Documento de Trabajo

9220

"EL PROBLEMA DE LAS CONDICIONES INICIALES EN LOS ALGORITMOS DE ESTIMACION RECURSIVA DE MODELOS LINEALES"

Sonia Sotoca López

FACULTAD DE CIENCIAS ECONOMICAS Y EMPRESARIALES
UNIVERSIDAD COMPLUTENSE DE MADRID
Campus de Somosaguas. 28.233 MADRID.

EL PROBLEMA DE LAS CONDICIONES INICIALES EN LOS ALGORITMOS DE ESTIMACION RECURSIVA DE MODELOS LINEALES.

Sonia Sotoca López*

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

Junio 1992

* Quiero agradecer especialmente a Miguel Jerez por sus interesantes comentarios y sugerencias. También deseo agradecer los comentarios recibidos de Rafael Flores, Mercedes Gracia-Díez, Alfonso Novales, Teodosio Pérez y Gregorio R. Serrano.

RESUMEN:

En este artículo se proponen dos soluciones para independizar los resultados del criterio de estimación recursiva estándar de la influencia de condiciones iniciales arbitrarias. La primera solución consiste en utilizar un algoritmo recursivo corregido que descuenta el efecto de condiciones iniciales elegidas arbitrariamente sobre la estimación final de los parámetros de un modelo de regresión. La segunda solución consiste en la utilización de los filtros basados en la propagación de la matriz de información en lugar de la matriz de covarianzas. Estos algoritmos disponen, por construcción, de condiciones iniciales exactas y son robustos numéricamente.

ABSTRACT:

This article proposes two solutions to allow the final recursive estimates of a regression model to be independent from arbitrary initial conditions. The first solution uses a recursive algorithm which discounts the effect of arbitrary initial conditions on the final parameter estimates. The second solution uses filters based on propagating the information matrix rather than the covariance matrix. By construction, this algorithm has exact initial conditions and is numerically robust.

1. Introducción.

Como es bien sabido, el algoritmo de estimación por mínimos cuadrados recursivos (MCR) es un caso particular del filtro de Kalman (ver, por ejemplo, Jazwinsky, 1970) y, por consiguiente, un aspecto crucial para su uso es la elección de condiciones iniciales.

En el contexto de los métodos de estimación recursiva, resulta frecuente utilizar condiciones iniciales arbitrarias o, a lo sumo, diseñadas de acuerdo con una regla heurística (ver Young, 1984). Este tipo de criterios *ad-hoc* puede dar buenos resultados cuando se dispone de muestras grandes y/o en situaciones de estimación bien condicionadas. Sin embargo, la práctica econométrica nos enfrenta a menudo con problemas de estimación mal condicionados (debidos, por ejemplo, a la presencia de multicolinealidad aproximada) que deben resolverse a partir de muestras muy limitadas. En estos casos, el uso de condiciones iniciales arbitrarias puede dar lugar a la degradación numérica del proceso recursivo, haciendo que las estimaciones finales del procedimiento no coincidan con los resultados del método de mínimos cuadrados ordinarios (MCO).

En este artículo se proponen dos soluciones para independizar los resultados del criterio MCR de la influencia de condiciones iniciales elegidas *ad-hoc*: a) utilizar un algoritmo de estimación MCR corregido (MCRC) y b) utilizar un algoritmo de estimación MCR basado en el filtro de la matriz de información[1]. El criterio MCRC descuenta el efecto de condiciones iniciales elegidas arbitrariamente sobre la estimación final de los parámetros de un modelo. Este criterio se inicializa también con arbitrariedad, pero cuando se dispone de cero grados de libertad se lleva a cabo un paso intermedio de filtrado, de forma que los resultados de este filtro intermedio se convierten en las nuevas condiciones iniciales del algoritmo MCR habitual. El algoritmo de la matriz de información puede inicializarse usando condiciones iniciales exactas, es analíticamente idéntico al estimador por MCO y su comportamiento numérico es sumamente estable, incluso en situaciones como las anteriormente mencionadas.

Es importante señalar que éstas son dos soluciones alternativas. En general, será preferible el criterio basado en la matriz de información cuando exista una gran incertidumbre inicial acerca del valor de los parámetros y sólo se requiera la estimación final de los mismos. Por el contrario, cuando sea interesante estudiar la evolución temporal de las estimaciones, es más conveniente utilizar el método MCRC.

La estructura del artículo es la siguiente:

En el apartado 2, se describe el algoritmo de estimación MCR estándar, basado en el filtro de Kalman y se demuestra que este método genera estimaciones comparables a las de una regresión cresta. También, se describen las formas de cálculo habituales de las condiciones iniciales del criterio MCR y se propone un método heurístico para su estimación. Aunque este método es también arbitrario, evita la degradación numérica del algoritmo recursivo, en algunos casos. Por último, se lleva a cabo un análisis de la divergencia existente entre las estimaciones finales obtenidas con MCR y con MCO.

En el apartado 3, se describen las dos soluciones propuestas al problema de inicialización del criterio MCR. En primer lugar, se describe el algoritmo de mínimos cuadrados recursivos corregidos (MCRC), que independiza los resultados del método MCR del efecto de condiciones iniciales elegidas arbitrariamente. La segunda solución surge al particularizar el filtro de información para el modelo de regresión lineal con parámetros constantes, lo que permite obtener las ecuaciones de un estimador recursivo alternativo (MCRI). Se demuestra que este método puede inicializarse a partir de condiciones iniciales exactas y se argumenta su robustez numérica.

En el apartado 4 se aplican las tres estrategias de estimación recursiva consideradas a datos simulados. En los ejemplos, se consideran distintos tamaños muestrales y distintas situaciones de condicionamiento. Por último, en el apartado 5, se resumen las conclusiones más importantes del trabajo.

2. Estimación recursiva mediante el filtro de Kalman (MCRK).

Sea el modelo de regresión lineal (MRL):

$$y_t = x_t^T \beta + \epsilon_t \qquad (t = 1,2,\ldots,n) \tag{1}$$

donde y_t es la observación t-ésima de la variable dependiente, x_t es un vector ($k \times 1$) que contiene la observación t-ésima de las variables explicativas, β es un vector ($k \times 1$) de parámetros desconocidos y ϵ_t es una perturbación aleatoria que se distribuye idéntica e independientemente con esperanza nula y varianza constante σ_ϵ^2.

Es sabido que la estimación de (1) puede llevarse a cabo utilizando MCO con toda la muestra o bien recursivamente, aplicando el filtro de Kalman a las dos ecuaciones siguientes:

$$y_t = x_t^T \beta_t + \epsilon_t \tag{2}$$

$$\beta_t = \beta_{t-1} \tag{3}$$

en donde β_t representa el vector de parámetros correspondiente a una muestra de tamaño t y (3) refleja la hipótesis de parámetros constantes. Desde el punto de vista de la formulación en espacio de los estados, (2) es la ecuación de observación del modelo y (3) es la ecuación de estado. Por tanto, desde este enfoque, la variable endógena del MRL puede interpretarse como una señal observable del vector de estado que es, a su vez, el vector de parámetros que se desea estimar.

Las ecuaciones que forman el criterio MCRK surgen al particularizar el filtro de Kalman para el modelo dado por (2) y (3) (ver Young, 1984):

$$\hat{\beta}_t = \hat{\beta}_{t-1} + k_t(y_t - x_t^T \hat{\beta}_{t-1}) \tag{4}$$

$$k_t = V_{t-1} x_t (x_t^T V_{t-1} x_t + 1)^{-1} \tag{5}$$

$$V_t = (I_k - k_t x_t^T) V_{t-1} \tag{6}$$

La ecuación (4) indica que la estimación del vector de parámetros se actualiza mediante una función lineal del error de predicción de y_t cometido al estimar β con la información disponible hasta el instante *t-1*. El factor de ponderación es el vector k_t, que se conoce como *ganancia* del filtro y cuya expresión figura en (5).

La matriz V_t, propagada a través de (6), es proporcional a la matriz de covarianzas de los parámetros estimados con la información disponible hasta el instante t, que denotamos por P_t. La relación entre ambas matrices es la siguiente:

$$V_t = \frac{1}{\sigma_\epsilon^2} P_t = (X_t^T X_t)^{-1} \tag{7}$$

en donde X_t es una matriz $(t \times k)$ que recoge, por columnas, la información muestral de cada variable explicativa hasta el instante t.

2.1. Formas de cálculo de las condiciones iniciales del criterio MCRK.

Evidentemente, es necesario inicializar el algoritmo dado por (4)-(6) con unas condiciones iniciales ($\hat{\beta}_0$, V_0). Existen distintas formas de estimar las mismas.

(a) Condiciones iniciales arbitrarias.

El criterio utilizado habitualmente (ver, entre otros, Young (1984), pág. 27 y Harvey (1989), pág. 107), consiste en fijar:

$$\hat{\beta}_0 = 0 \tag{8}$$

$$V_0 = \tau I_k \tag{9}$$

donde τ es un escalar positivo y arbitrariamente grande.

Las condiciones iniciales (8)-(9), expresan un elevado grado de incertidumbre acerca de la magnitud y el signo de los parámetros a estimar. Este criterio responde a la idea de que, si no se dispone de información *a priori*, $\hat{\beta}_0$ es una variable aleatoria, afectada por una incertidumbre infinita. Obsérvese que las condiciones (8)-(9) convergen a los dos primeros momentos de esta distribución cuando $\tau \rightarrow \infty$. Puesto que el algoritmo MCRK no puede propagarse comenzando con un valor infinito de τ, en la práctica se utiliza un valor finito, arbitrario y suficientemente grande del mismo.

La expresión (4) puede escribirse en función de las condiciones iniciales β_0 y V_0, de la forma[2]:

$$\hat{\beta}_t = (V_0^{-1} + \sum_{i=1}^{t} x_i x_i^T)^{-1}(V_0^{-1}\hat{\beta}_0 + \sum_{i=1}^{t} x_i y_i) \tag{10}$$

y teniendo en cuenta (8) y (9), la expresión (10) se reduce a:

$$\hat{\beta}_{t/MCRK} = (\xi I_k + \sum_{i=1}^{t} x_i x_i^T)^{-1}(\sum_{i=1}^{t} x_i y_i) \tag{11}$$

en donde $\xi = 1/\tau$ es un escalar positivo y arbitrariamente pequeño y $\hat{\beta}_{t/MCRK}$ denota la estimación obtenida con el criterio MCRK después de haber procesado t observaciones. La expresión (11) muestra que el estimador MCRK puede interpretarse como un estimador cresta (ver Gruber, 1990). Como es bien sabido, este estimador es sesgado, aunque su varianza puede ser menor que la del MCO si se elige adecuadamente la constante ξ que se añade a los elementos de la diagonal principal de $X_t^T X_t$. Sin embargo, el problema es que existen infinitos valores de ξ que cumplen dicha propiedad y no resulta obvio cúal de ellos debe seleccionarse.

Hoel y Kennard (1970) sugieren probar con distintos valores de ξ a partir de uno que mejore al MCO (es decir, con un error cuadrático medio más pequeño que la estimación por MCO). Si a partir de un cierto rango de valores, el estimador cresta no cambia mucho, entonces elegir el valor superior de dicho rango. Schmidt (1976) propone elegir el valor de ξ que minimice el error cuadrático medio del estimador cresta. El problema de este método es que el valor de ξ que cumple esta condición no es computable, ya que depende del verdadero valor del vector desconocido β.

Young (1984) sugiere que un valor de ξ del orden de 10^{-6} es suficiente para que el estimador MCRK converga al MCO al final de la muestra. No obstante, aunque es evidente que (11) converge a la expresión MCO cuando $\xi \rightarrow 0$, éste también es un criterio arbitrario. El estimador cresta es consistente (ver Apéndice B) y por tanto, para muestras grandes, un valor de ξ del orden que sugiere Young puede ser suficiente. Sin embargo, para muestras pequeñas, ξ puede no ser una cantidad despreciable con respecto a $X_t^T X_t$, dependiendo de cúal sea el orden de magnitud de las variables explicativas.

En general, la selección del factor ξ sin tener en cuenta la métrica del problema de estimación, puede dar lugar a la acumulación de errores numéricos en la propagación de V_t a través de (6). Esto ocurre, sobre todo, si se escoge un valor de τ demasiado grande. Por otra parte, si τ es excesivamente pequeño, las condiciones iniciales no recogerían suficientemente la incertidumbre existente *a priori*, haciendo que las estimaciones finales de los parámetros sean parecidas a sus valores iniciales. Es decir, el efecto sería similar a implantar la restricción de que cada parámetro sea "casi" igual a su condición inicial.

(b) Método heurístico de cálculo de las condiciones iniciales.

Proponemos un método heurístico muy sencillo para elegir ξ, que consiste en comparar la métrica de las matrices $X_t^T X_t$ y ξI_k. La escala de $X_t^T X_t$ puede medirse por su traza o por una norma de dicha matriz. En ambos casos, se puede fijar una tolerancia ϵ tan pequeña como se quiera, de forma que:

$$\frac{\xi}{tr(X_t^T X_t)} < \epsilon \tag{12}$$

o bien:

$$\frac{\xi}{\| X_t^T X_t \|_2} < \epsilon \tag{13}$$

siendo $\| X_t^T X_t \|_2 = (a_{11}^2 + a_{22}^2 + \ldots + a_{kk}^2)^{1/2}$ y a_{ii} $(i = 1, \ldots, k)$ los elementos de la diagonal principal de $X_t^T X_t$.

Fijado ϵ, puede encontrarse el valor de τ que hace que el estimador MCRK de β tienda al estimador MCO con toda la muestra. De esta forma, si el valor de τ es demasiado grande, lo cual podría dar lugar a la degradación numérica del algoritmo, puede fijarse un valor de τ más pequeño y buscar un factor de escala adecuado para las variables independientes del modelo. El inconveniente de este método es que exige procesar toda la muestra para medir el tamaño de la matriz $X_t^T X_t$ o, en su caso, para escalar adecuadamente los datos de las variables explicativas, con lo que se desvirtúa el interés por la recursividad del algoritmo MCRK. No obstante, aunque este método no resuelve el problema de la arbitrariedad de las condiciones iniciales, permite acotar los valores del factor τ que evitan la inestabilidad numérica del algoritmo.

En este mismo contexto, también se puede utilizar alguna normalización de la matriz $X_t^T X_t$, como por ejemplo, estandarizar los órdenes de magnitud de sus elementos.

(c) **Condiciones iniciales estimadas por MCO.**

Otra posibilidad, ya conocida en la literatura, para conseguir condiciones iniciales del algoritmo recursivo consiste en utilizar parte de la información muestral disponible para su estimación. En concreto, se usan las $k+1$ primeras observaciones para inicializar el filtro:

$$\hat{\beta}_0 = (X_{k+1}^T X_{k+1})^{-1} X_{k+1}^T y_{k+1} \tag{14}$$

$$V_0 = (X_{k+1}^T X_{k+1})^{-1} \tag{15}$$

Es decir, se obtiene una primera estimación del vector β aplicando MCO con un grado de libertad y se actualiza la estimación de este vector de parámetros para $t = k+2, \dots, n$. La desventaja de este método consiste en que, de nuevo, se desvirtúa la recursividad del procedimiento MCRK.

2.2. Análisis de la divergencia entre las estimaciones finales obtenidas con MCRK y MCO.

A partir de (11) y la expresión del estimador MCO, es fácil ver que una vez procesada una muestra de tamaño t, la diferencia resultante en las estimaciones obtenidas con los algoritmos MCRK y MCO, depende exclusivamente del tamaño de una matriz W_t definida como:

$$W_t = (\xi I_k + X_t^T X_t)^{-1} - (X_t^T X_t)^{-1} \tag{16}$$

Lema 1: La traza de la matriz W_t, depende positivamente del tamaño del escalar τ (o bien, negativamente del valor de ξ).

Demostración: Sea la factorización de la matriz $X_t^T X_t$:

$$X_t^T X_t = M \Omega M^T \tag{17}$$

donde Ω es una matriz diagonal que contiene los autovalores de $X_t^T X_t$ y M es una matriz ortogonal que contiene por columnas los autovectores de $X_t^T X_t$. Esta factorización siempre existe al ser $X_t^T X_t$ una matriz definida positiva.

Puesto que las matrices $(X_t^T X_t)^{-1}$ y $(\xi I + X_t^T X_t)^{-1}$ tienen los mismos autovectores (ver Schmidt, 1976), es fácil demostrar que W_t puede factorizarse de la forma:

$$W_t = M[(\xi I + \Omega)^{-1} - \Omega^{-1}]M^T = M diag\left[\frac{-\xi}{\lambda_i(\lambda_i + \xi)}\right]M^T \tag{18}$$

donde λ_i es el autovalor i-ésimo de la matriz $X_t^TX_t$ y el escalar $-\xi/\lambda_i(\lambda_i + \xi)$ el autovalor i-ésimo de la matriz diferencia W_t, siendo esta matriz definida negativa [la matriz W_t es definida negativa si y sólo si $(X_t^TX_t)$ - $(\xi I + X_t^TX_t)$ es definida negativa].

A partir de (18), la traza de W_t es:

$$tr(W_t) = -\sum_{i=1}^{k} \frac{\xi}{\lambda_i(\lambda_i + \xi)} \leq 0 \tag{19}$$

y derivando (19) con respecto al escalar ξ, se obtiene:

$$\frac{\partial tr(W_t)}{\partial \xi} = -\sum_{i=1}^{k} \frac{1}{(\lambda_i + \xi)^2} \tag{20}$$

El signo negativo de la derivada en (20) indica que, a medida que disminuye el valor de ξ (o bien, a medida que aumenta el valor del escalar τ), aumenta el tamaño de la matriz W_t. ■

Lema 2: Si algún autovalor de $X_t^TX_t$ es cercano a cero, entonces $\mathrm{tr}(W_t) \rightarrow -\infty$.

Demostración: Directamente de la expresión (19). ■

Las implicaciones de estos dos lemas son inmediatas. El Lema 1 establece que la diferencia entre los resultados de las estimaciones MCRK y MCO es una función directa del tamaño del factor τ. Por consiguiente, si no se dispone de suficientes datos, la diferencia entre ambas estimaciones puede ser sustancial. Por otra parte, según el Lema 2, el mal condicionamiento de la matriz $X_t^TX_t$ afecta directa y positivamente a la magnitud de la matriz W_t. Es decir, en las situaciones en que exista un cierto grado de multicolinealidad, el estimador MCRK puede dar resultados muy diferentes a los de MCO.

3. Soluciones al problema de condiciones iniciales en el criterio MCRK.

En este apartado se plantean dos métodos que permiten resolver el problema de degradación de los resultados de MCRK en presencia de colinealidad o cuando la muestra es pequeña.

3.1. Estimación mediante mínimos cuadrados recursivos corregidos (MCRC).

Como se ha visto en el apartado 2, el algoritmo MCRK inicializado arbitrariamente genera estimaciones comparables a una regresión cresta. En este sentido, no está garantizada la convergencia del estimador MCRK a su expresión MCO, una vez procesada toda la muestra.

La idea que se propone para resolver este problema es la que da lugar al nuevo criterio MCRC y se expone a continuación. Las condiciones iniciales del tipo (8) y (9) pueden interpretarse como el resultado de aplicar MCO a una muestra ficticia que forma el conjunto de información denotado por I^0. Las observaciones que generan las estimaciones de β_0 y V_0 dadas por (8) y (9), son las siguientes:

$$y^0 = 0 \qquad X^0 = \frac{1}{\sqrt{\tau}} I \tag{21}$$

donde y^0 es un vector $(k \times 1)$ que contiene las observaciones extramuestrales formadas por el vector $(y_{-k}, y_{-(k+1)}, \dots, y_{-1})^T$ y X^0 es una matriz $(k \times k)$ que contiene por columnas la información muestral ficticia de las k variables explicativas del modelo. Es decir, si denotamos por x_j^0 la columna j-ésima de X^0, el vector x_j^{0T} estará formado por los elementos $(x_{-k,j}, x_{-(k+1),j}, \dots, x_{-1,j})$.

Dada la información representada por (21) es fácil ver que la estimación MCO del vector β del MRL será:

$$\hat{\beta}_0 = (X^{0T}X^0)^{-1}X^{0T}y^0 = 0 \tag{22}$$

$$V_0 = (X^{0T}X^0)^{-1} = \tau I \tag{23}$$

Por tanto, si las condiciones iniciales dadas en (8) y (9) provienen de procesar el conjunto de información I^0, la idea sería eliminar el efecto que esas observaciones ficticias tienen sobre las estimaciones MCRK cuando se han procesado k observaciones reales.

En resumen, para independizar el criterio MCRK de la influencia de las condiciones iniciales (8) y (9), pueden llevarse a cabo los siguientes pasos:

(a) Inicializar el algoritmo MCRK con condiciones iniciales arbitrarias del tipo (8) y (9) hasta el instante $t=k$ (es decir, hasta disponer de cero grados de libertad):

$$\forall t = 1,2,\ldots,k$$

$$k_t = V_{t-1}x_t(x_t^T V_{t-1}x_t + 1)^{-1} \tag{24}$$

$$\hat{\beta}_t = \hat{\beta}_{t-1} + k_t(y_t - x_t^T\hat{\beta}_{t-1}) \tag{25}$$

$$V_t = (I - k_t x_t^T)V_{t-1}(I - k_t x_t^T)^T + k_t k_t^T \tag{26}$$

(b) Eliminar el efecto del conjunto de información I^0 sobre la estimación $\hat{\beta}_k$. Para ello, se pueden utilizar las siguientes ecuaciones de corrección[3] (ver Young (1984), pág. 60):

$$\forall s = k+1, k+2,\ldots,k+k$$

$$k_k^* = V_k x_{k-s}(x_{k-s}^T V_k x_{k-s} - 1)^{-1} \tag{27}$$

$$\hat{\beta}_k^* = \hat{\beta}_k + k_k^*(y_{k-s} - x_{k-s}^T\hat{\beta}_k) \tag{28}$$

$$V_k^* = (I - k_k^* x_{k-s}^T)V_k \tag{29}$$

(c) Utilizar las estimaciones $\hat{\beta}_k^*$ y V_k^* como condiciones iniciales del algoritmo recursivo (24)-(26), $\forall t = k+1, k+2, \ldots, n$.

Por tanto, esta estrategia supone llevar a cabo un paso intermedio de filtrado en el instante $t=k$. El filtro dado por las ecuaciones (27)-(29), descuenta el efecto de cada una de las observaciones ficticias generadas por las condiciones iniciales (8)-(9), sobre la estimación de los parámetros obtenida hasta ese momento. Los resultados de ese filtro intermedio se convierten en las nuevas condiciones iniciales del algoritmo MCRK habitual.

Con este mecanismo de corrección, no está garantizado que la expresión analítica del estimador MCRC coincida con la MCO al final de la muestra. Sin embargo, descontar el efecto de condiciones iniciales arbitrarias es suficiente para que las estimaciones MCRC converjan numéricamente a las MCO, aún en circunstancias de mal condicionamiento o con muestras cortas.

3.2. Estimación recursiva mediante el filtro de la matriz de información (MCRI).

El filtro de información se obtiene aplicando el lema de inversión de matrices a las ecuaciones del filtro de Kalman (ver Anderson y Moore (1979), pág. 138). Por tanto, ambos filtros son algebraicamente equivalentes. Las ventajas del filtro de información con respecto al criterio MCRK son dos:

1) Permite instrumentar exactamente las condiciones iniciales que se aproximan en el filtro de Kalman mediante $V_0 = \tau I$.

2) Las matrices del filtro se propagan de forma estable a partir de las condiciones iniciales.

El filtro de información propaga la inversa de V_t y un vector de estado (a_t^*) que es combinación lineal de los elementos del vector de estado original (β_t). Las condiciones iniciales de este algoritmo son siempre:

$$V_0^{-1} = 0 \tag{30}$$

$$\hat{a}_0^* = 0 \tag{31}$$

y las ecuaciones del filtro de información particularizadas para el modelo dado por (2)-(3), son las siguientes:

$$V_t^{-1} = V_{t-1}^{-1} + x_t x_t^T \tag{32}$$

$$\hat{a}_t^* = \hat{a}_{t-1}^* + x_t y_t \tag{33}$$

donde el vector de estimaciones de los parámetros originales del modelo se recupera mediante la siguiente relación:

$$\hat{a}_t^* = V_t^{-1}\hat{\beta}_t \;\Rightarrow\; \hat{\beta}_t = V_t \hat{a}_t^* \tag{34}$$

Escribiendo las ecuaciones (32) y (33) en función de las condiciones iniciales (30) y (31), se obtiene:

$$V_t^{-1} = V_0^{-1} + \sum_{i=1}^{t} x_i x_i^T \tag{35}$$

$$\hat{a}_t^* = \hat{a}_0^* + \sum_{i=1}^{t} x_i y_i \tag{36}$$

y sustituyendo (30), (31), (35) y (36) en (34), resulta:

$$\hat{\beta}_{t/MCRI} = (\sum_{i=1}^{t} x_i x_i^T)^{-1} (\sum_{i=1}^{t} x_i y_i) \qquad \forall t = k, k+1, \ldots, n \tag{37}$$

en donde $\hat{\beta}_{t/MCRI}$ denota la estimación de β obtenida con el criterio MCRI, cuando se han procesado t observaciones. La expresión (37) coincide exactamente con la del estimador MCO para una muestra de tamaño t. Esto prueba la independencia del filtro de información de las condiciones iniciales. La robustez numérica del filtro surge al propagar la matriz de información como una suma de matrices definidas positivas [ver (35)]. Además, el estimador MCRI es analíticamente idéntico a la expresión MCO.

La desventaja de este método con respecto a MCRK y MCRC es que para recuperar la evolución de los parámetros originales, es necesario utilizar la relación (34) que supone invertir la matriz de información cada vez que se añade un nuevo dato.

4. Resultados empíricos con datos simulados.

En este apartado se ilustran las conclusiones teóricas de los apartados anteriores, utilizando datos simulados. En concreto, se han aplicado los criterios de estimación MCRK, MCRI y MCRC a dos modelos: el primero bien condicionado (modelo I) y el segundo mal condicionado (modelo II), para distintos tamaños muestrales.

En la Tabla 1 se presentan los resultados de la estimación del modelo I mediante MCRK, MCRI y MCRC. Los valores entre paréntesis son las desviaciones típicas de las distribuciones empíricas. En la segunda columna de la Tabla, se ofrece el número de condición de la matriz $X_t^T X_t$ para los distintos tamaños muestrales considerados. Este número es, en todos los casos, del orden de magnitud de la unidad, indicando el buen condicionamiento del modelo.

A la vista de los resultados, en esta Tabla se observa que, con cualquiera de los tres procedimientos recursivos utilizados, a medida que aumenta el tamaño muestral, el valor medio de las estimaciones finales se acerca cada vez más a su correspondiente valor teórico. Sin embargo, independientemente del tamaño muestral, el error cometido[4] al estimar cualquier parámetro con MCRK es siempre mayor que el correspondiente a MCRI y MCRC. Además, se observa que para todas las muestras consideradas, los resultados obtenidos con MCRI convergen numéricamente a los obtenidos con MCRC. Es decir, el algoritmo MCRC realmente corrige el efecto de condiciones iniciales arbitrarias ($\tau = 10^6$, como sugiere Young, 1984) sobre el valor final de las estimaciones, haciendo que éstas convergan a la MCO.

En la Tabla 2 se presentan los resultados de la simulación del modelo II. En este caso, el número de condición de la matriz $X_t^T X_t$ es del orden de 10^4. Este modelo se ha generado incluyendo un término constante y dos variables exógenas que evolucionan de acuerdo con la siguiente relación:

$$x_{2t} = 2x_{3t} + \epsilon_t \qquad \forall t = 1,2,\ldots,n$$

donde ϵ_t y x_{3t} son procesos ruido blanco independientes entre sí y el ratio señal/ruido de esa relación es de 5.03.

En primer lugar, se observa que los algoritmos MCRK, MCRI y MCRC estiman adecuadamente el término constante. Sin embargo, el error cometido en la estimación de los parámetros asociados a las variables correlacionadas es siempre superior con MCRK que con los otros dos procedimientos recursivos, tendiendo a cero más despacio que al utilizar MCRI y/o MCRC. Al comparar los resultados de los criterios MCRI y MCRC, se aprecia, de nuevo, la convergencia numérica de las estimaciones finales obtenidas con ambos procedimientos.

Las Figuras 1 a 5, muestran el promedio de las estimaciones obtenidas con los tres métodos recursivos para los distintos tamaños muestrales utilizados. Con ellas, se intenta resumir gráficamente la información contenida en las Tablas 1 y 2.

Por último, en las Tablas 3 a 7 se ofrecen las medias de los errores cuadráticos medios (ECM) resultantes de la estimación de los parámetros de cada modelo considerado, cuando se utilizan alternativamente los criterios MCRK, MCRI y MCRC. Como puede observarse, en todos los casos, el ECM de cada coeficiente estimado con los criterios MCRC y MCRI es considerablemente menor que el obtenido con el criterio MCRK.

5. Conclusiones.

Las principales conclusiones del trabajo son las siguientes:

- En un procedimiento recursivo, la estimación final de los parámetros puede depender mucho de las condiciones iniciales. Sin embargo, en la mayoría de las aplicaciones económicas, apenas se cuenta con información acerca de los valores de los parámetros, por lo que las condiciones iniciales suelen ser arbitrarias. Por esta razón, conviene utilizar algoritmos para los que se disponga de condiciones iniciales exactas, como es el caso del criterio MCRI o bien, algoritmos que eliminen de alguna forma la influencia de condiciones iniciales arbitrarias sobre la estimación final de los parámetros, como es el caso del método MCRC.

- La estimación recursiva es particularmente sensible a errores de redondeo que se acumulan en el proceso recursivo y pueden llegar a invalidar completamente los resultados. Por ello, la estabilidad numérica debe ser un aspecto crucial a la hora de elegir un algoritmo de estimación recursiva. El algoritmo MCRI es robusto numéricamente, mientras que el método MCRK tiene problemas numéricos en situaciones que se presentan habitualmente en la práctica econométrica.

- Se ha demostrado que el estimador MCRK puede interpretarse como un estimador cresta, mientras que el estimador MCRI coincide analíticamente con el MCO. Por tanto, bajo las hipótesis habituales, el estimador MCRK es sesgado aunque consistente mientras que el MCRI es insesgado.

- Se ha demostrado también que la divergencia entre las estimaciones finales obtenidas con MCRK y MCRI, depende exclusivamente de la magnitud de una matriz denotada como W_t. El tamaño de W_t depende positivamente del mal condicionamiento de la matriz $X_t^T X_t$ y del valor del escalar que se añade a la diagonal principal de la misma en las condiciones iniciales habituales del método MCRK.

- Los resultados con datos simulados muestran que, incluso para muestras suficientemente grandes y problemas bien condicionados, pueden producirse sesgos importantes en la estimación MCRK. Sin embargo, independientemente del tamaño muestral disponible, el estimador MCRI proporciona siempre estimaciones finales de los parámetros que coinciden exactamente con los resultados obtenidos con MCO. Por otra parte, el estimador MCRC genera estimaciones finales que convergen numéricamente a los obtenidos con MCRI/MCO.

- Si existe una gran incertidumbre inicial acerca del valor de los parámetros y sólo se desea la estimación final de los mismos, la mejor alternativa para estimarlos recursivamente será MCRI. La razón es que este criterio presenta importantes ventajas frente al algoritmo MCRK: a) independencia de condiciones iniciales, b) su robustez numérica y c) la convergencia exacta de sus resultados finales a los obtenidos con MCO para toda la muestra.

- Cuando tenga interés obtener la senda de evolución de los parámetros estimados, así como su matriz de covarianzas en cada instante, puede ser más conveniente utilizar el criterio MCRC. Este criterio tiene la ventaja, con respecto al MCRI, de que se estiman directamente los parámetros originales y que, por tanto, no es necesario invertir la matriz de información cada vez que se añade un nuevo dato, con el fin de obtener una senda de evolución de las estimaciones de dichos parámetros.

- Por último, los métodos MCRK y MCRI pueden combinarse. Así, es posible utilizar el criterio MCRI para obtener condiciones iniciales estimadas por MCO con un grado de libertad y después, actualizar las estimaciones de los parámetros con MCRK.

Tabla 1: Resultados de la simulación del modelo I: $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \epsilon_t$ con 20 realizaciones, siendo $\beta_1 = 0.5$, $\beta_2 = 0.8$, $\beta_3 = -0.3$ y $var(\epsilon_t) = 10^{-8}$.

N	NC	MCRK $\tau=10^6$			MCRI MCO			MCRC $\tau=10^6$		
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
30	1.346	0.148 (0.036)	0.139 (0.043)	-0.092 (0.047)	0.462 (0.121)	0.668 (0.218)	-0.296 (0.171)	0.465 (0.122)	0.666 (0.218)	-0.295 (0.170)
50	1.329	0.169 (0.037)	0.181 (0.051)	-0.106 (0.060)	0.469 (0.100)	0.688 (0.192)	-0.295 (0.155)	0.467 (0.103)	0.688 (0.191)	-0.294 (0.155)
100	1.156	0.257 (0.042)	0.374 (0.048)	-0.181 (0.059)	0.477 (0.078)	0.784 (0.103)	-0.317 (0.114)	0.475 (0.079)	0.784 (0.101)	-0.301 (0.119)
200	1.155	0.322 (0.042)	0.507 (0.039)	-0.216 (0.056)	0.484 (0.063)	0.781 (0.060)	-0.309 (0.078)	0.480 (0.062)	0.778 (0.060)	-0.302 (0.080)
500	1.054	0.409 (0.030)	0.666 (0.033)	-0.258 (0.035)	0.491 (0.036)	0.799 (0.040)	-0.304 (0.042)	0.487 (0.033)	0.800 (0.040)	-0.301 (0.040)

Notas:

- Las cifras entre paréntesis son las desviaciones típicas muestrales del promedio de las estimaciones de los parámetros.
- N: Tamaño de la muestra.
- NC: Número de condición de la matriz $X^T X$.
- MCRK: Mínimos Cuadrados Recursivos.
- MCRI: Filtro de la matriz de información.
- MCRC: Mínimos Cuadrados Corregidos.

Tabla 2: Resultados de la simulación del modelo II: $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \epsilon_t$ con 20 realizaciones, siendo $\beta_1 = 0.5$, $\beta_2 = 0.8$, $\beta_3 = -0.3$ y $var(\epsilon_t) = 10^{-8}$.

N	NC	MCRK $\tau=10^6$			MCRI MCO			MCRC $\tau=10^6$		
		$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
30	26353.38	0.500 (0.000)	0.047 (0.024)	-0.007 (0.032)	0.500 (0.000)	1.015 (0.472)	-0.427 (0.213)	0.500 (0.000)	0.781 (0.060)	-0.404 (0.236)
50	27379.83	0.500 (0.000)	0.070 (0.026)	-0.005 (0.032)	0.500 (0.000)	0.978 (0.391)	-0.403 (0.184)	0.500 (0.000)	0.787 (0.049)	-0.386 (0.196)
100	25173.57	0.500 (0.000)	0.120 (0.029)	-0.021 (0.030)	0.500 (0.000)	0.814 (0.200)	-0.316 (0.095)	0.500 (0.000)	0.789 (0.038)	-0.301 (0.103)
200	24690.29	0.500 (0.000)	0.210 (0.038)	-0.060 (0.022)	0.500 (0.000)	0.821 (0.145)	-0.317 (0.059)	0.500 (0.000)	0.794 (0.031)	-0.309 (0.072)
500	23981.05	0.500 (0.000)	0.373 (0.043)	-0.126 (0.021)	0.500 (0.000)	0.790 (0.091)	-0.299 (0.040)	0.500 (0.000)	0.796 (0.018)	-0.295 (0.045)

Notas: Las cifras entre paréntesis son las desviaciones típicas muestrales del promedio de las estimaciones de los parámetros.

- N: Tamaño de la muestra.
- NC: Número de condición de la matriz X^TX.
- MCRK: Mínimos Cuadrados Recursivos.
- MCRI: Filtro de la matriz de información.
- MCRC: Mínimos Cuadrados Corregidos.

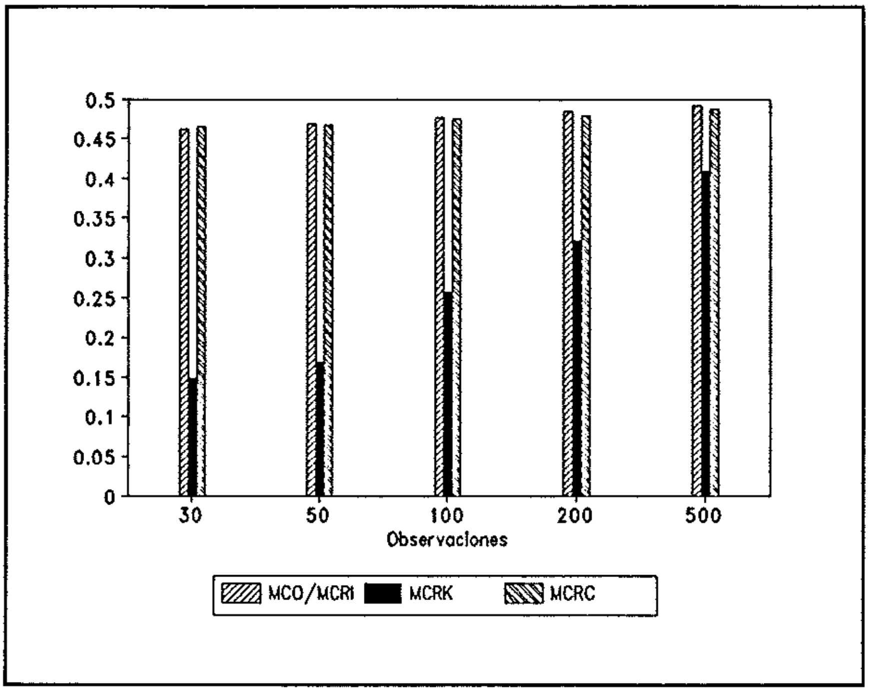

Figura 1: Valor medio de la estimación del parámetro β_1 del modelo I para distintos tamaños muestrales (valor teórico $\beta_1 = 0.5$).

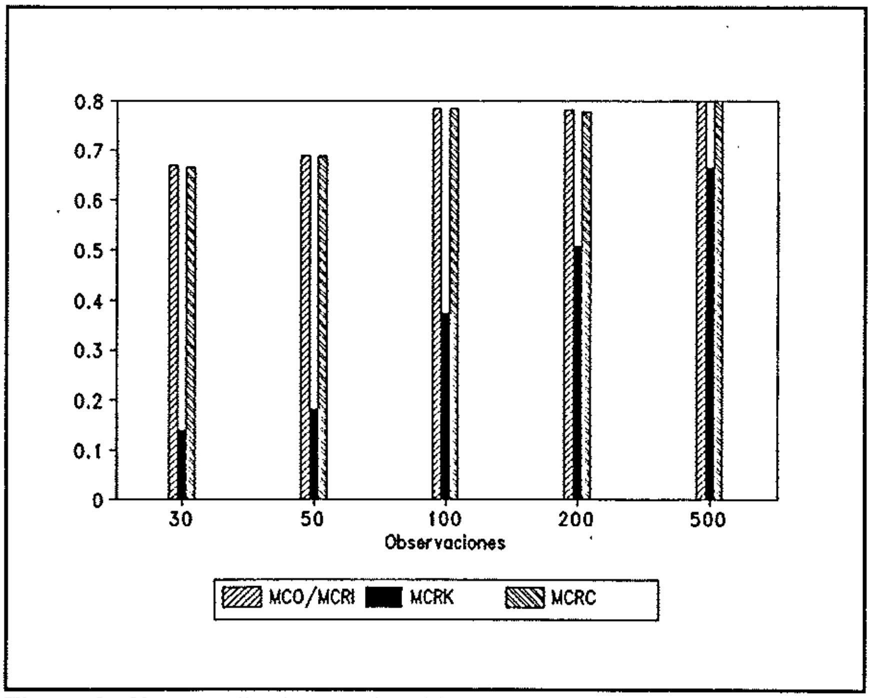

Figura 2: Valor medio de la estimación del parámetro β_2 del modelo I para distintos tamaños muestrales (valor teórico $\beta_2 = 0.8$).

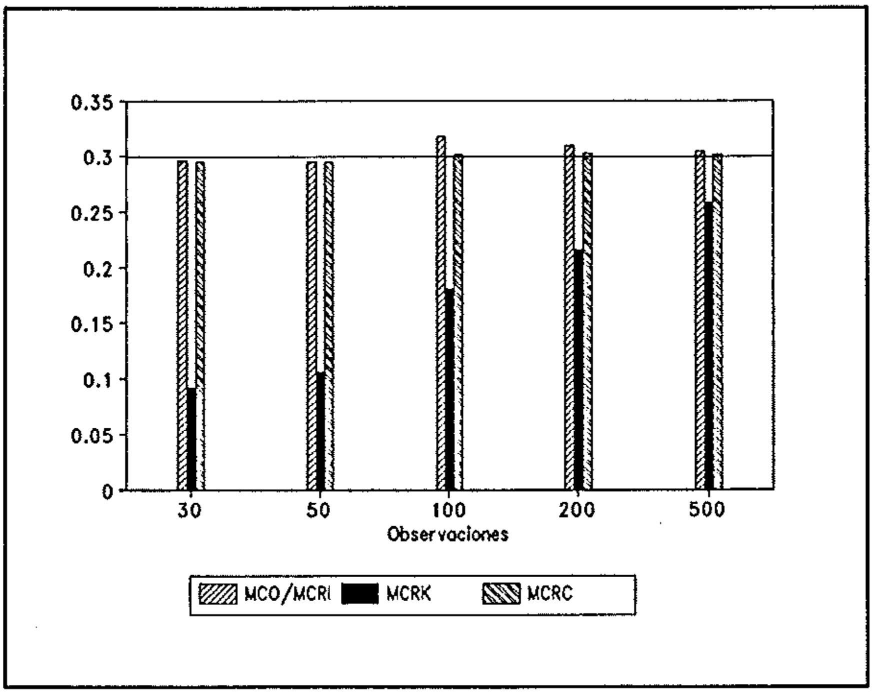

Figura 3: Valor medio de la estimación del parámetro β_3 del modelo I para distintos tamaños muestrales (valor teórico $|\beta_3| = 0.3$).

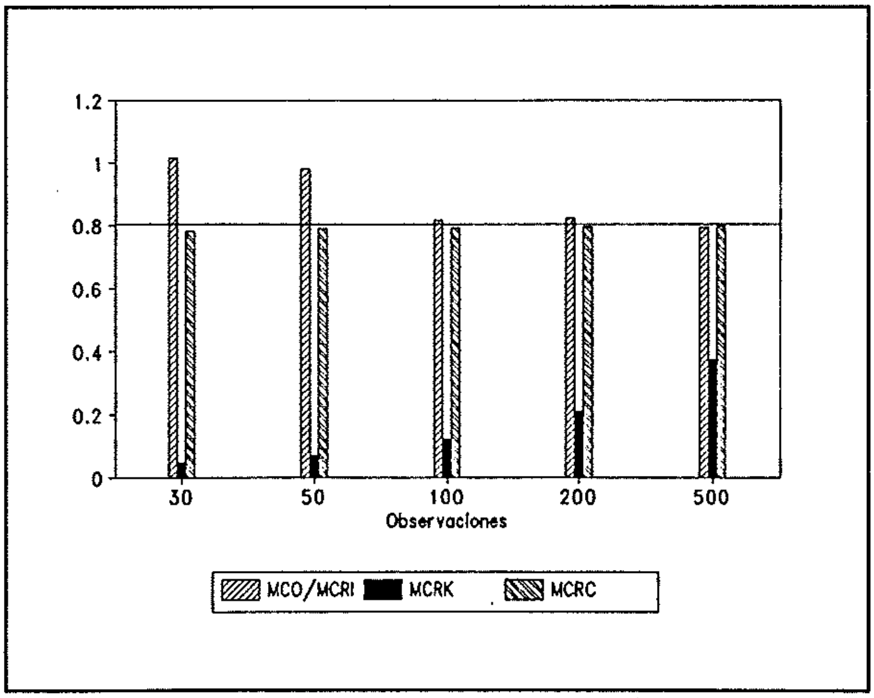

Figura 4: Valor medio de la estimación del parámetro β_2 del modelo II para distintos tamaños muestrales (valor teórico $\beta_2 = 0.8$).

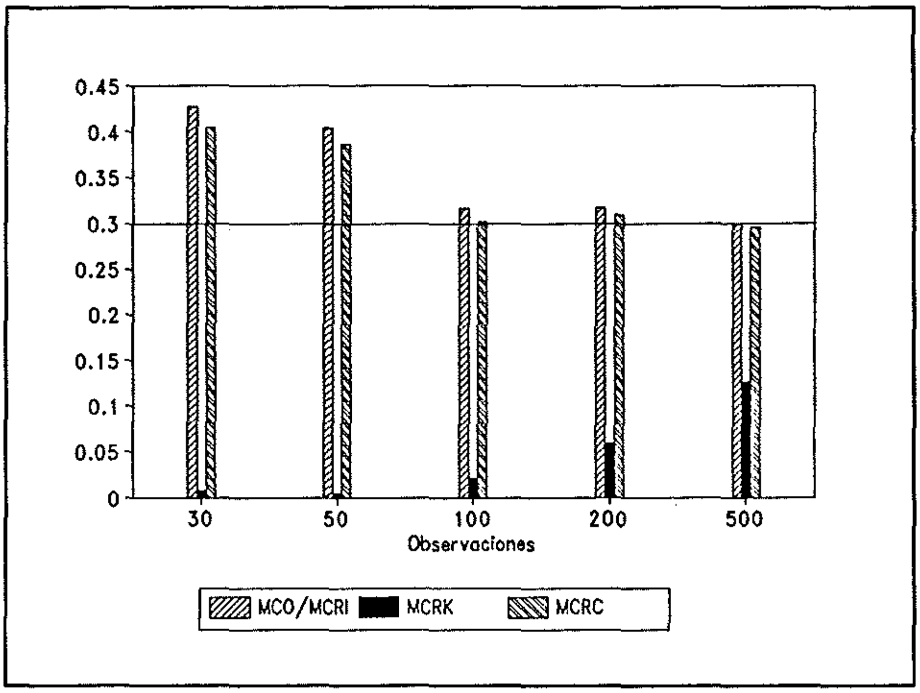

Figura 5: Valor medio de la estimación del parámetro β_3 del modelo II para distintos tamaños muestrales (valor teórico $|\beta_3| = 0.3$).

Tabla 3: ECM resultantes de la estimación del parámetro β_1 del modelo I con los criterios MCRK, MCRI y MCRC.

N	ECM($\hat{\beta}_1$)		
	MCRK	MCRI	MCRC
30	0.125	0.016	0.016
50	0.111	0.011	0.012
100	0.061	0.007	0.007
200	0.033	0.004	0.004
500	0.009	0.001	0.001

Tabla 4: ECM resultantes de la estimación del parámetro β_2 del modelo I con los criterios MCRK, MCRI y MCRC.

N	ECM($\hat{\beta}_2$)		
	MCRK	MCRI	MCRC
30	0.439	0.065	0.065
50	0.386	0.049	0.049
100	0.184	0.011	0.011
200	0.087	0.004	0.004
500	0.019	0.002	0.002

Tabla 5: ECM resultantes de la estimación del parámetro β_3 del modelo I con los criterios MCRK, MCRI y MCRC.

N	ECM($\hat{\beta}_3$)		
	MCRK	MCRI	MCRC
30	0.046	0.029	0.029
50	0.041	0.024	0.024
100	0.018	0.013	0.014
200	0.010	0.006	0.006
500	0.003	0.002	0.002

Tabla 6: ECM resultantes de la estimación del parámetro β_2 del modelo II con los criterios MCRK, MCRI y MCRC.

N	$\text{ECM}(\hat{\beta}_2)$		
	MCRK	MCRI	MCRC
30	0.568	0.269	0.004
50	0.534	0.185	0.003
100	0.463	0.040	0.002
200	0.350	0.022	0.001
500	0.184	0.008	0.000

Tabla 7: ECM resultantes de la estimación del parámetro β_3 del modelo II con los criterios MCRK, MCRI y MCRC.

N	$\text{ECM}(\hat{\beta}_3)$		
	MCRK	MCRI	MCRC
30	0.087	0.062	0.067
50	0.088	0.044	0.046
100	0.079	0.009	0.010
200	0.058	0.004	0.005
500	0.031	0.002	0.002

Notas:

[1] Para un estudio detallado de este filtro ver Anderson y Moore (1979).

[2] Puede verse una demostración de **(10)** en el Apéndice A.

[3] Estas ecuaciones están pensadas para eliminar de la estimación final de los parámetros el efecto de una o varias observaciones. Por ejemplo, podrían usarse para descontar el efecto de observaciones atípicas y/o influyentes. La aplicación de este filtro en nuestro contexto supone considerar como observaciones influyentes las generadas por condiciones iniciales arbitrarias. Puede encontrarse una demostración de las ecuaciones **(27)**-**(29)** en el Apéndice C.

[4] Error en términos de la desviación entre el promedio de las estimaciones y el verdadero valor del parámetro.

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

Bibliografía

Anderson, B.D.O. y J.B. Moore (1979). *Optimal filtering*. Prentice-all, Inc., Englewood Cliffs, N.J.

Anderson, J.E. (1981). "Ridge estimation of house value determinants". *Journal of Urban Economics*, nº9, pág. 286-297.

Bock, M.E. (1975). "Minimax estimators of the mean of a multivariate distribution". *Annals of Statistics*, nº3, pág. 209-218.

Cooley, T.F, B. Rosenberg y K.D. Wall (1977). "A note on optimal smoothing for time varying coefficient problems". *Annals of Economic and Social measurement*, 6, nº4, pág. 453-456.

Garbade, K. (1977). "Two methods for examining the stability of regression coefficients". *Journal of the American Statistical Association*, vol. 72, nº357, pág. 54-63.

Gruber, M.H.J. (1990). *Regression estimators. A comparative study*. Academic Press, INC., New York.

Hoerl, A.E. y R.W. Kennard (1970). "Ridge regression: biased estimation of nonorthogonal problems". *Technometrics*, vol. 12, pág. 55-67.

Jazwinsky, A.H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Ljung, S. y L. Ljung (1985). "Error propagation properties of recursive least-squares adaptation algorithms". *Automatica*, vol. 21, pág. 157-167.

Marquardt, D.W. y R.D. Snee (1975). "Ridge regression in practice". *American Statistician*, nº29, pág. 3-20.

Schmidt, P. (1976). *Econometrics*. Marcel Dekker, Inc., New York.

Young, P. (1984). *Recursive estimation and time-series analysis. An introduction*. Springer-Verlag, Heidelberg.

Apéndice A:

La expresión (10) puede demostrarse por inducción. En el instante $t=1$, la ecuación (4) puede escribirse como:

$$\hat{\beta}_1 = (I - k_1 x_1^T)\hat{\beta}_0 + k_1 y_1 \tag{A.1}$$

Teniendo en cuenta que $k_1 = V_1 x_1$ y $(I - k_1 x_1^T) = V_1 V_0^{-1}$, la expresión (A.1) queda como:

$$\hat{\beta}_1 = V_1 V_0^{-1}\hat{\beta}_0 + V_1 x_1 y_1 = V_1(V_0^{-1}\hat{\beta}_0 + x_1 y_1) \tag{A.2}$$

Sustituyendo (5) en (6) y particularizando en el instante $t=1$, se obtiene:

$$V_1 = V_0 - V_0(x_1^T V_0 x_1 + 1)^{-1} x_1 V_0 \tag{A.3}$$

y aplicando el lema de inversión de matrices[5] a la expresión (A.3):

$$V_1 = (V_0^{-1} + x_1 x_1^T)^{-1} \tag{A.4}$$

Por tanto, (A.2) queda como:

$$\hat{\beta}_1 = (V_0^{-1} + x_1 x_1^T)^{-1}(V_0^{-1}\hat{\beta}_0 + x_1 y_1) \tag{A.5}$$

En el instante $t=2$, la ecuación (4) será:

$$\begin{aligned}\hat{\beta}_2 &= (I - k_2 x_2^T)\hat{\beta}_1 + k_2 y_2 = V_2 V_1^{-1}(V_1 V_0^{-1}\hat{\beta}_0 + V_1 x_1 y_1) + V_2 x_2 y_2 = \\ &= V_2(V_0^{-1}\hat{\beta}_0 + x_1 y_1 + x_2 y_2)\end{aligned} \tag{A.6}$$

donde:

$$V_2 = (V_1^{-1} + x_2 x_2^T)^{-1} = (V_0^{-1} + x_1 x_1^T + x_2 x_2^T)^{-1} \tag{A.7}$$

y por tanto, (A.6) queda como:

$$\hat{\beta}_2 = (V_0^{-1} + x_1 x_1^T + x_2 x_2^T)^{-1}(V_0^{-1}\hat{\beta}_0 + x_1 y_1 + x_2 y_2)$$

Por inducción, después de haber procesado t observaciones, se obtendrá que:

$$\hat{\beta}_t = (V_0^{-1} + \sum_{i=1}^{t} x_i x_i^T)^{-1}(V_0^{-1}\hat{\beta}_0 + \sum_{i=1}^{t} x_i y_i)$$

Otra forma de demostrar (10) consiste en utilizar el resultado de que las ecuaciones del algoritmo MCRK son algebraicamente equivalentes a las del MCRI. De este último filtro sabemos que $\hat{\beta}_t = V_t\, \hat{a}_t^*$ y sustituyendo (35) y (36) en la expresión anterior, se obtiene:

$$\hat{\beta}_t = (V_0^{-1} + \sum_{i=1}^{t} x_i x_i^T)^{-1} (\hat{a}_0^* + \sum_{i=1}^{t} x_i y_i)$$

o bien:

$$\hat{\beta}_t = (V_0^{-1} + \sum_{i=1}^{t} x_i x_i^T)^{-1} (V_0^{-1} \hat{\beta}_0 + \sum_{i=1}^{t} x_i y_i)$$

Apéndice B:

La consistencia del estimador cresta ($\hat{\beta}_c$) se demuestra expresándolo en función del estimador MCO ($\hat{\beta}_{MCO}$)

$$\hat{\beta}_c = (\xi I_k + X^T X)^{-1} X^T X \hat{\beta}_{MCO}$$

y tomando límites en probabilidad a ambos lados de la anterior ecuación, se obtiene:

$$plím\hat{\beta}_c = plím\left[\frac{X^T X + \xi I_k}{N}\right]^{-1} plím\left[\frac{X^T X}{N}\right] plím\hat{\beta}_{MCO} = \beta$$

donde se ha tenido en cuenta que:

$$plím\hat{\beta}_{MCO} = \beta$$

$$plím\frac{I_k}{N} = 0$$

$$plím\frac{X^T X}{N} = \Sigma_{xx}$$

siendo Σ_{xx} una matriz no singular.

Apéndice C:

Sea $\hat{\beta}_k$ la estimación por MCO del vector de parámetros β:

$$\hat{\beta}_k = (X_k^T X_k)^{-1} X_k^T y_k \qquad \textbf{(C.1)}$$

denotamos por $\hat{\beta}_k^*$ a la estimación por MCO de β cuando se elimina la observación k-s de la uestra:

$$\hat{\beta}_k^* = (X_k^T X_k - x_{k-s} x_{k-s}^T)^{-1} (X_k^T y_k - x_{k-s} y_{k-s}) \qquad \textbf{(C.2)}$$

plicando el lema de inversión de inversión de matrices a la expresión (C.2) y llamando $V_k = (X_k^T X_k)^{-1}$, se obtiene que:

$$\hat{\beta}_k^* = [V_k + V_k x_{k-s} (1 - x_{k-s}^T V_k x_{k-s})^{-1} x_{k-s}^T V_k] [X_k^T y_k - x_{k-s} y_{k-s}]$$

bien:

$$\hat{\beta}_k^* = \hat{\beta}_k + V_k x_{k-s} (1 - x_{k-s}^T V_k x_{k-s})^{-1} (x_{k-s}^T \hat{\beta}_k - y_{k-s})$$

Llamando $k_k^* = V_k x_{k-s} (x_{k-s}^T V_k x_{k-s} - 1)^{-1}$, la expresión anterior puede escribirse como:

$$\hat{\beta}_k^* = \hat{\beta}_k + k_k^* (y_{k-s} - x_{k-s}^T \hat{\beta}_k)$$

y aplicando, de nuevo, el lema de inversión de matrices, la matriz de covarianzas de $\hat{\beta}_k^*$ tiene la expresión:

$$V_k^* = V_k + V_k x_{k-s} (1 - x_{k-s}^T V_k x_{k-s})^{-1} x_{k-s}^T V_k = (I - k_k^* x_{k-s}^T) V_k$$