Inferencia en modelos dinámicos uniecuacionales con variables integradas*

Juan J. Dolado

Servicio de Estudios, Banco de España

Javier Andrés

Departamento de Análisis Económico, Universidad de Valencia

Rafael Domenech

Departamento de Análisis Económico, Universidad de Valencia

1. Introducción

En este trabajo se discute bajo qué condiciones se puede llevar a cabo inferencia correcta, utilizando procedimientos estándar, al estimar modelos dinámicos condicionales por mínimos cuadrados ordinarios, en el caso en que las variables tengan el carácter de integradas y estén o no estén cointegradas. Nuestra principal preocupación reside en examinar la forma en que la validez de la forma condicional en la estimación afecta los diferentes procedimientos de inferencia, dado que este modo de estimar modelos uniecuacionales constituye una práctica ordinaria en econometría empírica. En el marco tradicional de modelos de regresión lineal con series temporales estacionarias, la verificación del supuesto de exogeneidad débil por parte de las variables explicativas, resulta ser condición necesaria y suficiente para la aplicación válida de los procedimientos habituales de inferencia. Sin embargo, cuando algunas de dichas variables explicativas son integradas, la existencia de un vector de cointegración y su forma de estimación puede alterar las propiedades asintóticas de los estimadores obtenidos, así como la condición requerida para poder efectuar la estimación en el modelo condicional.

En concreto, nos concentramos, a través del uso de diferentes ejemplos, en los méritos relativos de los procedimientos bietápicos (véase Engle y Granger (1987)) y unietápicos (véase Banerjee et al. (1986) y Sims et al. (1986)) de estimación del vector de cointegración. Nuestro análisis favorece la forma

* Este articulo ha sido escrito para el número especial de esta revista dedicado a «Cointegración», coordinado por Alvaro Escribano, al que agradecemos sus comentarios. Parte de su contenido recoge resultados de la investigación en curso llevada a cabo por el primer autor junto con Neil R. Ericsson, de la Reserva Federal, a quien estamos muy agradecidos por su colaboración, así como a los comentarios de numerosos colegas.

unietápica en base a su normalidad asintótica. De forma similar se analizan las propiedades asintóticas de los contrastes de cointegración bajo la hipótesis nula de ausencia de cointegración. El análisis se centra en dos contrastes: i) en el t-ratio de Dickey-Fuller que contrasta la existencia de una raíz unitaria en los residuos derivados de la relación estática de cointegración y ii) en el t-ratio del término de corrección del error en un modelo dinámico reparametrizado como un mecanismo de corrección del error (ECM) (véase Engle y Granger (1987)). El comportamiento de ambos contrastes en muestras finitas viene ilustrado a través de un pequeño estudio de Monte Carlo, que muestra la superioridad del test ECM.

Resulta importante señalar en este punto que la validez de los resultados obtenidos se encuentra restringida a la estimación de modelos en donde las variables se interpreten en desviaciones de sus componentes determinísticos (medias, tendencias, etc.). Si los componentes determinísticos se modelizan conjuntamente con los estocásticos, algunos de estos resultados pueden verse alterados (véase, por ejemplo, West (1988) y Park y Phillips (1988)).

La estructura organizativa del papel es la siguiente. En la sección 2 se analizan las propiedades de un modelo bivariante con una relación condicional estática, donde la variable explicativa es o bien endógena o bien débil o fuertemente exógena, de acuerdo con la terminología tradicional de Engle et al. (1983). En la sección 3 se extiende el análisis anterior a modelos condicionales dinámicos. En la sección 4 se analiza la ventaja comparativa de utilizar diferentes contrastes de cointegración bajo la hipótesis nula de no cointegración (tamaño) y bajo la hipótesis alternativa de cointegración (potencia). En la sección 5 se presenta un ejercicio ilustrativo de Monte Carlo que sirve para analizar el comportamiento en muestras finitas de las propiedades asintóticas de los contrastes analizados previamente. Finalmente, la sección 6 presenta un conjunto de conclusiones.

2. Inferencia en sistemas cointegrados: El modelo estático

Dado que el marco estadístico que a continuación se analiza ha sido descrito en forma extensiva por numerosos autores (véase, por ejemplo, Hendry y Richard (1982, 1983) y Hendry y Ericson (1987)) sólo se ofrecerá a continuación una exposición sucinta del mismo que sirva como soporte a la discusión posterior. En la breve exposición se subraya como cualquier modelo empírico condicional es una representación derivada del proceso generador de los datos, destacando las etapas más importantes en el proceso de reducción de la función de probabilidad conjunta de los datos.

Comenzaremos por definir la función de densidad conjunta o proceso de generación de los datos (DGP) de un vector, x_t, de realizaciones de una variable aleatoria n-dimensional, como $D(x_t/I_{t-1}, \theta)$ donde $\theta \in \Theta$ es un vector identificable de parámetros desconocidos pertenecientes al interior de un espacio paramétrico de dimensión finita; I_{t-1} denota el conjunto de información disponible en el momento $(t - 1)$. Si se considera una muestra de T observaciones de x_t, la función de densidad conjunta de los datos puede factorizarse secuencialmente en la forma siguiente

$$D(x_1, \ldots, x_T/X_0, \theta) = \prod_{t=1}^{T} D(x_t/X_{t-1}, \theta) \tag{1}$$

donde $X_j = (X_0, x_1, \ldots, x_j)$ y X_0 representa el conjunto de condiciones iniciales.

Se supone que, después de las transformaciones oportunas (e.g. logaritmos, ratios), las funciones condicionales de densidad en (1) presentan la forma funcional común

$$x_t/X_{t-1} \sim N(\mu_t, \Omega) \quad \text{con} \quad \mu_t = E(x_t/X_{t-1}, \theta) = \sum_{i=i}^{m} \pi_i x_{t-i} \tag{2}$$

donde $x_t - \mu_t = u_t$ es una «innovación» obtenida por construcción dado el desfase máximo m. Por ejemplo, si los agentes económicos forman planes contingentes basados en información limitada, de manera que tales planes definan relaciones de comportamiento como las sugeridas por la teoría económica, la representación estocástica de dicha conducta dará lugar a modelos condicionales. Por tanto, si se particiona x_t en $(y_t, z_t')'$, podemos factorizar secuencialmente la función de densidad conjunta de los datos en t de la siguiente forma

$$D(x_t/X_{t-1}, \theta) = D_1(y_t/z_t, X_{t-1}, \phi_1)D_2(z_t/X_{t-1}, \phi_2) \tag{3}$$

donde (ϕ_1, ϕ_2) representa una reparametrización apropiada de θ.

Sea ψ el conjunto de parámetros de interés. De acuerdo con las definiciones de Engle et al. (1983), si ψ es una función de ϕ_1 y los parámetros ϕ_1 y ϕ_2 no tienen elementos en común («variación libre»), entonces diremos que z_t es *débilmente exógena* con respecto a ψ. Bajo dicha condición puede llevarse a cabo inferencia eficiente sobre ψ en la distribución condicional D_1, a la que denotaremos como *modelo*, sin tener que modelar la distribución marginal D_2. Si además verificarse la *exogeneidad débil*, se cumple la condición

$$D_2(z_t/X_{t-1}, \phi_2) \equiv D_2(z_t/Z_{t-1}, \phi_2)(t = 1, \ldots, T) \tag{4}$$

de forma que y «no causa» a z en sentido de Granger, entonces diremos que z_t es *fuertemente exógena* con respecto a ψ.

Una vez que el marco estadístico conceptual ha sido establecido convenientemente comenzaremos analizando un simple ejemplo que contiene los ingredientes básicos del caso más general a tratar posteriormente. Se trata de un DGP bivariante donde las dos variables presentes (y_t, z_t), en desviaciones de sus componentes determinísticos, son integradas de orden uno, $I(1)$, se encuentran cointegradas y el vector de cointegración sólo aparece en la primera ecuación. Algunos de estos supuestos se relajarán posteriormente, ofreciendo generalizaciones de los resultados en el caso multivariante. Una versión que contiene un modelo condicional estático viene dado por el siguiente ejemplo

EJEMPLO 1

$$y_t = \beta z_t + \varepsilon_{1t} \tag{5a}$$

$$\varepsilon_t \sim NI[0, \Omega], \Omega = \begin{pmatrix} \sigma_1^2 & \rho\,\sigma_1\sigma_2 \\ & \sigma_2^2 \end{pmatrix}$$

$$\Delta z_t = \varepsilon_{2t}; z_0 = 0 \tag{5b}$$

Es fácil observar que tanto z_t como y_t son procesos $I(1)$ y además están cointegrados con un vector de cointegración $(1, -\beta)$.

Si se estima (5a) mediante mínimos cuadrados ordinarios (MCO), se obtiene

$$T(\hat{\beta} - \beta) = \frac{\Sigma\, z_t \varepsilon_{1t}}{\Sigma\, z_t^2} \frac{T^2}{T} \tag{6}$$

La formulación (6) posibilita la derivación de la distribución asintótica de $\hat{\beta}$, cuyos detalles pueden analizarse en Phillips (1987) y Stock (1987). Para obtener dicha derivación es conveniente condicionar $(y_t - \beta z_t)$ sobre Δz_t, o equivalentemente ε_{1t} sobre ε_{2t}, obteniendo

$$\varepsilon_{1t} = \gamma\varepsilon_{2t} + \varepsilon_{3t} \quad ; \quad \gamma = \rho\sigma_1/\sigma_2 \quad ; \quad \sigma_3^2 = (1 - \rho^2)\sigma_1^2 \tag{7}$$

Si se definen $S_{it} = \sum_{j=1}^{t} \varepsilon_{ij}\, (i = 1, 2, 3)$ como los procesos integrados formados a partir de las sumas parciales de las perturbaciones $\varepsilon_i's$, entonces, cuando $T \uparrow \infty$, los siguientes momentos muestrales tienden a las siguientes variables aleatorias, definidas en términos de procesos de Wiener $(B_i(i = 1, 2, 3))$ con varianzas unitarias (véase, e.g., Phillips (1987))[1]

$$\begin{aligned} &\text{(i)} \quad T^{-1}\, \Sigma\, S_{it}\varepsilon_{it} \to \sigma_i^2[\textstyle\int B_i dB_i + 1](i = 1, 2, 3) \\ &\text{(ii)} \quad T^{-1}\, \Sigma\, S_{it}\varepsilon_{jt} \to \sigma_i\sigma_j \textstyle\int B_i dB_j\, (i \neq j = 1, 2, 3) \\ &\text{(iii)} \quad T^{-2}\, \Sigma\, S_{it}S_{jt} \to \sigma_i\sigma_j \textstyle\int B_i B_j\, (i, j = 1, 2, 3) \end{aligned} \tag{8}$$

donde «→» denota convergencia débil en distribución y los límites de las integrales son (0, 1), siendo omitidos en las derivaciones posteriores. Utilizando las expresiones relevantes de (8) en (6), se obtiene

$$T(\hat{\beta} - \beta) = \frac{\Sigma\, S_{2t}(\gamma\varepsilon_{2t} + \varepsilon_{3t})}{\Sigma\, S_{2t}^2} \frac{T^2}{T} \to \gamma \frac{\int B_2 dB_2 + 1}{\int B_2^2} + \frac{\sigma_3 \int B_2 dB_3}{\sigma_2 \int B_2^2} \tag{9}$$

---

[1] En general, si $\varepsilon_t \sim \text{nid}\,(0, \sigma_\varepsilon^2)$, definimos la sucesión de sumas parciales $S_T(r) = T^{-1/2}\sum_{i}^{[Tr]} \varepsilon_i$, donde $[x]$ representa la parte entera de x, y $r \in [0, 1)$. Se puede demostrar que $S_T(r) \to B(r)$, donde $B(r)$ es un proceso de Wiener con varianza σ_ε^2. Por simplicidad, los procesos $B(r)$ se escriben sin el argumento r en la notación del texto.

Así pues, $T(\hat{\beta} - \beta)$ presenta una distribución asintótica no degenerada y por tanto $\hat{\beta}$ tiende a β a la tasa $\mathrm{Op}(1/T)$. Se trata, por consiguiente, de un estimador «super-consistente» de β (véase Stock (1987)) cuya distribución asintótica no es normal y ha de ser tabulada mediante procedimientos de simulación numérica. Si $\rho = 0$, entonces z_t es tanto débil como fuertemente exógena con respecto a β en este DGP, lo que implica $\gamma = 0$ y $\varepsilon_1 = \varepsilon_3$, en cuyo caso (9) se reduce a

$$\frac{\sigma_1 \int B_2 dB_1}{\sigma_2 \int B_2^2} \equiv \frac{\sigma_3 \int B_2 dB_3}{\sigma_2 \int B_2^2} \tag{10}$$

Condicionando en el sigma-álgebra $F_j = \sigma[B_j(r);\ r \in [[0, 1)]$, el teorema de la Correspondencia Continua (véase Billingsley (1968) y Phillips y Durlauf (1986)) afirma que cuando B_i y B_j son procesos de Wiener independientes se verifica que

$$\int B_j dB_i \equiv N(0, \int B_j^2) \tag{11a}$$

y, consecuentemente,

$$(\int B_j^2)^{-1/2} \int B_j dB_i \equiv N(0, 1) \tag{11b}$$

Por tanto, de acuerdo con (11b), el t-ratio de $\hat{\beta}$, definido por

$$(\Sigma\, z_t^2)^{1/2}(\hat{\beta} - \beta)/\hat{\sigma}_1 \to N(0, 1)$$

presenta una distribución asintótica estándar.

En general, siempre que en un modelo del tipo anterior, el t-ratio de $\hat{\beta}$ se comporte asintóticamente como la distribución normal estandarizada, diremos que $T(\hat{\beta} - \beta)$ se distribuye asintóticamente como una «mezcla de normales» (MN), notación que se empleará posteriormente.

El ejemplo anterior sirve para poner de manifiesto el hecho de que al menos una de las condiciones para exogeneidad débil en este modelo, i.e. $\rho = 0$, resulta ser condición necesaria y suficiente para lograr normalidad asintótica, y, consecuentemente, para poder aplicar procedimientos estándar de inferencia en dicho modelo. Alternativamente, tal como sugieren Johansen (1988), Johansen y Juselius (1988) y Phillips (1988a), podría estimarse (5a) por máxima verosimilitud (ML). Dicho procedimiento de estimación resulta equivalente a estimar el siguiente modelo mediante MCO

$$y_t = \beta z_t + \gamma \Delta z_t + \varepsilon_{3t} \tag{12}$$

donde el estimador ML de β viene dado por

$$T(\tilde{\beta} - \beta) = \frac{(\Sigma\, z_t \varepsilon_{3t})(\Sigma\, \varepsilon_{2t}^2) - (\Sigma\, z_t \varepsilon_{2t})(\Sigma\, \varepsilon_{2t}\varepsilon_{3t})}{(\Sigma\, z_t^2)(\Sigma\, \varepsilon_{2t}^2) - \Sigma\,(z_t \varepsilon_{2t})^2} \cdot \frac{T^3}{T^2} \tag{13}$$

Puesto que $E(\Sigma\, \varepsilon_{2t}\varepsilon_{3t}) = 0$ y $\Sigma\, z_t\varepsilon_{2t}$ es Op (T), (13) es equivalente al ratio de los primeros términos en el numerador y denominador, obteniéndose

$$T(\tilde{\beta} - \beta) \rightarrow \frac{\sigma_3 \int B_2 dB_3}{\sigma_2 \int B_2^2} \equiv MN \tag{14}$$

que, de acuerdo con el análisis en (10), es una «mezcla de normales». El modelo (12) también puede utilizarse para ilustrar la derivación de la distribución asintótica del estimador γ, asociado a una variable $I(0)$, en una regresión donde aparecen variables $I(1)$. Dicho estimador viene dado por

$$T^{1/2}(\tilde{\gamma} - \gamma) = \frac{(\Sigma\, \varepsilon_{2t}\varepsilon_{3t})(\Sigma\, z_t^2) - (\Sigma\, \varepsilon_{2t}z_t)(\Sigma\, \varepsilon_{3t}z_t)}{(\Sigma\, \varepsilon_{2t}^2)(\Sigma\, z_t^2) - (\Sigma\, z_t\varepsilon_{2t})^2} \cdot \frac{T^3}{T^{5/2}} \rightarrow N(0, \sigma_3^2/\sigma_2^2) \tag{15}$$

dado que $\Sigma\, z_t\varepsilon_{it}$ es Op (T) y $T^{-1/2}(\Sigma\, \varepsilon_{2t}\varepsilon_{3t})$ es Op (1). Así pues, los coeficientes de regresores $I(0)$ presentan distribuciones estándar, independientemente del carácter del resto de las variables explicativas (véase Sims et al. (1986)). Concluimos, por tanto, que la estimación ML consigue la normalidad asintótica cuando la variable explicativa no es débilmente exógena mientras que las variables estacionarias presentes en la ecuación siempre tendrán coeficientes asintóticamente normales. Resulta importante destacar que el primer resultado depende, véase (5b), de la imposición de la raíz unitaria de z_t en el modelo (12). Si dicha raíz unitaria no hubiese sido impuesta, sino estimada, como en el enfoque VAR irrestringido propuesto por Sims et al. (1986), el resultado no se seguiría. Para examinar dicho fenómeno, modificaremos la ecuación (5b) en el DGP (5) de forma que ahora sea

EJEMPLO 2

$$z_t = \alpha z_{t-1} + \varepsilon_{2t} \quad ; \quad \alpha = 1 \tag{5b'}$$

El estimador ML de β en este caso es equivalente al obtenido al estimar por MCO el modelo

$$y_t = \beta z_t + \gamma\varepsilon_{2t} + [\varepsilon_{3t} + \gamma(\hat{\alpha} - 1)z_{t-1}] \tag{16}$$

Por consiguiente,

$$T(\tilde{\beta} - \beta) = \gamma \frac{(\hat{\alpha} - 1)(\Sigma\, z_t z_{t-1})(\Sigma\, \hat{\varepsilon}_{2t}^2)}{(\Sigma\, z_t)(\Sigma\, \hat{\varepsilon}_{2t}) - (\Sigma z_t\hat{\varepsilon}_{2t})^2} \cdot \frac{T^3}{T^2}$$

$$+ \frac{(\Sigma\, z_t\varepsilon_{3t})(\Sigma\, \hat{\varepsilon}_{2t}^2) - (\Sigma\, z_t\varepsilon_{2t})(\Sigma\, \hat{\varepsilon}_{2t}\varepsilon_{3t})}{(\Sigma\, z_t^2)(\Sigma\, \hat{\varepsilon}_{2t}^2) - (\Sigma\, z_t\hat{\varepsilon}_{2t})^2} \cdot \frac{T^3}{T^2} \tag{17}$$

Puesto que $T^{-1}\Sigma z_t\hat{\varepsilon}_{2t} = T^{-1}\Sigma \varepsilon_{2t}\hat{\varepsilon}_{2t} \to \sigma_2^2$, la distribución asintótica de (17) viene dado por

$$T(\tilde{\beta} - \beta) = \gamma \frac{\int B_2 dB_2}{\int B_2^2} + \frac{\sigma_3 \int B_2 dB_3}{\sigma_2 \int B_2^2} \not\equiv MN \tag{18}$$

donde el primer término de (18) recoge la distribución de la raíz unitaria estimada en (5b'), mientras que el segundo término representa una MN, tal como se describía en (10). Así pues, aunque, tal como Sims et al. (1986) demuestran, la estimación del sistema irrestringido no afecta la distribución de las variables que pueden escribirse como estacionarias con media cero, si que afecta la distribución del estimador del vector de cointegración.

Finalmente, obsérvese que la estimación por variables instrumentales (IV) de (5a), dada la endogeneidad de z_t no resulta un procedimiento apropiado para recuperar la normalidad asintótica. Por ejemplo, si z_t es instrumentada mediante z_{t-1}, se obtiene

$$T(\hat{\beta}_{IV} - \beta) = \frac{\Sigma z_{t-1}\varepsilon_{1t}}{\Sigma z_t z_{t-1}} \cdot \frac{T^2}{T} =$$

$$= \frac{\Sigma z_{t-1}(\gamma\varepsilon_{2t} + \varepsilon_{3t})}{\Sigma z_t z_{t-1}} \frac{T^2}{T} \to \frac{\gamma\sigma_2 \int B_2 dB_2 + \sigma_3 \int B_2 dB_3}{\sigma_2 \int B_2^2} \tag{19}$$

cuya distribución asintótica es idéntica a (18). Igualmente, si se utiliza un instrumento w_t correlacionado con z_t pero incorrelacionado con ε_{1t} tal que

$$\Delta w_t = \varepsilon_{4t} \quad ; \quad E(\varepsilon_{2t}\varepsilon_{4s}) \neq 0 \quad ; \quad E(\varepsilon_{1t}\varepsilon_{4s}) = 0$$

puede demostrarse, utilizando la distribución condicional $\Delta w_t = \phi\varepsilon_{2t} + \varepsilon_{3t}$, que

$$T(\hat{\beta}_{IV} - \beta) = \frac{\Sigma w_t\varepsilon_{1t}}{\Sigma w_t x_t} \frac{T^2}{T} \to \frac{\phi\sigma_1\sigma_2 \int B_2 dB_1 + \sigma_1\sigma_3 \int B_3 dB_1}{\phi^2\sigma_2^2 \int B_2^2 + \sigma_2\sigma_3 \int B_2 B_3} \not\equiv MN \tag{20}$$

i.e. (20) tampoco es una MN. Nótese que si $\phi = 0$, i.e. w_t es un instrumento «espúreo», $\hat{\beta}_{IV}$ todavía es un estimador «super-consistente» de β, a diferencia del caso estacionario en que $(\hat{\beta}_{IV} - \beta)$ sigue una distribución de Cauchy carente de momentos (véase Phillips y Hansen (1988)). La validez de instrumentos espúreos en la estimación por IV también se extiende al uso de tendencias deterministicas. En efecto si $\Delta w_t = \mu$ o $w_t = \mu t$ (con $\omega_0 = 0$) tendríamos que[2]

$$T(\hat{\beta}_{IV} - \beta) = \frac{\Sigma t\varepsilon_{1t}}{\Sigma t x_t} \frac{T^{5/2}}{T^{3/2}} \to \frac{N(0, \sigma_1^2/3)}{\sigma_1^2 \int t B_2} \tag{21}$$

[2] Park y Phillips (1988) demuestran que $\Sigma t z_t$ y $\Sigma t\varepsilon_t$ son $Op(T^{5/2})$ y $Op(T^{3/2})$, respectivamente.

de forma que $\hat{\beta}_{IV}$ vuelve a tener el carácter de «super-consistente» pese a no ser MN.

Habiendo llegado a este punto, es importante señalar que en el DGP (5) objeto de análisis, z_t es fuertemente exógena con respecto a β cuando ε_{1t} y ε_{2t} son independientes ($\rho = 0$) lo cual es una simplificación excesiva, si como en el análisis de regresión con series estacionarias, lo que nos interesa es únicamente exogeneidad débil a la hora de hacer inferencia. En efecto, de acuerdo con la discusión en (4), (5b) puede ser generalizado a

$$\Delta z_t = \delta_1(L)\Delta z_{t-1} + \delta_2(L)\Delta y_{t-1} + \varepsilon_{2t} \tag{5b''}$$

donde ahora $\rho = 0$ y $\delta_2(L) \neq 0$ sólo implica exogeneidad débil. En este caso los resultados anteriores se generalizan de acuerdo con la siguiente proposición:

PROPOSICIÓN 1: Si $x_t = (y_t, z_t)'$ está generado por el DGP compuesto por (5a) y (5b''), con $\rho = 0$ y $\delta_i(1) < \infty (i = 1, 2)$, el estimador MCO de β en (5a) tiene la siguiente distribución asintótica

$$T(\hat{\beta} - \beta) \rightarrow [1 - \delta_1(1) - \delta_2(1)\beta] \frac{\sigma_1 \int B_2 dB_1}{\sigma_2 \int B_2^2} \equiv MN \tag{22}$$

si el término entre corchetes es distinto de cero.

DEMOSTRACIÓN: (Véase apéndice.)

Así pues, de acuerdo con (22), la exogeneidad débil es una condición suficiente para conseguir normalidad asintótica. Sin embargo, la exogeneidad débil ordinariamente se contrasta en el trabajo empírico mediante contrastes de la hipótesis nula $H_0 : \rho = 0$ (test de Hausman) que carecen de potencia frente a la posibilidad de que existan dependencias directas entre las funciones condicionales y marginal, dadas por restricciones directas entre subconjuntos de los vectores de parámetros ϕ_1 y ϕ_2 en (3). Para examinar este caso, modificaremos la distribución marginal permitiendo que el mecanismo de corrección del error aparezca tanto en (5a) como en (5b). Si, por ejemplo, $\delta_1(L) = -\delta\beta(1 - L)^{-1}$ y $\delta_2(L) = \delta(1 - L)^{-1}$, entonces se tiene que (5b'') puede representarse como sigue:

EJEMPLO 3

$$\Delta z_t = \delta\varepsilon_{1t-1} + \varepsilon_{2t} = \delta(y_{t-1} - \beta x_{t-1}) + \varepsilon_{2t} \tag{5b'''}$$

En este caso z_t no es débilmente exógena con respecto a β, a pesar de que $\rho = 0$, y consecuentemente la proposición anterior no se cumple (nótese no se verifica que $\delta_1(1)$ y $\delta_2(1)$ sean finitos). En efecto, la distribución asintótica de $\hat{\beta}$ en (5a) ahora será

$$T(\hat{\beta} - \beta) = \frac{\Sigma z_t \varepsilon_{1t}}{\Sigma z_t^2} \frac{T^2}{T} \rightarrow \frac{\delta\sigma_1^2 \int B_1 dB_2 + \sigma_1\sigma_2 \int B_2 dB_1}{\sigma_2^2 \int B_2^2 + \delta^2\sigma_1^2 \int B_1^2 + 2\delta\sigma_1\sigma_2 \int B_1 B_2} \not\equiv MN \tag{23}$$

que sólo es MN cuando $\delta = 0$. En este caso, tal como se demostró previamente, se necesita estimador por ML a fin de obtener normalidad asintótica, aunque MCO nos provea con un estimador «superconsistente» de β. Resulta interesante hacer notar que la ausencia de normalidad asintótica en este caso proviene del hecho de que las perturbaciones de las distribuciones condicional y marginal no se encuentran incorreladas en el «largo plazo». En efecto la perturbación de (5a) en el «largo plazo» es ε_1 mientras que la de (5b''') es $\delta\varepsilon_1 + \varepsilon_2$. Por tanto, su correspondiente covarianza en el «largo plazo», entendiendo por tal la covarianza en la función de densidad espectral evaluada en la frecuencia cero, es $\delta\sigma_1^2$. Ello sugiere, de acuerdo con la discusión en (13), un procedimiento asintóticamente equivalente al ML que consiste en efectuar la regresión de $y_t - \hat{\gamma}\Delta z_{t+1}$ sobre z_t, donde $\hat{\gamma}$ es un estimador de la correlación a largo $(\delta\sigma_1^2/\delta^2\sigma_1^2 + \sigma_2^2)$ obtenido a partir de los residuos mínimo cuadráticos de ambas ecuaciones. En dicho caso

$$T(\hat{\beta} - \beta) = \frac{\Sigma z_t \varepsilon_{3t} T^2}{\Sigma z_t^2 T} - \frac{(\hat{\gamma} - \gamma)\Sigma z_t \Delta z_{t+1}}{\Sigma z_t^2} \cdot \frac{T^2}{T} \rightarrow$$

$$\rightarrow \frac{\delta\sigma_3\sigma_1 \int B_1 dB_3 + \sigma_3\sigma_2 \int B_2 dB_3}{\sigma_2^2 \int B_2^2 + \delta^2\sigma_1^2 \int B_1^2 + 2\sigma_1\sigma_2\delta \int B_1 B_2} \equiv MN \qquad (24)$$

dado que $\Sigma z_t \Delta z_{t+1}$ es $\mathrm{Op}(T)$ y $(\hat{\gamma} - \gamma)$ es $\mathrm{Op}(T^{-1/2})$.

La discusión del párrafo anterior sirve también para ilustrar un caso en que no se da exogeneidad débil y en el que, sin embargo, se alcanza normalidad asintótica. Por ejemplo, supóngase que, en (5b''), $\delta_1(L) = -\delta\beta$ y $\delta_2(L) = \delta$, entonces

$$\Delta z_t = \delta\Delta\varepsilon_{1t-1} + \varepsilon_{2t} = \delta\Delta(y_{t-1} - \beta z_{t-1}) + \varepsilon_{2t} \qquad (5b^{iv})$$

En este caso existen restricciones entre los parámetros de las distribuciones condicional y marginal, ya que el parámetro β aparece en ambas. Sin embargo, la correlación a largo plazo es cero, dado que $\rho = 0$. Resulta, por tanto, que aplicando MCO a (5a) se obtiene de nuevo

$$T(\hat{\beta} - \beta) = \frac{\sigma_1 \int B_2 dB_1}{\sigma_2 \int B_2^2} \equiv MN \qquad (25)$$

Del análisis anterior se deduce que incluso el concepto tradicional de exogeneidad débil con series estacionarias, resulta ser una característica demasiado restrictiva en la formulación de modelos condicionales con variables cointegradas. La condición suficiente mínima para dichos modelos, que incluye casos como (5b^{iv}), es la de «*exogeneidad débil en el largo plazo*», entendiendo por tal aquella situación en que las perturbaciones asociadas a la distribución condicional de y_t sobre z_t y a la distribución marginal de Δz_t, denotadas en general como u_1 y u_2, respectivamente, agrupadas en el vector $u = (u_1, u_2)$ presenta una función de densidad espectral continua $f_u(\lambda)$, con $\Omega = 2\pi f_u(0)$, de forma que $\Omega_{12} = 0$. Resulta fácil comprobar que esta restricción se verifica en todos los ejemplos en que se ha obtenido un estimador con carácter *MN*. Por ejemplo, en el caso (5b'''), dado que $u_{1t} = \varepsilon_{1t}$ y $u_{2t} = \delta\varepsilon_{1t-1} + \varepsilon_{2t}$, se tiene que $\Omega_{12} = \delta\sigma_1^2 \neq 0$, mientras que en el caso (5b^{IV}), dado que $u_{1t} = \varepsilon_{1t}$ y $u_{2t} = \delta\Delta\varepsilon_{1t-1} + \varepsilon_{2t}$, se tiene que $\Omega_{12} = 0$.

3. Inferencia en sistemas cointegrados: El modelo dinámico

En el DGP (5), la primera ecuación, que recoge el modelo condicional a estimar, viene expresada en términos estáticos, reproduciendo la relación de cointegración directamente. Como es bien sabido, este supuesto es excesivamente restringido puesto que, de acuerdo con (3), ambas ecuaciones normalmente presentan efectos dinámicos en cualquiera de sus representaciones. Engle y Granger (1987) proponen la utilización de un procedimiento bietápico de estimación que se ha popularizado en la práctica empírica. En la primera etapa, se obtiene un estimador MCO «super-consistente» del vector de cointegración en un modelo estático del tipo (5a); a continuación, en la segunda etapa, tomando dicho estimador como dado, se modeliza la dinámica a corto plazo en la segunda etapa. Dichos autores demuestran que el procedimiento bietápico proporciona estimadores de los coeficientes de la variable $I(0)$ con idéntica distribución asintótica a la de los estimadores de dichas variables en la ecuación dinámica, tomando el vector de cointegración como conocido. Un procedimiento alternativo consiste en estimar conjuntamente el vector de cointegración, conjuntamente con los coeficientes de las variables $I(0)$, representativas de la dinámica a corto plazo, a partir del modelo dinámico irrestringido en una sola etapa (véase Banerjee et al. (1986) y Sims et al. (1986)).

El siguiente ejemplo sencillo, sirve para ilustrar el funcionamiento del procedimiento «unietápico». Supongamos que el DGP de $x_t = (y_t, z)'$ viene ahora dado por el

EJEMPLO 4

$$y_t = \beta_1 z_t + \beta_2 y_{t-1} + \varepsilon_{1t} \quad , \quad |\beta_2| < 1 \tag{26a}$$

$$\Delta z_t = \varepsilon_{2t} \quad ; \quad z_0 = 0 \quad ; \quad E(\varepsilon_{1s}\varepsilon_{2t}) = 0 \tag{26b}$$

donde de acuerdo con la discusión en la sección 2 (26a) tiene la interpretación de un modelo condicional y, por simplicidad, se ha supuesto que z_t es fuertemente exógena con respecto a β_1 y β_2. Este supuesto se relajará posteriormente.

Resulta fácil comprobar que y_t y z_t son $I(1)$ y están cointegrados con un vector de cointegración $(1, -\beta)$, donde β puede estimarse como la relación a largo plazo

$$\beta = (1 - \beta_2)^{-1}\beta_1 \tag{27}$$

Siguiendo la metodología propuesta por Wickens y Breusch (1988), puede utilizarse la denominada «transformación de Bewley» con el fin de obtener un estimador directo de β a partir de (26a). En efecto, sustrayendo $\beta_2 y_t$ de ambos lados de (26a) y reparametrizando, se obtiene

$$y_t = \beta z_t - (1 - \beta_2)^{-1}\beta_2[\Delta y_t + \varepsilon_{1t}] \tag{26a'}$$

que ha de ser estimada por IV, dada la presencia de y_t en ambos lados de la ecuación. Wickens y Breusch (1988) y Galbraith et al. (1988) demuestran que el conjunto de instrumentos (z_t, y_{t-1}) proporciona un estimador de β y de su desviación típica idénticos a los que se obtendrían de la estimación directa de los $\beta_i's$ en (26a) y su sustitución en (27). Utilizando dicho conjunto de instrumentos, se obtiene

$$T(\hat{\beta}_{\mathrm{IV}} - \beta) = (1 - \beta_2)^{-1}\frac{(\Sigma y_{t-1}\Delta y_t)(\Sigma z_t\varepsilon_{1t}) - (\Sigma z_t\Delta y_t)(\Sigma y_{t-1}\varepsilon_{1t})}{(\Sigma z_t^2)(\Sigma y_{t-1}^2\Delta y_t) - (\Sigma z_t\Delta y_t)(\Sigma z_t y_{t-1})}\cdot\frac{T^3}{T^2} \tag{28}$$

Utilizando las distribuciones límites presentadas en (8), se obtiene

$$\begin{aligned}
&T^{-2}\Sigma z_t^2 \to \sigma_2^2 \textstyle\int B_2^2\\
&T^{-1}\Sigma z_t\varepsilon_{1t} \to \sigma_1\sigma_2 \textstyle\int B_2 dB_1\\
&T^{-1}\Sigma y_{t-1}\varepsilon_{1t} \to (1 - \beta_2)^{-1}\beta_1\sigma_1\sigma_2 \textstyle\int B_2 dB_1\\
&T^{-1}\Sigma z_t\Delta y_t \to (1 - \beta_2)^{-1}\beta_1\sigma_2^2[\textstyle\int B_2 dB_2 + 1]\\
&T^{-1}\Sigma y_{t-1}\Delta y_t = [(1 - \beta_2)^{-1}\beta_1]^2\sigma_2^2 \textstyle\int B_2 dB_2 - (1 + \beta_2)^{-1}\sigma_1^2\\
&T^{-1}\Sigma z_y y_{t-1} \to (1 - \beta_2)^{-1}\beta_1\sigma_2^2 \textstyle\int B_2^2
\end{aligned} \tag{29}$$

Sustituyendo las expresiones (29) en (28), se obtiene a su vez

$$T(\hat{\beta}_{\mathrm{IV}} - \beta) \to (1 - \beta_2)^{-1}\frac{\sigma^1 \int B_2 dB_1}{\sigma_2 \int B_2^2} \equiv MN \tag{30}$$

Por tanto, el t-ratio de $\hat{\beta}_{\mathrm{IV}}$ se comporta asintóticamente como la normal estandarizada. Puesto que Wickens y Breusch (1988) demuestran que $\hat{\beta}_{\mathrm{IV}}$ es idéntico al estimar MCO de β, $\hat{\beta}$, obtenido al sustituir en (27) los estimadores mínimo cuadráticos de β, y β_2, se sigue que $\hat{\beta}$ se distribuye asintóticamente de forma MN. Resulta importante destacar que si en lugar de utilizar el procedimiento anterior, se utilizase el procedimiento bietápico de Engle y Granger, la primera etapa correspondería a la estimación por MCO de

$$y_t = \beta z_t + u_t \quad ; \quad u_t = [(1 - \beta_2 L)(1 - \beta_2)]^{-1}[\varepsilon_{1t} - \beta_1\beta_2\varepsilon_{2t}] \tag{31}$$

Por tanto, z_t no es débilmente exógena con respecto a β en (31), puesto que $E(\Delta z_t u_t) = -(1 - \beta_2)^{-1}\beta_1\beta_2\sigma_2^2$, y, consecuentemente con el análisis del DGP (5) en el caso de que $\rho \neq 0$, la distribución de $\hat{\beta}$ ya no será MN, impidiendo el uso de técnicas estándar de inferencia.

Tal como se discutió en la sección 2, resulta posible relajar el supuesto de exogeneidad fuerte de z_t, sustituyendo (26b) por (5b'') donde solamente se supone exogeneidad débil en el «largo plazo». El siguiente DGP generaliza también la dinámica del modelo condicional,

$$y_t = \alpha(L)z_t + \beta(L)y_{t-1} + \varepsilon_{1t} \quad ; \quad |\beta(1)| < 1 \tag{32a}$$

$$\Delta z_t = \delta_1(L)\Delta z_{t-1} + \delta_2(L)\Delta y_{t-1} + \varepsilon_{2t} \quad ; \quad |\delta_1(1)| < \infty\,, \quad |\delta_2(1)| < \infty \tag{32b}$$

$$E(\varepsilon_{1t}\varepsilon_{2s}) = 0$$

dando lugar a la siguiente proposición:

PROPOSICIÓN 2: Si $x_t = (y_t, z_t)'$ está generada por el DGP (32) y β se define como $\beta = [1 - \beta(1)]^{-1}\,\alpha(1)$, entonces el estimador MCO de β en (32a), definido previamente, presenta la siguiente distribución asintótica

$$T(\hat{\beta} - \beta)^2 \rightarrow (1 - \beta(1))^{-2}[1 - (\beta(1) + \delta_2(1)\alpha(1)) -$$
$$- \delta_1(1)(1 - \beta(1))] \frac{\sigma_1 \int B_2 dB_1}{\sigma_2 \int B_2^2} \equiv MN \tag{33}$$

si el término entre corchetes es distinto de cero.

DEMOSTRACIÓN: (Véase apéndice.)

Tanto en DGP (26) como en el caso más general, DGP (32), la presencia de desfases de z e y en el modelo condicional está reflejando el proceso de reducción de la función de densidad conjunta de la variable $x_t = (y_t, z_t)'$, de forma que se obtenga una forma paramétrica de la función condicional en que el término de error sea una «innovación». Un ejemplo bastará para clarificar este concepto. Supongamos que el DGP es el siguiente

EJEMPLO 5

$$y_t = \beta z_t + u_{1t} \quad ; \quad u_{1t} = \varepsilon_{1t} \tag{34a}$$

$$\Delta z_t = u_{2t} \quad ; \quad u_{2t} = (1 - bL)^{-1}\varepsilon_{2t} \quad ; \quad |b| < 1 \tag{34b}$$

$$E(\varepsilon_{1t}\varepsilon_{2t}) = \rho\sigma_1\sigma_2 \quad ; \quad \varepsilon_{it} \sim \text{nid}\,(0, \sigma_i^2)$$

Condicionando ε_{1t} sobre ε_{2t} se obtiene

$$\varepsilon_{1t} = \gamma(1 - bL)\Delta z_t + \varepsilon_{3t} \quad ; \quad E(\Delta z_s \varepsilon_{3t}) = 0$$

por lo que el modelo se puede escribir en forma paramétrica como

$$y_t = \beta z_t + \gamma(1 - bL)\Delta z_t + \varepsilon_{3t} \quad ; \quad E(\Delta z_t \varepsilon_{3s}) = 0$$

Al ser z_t al menos débilmente exógena con respecto a β en este modelo (no lo es en (34a)), los resultados de la proposición 2 se pueden aplicar, obteniendo normalidad asintótica para el estimador $\hat{\beta}$.

Existen, sin embargo, DGP's para los que dicha transformación paramétrica no es posible. En efecto, supongamos ahora que (34b) tuviese una forma de media móvil en vez de autorregresiva, esto es

$$\Delta z_t = (1 - bL)\varepsilon_{2t} \quad ; \quad |b| < 1 \tag{34b'}$$

En este caso una transformación paramétrica del tipo (35) no resulta adecuada, ya que el término de error sería MA(1) en vez de una innovación. Por ello resulta de gran interés examinar la posibilidad de obtener un estimador de β con características no parámetricas, en el sentido descrito previamente, con la propiedad MN. La distribución asintótica de $\hat{\beta}$ en (34a) nos ayuda en la construcción de dicho estimador ya que

$$T(\hat{\beta} - \beta) = \frac{\Sigma z_t \varepsilon_{1t}}{\Sigma z_t^2} \cdot \frac{T^2}{T} = \frac{\Sigma z_t[\gamma(1 - bL)^{-1}\Delta z_t + \varepsilon_{3t}]}{\Sigma z_t^2} \cdot \frac{T^2}{T} \to$$

$$\to (1 - b)^{-1}\left[\frac{\gamma\sigma_2(\int B_2 dB_2 + 1) + \sigma_3 \int B_2 dB_3}{\sigma_2 \int B_2^2}\right] \tag{35}$$

Como el segundo término en (35) es MN, todo lo que se necesita es un estimador no paramétrico del primer término. La matriz de covarianzas a «largo plazo» de u_{1t} y u_{2t} proporciona dicho estimador ya que, denotando sus elementos como $\{w_{ij}\}$ sabemos que $(1 - b)^{-1}\gamma = \omega_{12}\omega_{22}^{-1}$. Utilizando el estimador muestral de $\hat{\gamma}$ a partir de Δz_t y $\hat{u}_{1t}$, puede construirse un nuevo estimador $\bar{\beta}$ de la forma siguiente

$$\bar{\beta} = \frac{\Sigma z_t(y_t - \hat{\gamma}\Delta z_t)}{\Sigma z_t^2} \tag{36}$$

con lo que desaparece el primer término en (35), de manera que

$$T(\bar{\beta} - \beta) \equiv MN$$

Resulta importante destacar que, de acuerdo con la discusión acerca del estimador ML en la sección 2, el estimador $\bar{\beta}$ es asintóticamente equivalente al estimador ML y, por tanto, comparte sus propiedades de optimalidad asintótica.

En el caso más general en que u_1 y u_2 en (34) presenten correlación serial y z_t no sea débilmente exógena, Phillips y Hansen (1988) han propuesto un estimador similar a (35) al que denominan estimador FM («fully modified»). Dicho estimador se construye de la forma siguiente

$$\bar{\beta} = \frac{\Sigma z_t y_t^+ - \hat{\delta}^+ T}{\Sigma z_t^2} \tag{37}$$

donde

$$y_t^+ = y_t - \hat{\gamma}\Delta z_t \quad ; \quad \hat{\gamma} = \hat{\omega}_{12}/\hat{\omega}_{22}$$

y $\hat{\delta}^+$ es un estimador consistente de δ, definido por

$$\delta = \sum_{h=0}^{\infty} [E(\Delta z_t u_{1t+k}) - \gamma E(\Delta z_t \Delta z_{t+k})]$$

donde δ representa el posible sesgo a «largo plazo» existente entre regresores y la perturbación[3].

Puede demostrarse de nuevo que

$$T(\bar{\beta} - \beta) \equiv MN$$

si bien en este caso la varianza de u_1 en el cómputo del t-ratio de $\bar{\beta}$, habrá de ser sustituida por un estimador consistente de σ_3^2 dado por $(\hat{\omega}_{11} - \hat{\gamma}^2\hat{\omega}_{22})$. Nótese que en el caso en que $\delta = 0$, el estimador corresponde a (35). Resumiendo, debería quedar claro de los ejemplos anteriores que el procedimiento unietápico, correspondiente a la metodología de Hendry y Richard en la mayoría de los casos presenta las mismas propiedades que en la estimación FM y, por tanto, es óptima. Sólo cuando existe evidencia de medias móviles en la perturbación (e.g. Hall y Pagan (1981)), después de haber introducido desfases de la variable exógena y endógena, las correcciones del método FM se hacen necesarias.

Una limitación del marco bivariante adoptado hasta ahora en los diferentes ejemplos es que no permite la existencia de más de un vector de cointegración, ya que de no ser único las variables y_t y z_t serían $I(0)$ en vez de $I(1)$. Esta limitación desaparece en un marco n-variante (x_t ahora es un vector $(n \times 1)$) donde se permita la existencia de $r(r \leqslant n - 1)$ relaciones de cointegración. Con el fin de analizar las condiciones bajo las que los procedimientos uniecuacionales descritos previamente conducen, tras la normalización apropiada, a estimadores de vectores de cointegración que son MN, resulta conveniente adoptar la representación ECM del sistema (véase Engle y Granger (1987)) dada por

$$A^*(L)\Delta x_t = - B\Gamma' x_{t-1} + \varepsilon_t \quad ; \quad E(\varepsilon_t \varepsilon_t') = \sigma_\varepsilon^2 \Sigma \tag{38}$$

donde se supone que $A^*(L)$ es una matriz de polinomios en desfases con orden finito, tal que $A^*(\cdot)$ tiene todas sus raíces fuera del círculo unitario; Γ es una matriz $(n \times r)$ cuyas columnas contienen los r diferentes factores de cointegración; y B es una matriz $(n \times r)$ cuyas filas representan los coeficientes de los términos de corrección del error en cada ecuación. Finalmente, se supone que la matriz de var-covarianzas de ε_t es diagonal, tras un proceso de condicionamiento progresivo. Condiciones adicionales para que se verifique la exogeneidad débil requieren que $A^*(0)$ sea una matriz triangular superior, propiedad que junto con la diagonalidad de la matriz de var-covarianzas dota al sistema de estructura recursiva. Supongamos, sin pérdida de generalidad, que la i-sima

[3] El parámetro δ recoge el denominado «sesgo de segundo orden» que, por ejemplo, genera los sesgos en muestras finitas encontrados por Banerjee et al. (1986). Dicho sesgo es independiente del sesgo de endogeneidad debido a la correlación entre ε_{1t} y ε_{2t}, que conlleva la aparición de la distribución del estimador de la raíz unitaria en la distribución límite del estimador del vector de cointegración.

ecuación del sistema representa el modelo condicional de interés. De acuerdo con la discusión referente al ejemplo 3, ha de excluirse la presencia de términos de corrección del error idénticos en más de una ecuación cuando las variables con coeficientes normalizados a la unidad en cada una de dichas ecuaciones entre como variable explicativa en la i-sima ecuación. Ello implica que, con $B = \{b_{ij}\}$, es necesario que si $b_{ij} \neq 0$, se verifique $b_{kj} = 0$ (para todo $k \neq i$). Adicionalmente han de eliminarse los casos en que existe perfecta colinealidad asintótica, producida por la existencia de diferentes vectores de cointegración que comprendan todos o un subconjunto de las variables que aparezcan en la i-sima ecuación. El siguiente DGP sirve de ilustración al problema:

EJEMPLO 6

$$y_t = \beta_1 z_{1t} + \beta_2 z_{2t} + \varepsilon_{1t} \tag{39a}$$

$$z_{1t} = \alpha z_{2t-1} + \varepsilon_{2t} \tag{39b}$$

$$\Delta z_{2t} = \varepsilon_{3t} \tag{39c}$$

$$E(\varepsilon_{is}\varepsilon_{jt}) = 0(i, j = 1, 2, 3)$$

Resulta sencillo comprobar que y_t y las z_{it}'s son variables $I(1)$ y que $(1, -\beta_1, -\beta_2)$ y $(0, 1, -\alpha)$ son los vectores de cointegración existentes. Bajo el supuesto de independencia de las perturbaciones, puede aplicarse MCO al modelo condicional (39a) pero la matriz $(Z'Z)(z = (z_1, z_2)')$ necesaria para evaluar $(\hat{\beta}_1, \hat{\beta}_2)$ es tal que

$$T^{-2} \Sigma z_{2t}^2 \to \sigma_3^2 \int B_3^2 \quad ; \quad T^{-2} \Sigma z_{1t} z_{2t} \to \alpha\sigma_3^2 \int B_3^2 \quad ; \quad T^{-2} \Sigma z_{1t}^2 \to \alpha^2\sigma_3^2 \int B_3^2$$

Por tanto el rango de $(Z'Z) \to 1$ y diremos que z_{1t} y z_{2t} son colineales asintóticamente. Alternativamente podemos afirmar que $T(\hat{\beta}_1 - \beta_1)$ y $T(\hat{\beta}_2 - \beta_2)$ tienen una distribución conjunta degenerada. Nótese que, por ejemplo, las distribuciones anteriores serán MN si z_{2t-1} se sustituye por y_{t-1} en (39b) dado que el segundo vector de cointegración no incluye un subconjunto de las variables explicativas del modelo condicional[4]. Con el fin de excluir este problema, con $\Gamma = \{\gamma_{ij}\}$ y normalizando $\gamma_{ii} = 1(i = 1, ..., r)$, es necesario que si $\gamma_{ij} \neq 0$ $(j = 1, ..., k < n)$ se verifique que $\gamma_{ij} = 0$ $(i = 2, ..., r, j = 1, ..., k)$.

Con el fin de contrastar todas las condiciones anteriores, se puede aplicar inferencia ordinaria a los elementos de $A^*(L)$, de acuerdo con los resultados de Sims et al. (1986), y también a los elementos de B y Γ, construyendo contrastes

[4] Alternativamente (39a) puede reparametrizarse como

$$y_t = \beta_1(z_{1t} - \alpha z_{2t}) + (\beta_2 + \alpha\beta_1)z_{2t} + \varepsilon_{1t}$$

donde $T^{1/2}(\hat{\beta}_1 - \beta_1)$ y $T(\hat{\gamma} - \gamma)$ (con $\gamma = (\beta_2 + \alpha\beta_1))$ tiene distribuciones límites no degeneradas. Sin embargo, para identificar los β_i's se necesita un estimador de α obtenido de (39b). Por tanto z_t no es débilmente exógena a no ser que α sea conocido.

de Wald como los propuestos por Johansen y Juselius (1988). Los contrastes sobre la diagonalidad de la matriz de var-covarianzas de las perturbaciones corresponden a la clase de contrastes de Hausman.

4. Inferencia en sistemas posiblemente no cointegrados

A diferencia de las secciones anteriores donde se ha tomado como un supuesto de partida la existencia de cointegración, en esta sección se examina el comportamiento relativo de contrastes de cointegración, en lo que respecta a su tamaño y potencia bajo las hipótesis nula y alternativa, respectivamente. El análisis se concentra al igual que en los casos anteriores en un marco uniecuacional donde las condiciones relevantes de cumplimiento del requisito de exogeneidad débil por parte de las variables explicativas se ha verificado previamente.

Con el fin de simplificar la discusión, se analiza un ejemplo muy sencillo de carácter bivariante en el que el posible vector de cointegración se supone conocido e igual a $(1, -\lambda)$ tal como ocurre en numerosas aplicaciones empíricas (e.g., $\lambda = 1$ corresponde a logs. de consumo y renta disponible, logs. de salario real y productividad, logs. de cantidad de dinero y renta nominal, logs. de tipo de cambio nominal y precios relativos, etc.). El DGP correspondiente viene definido por

EJEMPLO 7

$$\Delta y_t = \alpha_1 \Delta z_t + \alpha_2 (y - \lambda z)_{t-1} + \varepsilon_{1t} \quad (40a)$$

$$\Delta z_t = \varepsilon_{2t} \quad (40b)$$

$$E(\varepsilon_{1t}\varepsilon_{2s}) = 0 \quad ; \quad s = \sigma_2/\sigma_1$$

donde y_t y z_t son $I(1)$ y s representa la razón «señal-ruido». Bajo la hipótesis nula de no cointegración se verifica que $\alpha_2 = 0$. Generalmente se han supuesto dos contrastes de dicha hipótesis. El primer contraste (ECM) está basado en el t-ratio de $\hat{\gamma}_2$ en el modelo

$$\Delta y_t = \gamma_1 \Delta z_t + \gamma_2 (y - \lambda z)_{t-1} + \eta_t \quad (41)$$

basado en la idea de que si y_t y z_t no están cointegradas, el término en niveles debe desaparecer de la ecuación. El segundo contraste es el de Dickey y Fuller (DF) (véase Dickey y Fuller (1979)), el cual, bajo el supuesto del conocimiento del vector de cointegración, se basa en el t-ratio de $\hat{\delta}$ en el modelo

$$\Delta(y_t - \lambda z_t) = \delta (y - \lambda z)_{t-1} + v_t \quad (42)$$

Por tanto, el contraste ECM, en notación matricial, viene dado por

$$t_{\gamma_2=0}\frac{\Delta y' M_{\Delta z} w_{-1}}{[w'_{-1} M_{\Delta z} w_{-1}]^{1/2} \hat{s}_\eta} \tag{43}$$

donde $M_{\Delta z} = 1 - \Delta z(\Delta z'\Delta z)^{-1}\Delta z$; $w = (y - \lambda z)$ y $\hat{s}_\eta^2 = \Sigma\, \hat{\eta}_t^2/T - 2$.

El contraste DF a su vez, viene dado por

$$t_{\delta=0} = \frac{\Delta w' w_{-1}}{[w'_{-1} w_{-1}]^{1/2} \hat{s}_v} \tag{44}$$

con $\hat{s}_v^2 = \Sigma\, \hat{v}'\hat{v}/T - 1$.

Utilizando las distribuciones límites presentadas en (8), puede demostrarse la siguiente proposición concerniente al comportamiento relativo de ambos contrastes:

PROPOSICIÓN 3: Si $z_t = (y_t, z_t)'$ está generado por el DGP (40) con $\alpha_2 = 0$, las distribuciones asintóticas de los contrastes ECM y DF vienen dados por

$$t_{\gamma_2} = 0 \to \frac{(\alpha_1 - \lambda)\int B_2 dB_1 + s^{-1}\int B_1 dB_1}{[(\alpha_1 - \lambda)^2 \int B_2^2 + 2s^{-1}(\alpha_1 - \lambda)\int B_1 B_2 + s^{-2}\int B_1^2]^{1/2}} \tag{45}$$

$$t_{\delta=0} \to \frac{(\alpha_1 - \lambda)^2 \int B_2 dB_2 + 2s^{-1}(\alpha_1 - \lambda)\int B_2 dB_1 + s^{-2}\int B_1 dB_1}{[s^{-2} + (\alpha_1 - \lambda)^2]^{1/2}[(\alpha_1 - \lambda)^2 \int B_2^2 + 2s^{-1}(\alpha_1 - \lambda)\int B_1 B_2 + s^{-2}\int B_1^2]^{1/2}} \tag{46}$$

DEMOSTRACIÓN: (Véase apéndice.)

Tomando una aproximación de «pequeño sigma», i.e., $s \uparrow \infty$, conjuntamente con $T \uparrow \infty$, se obtiene una doble aproximación asintótica tal que

$$t_{\gamma_2=0} \to N(0, 1) + \mathrm{Op}(s^{-1}) \tag{47}$$

donde aquellos términos afectados por $(1/s^2)$ desaparecen a mayor velocidad que los restantes y, por tanto, sólo los términos que definen una MN permanecen en la aproximación. Nótese que $\alpha_1 \neq \lambda$ es condición necesaria para la validez de la aproximación.

Una doble aproximación asintótica similar a la anterior, produce en el caso del test DF

$$t_{\delta=0} \to [\int B_2^2]^{-1/2} \int B_2 d\, B_2 + \mathrm{Op}(s^{-1}) \tag{48}$$

que resulta ser la conocida distribución límite del t-ratio de $\hat{\rho}$ en el proceso AR (1)

$$\Delta y_t = -\rho y_{t-1} + \varepsilon_t \quad ; \quad \rho = 0 \tag{49}$$

tal como ha sido demostrado por Phillips (1987).

Por tanto, de acuerdo con la discusión anterior, existen circunstancias en que el test ECM es asintóticamente normal, mientras que el test DF nunca lo será. Nótese que dicha característica depende del tamaño del ratio «señal-ruido», s, a cuyo análisis en muestras finitas se dedica la sección siguiente.

La intuición subyacente a los resultados anteriores es la siguiente. Tomando la ecuación (40a) con $\alpha_2 = 0$ y sustrayendo $\lambda\Delta z_t$ de ambos lados, se obtiene

$$\Delta(y_t - \lambda z_t) = (\alpha_1 - \lambda)\Delta z_t + \varepsilon_{1t} \tag{50}$$

lo que implica que las varianzas de las perturbaciones en (41) y (42), son

$$\operatorname{var}(\eta_t) = \sigma_1^2 \tag{51}$$

$$\operatorname{var}(v_t) = \sigma_1^2 + (\alpha_1 - \lambda)^2\sigma_2^2 \tag{52}$$

Cuando $s \uparrow \infty$, y_t y z_t parecen cointegradas, puesto que al integrar (40a), con $\alpha_2 = 0$, la varianza de z_t dominará a la varianza de $(1 - L)^{-1}\varepsilon_{1t}$. Esto no ocurre con el test DF, puesto que el término de error, de acuerdo con (50), es $(1 - L)^{-1}[\alpha_1 - \lambda)\varepsilon_{2t} + \varepsilon_{1t}]$, cuya varianza, al contener ε_{2t}, ya no es asintóticamente despreciable. Esta intuición, sin embargo, sólo es válida cuando $\alpha_1 \neq \lambda$. Cuando $\alpha_1 = \lambda$ y s sea finito se comprueba a partir de (45) y (46) que ambos contrastes coinciden.

Otra manera alternativa de interpretar los resultados anteriores consiste en observar que el test DF *impone una restricción de «raíz común»* en el modelo (42) (véase Sargan (1980)), i.e., $\alpha_1 = \lambda$, restricción que no tiene por qué verificarse en el DGP. Generalizaciones del test DF, tales como el test aumentado de Dickey y Fuller (ADF), que incluiría desfases de $\Delta(y_t - \lambda z_t)$ en (42), tampoco solucionan este problema ya que continúan imponiendo una restricción potencialmente inválida, i.e., la igualdad de las elasticidades a corto y largo plazo.

La teoría asintótica desarrollada hasta el momento puede utilizarse también para derivar las funciones de potencia asintótica de los contrastes ECM y DF. Con el fin de ilustrar su desarrollo en el ejemplo 7, utilizaremos el marco estadístico propuesto recientemente por Phillips (1988b) para el análisis de series temporales «casi integradas», esto es, series estacionarias pero con raíces en un entorno muy cercano a la unidad. En nuestro caso, la hipótesis alternativa local relevante será $\alpha_2 = T^{-1}\pi$ con π fijo, de forma que cuando $\pi = 0$ obtendremos los resultados de la Proposición 3, mientras que cuando $T \uparrow \infty$ y $\pi \neq 0$, obtendremos los resultados relevantes a la potencial local de los contrastes. Nótese que la velocidad de convergencia del parámetro α_2 bajo la hipótesis alternativa es $\mathrm{Op}(T)$ en vez de $\mathrm{Op}(T^{1/2})$ que es la velocidad de convergencia habitual en la función de potencia local de contrastes basados en series estacionarias. Ello es debido a que, de acuerdo con el análisis de series «cuasi integradas» los segundos momentos muestrales convergen a velocidad $\mathrm{Op}(T^2)$. Por ejemplo, si en el proceso AR (1) recogido en (49), $\rho = T^{-1}\pi$, Phillips (1988b) demuestra que

$$T\hat{\rho} \rightarrow \pi + [\int K^2]^{-1}[\int K dB] \tag{53}$$

donde B es el proceso de Wiener asociado a ε y K es un proceso de difusión tal que

$$K(r) = B(r) + \pi \int_0^r e^{(r-s)} dB(s) \quad , \quad r \in [0, 1)$$

de forma que cuando $\pi = 0$, $K(r)$ es igual a $B(r)$.

Utilizando los resultados anteriores, Kremers et al. (1989) y Dolado y Ericsson (1989) demuestran que los factores de no centralidad (μ) de los tests ECM y DF vienen dados, respectivamente, por

$$\mu_{\mathrm{ECM}} = \pi^2[1 + (\alpha_1 - \lambda)s]^2 \int K^2 \tag{54}$$

$$\mu_{\mathrm{DF}} = \pi^2 \int K^2 \tag{55}$$

donde K es el proceso de difusión asociado a una perturbación con varianza unitaria. Por tanto, cuando el término entre corchetes en (54) sea, en valor absoluto, superior a la unidad, la potencia del test ECM será superior a la del test DF. Cuando $s \uparrow 0$, o cuando $\alpha_1 = \lambda$ ambos contrastes tienen la misma potencia.

Finalmente, es importante hacer notar que los resultados relativos a la normalidad asintótica del test ECM, cuando $s \uparrow \infty$, no se verifican si el vector de cointegración se estima como en el proceso «bietápico» de Engle y Granger, tal como se demuestra en Dolado y Ericsson (1989). El resultado es similar a lo que ocurre con la no imposición de las raíces unitarias en el ejemplo 2, de ahí que la teoría económica juegue un papel fundamental en la imposición del potencial vector de cointegración.

5. Resultados de un estudio de Monte Carlo

En esta sección se analiza el comportamiento en muestras finitas de las aproximaciones asintóticas de tamaño y potencia sugeridas en la sección anterior. Con el fin de analizar el tamaño, se simula un DGP como el expresado en (40), mediante un pequeño estudio de Monte Carlo con un número de simulaciones $N = 1000$ y un tamaño muestral $T = 100$ en el espacio paramétrico definido por: $s = (1, 5, 20, 200, 1/5, 1/20, 1/200)$, $\alpha_1 = (0{,}1, 0{,}5, 0{,}9)$, $\lambda = 1$. El rango del ratio «señal-ruido» es bastante extenso, incluyendo tres casos favorables a la normalidad asintótica del test ECM, tres casos desfavorables, cuando s es pequeño y el caso en que s es la unidad. De forma similar, el rango del parámetro a corto plazo α_1 trata de capturar un caso favorable, cuando α_1 está muy por debajo de la unidad, valor adoptado por λ, un caso desfavorable, cuando α_1 está cerca de la unidad, y un caso intermedio, cuando α_1 es 0,5

Los dos bloques en el cuadro 1 presentan la media, t-ratio de la media y un contraste de normalidad basado en los coeficientes de kurtosis y asimetría de la distribución simulada (véase Jarque y Bera (1980)) en el caso de los dos contrastes propuestos, con el fin de examinar la semejanza existente entre las distribuciones empíricas de los contrastes y la distribución $N(0, 1)$. Los resultados se resumen a continuación. En la mayoría de los casos relativos al test

CUADRO 1

CARACTERÍSTICAS DE LA DISTRIBUCIÓN BAJO H_0

Modelo (41): Test ECM

	s	5	20	200	1	1/5	1/20	1/200
α_1	0,1	−0,18	−0,08	−0,07	−0,30	−0,53	−0,54	−0,52
		(1,68)**	(0,75)	(0,66)	(3,02)*	(5,32)*	(5,47)*	(5,22)*
		0,67	0,18	0,14	1,49	5,23**	1,58	2,74
	0,5	−0,16	−0,04	0,00	−0,37	−0,42	−0,43	−0,43
		(1,59)	(0,43)	(0,03)	(2,18)*	(4,11)*	(4,17)*	(4,17)*
		0,19	0,15	0,12	0,36	0,79	1,00	1,12
	0,9	−0,48	−0,19	−0,10	−0,43	−0,40	−0,49	−0,52
		(4,88)*	(1,92)**	(1,04)	(4,11)*	(4,03)*	(4,93)*	(5,22)*
		2,30	2,30	0,42	0,73	1,44	1,79	1,86

Modelo (42): Test DF

	s	5	20	200	1	1/5	1/20	1/200
α_1	0,1	−0,54	−0,54	−0,52	−0,51	−0,48	−0,52	−0,50
		(5,43)*	(5,37)*	(5,24)*	(5,15)*	(4,82)*	(5,16)*	(5,08)*
		1,47	2,34	2,15	1,92	1,64	3,90	2,16
	0,5	−0,36	−0,39	−0,40	−0,40	−0,42	−0,43	−0,43
		(3,60)*	(3,90)*	(3,80)*	(4,05)*	(4,12)*	(4,23)*	(4,22)*
		0,59	0,38	0,43	0,88	0,60	0,82	0,96
	0,9	−0,50	−0,53	−0,50	−0,50	−0,50	−0,48	−0,49
		(5,15)*	(5,46)*	(5,06)*	(4,97)*	(5,01)*	(4,89)*	(4,86)*
		2,57	4,14	1,77	1,53	1,48	2,00	1,43

Nota: El DGP viene dado por las ecuaciones (40) con $\lambda = 1$. Las cifras superiores representan las medias de las distribuciones del contraste correspondiente; las cifras intermedias (entre paréntesis) representan los t-ratios de las medias anteriores; las cifras inferiores representan los valores del contraste de normalidad descrito en el texto, cuya distribución asintótica es una chi-cuadrado con 2 grados de libertad; (*) y (**) representan valores significativos a niveles del 5 y 10 por 100, respectivamente.

ECM, el contraste de normalidad no captura desviaciones significativas. Lo mismo ocurre cuando se trata del contraste DF. Sin embargo, cuando s es pequeño, con independencia del valor tomado por α_1, la media de la distribución del test ECM aparece sesgada a la baja, lo que implica la existencia de valores críticos distintos de los ordinariamente utilizados. Un sesgo similar ocurre en la distribución del test DF, en este caso, como es esperable, con independencia de ciertas combinaciones de los parámetros s y α_1. Así pues, la no centralidad de las distribuciones en el caso no estándar aparece como un rasgo más importante que el exceso de kurtosis y/o asimetría, una característica que parece concordar con los resultados de Banerjee y Dolado (1987, 1988). La ausencia de desviaciones significativas con respecto a la normal estandarizada en el caso en que s sea grande ($s > 5$) y α_1 pequeño ($\alpha_1 < 0,5$) apoya en muestras finitas los resultados asintóticos obtenidos en la sección anterior, apoyando el uso de dicha distribución cuando exista evidencia de que s es suficientemente grande.

CUADRO 2

POTENCIA

Modelo (41): Test ECM

	s	5	20	1/5	1/20
$\alpha_2 = -0{,}05, \alpha_1$	0,1	0,998	1,000	0,340	0,340
	0,5	0,995	1,000	0,342	0,341
	0,9	0,429	0,911	0,340	0,343
$\alpha_2 = -0{,}10, \alpha_1$	0,1	1,000	1,000	0,790	0,787
	0,5	1,000	1,000	0,782	0,780
	0,9	0,856	0,998	0,761	0,765

Modelo (42): Test DF

	s	5	20	1/5	1/20
$\alpha_2 = -0{,}05, \alpha_1$	0,1	0,321	0,327	0,314	0,334
	0,5	0,314	0,326	0,332	0,333
	0,9	0,306	0,310	0,334	0,335
$\alpha_2 = -0{,}10, \alpha_1$	0,1	0,752	0,761	0,787	0,790
	0,5	0,763	0,772	0,781	0,782
	0,9	0,776	0,751	0,776	0,775

Nota: El DGP viene dado por las ecuaciones (40. con $\lambda = 1$. Los valores críticos al nivel del 5 por 100 son $-1{,}95$ (tomado de Fuller (1976), tabla 8.5.2) y $-1{,}96$ (tomado de la normal estandarizada).

Con el fin de cuantificar la potencia de los contrastes ECM y DF, se ha simulado de nuevo un DGP como el anterior, pero ahora con $\alpha_2 \neq 0$. De acuerdo con el análisis de potencia local llevado a cabo en la sección anterior, el espacio paramétrico se ha reducido a $s = (5, 20, 1/5, 1/20)$, $\alpha_1 = (0{,}1, 0{,}5, 0{,}9)$ y $\alpha_2 = (-0{,}05, -0{,}10)$, ya que para valores superiores de α_2 la potencia de ambos contrastes alcanza la unidad. El cuadro 2 presenta los resultados relativos a la frecuencia de rechazo de la hipótesis nula $\alpha_2 = 0$, cuando se utilizan valores críticos correspondientes al 5 por 100 de nivel de significación. Cuando s es bajo, la potencia de ambos contrastes es similar, mientras que cuando s es alto la potencia del test ECM tiende a la unidad, lo que no ocurre con el test DF cuya potencia, de acuerdo con (55), es independiente de s. Nótese de nuevo que, este último contraste al imponer restricciones de factor común inválidas, i.e. $\alpha_1 = \lambda$, conlleva una pérdida de potencia relativa apreciable con respecto al test ECM.

6. Conclusiones

En este trabajo se han discutido condiciones bajo las que puede llevarse a cabo inferencia ordinaria en modelos condicionales con variables $I(1)$, estimados por MCO, tanto cuando existe cointegración como cuando no existe. Cuando existe cointegración, la condición de «*exogeneidad débil en el largo plazo*», definida con respecto a la matriz de densidad espectral evaluada en la frecuencia cero, es suficiente para lograr normalidad asintótica en los estimadores de todos los coeficientes que aparecen en la regresión, habiendo estimado el vector de cointegración como la solución a «largo plazo» del modelo condicional dinámico. Se demuestra, como corolario del resultado anterior, que la utilización del popular procedimiento bietápico de Engle y Granger, conlleva necesariamente la ausencia de normalidad asintótica en los coeficientes estimadores del vector de cointegración. Ello implica que la especificación y estimación del modelo dinámico, en una única etapa, resulta ser un procedimiento preferible desde ésta perspectiva. Además, cuando no se verifica la condición suficiente de exogeneidad y, consecuentemente, ha de aplicarse un procedimiento de estimación máximo-verosímil, la imposición de las raíces unitarias existentes en el DGP resulta necesaria para lograr la normalidad asintótica. En este sentido, la validez del enfoque VAR irrestringido propuesto por Sims et al., se ve limitada.

En el caso de que no exista cointegración, se han encontrado casos en que el contraste ECM, basado en el t-ratio del estimador del coeficiente del mecanismo de corrección del error, se distribuye asintóticamente como una normal estandarizada. Estos casos vienen representados por un alto ratio «señal-ruido». Esta propiedad no se verifica en el caso del contraste DF, basado en la aplicación del test de Dickey-Fuller a los residuos de la regresión estática en niveles, cuya distribución no depende del ratio «señal-ruido». Un pequeño estudio de Monte Carlo pone de manifiesto que la no centralidad es la característica dominante de la distribución no estándar de estos contrastes. La imposición de restricciones, en muchos casos inválidas, por parte del contraste DF y sus generalizaciones supone una pérdida apreciable de potencia local con respecto al contraste ECM, lo que aconseja el uso de este último en la práctica econométrica.

Apéndice

Las demostraciones de las Proposiciones (1)-(3) están basadas en el siguiente Lema, cuya demostración puede encontrarse en: e.g., Stock (1987).

LEMA: Sea z_t una serie temporal $I(1)$ con una representación de Wold dada por $\Delta z_t = c(L)\varepsilon_{2t}$ siendo $c(L)$ un polinomio en desfases invertible y $\varepsilon_{2t} \sim \text{nid}(0, \sigma_2^2)$. Entonces $c(L) = c(1) + c^*(L)(1 - L)$ con $c^*(\cdot)$ invertible. Si $\varepsilon_{1t} \sim \text{nid}(0, \sigma_1^2)$ está incorrelacionado con ε_{2t} en todos los períodos, se verifica

$$T^{-2}\,\Sigma\, z_t^2 \to c(1)^2\sigma^2{}_2 \int B_2^2 \qquad \text{(A.1)}$$

$$T^{-1}\,\Sigma\, z_t\varepsilon_{2t} \to \sigma_2^2[c(1)\int B_2 dB_2 + 1] \qquad \text{(A.2)}$$

$$T^{-1}\Sigma z_t\varepsilon_{1t} \to c(1)\int B_2 dB_1 \to c(1)N\left(0, \frac{\sigma_1^2\sigma_2^2}{2}\right) \quad (A.3)$$

donde B_1 y B_2 son procesos de Wiener con varianzas unitarias, respectivamente. Las proposiciones siguientes son casos particulares de aplicación del Lema anterior.

PROPOSICIÓN 1: (5.b'') puede escribirse como

$$z_t = [1 - \delta_1(L) - \beta\delta_2(L)L]^{-1}[\delta_2(L)L\varepsilon_{1t} + S_{2t}]$$

y aplicando el Lema, $c(L) = [1 - \delta_1(L) - \beta\delta_2(L)L]^{-1}$. Por tanto utilizando (A.1) y (A.3) se obtiene (22).

PROPOSICIÓN 2: Utilizando el Lema (32.a) y (32.b) pueden escribirse como

$$y_t = (1 - \beta(1))^{-1}[\alpha(1)z_t + \alpha^*(L)\Delta z_t - (\beta(1) - \beta^*(L)L)\Delta y_t + \varepsilon_{1t}] \quad (A.4)$$

y

$$z_t = [1 - (\beta(L) + \alpha(L)\delta_2(L)L) - \delta_1 L(1 - \beta(L))]^{-1}[\delta_2(L)L\varepsilon_{1t} + (1 - \beta(L))S_{2t}] \quad (A.5)$$

Aplicando (A.1) y (A.3) a (A.4) y a (A.5) obtenemos (33).

PROPOSICIÓN 3: El t-ratio de $\hat{\gamma}_2$ en (41) viene dado por

$$t_{\gamma_2} = A/B \quad (A.7)$$

donde

$$\begin{aligned} A = {} & [(\alpha_1\varepsilon_2 + \varepsilon_1)'(\alpha_1 - 1)S_{1-1} + S_{2-1})][\varepsilon_2'\varepsilon_2] - \\ & - [(\alpha_1\varepsilon_2 + \varepsilon_2][(\alpha_1 - 1)S_{1-1} + S_{2-1}]'\varepsilon_2 \\ B = {} & \sigma_1\{[(\alpha_1 - \lambda)S_{1-1} + S_{2-1}]'[(\alpha_1 - \lambda)S_{1-1} + S_{2-1}] - \\ & - [((\alpha_1 - \lambda)S_{1-1} + S_{2-1})'\varepsilon_2]'[((\alpha_1 - \lambda)S_{1-1} + S_{2-1})'\varepsilon_2]\} \end{aligned}$$

Aplicando (A.1) — (A.3) se obtiene (45) y multiplicando numerador y denominador de (A.7) por s, con $s \uparrow \infty$, se obtiene (47).

De forma similar, el t-ratio de $\hat{\delta}$ en (42) viene dado por

$$t_\delta = C/D \quad (A.8)$$

donde

$$\begin{aligned} C &= [(\alpha_1 - \lambda)\varepsilon_2 + \varepsilon_1]'[(\alpha_1 - \lambda)S_{1-1} + S_{2-1}] \\ D &= [(\alpha_1 - \lambda)^2\sigma_2^2 + \sigma_1^2]^{1/2}[(\alpha_1 - \lambda)S_{1-1} + S_{2-1}]'[(\alpha_1 - \lambda)S_{1-1} + S_{2-1}] \end{aligned}$$

Aplicando (A.1) — (A.3) se obtiene (46) y multiplicando numerador y denominador de (A.8) por s, con $s \uparrow \infty$, se obtiene (48).

Referencias

BANERJEE, A.; DOLADO, J.; HENDRY, D. F., y SMITH, G. (1986): «Exploring Equilibrium Relationships in Econometrics through Static Models: Some Monte Carlo Evidence», *Oxford Bulletin of Economics and Statistics*, 48, 253-279.

BANERJEE, A., y DOLADO, J. (1987): «Do We Reject Rational Expectations too Often?: Interpreting Evidence Using Nagar Expansions», *Economics Letters*, 24, 27-32.

— (1988): «Tests of the Life Cycle-Permanent Income Hypothesis in the Presence of Random Walks: Asymptotic Theory and Small Sample Interpretations», *Oxford Economic Papers*, 40, 610-633.

DICKEY, D. A., y FULLER, W. A. (1979): «Distribution of the Estimators for Autoregressive Time Series with a Unit Root», *Journal of the American Statistical Association*, 74, 427-431.

DOLADO, J., y ERICSSON, N. R. (1989): «Inference in Conditional Dynamic Models with Integrated Series» (preprint).

ENGLE, R. F.; HENDRY, D. F., y RICHARD, J. F. (1983): «Exogeneity», *Econometrica*, 51, 277-304.

ENGLE, R. F., y GRANGER, C. W. (1987): «Cointegration and Error Correction: Representation, Estimation and Testing», *Econometrica*, 55; 251-276.

FULLER, W. A. (1976): *Introduction Statistical Time Series*, John Wiley, New York.

GALBRAITH, J. W.; BANERJEE, A., y DOLADO, J. (1988): «Dynamic Specification with the General Error-Correction Form», *Oxford Bulletin of Economics and Statistics*, 52, 95-105.

HALL, A. D., y PAGAN, A. R. (1981): «The LIML and Related Estimators of an Equation with Moving Average Disturbances», *International Economic Review*, 22, 719-730.

HENDRY, D. F., y RICHARD, J. F. (1982): «On the Formulation of Empirical Models in Dynamic Econometrics», *Journal of Econometrics*, 20, 549-563.

HENDRY, D. F., y RICHARD, J. F. (1983): «The Econometric Analysis of Economic Time Series» (with Discussion), *International Statistical Review*, 51, 111-163.

HENDRY, D. F., y ERICSSON, N. R. (1987): «Assertion Without Empirical Basis: An Econometric Appraisal of Monetary Trends in the United Kingdom by Milton Friedman and Anna J. Schwartz», *Applied Economics Discussion Paper Series*, no. 25, University of Oxford (forthcoming in *American Economic Review*).

JARQUE, C., y BERA, A. (1980): «Efficient Tests for Normality, Homoskedasticity and Serial Independence of Regression Residuals», *Economics Letters*, 6, 255-259.

JOHANSEN, S. (1988): «Statistical Analysis of Cointegration Vectors», *Journal of Economic Dynamics and Control*, 12, 231-254.

JOHANSEN. S., y JUSELIUS, K. (1988): «Hypothesis Testing for Cointegration Vectors with and Application to the Demand for Money in Denmark and Finland», *Discussion Paper*, no. 8805, University of Copenhaguen.

KREMERS, J. J.; ERICSSON, N. R., y DOLADO, J.: «The Power of Cointegration Tests», Federal Reserve Board (preprint).

PARK, J. Y., y PHILLIPS, P. C. B. (1988): «Statistical Inference in Regressions with Integrated Processes: Part 1», *Econometric Theory*, 4, 468-497.

PHILLIPS, P. C. B. (1987): «Time Series Regression with Unit Roots», *Econometrica*, 55, 277-301.

PHILLIPS, P. C. B. (1988a): «Optimal Inference in Cointegrated Systems», Cowles Foundation. *Discussion Paper*, no. 866, Yale University.

PHILLIPS, P. C. B. (1988b): «Regression Theory for Near-Integrated Time Series», *Econometrica*, 56, 1021-1044.

PHILLIPS, P. C. B., PARK, J. Y. (1988): «Asymptotic Equivalence of Ordinary Least Squares and Generalised Least Squares in Regressions with Integrated Variables», *Journal of the American Statistical Association,* 83, 111-115.

PHILLIPS, P. C. B., y HANSEN, B. E. (1988): «Statistical Inference in Instrumental Variable Regression with $I(1)$ Processes», Cowles Foundation Paper 869.

SARGAN, J. D. (1980): «Some Tests of Dynamic Specification for a Single Equation», *Econometrica,* 48, 879-897.

SIMS, C. A.; STOCK, J. H., y WATSON, M. W. (1986): «Inference in Linear Time Series Models with some Unit Roots» (mimeo), Harvard University.

STOCK, J. H. (1987): «Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors», *Econometrica,* 55, 1035-1056.

WICKENS, M. R., y BREUSCH, T. S. (1988): «Dynamic Specification, the Long Run and the Estimation of Transformed Regression Models», *Economic Journal,* 98, 189-205.

WEST, K. D. (1988): «Asymptotic Normality when Regressors have a Unit Root», *Econometrica,* 56, 1397-1417.