Qüestiió - V. 7, n.º 3 (setembre 1983) pp. 467-477

UN ALGORITMO PARA LA ESTIMACIÓN MAXIMOVEROSÍMIL DE MODELOS ECONOMÉTRICOS DE DESEQUILIBRIO DE LADO CORTO

CÉSAR MOLINAS SANS

UNIVERSIDAD DE BARCELONA

La estimación maximoverosimil de modelos econométricos de desequilibrio de lado corto presenta dificultades debido a la no acotación de la función de verosimilitud. En este artículo se describen brevemente dicho tipo de modelos y se propone un sencillo procedimiento, basado en el algoritmo E.M., para su estimación por máxima verosimilitud.

AN ALGORITHM FOR MAXIMUM LIKELIHOOD ESTIMATION OF SHORT SIDE DISEQUILIBRIUM ECONOMETRIC MODELS.

Keywords: MAXIMUM LIKELIHOOD, ESTIMATION, ECONOMETRIC MODELS, SHORD SIDE DISEQUILIBRIUM, E.M. ALGORITHM.

1. INTRODUCCION.

En este trabajo se propone un algoritmo, basado en el algoritmo E.M. de Dempster, Laird y Rubin (1977), para la estimación maximoverosímil de los parámetros de modelos econométricos de desequilibrio de lado corto. Dicho algoritmo, al consistir en una sucesión de regresiones ponderadas es de instrumentación muy sencilla en el marco de los paquetes estadísticos disponibles en la mayoría de los centros de cálculo ; concretamente, se estudia aquí su instrumentación mediante el Time Series Processor (TSP). Con ligeras modificaciones, el algoritmo resulta también útil para estimar modelos de variables latentes, regresiones oscilantes, etc., que tengan funciones de verosimilitud no acotadas.

En la Sección 2 se introducen los modelos econométricos de desequilibrio de lado corto, se deduce su función de verosimilitud y se comenta su no acotación, originada por la existencia de singularidades; así mismo se establece el paralelismo con otros modelos econométricos que presentan una problemática estadística similar. En la Sección 3 se introduce el algoritmo E.M. revisándose brevemente sus principales propiedades. En la Sección 4 se propone un algoritmo de la clase E.M. para la estimación de los modelos expuestos en la Sección 2 y, finalmente, en el Apéndice, se estudia su ejecución mediante el TSP.

2. MODELOS ECONOMETRICOS DE DESEQUILIBRIO.

La demanda y la oferta de un bien en un mercado determinado se representan, habitualmente, mediante ecuaciones del tipo

- César Molinas Sans - Departament d'Estadística i Econometria de la Universitat de Barcelona Av. Diagonal, s/n. - Barcelona

- Article rebut el Juliol del 1983.

$$D_t = \alpha_1 P_t + \beta_1' X_{1t} + u_{1t}$$
$$S_t = \alpha_2 P_t + \beta_2' X_{2t} + u_{2t} \quad (1)$$

en donde D_t y S_t son, respectivamente las cantidades demandada y ofrecida del bien de dicho mercado en el tiempo t, P_t es su precio en dicho tiempo, X_{1t} y X_{2t} son sendos vectores de observaciones de variables predeterminadas en t y u_{1t}, u_{2t} son términos de perturbación aleatorios que se supondrán de media cero, normalmente distribuidos y no correlacionados ni en el tiempo ni entre sí.

Bajo la hipótesis de equilibrio, los precios se mueven con suficiente agilidad para igualar, en todo momento, la oferta con la demanda, de modo que si Q_t denota la cantidad del bien que ha sido objeto de transacción en el tiempo t, se tiene

$$Q_t = D_t = S_t \quad (2)$$

Las ecuaciones (1) y (2) constituyen un modelo de ecuaciones simultáneas en el que las variables endógenas, observables ambas, son Q_t y P_t. Dicho modelo es estimable por las técnicas econométricas habituales.

La hipótesis de equilibrio (2) resulta poco satisfactoria en diversas ocasiones, concretamente en aquellas en las que los precios, por condicionantes institucionales o de otro tipo, no reaccionan de manera instantánea para igualar ambos lados del mercado; se producen entonces, situaciones de exceso de demanda o de oferta que pueden modelizarse de maneras diversas (véase /2/ para una sistematización del tratamiento de dichos modelos); se tratarán aquí, exclusivamente, los llamados modelos de desequilibrio de lado corto, en los que se supone que Q_t es la mínima entre las cantidades ofrecida y demandada.

$$Q_t = \min (D_t, S_t) \quad (3)$$

es decir, que la cantidad observada Q_t coincide con el lado corto del mercado.

Las ecuaciones (1) y (3) constituyen un modelo econométrico de desequilibrio que es llamado habitualmente el "modelo canónico". Desde un punto de vista estadístico, sus propiedades son muy diferentes de las del modelo de equilibrio compuesto por (1) y (2). Las variables D_t y S_t dejan de ser observables, puesto que la muestra de cantidades está constituida por observaciones de las que a priori no se sabe si han sido generadas por la ecuación de demanda o por la de oferta. Esta es una situación de "datos incompletos", como las caracterizadas por Dempster, Laird y ubín (1977), en la que existen dos espacios muestrales X e Y y una aplicación exhaustiva entre ellos

$$X \xrightarrow{\alpha} Y$$

Los elementos de X, en nuestro caso serían los pares demanda-oferta (D_t, S_t), no son observables, sino que solamente se observan los elementos de Y, que en nuestro caso serían las cantidades observadas

$$Q_t = \min (D_t, S_t).$$

Designado por $f(X \mid \theta)$ a la familia de densidades de probabilidad, dependientes del vector de parámetros θ, de los elementos de X, podemos derivar la densidad de probabilidad del observable $y \in Y$ como

$$g(y|\theta) = \int_{\alpha^{-1}(y)} f(X|\theta)\, d x \quad (4)$$

con la integral extendida sobre todos aquellos $x \epsilon X$ que pueden producir el observable y. En el modelo canónico de desequilibrio (4) puede escribirse como

$$g(Q_t|\theta) = \int_{Q_t}^{\infty} f(Q_t,S_t|\theta)\, d S_t + \int_{Q_t}^{\infty} f(D_t,Q_t|\theta)\, d D_t \quad (5)$$

que es un punto de partida conveniente para escribir la función de verosimilitud de los observables Q_t. Prescindiendo, en la notación, de la explicitación de la dependencia respecto a θ, se puede escribir (5) como

$$g(Q_t) = g(Q_t/D_t<S_t) \text{ prob } (D_t<S_t) + g(Q_t/S_t\leq D_t)\ (1 - \text{prob}(D_t<S_t)) \quad (6)$$

En virtud de la definición de probabilidad -

condicionada,

$$g(Q_t/D_t<S_t) = \frac{\int_{Q_t}^{\infty} f(Q_t,S_t)\, dS_t}{\text{prob } (D_t<S_t)} \qquad (7)$$

y, teniendo presente la normalidad y la independencia de u_{1t} y de u_{2t}, podemos escribir, denotando por σ_1^2 y σ_2^2 sus varianzas respectivas,

$$\int_{Q_t}^{\infty} f(Q_t,S_t)\, dS_t =$$

$$\int_{Q_t}^{\infty} \frac{1}{2\Pi\sigma_1\sigma_2} \exp \left\{ -\frac{1}{2\sigma_1^2} (Q_t-\alpha_1P_t-\beta_1'X_{1t})^2 - \right.$$

$$\left. -\frac{1}{2\sigma_2^2} (S_t-\alpha_2P_t-\beta_2'X_{2t})^2 \right\} dS_t = \qquad (8)$$

$$= f_1(Q_t)\, F_2(Q_t)$$

en donde

$$f_1(Q_t) = \frac{1}{\sqrt{2\Pi}\sigma_1} \exp \left\{ -\frac{1}{2\sigma_1^2} (Q_t-\alpha_1P_t-\beta_1'X_{1t})^2 \right\}$$

$$F_2(Q_t) = \int_{Q_t}^{\infty} \frac{1}{\sqrt{2\Pi}\sigma_2} \exp \left\{ -\frac{1}{2\sigma_2^2} (S_t-\alpha_2P_t-\beta_2'X_{2t})^2 \right\} dS_t \qquad (9)$$

de modo que (7) puede escribirse como

$$g(Q_t/D_t< S_t) = \frac{f_1(Q_t)F_2(Q_t)}{\text{prob}(D_t\ S_t)} \qquad (10)$$

Por un procedimiento similar al seguido para llegar a (10) se obtiene

$$g(Q_t/D_t\geqslant S_t) = \frac{f_2(Q_t)F_1(Q_t)}{1-\text{prob}(D_t\ S_t)} \qquad (11)$$

en donde

$$f_2(Q_t) = \frac{1}{\sqrt{2\Pi}\sigma_2} \exp \left\{ -\frac{1}{2\,\sigma_2^2} (Q_t-\alpha_2P_t-\beta_2'X_{2t})^2 \right\}$$

$$F_1(Q_t) = \int_{Q_t}^{\infty} \frac{1}{\sqrt{2\Pi}\sigma_1} \exp \left\{ -\frac{1}{2\sigma_1^2} (D_t-\alpha_1P_t-\beta_1'\ X_{1t})^2 \right\} dD_t \qquad (12)$$

Sustituyendo (10) y (11) en (6), se obtiene

$$g(Q_t) = f_1(Q_t)F_2(Q_t)+f_2(Q_t)F_1(Q_t)$$

con lo que la función de verosimilitud es, para una muestra de tamaño T,

$$L = \prod_{t=1}^{T} g(Q_t) =$$

$$= \prod_{t=1}^{T} f_1(Q_t)F_2(Q_t)+f_2(Q_t)F_1(Q_t) \qquad (13)$$

que es la propuesta por Maddala y Nelson (1974).

La función de verosimilitud (13) no está acotada, puesto que presenta singularidades contenidas en las variedades del espacio de parámetros definidas por $\sigma_1^2 = 0$ y por $\sigma_2^2 = 0$. Para ponerlo de manifiesto, escribimos (13) como

$$L = \prod_{t=1}^{T} \left\{ \frac{1}{\sqrt{2\Pi}\sigma_1} \exp\left[-\frac{1}{2} \frac{(Q_t-\alpha_1 P_t-\beta_1' X_{1t})^2}{\sigma_1^2} \right] \left[1-\Phi\left(\frac{Q_t-\alpha_2 P_t-\beta_2' X_{2t}}{\sigma_2}\right) \right] + \right.$$

$$\left. + \frac{1}{\sqrt{2\Pi}\sigma_2} \exp\left[-\frac{1}{2} \frac{(Q_t-\alpha_2 P_t-\beta_2' X_{2t})^2}{\sigma_2^2} \right] \left[1-\Phi\left(\frac{Q_t-\alpha_1 P_t-\beta_1' X_{1t}}{\sigma_1}\right) \right] \right\} \quad (14)$$

en donde Φ denota la función acumulativa de distribución N(0,1). Se puede, ahora, hacer saltar L al infinito del siguiente modo. Para $1 \leq k \leq T$, escogemos α_1 y β_3 de manera que

$$Q_k = \alpha_1 P_k + \beta_1' X_{1k}$$

Para valores de α_2, β_2 y σ_2^2 que no anulen o hagan infinitos a $F_2(Q_t)$ y a $f_2(Q_t)$, tomemos una sucesión de valores de σ_t^2, que tienda a cero. En este caso $f_1(Q_k)$ tiende a infinito y es suficiente haber escogido α_1 y β_1 de manera que

$$Q_t-\alpha_1 P_t-\beta_1' X_{1t}<0$$

o sea

$$\Phi\left(\frac{Q_t-\alpha_1 P_t-\beta_1' X_{1t}}{\sigma_1}\right)<1$$

para todo $t \neq k$ para que L tienda a infinito, puesto que para $t = k$ $g(Q_t)$ es infinito y para $t \neq k$ los $g(Q_t)$ son valores finitos.

La función de verosimilitud tiene, por tanto, infinitas singularidades, puesto que el proceso anterior puede llevarse a cabo para infinitas combinaciones de valores de los parámetros del modelo. Este hecho implica dificultades técnicas importantes para el cálculo efectivo de los estimadores maximoverosimiles de los parámetros a partir de la maximización de (13) puesto que cuanto más precisa sea la estimación, menores serán las varianzas y más cerca se estará de una singularidad; en la práctica, los algoritmos iterativos de maximización habituales muestran una fuerte tendencia a meterse en la "chimenea" o a quedarse en la frontera del espacio de parámetros en el caso de que éste se acote previamente para evitar las singularidades.

Las pecularieades de la función de verosimilitud de los modelos de desequilibrio son compartidas por otros modelos de interés estadístico y econométrico. Kiefer, /5/, propone una clase general de modelos de la cual los modelos de desequilibrio, regresión oscilante, Tobit, Probit y Logit son casos particulares. La problemática de la estimación maximoverosimil de los modelos canónicos de desequilibrio no es patrimonio exclusivo suyo sino que puede reproducirse en los otros modelos de la clase general.

3. EL ALGORITMO E.M.

El algoritmo E.M., propuesto por Dempster, Laird y Rubín /1/, compite ventajosamente con los métodos iterativos tradicionales de maximización de la función de verosimilitud en los casos que generan situaciones de "datos incompletos" y es capaz de superar obstáculos computacionales tales como la no acotación de la función de verosimilitud o la potencial singularidad de la matriz de información. Con la notación de la Sección 2ª, el algoritmo E.M. utiliza la familia de densidades $f(x/\theta)$ para hallar el valor de que maximiza la función $g(y/\theta)$. El algoritmo consiste en la iteración de dos etapas: una etapa "esperanza" y una etapa "maximización" (nótese que de las iniciales de las dos etapas proviene el nombre del algoritmo). A grandes rasgos, el funcionamiento del algoritmo en la iteración n-ésima es el siguiente. Dado $\theta^{(n-1)}$, valor de θ obtenido en la iteración (n-1), se calcula la esperanza de un conjunto de estadísticos suficientes de los datos completos; a continuación, la segunda etapa consiste en calcular el valor $\theta^{(n)}$ que maximiza la esperanza anteriormente computada.

Una definición del algoritmo E.M. es la siguiente. Denotemos por Ω el espacio de parámetros y definamos la función

$$\tilde{K}(\theta|\theta)=E(\log f(X|\theta)| \quad y,\theta) \quad (15)$$

Asumamos la existencia de K para todos los pares $(\theta,\tilde{\theta})$ de $\Omega\times\Omega$. La n-ésima iteración del algoritmo procede así: en primer lugar se calcula $K(\theta|\theta^{(n-1)})$ (etapa "esperanza") y, en segundo lugar se calcula el valor $\theta^{(n)}$ que maximiza $K(\theta|\theta^{(n-1)})$.

Llamemos $L(\theta)$ al logaritmo de la función de verosimilitud

$$L(\theta) = \log g(y|\theta) \tag{16}$$

Definamos la función de densidad condicional de x , dados θ e y, como

$$h(x|y,\theta) = \frac{f(x|\theta)}{g(y|\theta)} \tag{17}$$

así como la función

$$R(\tilde{\theta}|\theta) = E(\log h(x/y,\tilde{\theta}) \quad |y,\theta) \tag{18}$$

Nótese que, en virtud de la definición de h, se tiene que

$$K(\tilde{\theta}|\theta) = L(\theta) + R(\tilde{\theta}|\theta) \tag{19}$$

Llamaremos algoritmo E.M. generalizado (E.M.G.) a una aplicación

$$\Omega \xrightarrow{\quad M \quad} \Omega$$

tal que, para todo $\theta\epsilon\Omega$ se verifique que

$$K(M(\theta)|\theta) \geq K(\theta|\theta) \tag{20}$$

En este sentido, una sucesión de valores del vector de parámetros $\theta^{(0)}, \theta^{(1)}, \ldots, \theta^{(n)}, \ldots$ puede considerarse como resultado de la aplicación reiterada de M al valor inicial $\theta^{(0)}$.

Los siguientes teoremas establecen el marco general de convergencia del algoritmo E.M. hacia los estimadores maximoverosímiles de θ . Las demostraciones, que aquí se emiten, pueden hallarse en /1/.

El Teorema 1 establece que los valores de la función de verosimilitud asociados a la sucesión $\theta^{(n)}$ forman una sucesión monótona creciente y el Corolario caracteriza a los estimadores maximoverosímiles como puntos fijos del algoritmo E.M.

<u>TEOREMA 1</u>. Sea M un algoritmo E.M.G.; entonces, para todo $\theta\epsilon\Omega$ se tiene que $L(M(\theta))\geq L(\theta)$ con igualdad si y sólo si $K(M(\theta)|\theta)=K(\theta|\theta)$ y $(x|y,M(\theta))=h(x|y,\theta)$.

<u>Corolario</u>. Si existe algún $\theta^*\epsilon\Omega$ tal que $L(\theta^*)\geq L(\theta)$ para todo $\theta\epsilon\Omega$, entonces

a) $L(M(\theta^*)) = L(\theta^*)$

b) $K(M(\theta^*)|\theta^*) = K(\theta^*|\theta^*)$

c) Si, además, $L(\theta^*)>L(\theta)$ para $\theta^*\neq\theta$, entonces $M(\theta^*) = \theta^*$

El teorema 2 establece condiciones suficientes para la convergencia del algoritmo; la primera de ellas hace referencia a la acotación de la sucesión de valores de la función de verosimilitud para los $\theta^{(n)}$, condición necesaria para su convergencia (nótese que no se requiere la acotación de la función de verosimilitud). La segunda condición está íntimamente relacionada con el Teorema 3.

<u>TEOREMA 2</u>. Sea la sucesión $\theta^{(n)}$, n=1,2,... obtenida por aplicación reiterada de un algoritmo E.M.G. Supóngase que

a) La sucesión $L(\theta^{(n)})$ está acotada

b) Existe un $\lambda\epsilon R$ tal que, para todo n,

$$K(\theta^{(n+1)}|\theta^{(n)}) - -K(\theta^{(n)}|\theta^{(n)})\geq\lambda(\theta^{(n+1)}-\theta^{(n)})'(\theta^{(n+1)}-\theta^{(n)})$$

entonces la sucesión $\theta^{(n)}$ converge hacia algún θ^* en la clausura de Ω.

Suponiendo las habituales condiciones de regularidad de las funciones K,L,R y M y la permutabilidad de las operaciones esperanza y diferenciación puede demostrarse el

<u>TEOREMA 3.</u> Si

$$\frac{\partial}{\partial\theta} K(\theta^{(n+1)}|\theta^{(n)}) = 0$$

entonces, para todo n, existe un punto $\theta_0^{(u+1)}$ en el segmento que une a $\theta^{(u)}$ con $\theta^{(u+1)}$ tal que

$$K(\theta^{(n+1)}|\theta^{(n)} - K(\theta^{(n)}|\theta^{(n)}) =$$

$$=-(\theta^{(n+1)}-\theta^{(n)})'\frac{\partial^2}{\partial\theta\partial\theta'} K(\theta_0^{(n+1)}|\theta^{(n)})(\theta^{(n+1)}|\theta^{(n)})$$

(Nótese que para que se cumpla la segunda hipótesis del Teorema 2 es suficiente que la matriz de segundas derivadas de K sea negativa definida con valores propios que no estén dentro de un entorno lo suficientemente pequeño de cero).

El Teorema 4 establece condiciones suficientes para la convergencia hacia los estimadores maximoverosímiles.

TEOREMA 4. Sea $\theta^{(n)}$ una sucesión obtenida mediante la iteración de un algoritmo E.M.G. Supongamos que

a) $\theta^{(n)}$ converge hacia un θ^* de la clausura de Ω.

b) $\frac{\partial}{\partial\theta} K(\theta^{(n+1)}|\theta^{(n)}) = 0$

c) $\frac{\partial^2}{\partial\theta\partial\theta'} K(\theta^{(n+1)}|\theta^{(n)})$ es negativa definida y sus valores propios están todos fuera de un entorno lo suficientemente pequeño de cero.

Se verifica entonces que

a) $\frac{\partial}{\partial\theta} L(\theta^*) = 0$

b) $\frac{\partial^2}{\partial\theta\partial\theta'} K(\theta^*|\theta^*)$ es negativa definida.

El comportamiento del algoritmo E.M. frente a funciones de verosimilitud no acotadas no queda aclarado por los teoremas anteriores, en los que la acotación de la sucesión $L(\theta^{(n)})$ es una hipótesis esencial. La capacidad del algoritmo para converger hacia los estimadores maximoverosímiles en tales situaciones ha sido puesta de manifiesto en experimentos de Montecarlo referenciados (/3/), en el contexto de la estimación de mezclas de distribuciones normales. Nuestra experiencia con modelos econométricos de desequilibrio es que el algoritmo E.M. propuesto en la Sección 4ª. converge sin problemas, mientras que métodos iterativos del tipo Newton-Raphson (utilizando las primeras y segundas derivadas propuestas por Maddala y Nelson /6/) y del tipo quasi-Newton (concretamente el algoritmo de Gill, Murray y Pitfield) no convergen incluso para valores iniciales de los parámetros muy cercanos a los estimadores maximoverosímiles del modelo.

4. UN ALGORITMO E.M. PARA LA ESTIMACION DEL MODELO CANONICO DE DESEQUILIBRIO.

Para comodidad de notación, volvamos a escribir el modelo compuesto por las ecuaciones (1) y (3) incluyendo la variable precio como una más dentro de las matrices X_1 y X_2.

$$\begin{aligned} D_t &= \beta_1' X_{1t} + u_{1t} \\ S_t &= \beta_2' X_{2t} + u_{2t} \qquad (21) \\ Q_t &= \min(D_t, S_t) \end{aligned}$$

El modelo de desequilibrio genera una situación de "datos incompletos" en la que, partiendo de los datos completos (D_t, S_t) se observa sólo Q_t.

En la situación de datos completos, la función de densidad conjunta de (D_t, S_t) es

$$f(D_t, S_t) = f_1(D_t) \cdot f_2(S_t) \qquad (22)$$

en donde f_1 y f_2 se definen como en (9) y en (12) respectivamente. La función de densidad condicional, dado el valor observado Q_t, es

$$h(D_t, S_t/Q_t) = \frac{f(D_t, S_t)}{g(Q_t)} \qquad (23)$$

en donde

$$g(Q_t) = f_1(Q_t)F_2(Q_t) + f_2(Q_t)F_1(Q_t) \qquad (24)$$

pudiendo escribir

$$h(D_t, S_t|Q_t) = \begin{cases} \dfrac{f(Q_t, S_t)}{g(Q_t)} & \text{si } D_t < S_t \\[2ex] \dfrac{f(D_t, Q_t)}{g(Q_t)} & \text{si } D_t > S_t \end{cases} \qquad (25)$$

La función de verosimilitud en la situación de datos completos es

$$L^*(\theta) = \sum_{t=1}^{T} \left[\log f_1(D_t) + \log f_2(S_t)\right] \qquad (26)$$

y, dada una sucesión $\theta^{(n)}$, podemos escribir la expresión equivalente a (15) relativa a $\theta^{(n-1)}$ como

$$K(\theta|\theta^{(n-1)}) = E^{(n-1)}\left[L(\theta^*)/Q\right] =$$

$$= \sum_{t=1}^{T} \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left[\log f_1(D_t)+\log f_2(S_t)\right] h^{(n-1)}(D_t,S_t|Q_t)\, dD_t\, dS_t \qquad (27$$

en donde los superíndices (n-1) indican que las funciones han sido calculadas con el valor $\theta^{(n-1)}$ de los parámetros.

Teniendo presente (25), podemos escribir

$$K(\theta|\theta^{(n-1)}) =$$

$$= \sum_{t=1}^{T} \int_{Q_t}^{\infty} \left[\log f_1(Q_t)+\log f_2(S_t)\right] \frac{f_1^{(n-1)}(Q_t)\, f_2^{(n-1)}(S_t)}{g^{(n-1)}(Q_t)}\, dS_t +$$

$$+ \sum_{t=1}^{T} \int_{Q_t}^{\infty} \left[\log f_1(D_t)+\log f_2(Q_t)\right] \cdot \frac{f_1^{(n-1)}(D_t)\,.\, f_2^{(n-1)}(Q_t)}{g^{(n-1)}(Q_t)}\, dD_t =$$

$$= \sum_{t=1}^{T}\left[-\frac{1}{2}\log 2\pi-\frac{1}{2}\log\sigma_1^2-\frac{1}{2\sigma_1^2}(Q_t-\beta_1'X_{1t})^2\right]\frac{f_1^{(n-1)}(Q_t)}{g^{(n-1)}(Q_t)}\int_{Q_t}^{\infty} f_2^{(n-1)}(S_t)\, dS_t +$$

$$+ \sum_{t=1}^{T}\int_{Q_t}^{\infty} \log f_2(S_t)\frac{f_1^{(n-1)}(Q_t)\, f_2^{(n-1)}(S_t)}{g^{(n-1)}(Q_t)}\, dS_t +$$

$$+ \sum_{t=1}^{T}\left[-\frac{1}{2}\log 2\pi-\frac{1}{2}\log\sigma_2^2-\frac{1}{2\sigma_2^2}(Q_t-\beta_2'X_{2t})^2\right]\frac{f_2^{(n-1)}(Q_t)}{g^{(n-1)}(Q_t)}\int_{Q_t}^{\infty} f_1^{(n-1)}(D_t)\, dD_t +$$

$$+ \sum_{t=1}^{T}\int_{Q_t}^{\infty} \log f_1(D_t)\,\frac{f_1^{(n-1)}(D_t)\, f_2^{(n-1)}(Q_t)}{g^{(n-1)}(Q_t)}\, dD_t \qquad (28)$$

Definamos

$$W_1^{(n-1)}(Q_t) = \frac{f_1^{(n-1)}(Q_t)}{g^{(n-1)}(Q_t)}\int_{Q_t}^{\infty} f_2^{(n-1)}(S_t)\, dS_t =$$

$$= \frac{f_1^{(n-1)}(Q_t)\ F_2^{(n-1)}(Q_t)}{g^{(n-1)}(Q_t)}$$

$$W_2^{(n-1)}(Q_t) = \frac{f_2^{(n-1)}(Q_t)}{g^{(n-1)}(Q_t)} \int_{Q_t}^{\infty} f_1^{(n-1)}(D_t)\, dD_t =$$

$$= \frac{f_2^{(n-1)}(Q_t)\ F_1^{(n-1)}(Q_t)}{g^{(n-1)}(Q_t)}$$

(Nótese que $W_1^{(n-1)}(Q_t) + W_2^{(n-1)}(Q_t) = 1$).

Las condiciones de primer orden para la maximización de $K(\theta/\theta^{(n-1)})$ son, a partir de (28)

$$\frac{\partial}{\partial \beta_1} K(\theta|\theta^{(n-1)}) =$$

$$= \frac{1}{\sigma_1^2} \sum_{t=1}^{T} W_1^{(n=1)}(Q_t)(Q_t - \beta_1' X_{1t}).X_{1t} = 0 \qquad (29)$$

$$\frac{\partial}{\partial \beta_2} K(\theta|\theta^{(n-1)}) =$$

$$\frac{1}{\sigma^2} \sum_{t=1}^{T} W_2^{(n-1)}(Q_t)(Q_t - \beta_2' X_{2t}).S_{2t} = 0 \qquad (30)$$

$$\frac{\partial}{\partial \sigma_1^2} K(\theta|\theta^{(n-1)}) =$$

$$= \frac{1}{2\sigma_1^4} \sum_{t=1}^{T} W_1^{(n-1)}(Q_t) \left[\sigma_1^2 - (Q_t - \beta_1' X_{1t})^2\right] = 0$$

$$\frac{\partial}{\partial \sigma_2^2} K(\theta|\theta^{(n-1)}) =$$

$$= \frac{1}{2\sigma_2^4} \sum_{t=1}^{T} W_2^{(n-2)}(Q_t) \left[\sigma_2^2 - (Q_t - \beta_2' X_{2t})^2\right] = 0$$

Es decir, la maximización de $K(\theta/\theta^{(n-1)})$ consiste en las dos regresiones ponderadas (29) y (30) con matrices de ponderación.

$$W_1^{(n-1)} = \text{diag} \quad W_1^{(n-1)}(Q_1), \ldots, W_1^{(n-1)}(Q_T)$$

$$W_2^{(n-1)} = \text{diag} \quad W_2^{(n-1)}(Q_1), \ldots, W_2^{(n-1)}(Q_T)$$

y a la estimación de la varianza de los residuos ponderados de dichas regresiones.

El algoritmo E.M. para la estimación del modelo canónico de desequilibrio queda, pues, configurado como

- Etapa esperanza: Dado el vector

$$\theta^{(n-1)} = (\beta_1^{(n-1}, \beta_2^{(n-1)}, \sigma_1^{2\,(n-1)}, \sigma_2^{2\,(n-1)})$$

calcular las matrices de ponderación $W_1^{(n-1)}$ y $W_2^{(n-2)}$.

- Etapa maximización: Dadas $W_1^{(n-1)}$ y $W_2^{(n-2)}$ estimar

$\beta_1^{(n)}$, $\beta_2^{(n)}$, $\sigma_1^{2\,(n)}$, $\sigma_2^{2\,(n)}$ mediante

$$\beta_1^{(n)} = (X_1' \ W_1^{(n-1)} \ X_1)^{-1} \ X_1' \ W_1^{(n-1)} \ Q$$

$$\beta_2^{(n)} = (X_2' \ W_2^{(n-1)} \ X_1)^{-1} \ X_1' \ W_2^{(n-1)} \ Q$$

$$\sigma_1^{2\,(n)} = \frac{1}{\sum_{t=1}^{T} W_1^{(n-1)}(Q_t)} (Q - X_1\beta_2^{(n)})' W_1^{(n-1)} (Q - X_1\beta_1^{(n)})$$

$$\sigma_2^{2\,(n)} = \frac{1}{\sum_{t=1}^{T} W_2^{(n-1)}(Q_t)} (Q - X_2\beta_2^{(n)})' W_2^{(n-1)} (Q - X_2\,{}_1^{(n)})$$

en donde Q, X_1 y X_2 denotan, respectivamente, los vectores de observaciones de Q_t y las matrices de observaciones de X_{1t} y X_{2t} para -- t=1,...,T.

Hay que señalar que el algoritmo E.M. no proporciona directamente una estimación de la -

matriz de varianzas y covarianzas asintótica de los estimadores maximoverosímiles. Esta puede calcularse, sin embargo, de manera indirecta sustituyendo las estimaciones de los parámetros en las segundas derivadas de la función de verosimilitud, calculadas éstas como en Maddala y Nelson /6/.

5. CONCLUSIONES.

Los modelos econométricos de desequilibrio de lado corto presentan dificultades para ser estimados por máxima verosimilitud por los métodos iterativos habituales debido a la no acotación de la función de verosimilitud. Dichos modelos generan situaciones de "datos incompletos" para las cuales el algoritmo E.M. de Dempster, Laird y Rubín /1/ es un instrumento versátil y potente para calcular los estimadores maximoverosímiles. Dicho algoritmo, en el caso de los modelos de desequilibrio, tiene una forma particularmente sencilla, puesto que se reduce a una sucesión de regresiones ponderadas en las que los regresores son las variables originales del modelo y en las que, en cada iteración, se actualizan sólo las matrices de ponderación.

Aunque no está demostrado que el algoritmo E.M. converja siempre en el caso de funciones de verosimilitud no acotadas, en modelos de desequilibrio la limitada experiencia que poseemos de él parece indicarlo así. Una posible razón para ello es el carácter derivado que tiene, en cada iteración, la estimación de las varianzas de las perturbaciones, que se estiman mediante las varianzas de los residuos de las regresiones. Esto, aparentemente, mantiene el algoritmo lo suficientemente alejado de los hiperplanos que contienen las asíntotas, $\sigma_1^2 = 0$ y $\sigma_2^2 = 0$, para que la convergencia pueda tener lugar. La posible existencia de múltiples máximos locales es un problema que el algoritmo E.M. comparte con los métodos habituales de optimización, por lo que resulta prudente comenzar las iteraciones con distintos valores iniciales. Nuestra experiencia va en el sentido de que los estimadores por Mínimos Quadrados son un buen punto de partida (nótese que ésto equivale a tomar las matrices de ponderación iguales a la identidad en la primera iteración).

6. APENDICE.

EJECUCION DEL ALGORITMO E.M. MEDIANTE EL T.S.P.

El Time Series Processor (T.S.P.) es un instrumento habitual en la investigación económica aplicada. La única dificultad, a priori, que surge al intentar plantear el algoritmo descrito en la Sección 4 en el T.S.P. es la imposibilidad de calcular directamente funciones de distribución o de acudir desde el programa a paquetes de subrutinas tales como el Scientific Subroutine Package (S.S.P.) que sí ofrecen esta posibilidad.

Concretamente, en cada iteración del algoritmo E.M. se hacen necesarios los cálculos de las áreas bajo la curva normal $F_1^{(n)}(Q_t)$ y $F_2^{(n)}(Q_t)$. La alternativa más sencilla, si no es necesario un grado de aproximación superior a 10^{-6}, es usar la misma aproximación al área bajo la normal que en la subrutina NDTR del S.S.P. Si denotamos por $f(X)$ la función de densidad normal y por $F(X)$ la correspondiente función de distribución, se tiene, para $X \geq 0$

$$F(X) = 1 - f(X) \sum_{i=1}^{5} a_i W^i$$

en donde

$$W = \frac{1}{1+PX}$$

$$P = 0.2316491$$

$$a_1 = 0.3193815$$

$$a_2 = -0.3565638$$

$$a_3 = 1.781478$$

$$a_4 = -1.821256$$

$$a_5 = 1.330274$$

Si se desea un grado mayor de aproximación hay que recurrir, simultaneamente, a un desarrollo de Mac Laurin y a un desarrollo asintótico de la distribución normal puesto que el primero de ellos sólo converge con rapidez para valores pequeños del argumento mientras que el segundo sólo lo hace para valores grandes del mismo (véase /4/ pp.136 y 138 para una discusión de los mencionados desarrollos). El desarrollo de Mac Laurin no plantea especiales problemas y se necesitan ocho términos para superar la aproximación de 10^{-6}. El desarrollo asintótico viene dado en función de la razón de Mills, la cual admite a su vez un desarrollo en fracción contínua que puede aproximarse mediante las reducidas correspondientes (véase /7/ pp 344-348) de fácil cálculo recurrente. Se necesitan, no obstante, las veinte primeras reducidas para conseguir aproximaciones del orden de 10^{-6}.

El siguiente programa produce k iteraciones del algoritmo.

```
- LOAD $
- SMPL 1 T $
- SET IT = 1 $
- GENR POP1 = 1 $
- GENR POP2 = 1 $
- 1000 PLOTS $
- OLSQ (WEIGHT = POP 1, UNNORM)
        Q,c,X11,...,Xik1 $
- RETRV CRF1 33 T $
- RETRV COF1 1 23 $
- GENR RESI = Q - CRF 1 $
- GENR RESI2 = RESI**2 $
- INPROD T 11 RESI2 POPI VAR1 $
- INPROD T 11 POPI UNO DIV1 $
- SET s1 = (VAR1 / DIV1) ** .5 $
  . . . . . . . . . . . . . . .
  . . . . . . . . . . . . . . .
- GENR S2 = (VAR2/DIV2) **.5 $
- GENR F14 = (.398942/S1(1)) * EXP ((RESI **2)/
  (-.5 * S1(1) ** 2)) $
- GENR F24 = (.398942/S2(1)) * EXP ((RES2 **2)/
  (-.5 * S2(1) ** 2)) $

- GENR H1 = RES1/S1 $
- GENR H2 = RES2/S2 $
- GENR H1G = (H1.GE.0) *H1 $
- GENR H1P = ABS((H1.LT.0)*H1) $
- GENR H2G = (H2.GE.0)*H2 $
- GENR H2P = ........... $
- GENR DH1G = .398942 * EXP(-.5(H1G**2)) $
- GENR DH1P = .398942* EXP(-.5(H1P**2)) $
- GENR DH2G = ......................... $
- GENR DH2P = ......................... $
- GENR FF1G = 1/(1+ (.2316419*H1G)) $
- GENR FF1P = ..................... $
- GENR FF2G = ..................... $
- GENR FF2P = ..................... $
- GENR FZ1G = 1-DH1G (.3193815 FF1G-.3565638
    (FF1G  2) 1.781478 (FFIG  3)-1.821256
    (FF1G  4) 1.330274 (FF2  5) $
- GENR FZ1P = .............................
              .............................
              ................ $
- GENR FZ2G = .............................
              .............................
              ................ $
- GENR FZ2P = .............................
              .............................
              ................ $
- GENR FZ1 = FZ1G + FZ1P -.5 $
- GENR FZ2 = FZ2G + FZ2P -.5 $
- GENR H = F14 * FZ1 + F24 * FZ2 $
- GENR POP1 = FZ1 * F14/H $
- GENR POP2 = FZ2 * F24/H $
- SET IT = IT 1
- IF IT .LT. K $
- THEN $
- GO TO 1000 $
- STOP $
- END $
```

7. BIBLIOGRAFIA.

/1/ DEMPSTER,A., LAIRD, N.M., RUBIN,D.B. --- (1977): "Maximum Likelihood from Incom-- plete Data via the E.M. Algorithm". Journal of the Royal Statistical Society, -- sre.B., vol. 39 pp. 1-38.

/2/ GARCIA, J. y MOLINAS, C. (1981): "El tratamiento econométrico de los modelos de desequilibrio de lado corto". Documento de trabajo, Departamento de Estadística y Econometría de la Universidad de Barcelona.

/3/ HARTLEY, M.J. (1978): "Estimating Mixtures of Normal Distributions: A comment". Journal of the American Statistical Association, vol 73 pp.738-741.

/4/ KENDALL, E.J. y STUART, A. (1969): "The Advanced Theory of Statistics" vol. 1,3ª ed. Griffin, Londres.

/5/ KIEFER, N.M. (1977): "Models of Switch-- ing Regression, Disequilibrium and Endogenous Structural Change and the Value - of Information". Report 7757, Center of Mathematical Studies in Bussiness and -- Economics. University of Chicago.

/6/ MADDALA, G.S. y NELSON, F.D. (1974): "Maximum Likelihood Methods for Models of Markets in Disequilibrium", Econometrica, vol. 42, pp. 1013-1030.

/7/ REY PASTOR, J. PI CALLEJA, P. TREJO,C. - "Análisis matemático", vol. 1. Kapelusz, Buenos Aires.