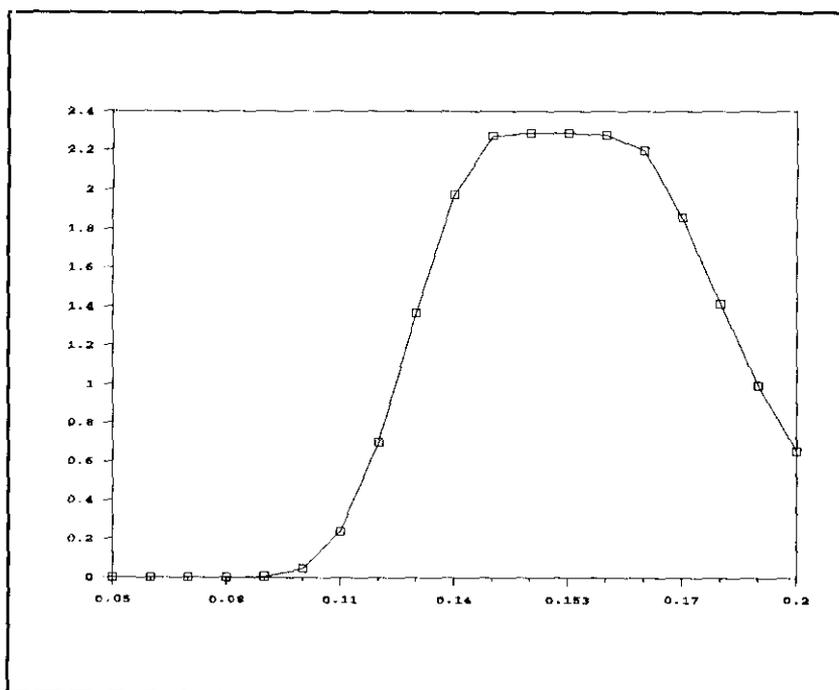


Fig 4.2.2.3: Perfil de la función de verosimilitud con respecto a σ_a^2 .

En la Tabla 4.2.2.4 se ofrece el tiempo usado en la estimación de este modelo por los dos métodos de máxima verosimilitud considerados. En este caso, también son escalares las dimensiones del vector de estado y de observación. Sin embargo, existe una pequeña ganancia computacional cuando se aplica el método de máxima verosimilitud con el filtro de Chandrasekhar.

CPU	Máx.verosim. con el F. Kalman	Máx.verosim. con el F.Chandrasek.	Ganancia (%)
Proc.INTEL 286 (10Mhz)	103.48	101.72	+ 1.73
Proc.INTEL 386 (20Mhz)	35.70	35.04	+ 1.88

Tabla 4.2.2.4: Tiempos de cálculo en segundos en la estimación del modelo [4.2.2].

La Figura 4.2.2.5 muestra que éste es un ejemplo en el que se obtiene una varianza del vector de estado estrictamente positiva con los dos filtros utilizados para su cálculo. Además, esa varianza converge muy pronto a un valor estacionario cercano a cero.

La Figura 4.2.2.6 representa la evolución temporal de la varianza del vector de observación calculada con ambos filtros. Se observa que esta varianza conserva siempre su propiedad de semidefinida positiva con los dos métodos. La única diferencia apreciable es que el filtro de Kalman calcula una varianza mayor en los primeros instantes de tiempo que el de Chandrasekhar, convirtiéndose después en una constante positiva e idéntica para los dos filtros. La divergencia apreciada en este caso, se explica porque la varianza del vector de observación en los dos primeros instantes de tiempo es:

- Según el filtro de Kalman:

$$B_1 = \sigma_a^2(1 + \theta^2)$$

$$B_2 = \frac{\sigma_a^2(1 + \theta^2 + \theta^4)}{(1 + \theta^2)}$$

• Según el filtro de Chandrasekhar:

$$B_1 = \frac{\sigma_a^2(1 + \theta^2 + \theta^4)}{(1 + \theta^2)} = \sigma_a^2(1 + \theta^2) - \frac{\theta^2\sigma_a^2}{(1 + \theta^2)}$$

$$B_2 = \frac{\sigma_a^2(1 + \theta^2 + \theta^4)}{(1 + \theta^2)} - \frac{\theta^4\sigma_a^2}{(1 + \theta^2)(1 + \theta^2 + \theta^4)}$$

Iterando sucesivamente, puede comprobarse que esto ocurre hasta un determinado momento, a partir del cual, esta varianza se convierte en una constante numéricamente idéntica para los dos filtros. Esto se debe al siguiente hecho: dado que la varianza del vector de observación con el filtro de Kalman se calcula a partir de la expresión $B_t = P_{t/t-1} + \sigma^2$ y $P_{t/t-1}$ se hace casi nula a partir de un instante de tiempo, B_t convergerá también a una constante positiva que coincide con la varianza residual del modelo. Lo mismo ocurre para el filtro de Chandrasekhar, al propagar B_t de acuerdo con $B_t = B_{t-1} + P_t - P_{t-1}$.

La Figura 4.2.2.7 muestra la propagación del vector de estado estimado con los dos procedimientos de máxima verosimilitud considerados. Se observa que esta estimación prácticamente coincide con los dos filtros. Además, se aprecia que el vector de estado es más estacionario en media que en el ejemplo anterior. La razón es que la estimación del vector de estado depende, entre otras cosas, de los datos, y en este caso, la serie del Índice Dow-Jones es más estacionaria al haber sido diferenciada dos veces.

La Figura 4.2.2.8 representa la evolución de la ganancia calculada por los dos filtros a lo largo de la muestra, observándose cómo alcanza muy pronto un estado estacionario.

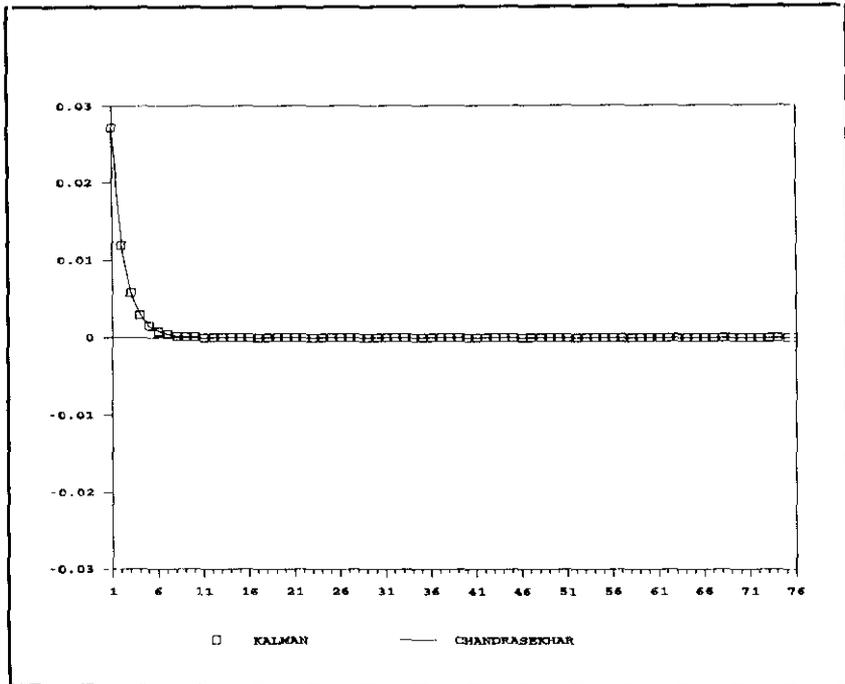


Fig. 4.2.2.5: Varianza estimada del vector de estado del modelo [4.2.2].

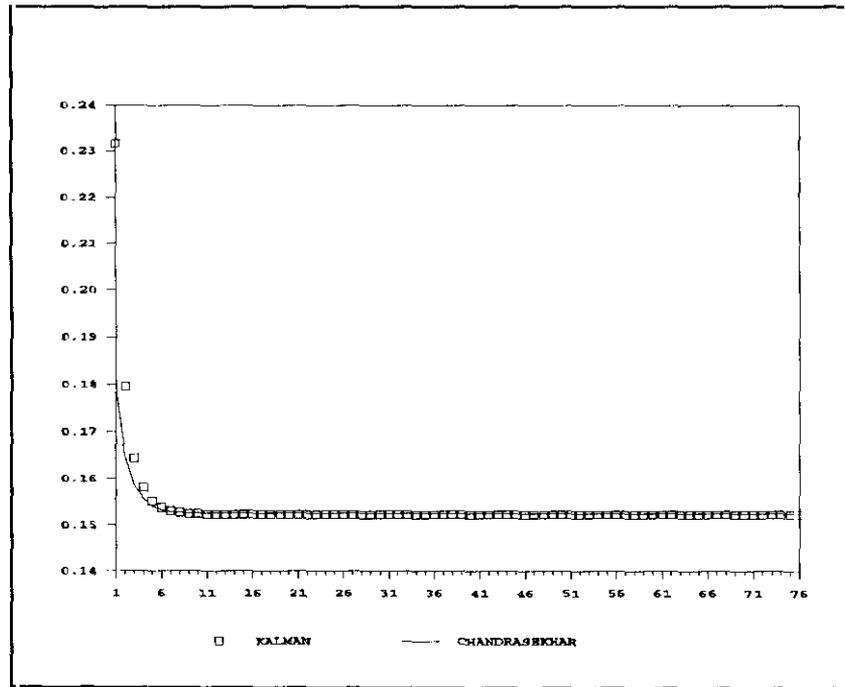


Fig. 4.2.2.6: Varianza estimada del vector de observación del modelo [4.2.2].

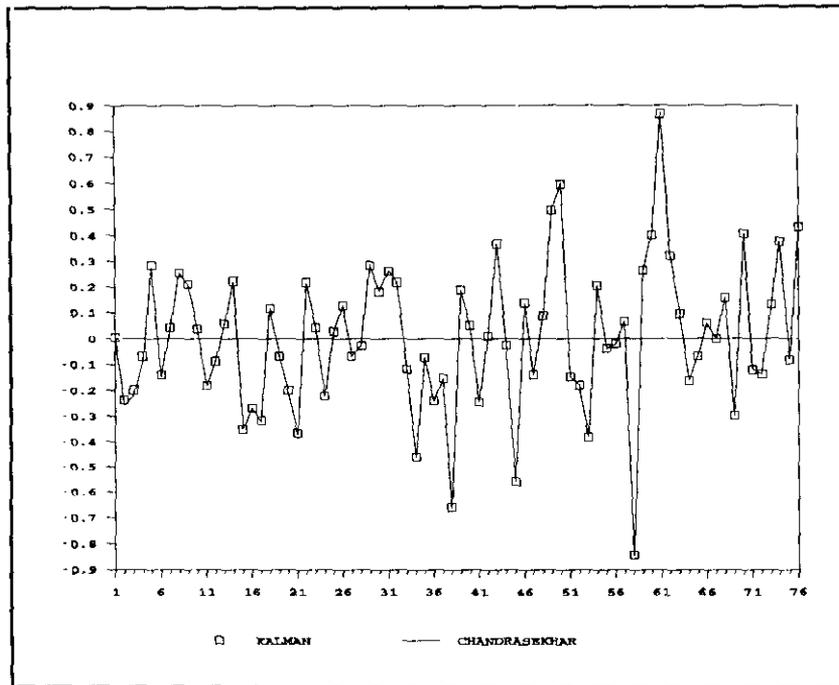


Fig. 4.2.2.7: Estimación del vector de estado del modelo [4.2.2].

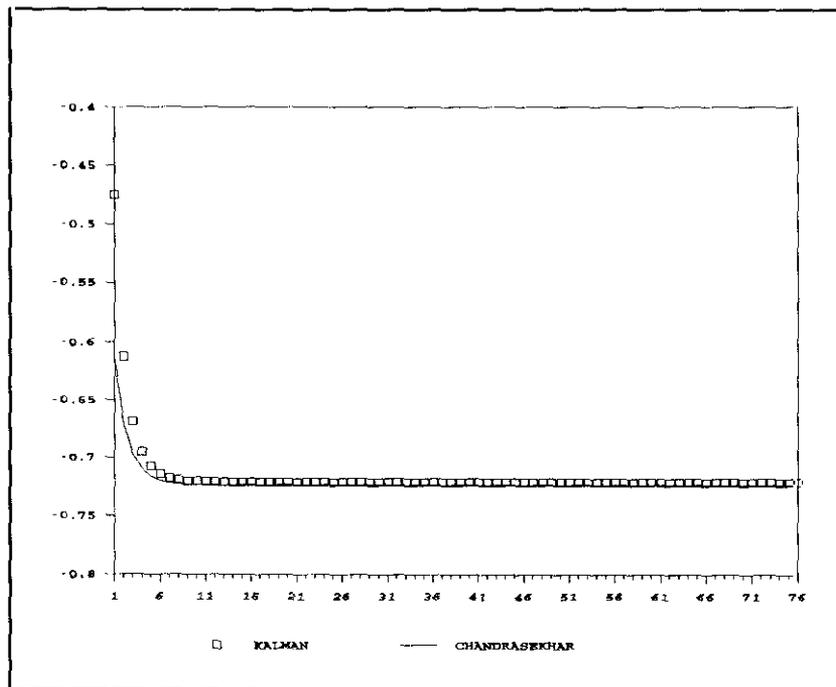


Fig. 4.2.2.8: Evolución de la ganancia calculada del modelo [4.2.2].

En la Figura 4.2.2.9 se ofrece el perfil de la serie residual cuando se ha llegado al vector de parámetros que maximizan la función de verosimilitud del modelo. Es decir, se presenta la evolución de los residuos resultantes del ajuste del Índice Dow-Jones a través de un proceso MA(1). En él, se observa la existencia de dos residuos que superan las bandas de significación de $\pm 2DS$ fechados en las observaciones 60 y 61. Esto sugiere la necesidad de intervenir dichos valores atípicos introduciendo en el modelo una variable determinista de tipo escalón en la observación 60.

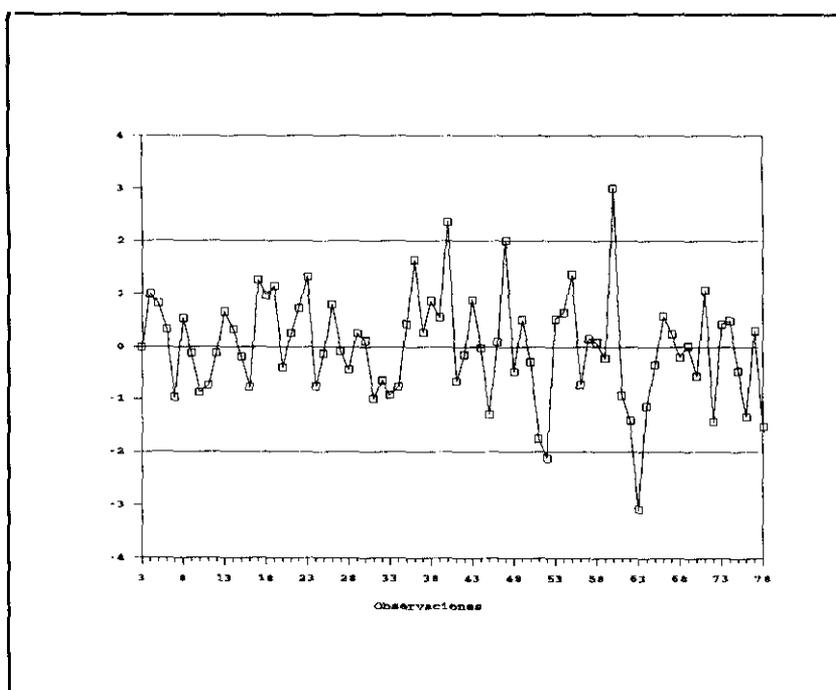


Fig. 4.2.2.9: Residuos estandarizados del modelo [4.2.2].

Es decir, el modelo univariante de intervención a estimar es:

$$DWJ_t = \omega_0 \xi_t^{S/60} + N_t$$

con una estructura para el ruido:

$$\nabla^2 N_t = (1 - \theta B) a_t$$

en donde:

a_t es un proceso de ruido blanco,

$$\xi_t^{S/60} = \begin{cases} 1 & \text{si } t \geq 60 \\ 0 & \text{en otro caso} \end{cases}$$

Esta especificación puede escribirse alternativamente de la forma:

$$\nabla^2 \text{DWJ}_t = \omega_0 \xi_t^{\text{IC}/60} + (1 - \theta B) a_t \quad [4.2.3]$$

siendo:

$$\xi_t^{\text{IC}/60} = \begin{cases} 1 & \text{si } t = 60 \\ -1 & \text{si } t = 61 \\ 0 & \text{en otro caso} \end{cases}$$

que da lugar a la siguiente formulación en espacio de los estados:

$$\begin{aligned} x_{t+1} &= -\theta a_t \\ \nabla^2 \text{DWJ}_t &= x_t + \omega_0 \xi_t^{\text{IC}/60} + a_t \end{aligned}$$

donde:

$$z_t = \nabla^2 \text{DWJ}_t \quad ; \quad v_t = w_t = a_t \quad ; \quad Q = R = S = \sigma_a^2$$

Los resultados de la estimación de este modelo univariante de intervención se ofrecen en la siguiente tabla:

Método de Estim.	Mín. cuadrados con retrovisión	Máx. verosimil. con el F. Kalman	Máx. verosimil. con el F. Chandrasekhar
Modelo [4.2.3]	$\hat{\omega} = 1.3743$ $0 (0.3220)$ $\hat{\theta} = 0.6985$ (0.0851) $\hat{\sigma}_a^2 = 0.1212$	$\hat{\omega} = 1.3726$ $0 (0.3220)$ $\hat{\theta} = 0.6879$ (0.0860) $\hat{\sigma}_a^2 = 0.1229$ $a (0.0201)$	$\hat{\omega} = 1.3729$ $0 (0.3225)$ $\hat{\theta} = 0.6896$ (0.0859) $\hat{\sigma}_a^2 = 0.1231$ $a (0.0201)$

Tabla 4.2.2.10: Resultados de la estimación del modelo [4.2.3].

donde la correlación existente entre el parámetro media móvil y el asociado a la variable de intervención de 0.08 por el método de mínimos cuadrados con retrovisión. Sin embargo, la matriz de correlaciones resultante cuando se utiliza cualquiera de los métodos de máxima verosimilitud es la siguiente:

$$\begin{array}{ccc}
 \hat{\omega}_0 & \hat{\theta} & \hat{\sigma}_a^2 \\
 \left[\begin{array}{ccc}
 1.0 & & \\
 0.0 & 1.0 & \\
 0.0 & -0.01 & 1.0
 \end{array} \right]
 \end{array}$$

Por tanto, se observa que sólo al utilizar la expresión exacta de la matriz de información se obtiene una correlación estrictamente nula entre los parámetros $\hat{\omega}_0$ y $\hat{\theta}$, como es de esperar teóricamente.

La Tabla 4.2.2.10 muestra que el coeficiente asociado a la variable de intervención es altamente significativo y su estimación es muy similar con los tres métodos. Además, al haber incluido esta variable determinista, la estimación del parámetro media móvil es más parecida con los tres criterios que en la especificación [4.2.2]. Esto se debe a que la corrección de ese efecto impulso transitorio ha mejorado el condicionamiento de la función de verosimilitud de la muestra (vid. Figuras 4.2.2.11, 4.2.2.12 y 4.2.2.13). Lógicamente, la varianza residual de este ajuste también ha disminuido con respecto a la especificación [4.2.2].

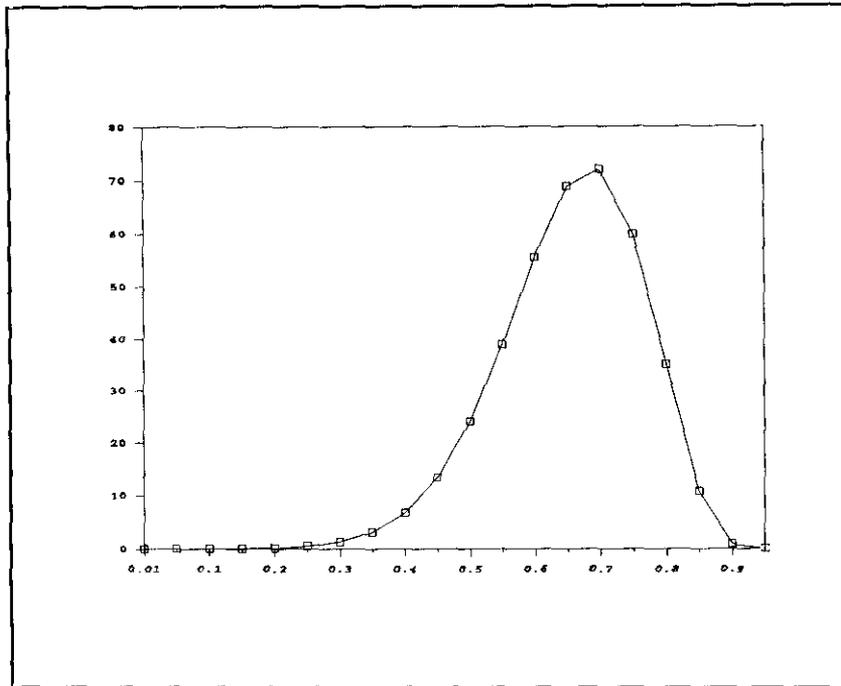


Fig. 4.2.2.11: Perfil de la función de verosimilitud con respecto a θ .

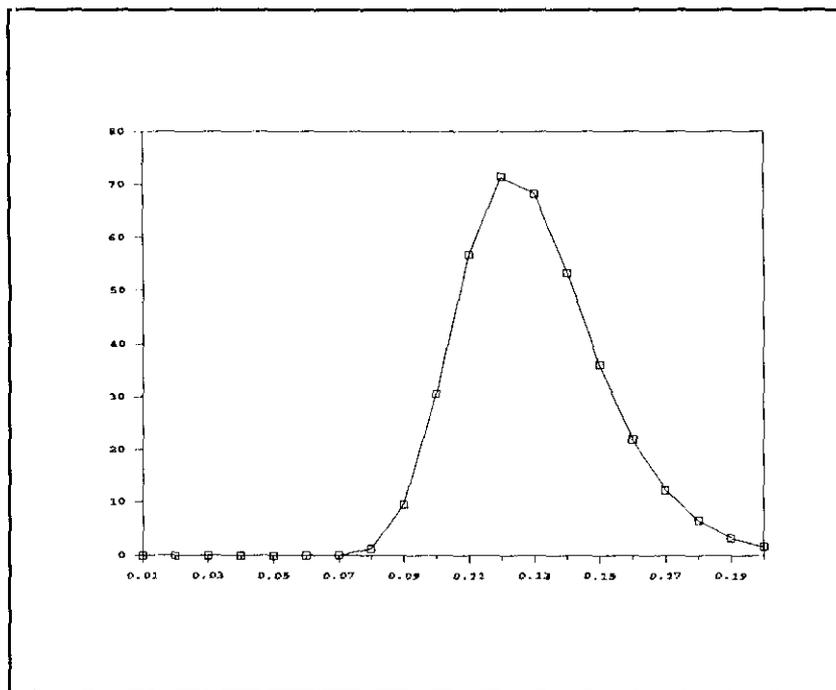


Fig. 4.2.2.12: Perfil de la función de verosimilitud con respecto a σ_a^2 .

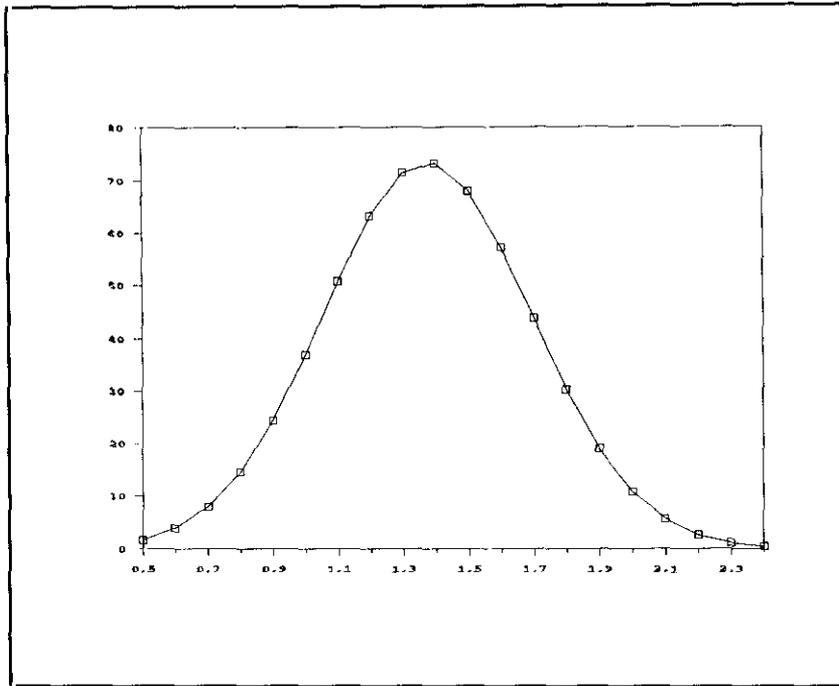


Fig. 4.2.2.13: Perfil de la función de verosimilitud con respecto a ω_0 .

4.3. Modelos del Desempleo Femenino del Reino Unido.

Esta serie (a partir de ahora denotada como DF_t) recoge el número de mujeres desempleadas el primer día de cada mes en el Reino Unido. La muestra analizada es de 67 datos mensuales que abarcan el período de Enero de 1967 a Julio de 1972.

En la Figura 4.3.1 se muestra el gráfico de la serie original, observando la no estacionariedad de la misma con respecto a la media. En cuanto a la estacionariedad en varianza, existe una duda razonable acerca de la conveniencia de su transformación logarítmica (vid. Figura 4.3.3). Aunque el gráfico rango-media (vid. Figura 4.3.2) mostraba una configuración creciente, la información que proporciona este instrumento no es demasiado concluyente ante una muestra tan corta. Por ello, se optó por estimar después el valor óptimo del parámetro de la transformación Box-Cox. No se pudo rechazar la hipótesis nula $H_0: \lambda = 0$ al ser $\hat{\lambda} = -0.0695$ con una desviación típica de 0.3827.

La Figura 4.3.4 muestra claramente la estacionalidad de la serie al estar, por ejemplo, el Desempleo Femenino sistemáticamente por debajo de la media en todos los meses de Junio de la muestra utilizada. Si se divide la muestra en intervalos de 12 meses desde el mes de Junio de cada año, se observa que la pauta de comportamiento de la serie es muy similar en cada intervalo de un año.

La necesidad de tomar dos diferencias regulares se aprecia en la Figura 4.3.5 donde la serie deambula alrededor de la media sobre todo, en los primeros y últimos meses de la muestra escogida. Por último, la Figura 4.3.6 muestra que la serie puede ser considerada estacionaria en media con ese orden de diferenciación.

A partir de las funciones de autocorrelación simple y parcial de la serie $\nabla_{12}\nabla^2\ln DF_t$, el modelo identificado fue un proceso MA(1) regular, no encontrándose ninguna estructura

estacional modelizable. Se estimó este modelo y la acf y pacf residuales mostraron la necesidad de incorporar estructura MA(1) estacional. El modelo que propone Anderson difiere de éste en que no transforma logarítmicamente la serie, sólo considera necesaria una diferencia regular para lograr estacionariedad en media, no incorpora ninguna estructura estacional y modeliza en la parte regular un proceso AR(1). En resumen, se ha seleccionado la siguiente especificación:

$$\nabla^2 \nabla_{12} DF_t = (1 - \theta B)(1 - \theta B^{12}) a_t \quad [4.3.1]$$

Este modelo expresado en forma de espacio de los estados, tiene un vector de estado de dimensión mucho mayor a la del vector de observación. En concreto, $n=13$ mientras que $m=1$. Por ello, éste es un ejemplo adecuado para mostrar la importante reducción de carga computacional que supone estimar este tipo de especificaciones con el filtro de Chandrasekhar.

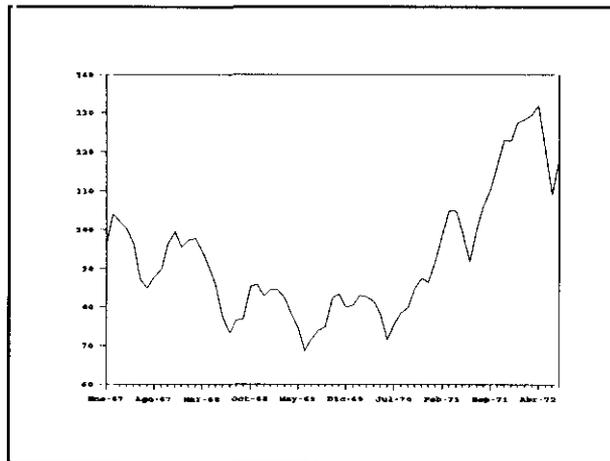


Fig. 4.3.1: Serie original de desempleo femenino.

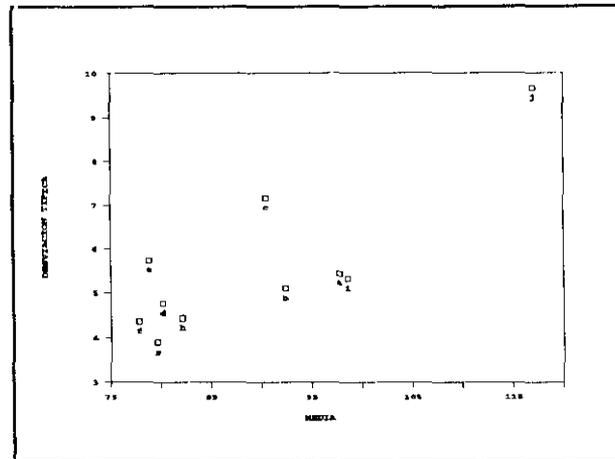


Fig. 4.3.2: Gráfico rango media-desviación típica de la serie DF_t .

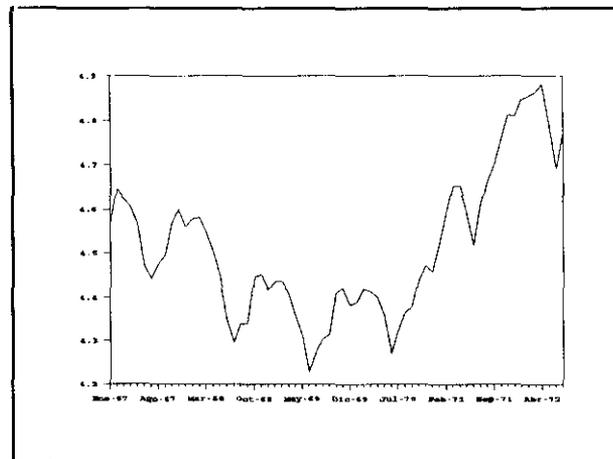


Fig. 4.3.3: Serie $\ln DF_t$.

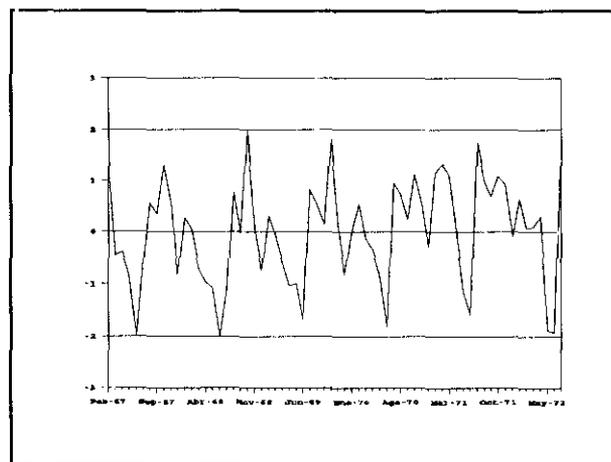


Fig. 4.3.4: Serie $\nabla \ln DF_t$ estandarizada.

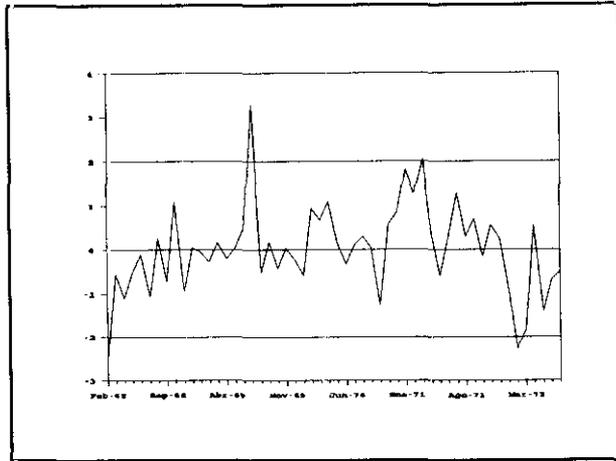


Fig. 4.3.5: Serie $\nabla_{12} \nabla \ln DF_t$ estandarizada.

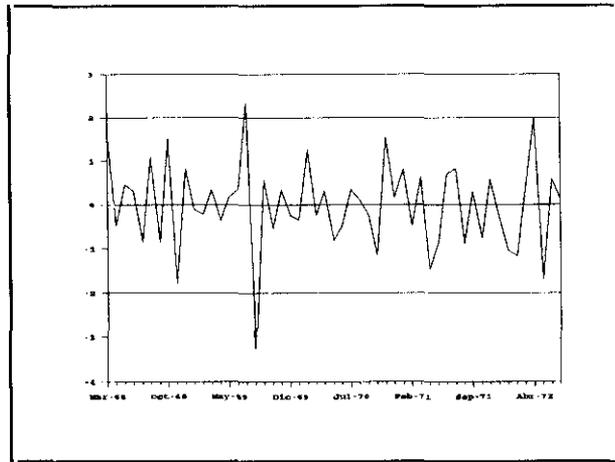


Fig. 4.3.6: Serie $\nabla_{12} \nabla^2 \ln DF_t$ estandarizada.

En la siguiente tabla, se ofrece el resultado de la estimación de [4.3.1] por los tres criterios usados:

Método de Estim.	Mín. cuadrados con retrovisión	Máx. verosimil. con el F. Kalman	Máx. verosimil. con el F. Chandrasekhar
Modelo [4.3.1]	$\hat{\theta} = 0.778$ (0.080)	$\hat{\theta} = 0.764$ (0.083)	$\hat{\theta} = 0.747$ (0.082)
	$\hat{\theta} = 0.286$ (0.137)	$\hat{\theta} = 0.202$ (0.155)	$\hat{\theta} = 0.185$ (0.156)
	$\hat{\sigma}^2 = 0.803 \times 10^{-3}$ a	$\hat{\sigma}^2 = 0.860 \times 10^{-3}$ a (0.193 $\times 10^{-3}$)	$\hat{\sigma}^2 = 0.872 \times 10^{-3}$ a (0.195 $\times 10^{-3}$)

Tabla 4.3.7: Resultados de la estimación del modelo [4.3.1].

A la vista de los resultados de la Tabla 4.3.7, cabe resaltar los siguientes comentarios:

- De nuevo, las estimaciones del parámetro media móvil, tanto regular como estacional, son distintas con las tres aproximaciones. En este ejemplo, las estimaciones obtenidas por máxima verosimilitud con el filtro de Chandrasekhar son ligeramente más bajas que con los otros dos procedimientos, sobre todo, para la estructura media móvil estacional.
- La estimación por mínimos cuadrados con retrovisión admite la estructura estacional, mientras que los otros dos mecanismos de máxima verosimilitud rechazan su inclusión, sobre todo, la aproximación de Chandrasekhar. Esto ocurre aunque la correlación existente entre los dos parámetros media móviles es bastante baja (según el método de mínimos cuadrados es -0.12, según el método de máxima verosimilitud con el filtro de Kalman -0.07 y según la aproximación de Chandrasekhar -0.05).

En cuanto al tiempo de cálculo, éste es un modelo donde su estimación por la aproximación de Chandrasekhar supone una importante ventaja computacional con respecto a la aproximación de Kalman. Esto es así, puesto que el vector de observación es un escalar y el vector de estado es de orden (13x1). En este caso, la ecuación de Riccati asociada al filtro de Kalman y que propaga la matriz P_t , ha de actualizar 91 elementos distintos de la misma en cada instante de tiempo, mientras que en la aproximación de Chandrasekhar sólo hay que actualizar 13 componentes de la matriz factor Y_t y el único elemento existente del factor M_t . La siguiente tabla muestra la citada ganancia en tiempo de estimación:



CPU	Máx.verosim. con el F. Kalman	Máx.verosim. con el F.Chandrasek.	Ganancia (%)
Proc.INTEL 286 (10Mhz)	8573.54	4313.52	+ 98.76
Proc.INTEL 386 (20Mhz)	2074.98	1110.20	+ 86.90

Tabla 4.3.8: Tiempos de cálculo en segundos en la estimación del modelo [4.3.1].

Con objeto de estudiar si en este modelo el filtro de Kalman calcula matrices de covarianzas del vector de estado semidefinidas positivas en todo instante de tiempo, se ha representado en la Figura 4.3.9 la evolución de la traza de esta matriz cuando es propagada por los dos filtros, dado que la dimensión de la misma es de orden (13x13). Se observa que la traza es estrictamente positiva con las dos aproximaciones.

Con respecto a la estimación del vector de estado obtenida por ambos filtros, sólo se han representado la primera y las dos últimas componentes de dicho vector con respecto al tiempo. El motivo es que al expresar este modelo en forma de espacio de los estados, se obtiene que sólo estos tres elementos del vector de estado dependen de los parámetros del modelo econométrico. Para el resto de sus componentes, se comprueba que el elemento i -ésimo del vector de estado en un instante cualquiera, es siempre igual a la componente $i+1$ del vector de estado en el instante anterior. Las Figuras 4.3.10, 4.3.11 y 4.3.12 ofrecen la evolución de las componentes estimadas del vector de estado que interesan. Estos gráficos muestran que la propagación de esas tres componentes del vector de estado es muy similar cuando dicho vector se estima con el filtro de Kalman y el de Chandrasekhar. Solamente se aprecian ligeras divergencias en los primeros instantes de tiempo, debido a que el filtro de Chandrasekhar es inicializado de forma distinta al filtro de Kalman.

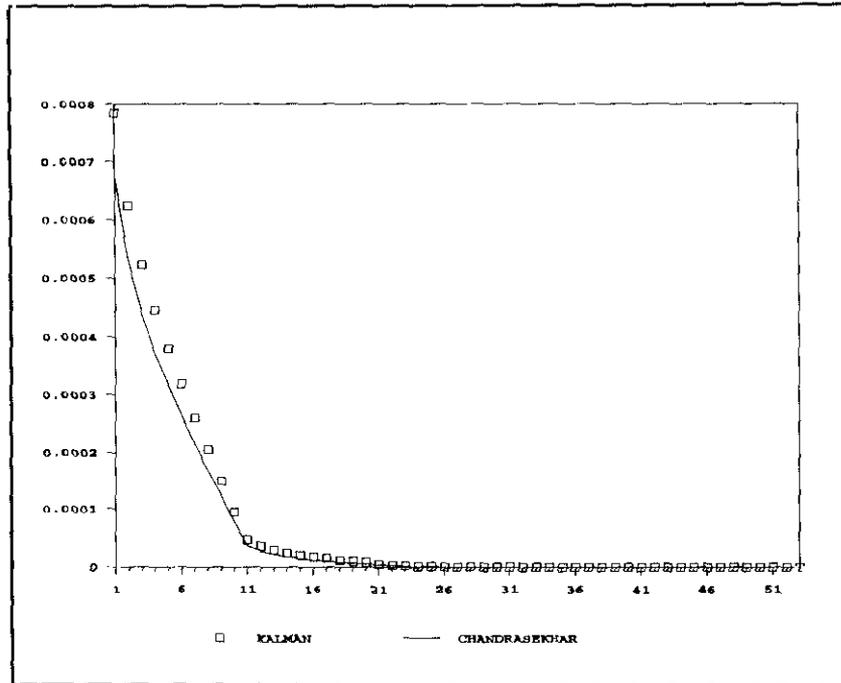


Fig. 4.3.9: Trazo de la matriz de covarianzas estimada del vector de estado del modelo [4.3.1].

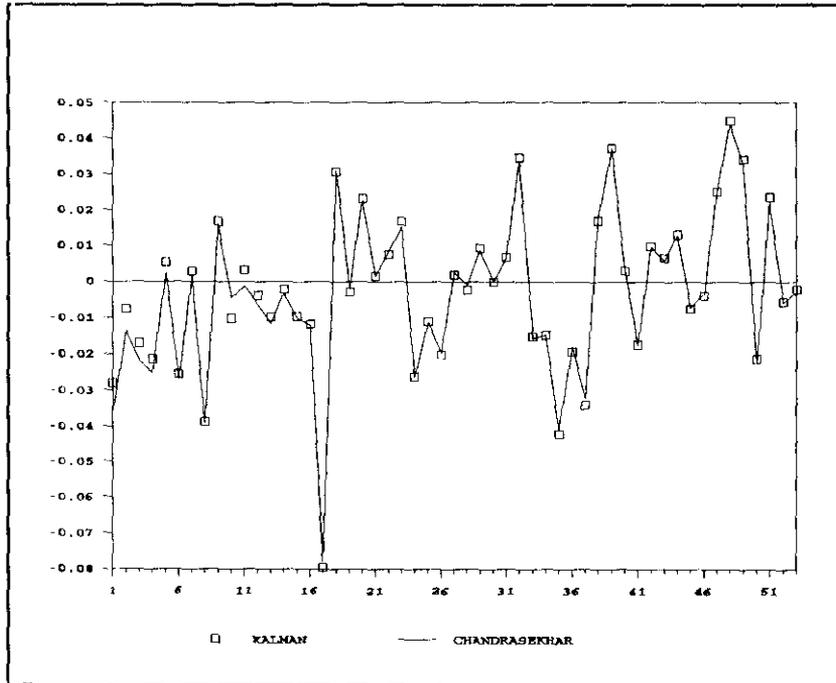


Fig. 4.3.10: Primera componente estimada del vector de estado del modelo [4.3.1].

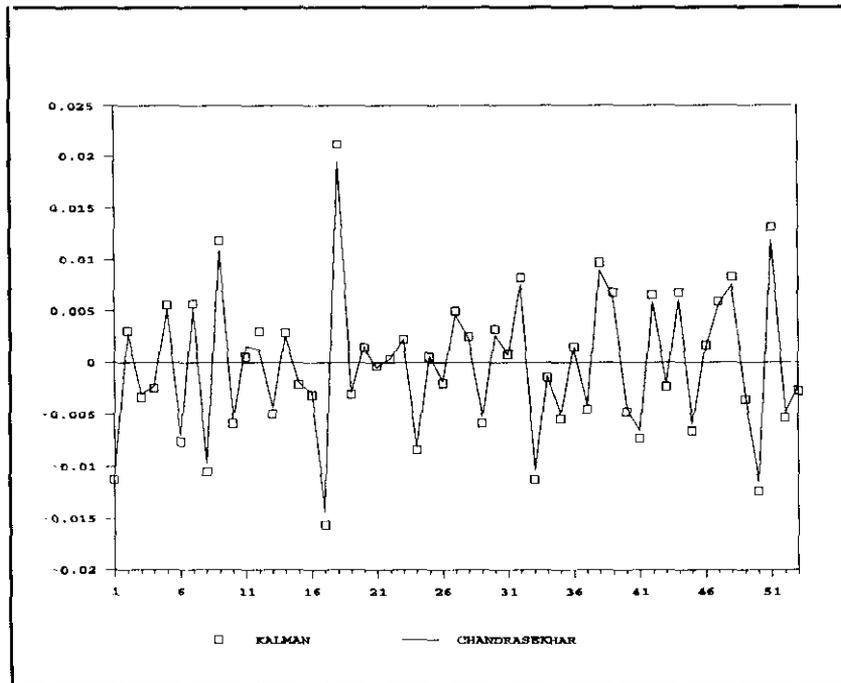


Fig. 4.3.11: Penúltima componente estimada del vector de estado del modelo [4.3.1].

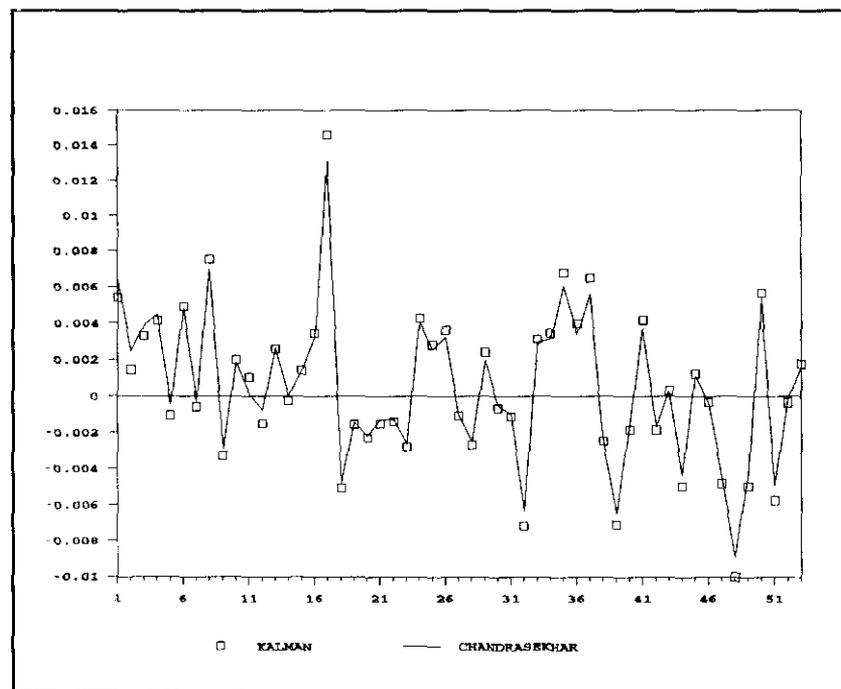


Fig. 4.3.12: Última componente estimada del vector de estado del modelo [4.3.1].

4.4. Modelos de relación entre el Consumo Energético y el PIB de Dinamarca.

En este apartado, se estudia la relación existente entre el Consumo de Energía (CE) y el PIB de Dinamarca a través de una función de transferencia. Esta relación fue analizada por Ramain (1986) para 18 países de la OCDE, estimando un modelo de regresión simple para cada país. La diagnosis de estos modelos estimados muestra claramente que un análisis de regresión no recoge adecuadamente la relación entre CE y PIB (residuos no estacionarios, etc.)

El análisis se ha realizado utilizando una muestra anual de 30 observaciones que abarca el período 1951-1980 y sólo se ha analizado el caso de Dinamarca. Los datos de PIB vienen dados en números índice con base en 1970 y el Consumo Energético está agregado y medido en millones de Tm. equivalentes de petróleo.

Después de haber conseguido una representación univariante adecuada tanto para el CE_t como para el PIB, la identificación del modelo de función de transferencia dio lugar a la siguiente especificación:

$$\begin{aligned} \ln CE_t &= (\omega_0 - \omega_1 B) \ln PIB_t + N_t \\ (1 - \phi_1 B - \phi_2 B^2) \nabla^2 N_t &= a_t \end{aligned} \quad [4.4.1]$$

siendo la ganancia en estado estacionario $g = \omega_0 - \omega_1$. La forma en que ha de escribirse el modelo [4.4.1] para ser expresado en forma de espacio de los estados es:

$$(1 - \phi_1 B - \phi_2 B^2) \nabla^2 \ln CE_t = (1 - \phi_1 B - \phi_2 B^2) (\omega_0 - \omega_1 B) \nabla^2 \ln PIB_t + a_t$$

donde a_t es un proceso de ruido blanco.

El resultado de la estimación de este modelo por los tres procedimientos utilizados se ofrece en la siguiente Tabla:

Min. cuadrados con retrovisión	Máx. verosimil. con el F. Kalman	Máx. verosimil. con el F. Chandrasekhar
$\hat{\omega}_0 = 0.950$ (0.520)	$\hat{\omega}_0 = 0.930$ (0.483)	$\hat{\omega}_0 = 0.920$ (0.483)
$\hat{\omega}_1 = -0.950$ (0.520)	$\hat{\omega}_1 = -0.922$ (0.488)	$\hat{\omega}_1 = -0.912$ (0.488)
$\hat{\phi}_1 = -0.820$ (0.190)	$\hat{\phi}_1 = -0.796$ (0.174)	$\hat{\phi}_1 = -0.813$ (0.172)
$\hat{\phi}_2 = -0.450$ (0.180)	$\hat{\phi}_2 = -0.419$ (0.178)	$\hat{\phi}_2 = -0.435$ (0.176)
$\hat{\sigma}_a^2 = 0.50 \times 10^{-2}$	$\hat{\sigma}_a^2 = 0.50 \times 10^{-2}$ (0.13 $\times 10^{-2}$)	$\hat{\sigma}_a^2 = 0.50 \times 10^{-2}$ (0.13 $\times 10^{-2}$)

Tabla 4.4.1: Resultados de la estimación del modelo [4.4.1].

La matriz de correlaciones entre los parámetros estimados según mínimos cuadrados con retrovisión es la siguiente:

$$\begin{matrix}
 & \hat{\phi}_1 & \hat{\phi}_2 & \hat{\omega}_0 & \hat{\omega}_1 \\
 \begin{bmatrix}
 1.00 & & & & \\
 .53 & 1.00 & & & \\
 -.31 & -.07 & 1.00 & & \\
 -.10 & .21 & .07 & 1.00 &
 \end{bmatrix}
 \end{matrix}$$

Según máxima verosimilitud con el filtro de Kalman es:

$$\begin{matrix}
 & \hat{\phi}_1 & \hat{\phi}_2 & \hat{\omega}_0 & \hat{\omega}_1 & \hat{\sigma}_a^2 \\
 \begin{bmatrix}
 1.00 & & & & & \\
 .55 & 1.00 & & & & \\
 .00 & .00 & 1.00 & & & \\
 .00 & .00 & .00 & 1.00 & & \\
 .10 & .08 & .00 & .00 & 1.00 &
 \end{bmatrix}
 \end{matrix}$$

Y según el filtro de Chandrasekhar:

$$\begin{matrix}
 & \hat{\phi}_1 & \hat{\phi}_2 & \hat{\omega}_0 & \hat{\omega}_1 & \hat{\sigma}_a^2 \\
 \begin{bmatrix}
 1.00 & & & & & \\
 .56 & 1.00 & & & & \\
 .00 & .00 & 1.00 & & & \\
 .00 & .00 & .01 & 1.00 & & \\
 .11 & .08 & .00 & .00 & 1.00 &
 \end{bmatrix}
 \end{matrix}$$

La única diferencia apreciable en la estimación por los tres métodos consiste en que se obtiene una correlación más alta entre los parámetros autorregresivos cuando se estiman por máxima verosimilitud que por mínimos cuadrados con retrovisión. Además, no existe ninguna correlación entre los parámetros estimados de la función de transferencia y los autorregresivos por cualquiera de los métodos de máxima verosimilitud, mientras que sí se detectaban cuando se utilizaba mínimos cuadrados con retrovisión. Este resultado se debe al empleo de una aproximación al gradiente de la función de verosimilitud que, a su vez, se utiliza para calcular una aproximación al hessiano y por último, esta aproximación se usa para calcular una aproximación a la matriz de información.

El tiempo de cálculo utilizado en la estimación de esta función de transferencia fue de 135.83 segundos utilizando el filtro de Kalman y de 129.68 segundos según la aproximación de Chandrasekhar. La pequeña ganancia porcentual en tiempo de cálculo (4.74%), es debida a que en esta especificación el ratio de dimensiones n/m es 3/1. Es decir, según el filtro de Kalman se han de actualizar 6 elementos distintos de la matriz P_t cada vez que se procesa una observación, mientras que con el filtro de Chandrasekhar sólo hay que actualizar 3 elementos del vector factor Y_t y el único componente del factor M_t .

Los gráficos 4.4.2 y 4.4.3 muestran la evolución de los elementos de la diagonal de la matriz P_t , es decir, las varianzas estimadas de cada componente del vector de estado cuando se ha llegado al vector de parámetros óptimo. Del examen de los mismos, se pueden sacar las siguientes conclusiones:

- (a) El deterioro numérico del filtro de Kalman se refleja en la aparición de elementos negativos en la diagonal principal de la matriz de covarianzas P_t (vid. Figura 4.4.2).

- (b) La factorización de Chandrasekhar asegura el carácter estrictamente positivo de dichas varianzas en todo instante de tiempo (vid. Figura 4.4.3).
- (c) Las varianzas de los dos primeros componentes del vector de estado calculadas con el filtro de Kalman son intermitentemente negativas, en la línea de los resultados obtenidos por Bierman y Thornton (1980).
- (d) Los errores numéricos en el cálculo de la matriz P_t hacen que las ganancias calculadas con el filtro de Kalman no sean fiables.
- (e) Como el modelo estimado es estacionario, tanto la matriz de covarianzas P_t como la matriz ganancia K_t , deberían converger a un valor estacionario en el vector de parámetros óptimo. Esto sólo se observa con la aproximación de Chandrasekhar, pues con el filtro de Kalman dichas variables oscilan constantemente entre dos valores.
- (f) El empleo de la versión estable de la ecuación de propagación de P_t (como se ha hecho en este caso), no garantiza la estabilidad numérica del filtro de Kalman. La utilización de esta versión hace que los errores numéricos sean menores, pero no se eliminan.
- (g) La inestabilidad numérica no desaparece aún trabajando con números reales en precisión doble.
- (h) La idea de acotar los elementos de la diagonal de P_t para asegurar su condicionamiento (vid. Capítulo 1), es un método ad-hoc que, en teoría, producirá resultados subóptimos. Además, la forma adecuada de acotar dichos elementos de la matriz dependerá del problema en concreto y por supuesto, esta técnica aumentaría inevitablemente el tiempo de cálculo.

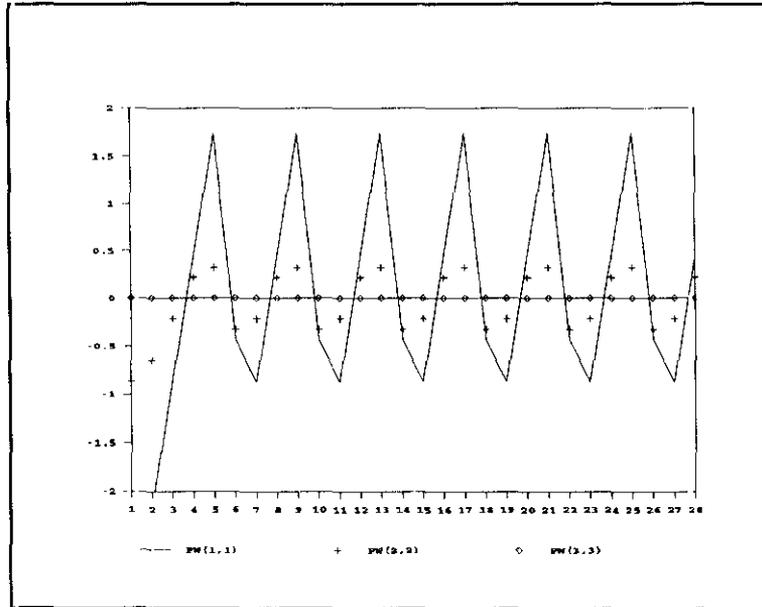


Fig. 4.4.2: Varianzas estimadas de los componentes del vector de estado calculadas con el filtro de Kalman. [PW(i,i) es la varianza del componente i-ésimo].

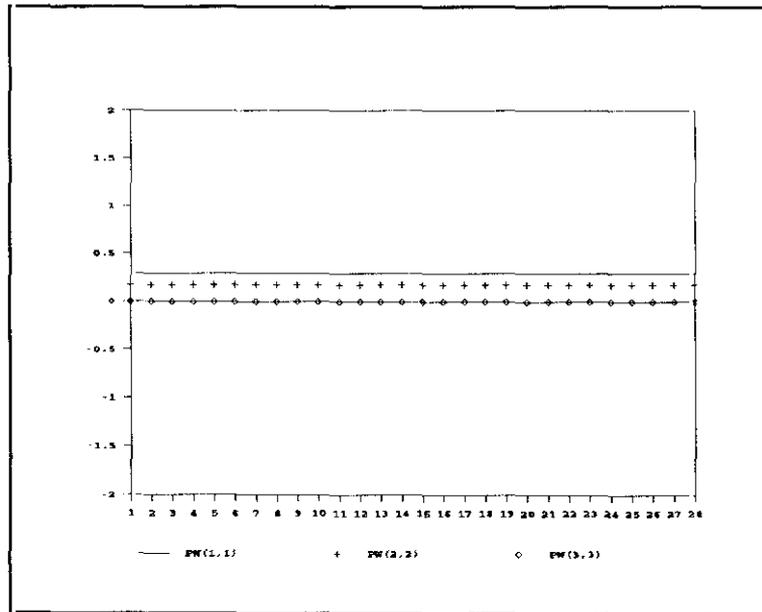


Fig. 4.4.3: Varianzas estimadas de los componentes del vector de estado calculadas con el filtro de Chandrasekhar. [PW(i,i) es la varianza del componente i-ésimo].

En algunas ocasiones, puede ser más interesante la estimación de la ganancia en estado estacionario de una función de transferencia que la de los parámetros individuales de la misma. Incluso, en determinadas situaciones, puede interesar imponer la restricción de ganancia unitaria en estado estacionario.

Es fácil imponer restricciones lineales o no lineales entre los parámetros de un modelo cuando éste es formulado en espacio de los estados y estimado por cualquier criterio de máxima verosimilitud. La contrastación de estas hipótesis puede llevarse a cabo mediante el test de Wald, el test de razón de verosimilitudes o el test del multiplicador de Lagrange (ML). La ventaja de utilizar este último es que sólo hay que estimar el modelo bajo la hipótesis nula. La forma de este test es la siguiente:

$$ML = \left[\frac{\partial \ell(\hat{\theta}_r)}{\partial \hat{\theta}_r} \right]^T [M(\hat{\theta}_r)]^{-1} \left[\frac{\partial \ell(\hat{\theta}_r)}{\partial \hat{\theta}_r} \right] \sim \chi_r^2$$

donde r es el número de restricciones impuestas, $\hat{\theta}_r$ es el vector de estimaciones de los parámetros en el modelo restringido y $M(\hat{\theta}_r)$ es la matriz de información calculada bajo la hipótesis nula. Debe tenerse en cuenta que, tal como señala Terceiro (1990), para la utilización de este test, sólo se necesitan variables calculadas en el proceso de estimación y que, en nuestro caso, se dispone de las expresiones exactas del gradiente y de la matriz de información.

El modelo que nos ocupa se ha estimado bajo la hipótesis nula $H_0: g = 1$, donde el vector gradiente y la matriz de información calculados mediante la aproximación de Kalman son:

$$\left[\frac{\partial \ell(\hat{\theta}_r)}{\partial \hat{\theta}_r} \right]^T = [-.133 \times 10^{-11} \quad -.113 \times 10^{-11} \quad .290 \times 10^{-11} \quad -.156 \times 10^{-8}]$$

$$M(\hat{\theta}_r)^{-1} = \begin{bmatrix} 3.367 \times 10^{-2} & & & & \\ 1.646 \times 10^{-2} & 3.480 \times 10^{-2} & & & \\ -5.494 \times 10^{-22} & 9.448 \times 10^{-21} & 1.300 \times 10^{-1} & & \\ 1.676 \times 10^{-5} & 1.340 \times 10^{-5} & 1.617 \times 10^{-24} & 2.592 \times 10^{-6} & \end{bmatrix}$$

y el valor obtenido del test es $ML = 7.671 \times 10^{-24}$.

Y según la aproximación de Chandrasekhar:

$$\left[\frac{\partial \ell(\hat{\theta}_r)}{\partial \hat{\theta}_r} \right]^T = [.804 \times 10^{-11} \quad -.417 \times 10^{-11} \quad -.645 \times 10^{-12} \quad -.186 \times 10^{-8}]$$

$$M(\hat{\theta}_r)^{-1} = \begin{bmatrix} 3.344 \times 10^{-2} & & & & \\ 1.673 \times 10^{-2} & 3.460 \times 10^{-2} & & & \\ 2.714 \times 10^{-21} & 3.444 \times 10^{-21} & 1.316 \times 10^{-1} & & \\ 2.582 \times 10^{-6} & 1.395 \times 10^{-5} & 1.840 \times 10^{-24} & 2.582 \times 10^{-6} & \end{bmatrix}$$

siendo, en este caso, el valor calculado del estadístico de 1.077×10^{-23} .

Por tanto, a la vista de los resultados, no puede rechazarse la hipótesis nula de ganancia unitaria con los dos criterios de estimación.

4.5. Modelos multivariantes.

En Grandal (1986) se desarrolla un marco teórico general para el estudio de mercados con retardo fijo de oferta. También se lleva a cabo una aplicación econométrica de este modelo al mercado de la patata en España.

Este modelo generaliza en muchos aspectos los formulados hasta entonces para este tipo de mercados. Además, permite contrastar no sólo la conocida hipótesis de expectativas racionales de Muth (1961), sino también una amplia gama de hipótesis alternativas de racionalidad.

El modelo conceptual básico, aplicable a cualquier modelo con una función de oferta con retardo fijo, está formado por las siguientes relaciones:

(a) Relación de demanda.

$$c_t = \nu_c(B)p_t + \eta_{ct} \quad [4.5.1]$$

$$\nu_c(0) \equiv \alpha \neq 0 \quad [4.5.2]$$

La ecuación [4.5.1] indica que la cantidad demandada (c_t) puede responder de forma dinámica ante variaciones en el precio (p_t). $\nu_c(B)$ es una función de transferencia que recoge ese posible efecto dinámico del precio sobre cantidad. El término η_{ct} es el error estocástico en la demanda, que cumple las condiciones de estacionariedad e invertibilidad de un modelo univariante estocástico. Por último, α (vid. expresión [4.5.2]) es la elasticidad-precio de la demanda al más corto plazo. La condición $\alpha \neq 0$ es necesaria para que el precio se determine en el mercado.

(b) Relación de oferta.

$$q_t = \beta p_{t-1}^e(1) + \eta_{qt} \quad [4.5.3]$$

La ecuación [4.5.3] relaciona la cantidad ofrecida al mercado en el instante t , (q_t) , con la expectativa del precio que forman los productores en $t-1$ para el momento t , $[p_{t-1}^e(1)]$. β es la elasticidad-precio de la oferta que se supone distinta de cero (vid. [4.5.4]), ya que $\beta = 0$ implicaría que la única relación entre precio y cantidad vendría dada por el lado de la demanda. Por último, η_{qt} es el error estocástico en la oferta que satisface las condiciones de estacionariedad e invertibilidad de un modelo univariante estocástico.

(c) Condición de equilibrio del mercado.

$$c_t = q_t \quad \forall t \quad [4.5.5]$$

Es decir, [4.5.5] implica que no se acumulan stocks del bien. Esta es una hipótesis simplificatoria que resulta verosímil en el caso estudiado.

(d) Estructura de los términos de error e independencia entre las relaciones de oferta y demanda.

$$\Pi_c(B)\eta_{ct} = a_{ct} \quad [4.5.6]$$

$$\Pi_q(B)\eta_{qt} = a_{qt} \quad [4.5.7]$$

$$E(a_{ct}a_{qt}') = 0 \quad [4.5.8]$$

donde $\Pi(B) = \theta(B)^{-1}\phi(B)$, $\theta(B)$ es un polinomio media-móvil invertible y $\phi(B)$ es un polinomio autorregresivo estacionario. La condición [4.5.8] es necesaria para que el concepto de equilibrio en un mercado sea operativo. Esto implica que las variables recogidas en los términos de error de oferta y demanda no deben estar correlacionadas entre sí.

(e) Formación de expectativas.

Para completar el modelo, se supone que la expectativa $p_{t-1}^e(1)$ se relaciona con las variables del mercado de la forma:

$$p_{t-1}^e(1) = \nu_p(B)p_{t-1} + \nu_q(B)q_{t-1} \quad [4.5.8]$$

donde $\nu_p(B)$ y $\nu_q(B)$ son funciones de transferencia no explosivas. Es decir, el polinomio autorregresivo del denominador en $\nu(B)$, ha de tener sus raíces características fuera del círculo unidad. La relación [4.5.8] supone que la única información disponible para los agentes en el instante t , viene dada por la historia pasada de precios y cantidades hasta el instante $t-1$.

A partir de [4.5.8], es posible contrastar la hipótesis de racionalidad en el sentido de Muth (1961). Esta hipótesis supone que los agentes utilizan el conjunto de información máximo que, en este caso, incluye tanto la historia de precios y de cantidades, como la estructura y los valores de los parámetros del modelo dado por [4.5.1]-[4.5.8]. Es decir, $p_{t-1}^e(1)$ sería la esperanza condicional de p_t por el modelo y la historia de las variables. Otra hipótesis contrastable dentro de este marco, es la de racionalidad en sentido univariante, que supone que los agentes sólo utilizan la historia del precio para formar sus expectativas. En este caso, $p_{t-1}^e(1)$ sería la esperanza condicional de p_t por esta información y el modelo univariante del precio.

A partir de las ecuaciones [4.5.1]-[4.5.8], se obtiene el siguiente modelo bivariante estocástico:

$$\begin{bmatrix} \Pi_c(B) & -\nu_c(B)\Pi_c(B) \\ \Pi_q(B)[1 - \beta\nu_q(B)B] & -\beta\nu_p(B)B\Pi_q(B) \end{bmatrix} \begin{bmatrix} q_t \\ p_t \end{bmatrix} = \begin{bmatrix} a_{ct} \\ a_{qt} \end{bmatrix} \quad [4.5.9]$$

el cual, una vez normalizado queda:

$$\begin{bmatrix} \Pi_q(B)[1 - \beta\nu_q(B)B] & -\beta\nu_p(B)B\Pi_q(B) \\ \frac{\Pi_q(B)[1 - \beta\nu_q(B)B] - \Pi_c(B)}{\alpha} & \frac{\nu_c(B)\Pi_c(B) - \beta\nu_p(B)B\Pi_q(B)}{\alpha} \end{bmatrix} \begin{bmatrix} q_t \\ p_t \end{bmatrix} = \begin{bmatrix} a_{qt} \\ a_{pt} \end{bmatrix} \quad [4.5.10]$$

$$\text{donde } a_{pt} = \frac{a_{qt} - a_{ct}}{\alpha}$$

A partir de una estimación de [4.5.10], es posible estimar la elasticidad-precio de la demanda mediante la relación:

$$\hat{\alpha} = \frac{\hat{\sigma}_q^2}{\hat{\sigma}_{qp}} \quad [4.5.11]$$

y, una vez estimada esta elasticidad, ya es posible escribir el modelo en la forma [4.5.9]. La elasticidad-precio de la oferta y la función de transferencia que recoge el efecto dinámico del precio sobre cantidad, pueden estimarse a partir de [4.5.9]. Así, denotando Π_{ij} al elemento (i,j) de la matriz autorregresiva de [4.5.9], se tiene que:

$$\hat{\beta} = \frac{\hat{\Pi}_{22}(B)}{\hat{\Pi}_{21}(B)} \Big|_{B=1} \quad \hat{\nu}_c(B) = \frac{-\hat{\Pi}_{12}(B)}{\hat{\Pi}_{11}(B)} \quad [4.5.12]$$

Grandal aplica este modelo teórico al mercado de la patata en España. Para ello, utiliza una muestra trimestral de 64 observaciones que abarcan el período del primer trimestre de 1965 al último de 1980. Las series temporales utilizadas son la Cantidad de Patata Producida trimestralmente (en miles de Tm.) y la serie de Precios de la Patata Percibidos por los Agricultores Ponderados (en Pts/Kg.). En el Apéndice D se ofrecen los datos de ambas series reelaborados a partir de los datos estadísticos de base.

Como ejemplo multivariante para aplicar nuestro mecanismo de estimación basado en el filtro de Chandrasekhar, se ha escogido uno de los modelos que se presentan en el trabajo de Grandal por diversas razones:

- El modelo conceptual básico en el que se apoya es muy general. En principio, es aplicable a cualquier modelo con una función de oferta con retardo fijo, por lo que puede ser interesante su futura utilización para el estudio de otros mercados distintos al de la patata en España. Aunque en este trabajo no se ha abordado este aspecto, es evidente que puede ser útil disponer de mecanismos de estimación robustos y flexibles para una posible nueva aplicación del modelo teórico.
- El modelo que se ha seleccionado fue estimado por máxima verosimilitud utilizando el programa SCA (1986). Este paquete sólo calcula la expresión analítica exacta de la función de verosimilitud para modelos de medias-móviles puros. En el caso de modelos mixtos, como el del ejemplo, se utiliza la aproximación dada por Hillmer y Tiao (1979).
- Algunos parámetros del modelo teórico han de ser estimados a posteriori, es decir, usando las estimaciones resultantes de otros coeficientes. Además, sus desviaciones típicas son también recuperadas mediante

aproximaciones. Por otro lado, el cálculo de la matriz de información exacta es particularmente importante en este ejemplo, donde el contraste de hipótesis permite estudiar el tipo de mecanismo de formación de expectativas existente en el mercado.

Los resultados de la estimación del modelo bivalente por máxima verosimilitud exacta con SCA son los siguientes:

$$c_t = \frac{\hat{\alpha} (1 + 0.14B^4 + 0.05B^8 - 0.17B^{12})}{1 - 0.69B^4} \nabla \ln P_t + \frac{1 - 0.44B^4}{\nabla_4} a_{ct} \quad [4.5.13]$$

$$\hat{\alpha} = \begin{matrix} -4.4 \\ (4.6) \end{matrix}$$

$$\hat{\sigma}_c \times 100 = 70.6$$

$$q_t = \hat{\beta} p_{t-1}^e + \frac{1 - 0.44B^4}{\nabla_4} a_{qt} \quad [4.5.14]$$

$$\hat{\beta} = \begin{matrix} 1.16 \\ (0.57) \end{matrix}$$

$$\hat{\sigma}_p \times 100 = 9.3$$

$$p_{t-1}^e(1) = \left[\frac{-0.21B^4 + 0.19B^8 + 0.30B^{12}}{1 - 0.69B^4} \right] \nabla \ln p_t \quad [4.5.15]$$

Las relaciones [4.5.13]-[4.5.15], junto con la condición de equilibrio del mercado, pueden escribirse de nuevo en forma de una estructura bivalente:

$$\begin{bmatrix} (1 - 0.69B^4) & -\hat{\alpha} (1 + 0.14B^4 + 0.05B^8 - 0.17B^{12}) \\ (1 - 0.69B^4) & -\hat{\beta} (-0.21B^4 + 0.19B^8 + 0.30B^{12}) \end{bmatrix} \begin{bmatrix} \nabla_4 \ln q_t \\ \nabla \nabla_4 \ln p_t \end{bmatrix} =$$

$$\begin{bmatrix} (1 - 0.44B^4)(1 - 0.69B^4) & 0 \\ 0 & (1 - 0.44B^4)(1 - 0.69B^4) \end{bmatrix} \begin{bmatrix} a_{ct} \\ a_{qt} \end{bmatrix} \quad [4.5.16]$$

y normalizando [4.5.16], las matrices correspondientes a su representación en espacio de los estados son las siguientes:

$$x_{t+1} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ \bar{\pi}_1 & 0 & 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \bar{\pi}_2 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 \\ \bar{\pi}_3 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \end{bmatrix} x_t + \begin{bmatrix} 0 \\ 0 \\ 0 \\ E_1 \\ 0 \\ 0 \\ 0 \\ E_2 \\ 0 \\ 0 \\ 0 \\ E_3 \end{bmatrix} w_t$$

(24x24) (24x2)

$$\begin{bmatrix} \nabla_4 c_t \\ \nabla_4 p_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \end{bmatrix} x_t + \begin{bmatrix} 0 & 1 \\ -1/\alpha & 1/\alpha \end{bmatrix} v_t$$

(2x24)

donde:

$$\bar{\pi}_1 = \begin{bmatrix} -\phi_{11} & \beta\phi_{22} \\ 0 & -\phi_{12} + \frac{\beta\phi_{22}}{\alpha} \end{bmatrix}; \quad \bar{\pi}_2 = \begin{bmatrix} 0 & \beta\phi_{23} \\ 0 & -\phi_{13} + \frac{\beta\phi_{23}}{\alpha} \end{bmatrix}; \quad \bar{\pi}_3 = \begin{bmatrix} 0 & \beta\phi_{24} \\ 0 & -\phi_{14} + \frac{\beta\phi_{24}}{\alpha} \end{bmatrix}$$

$$E_1 = \begin{bmatrix} \frac{-\beta\phi_{22}}{\alpha} & \theta_{11} + \frac{\beta\phi_{22}}{\alpha} \\ \frac{-\phi_{11} - \theta_{11} + \phi_{12}}{\alpha} - \frac{\beta\phi_{22}}{\alpha^2} & \frac{\phi_{11} + \theta_{11} - \phi_{12}}{\alpha} + \frac{\beta\phi_{22}}{\alpha^2} \end{bmatrix}$$

$$E_2 = \begin{bmatrix} \frac{-\beta\phi_{23}}{\alpha} & \phi_{11}\theta_{11} - \frac{\beta\phi_{23}}{\alpha} \\ \frac{-\phi_{11}\theta_{11} + \phi_{13} - \beta\phi_{23}}{\alpha} - \frac{\beta\phi_{23}}{\alpha^2} & \frac{\phi_{11}\theta_{11} - \phi_{13}}{\alpha} + \frac{\beta\phi_{23}}{\alpha^2} \end{bmatrix}$$

$$E_3 = \begin{bmatrix} -\beta\phi_{24} & \frac{\beta\phi_{24}}{\alpha} \\ \frac{\phi_{14}}{\alpha} - \frac{\beta\phi_{24}}{\alpha^2} & \frac{-\phi_{14}}{\alpha} + \frac{\beta\phi_{24}}{\alpha^2} \end{bmatrix}; \quad Q = R = S = \begin{bmatrix} \sigma_{ct}^2 & 0 \\ 0 & \sigma_{qt}^2 \end{bmatrix}; \quad w_t = v_t = \begin{bmatrix} a_{ct} \\ a_{qt} \end{bmatrix}$$

Los resultados de la estimación de este modelo por máxima verosimilitud exacta, utilizando el filtro de Chandrasekhar y la matriz de información exacta, son los siguientes:

$$c_t = \frac{\hat{\alpha}(1 + 0.13B^4 + 0.051B^8 - 0.19B^{12})}{1 - 0.55B^4} \nabla \ln P_t + \frac{(0.055)}{\nabla_4} a_{ct} \quad [4.5.16]$$

$$\hat{\alpha} = -2.63 \quad (0.42)$$

$$\hat{\sigma}_c^2 = 0.20 \quad (0.0071)$$

$$q_t = \hat{\beta} p_{t-1}^e + \frac{(0.055)}{\nabla_4} a_{qt} \quad [4.5.17]$$

$$\hat{\beta} = 0.68 \quad (0.01)$$

$$\hat{\sigma}_q^2 = 0.0089 \quad (0.0012)$$

$$p_{t-1}^e = \left[\begin{array}{c} (0.071) \quad (0.072) \quad (0.070) \\ - 0.24B^4 + 0.20B^8 + 0.33B^{12} \\ \hline 1 - 0.55B^4 \\ (0.067) \end{array} \right] \nabla \ln p_t \quad [4.5.18]$$

Los resultados muestran que todas las estimaciones coinciden aproximadamente, excepto las correspondientes a las elasticidades-precio de demanda y oferta. En concreto, la elasticidad-precio de la demanda estimada es negativa y significativamente distinta de cero. Y con respecto a la elasticidad-precio de la oferta, se rechaza la hipótesis nula de que sea igual a la unidad.

A continuación, se ofrecen los perfiles de la función de verosimilitud transformada logarítmicamente y cambiada de signo, para algunos parámetros de esta especificación. Los gráficos 4.5.1 y 4.5.2 muestran que la función $-\ln L(\theta)$ está muy bien condicionada con respecto a los parámetros ϕ_{11} y θ_{11} . Con respecto a las varianzas residuales, el perfil de dicha función es ligeramente más plano en un entorno de puntos que contiene al óptimo (ver Figuras 4.5.3 y 4.5.4). La función $-\ln L(\theta)$ con respecto a la elasticidad-precio de la demanda (α) es la peor condicionada (ver Figura 4.5.5). En concreto, este gráfico muestra que la función presenta un pico en el punto en el que el valor de la función es mínimo. A partir de ese punto, existe un intervalo de valores de α para los cuales la función $-\ln L(\theta)$ toma aproximadamente el mismo valor. Por supuesto, el perfil mostrado sería bastante más suave si se hubiesen dibujado más puntos, pero en cualquier caso, en Grandal la estimación final de dicho parámetro cae dentro de esa zona plana de la función, mientras que con nuestro mecanismo se converge a un punto donde el valor de la función $-\ln L(\theta)$ es menor.

Por último, señalar como conclusión que además de ser importante la utilización de procedimientos de estimación robustos y flexibles, también puede ser útil el uso de instrumentos gráficos muy sencillos como los perfiles de la función de

verosimilitud en la dirección de todos los parámetros. Esto permite detectar a veces un mal condicionamiento de la función a optimizar en algunas direcciones, así como posibles no linealidades de la misma que hagan que el algoritmo de optimización empleado no llegue a un óptimo en sentido estricto.

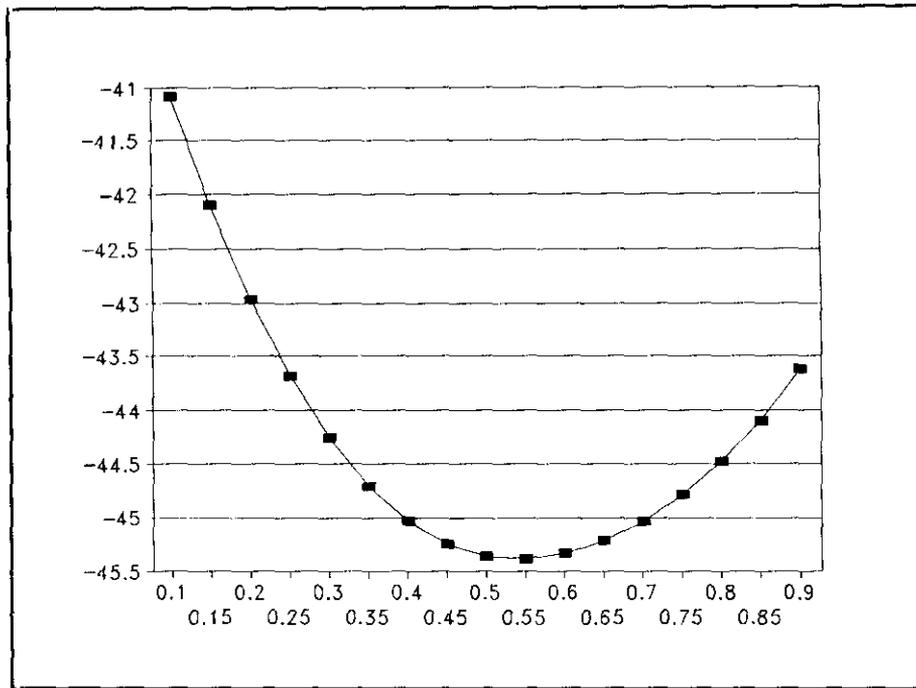


Figura 4.5.1: Perfil de la función $-\ln L(\theta)$ con respecto al parámetro ϕ_{11} .

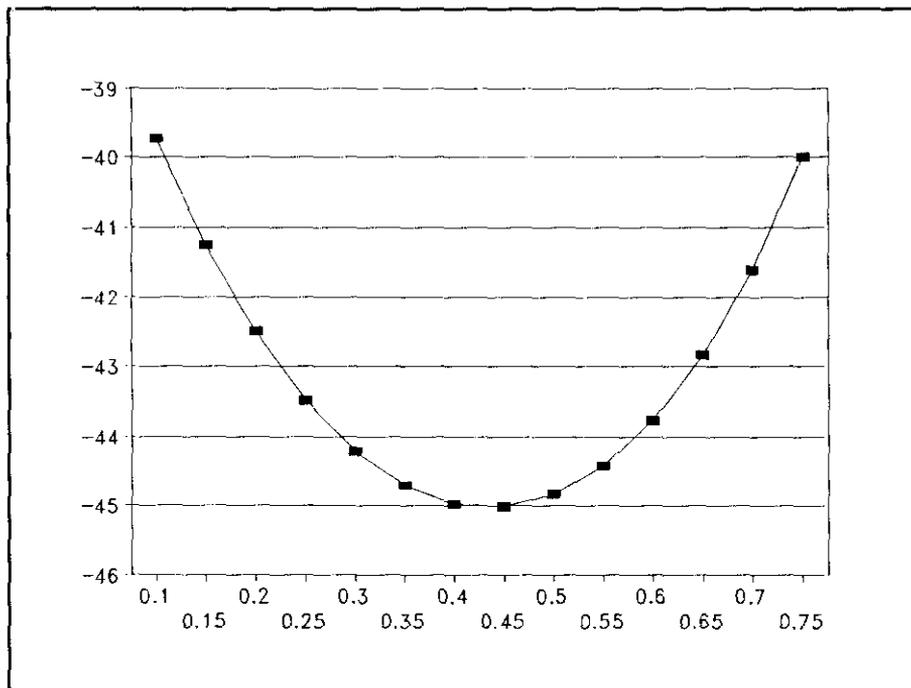


Figura 4.5.2: Perfil de la función $-\ln L(\theta)$ con respecto al parámetro θ_{11} .

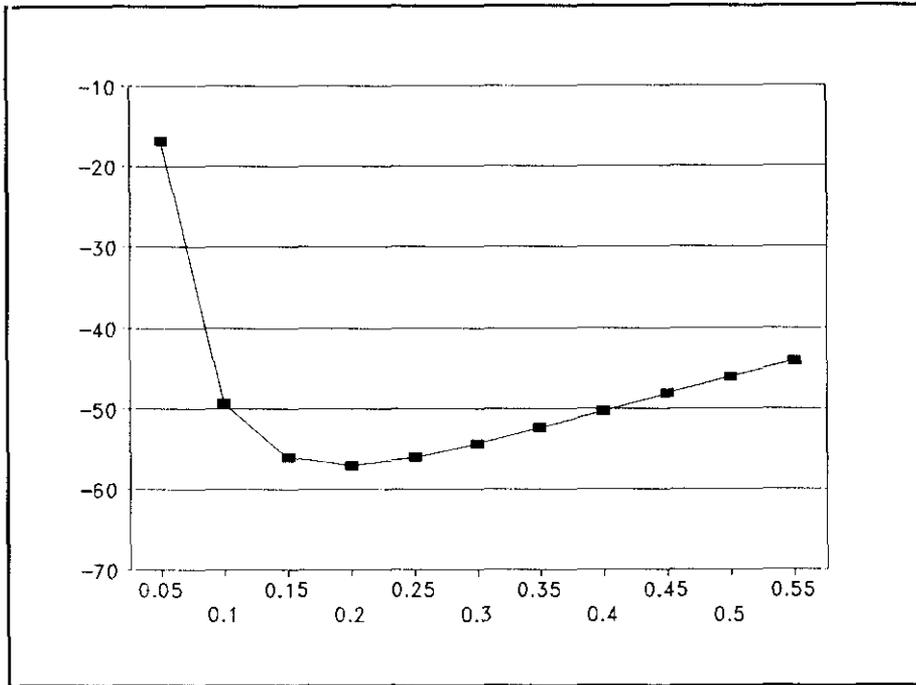


Figura 4.5.3: Perfil de la función $-\ln L(\theta)$ con respecto al parámetro σ_c^2 .

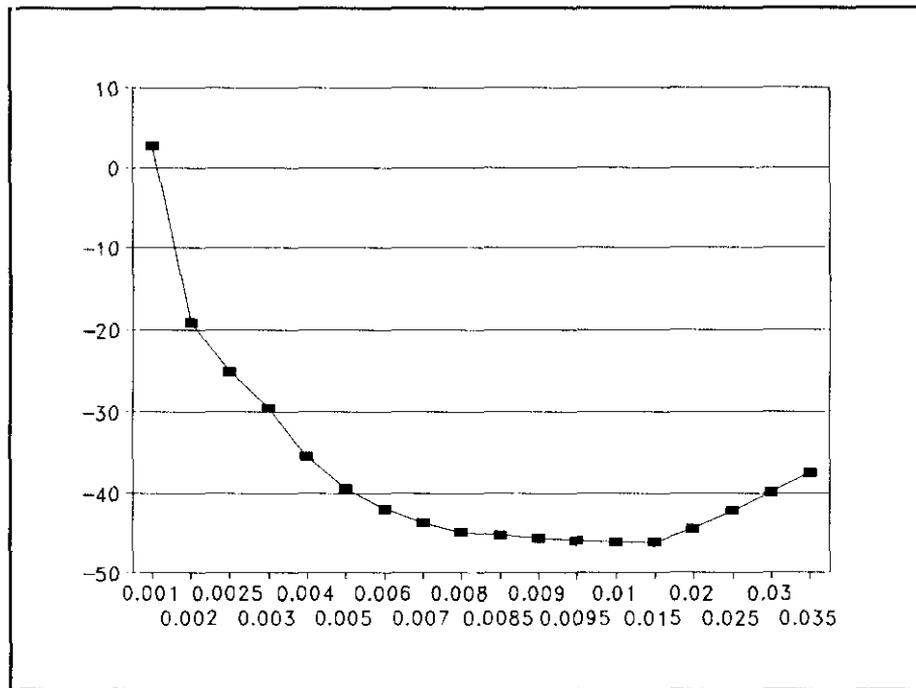


Figura 4.5.4: Perfil de la función $-\ln L(\theta)$ con respecto al parámetro σ_q^2 .

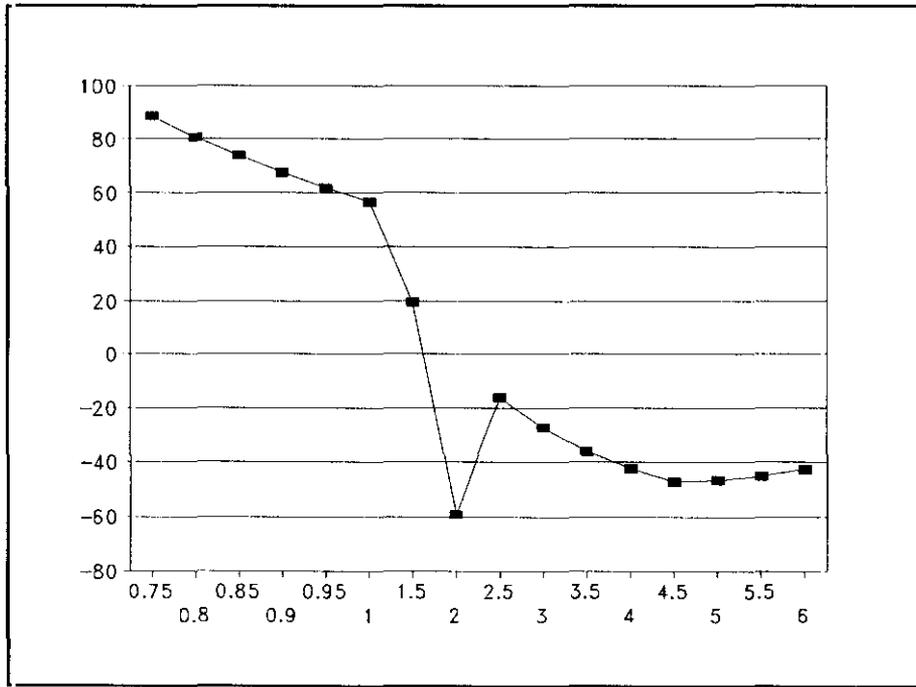


Figura 4.5.5: Perfil de la función $-\ln L(\theta)$ con respecto al parámetro α .

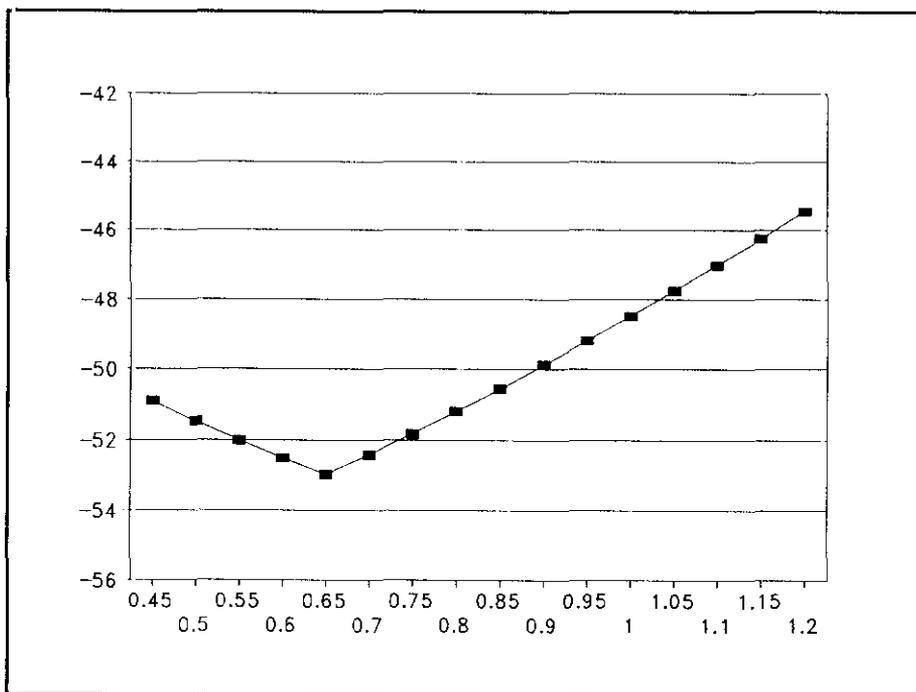


Figura 4.5.6: Perfil de la función $-\ln L(\theta)$ con respecto al parámetro β .

CAPITULO 5
CONCLUSIONES



Las principales conclusiones de este trabajo son las siguientes:

En primer lugar, se ha tratado el problema de estimación por máxima verosimilitud exacta de modelos econométricos utilizando un filtro de Chandrasekhar. En este punto, se ha llevado a cabo un estudio exhaustivo del funcionamiento del procedimiento de estimación, al ser aplicado tanto a modelos con datos simulados como con datos económicos reales. Además, las especificaciones consideradas presentan distintas estructuras dinámicas y situaciones de condicionamiento. Las ventajas de esta aproximación son las siguientes:

- Los resultados confirman la mayor robustez numérica del filtro de Chandrasekhar frente al filtro de Kalman convencional. En concreto, la utilización de versiones estables del filtro de Kalman no es suficiente para eliminar los errores numéricos inherentes al mismo y que dan lugar a la obtención de matrices de covarianzas con autovalores estrictamente negativos. Sin embargo, en todas las pruebas realizadas con el filtro de Chandrasekhar, se muestra la robustez numérica de este algoritmo proporcionando matrices de covarianzas semidefinidas positivas en todo instante de tiempo.
- Por otra parte, el algoritmo de Chandrasekhar presenta importantes ventajas computacionales con respecto al filtro de Kalman cuando se aplica a modelos con dimensión del vector de estado (n) mayor a la dimensión del vector de observación (m). Esto ocurre en la mayoría de los modelos econométricos que presentan estructura estacional. La ganancia porcentual en tiempo de cálculo puede llegar a ser superior al 80-90% con respecto al filtro de Kalman, creciendo de forma exponencial a medida que aumenta el ratio de dimensiones n/m .

- A partir de las ecuaciones de Chandrasekhar, pueden derivarse las expresiones analíticas exactas de la función de verosimilitud de la muestra, de su gradiente, su hessiano y su matriz de información. Calcular de forma exacta estas funciones resulta importante a la hora de estimar y especialmente, para contrastar cualquier hipótesis (lineal o no lineal) sobre los parámetros.

Un aspecto crucial a la hora de utilizar un filtro, consiste en estimar adecuadamente las condiciones iniciales del mismo ya que, cuando la muestra es corta, las estimaciones de los parámetros del modelo econométrico pueden ser sensibles a las condiciones iniciales del. La aportación de este trabajo en este aspecto es la siguiente:

- Se ha aplicado un algoritmo de estimación por máxima verosimilitud de las condiciones iniciales del filtro. Esto evita tanto la arbitrariedad con que se eligen éstas en la mayoría de los casos, como los posibles errores numéricos derivados de una mala elección de condiciones iniciales (vid. Apéndice B).

Por otra parte, cabe destacar las ventajas de usar una representación en forma de espacio de los estados con dimensión mínima del vector de estado, ya que:

- Usando dicha formulación, se simplifica la aplicación de las ecuaciones de Chandrasekhar a la estimación de los parámetros del modelo econométrico.
- Proporciona ventajas computacionales con respecto a otras formulaciones de dimensión no mínima.
- Evita que alguna(s) componente(s) del vector de estado inicial sea(n) no identificable(s), cuando se estiman

los elementos de dicho vector por el algoritmo de máxima verosimilitud ya citado.

Finalmente, en este trabajo se muestra que los filtros derivados de la factorización de matrices (como es el de Chandrasekhar) no son difíciles de aplicar en la práctica y que su construcción garantiza la estabilidad del filtro. En nuestro contexto, es conveniente utilizar algoritmos con esta característica, ya que las variables que se propagan recursivamente en el filtro son usadas para evaluar, en cada vector de parámetros, la función de verosimilitud del modelo, su gradiente y su hessiano. Como es bien sabido, el proceso recursivo es sensible a la acumulación de errores de redondeo.

Por otro lado, es importante disponer de las expresiones exactas de dichas funciones para evitar el uso de aproximaciones a las mismas. Pero, igualmente importante es utilizar después un filtro numéricamente robusto para evaluar con la mayor precisión posible los valores exactos de dichas funciones. La robustez numérica del filtro de Chandrasekhar hace que dicho algoritmo sea menos susceptible a los efectos de posibles errores numéricos.

Puesto que tenemos disponibles algoritmos superiores al filtro de Kalman en cuanto a robustez numérica y eficiencia computacional, una extensión lógica de este trabajo sería implementarlos también de forma eficiente mediante el desarrollo de programas informáticos de fácil manejo para el usuario. Este aspecto es cada vez más importante al ser cada vez mayor la gama de aplicaciones de las técnicas de filtrado (casi siempre del filtro de Kalman) y smoothing a problemas econométricos ya tratados en la literatura. Algunos ejemplos son la estimación de modelos que contienen expectativas racionales de algunas variables [vid. Watson (1989)], la estimación de modelos bajo estructuras de heterocedasticidad dinámica (ARCH) [vid. Harvey (1989)], la interpolación óptima de datos [vid. Kohn y Ansley (1983)], la estimación de modelos con errores de observación en las variables o con distinta periodicidad muestral [vid. Terceiro

(1990)], la estimación de modelos con parámetros cambiantes [vid. Young (1984), los procedimientos de desestacionalización, etc.

APENDICE A

El filtro de Kalman en tiempo discreto

La derivación de este filtro parte de las siguientes hipótesis sobre la formulación de un modelo en espacio de los estados dado por [1.1.1] y [1.1.2]:

Hipótesis:

- (a) Los ruidos \mathbf{v}_t y \mathbf{w}_t son procesos de ruido blanco con esperanza nula y matrices de covarianzas \mathbf{R}_t y \mathbf{Q}_t , respectivamente:

$$\begin{aligned} E(\mathbf{v}_t) &= \mathbf{0} & ; & & \text{COV}(\mathbf{v}_t) &= \mathbf{R}_t & & \forall t \in \{1, 2, \dots, n\} \\ E(\mathbf{w}_t) &= \mathbf{0} & ; & & \text{COV}(\mathbf{w}_t) &= \mathbf{Q}_t & & \forall t \in \{1, 2, \dots, n\} \end{aligned}$$

- (b) Los ruidos de estado y de observación no están correlacionados entre sí.

$$E(\mathbf{w}_t, \mathbf{v}_s') = 0 \quad \forall t, s \in \{1, 2, \dots, n\}$$

Este supuesto puede relajarse, permitiendo que los ruidos de estado y de observación estén correlacionados contemporáneamente. En ese caso, la correspondiente matriz de varianzas-covarianzas se denotará por \mathbf{S}_t :

$$E(\mathbf{w}_t, \mathbf{v}_s') = \mathbf{S}_t \delta_{ts} \quad \delta_{ts} = \begin{cases} 1 & \text{si } t=s \\ 0 & \text{si } t \neq s \end{cases} \\ \forall t, s \in \{1, 2, \dots, n\}$$

- (c) El vector de estado inicial \mathbf{x}_0 , tiene media $\bar{\mathbf{x}}_0$ y una matriz de varianzas-covarianzas \mathbf{P}_0 . Es decir:

$$E(\mathbf{x}_0) = \bar{\mathbf{x}}_0 \quad ; \quad \text{COV}(\mathbf{x}_0) = \mathbf{P}_0$$

- (d) Los ruidos de estado y de observación no están correlacionados con el vector de estado inicial.

$$\begin{aligned}
E(\mathbf{v}_t, \mathbf{x}_0') &= \mathbf{0} & \forall t \in \{1, 2, \dots, n\} \\
E(\mathbf{w}_t, \mathbf{x}_0') &= \mathbf{0} & \forall t \in \{1, 2, \dots, n\}
\end{aligned}$$

En el caso de modelos con ruido homoscedástico, el supuesto (a) se convierte en el siguiente:

$$\begin{aligned}
E(\mathbf{v}_t) &= \mathbf{0} & ; & \quad \text{COV}(\mathbf{v}_t) = \mathbf{R} & \quad \forall t \in \{1, 2, \dots, n\} \\
E(\mathbf{w}_t) &= \mathbf{0} & ; & \quad \text{COV}(\mathbf{w}_t) = \mathbf{Q} & \quad \forall t \in \{1, 2, \dots, n\}
\end{aligned}$$

Y si se permite que los ruidos de estado y de observación estén correlacionados contemporáneamente, éstos tendrán una matriz de covarianzas constante denotada por \mathbf{S} .

$$\begin{aligned}
E(\mathbf{w}_t, \mathbf{v}_s') &= \mathbf{S} \delta_{ts} & \delta_{ts} &= \begin{cases} 1 & \text{si } t=s \\ 0 & \text{si } t \neq s \end{cases} \\
& & & \forall t, s \in \{1, 2, \dots, n\}
\end{aligned}$$

La información disponible consiste en el vector de variables observadas \mathbf{z}_t , el vector de variables de control \mathbf{u}_t , el vector de estado inicial \mathbf{x}_0 y los elementos de las matrices \mathbf{F}_t , \mathbf{E}_t , \mathbf{H}_t , \mathbf{D}_t , \mathbf{C}_t , \mathbf{Q}_t , \mathbf{R}_t y \mathbf{S}_t (en su caso).

A.1. Derivación del filtro de Kalman.

En este apartado, se derivará la forma general del filtro de Kalman en dos casos: (a) cuando los ruidos de estado y de observación son independientes entre sí y (b) cuando se supone que están correlacionados contemporáneamente.

(a) Caso en que los ruidos \mathbf{v}_t y \mathbf{w}_t no están correlacionados.

Sea un modelo expresado en la forma [1.1.1]-[1.1.2], donde los ruidos \mathbf{v}_t y \mathbf{w}_t son independientes entre sí (es decir, $\mathbf{S}_t = \mathbf{0}$, $\forall t$). Se denota $\hat{\mathbf{x}}_{t-1/t-1}$ al estimador óptimo de \mathbf{x}_{t-1} basado en la información disponible hasta el instante $t-1$ y $\mathbf{P}_{t-1/t-1}$ a la matriz de covarianzas del error de esa estimación. Es decir:

$$P_{t-1/t-1} = E[(x_{t-1} - \hat{x}_{t-1/t-1})(x_{t-1} - \hat{x}_{t-1/t-1})']$$

Dado $\hat{x}_{t-1/t-1}$ y $P_{t-1/t-1}$, el estimador óptimo de x_t viene dado por la expresión:

$$\hat{x}_{t/t-1} = \Phi_t \hat{x}_{t-1/t-1} + \Gamma_t u_t \quad [A.1.1]$$

y la matriz de covarianzas del error de esa estimación tiene la expresión:

$$\begin{aligned} P_{t/t-1} &= E[(x_t - \hat{x}_{t/t-1})(x_t - \hat{x}_{t/t-1})'] = \\ &= E[\Phi_t(x_{t-1} - \hat{x}_{t-1/t-1})(x_{t-1} - \hat{x}_{t-1/t-1})'\Phi_t'] + \\ &+ E[E_t w_t w_t' E_t'] = \Phi_t P_{t-1/t-1} \Phi_t' + E_t Q_t E_t' \end{aligned} \quad [A.1.2]$$

Las ecuaciones [A.1.1] y [A.1.2] son las llamadas ecuaciones de predicción del filtro de Kalman.

La expresión [A.1.1] se obtiene tomando esperanzas condicionadas por la información disponible hasta $t-1$ en la ecuación [1.1.1] y aplicando la hipótesis de que la esperanza de w_t es nula.

La ecuación [A.1.2] se obtiene a partir de [1.1.1] y de [A.1.2], aplicando el supuesto de que w_t es independiente del vector de estado inicial.

Cuando está disponible un nuevo dato sobre las variables observadas (z_t), el filtro de Kalman actualiza la estimación denotada por $\hat{x}_{t/t-1}$, de la siguiente forma:

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + K_t \tilde{z}_t \quad [A.1.3]$$

$$\tilde{z}_t = z_t - H_t \hat{x}_{t/t-1} - D_t u_t \quad [A.1.4]$$

$$P_{t/t} = [I - K_t H_t] P_{t/t-1} \quad [A.1.5]$$

$$K_t = P_{t/t-1} H_t' B_t^{-1} \quad [A.1.6]$$

$$B_t = H_t P_{t/t-1} H_t' + C_t R_t C_t' \quad [A.1.7]$$

Las expresiones [A.1.3] y [A.1.5] son las llamadas ecuaciones de actualización del filtro de Kalman.

La ecuación [A.1.3] actualiza la estimación del vector de estado ponderando por la matriz ganancia K_t , el error cometido al estimar la medida z_t mediante su esperanza condicionada por la información disponible hasta el instante $t-1$. Ese vector de errores está definido en la ecuación [A.1.4] y es conocido como proceso de innovaciones, ya que representa la nueva información que añade la última observación procesada. El vector de innovaciones puede escribirse alternativamente como:

$$\tilde{z}_t = H_t(x_t - \hat{x}_{t/t-1}) + C_t v_t$$

y, de acuerdo con esta expresión, es fácil ver que dicho proceso tiene esperanza nula y que su matriz de covarianzas viene dada por la matriz B_t definida en [A.1.7]. Es decir:

$$B_t = E[\tilde{z}_t \tilde{z}_t']$$

Se supone que existe la inversa de B_t . Esto puede no ser cierto si el vector v_t tiene componentes deterministas.

La ecuación [A.1.5] proporciona la matriz de covarianzas del error de la estimación de x_t dada por la expresión [A.1.3]. Es decir¹:

¹ Es fácil mostrar que una expresión alternativa a [A.1.5] es la siguiente:

$$P_{t/t} = [I - K_t H_t] P_{t/t-1} [I - K_t H_t]' + K_t C_t R_t C_t' K_t' \quad [A.1.5']$$

Esta última expresión tiene la ventaja de ser más robusta numéricamente que [A.1.5] al componerse de la suma de dos matrices definidas positivas por construcción, aunque tiene el inconveniente de que su cálculo supone un mayor número de operaciones que la ecuación [A.1.5]. Las ecuaciones de predicción [A.1.1] y [A.1.2] junto con las de actualización [A.1.3] y [A.1.5'] describen la llamada versión estable del filtro de Kalman. Sin embargo, Thornton y Bierman (1980) muestran con

$$P_{t/t} = E[(x_t - \hat{x}_{t/t})(x_t - \hat{x}_{t/t})']$$

La ecuación [A.1.6] define la matriz K_t , que se conoce como la ganancia del filtro de Kalman. Esta matriz mide el efecto de una nueva observación sobre la estimación del vector de estado. De acuerdo con [A.1.6], esta matriz es directamente proporcional a la incertidumbre asociada a la estimación del vector de estado e inversamente proporcional a la incertidumbre asociada al vector de observaciones. Se demostrará más adelante que la expresión de la ganancia dada por [A.1.6] es la que minimiza la traza de $P_{t/t}$.

Las ecuaciones de predicción y actualización del filtro de Kalman pueden expresarse conjuntamente, de la siguiente forma:

$$\hat{x}_{t+1/t} = (\Phi_{t+1} - K_t H_t) \hat{x}_{t/t-1} + K_t (z_t - D_t u_t) + \Gamma_{t+1} u_{t+1} \quad [A.1.8]$$

$$\text{donde } K_t = \Phi_{t+1} P_{t/t-1} H_t' B_t^{-1} \quad [A.1.9]$$

y la ecuación de propagación de la matriz de covarianzas del error de estimación del vector de estado es ahora:

$$P_{t+1/t} = \Phi_{t+1} P_{t/t-1} \Phi_{t+1}' + E_{t+1} Q_{t+1} E_{t+1}' - K_t B_t K_t' \quad [A.1.10]$$

o bien:

$$P_{t+1/t} = \Phi_{t+1} [P_{t/t-1} - P_{t/t-1} H_t' B_t^{-1} H_t P_{t/t-1}] \Phi_{t+1}' + E_{t+1} Q_{t+1} E_{t+1}'$$

Esta última expresión es conocida como la ecuación de Riccati del filtro de Kalman (vid. Apéndice C). La ecuación [A.1.10] es poco aconsejable numéricamente, pues se calcula como la diferencia de dos matrices definidas positivas y esto puede

diversos ejemplos que, incluso utilizando esta versión más estable, el filtro puede llegar a degradarse numéricamente.

hacer que $P_{t+1/t}$ pierda su condicionamiento en algún momento de la propagación. Existe también una versión estable de esa ecuación:

$$P_{t+1/t} = \bar{\hat{x}}_t P_{t/t-1} \bar{\hat{x}}_t' + \begin{bmatrix} E_t & -K_t C_t \end{bmatrix} \begin{bmatrix} Q_t & S_t \\ S_t' & R_t \end{bmatrix} \begin{bmatrix} E_t' \\ -C_t' K_t' \end{bmatrix} \quad [\text{A.1.11}]$$

siendo $\bar{\hat{x}}_t = \hat{x}_t - K_t H_t$.

Hay que señalar que los errores numéricos cometidos en el cálculo de la ganancia a través de [A.1.9] provocarían errores de primer orden en el cálculo de $P_{t+1/t}$ a través de [A.1.10] y de segundo orden si se utiliza en su lugar la expresión [A.1.11].

(b) Caso en que los ruidos v_t y w_t están correlacionados.

Cuando la formulación del modelo en espacio de los estados permita que los ruidos v_t y w_t estén correlacionados contemporáneamente, las ecuaciones de predicción del filtro siguen siendo las mismas, mientras que las de actualización se convierten en las siguientes:

$$\hat{x}_{t/t} = \hat{x}_{t/t-1} + K_t \bar{z}_t \quad [\text{A.1.3}']$$

$$\bar{z}_t = z_t - H_t \hat{x}_{t/t-1} - D_t u_t \quad [\text{A.1.4}']$$

$$K_t = (P_{t/t-1} H_t' + E_t S_t C_t') B_t^{-1} \quad [\text{A.1.5}']$$

$$B_t = H_t P_{t/t-1} H_t' + C_t R_t C_t' + H_t E_t S_t C_t' + C_t S_t E_t' H_t' \quad [\text{A.1.6}']$$

$$P_{t/t} = P_{t/t-1} - (P_{t/t-1} H_t' + E_t S_t C_t') B_t^{-1} (H_t P_{t/t-1} + C_t S_t E_t') \quad [\text{A.1.7}']$$

donde las variables del filtro siguen teniendo la misma interpretación estadística que en el caso (a).

Por último, si se utiliza la formulación [1.1.1']-[1.1.2], pueden unirse las ecuaciones de predicción y de actualización resultando las expresiones:

$$\hat{\mathbf{x}}_{t+1/t} = \Phi_t \hat{\mathbf{x}}_{t/t-1} + \Gamma_t \mathbf{u}_t + \mathbf{K}_t \tilde{\mathbf{z}}_t \quad [\text{A.1.3}''']$$

$$\tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{H}_t \hat{\mathbf{x}}_{t/t-1} - \mathbf{D}_t \mathbf{u}_t \quad [\text{A.1.4}''']$$

$$\mathbf{K}_t = (\Phi_t \mathbf{P}_{t/t-1} \mathbf{H}_t' + \mathbf{E}_t \mathbf{S}_t \mathbf{C}_t') \mathbf{B}_t^{-1} \quad [\text{A.1.5}''']$$

$$\mathbf{P}_{t+1/t} = \Phi_t \mathbf{P}_{t/t-1} \Phi_t' + \mathbf{E}_t \mathbf{Q}_t \mathbf{E}_t' - \mathbf{K}_t \mathbf{B}_t \mathbf{K}_t' \quad [\text{A.1.6}''']$$

$$\mathbf{B}_t = \mathbf{H}_t \mathbf{P}_{t/t-1} \mathbf{H}_t' + \mathbf{C}_t \mathbf{R}_t \mathbf{C}_t' \quad [\text{A.1.7}''']$$

Esta última forma del filtro de Kalman será la más utilizada en este trabajo.

A.2. Propiedades del filtro de Kalman.

- Es un filtro lineal, aplicable a cualquier modelo dinámico de orden finito, y computable de forma recursiva. Por tanto, la estimación del vector de estado puede actualizarse a medida que llegue nueva información de forma secuencial.
- El filtro proporciona una estimación del vector de estado, $\hat{\mathbf{x}}_{t+1/t}$, que depende del vector de observaciones \mathbf{z}_t . Sin embargo, la ecuación que propaga la matriz de covarianzas del estimador del vector de estado muestra que ésta es independiente del vector de medidas. Es decir, ninguna observación ayuda más que otra a reducir la incertidumbre existente acerca del vector de estado, aunque está garantizado el no crecimiento de dicha incertidumbre en el ciclo de actualización².
- La ganancia del filtro \mathbf{K}_t es también independiente de la información disponible hasta el instante $t-1$. Por tanto, las expresiones [A.1.5''']-[A.1.7'''] pueden propagarse independientemente del resto de las ecuaciones del filtro. Esta propiedad no se cumple para cualquier filtro [vid. Anderson y Moore (1979)].

² Esto se debe a que $\text{tr}(\mathbf{P}_{t/t}) \leq \text{tr}(\mathbf{P}_{t/t-1})$.

- En sistemas con matrices constantes, la ganancia del filtro y la matriz P_t variarán en el tiempo, incluso si el estado del sistema es estacionario con respecto a su media y varianza. Esto no quiere decir que, a partir de un instante determinado, ambas matrices no puedan alcanzar un valor estacionario.
- El filtro de Kalman es óptimo porque la ganancia (K_t) se elige de forma que minimiza la traza de la matriz $P_{t/t}$. Demetry (1970) demostró que la ganancia del filtro de Kalman no sólo minimiza la traza de $P_{t/t}$, sino cualquier combinación lineal de los elementos de la diagonal principal de dicha matriz. La optimalidad del filtro se demuestra en el siguiente teorema [vid. Jazwinsky (1970)].

Teorema: La expresión de la ganancia dada por la ecuación [A.1.6] es la que minimiza la traza de $P_{t/t}$.

Demostración: Supongamos que existe otra matriz ganancia definida como $K_t^* = K_t + \tilde{K}_t$ tal que se cumpla que $\text{tr}(P_{t/t}^*) < \text{tr}(P_{t/t})$, siendo K_t una matriz cualquiera no nula y $K_t = P_{t/t-1}H_t'B_t^{-1}$. En este caso, la matriz de covarianzas del estimador del vector de estado tiene la expresión:

$$\begin{aligned}
 P_{t/t}^* &= [I - (K_t + \tilde{K}_t)H_t]P_{t/t-1}[I - (K_t + \tilde{K}_t)H_t]' + \\
 &+ (K_t + \tilde{K}_t)C_tR_tC_t'(K_t + \tilde{K}_t)' = \\
 &= [I - K_tH_t]P_{t/t-1}[I - K_tH_t]' + [K_tH_t - I]P_{t/t-1}H_t'\tilde{K}_t' + \\
 &+ \tilde{K}_tH_tP_{t/t-1}[H_t'K_t' - I] + \tilde{K}_tC_tR_tC_t'K_t + K_tC_tR_tC_t'\tilde{K}_t + \\
 &+ \tilde{K}_tH_tP_{t/t-1}H_t'\tilde{K}_t' + K_tC_tR_tC_t'K_t + \tilde{K}_tC_tR_tC_t'\tilde{K}_t'
 \end{aligned}$$

y sustituyendo K_t por su expresión en los siguientes sumandos: $[K_tH_t - I]P_{t/t-1}H_t'\tilde{K}_t'$ y $\tilde{K}_tH_tP_{t/t-1}[H_t'K_t' - I]$, se obtiene:

$$[K_tH_t - I]P_{t/t-1}H_t'\tilde{K}_t' = -K_tC_tR_tC_t'\tilde{K}_t'$$

$$\tilde{K}_t H_t P_{t/t-1} [H_t' K_t' - I] = -\tilde{K}_t C_t R_t C_t' K_t'$$

Por tanto:

$$P_{t/t}^* = P_{t/t} + \tilde{K}_t B_t \tilde{K}_t'$$

De esta última ecuación se deduce que $\text{tr}(P_{t/t}^*) > \text{tr}(P_{t/t})$ al ser $\tilde{K}_t B_t \tilde{K}_t'$ una matriz definida positiva $\forall \tilde{K}_t \neq 0$. Pero esto está en contradicción con la hipótesis de partida. Por tanto, no existe una matriz distinta de K_t que reduzca más la $\text{tr}(P_{t/t})$.

C.Q.D.

A.3. Interpretación de los resultados del filtro de Kalman para modelos gaussianos.

Un aspecto importante del filtro de Kalman es que cuando el modelo es gaussiano (es decir, cuando la distribución del vector de estado inicial x_0 y de los ruidos v_t y w_t sea normal y ambos ruidos sean independientes de x_0), las ecuaciones que forman el filtro de Kalman son las mismas que en el caso no gaussiano, pero la interpretación de sus resultados es distinta. Una derivación de las ecuaciones del filtro de Kalman bajo la hipótesis de normalidad puede verse en Anderson y Moore (1979) y Harvey (1989).

En el caso no gaussiano, $\hat{x}_{t/t}$ es un estimador lineal, insesgado y óptimo de x_t basado en la información disponible hasta el instante t y $P_{t/t}$ es la matriz de covarianzas del error de esa estimación.

En el caso gaussiano, el filtro de Kalman proporciona, además, la media y la matriz de covarianzas de la distribución de x_t condicionada por el conjunto de información disponible hasta ese instante. Es decir:

$$\hat{x}_{t/t} = E(x_t / z^t)$$

$$\hat{P}_{t/t} = E[(\mathbf{x}_t - \hat{\mathbf{x}}_{t/t})(\mathbf{x}_t - \hat{\mathbf{x}}_{t/t})' | \mathbf{z}^t]$$

donde $\hat{\mathbf{x}}_{t/t}$ se interpreta como la media de la distribución condicionada de \mathbf{x}_t , $P_{t/t}$ la matriz de covarianzas de esa distribución y \mathbf{z}^t es el conjunto de información disponible hasta ese momento. En este caso, la distribución del vector de estado está completamente caracterizada.

Si son conocidos todos los elementos del vector inicial \mathbf{x}_0 , entonces éste tendrá una distribución a priori³ con media \mathbf{x}_0 y matriz de covarianzas P_0 . En este caso, el filtro de Kalman permite calcular la función de verosimilitud exacta de las observaciones. Esta será la vía para poder estimar cualquier parámetro desconocido de un modelo, así como para construir los test estadísticos para la validación del mismo una vez que ha sido estimado.

A.4. Inicialización del filtro de Kalman.

Cualquiera que sea la forma concreta que tome el filtro de Kalman, siempre necesita ser inicializado con los valores \mathbf{x}_0 y P_0 , que pueden ser desconocidos. Una forma de fijar estas condiciones iniciales consiste en tomar $\mathbf{x}_0 = \mathbf{0}$ y $P_0 = \tau \mathbf{I}$, donde τ es un escalar positivo y arbitrariamente grande. Estas condiciones indican un elevado grado de incertidumbre acerca del estado inicial del sistema. Esta estrategia para determinar las condiciones iniciales puede ser una fuente de inestabilidad numérica del filtro de Kalman, existiendo métodos alternativos para su estimación. El criterio más habitual es utilizar un subconjunto inicial de observaciones para obtener un conocimiento previo acerca del vector de estado. También es posible considerar que todos los elementos de \mathbf{x}_0 son parámetros constantes pero des-

³ Es usual no disponer de información sobre \mathbf{x}_0 , pero en estos casos, puede utilizarse un subconjunto de observaciones iniciales para formar una distribución a priori sobre \mathbf{x}_t .

conocidos, que hay que estimar como parte del procedimiento de máxima verosimilitud (vid. Apéndice B).

En el caso de total o casi total incertidumbre acerca del vector de estado inicial, puede ser aconsejable utilizar el filtro de la matriz de información. Este filtro es equivalente al filtro de Kalman y la idea básica del mismo consiste en propagar la inversa de P_t (o matriz de información) en lugar de ella misma. En este algoritmo, no se obtiene una ecuación de propagación para el estimador del vector de estado, sino para un vector a_t que es combinación lineal de las componentes del vector de estado original. Las condiciones iniciales de este filtro son $a_0 = 0$ y $P_0^{-1} = 0$ y llamando $P_t^{-1} = I_t$, las correspondientes ecuaciones de propagación y actualización son las siguientes:

$$I_{t/t} = I_{t/t-1} + H_t' R_t^{-1} H_t \quad [A.4.1]$$

$$I_{t+1/t} = [I - L_t E_t'] A_t \quad [A.4.2]$$

$$A_t = [\xi_t^{-1}]' I_{t/t} \xi_t^{-1} \quad [A.4.3]$$

$$L_t = A_t E_t [E_t' A_t E_t + Q_t^{-1}]^{-1} \quad [A.4.4]$$

$$K_t = \xi_t I_{t/t} H_t' R_t^{-1} \quad [A.4.5]$$

$$\hat{a}_{t/t-1} = I_{t/t-1} \hat{x}_{t/t-1} \quad [A.4.6]$$

$$\hat{a}_{t/t} = I_{t/t} \hat{x}_{t/t} \quad [A.4.7]$$

$$\hat{a}_{t+1/t} = [I - L_t E_t'] [\xi_t^{-1}]' \hat{a}_{t/t} \quad [A.4.8]$$

$$\hat{a}_{t/t} = \hat{a}_{t/t-1} + H_t' R_t^{-1} z_t \quad [A.4.9]$$

Desde el punto de vista computacional, es más aconsejable que la versión convencional del filtro de Kalman cuando la dimensión del vector de observación es mayor que la del vector de estado⁴. Este es el caso menos frecuente en Econometría, donde

⁴ Esto se debe a que en el filtro de Kalman se ha de calcular la inversa de la matriz $H_t P_{t/t-1} H_t' + R_t$, mientras que en el filtro de información se ha de invertir $E_t' A_t E_t + Q_t^{-1}$. Por otro lado, el algoritmo de la matriz de información requiere el cálculo de las inversas de las matrices de ruido (R_t y Q_t), no necesario en el filtro de Kalman. Por tanto, las dimensiones del

lo más usual es que el orden del vector de estado supere en gran medida al del vector de observaciones. Por esta razón, en este trabajo no se presta demasiada atención a este algoritmo, cuya aplicación puede ser aconsejable en otros casos⁵.

En cuanto a la robustez numérica del filtro, es fácil ver que el uso de la expresión [A.4.2] para propagar la matriz de información, puede llevar a que dicha matriz pierda su simetría y su carácter de definida positiva. Aunque existe una versión estable de [A.4.2], la forma más eficiente de resolver dichos problemas numéricos consiste en utilizar el filtro de la raíz cuadrada de la matriz de información⁶.

problema y la estructura particular de las matrices de ruido son los dos aspectos que pueden hacer más atractivo, desde un punto de vista computacional, un filtro u otro. Así, el filtro de la matriz de información sería más eficiente cuando la dimensión del vector z_t fuera grande, las matrices de covarianzas de los ruidos fuesen diagonales y sólo se necesitase la matriz de covarianzas del vector de estado, después de haber procesado toda la información disponible.

⁵ En concreto, la aplicación de este filtro es muy conveniente en la estimación recursiva de modelos de regresión lineales, básicamente por tres razones: a) la independencia del filtro de condiciones iniciales arbitrarias, b) la robustez numérica del algoritmo al propagarse la matriz de información como suma de matrices definidas positivas y c) las estimaciones finales resultantes del filtro coinciden analítica y numéricamente con la estimación M.C.O. al final de la muestra.

⁶ Vid. Morf, Sidhu y Kailath (1974) y Bierman, Belzer y Vandergraft (1977).

APENDICE B

Estimación de las condiciones iniciales óptimas del vector de estado.

El interés del desarrollo de este apéndice se debe a que el procedimiento de estimación descrito en los Capítulos 2 y 3 necesita dos tipos de condiciones iniciales: a) las condiciones iniciales del vector de estado y b) un conjunto de preestimaciones de los parámetros del modelo econométrico.

Con respecto al primer tipo de condiciones iniciales, se obtiene el estimador máximo-verosímil del estado inicial, condicionado por el valor de los parámetros propios del modelo econométrico. Este método puede utilizarse para inicializar de forma óptima tanto un filtro de Kalman como uno de tipo Chandra-sekhar.

Caben dos posibilidades para obtener una estimación del vector de estado inicial \mathbf{x}_0 :

- (a) Estimar el vector inicial de estado basándose en el comportamiento posterior de dicha variable. En este caso, se obtiene una aproximación a la función de verosimilitud al considerar la estimación de \mathbf{x}_0 como su verdadero valor. Esta aproximación puede resultar eficiente en los casos en que la función de verosimilitud sea poco sensible a las condiciones iniciales del sistema. No es conveniente, por ejemplo, en los casos en que el tamaño muestral es pequeño, ya que las estimaciones de los parámetros del modelo pueden ser muy sensibles a las condiciones iniciales del vector de estado de las que se parta. Existen algoritmos basados en esta aproximación, como son los llamados algoritmos de punto fijo. Estos calculan un estimador óptimo de \mathbf{x}_0 por retrovisión. Un estudio más

detallado puede verse en Anderson y Moore (1979) y Parzen (1984).

- (b) Considerar los elementos del vector \mathbf{x}_0 como parámetros constantes pero desconocidos, que necesitan ser estimados como parte del procedimiento de máxima verosimilitud. Al considerar constantes los elementos de \mathbf{x}_0 , es razonable suponer que $\mathbf{P}_0 = \mathbf{0}$. Con esta idea, se obtiene que el estimador máximo-verosímil de \mathbf{x}_0 , condicional al resto de parámetros, es una función lineal de las observaciones. Este estimador puede concentrarse fuera de la función de verosimilitud que sólo necesitará ser maximizada con respecto a los parámetros propios del modelo. Un ejemplo de esta aproximación ha sido el algoritmo desarrollado por Terceiro (1990), que incluye como caso particular al de Rosenberg (1973).

En este trabajo, se utiliza el algoritmo de condiciones iniciales debido a Terceiro (1990) para inicializar un filtro de Chandrasekhar y, mediante el mismo, estimar los parámetros propios del modelo econométrico.

B.1. Estimador máximo-verosímil del vector de estado inicial.

El vector de parámetros que se desea estimar ahora es:

$$\theta = [\theta_1' \ \mathbf{x}_0']'$$

donde:

θ_1 es un vector que contiene todos los elementos desconocidos de las matrices \mathbf{F} , \mathbf{E} , \mathbf{C} , \mathbf{D} , \mathbf{Q} , \mathbf{R} y \mathbf{S} .

\mathbf{x}_0 es el vector ($n \times 1$) de las condiciones iniciales del vector de estado \mathbf{x}_t .

La función a minimizar con respecto al vector de parámetros es:

$$\ell(\theta) = \frac{1}{2} \sum_{t=1}^n (\ln 2\pi + \ln |B_t| + \tilde{z}_t' B_t^{-1} \tilde{z}_t)$$

siendo \tilde{z}_t el proceso de innovaciones (o vector de errores de predicción) y B_t su matriz de covarianzas.

Sustituyendo [A.4'''] en [A.3'''] se obtiene:

$$\hat{x}_{t+1/t} = \bar{\Phi}_t \hat{x}_{t/t-1} + \bar{\Gamma}_t u_t + K_t z_t$$

donde:

$$\bar{\Phi}_t = \Phi - K_t H \quad ; \quad \bar{\Gamma}_t = \Gamma - K_t H$$

Dadas unas condiciones iniciales x_0^* y $P_0^* = 0$, el filtro de Chandrasekhar proporcionará una estimación del vector de estado $\hat{x}_{t+1/t}^*$ y un proceso de innovaciones \tilde{z}_t^* . Con unas condiciones iniciales distintas, x_0 y $P_0 = 0$, los resultados del filtro serán $\hat{x}_{t+1/t}$ y \tilde{z}_t . Para evaluar la función de verosimilitud, en cada vector de parámetros, es necesario calcular este último vector de errores de predicción y su matriz de covarianzas B_t . Teniendo en cuenta que las matrices $P_{t/t-1}$ son las mismas en los dos filtros, la idea es obtener \tilde{z}_t del output del primer filtro.

Es fácil demostrar que la relación existente entre el estimador del vector de estado y del proceso de innovaciones, ante un cambio en las condiciones iniciales del vector de estado, es la siguiente:

$$\tilde{z}_t = \tilde{z}_t^* - H \bar{\Phi}_{t-1} (x_0 - x_0^*) \quad [B.1.1]$$

donde $\bar{\Phi}_t$ se calcula recursivamente de la forma:

$$\bar{\Phi}_t = (\Phi - K_t H) \bar{\Phi}_{t-1} \quad [B.1.2]$$

comenzando con $\bar{\Sigma}_0 = I_n$.

$$\hat{\mathbf{x}}_{t+1/t} = \hat{\mathbf{x}}_{t+1/t}^* + \bar{\Sigma}_t(\mathbf{x}_0 - \mathbf{x}_t^*) \quad [\text{B.1.3}]$$

La expresión [B.1.3] se obtiene observando que las matrices $\bar{\Sigma}$, Γ , H , D y K_t sólo dependen del vector θ_1 y no de \mathbf{x}_0 . Por tanto, estas matrices permanecerán invariantes ante un cambio en las condiciones iniciales del vector de estado, desde \mathbf{x}_0^* a \mathbf{x}_0 .

Como la matriz B_t tampoco depende de las condiciones iniciales de \mathbf{x}_t , $\ell(\theta)$ es una función cuadrática en \mathbf{x}_0 , pudiendo expresarse como:

$$\begin{aligned} \ell(\theta_1, \mathbf{x}_0) = & \frac{1}{2} \left[\sum_{t=1}^n m \ln 2\pi + \ln |B_t| + \right. \\ & \left. + [\tilde{\mathbf{z}}_t^* - H\bar{\Sigma}_{t-1}(\mathbf{x}_0 - \mathbf{x}_0^*)]' B_t^{-1} [\tilde{\mathbf{z}}_t^* - H\bar{\Sigma}_{t-1}(\mathbf{x}_0 - \mathbf{x}_0^*)] \right] \end{aligned} \quad [\text{B.1.4}]$$

y la derivada de [B.1.4] con respecto a \mathbf{x}_0 , es una función lineal de las observaciones:

$$\frac{\partial \ell(\theta_1, \mathbf{x}_0)}{\partial \mathbf{x}_0} = \sum_{t=1}^n \tilde{\mathbf{z}}_t^*{}' B_t^{-1} \frac{\partial \tilde{\mathbf{z}}_t^*}{\partial \mathbf{x}_0} \quad [\text{B.1.5}]$$

En la expresión [B.1.5] se desconoce la derivada del proceso de innovaciones con respecto al vector \mathbf{x}_0 . Sin embargo, derivando [B.1.1] con respecto a \mathbf{x}_0 se obtiene:

$$\frac{\partial \tilde{\mathbf{z}}_t^*}{\partial \mathbf{x}_0} = - H\bar{\Sigma}_{t-1} \quad [\text{B.1.6}]$$

Sustituyendo [B.1.1] y [B.1.6] en [B.1.5] e igualando esta última expresión a cero, se obtiene el estimador máximo-verosímil de \mathbf{x}_0 , condicional al valor de cualquier otro parámetro del modelo:

$$\hat{\mathbf{x}}_0 = \left[\sum_{t=1}^n \bar{\mathbf{v}}_{t-1}' \mathbf{H}' \mathbf{B}_t^{-1} \mathbf{H} \bar{\mathbf{v}}_{t-1} \right]^{-1} \left[\sum_{t=1}^n \bar{\mathbf{v}}_{t-1}' \mathbf{H}' \mathbf{B}_t^{-1} (\tilde{\mathbf{z}}_t^* + \mathbf{H} \bar{\mathbf{v}}_{t-1} \mathbf{x}_0^*) \right] \quad [\text{B.1.7}]$$

En el caso particular en que $\mathbf{x}_0^* = \mathbf{0}$, este algoritmo es equivalente al de Rosenberg (1973).

La solución [B.1.7] es un mínimo de [B.1.4], ya que la derivada segunda de [B.1.4] con respecto a \mathbf{x}_0 :

$$\frac{\partial^2 \ell(\theta_1, \mathbf{x}_0)}{\partial \mathbf{x}_0 \partial \mathbf{x}_0'} = \sum_{t=1}^n \bar{\mathbf{v}}_{t-1}' \mathbf{H}' \mathbf{B}_t^{-1} \mathbf{H} \bar{\mathbf{v}}_{t-1} \quad [\text{B.1.8}]$$

es una matriz definida positiva al serlo también \mathbf{B}_t (o bien, su inversa).

A partir de [B.1.7], se observa que para obtener una estimación óptima del vector de estado inicial, es necesario que la matriz [B.1.8] tenga rango completo. Sin embargo, como \mathbf{B}_t^{-1} es definida positiva y el $\text{rang}(\mathbf{H}) = m$, el rango de [B.1.8] es generalmente igual a m . Esto significa que no tendrá inversa en los casos en que n sea mayor que m , que como ya se ha dicho son los más frecuentes en Econometría. En este caso, sólo serían identificables m componentes del vector de estado inicial y para obtener una solución única habría que incorporar información sobre $(n-m)$ elementos de dicho vector. Dado que el objetivo primordial es la estimación de los parámetros contenidos en θ_1 , la solución que se ha dado a este problema es utilizar una inversa generalizada de la matriz. Si $m \geq n$, todos los componentes de \mathbf{x}_0 son identificables, ya que la matriz [B.1.8] será no singular.

Por último, cabe señalar que este método evita estimar simultáneamente todos los parámetros contenidos en $[\theta_1' \mathbf{x}_0']'$. De esta forma, se puede usar el algoritmo descrito en el Capítulo 2 para minimizar $\ell(\theta)$ con respecto a θ_1 . La minimización de esta función con respecto a \mathbf{x}_0 se resuelve analíticamente utilizando la expresión [B.1.7].

APENDICE C

Ecuaciones de Riccati

De las ecuaciones del filtro de Kalman (vid. Apéndice A), se observa que $P_{t+1/t}$ satisface la siguiente ecuación de Riccati:

$$P_{t+1/t} = \Phi P_{t/t-1} \Phi' - [\Phi P_{t/t-1} H' + E S C'] [H P_{t/t-1} H' + C R C']^{-1} \cdot [\Phi P_{t/t-1} H' + E S C']' + E Q E' \quad [C.1]$$

Sin pérdida de generalidad, se puede definir $C R C' = R$, $E Q E' = Q$ y suponer que $S = 0$ ¹. En este caso, la ecuación [C.1] queda simplificada a la expresión:

$$P_{t+1/t} = \Phi P_{t/t-1} \Phi' - \Phi P_{t/t-1} H' [H P_{t/t-1} H' + R]^{-1} H P_{t/t-1} \Phi' + Q \quad [C.2]$$

Si la secuencia $\{P_{t/t-1}\}$ converge a un estado estacionario cuando $t \rightarrow \infty$, la solución en el límite (denotada por P) cumplirá la siguiente ecuación algebraica de Riccati (EAR) en tiempo discreto²:

$$P - \Phi P \Phi' + \Phi P H' [H P H' + R]^{-1} H P \Phi' - Q = 0 \quad [C.3]$$

¹ Un sistema lineal dinámico dado por las ecuaciones:

$$\begin{aligned} x_{t+1} &= \Phi x_t + \Gamma u_t + w_t \\ z_t &= H x_t + D u_t + v_t \end{aligned}$$

donde $S \neq 0$, se puede transformar en otro donde $S = 0$:

$$\begin{aligned} x_{t+1} &= \Phi^* x_t + \Gamma^* u_t + w_t^* + S R^{-1} z_t \\ z_t &= H x_t + D u_t + v_t \end{aligned}$$

siendo:

$$\begin{aligned} \Phi^* &= \Phi - S R^{-1} H & \Gamma^* &= \Gamma - S R^{-1} D & w_t^* &= w_t - S R^{-1} v_t & ; \\ Q^* &= Q - S R^{-1} S' \end{aligned}$$

² Para estudiar en detalle una teoría global sobre la ecuación de Riccati vid. Bucy (1967).

que se obtiene haciendo $P_{t+1/t} = P_{t/t-1} = P$ en la ecuación [C.2]. La ecuación matricial [C.3] representa la solución estacionaria de $P_{t+1/t}$ definida en [C.2].

Asimismo, se puede definir la matriz de transición en estado estacionario de la forma:

$$\bar{x} = x - KH$$

y la matriz ganancia en estado estacionario como:

$$K = PH'[HPH' + R]^{-1}$$

Por conveniencia, sea la factorización de las matrices R y Q :

$$R = (R^{\frac{1}{2}})(R^{\frac{1}{2}})'$$

$$Q = DD'$$

$$\text{y } \bar{H} = R^{-\frac{1}{2}}H.$$

Bajo los supuestos anteriores, la EAR puede escribirse ahora como:

$$P - PH'P' + PH'[\bar{H}P\bar{H}' + I]^{-1}HPH' - DD' = 0 \quad [C.4]$$

y se puede definir:

$$\bar{x} = x - PH'[\bar{H}P\bar{H}' + I]^{-1}\bar{H}$$

En un problema de filtrado, $P_{t/t-1}$ es la matriz de covarianzas del error de estimación de x_t . Esto implica que dicha matriz ha de ser simétrica y definida no negativa. Por tanto, sólo son interesantes las soluciones de la EAR que cumplan dichas propiedades de $P_{t/t-1}$.

Los métodos existentes para encontrar las soluciones de una EAR pueden clasificarse básicamente en dos tipos:

- métodos de cálculo directos.
- métodos de cálculo iterativos.

C.1. Métodos de cálculo directos.

En estos métodos, las soluciones de la EAR están relacionadas con el cálculo de los autovectores generalizados³ de una matriz M de orden $(2n \times 2n)$, definida como:

$$M = \begin{bmatrix} \Phi + \bar{H}'\bar{H}\Phi^{-1}Q & -\bar{H}'\bar{H}\Phi^{-1} \\ -\Phi^{-1}Q & \Phi^{-1} \end{bmatrix} \quad [C.1.1]$$

donde se supone que existe la inversa de la matriz Φ .

Particionando cada autovector generalizado (denotado por z_i) de la matriz M en dos vectores de orden n (denotados como x_i e y_i), de la forma:

$$z_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

y llamando $[X' \ Y']'$ a una selección de tales autovectores, se obtienen los siguientes resultados⁴:

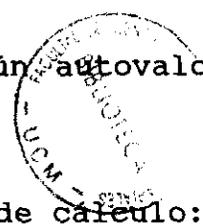
³ Sea λ_j un autovalor de la matriz M de multiplicidad p_j . Los autovectores generalizados z_{j+r} , $\forall r = 0, 1, \dots, p_j - 1$ con rango $1, 2, \dots, p_j$, asociados a λ_j han de cumplir que:

$$(M - \lambda_j I) z_j = 0$$

$$(M - \lambda_j I) z_{j+r} = z_{j+r-1} \quad (r = 1, 2, \dots, p_j - 1)$$

⁴ Puede encontrarse una demostración de los mismos en Chan et al. (1984).

- (1) Si P es una solución de la EAR, existe una selección de autovectores $[X' Y']'$ tal que $P = YX^{-1}$, siendo X no singular.
- (2) Para cada posible selección $[X' Y']'$ tal que X sea no singular, entonces $P = YX^{-1}$ es una solución de la EAR⁵.
- (3) Si M es no singular, M no tendrá ningún autovalor nulo.



Este método supone llevar a cabo dos etapas de cálculo:

- (a) Encontrar la matriz M de autovectores generalizados dada en [C.1.1].
- (b) Resolver la ecuación matricial de orden n , $P = YX^{-1}$.

En la práctica, esta aproximación no es muy utilizada por tener claras desventajas:

- Exige calcular la matriz M de orden $(2n \times 2n)$ y cuando n es mucho mayor que m , esto supone una considerable carga computacional.
- No es un método seguro en el caso de que algún autovalor de M tenga multiplicidad mayor que uno. Para resolver este problema existe otro criterio directo, consistente en calcular transformaciones ortogonales que convierten a M en una matriz en forma Schur⁶, siendo muy conveniente esta forma si se quieren evitar problemas numéricos.

⁵ Puede verse un estudio sobre la existencia y unicidad de las soluciones de una EAR en Lancaster y Rodman (1980).

⁶ Vid. Pappas, Laub y Sandell (1980).

Ambas aproximaciones son generalizables al caso en que \bar{K} sea singular o casi singular [vid. Pappas et alter (1980)]. Esto elimina los posibles problemas numéricos inherentes a otros algoritmos no recursivos, como el de Vaughn (1970), cuando la matriz \bar{K} está mal condicionada con respecto a su inversión.

C.2. Métodos de cálculo iterativos.

Algunos de estos métodos consisten en comenzar con una matriz inicial P_0 , semidefinida positiva, tal que se cumpla que:

$$\lim_{t \rightarrow \infty} P_{t+1/t} = P$$

Para obtener la solución P se puede iterar, desde la condición inicial, sobre [C.1] o sobre la versión más estable de dicha ecuación:

$$P_{t+1/t} = \bar{K}_t P_{t/t-1} \bar{K}_t' + [E \quad -K_t C] \begin{bmatrix} Q & S \\ S' & R \end{bmatrix} \begin{bmatrix} E' \\ -C' K_t' \end{bmatrix}$$

Otro método iterativo es una generalización del método de Hewer (1971) y está basado en el siguiente resultado:

Sea P_t , $\forall t = 0, 1, \dots$ la única solución semidefinida positiva de la ecuación:

$$P_t = \bar{K}_t P_t \bar{K}_t' + EQE' + K_t C R C' K_t' - K_t C S' E' - E S C' K_t' \quad [C.2.1]$$

donde:

$$K_t = [\bar{K} P_{t-1} H' + E S C'] [H P_{t-1} H' + C R C']^{-1} \quad [C.2.2]$$

Si se elige K_0 de forma que \bar{K}_0 tenga raíces en valor absoluto menores que la unidad, puede mostrarse que:

$$(1) P_0 > \dots \geq P_t \geq P_{t+1} \geq P \geq 0$$

$$(2) \mathbf{P} = \lim_{t \rightarrow \infty} \mathbf{P}_t$$

$$(3) \|\mathbf{P}_{t+1} - \mathbf{P}\| \leq \delta \|\mathbf{P}_t - \mathbf{P}\|^2 \text{ siendo } \delta \text{ una constante positiva y pequeña.}$$

Con este resultado, los pasos del proceso iterativo para solucionar la EAR son los siguientes:

- (1) Elegir \mathbf{K}_0 de forma que $\bar{\mathbf{K}}_0$ sea una matriz estable.
- (2) Calcular \mathbf{P}_0 de acuerdo con la ecuación [C.2.1].
- (3) Calcular \mathbf{K}_1 de acuerdo con la ecuación [C.2.2].
- (4) Calcular \mathbf{P}_1 de acuerdo con [C.2.1] y así sucesivamente, hasta que se cumpla un criterio de convergencia.

Este algoritmo iterativo finaliza cuando se cumple la condición $\|\mathbf{K}_{t+1} - \mathbf{K}_t\| < \eta$ siendo η una tolerancia positiva, fijada de antemano.

Cualquiera de los métodos iterativos no exigen el cálculo explícito de la inversa de \mathbf{K} , pero requieren calcular, en cada iteración, la ecuación [C.2.1]. Este cálculo será tanto más costoso cuanto mayor sea la dimensión n . Como contrapartida, no exigen el uso de rutinas que calculen los autovectores generalizados asociados a los autovalores estables de una matriz de orden $(2n \times 2n)$, o en su caso, de rutinas para el cálculo de transformaciones ortogonales.

Por último, señalar que puede verse un estudio detallado sobre las propiedades de convergencia de las soluciones de una EAR en Chan, Goodwin y Sin (1984). Estos últimos, dan un especial énfasis al estudio de dichas propiedades cuando el sistema es observable y detectable (vid. Capítulo 2), pero no necesariamente estable (en el sentido de que \mathbf{K} tenga raíces fuera del círculo unidad).

APENDICE D

Datos utilizados

A continuación, se ofrecen los datos utilizados en las estimaciones de los modelos con datos reales presentados en el Capítulo 4.

TABLA D.1: Serie del Índice Dow-Jones.

Obs.	Serie origin.	Obs.	Serie origin.
1	110.94	40	114.36
2	110.69	41	114.65
3	110.43	42	115.06
4	110.56	43	115.86
5	110.75	44	116.40
6	110.84	45	116.44
7	110.46	46	116.88
8	110.56	47	118.07
9	110.46	48	118.51
10	110.05	49	119.28
11	109.60	50	119.79
12	109.31	51	119.70
13	109.31	52	119.28
14	109.25	53	119.66
15	109.02	54	120.14
16	108.54	55	120.97
17	108.77	56	121.13
18	109.02	57	121.55
19	109.44	58	121.96
20	109.38	59	122.26
21	109.53	60	123.79
22	109.89	61	124.11
23	110.56	62	124.14
24	110.56	63	123.37
25	110.72	64	123.02
26	111.23	65	122.86
27	111.48	66	123.02
28	111.58	67	123.11
29	111.90	68	123.05
30	112.19	69	123.05
31	112.06	70	122.83
32	111.96	71	123.18
33	111.68	72	122.67
34	111.36	73	122.73
35	111.42	74	122.86
36	112.00	75	122.67
37	112.22	76	122.09
38	112.70	77	122.00
39	113.15	78	121.23

Fuente: Anderson (1976).

TABLA D.2: Serie de Desempleo Femenino en el Reino Unido.

Fecha	Serie origin.	Fecha	Serie origin.
Ene-67	96.40	Nov-69	83.30
Feb-67	104.10	Dic-69	80.10
Mar-67	102.00	Ene-70	80.50
Abr-67	100.30	Feb-70	83.00
May-67	96.40	Mar-70	82.70
Jun-67	87.30	Abr-70	81.50
Jul-67	85.10	May-70	78.30
Ago-67	87.90	Jun-70	71.60
Sep-67	89.80	Jul-70	75.40
Oct-67	96.30	Ago-70	78.60
Nov-67	99.50	Sep-70	79.80
Dic-67	95.60	Oct-70	84.80
Ene-68	97.30	Nov-70	87.50
Feb-68	97.80	Dic-70	86.50
Mar-68	94.50	Ene-71	92.00
Abr-68	90.20	Feb-71	98.80
May-68	85.60	Mar-71	104.90
Jun-68	77.50	Abr-71	104.80
Jul-68	73.40	May-71	99.20
Ago-68	76.60	Jun-71	91.70
Sep-68	76.80	Jul-71	100.60
Oct-68	85.30	Ago-71	106.10
Nov-68	85.90	Sep-71	110.30
Dic-68	82.90	Oct-71	117.00
Ene-69	84.50	Nov-71	123.10
Feb-69	84.50	Dic-71	122.90
Mar-69	82.30	Ene-72	127.50
Abr-69	78.30	Feb-72	128.40
May-69	74.60	Mar-72	129.50
Jun-69	68.70	Abr-72	131.90
Jul-69	71.90	May-72	120.10
Ago-69	74.10	Jun-72	109.10
Sep-69	74.90	Jul-72	118.30
Oct-69	82.40		

Fuente: "Unemployment Flow Statistics", Dept. of Employment Gazette, HMSO, Sept. 1972, pág. 793-795.

Unidades: Miles de personas a día 1 de cada mes.

TABLA D.3: Series de Consumo Energético (CE) y PIB de Dinamarca.

Año	CE	PIB
1951	4.49	46.69
1952	4.81	47.56
1953	5.11	50.31
1954	5.70	52.07
1955	6.09	51.87
1956	6.14	52.91
1957	5.43	55.21
1958	6.15	56.75
1959	6.06	60.64
1960	6.93	64.24
1961	7.55	68.33
1962	8.59	72.20
1963	9.42	72.67
1964	9.85	79.40
1965	10.86	83.02
1966	11.64	85.29
1967	11.81	88.42
1968	12.69	91.80
1969	14.46	97.78
1970	15.39	100.00
1971	14.80	102.45
1972	15.05	108.03
1973	15.14	112.17
1974	13.08	111.35
1975	13.43	110.24
1976	14.71	117.81
1977	14.81	120.07
1978	15.07	121.69
1979	15.43	125.39
1980	13.83	125.12

Fuente: Ramain (1986). Energy Economics.

Unidades: PIB: número índice con base 1970.

CE: millones de Tm. equivalentes de petróleo.

TABLA D.4: Series de Cantidad producida y Precios de la patata.

Fecha	q_t	p_t	Fecha	q_t	p_t
I/65	32	4.37	I/73	103	4.92
II/65	863	4.77	II/73	1000	6.70
III/65	1573	4.18	III/73	2363	5.74
IV/65	1597	4.47	IV/73	2114	5.08
I/66	66	4.68	I/74	85	6.11
II/66	818	5.63	II/74	1026	5.42
III/66	1757	4.78	III/74	2387	4.97
IV/66	1781	4.25	IV/74	2194	5.23
I/67	83	4.92	I/75	84	6.06
II/67	850	4.14	II/75	930	6.90
III/67	1733	2.63	III/75	2286	7.48
IV/67	1824	2.98	IV/75	2039	7.79
I/68	76	4.04	I/76	89	10.69
II/68	900	4.71	II/76	1041	14.64
III/68	1732	4.63	III/76	2416	9.07
IV/68	1836	4.40	IV/76	2110	8.56
I/69	76	4.90	I/77	106	9.75
II/69	794	6.82	II/77	935	10.67
III/69	1884	4.05	III/77	2651	10.43
IV/69	2033	3.72	IV/77	2187	7.05
I/70	66	4.43	I/78	85	6.76
II/70	896	4.64	II/78	1000	9.25
III/70	2206	3.65	III/78	2494	7.35
IV/70	2130	3.40	IV/78	1787	7.76
I/71	103	4.27	I/79	67	11.65
II/71	801	4.36	II/79	867	17.96
III/71	1979	3.62	III/79	2556	13.86
IV/71	1982	3.81	IV/79	2144	11.28
I/72	108	4.36	I/80	74	14.28
II/72	949	5.09	II/80	1059	10.84
III/72	2119	4.50	III/80	2704	8.49
IV/72	2098	4.44	IV/80	1901	10.06

Fuente: Grandal (1986). "Mecanismos de formación de expectativas en mercados con retardo fijo de oferta: el mercado de la patata en España. Banco de España. Servicio de Estudios. Estudios Económicos, nº 41.

Unidades: q_t : Cantidad de patata producida en miles de Tm.
 p_t : Precios de la patata percibidos por los agricultores en Pts/kg.

BIBLIOGRAFIA

Agee, W.S. y R.H. Turner (1972). "Triangular decomposition of a positive definite matrix plus a symmetric dyad with application to Kalman filtering". White Sands Missile Range Tech. Rep., n° 38.

Anderson, O.D. (1976). Time series analysis and forecasting (The Box-Jenkins approach). Butterworths, Londres.

Anderson, B.D.O. y J.B. Moore (1979). Optimal filtering. Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

Andrews, A. (1968). "A square root formulation of the Kalman covariance equations". AIAA J., n°6, pág. 1165-1166.

Ansley, C.F y R. Kohn (1983). "Exact likelihood of vector autorregresive-moving average process with missing or aggregate data". Biometrika, vol. 70, n°1, pág. 275-278.

Ansley, C.F. y R. Kohn (1985). "Estimation, filtering and smoothing in state space models with incompletely specified initial conditions". Annals of Statistics, n°13, pág. 1286-1316.

Ansley, C.F. y R. Kohn (1988). "A note on square root filtering for vector autorregresive moving-average model". Journal of Time Series Analysis, vol. 11, n°3, pág. 181-183.

Aoki, M. (1987). State Space Modelling of Time Series. Springer-Verlag. Heidelberg.

Athans, M. (1974). "The importance of Kalman filtering methods for economic systems". Annals of Economic and Social Measurement, n°3, pág. 49-64.

Bellantoni, J.F. y K.W. Dodge (1967). "Square root formulation of the Kalman-Schmidt filter". AIAA J., vol. 5, pág. 1309-1314.

Bierman, G.J. (1973). "A comparison of discrete linear filtering algorithms". IEEE Trans. Aero. Elect. Systems, AES-9, n°1, pág. 28-37.

Bierman, G.J. (1973). "Fixed interval smoothing with discrete measurements". International Journal of Control, vol. 18, n°1, pág. 65-72.

Bierman, G.J. (1977). Factorization methods for discrete sequential estimation. Academic Press, New York.

Bierman, G.J. y C.L. Thornton (1977). "Numerical comparison of Kalman filter algorithms: Orbit determination case study". Automatica, n°13, pág. 23-35.

Bierman, G.J., M.R. Belzer y J.S. Vandergraft (1990). "Maximum likelihood estimation using square root information filters". IEEE Transactions on Automatic Control, vol. 35, n°12, pág. 1293-1298.

Bittanti, S., A.J. Laub y J.C. Willems (Eds.) (1991). The Riccati Equation. Springer-Verlag, Heidelberg.

Bucy, R.S. (1967). "Global theory of the Riccati equation". J. Comput. Syst. Sci., vol. 1, pág. 349-361.

Caines, P.E. y Rissanen, J. (1974). "Maximum likelihood estimation of parameters in multivariable gaussian stochastic processes". IEEE Transactions on Information Theory, vol. IT-20, n°1, pág. 102-104.

Caines, P.E. (1988). Linear Stochastic Systems. John Wiley, New York.

Cancio, R. (1989). "Estimación máximo verosímil y predicción de modelos econométricos lineales dinámicos en espacio de los estados". Tesis Doctoral. Universidad Complutense de Madrid.

Chan, S.W., G.C. Goodwin y K.S. Sin (1984). "Convergence properties of the Riccati difference equation in optimal filtering of nonstabilizable systems". IEEE Transactions on Automatic Control, vol. AC-29, n°2, pág. 110-118.

Chin, L. (1983). "Advances in computational efficiencies of linear filtering". Control and Dynamic Systems, vol. 19, pág. 125-192.

Cramer, H. (1946). Mathematical Methods of Statistics. Princeton University Press, New Jersey.

Crowder, M.J. (1976). "Maximum likelihood estimation for dependent observations". Journal of the Royal Statistical Society, series B, vol. 38, pág. 45-53.

Demetry, J.S. (1970). "A note on the nature of optimality in the discrete Kalman filter". IEEE Transactions on Automatic Control, vol. AC-15, pág. 603-604.

Dennis, J.E. y R.B. Schnabel (1983). "Numerical methods for unconstrained optimization and nonlinear equations". Prentice-Hall Series in Computational Mathematics, New Jersey.

Dyer, P. y S.R. McReynolds (1969). "Extension of square root filtering to include process noise". J. Opt. Theory Appl., vol. 3, n°6, pág. 444-459.

Fitzgerald, R.J (1971). "Divergence of the Kalman filter". IEEE Transactions on Automatic Control, vol. AC-16, n°6, pág. 736-747.

Friedlander, B., T. Kailath, M. Morf y L. Ljung (1978). "Extended Levinson and Chandrasekhar equations for general discrete-time linear estimation problems". IEEE Transactions on Automatic Control, vol. AC-23, pág. 653-659.

Gardner, G., A.C. Harvey y G.D.A. Phillips (1980). "An algorithm for exact maximum likelihood estimation by means of Kalman filtering". Applied Statistics, nº29, pág. 311-322.

Gill, P.E. y W. Murray (1972). "Quasi-Newton methods for unconstrained optimization". J.I.M.A., vol. 9, pág. 91-108.

Grandal, M.D. (1986). "Mecanismos de formación de expectativas en mercados con retardo fijo de oferta: el mercado de la patata en España". Banco de España. Servicio de Estudios. Estudios Económicos, nº 41.

Gupta, N.K. y R.K. Mehra (1974). "Computational aspects of maximum likelihood estimation and reduction in sensitivity function calculations". IEEE Transactions on Automatic Control, vol. AC-19, nº6, pág. 774-783.

Hannan, E.J. y J. Rissanen (1982). "Recursive estimation of mixed autorregressive moving average order". Biometrika, vol. 69, pág. 81-94.

Harvey, A.C. (1981). Time Series Models. Philip Allan Publishers, Oxford.

Harvey, A.C y P.H.J. Tood (1983). "Forecasting economic time series with structural and Box-Jenkins models: A case study". Journal of Business and Economic Statistics, vol. 1, nº4, pág. 299-315.

Harvey, A.C. (1989). Forecasting, structural time series models and the Kalman filter. University Press, Cambridge.

Heijmans, R.D.H. y Magnus, J.R. (1986a). "On the first order efficiency and asymptotic normality of maximum likelihood estimators obtained from dependent observations". Statistica Neerlandica, vol. 40, nº3, pág. 169-187.

Heijmans, R.D.H. y Magnus, J.R. (1986b). "Consistent maximum likelihood estimation with dependent observations: The general (non-normal) case and the normal case". Journal of Econometrics, vol. 32, pág. 253-285.

Hendry, D.F. y J.F. Richard (1983). "The econometric analysis of economic time series". International Statistical Review, n°51, pág. 11-163.

Hewer, G.A. (1971). "An iterative technique for the computation of the steady state gains for the discrete optimal regulator". IEEE Transactions on Automatic Control, vol. AC-16, n°4, pág. 382-384.

Hillmer, S.C. y G.C. Tiao (1979). "Likelihood function of stationary multiple autoregressive moving average models". Journal of the American Statistical Association, vol. 74, n°367, pág. 652-660.

Jazwinsky, A.H. (1970). Stochastic Processes and Filtering Theory. Academic Press, New York.

Kailath, T. (1972). "Some Chandrasekhar type algorithms for quadratic regulators". Proc. 1972 IEEE Conf. Decision Control, New Orleans, pág. 219-223.

Kailath, T. (1974). "A view of three decades of linear filtering theory". IEEE Transactions of Information Theory, vol. IT-20, n°2, pág. 146-181.

Kailath, T. (1980). "Some alternatives in recursive estimation". International Journal of Control, vol. 32, n°2, pág. 311-328.

Kailath, T. (1980). Linear System Theory. Prentice-Hall, Nueva Jersey.

Kaminski, P.G. (1971). "Square root filtering and smoothing for discrete processes". Ph.D. dissertation, Dep. Aerosp. Astronaut., Stanford.

Kaminski, P.G., A.E. Bryson y S.F. Schmidt (1971). "Discrete square root filtering: A survey of current techniques". IEEE Transactions on Automatic Control, vol. AC-16, n°6, pág. 727-735.

Kaminski, P.G. y A.E. Bryson (1972). "Discrete square root smoothing" Proc. AIAA Guidance and Control Conference, paper 72-877.

Kalman, R.E. (1960). "A new approach to linear filtering and prediction problems". Transactions of the ASME, Journal of Basic Engineering, vol. 82, pág. 35-45.

Kalman, R.E. y R.S. Bucy (1961). "New results in linear filtering and prediction theory". Transactions of the ASME, Journal of Basic Engineering, pág. 95-108.

Kitagawa, G. (1981). "A nonstationary time series model and its fitting by a recursive filter". Journal of Time Series Analysis, n°2, pág. 103-116.

Kohn, R. y C.F. Ansley (1983). "Fixed interval estimation in state space models when some the data are missing or aggregated". Biometrika, vol. 70, n°3, pág. 683-688.

Lancaster, P. y L. Rodman (1980). "Existence and uniqueness theorems for the algebraic Riccati equation". International Journal of Control, vol. 32, n°2.

Laub, A.J. (1979). "A Schur method for solving algebraic Riccati equations". IEEE Transactions Automatic Control, vol. AC-24, pág. 913-921.

Laub, A.J. (1983). "Numerical aspects of solving algebraic Riccati equations". Proc. IEEE Conf. Decision and Control, pág. 184-186.

Leondes, C.T. (1970). "Theory and applications of Kalman filtering". NATO Advisory Group for Aerospace Res. Develop., AGARDograph 139.

Lindquist, A. (1974). "A new algorithm for optimal filtering of discrete time stationary processes". SIAM J. Control, vol. 12, n°4, pág. 737-746.

Ljung, L. y T. Kailath (1977). "Efficient change of initial conditions, dual Chandrasekhar equations and some applications". IEEE Transactions of Automatic Control, vol. 22, n°3, pág. 443-447.

Ljung, S. y L. Ljung (1985). "Error propagation properties of recursive least-squares adaptation algorithms". Automatica, vol. 21, pág. 158-167.

McGiffin, P.B. y D.N.P. Murthey (1980). "Parameter estimation for autorregressive systems with missing observations". International Journal of Systems Science, vol. 11, n°9, pág. 1021-1034.

Mehra, R.K. (1976). Identification and estimation of the error-in-variables model (EVM) in structural form, en R.J.B. Web (editor), Stochastic Systems: Modeling Identification and Optimization, vol.1, North-Holland, Amsterdam.

Mendel, J. (1971). "Computational requirements for a discrete Kalman filter". IEEE Transactions on Automatic Control, vol. AC-16, n°6, pág. 748-758.

Morf, M. y T. Kailath (1974). "Square root algorithms for least squares estimation". IEEE Transactions on Automatic Control, vol. AC-20, n°4, pág. 483-497.

Morf, M., G.S. Sidhu y T. Kailath (1974). "Some new algorithms for recursive estimation in constant, linear, discrete-time systems". IEEE Transactions on Automatic Control, vol. AC-19, n° 4, pág. 315-323.

Pagan, A. (1980). "Some identification and estimation results for regression models with stochastically varying coefficients". Journal of Econometrics, vol. 13, n°3, pág. 341-363.

Pappas, T., A.T. Laub y N.R. Sandell (1980). "On numerical solution of the algebraic Riccati equation". IEEE Transactions on Automatic Control, vol. AC-25, n°4, pág. 631-641.

Parzen, E. (1984). Time series analysis of irregularly observed data. Springer-Verlag, New York.

Ralston, A. y P. Rabinowitz (1978). A first course in numerical analysis. McGraw-Hill International Editions, Singapore.

Ramain, P. (1986). "The energy demand elasticity in relation to gross domestic product (A relevant indicator?)". Energy Economics, pág. 29-38.

Rosenberg, B. (1973). "The analysis of a cross section of time series by stochastically convergent parameter regression". Annals of Economic and Social Measurement, vol. 2, n°4, pág. 399-428.

Rosenbrock, H.H. (1970). State-Space and Multivariable Theory. John Wiley, New York.

Schmidt, G.T. (1976). "Linear and Nonlinear Filtering Techniques". Control and Dynamic Systems, vol. 12, pág. 63-95.

Schlee, F.H., C.J. Standish y N.F. Toda (1967). "Divergence in the Kalman filter". AIAA J., vol. 5, pág. 1114-1120.

Schweppe, F.C. (1965). "Evaluation of likelihood functions for gaussian signals". IEEE Transactions of Information Theory, vol. IT-11, n°1, pág. 61-70.

Shea, B.L. (1983). "Maximum Likelihood Estimation of Multivariate Time Series via the Kalman Filter". PhD dissertation, University College of Wales, Aberystwyth.

Shea, B.L. (1989). "The Exact Likelihood of a Vector Autoregressive Moving Average Model". Applied Statistics, vol. 38, n°1, pág. 161-204.

Silverman, L.M. (1976). "Discrete Riccati equations: alternative algorithms, asymptotic properties, and system theory interpretations". Control and Dynamic Systems, vol. 12, pág. 313-384.

Sorenson, H.W. (1976). "An overview of filtering and stochastic control in dynamic systems". Control and Dynamic Systems, vol. 12, pág. 1-56.

Tanaka, M. y T. Katayama (1987). "Robust Kalman filter for linear discrete-time system with gaussian sum noises". International Journal of Systems, vol. 18, n°9, pág. 1721-1731.

Terceiro, J. (1975). "Error analysis of the linear stochastic control problem". Proceedings of the eight Hawaii International Conference in Systems Sciences. Western Periodical Company, pág. 66-78.

Terceiro, J. (1990). Estimation of dynamic econometric models with errors in variables. Springer-Verlag, Heidelberg.

Thornton, C.L y G.J. Bierman (1980). " UDU^T covariance factorization of Kalman filtering". Control and Dynamic Systems, vol. 16, pág. 177-248.

Vaughn, D.R. (1970). "A nonrecursive algebraic solution for the discrete Riccati equation". IEEE Transactions on Automatic Control, vol. AC-15, n°5, pág. 597-599.

Verhaegen, M. y P. Van Dooren (1986). "Numerical aspects of different Kalman filter implementations". IEEE Transactions on Automatic Control, vol. AC-31, n°10, pág. 907-917.

Verhaegen, M. y P. Van Dooren (1988). "New insights in the numerical reliability properties of existing Kalman filter implementations". Control and Dynamic Systems, vol. 29, pág. 1-45.

Wald, A. (1949). "Note on the consistency of the maximum likelihood estimate". Annals of Mathematical Statistics, vol. 20, pág. 595-601.

Watanabe, K. (1985). "A note on the Kalman filter with estimated parameters". Journal of Time Series Analysis, vol. 6, n°4, pág. 269-278.

Watanabe, K. (1987). "Simple method for solving the constant gains of Kalman filters with single output". International Journal of Control, vol. 46, n°3, pág. 849-864.

Watson, M.W. y R.F. Engle (1983). "Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models". Journal of Econometrics, vol. 23, n°3, pág. 385-400.

Wertz, V., M. Gevers y E.J. Hannan (1982). "The determination of optimum structures for the state-space representation of multivariate stochastic processes". IEEE Transactions on Automatic Control, vol. AC-27, n°6, pág. 1200-1210.

Wiener, N. (1949). The Extrapolation, Interpolation and Smoothing of Stationary Time Series. John Wiley, Nueva York.

Wold, H. (1954). A study in the analysis of stationary time series. Almqvist & Wiksell, Stockholm.

Womble M.E. y J.E. Potter (1973). "A prefiltering version of the Kalman filter with new numerical integration formulas for Riccati equations". Proc. 1973 IEEE Conference Decision and Control, San Diego, pág. 63-67.

Young, P. (1984). Recursive estimation and time-series analysis. An introduction. Springer-Verlag, Heidelberg.

INDICE DE AUTORES

- Agee 3, 22, 210
- Anderson 11, 15, 129, 139,
151, 189, 191, 195,
206, 210
- Andrews 3, 22, 210
- Ansley 11, 181, 210, 215
- Aoki 42, 43, 210
- Bellantoni 2, 210
- Belzer 194, 211
- Bierman 2, 3, 11, 17,
22-24, 38, 39, 71, 161,
186, 194, 210, 211, 218
- Bittanti 211
- Box 108, 129, 150, 210, 213
- Broyden 101
- Bucy 200, 211, 215
- Caines 72, 211
- Cancio ii, 2, 211
- Chan 202, 205, 211
- Demetry 190, 212
- Dennis 84, 85, 87, 88, 99,
101, 212
- Dodge 2, 210
- Engle 71, 219
- Fitzgerald 2, 212
- Fletcher 101
- Gardner 2, 212
- Gill 85, 212
- Goldfarb 101
- Goodwin 205, 211
- Grandal 165, 168, 169, 173,
209, 213
- Harvey 44, 181, 191, 212,
213
- Heijmans 72, 213
- Hendry 213
- Hewer 204, 214
- Hillmer 169, 214
- Jazwinsky 11, 15, 190, 214
- Jenkins 108, 129, 210, 213
- Kailath 3, 24-26, 194, 212,
214-216
- Kalman 1, iii, v, viii, 2,
3, 4, 5, 11-18, 20-24,
26-29, 31, 37-39, 46,
108-127, 131, 133, 134,
135, 136, 139, 141,
142, 146, 153-155, 159,
160, 161-163, 179, 181,
183-193, 195, 200,
210-213, 215-219
- Kaminski 3, 214
- Kohn 11, 181, 210, 215
- Lancaster 203, 215
- Laub 203, 211, 215, 216
- Leondes 2, 3, 215
- Ljung 212, 215, 216
- Magnus 72, 213
- Moore 11, 15, 189, 191,
195, 210
- Morf 3, 24-26, 29, 194,
212, 216

Pagan 72, 216
 Pappas 203, 216
 Parzen 195, 216
 Potter 3, 19, 21, 22, 24,
 38, 219
 Ramain 158, 208, 217
 Richard 213
 Rissanen 72, 211, 213
 Rodman 203, 215
 Rosenberg 196, 198, 217
 Rosenbrock 43, 217
 Sandell 203, 216
 Schlee 2, 217
 Schnabel 84, 85, 87, 99,
 101, 212
 Schweppe 44, 217
 Shea 3, 217
 Sidhu 24, 26, 194, 216
 Sin 2, 8, 9, 17-19, 22, 23,
 25, 27, 33, 36-38, 42,
 44, 48, 68, 71, 72,
 80-82, 84, 87, 93, 95,
 96, 98, 100, 104, 108,
 113, 114, 129, 134,
 141, 146, 179, 186,
 189, 198-200, 205, 211
 Terceiro 1, ii, 2, 4, 10,
 43, 68, 113, 163, 181,
 196, 218
 Thornton 2, 17, 23, 25, 71,
 161, 186, 211, 218
 Tiao 169, 214
 Turner 3, 22, 210
 Van Dooren 3, 39, 218
 Vandergraft 194, 211
 Vaughn 203, 218
 Verhaegen 3, 39, 218
 Wald 72, 163, 218
 Watanabe 46, 218, 219
 Watson 71, 181, 219
 Wertz 9, 219
 Willems 211
 Young 181, 219

INDICE DE PALABRAS

- actualización 12, 13, 83,
 93, 185-189, 193
- algoritmo de Bierman 22, 39
- algoritmo de la raíz
 cuadrada 37
- algoritmo de Potter 3, 22,
 38
- aproximación por diferencias
 finitas 98, 99
- asintóticamente estable 135
- autovectores generalizados
 202, 203, 205
- backtracking iv, 79, 80,
 86, 88, 94
- condicionamiento iv, 3, 16,
 83, 85, 100, 136, 147,
 161, 173, 179, 187
- condiciones iniciales v, 2,
 5, 15, 21, 26, 30-33,
 35, 43, 51, 52, 60, 64,
 70, 82, 101, 103, 104,
 110, 133, 134, 180,
 192-198
- consistencia 72
- convergencia iv, 77, 78,
 82, 85, 87, 95, 96, 99,
 102, 205
- descomposición de Cholesky
 22, 84
- desestacionalización 181
- dimensión mínima 4, 42, 180
- dimensión no mínima 43, 180
- dirección de ascenso 81
- dirección de búsqueda viii,
 77, 79
- dirección de Cauchy iv, 81,
 82, 93, 97
- dirección de cuasi-Newton
 87, 91, 93
- dirección de descenso 81,
 82, 87, 92
- dirección de Newton iv,
 82-84, 86, 93, 97
- double dogleg 93, 100
- ecuación algebraica de
 Riccati ix, 200
- ecuación de estado 7, 9
- ecuación de Lyapunov 31, 51
- ecuación de observación
 7-10, 19, 21, 22
- ecuaciones de Chandrasekhar
 3, 4, 24, 26, 29, 37,
 43, 111, 112, 113, 136,
 179, 180
- eficiencia asintótica 73
- errores de observación vii,
 8, 43, 113, 181
- escalamiento iv, 97, 98
- espacio de los estados iii,
 2, 4, 5, 7-10, 12, 31,
 40, 41, 42-44, 72, 103,
 104, 108, 133, 146,
 151, 155, 158, 163,
 170, 180, 183, 188, 211
- estabilidad numérica 2, 4,
 18, 23, 39, 161
- estacionalidad 150
- estacionariedad 108, 150,
 151, 165, 166
- estado v, vii, ix, 7-13,
 15, 17-22, 25, 27, 28,

33, 37, 39, 42, 43, 45,
 47, 48, 69, 103, 104,
 111, 133-138, 141-144,
 151, 154-158, 160-163,
 179, 180, 183, 184-187,
 189-193, 195-201

estado estacionario ix,
 142, 158, 163, 200, 201

estimación máximo verosímil
 211

estimación recursiva 12,
 37, 193

factorización iii, 3, 4,
 18-26, 28, 29, 32, 33,
 37, 38, 70, 71, 84, 85,
 160, 180, 201

filtrado 11, 12, 181, 201

filtro de Chandrasekhar 1,
 iii, 2, 3, 5, 11, 26,
 27, 35, 36-38, 45-49,
 54, 70, 71, 103,
 109-113, 133, 134, 135,
 141, 142, 151, 154,
 155, 159, 160, 162,
 169, 171, 179, 181,
 196, 197

filtro de Kalman 1, iii, v,
 viii, 2, 3, 5, 11-18,
 20, 21, 22-24, 26-29,
 31, 37-39, 46, 108-122,
 134-136, 139, 141, 142,
 154, 155, 159-162, 179,
 181, 183-193, 195, 200

filtro de la matriz de
 información 192, 193

filtro de la raíz cuadrada
 19, 194

forma de espacio de los
 estados iii, 2, 4, 5,
 7-10, 12, 40-44, 103,
 108, 151, 155, 158, 180

forma estructural 41

forma reducida 41

forma simétrica 36

formulación impropia 10

formulación propia 10

función cuadrática 83, 86,
 198

función de autocorrelación
 parcial ix

función de autocorrelación
 simple ix

función de densidad
 condicional 44

función de densidad conjunta
 44

función de verosimilitud
 iii, iv, viii, 2, 4, 5,
 37, 44, 45, 46, 49, 54,
 68, 71, 86, 131, 132,
 136, 139, 140, 145,
 147-149, 160, 169, 172,
 173, 179, 181, 192,
 195-197

función objetivo 47-49, 76,
 78, 79, 81-83, 100,
 101, 103, 104

ganancia del filtro de
 Kalman viii, 21, 22,
 27, 31, 37, 187, 190

gradiente de la función de
 verosimilitud 49, 160

hessiano analítico 101

hessiano de la función de
 verosimilitud viii, 2,
 54, 68

heterocedasticidad dinámica
 181

hook step 92, 93, 100

incertidumbre 16, 187, 189,
 192

inestabilidad numérica 161,
 192

inicialización v, 5, 192

insesgadez asintótica 72
 interpolación óptima 181
 inversa generalizada 199
 invertibilidad 100, 111, 139, 165, 166
 line-search iv, 79, 86-88, 91, 94, 100, 110
 longitud de paso iv, ix, 77, 79, 80, 86, 88, 90, 91
 matriz de covarianzas vii, viii, 2, 13, 15, 16, 18-22, 27, 28, 31, 33, 37, 38, 48, 68, 69, 76, 134, 135, 156, 160, 161, 184-187, 189-193, 197, 201
 matriz de incrementos viii, 26, 37, 38
 matriz de información viii, 2, 4, 5, 19, 37, 68, 70, 71, 103, 110, 147, 160, 163, 169, 172, 180, 192, 193, 194
 matriz de transición 201
 matriz de varianzas-covarianzas 16, 28, 49, 54, 183
 matriz definida positiva 84, 190, 199
 matriz en forma Schur 19, 22, 24, 38, 203
 mínimos cuadrados con retrovisión 108, 109, 111-113, 131, 139, 146, 154, 159, 160
 mínimos cuadrados ordinarios ix
 mínimos cuadrados ponderados iv, 76, 91
 modelos ARMA univariantes 108
 modelos de función de transferencia 5, 30-33, 35, 39, 46, 48, 49, 51, 54, 60, 108, 133, 134
 modelos estacionarios v, 5, 12, 191
 modelos gaussianos v, 108, 110, 113, 165
 modelos multivariantes 4, 32, 35, 45, 47, 52, 62, 72, 73, 191
 normalidad asintótica 72, 73
 número de condición 18, 39, 85
 optimalidad 17, 37, 190
 optimización iv, v, 5, 46, 48, 71, 76, 78, 79, 86, 98, 100, 102, 110, 135, 136, 173
 parámetros cambiantes 181
 predicción 11-13, 17, 45, 76, 185-188, 197, 211
 proceso de innovaciones vii, 13, 23, 27, 48, 49, 54, 68, 69, 71, 186, 197, 198
 proceso de ruido blanco 145, 158
 propagación 3, 15, 21, 32, 33, 37, 39, 60, 71, 135, 142, 155, 161, 187, 192, 193
 región de confianza 93, 94, 97, 102
 restricciones lineales 163

retroprevisión 108, 109,
111-113, 115-122, 131,
139, 146, 153, 154,
159, 160, 195

simulación 111-127

sistema dinámico 7, 8

sistema dinámico 11, 181,
210, 214, 219

test de Wald 163

test del multiplicador de
Lagrange ix, 71, 163

tolerancia 95-97, 102, 205

transformación Box-Cox 150

transformaciones ortogonales
23, 203, 205