ESTADÍSTICA ESPAÑOLA
Vol. 37, Núm. 140, 1997, págs. 363 a 380

Una nota sobre la estimacion eficiente de modelos con parametros cambiantes

por
SONIA SOTOCA LÓPEZ (1)
Departamento de Economía Cuantitativa
Fac. de CC. Económicas y Empresariales
Universidad Complutense de Madrid

RESUMEN

Los procedimientos estándar para estimar modelos de parámetros cambiantes suponen conocidas las varianzas de los términos de error presentes en el modelo. Obviamente, éste no es un supuesto realista en la mayor parte de las aplicaciones econométricas. Por otra parte, los resultados que proporcionan estos métodos son sensibles a las condiciones iniciales, hecho que habitualmente es ignorado por la literatura. En este trabajo se propone una extensión del algoritmo recursivo debido a Cooley, Rosenberg y Wall (1977), que también es independiente de condiciones iniciales e incorpora la estimación *on-line* de todas las varianzas relevantes. Los resultados obtenidos con este procedimiento se comparan favorablemente con los obtenidos usando los métodos habituales.

Palabras clave: condiciones iniciales, modelos de parámetros cambiantes, filtro de información, filtro de Kalman, ratio de varianzas, *smoother* de intervalo fijo.

Clasificación AMS: 62J05, 62J07.

---

(1) Quiero agradecer las sugerencias recibidas de Rafael Flores, Miguel Jerez y Alfonso Novales. Los comentarios de dos evaluadores anónimos han ayudado a detectar errores y por tanto, a mejorar considerablemente la primera versión de este trabajo.

INTRODUCCIÓN

En ocasiones, una estructura lineal de parámetros fijos no permite modelizar adecuadamente determinadas relaciones causales. Esto puede deberse a diversos factores, como son cambios en la estructura de la relación, existencia de no linealidades, así como el uso de una especificación incorrecta, que puede provocar variación en los parámetros. En este contexto, diversos autores como Cooley y Prescott (1973, 1976), Pagan (1980) y Young (1984), proponen usar modelos de parámetros cambiantes, que pueden reducirse al siguiente modelo en forma de espacio de los estados:

$$y_t = x_t^T \beta_t + \varepsilon_t$$

$$\beta_t = \Phi \beta_{t-1} + u_t$$

donde la ley de evolución del vector β_t es la ecuación de estado del sistema y ε_t y $\boldsymbol{u_t}$ son variables aleatorias independiente e idénticamente distribuidas con esperanza nula y varianzas σ^2 y $\boldsymbol{Q}$, respectivamente. Estas varianzas suelen ser desconocidas y el tamaño relativo de las mismas indica lo cerca que estamos de un modelo de parámetros fijos o de un modelo de parámetros cambiante matriz Φ y modificando el vector de estado como sea necesario, pueden formularse distintos modelos, como es el modelo de constante adaptativa de Cooley y Prescott (1973) o los modelos ARCH [ver Wolff (1988)].

Este trabajo se centra en la estimación de una clase de modelos de parámetros cambiantes. Para estimar éstos, suele utilizarse un algoritmo en dos etapas en el que se aplican, sucesivamente, el filtro de Kalman y el *smoother* óptimo de intervalo fijo [ver Harvey (1989), cap.3]. Los resultados de este algoritmo (que, en adelante denominaremos FK-SIF) dependen crucialmente de disponer de estimaciones "adecuadas" del vector inicial de coeficientes y de las varianzas de los términos de error presentes en el modelo o, al menos, de la proporción entre ellas.

El primer problema(2) fue abordado por Cooley, Rosenberg y Wall (1977), quienes propusieron un algoritmo (a partir de ahora, algoritmo CRW) independiente de condiciones iniciales, pero que requiere conocer las varianzas de las perturbaciones que intervienen en el modelo, lo que evidentemente no es realista en la mayoría de las situaciones reales. Una posible solución a este problema es estimar el modelo por máxima verosimilitud mediante el filtro de Kalman [ver Cooley y Prescott (1976)]. Sin embargo, la experiencia práctica sugiere que: a) evaluar la función de verosimilitud de este tipo de modelos resulta costoso computacionalmente y b)

(2) Para una discusión detallada, ver Sotoca (1993).

el perfil de dicha función suele ser bastante plano en un entorno del máximo [ver García-Ferrer et al. (1993)]. Estos problemas han dado lugar a que algunos autores [ver Young (1984), cap. 5] sugieran la utilización de un ratio de varianzas arbitrario o “manual”. Sin embargo, nuestra experiencia indica que dicha arbitrariedad afecta tanto a la eficiencia de los estimadores del resto de los parámetros, como a la secuencia de estimaciones.

En este trabajo se propone una extensión del algoritmo CRW (a partir de ahora CRW1) que incorpora una estimación recursiva de dichas varianzas, además de la trayectoria óptima de los parámetros propios del modelo. Esta nueva versión también es independiente de condiciones iniciales, al estar basado en un filtro de información en lugar del filtro de Kalman.

La estructura del trabajo es la siguiente. En el apartado 1 se plantea el modelo objeto de estudio y el procedimiento de estimación CRW, poniendo especial énfasis en las ventajas de este algoritmo con respecto a otros alternativos.

En el apartado 2 se describe la forma de incorporar en el algoritmo CRW la estimación recursiva de las varianzas de los términos de error del modelo, dando lugar al criterio llamado CRW1. En el apartado 3 se presenta el algoritmo FK-SIF, siendo también posible incorporar en el mismo la estimación *on-line* de las varianzas de los ruidos, aunque este nuevo criterio, denotado por FK-SIF1, no es independiente de condiciones iniciales.

En el apartado 4 se presentan los resultados obtenidos con este procedimiento usando datos simulados. Estos resultados se comparan favorablemente con los obtenidos mediante los criterios recursivos FK-SIF1 y CRW, siendo éste último el punto de referencia para comparar, dado que en el mismo se fijan los verdaderos valores de todas las varianzas relevantes. Además, y a efectos de comparación, se ofrecen los resultados obtenidos con los mismos datos mediante el procedimiento de máxima verosimilitud, técnica no recursiva aunque use el filtro de Kalman en la evaluación de la función de verosimilitud.

Finalmente, en el apartado 5 se resumen las principales conclusiones del trabajo.

1. EL ALGORITMO CRW

Sea el siguiente modelo de regresión con parámetros cambiantes en el tiempo:

$$y_t = x_t^T \beta_t + \varepsilon_t \qquad [1]$$

$$\beta_{t+1} = \beta_t + u_t \qquad [2]$$

donde y_t y el vector x^T_t representan las variables observables del sistema, β_t es un vector de k parámetros desconocidos y las perturbaciones ε_t y u_t son variables aleatorias independiente e idénticamente distribuidas con esperanza nula y varianzas σ^2 y Q, respectivamente. El objetivo es estimar eficientemente el vector β_t a partir de las observaciones $\{y_1, y_2, ..., y_N\}$.

El algoritmo CRW proporciona una solución al problema de estimación óptima del modelo dado por (1) y (2), conocidas las varianzas σ^2 y Q. Este algoritmo se basa en la combinación de dos filtros de información(3), lo que da lugar a un procedimiento recursivo de estimación independiente del vector inicial de parámetros.

Denotando las matrices de información en las etapas de predicción y actualización por $H_{t/t-1}$ y $H_{t/t}$, respectivamente, las variables auxiliares que juegan el papel de los parámetros en un filtro de información se definen como:

$$f_{t/t-1} = H_{t/t-1}\hat{\beta}_{t/t-1} \;;\; f_{t/t} = H_{t/t}\hat{\beta}_{t/t} \qquad [3]$$

y las correspondientes matrices de covarianzas del vector $\hat{\beta}_t$ se definen como $P_{t/t-1} = H^{-1}_{t/t-1}$ y $P_{t/t} = H^{-1}_{t/t}$. El primer paso del algoritmo CRW consiste en propagar un filtro de información que recorre la muestra hacia adelante a partir de las condiciones iniciales $H_{1/0} = 0$ y $f_{1/0} = 0$. Con esta inicialización, las ecuaciones de la fase de predicción de un filtro de información aplicado al modelo (1)-(2) son:

$$K_{t-1} = [I + H_{t-1/t-1}Q]^{-1} \qquad [4]$$

$$H_{t/t-1} = K_{t-1}H_{t-1/t-1} \qquad [5]$$

$$f_{t/t-1} = K_{t-1}f_{t-1/t-1} \qquad [6]$$

donde K_t es la ganancia del filtro que se propaga hacia adelante. Cuando un nuevo dato está disponible, las predicciones anteriores se corrigen de la siguiente manera:

$$f_{t/t} = f_{t/t-1} + \frac{x_t y_t}{\sigma^2} \qquad [7]$$

$$H_{t/t} = H_{t/t-1} + \frac{x_t x_t^T}{\sigma^2} \qquad [8]$$

(3 En un filtro de información se propagan las inversas de las matrices de covarianzas en lugar de las propias matrices. Esto hace que el algoritmo disponga de condiciones iniciales exactas y que las variables del filtro se actualizen de forma estable a partir de su inicialización [ver Anderson y Moore (1979), cap. 6].

Evidentemente, los resultados de las ecuaciones (4)-(6) pueden sustituirse en (7)-(8), dando lugar a una única fase de propagación, lo que resulta ventajoso desde un punto de vista computacional.

El siguiente paso consiste en propagar un segundo filtro de información que recorre la muestra hacia atrás, es decir, en el orden $t = N, N-1, \ldots 1$. Si denotamos por $\boldsymbol{G}_{t/t+1}$ y $\boldsymbol{G}_{t/t}$ las matrices de información en las etapas de predicción y actualización de este nuevo filtro y las correspondientes variables auxiliares como $\boldsymbol{r}_{t/t+1}$ y $\boldsymbol{r}_{t/t}$, respectivamente, la inicialización del filtro es de nuevo $G_{N/N+1} = 0$ y $r_{N/N+1} = 0$. La fase de predicción de este segundo filtro viene dada por:

$$J_{t+1} = [I + G_{t+1/t+1}Q]^{-1} \tag{9}$$

$$G_{t/t+1} = J_{t+1}G_{t+1/t+1} \tag{10}$$

$$r_{t/t+1} = J_{t+1}r_{t+1/t+1} \tag{11}$$

siendo $\boldsymbol{J}_{t+1}$ la matriz ganancia del filtro propagado hacia atrás y la etapa de actualización:

$$r_{t/t} = r_{t/t+1} + \frac{x_t y_t}{\sigma^2} \tag{12}$$

$$G_{t/t} = G_{t/t+1} + \frac{x_t x_t^T}{\sigma^2} \tag{13}$$

Comparando las ecuaciones (4)-(8) con (9)-(13), se observa que:

1. La estructura general de ambos filtros es idéntica.

2. Ambos algoritmos son independientes; esto es, para aplicar uno de ellos no es necesario conocer los resultados del otro(4)

3. El algoritmo no requiere invertir la matriz de covarianzas $\boldsymbol{Q}$, como ocurre con algunas versiones de este tipo de filtros [ver Anderson y Moore (1979), cap.6]. Esto es importante ya que, en la práctica, es frecuente que dicha matriz sea singular(5).

4. A la vista de las ecuaciones (4)-(13), es fácil demostrar que el algoritmo puede escribirse en función de ratios de varianzas, sin más que redefinir las variables auxiliares $\boldsymbol{f}$ y $\boldsymbol{r}$ como $\sigma^2\boldsymbol{f}$ y $\sigma^2\boldsymbol{r}$, respectivamente.

(4) Esta propiedad es de gran interés computacional, ya que permite aplicar procesos de cálculo en paralelo

(5) Concretamente, cuando alguno de los parámetros del modelo se suponga constante en el tiempo. Ver, por ejemplo, Cooley y Prescott (1973)

El *smoothing* óptimo en el algoritmo CRW, se obtiene mediante una combinación lineal de los resultados de ambos filtros. Concretamente, la trayectoria óptima del vector β_t y sus correspondientes matrices de covarianzas pueden calcularse de la forma [ver Fraser y Potter (1969)]:

$$P_{t/N} = [P_{t/t}^{-1} + P_{t/t+1}^{-1}]^{-1} = [H_{t/t} + G_{t/t+1}]^{-1} \qquad [14]$$

$$b_{t/N} = P_{t/N}[P_{t/t}^{-1}\hat{\beta}_{t/t} + P_{t/t+1}^{-1}\hat{\beta}_{t/t+1}] = P_{t/N}[f_{t/t} + r_{t/t+1}] \qquad [15]$$

La interpretación del *smoother* lineal óptimo como la combinación de dos filtros lineales también óptimos es sencilla. El filtro procesado desde el principio de la muestra hasta algún instante *t* dentro del intervalo muestral, proporciona la mejor estimación del vector de parámetros en *t*, contando con toda la información desde el principio del intervalo hasta el momento *t*. Por otro lado, el filtro procesado desde el final de la muestra hacia atrás hasta un instante *t*, proporciona la mejor estimación del vector de parámetros en *t*, basándose en la información desde el momento *t* hasta el final del intervalo muestral. Por tanto, estos dos filtros unidos utilizan precisamente toda la información disponible. Si los ruidos del modelo son procesos de ruido blanco, las dos estimaciones estarán incorrelacionadas entre sí. Este hecho hace que puedan combinarse adecuadamente dos estimaciones óptimas, para obtener el *smoother* que utiliza toda la muestra disponible. Las expresiones (14) y (15) son las conocidas ecuaciones que combinan de forma óptima dos estimaciones independientes, donde $\boldsymbol{b}_{t/N}$ es tanto el estimador de máxima verosimilitud como el de mínima varianza [ver Fraser y Potter (1969)].

2. ESTIMACIÓN RECURSIVA DE LAS VARIANZAS DEL MODELO: EL ALGORITMO CRW1

El principal inconveniente del algoritmo CRW es que supone conocidas las varianzas σ^2 y $\boldsymbol{Q}$, lo que no es habitual en la práctica econométrica con datos reales. Por otra parte, la elección arbitraria del valor de estas varianzas influye en todo el proceso de estimación. En concreto, cuanto mayor sea σ^2 con respecto a las varianzas asociadas a la ley de variación de los parámetros, más cerca estaremos de un modelo de parámetros fijos. En el caso contrario, permitimos mucha más variabilidad al parámetro(s) cambiante(s), pudiendo llegar a que las fluctuaciones de la variable a explicar coincidan prácticamente con las fluctuaciones de los parámetros.

Siguiendo a Margaritis (1990), una forma de estimar recursivamente dichas varianzas es:

$$\sigma^2_{t/t} = \sigma^2_{t/t-1} + \frac{\tilde{z}^2_{t/t-1} - \sigma^2_{t/t-1}}{t} \qquad [16]$$

$$Q_{t/t} = Q_{t/t-1} + \frac{\tilde{\beta}_t \tilde{\beta}_t^T - Q_{t/t-1}}{t} \qquad [17]$$

siendo

$$\tilde{z}_{t/t-1} = y_t - x_t^T \hat{\beta}_{t/t-1} \qquad [18]$$

$$\tilde{\beta}_t = \hat{\beta}_{t/t} - \hat{\beta}_{t/t-1} \qquad [19]$$

Es decir, (18) define el error de predicción un período hacia adelante de la variable y_t y (19) la diferencia entre el vector de estimaciones de β en la etapa de actualización y predicción del filtro. Una simple observación de (16) y (17) muestra que ambas ecuaciones son independientes de la inicialización de la que se parta. De hecho, dado cualquier valor inicial $\sigma^2_{0/-1}$, cuando es procesado el primer dato, la estimación de la varianza de la perturbación del modelo es $\sigma^2_{1/1} = \tilde{z}^2_{1/0} = (y_1 - x_1^T \hat{\beta}_{1/0})^2$ y una vez procesadas las N observaciones disponibles:

$$\sigma^2_{N/N} = \frac{\sum_{t=1}^{N} \tilde{z}^2_{t/t-1}}{N} \qquad [20]$$

Es decir, (16) indica que un estimador de la varianza de la perturbación del modelo es la varianza muestral de los errores de predicción de y_t un período hacia adelante. Una interpretación similar tiene la ecuación (17), que estima recursivamente la matriz de covarianzas (***Q***) del vector de perturbaciones ($\boldsymbol{u}_t$) asociadas a la ley de evolución de los parámetros β. La expresión (20) es similar al estimador máximo-verosímil de la varianza de las perturbaciones de un modelo de parámetros fijos. Por tanto, dicho estimador será sesgado aunque consistente.

La idea fundamental del trabajo es incorporar la estimación recursiva de las varianzas σ^2 y ***Q*** en el algoritmo de CRW, con objeto de combinar de forma óptima la secuencia de estimaciones de las varianzas obtenidas en los dos filtros. Con esta idea, el nuevo filtro (que llamamos CRW1) necesita también condiciones iniciales de las varianzas(6): $\sigma^2_{1/0} = 1$ y $Q_{1/0}=0$. La etapa de predicción del filtro que recorre

(6) La condición inicial de σ^2 puede ser arbitraria, pero distinta de cero si queremos propagar el filtro en función de las variables $\boldsymbol{f}$ y $\boldsymbol{r}$ y no en función de $\sigma^2\boldsymbol{f}$ y $\sigma^2\boldsymbol{r}$ [ver ecuaciones (21) y (22)]

la muestra hacia adelante no cambia con respecto al criterio CRW, mientras que la etapa de actualización pasa a ser:

$$f_{t/t} = f_{t/t-1} + \frac{x_t y_t}{\sigma^2_{t/t-1}} \quad [21]$$

$$H_{t/t} = H_{t/t-1} + \frac{x_t x_t^T}{\sigma^2_{t/t-1}} \quad [22]$$

$$\sigma^2_{t/t} = \sigma^2_{t/t-1} + \frac{(y_t - x_t^T H^{-1}_{t/t-1} f_{t/t-1})^2 - \sigma^2_{t/t-1}}{t} \quad [23]$$

$$Q_{t/t} = Q_{t/t-1} + \frac{[(H^{-1}_{t/t} f_{t/t} - H^{-1}_{t/t-1} f_{t/t-1})(H^{-1}_{t/t} f_{t/t} - H^{-1}_{t/t-1} f_{t/t-1})^T] - Q_{t/t-1}}{t} \quad [24]$$

Las ecuaciones de predicción del filtro que recorre la muestra hacia atrás tampoco cambian con respecto al procedimiento CRW, siendo la condición inicial de las varianzas $\sigma^2_{N/N+1} = 1$ y $Q_{N/N+1} = 0$ y la fase de actualización:

$$r_{t/t} = r_{t/t+1} + \frac{x_t y_t}{\sigma^2_{t/t+1}} \quad [25]$$

$$G_{t/t} = G_{t/t+1} + \frac{x_t x_t^T}{\sigma^2_{t/t+1}} \quad [26]$$

$$\sigma^2_{t/t} = \sigma^2_{t/t+1} + \frac{(y_t - x_t^T G^{-1}_{t/t+1} r_{t/t+1})^2 - \sigma^2_{t/t+1}}{t} \quad [27]$$

$$Q_{t/t} = Q_{t/t+1} + \frac{[(G^{-1}_{t/t} r_{t/t} - G^{-1}_{t/t+1} r_{t/t+1})(G^{-1}_{t/t} r_{t/t} - G^{-1}_{t/t+1} r_{t/t+1})^T] - Q_{t/t+1}}{t} \quad [28]$$

El *smoother* óptimo se obtiene ponderando las secuencias de estimaciones obtenidas tanto de los parámetros β de acuerdo con (14)-(15), como las estimaciones de σ^2 y $\boldsymbol{Q}$:

$$\sigma^2_{t/N} = \frac{(x_t^T H^{-1}_{t/t} x_t)^{-1} \sigma^2_{t/t} + (x_t^T G^{-1}_{t/t+1} x_t)^{-1} \sigma^2_{t/t+1}}{(x_t^T H^{-1}_{t/t} x_t)^{-1} + (x_t^T G^{-1}_{t/t+1} x_t)^{-1}} \quad [29]$$

$$Q_{t/N} = [H_{t/t} + G_{t/t+1}]^{-1}[H_{t/t}Q_{t/t} + G_{t/t+1}Q_{t/t+1}] \qquad [30]$$

Las ecuaciones (29) y (30) muestran que las estimaciones de las varianzas σ^2 y $\boldsymbol{Q}$ se combinan linealmente ponderando la precisión relativa asociada a la estimación obtenida en la propagación de los dos filtros.

3. ESTIMACIÓN RECURSIVA DE LAS VARIANZAS DEL MODELO: EL ALGORITMO FK-SIF1

Usando las expresiones (16) y (17), también es posible incorporar una estimación *on-line* de las varianzas de los ruidos presentes en el algoritmo FK-SIF, que se denotará por FK-SIF1. Sin embargo, esta nueva versión del criterio FK-SIF no es posible independizarla de las condiciones iniciales de los parámetros propios del modelo.

El algoritmo FK-SIF consta de dos etapas. En la primera, se propaga el filtro de Kalman (FK) para el modelo (1)-(2) empezando en $t = 1,2,...N$ y con condiciones iniciales arbitrarias del tipo $\hat{\beta}_{1/0} = 0$ y $P_{1/0} = \tau I$, donde τ es un escalar positivo y arbitrariamente grande. La etapa de predicción viene dada por:

$$\hat{\beta}_{t/t-1} = \hat{\beta}_{t-1/t-1} \qquad [31]$$

$$P_{t/t-1} = P_{t-1/t-1} + Q \qquad [32]$$

y la etapa de actualización:

$$K_t = P_{t/t-1}x_t(x_t^T P_{t/t-1}x_t + \sigma^2)^{-1} \qquad [33]$$

$$\tilde{z}_{t/t-1} = y_t - x_t^T \hat{\beta}_{t/t-1}. \qquad [34]$$

$$\hat{\beta}_{t/t} = \hat{\beta}_{t/t-1} + K_t \tilde{z}_t \qquad [35]$$

$$P_{t/t} = (I - K_t x_t^T)P_{t/t-1} \qquad [36]$$

En la segunda etapa, se propaga un *smoother* de intervalo fijo empezando en el instante $t = N-1, N-2, \ldots 1$, con los resultados obtenidos del filtro de Kalman anterior:

$$\hat{\beta}_{t/N} = \hat{\beta}_{t/t} + P_{t/t}P_{t+1/t}^{-1}(\hat{\beta}_{t+1/N} - \hat{\beta}_{t/t}) \qquad [37]$$

$$P_{t/N} = P_{t/t} + P_{t/t}P^{-1}_{t+1/t}(P_{t+1/N} - P_{t+1/t})P^{-1}_{t+1/t}P_{t/t} \qquad [38]$$

El algoritmo FK-SIF1 viene dado por las ecuaciones (31)-(38) incorporando en la etapa de filtrado las siguientes fórmulas recursivas para las varianzas relevantes del modelo:

$$\sigma_t^2 = \sigma_{t-1}^2 + \frac{(\tilde{z}^2_{t/t-1} - \sigma^2_{t-1})}{t} \qquad [39]$$

$$Q_{t/t} = Q_{t-1/t-1} + \frac{[(K_t\tilde{z}_{t/t-1})(K_t\tilde{z}_{t/t-1})^T - Q_{t-1/t-1}]}{t} \qquad [40]$$

Como puede apreciarse, las diferencias entre el criterio CRW1 y el FK-SIF1 son fundamentalmente dos: i) el algoritmo FK-SIF1 consta de dos etapas que no son independientes entre sí, es decir, la condición inicial del *smoother* es el resultado de la etapa de actualización del filtro de Kalman y ii) mientras que el algoritmo CRW1 es independiente de condiciones iniciales, el FK-SIF1 no lo es y se verá cómo la inicialización arbitraria afecta a la estimación recursiva de las varianzas de las perturbaciones.

4. RESULTADOS CON DATOS SIMULADOS

Para validar el algoritmo descrito en el apartado anterior, se ha aplicado a la estimación de un modelo de regresión con un parámetro cambiante usando datos simulados. El experimento ha consistido en estimar para distintos tamaños muestrales (en concreto, para 100, 200, 500 y 1000 observaciones) la siguiente especificación:

$$y_t = \alpha_t + \beta x_t + \varepsilon_t \qquad [41]$$

$$\alpha_t = \phi\alpha_{t-1} + u_t \qquad [42]$$

donde $\beta = 0.5$, el parámetro autorregresivo ϕ toma los valores 1, 0.95 y 0.5, el valor inicial de α_t es cero y las perturbaciones ε_t y u_t se han generado como variables aleatorias independientes e idénticamente distribuidas como normales con los siguientes momentos: ε_t .~ N(0,9) y $u_t \sim$ N(0,1). La variable explicativa x_t también ha sido generada como una variable aleatoria independiente de ε_t y con distribución *N*(0, 25).

Por tanto, se han simulado tres modelos de acuerdo con la especificación dada por (41) y (42). El modelo I (ϕ = 1) es el modelo con constante adaptativa de Cooley y Prescott (1973). El modelo II supone que la ley de variación del parámetro está cerca de la no estacionariedad (ϕ = 0.95) y en el modelo III, la constante sigue un proceso AR(1) bien condicionado al fijar ϕ = 0.5. La razón de generar modelos donde el parámetro cambiante sigue una ley de evolución estacionaria y no estacionaria, es comprobar si este supuesto afecta a la estimación puntual del parámetro constante β, de la varianza de la perturbación del modelo, de la varianza del error asociado al parámetro cambiante y, sobre todo, a la precisión con que son estimados dichos coeficientes.

Los modelos I, II y III han sido estimados utilizando cuatro procedimientos: (1) el algoritmo CRW, considerando conocidos los valores teóricos de las varianzas σ_ε^2 y σ_u^2; (2) el procedimiento no recursivo de máxima verosimilitud usando el filtro de Kalman dado por (31)-(36) para evaluar la función de verosimilitud exacta(7) del modelo (41)-(42); (3) el algoritmo CRW1 y (4) el algoritmo FK-SIF1 eligiendo las siguientes condiciones iniciales para el vector de parámetros y su matriz de covarianzas: $\alpha_{1/0} = 0$, $\beta_{1/0} = 0$ y $\boldsymbol{P}_{1/0} = \tau \boldsymbol{I}$ con $\tau = 10^6$. La razón de utilizar este último criterio es mostrar la importancia de las condiciones iniciales de los parámetros α y β en la estimación recursiva de las varianzas σ_ε^2 y σ_u^2. De hecho, a medida que τ tiende a ser un número "muy" grande, la estimación final de las varianzas se acerca más al verdadero valor, pero el filtro se degrada numéricamente(8). Este hecho muestra que es necesario utilizar un *smoother* de tipo información para hacer independiente la estimación final de la inicialización.

Los resultados de este primer experimento se muestran en las Tablas 1, 2 y 3. Los modelos considerados han sido estimados usando distintos tamaños muestrales y se ofrece la media de los resultados obtenidos con 500 realizaciones. Los errores estándar de las estimaciones de los parámetros β, σ_ε^2 y σ_u^2, que se presentan entre paréntesis, se han calculado a partir de las varianzas muestrales.

---

(7) Bajo el supuesto de que los ruidos ε_t y u_t y el vector de parámetros inicial ($\hat{\beta}_{0/0}$) siguen distribuciones normales, la función de verosimilitud exacta de (41)-(42) en logaritmos (dejando aparte la constante) es:

$$L = -\frac{1}{2}\sum_{t=1}^{N} \ln b_t - \frac{1}{2}\sum_{t=1}^{N} \frac{\tilde{z}_{t/t-1}^2}{b_t}$$

donde $b_t = x_t^T P_{t/t-1} x_t + \sigma_\varepsilon^2$.

(8) Se ha elegido un valor de $\tau = 10^6$, porque es el habitualmente usado en los algoritmos de estimación recursiva de modelos con parámetros fijos [ver Young (1984)]

Tabla 1

RESULTADOS DE LA SIMULACIÓN DEL MODELO I: $y_t = \alpha_t + \beta_t + \varepsilon_t$ CON $\alpha_t = \alpha_{t-1} + u_t$. VALORES TEÓRICOS DE LOS PARÁMETROS: $\beta = 0.5$, $\sigma_\varepsilon^2 = 9$, $\sigma_u^2 = 1$

N	Algoritmo CRW	Máxima Verosimilitud con el FK			Algoritmo CRW1			Algoritmo FK-SIF1 con $\tau = 10^6$		
	$\hat\beta$	$\hat\beta$	$\hat\sigma_\varepsilon^2$	$\hat\sigma_u^2$	$\hat\beta$	$\hat\sigma_\varepsilon^2$	$\hat\sigma_u^2$	$\hat\beta$	$\hat\sigma_\varepsilon^2$	$\hat\sigma_u^2$
100	0.5050	0.5070	15.385	1.0603	0.5006	15.409	3.4759	0.4854	17.884	0.6935
	(0.1529)	(0.1620)	(2.7871)	(0.5571)	(0.1538)	(3.1062)	(2.5385)	(0.1668)	(9.8478)	(0.4633)
200	0.4914	0.4971	15.159	1.0124	0.4987	14.468	2.5771	0.5047	15.286	0.5441
	(0.1088)	(0.1166)	(1.8190)	(0.4259)	(0.1130)	(1.9638)	(1.6053)	(0.1108)	(4.3740)	(0.3258)
500	0.5011	0.5043	15.362	1.0203	0.4994	15.018	1.8350	0.5037	15.059	0.4101
	(0.0683)	(0.0771)	(1.1265)	(0.2605)	(0.0708)	(1.7763)	(0.8125)	(0.0714)	(5.0561)	(0.2846)
1000	0.5007	0.5017	15.165	1.0094	0.4991	13.136	1.3838	0.5003	14.395	0.3303
	(0.0493)	(0.0448)	(0.7555)	(0.1568)	(0.0472)	(0.6861)	(0.6806)	(0.0475)	(4.9476)	(0.1584)

Notas: N: Tamaño muestral utilizado. Las cifras entre paréntesis representan las desviaciones típicas muestrales.

Tabla 2

RESULTADOS DE LA SIMULACIÓN DEL MODELO II: $y_t = \alpha_t + \beta_t + \varepsilon_t$ CON $\alpha_t = 0.95\alpha_{t-1} + u_t$. VALORES TEÓRICOS DE LOS PARÁMETROS: $\beta = 0.5$, $\sigma_\varepsilon^2 = 9$, $\sigma_u^2 = 1$

N	Algorltmo CRW	Máxima Verosimilitud con el FK			Algoritmo CRW1			Algoritmo FK-SIF1 con $\tau = 10^6$		
	$\hat{\beta}$	$\hat{\beta}$	$\hat{\sigma}_\varepsilon^2$	$\hat{\sigma}_u^2$	$\hat{\beta}$	$\hat{\sigma}_\varepsilon^2$	$\hat{\sigma}_u^2$	$\hat{\beta}$	$\hat{\sigma}_\varepsilon^2$	$\hat{\sigma}_u^2$
100	0.5055	0.5165	15.408	1.0026	0.5022	14.677	2.3553	0.4939	15.162	0.3812
	(0.0806)	(0.0926)	(2.7701)	(0.5283)	(0.0886)	(6.6303)	(8.4771)	(0.0885)	(3.5969)	(0.3602)
200	0.5022	0.5105	15.461	0.9296	0.4989	14.343	1.8095	0.4999	15.020	0.2807
	(0.0652)	(0.0648)	(2.1142)	(0.4203)	(0.0689)	(5.5061)	(4.7114)	(0.0702)	(2.8747)	(0.2821)
500	0.5012	0.5040	15.320	1.0778	0.4978	13.169	1.1071	0.5021	14.617	0.1730
	(0.0391)	(0.0394)	(1.1685)	(0.2714)	(0.0393)	(3.2722)	(1.3058)	(0.0387)	(2.8620)	(0.1868)

Notas: N: Tamaño muestral utilizado. Las cifras entre paréntesis representan las desviaciones típicas muestrales.

Tabla 3

RESULTADOS DE LA SIMULACIÓN DEL MODELO III: $y_t = \alpha_t + \beta_t + \varepsilon_t$ CON. $\alpha_t = 0.5\alpha_{t-1} + u_t$ VALORES TEÓRICOS DE LOS PARÁMETROS: $\beta = 0.5$, $\sigma_\varepsilon^2 = 9$, $\sigma_u^2 = 1$

N	Algoritmo CRW	Máxima Verosimilitud con el FK			Algoritmo CRW1			Algoritmo FK-SIF1 con $t = 10^6$		
	$\hat\beta$	$\hat\beta$	$\hat\sigma_\varepsilon^2$	$\hat\sigma_u^2$	$\hat\beta$	$\hat\sigma_\varepsilon^2$	$\hat\sigma_u^2$	$\hat\beta$	$\hat\sigma_\varepsilon^2$	$\hat\sigma_u^2$
100	0.5001	0.5046	14.580	1.2647	0.5078	11.250	2.2305	0.5023	11.929	0.0095
	(0.0662)	(0.0652)	(2.9051)	(1.5527)	(0.0661)	(2.4502)	(5.4137)	(0.0670)	(2.7676)	(0.0115)
200	0.4982	0.4919	14.749	1.2475	0.4966	10.841	1.3398	0.4966	11.050	0.0051
	(0.0459)	(0.0493)	(2.2143)	(1.2651)	(0.0469)	(1.8720)	(1.2674)	(0.0479)	(1.4392)	(0.0059)
500	0.5002	0.5020	15.279	0.9678	0.5000	12.945	0.8876	0.4985	10.640	0.0019
	(0.0300)	(0.0312)	(1.4742)	(0.7787)	(0.0299)	(1.3882)	(0.0491)	(0.0299)	(0.8126)	(0.0022)

Notas: N: Tamaño muestral utilizado. Las cifras entre paréntesis representan las desviaciones típicas muestrales.

A la vista de los resultados, puede concluirse que, para las tres especificaciones consideradas, el valor medio del parámetro β es muy próximo a su verdadero valor con los cuatro procedimientos utilizados. Se observa que para muestras cortas (100 o 200 observaciones), el algoritmo propuesto (CRW1) genera estimaciones más eficientes del parámetro β que el procedimiento FK-SIF1. En estos casos, el procedimiento de referencia (CRW) es el que proporciona la menor desviación típica muestral del coeficiente β. No obstante, a medida que aumenta el tamaño muestral, se observa que la mejora en eficiencia al estimar dicho coeficiente con el criterio CRW1 disminuye o prácticamente desaparece con respecto al FK-SIF1.

En cuanto a la estimación de la varianza σ_ε^2, se aprecia que, tanto los procedimientos recursivos (CRW1 y FK-SIF1) como el método de máxima verosimilitud usando el FK, tienden a sobrestimar el verdadero valor de dicha varianza, aunque disminuya el sesgo según crece el tamaño muestral utilizado. La justificación de este hecho es que, incluso cuando las condiciones iniciales del filtro son extremadamente buenas (como en el caso del criterio CRW1), los errores de predicción $\tilde{z}_{t/t-1}$ están "inflados" al principio de la muestra y esto hace que la convergencia de dicho parámetro a un valor estacionario sea más lenta. Para resolver este problema caben, al menos, dos posibilidades. La primera, consiste en no actualizar la estimación de dicha varianza al procesar las primeras observaciones, haciendo que la varianza muestral de los primeros errores de predicción calculados con pocos datos no distorsionen la estimación final de la varianza σ_ε^2. La segunda solución se basa en estimar esta varianza usando sólo los *m* datos más recientes de la muestra, en lugar de toda la información pasada, siendo *m* la amplitud de la "ventana" utilizada para estimar este parámetro.

Con respecto a la varianza asociada a la perturbación u_t, el algoritmo CRW1 tiende a sobrestimar su valor teórico y por el contrario, el criterio FK-SIF1 siempre infraestima dicho parámetro. En este caso, el procedimiento de máxima verosimilitud usando el FK es el que proporciona estimaciones más eficientes de la varianza σ_u^2. La diferencia entre los resultados obtenidos con el método CRW1 y el FK-SIF1 en la estimación de esta varianza se basa en que, aunque en los dos métodos la expresión del estimador recursivo de dicha varianza coincide, en el primer método las matrices ganancias calculadas en los dos filtros propagados no dependen de las condiciones iniciales de las matrices de información, mientras que en la etapa de filtrado del criterio FK-SIF1 la matriz ganancia está afectada por la inicialización arbitrariamente elegida para la matriz de covarianzas (o para la inversa de la matriz de información).

A pesar de los buenos resultados obtenidos por máxima verosimilitud usando el FK, hay que indicar que esta técnica, al no ser recursiva, tiene la desventaja con respecto a los procedimientos CRW1 y FK-SIF1, de que consume mucho más tiempo de cálculo, sobre todo cuando se usan muestras largas y las condiciones iniciales de los parámetros (σ_ε^2 y σ_u^2) se eligen alejadas del óptimo. La razón de ser un criterio muy costoso en términos computacionales es que exige propagar el filtro de Kalman para toda la muestra en cada de vector de parámetros hasta el alcanzar el máximo de la función de verosimilitud de acuerdo con algún procedimiento de optimización numérica. En concreto, y como ejemplo, el criterio CRW1 necesita, por término medio, 1 minuto y medio para estimar el modelo (41)-(42) usando 1000 observaciones, mientras que por máxima verosimilitud exacta el tiempo medio de cálculo es de casi 22 minutos.

5. CONCLUSIONES

La estimación recursiva de modelos con parámetros cambiantes mediante el criterio habitual en dos etapas, consistente en propagar un filtro de Kalman y después un *smoother* de intervalo fijo (denotado por FK-SIF), exige conocer o fijar las varianzas de todos los términos de error presentes en el modelo. Dado que los resultados de este criterio son sensibles a la inicialización del mismo, es preferible utilizar un algoritmo como el de Cooley, Rosenberg y Wall (1977) (denotado por CRW), independiente de condiciones iniciales. Sin embargo, este algoritmo también supone conocidas las varianzas de las perturbaciones existentes. El cumplimiento de este supuesto no es habitual en la práctica econométrica y, por otra parte, en este trabajo se muestra que la elección arbitraria de dichas varianzas, o de los ratios entre ellas, puede afectar tanto a la eficiencia de los estimadores del resto de parámetros, como a la trayectoria de las estimaciones, sobre todo cuando se trabaja con muestras cortas.

En este trabajo, se ha derivado una versión del algoritmo CRW (1977) que permite obtener una estimación recursiva de las varianzas relevantes de un modelo de parámetros cambiantes, además de la secuencia óptima de estimaciones de los parámetros propios del modelo. Esta nueva versión del filtro CRW es independiente de las condiciones iniciales de todos los parámetros, considerando también como parámetros las varianzas de las distintas perturbaciones. Siguiendo la misma idea, se deriva una versión del algoritmo FK-SIF, denotada por FK-SIF1, que incorpora la estimación *on-line* de las varianzas de los ruidos. Sin embargo, los resultados de este nuevo algoritmo dependen del vector inicial de parámetros del que se parta, comprobando que la elección arbitraria de estas condiciones iniciales afecta a la estimación final de las varianzas de las perturbaciones presentes. Por tanto, si los métodos estándar de estimación recursiva de modelos de parámetros fijos son sensibles a las condiciones iniciales [ver Sotoca (1993)], este problema también es importante en el contexto de parámetros cambiantes.

Al igual que el CRW, este nuevo algoritmo está basado en la utilización de dos filtros de información que se propagan de forma independiente entre sí, pudiendo correr en paralelo para después combinar de forma óptima los resultados de ambos. Este hecho supone una ventaja computacional con respecto a otros algoritmos de *smoothing* más clásicos, como es el criterio FK-SIF o el FK-SIF1.

Se ha aplicado este procedimiento a la estimación de modelos de regresión con un único parámetro cambiante usando datos simulados. Los resultados obtenidos con este nuevo criterio indican un razonable funcionamiento del mismo, sobre todo para muestras largas. En concreto, para las especificaciones consideradas, los dos criterios propuestos (CRW1 y FK-SIF1) estiman por encima de su verdadero valor la varianza de las perturbaciones del modelo, debido a que en la estimación de dicho parámetro no se descuenta el efecto de los errores de predicción cometidos al principio de la muestra (que suelen estar "inflados" al disponer todavía de poca información muestral en el proceso recursivo). Por otro lado, el procedimiento CRW1 aproxima mejor la varianza asociada a la ley de evolución del parámetro cambiante que el FK-SIF1, el cual siempre infrestima este parámetro. En este caso, la diferencia en el resultado obtenido se debe a que el criterio CRW1 dispone de condiciones iniciales exactas para las variables del filtro, mientras que el FK-SIF1 no es independiente de la inicialización.

Por último, y a efectos de comparación, se han estimado los mismos modelos por máxima verosimilitud exacta usando el filtro de Kalman. Los resultados indican que, trabajando con muestras largas, el procedimiento recursivo CRW1 genera estimaciones comparables a las máximo-verosímiles, pero con la ventaja de ser una técnica mucho menos costosa en términos de tiempo de cálculo que el criterio de máxima verosimilitud.

REFERENCIAS

ANDERSON, B.D.O. y MOORE, J.B. (1979). «Optimal Filtering». *Prentice-Hall, Inc.*, New Jersey.

COOLEY, T.F. y PRESCOTT, E.C. (1973). «An Adaptive Regression Model»". *International Economic Review*, 14, 364-371.

COOLEY, T.F. y PRESCOTT, E.C. (1976). «Estimation in the Presence of Stochastic Parameter Variation». *Econometrica*, 44, 1, 167-184.

COOLEY, T.F., ROSENBERG, B. y WALL, K.D. (1977). «A Note on Optimal Smoothing for Time Varying Coefficient Problems». *Annals of Economic and Social Measurement*, 6, 4, 453-456.

FRASER, D.C. y POTTER, J.E. (1969). «The Optimum Linear Smoother as a Combination of two Optimum Linear Filters». *IEEE Transactions on Automatic Control*, AC-14, 4, 387-390.

GARCÍA-FERRER, A., HOYO, J. DEL,. NOVALES, A y. YOUNG, P.C (1993). «Further Evidence on Forecasting International GNP Growth Rates Using Unobserved Components Transfer Function Models» *Documento de Trabajo ICAE*, nº 9312.

HARVEY, A.C. (1989). «Forecasting Structural Time Series Models and the Kalman Filter». *Cambridge University Press.*

MARGARITIS D. (1990). «A Time-Varying Model of Rational Learning» *Economics Letters*, 33, 309-314.

PAGAN, A.R. (1980). «Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients». *Journal of Econometrics*, 13, 341-363.

SOTOCA, S. (1993). «El Problema de las Condiciones Iniciales en los Algoritmos de Estimación Recursiva de Modelos Lineales». *Estadística Española*, 35, 132, 89-115.

WOLFF, C. (1988). «Autoregressive Conditional Heteroscedasticity. A Comparison of ARCH and Random Coefficient Models». *Economic Letters*, 27, 141-143.

YOUNG, P. (1984). «Recursive Estimation and Time-Series Analysis. An Introduction». *Springer-Verlag, Heidelberg.*

A NOTE ON EFFICIENT ESTIMATION OF TIME VARYING COEFFICIENT MODELS

SUMMARY

Standard estimation procedures for the time-varying parameters model suppose that the variances of the noises in the model are known. Obviously, this assumption is not realistic in most econometric applications. Besides, the results of these methods are sensitive to the initial conditions of the algorithm, a fact that is often overlooked by the literature. In this paper, we propose an extension of the recursive algorithm proposed by Cooley, Rosenberg y Wall (1977), which is independent of initial conditions and includes on-line estimation of all the relevant variances. The results obtained with this method compare favourably with those obtained by standard procedures.

Key words initial conditions, time varying coefficient models, information filter, Kalman filter, noise variance ratio, fixed-interval smoother.

AMS Classification: 62J05, 62J07