



Universidad de la República
Facultad de Ciencias Sociales
DEPARTAMENTO DE ECONOMIA

Notas Docentes

Cointegración

Adriana Cassoni

Nota Docente No. 02

INDICE

	pág.
Introducción	1
Estacionariedad y relaciones de equilibrio	3
Modelos dinámicos	8
Regresiones espúreas	14
Procesos estocásticos: caracterización y propiedades	17
Pruebas de raíz unitaria	25
<i>El estadístico de Dickey-Fuller</i>	26
<i>Extensiones a la prueba de Dickey-Fuller</i>	30
<i>Extensiones al caso de pruebas de más de un parámetro</i>	37
<i>Extensión al caso de más de una raíz unitaria</i>	39
<i>Extensión al caso de raíces unitarias en el componente estacional</i>	41
Cointegración	43
<i>Definiciones y propiedades</i>	44
<i>Pruebas de cointegración</i>	50
<i>Estimación de modelos con variables cointegradas</i>	51
<i>Pruebas de hipótesis sobre los vectores de cointegración</i>	64
Comentarios finales	67
Referencias bibliográficas	68
Anexo	71

I. INTRODUCCION¹

Durante los años ochenta se produjeron importantes avances en la Teoría Econométrica que han permitido encauzar la práctica de la disciplina hacia nuevas direcciones. Uno de los aportes fundamentales es el relacionado al desarrollo de un cuerpo teórico aplicable al caso en que los procesos bajo estudio sean no estacionarios. Hasta ese momento, y dado que en realidad los fenómenos económicos en el tiempo son casi unánimemente de este tipo, el investigador debía realizar transformaciones de los mismos que fuesen estacionarias para poder llevar adelante análisis econométricos utilizando el cuerpo teórico existente, pero perdiendo información relevante en el proceso.

Dada la relativa contemporaneidad de dichos desarrollos, la literatura sobre el tema se encuentra dispersa en artículos de revistas y documentos de trabajo, existiendo algunas compilaciones más o menos extensas y un texto reciente de escasa difusión aún. Por ello, en este documento se pretende sintetizar los resultados más destacados del enfoque, de manera de facilitar la comprensión del mismo así como la lectura de la bibliografía existente.

El trabajo se organiza de la siguiente manera: en la próxima sección se definen los conceptos básicos que serán utilizados en el documento y que permiten caracterizar a las relaciones entre fenómenos económicos. Posteriormente, se exponen distintas formulaciones de modelos dinámicos y su interpretación, que más tarde se vincularán a la definición de cointegración. En la cuarta parte se plantea el problema de regresiones espúreas entre series no estacionarias. El capítulo V se dedica a la caracterización de procesos estocásticos en términos de sus raíces y a sintetizar algunos resultados de interés. Los capítulos mencionados hasta

¹ Este trabajo resume los temas desarrollados en el seminario llevado a cabo en el Area de Econometría. Por ello, los aportes realizados por Gabriela Fachola y Carlos Mendive han contribuido a mejorar sustancialmente su contenido. Asimismo, se agradecen los comentarios de Alvaro Forteza y Federico Echenique.

aquí permiten enmarcar la problemática general vinculada a procesos integrados. A partir de allí, se desarrollan diversas pruebas de raíz unitaria (capítulo VI), necesarias para determinar qué tipo de proceso se está analizando. Finalmente, en la sección VII se analizan las pruebas de cointegración, los métodos de estimación apropiados y algunos puntos de interés.

Es importante tener en cuenta que el grado de complejidad del documento aumenta sustancialmente en las últimas secciones. Así, mientras los capítulos I a V son de relativa sencillez, siendo su objetivo dar una interpretación mayormente intuitiva de los conceptos, los restantes requieren del manejo fluido de varios conceptos estadísticos. Sin embargo, aún en estos últimos, se intenta brindar una aproximación intuitiva y operativa de los temas.

II. ESTACIONARIEDAD Y RELACIONES DE EQUILIBRIO

Una clasificación de interés de las relaciones entre fenómenos económicos en el tiempo, es la que distingue entre aquellas de corto y largo plazo, de acuerdo a la duración de las mismas. Estas últimas son las que se asocian con el concepto de *equilibrio*, definido como un estado en el que no existe una tendencia inherente al cambio². Por contraposición, un estado de *desequilibrio* anida su propia destrucción. Un equilibrio puede, a su vez, ser *estable* o *inestable*, dependiendo de la respuesta del fenómeno ante una perturbación externa. Si luego de la perturbación existe una tendencia a retornar al punto de equilibrio inicial se estará en la primera situación, mientras que en otro caso se tratará de un equilibrio inestable.

La noción estadística de equilibrio es equivalente a la de un equilibrio estable, y se asocia por lo tanto al concepto de *estacionariedad*: existe una relación de equilibrio entre dos variables si y sólo si la desviación de esa relación en el tiempo es un proceso débilmente estacionario. Para definir este concepto formalmente conviene definir previamente un proceso estocástico:

Definición

Dado un espacio de probabilidad, en el que se denominará \mathcal{S} al conjunto de eventos posibles, y dado un subconjunto \mathbb{T} de los reales, es posible definir una función $z(.,.) : \mathcal{S} \times \mathbb{T} \rightarrow \mathbb{R}$. Así, se denomina proceso estocástico a la secuencia ordenada de variables aleatorias $\{z_t, t \in \mathbb{T}\}$.

Definición

Un proceso estocástico $\{z_t, t \in \mathbb{T}\}$ es débilmente estacionario si posee las siguientes características:

$$E(z_t) = \mu \quad \forall t \quad V(z_t) = \sigma^2 \quad \forall t \quad \text{cov}(z_t, z_s) = h(|t-s|) \quad \forall t, s \quad t \neq s$$

²Esta definición resulta incompleta, ya que no incorpora la posibilidad de estados de equilibrio dinámicos. Se mantiene en aras de la simplicidad del argumento.

Esto significa que en un proceso débilmente estacionario (o estacionario de segundo orden), la media y la varianza son constantes (independientes de t), y la covarianza no depende de t ni de s sino de la distancia entre ellos. A estos procesos se les denomina también *integrados de orden cero*, lo que se expresa sintéticamente como $I(0)$. Además, la estacionariedad de z_t implica que el proceso posee una representación *ARMA* (autorregresiva y de medias móviles) estacionaria y que carece de componentes determinísticos (Spanos, 1986).

Obsérvese que este concepto es menos restrictivo que el de idéntica distribución de las variables, que implica que todos los momentos son constantes y libres de t . Sin embargo, en el caso de distribuciones normales ambos coinciden, ya que la distribución queda totalmente definida por los momentos 1 y 2.

Con esta definición es posible expresar la noción de equilibrio expuesta más arriba como:

Definición

Sean x_1 y x_2 dos variables aleatorias tales que $f(x_1, x_2) = 0$. Se dice que $f(x_1, x_2) = 0$ es una relación de equilibrio si dada la relación: $E_t = f(x_{1t}, x_{2t})$, $\{E_t, t \in \mathbb{T}\}$ es un proceso débilmente estacionario.

Lo anterior muestra claramente que la consideración de una relación como de equilibrio implica que ante perturbaciones exógenas al sistema es posible observar desviaciones de la relación en el corto plazo, pero éstas tenderán a desaparecer: existen fuerzas que impiden que la discrepancia de corto plazo presente una tendencia sistemática a crecer. Por contraposición, si una perturbación tiene efectos permanentes sobre el sistema, la diferencia E_t observada en cada momento del tiempo no puede ser estacionaria.

Finalmente, cabe anotar que las discrepancias entre los valores observados y los valores de equilibrio proveen de información

relevante para el analista, ya que son un indicador útil de cómo es la trayectoria de las variables hacia el equilibrio. Por ello, en cierto tipo de modelos dinámicos, se los incorpora como una variable explicativa, denominándola *mecanismo de corrección de error*.

Definición

Un proceso no estacionario $\{x_t, t \in \mathbb{T}\}$ es *integrado de orden 1* - $x_t \sim I(1)$ - si su primera diferencia es un proceso estacionario, es decir, si $\{\Delta x_t\} \sim I(0)$.

En general, $x_t \sim I(k)$ si siendo la serie no estacionaria es necesario diferenciarla k veces y no menos de k veces para obtener una serie $I(0)$:

$$x_t \sim I(k) \text{ sii } \Delta^k x_t \sim I(0) \text{ y para } h < k \Delta^h x_t \sim I(a) \text{ } a > 0$$

Se deduce que el hecho que la k -ésima diferencia de una serie sea estacionaria no implica que la serie es integrada de orden k , ya que puede ocurrir que $\Delta^s x_t \sim I(0)$ con $s < k$.

Es importante recordar que el hecho que la primer diferencia de una serie integrada sea un proceso estacionario implica, por definición, que la serie original posee una raíz unitaria. A modo de ejemplo, en el caso en que se tiene un proceso autorregresivo de primer orden: $x_t = \alpha x_{t-1} + \varepsilon_t$ con $\varepsilon_t \sim iid$, se podría re-escribir la ecuación como $\Delta x_t = (\alpha - 1)x_{t-1} + \varepsilon_t$. Si x_t es no estacionaria, entonces la única forma para que Δx_t lo sea es que $(\alpha - 1)$ sea nulo, es decir, que α sea igual a 1 (para un desarrollo formal del tema, véase el capítulo 8 de Spanos, 1986).

De aquí en adelante, se denominará *procesos integrados* sólo a aquellos cuyo orden de integración sea estrictamente mayor que 0.

Algunas de las propiedades que cumplen los procesos $I(k)$ con $k \geq 0$ son las siguientes:

1. Si $x_t \sim I(0) \Rightarrow (a \pm bx_t) \sim I(0)$ para $(a, b) \neq (0, 0)$
2. Si $x_t \sim I(1) \Rightarrow (a \pm bx_t) \sim I(1)$ para $(a, b) \neq (0, 0)$

3. Si $x_t \sim I(0)$ y $y_t \sim I(0) \Rightarrow (ax_t \pm by_t) \sim I(0)$ para $(a,b) \neq (0,0)$
 4. Si $x_t \sim I(1)$ y $y_t \sim I(0) \Rightarrow (ax_t \pm by_t) \sim I(1)$ para $(a,b) \neq (0,0)$, es decir, la propiedad de ser $I(1)$ es dominante
 5. Si $x_t \sim I(1)$ y $y_t \sim I(1) \Rightarrow (ax_t \pm by_t) \sim I(1)$ para $(a,b) \neq (0,0)$
- en general

Definición

Sean $\{x_t, t \in \mathbb{T}\}$ y $\{y_t, t \in \mathbb{T}\}$ dos procesos integrados de primer orden. Se dice que $\{x_t, t \in \mathbb{T}\}$ y $\{y_t, t \in \mathbb{T}\}$ están *cointegradas* de orden $(1,1)$ - $(x_t, y_t) \sim CI(1,1)$ - si se cumple que una combinación lineal de los procesos es estacionaria:

$$x_t \sim I(1), y_t \sim I(1) \text{ y } (ax_t \pm by_t) \sim I(0) \text{ para algún } (a,b) \neq (0,0)$$

En general: sean $x_t \sim I(m)$, $y_t \sim I(n) \Rightarrow (x_t, y_t) \sim CI(m,n)$ con $m \geq n$, si $(ax_t + by_t) \sim I(m-n)$ para algún $(a,b) \neq (0,0)$

Con estas propiedades y la definición de cointegración es posible ahondar más en la noción de equilibrio en economía y su relación con el orden de integración de las series involucradas, que es la noción estadística del equilibrio.

Supóngase que se desea analizar la existencia de una relación de equilibrio entre dos variables, y y x , de la forma: $y + ax = 0$. Si las series involucradas - x_t y y_t - son estacionarias, la discrepancia entre los valores observados y los valores de equilibrio $E_t = y_t + ax_t$ será, por la propiedad 3, estacionaria *independientemente* de si la relación teórica es válida o no. Sin embargo, es posible estimar el parámetro a y aplicar todo el instrumental estadístico habitual de procesos estacionarios para buscar evidencia que sustente la existencia de dicha relación. Esto no resulta útil, por el contrario, si las series son no estacionarias ya que, de acuerdo a la propiedad 5, E_t será en general no estacionario y, por lo tanto, el uso de los estadísticos de prueba tradicionales no es válido. Una excepción a ello es el caso en que *las series están cointegradas*. Es evidente, entonces, que cuando las series son integradas el análisis de cointegración debe *preceder* a la inferencia estadística.

Finalmente, es posible obtener una alta correlación entre variables integradas que no se debe a la existencia de una relación entre ellas sino, en realidad, a que ambas series poseen patrones similares en la media o la varianza, dando lugar a lo que se conoce habitualmente como *regresiones espúreas*.

De lo anterior se desprende la importancia de realizar pruebas de hipótesis para identificar el orden de integración de las series involucradas antes de realizar cualquier estimación; la necesidad de analizar la posible existencia de cointegración en caso de que dichas series sean no estacionarias y la relevancia de conocer las propiedades de los procesos integrados y su interpretación económica, así como las implicaciones de las distintas formulaciones dinámicas de un mismo modelo. Estos tópicos se desarrollarán en los capítulos siguientes.

III. MODELOS DINAMICOS

La existencia de una relación de largo plazo entre un conjunto de variables $(y, x_1, x_2, \dots, x_k)$ se plantea generalmente en el mismo momento del tiempo: $y_t = x_t' \beta$. Al extender este razonamiento a los valores realizados, sin embargo, se estaría implicando que el efecto de las variables exógenas x_t sobre la endógena y_t se produce y se agota inmediatamente. En general, lo anterior no se ajusta a la realidad, ya que si bien la relación se ha de cumplir en el largo plazo, en el corto plazo existen desajustes y rezagos que determinan que los valores pasados de las variables cumplan un rol importante en la determinación del fenómeno en cada momento del tiempo. De ahí la importancia de las formulaciones dinámicas para la explicación de las relaciones de largo plazo. Estas, además, proveen de información relevante sobre la forma en que se producen los ajustes en el corto plazo, que dan lugar a la trayectoria hacia el equilibrio.

Los modelos dinámicos pueden plantearse de diversas formas, equivalentes entre sí en términos de los valores de los multiplicadores de largo plazo, pero que difieren respecto a los parámetros de corto plazo que se identifican a partir de cada una. El uso de una u otra formulación dependerá así, exclusivamente, del interés específico del investigador. A continuación se presentan cuatro modelos dinámicos: el modelo de rezagos distribuidos (MRD); el modelo de corrección de error (MCE); la formulación de Bewley y la de Bårdsen.

El modelo de rezagos distribuidos

El $MRD(m, n; p)$ es la formulación más general, siendo m el número de rezagos de la variable dependiente y n el máximo número de rezagos de las p exógenas. Así:

$$MRD(m, n; p): y_t = \alpha_0 + \sum_{i=1}^m \alpha_i y_{t-i} + \sum_{j=1}^p \sum_{i=0}^n \beta_{ji} x_{j,t-i} + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma^2)$$

o en forma polinómica: $\alpha(L)y_t = \alpha_0 + \sum_{j=1}^p \beta_j(L)x_{j,t} + \varepsilon_t$

con $\alpha(L) = 1 - \sum_{i=1}^m \alpha_i L^i$ y $\beta_j(L) = \sum_{i=0}^n \beta_{ji} L^i$

Si se supone que las variables involucradas son conjuntamente estacionarias (son estacionarias cada una individualmente o están cointegradas), los valores de largo plazo equivalen a la esperanza incondicional de las variables y el equilibrio se identifica con el estado en el que todo cambio ha cesado. Así, la solución de largo plazo para el $MRD(m,n;p)$ es:

$$E(y_t) = \left(1 - \sum_{i=1}^m \alpha_i\right)^{-1} \left[\alpha_0 + \sum_{j=1}^p \sum_{i=0}^n \beta_{ji} E(x_{jt})\right] =$$

$$= [\alpha(1)]^{-1} \left[\alpha_0 + \sum_{j=1}^p \beta_j(1) E(x_{jt})\right] = \theta_0 + \sum_{j=1}^p \theta_j E(x_{jt})$$

siendo $\theta_j = [\alpha(1)]^{-1} \beta_j(1)$ para $j=1,2,\dots,p$ y $\theta_0 = [\alpha(1)]^{-1} \alpha_0$

En esta formulación, entonces, para identificar los multiplicadores de largo plazo de y con respecto a cada x_j habrá que calcular los θ_j . Por otra parte, la estimación se puede realizar sin problemas por mínimos cuadrados ordinarios, aunque la varianza de los estimadores de los multiplicadores de largo plazo no se obtiene en forma exacta: al ser éstos transformaciones no lineales de los estimadores en la ecuación original, hay que utilizar una aproximación a la varianza del cociente, que viene dada por:

$var(\hat{\theta}) \approx J_f' var(\hat{\alpha}, \hat{\beta}) J_f$ siendo J_f el jacobiano de la transformación

$$f_j(\alpha, \beta) = \left(1 - \sum_{i=1}^m \alpha_i\right)^{-1} \cdot \sum_{i=0}^n \beta_{ji}$$
 para $j = 1, 2, \dots, p$

O sea: $J_f = (\partial f / \partial \alpha_1, \dots, \partial f / \partial \alpha_m, \partial f / \partial \beta_{10}, \dots, \partial f / \partial \beta_{1n}, \dots, \partial f / \partial \beta_{pn})$

El modelo de corrección de error

Esta formulación se puede derivar directamente del MRD restando a ambos lados y_{t-1} y sumando y restando del lado derecho rezagos de x_t :

$$MCE(r, n; p): \quad \Delta y_t = \alpha_0 + \sum_{i=1}^r \eta_i \left(y_{t-1} - \sum_{j=1}^p x_{jt-1} \right) + \sum_{j=1}^p \beta_{j0} \Delta x_{jt} + \\ + \sum_{j=1}^p \sum_{i=1}^r \zeta_{ji} x_{jt-1} + \sum_{j=1}^p \sum_{i=r+1}^n \beta_{ji} x_{jt-1} + \sum_{i=r+1}^m \alpha_i y_{t-1} + \varepsilon_t$$

con:

$$\varepsilon_t \sim iid(0, \sigma^2)$$

$$r \leq m, \quad r = \min(m, n)$$

$$\eta_1 = \alpha_1 - 1 \quad ; \quad \eta_i = \alpha_i \quad \text{para } i = 2, 3, \dots, r$$

$$\zeta_{j1} = \alpha_1 - 1 + \beta_{j0} + \beta_{j1} \quad ; \quad \zeta_{ji} = \alpha_i + \beta_{ji} \quad \text{para } i = 2, 3, \dots, r$$

Planteado de esta forma, los r términos $(y_{t-1} - \sum_{j=1}^p x_{jt-1})$ reflejarían el efecto de los desajustes en el corto plazo - desde $t=t-r$ a $t=t-1$ - sobre la variable de interés y_t . Esto no implica, sin embargo, que en equilibrio se observe: $y_{t-1} = \sum_{j=1}^p x_{jt-1}$, es decir, que los coeficientes de la relación de largo plazo sean todos iguales a 1 (condición de *homogeneidad*), ya que por cada uno de estos sumandos se ha añadido a la ecuación un rezago de x_{jt} , que es lo que permite que, al reordenarlos, se obtenga el término de corrección de error que mide efectivamente el desajuste de corto plazo:

$$\sum_{i=1}^r \eta_i \left(y_{t-1} - \sum_{j=1}^p x_{jt-1} \right) + \sum_{i=1}^r \sum_{j=1}^p \zeta_{ji} x_{jt-1} = \sum_{i=1}^r \eta_i \left(y_{t-1} - \sum_{j=1}^p (1 - \zeta_{ji} / \eta_i) x_{jt-1} \right) = \\ = \sum_{i=1}^r \eta_i \left(y_{t-1} - \sum_{j=1}^p \nu_{ji} x_{jt-1} \right) \quad \text{con } \nu_{ji} = 1 - \zeta_{ji} / \eta_i$$

$$\text{De esta forma, } \sum_{i=1}^r \eta_i \nu_{ji} = \sum_{i=1}^r \eta_i - \zeta_{ji} = -\beta_j(1)$$

Reordenando adecuadamente se obtienen los multiplicadores de largo plazo como:

$$\theta_j = - \left(\sum_{i=1}^r \eta_i + \sum_{i=r+1}^m \alpha_i \right)^{-1} \left[\sum_{i=1}^r (\zeta_{ji} - \eta_i) + \sum_{i=r+1}^n \beta_{ji} \right] = \beta_j(1) / \alpha(1)$$

La ventaja que presenta este modelo frente al MRD es que en su formulación aparecen explícitamente los coeficientes η_1 , que se interpretan como el "ajuste de corto plazo al desequilibrio", o sea, el impacto de los desequilibrios anteriores sobre la variable de interés, lo cual a su vez se puede asociar a la velocidad de ajuste de y a las discrepancias pasadas entre y y x . Sin embargo, los multiplicadores de largo plazo deben calcularse a partir de los coeficientes originales. Nuevamente, la variabilidad de los mismos debe ser aproximada y esto se realiza de acuerdo a:

$var(\hat{\theta}) \cong J_g' var(\hat{\eta}, \hat{\zeta}) J_g$ siendo J_g el jacobiano de la transformación

$$g_j(\eta, \zeta) = -\left(\sum_{i=1}^r \eta_i + \sum_{i=r+1}^m \alpha_i\right)^{-1} \left[\sum_{i=1}^r (\zeta_{j1} - \eta_i) + \sum_{i=r+1}^n \beta_{j1} \right]$$

El modelo de Bewley

Este modelo se plantea, al igual que el MRD, en términos del nivel de la variable dependiente y resulta de restarle a éste $\sum_{i=1}^m \alpha_i y_t$ en ambos lados de la igualdad y de sumarle y restarle del lado derecho $\sum_{j=1}^p \beta_j x_{jt}$:

$$y_t = \lambda \alpha_0 - \lambda \sum_{i=1}^m \alpha_i (y_t - y_{t-1}) + \sum_{j=1}^p (\lambda \sum_{i=0}^n \beta_{ji}) x_{jt} - \sum_{j=1}^p \lambda \left(\sum_{i=1}^n \beta_{ji} (x_{jt} - x_{jt-1}) \right) + \lambda \varepsilon_t$$

con:

$$\varepsilon_t \sim iid(0, \sigma^2) \quad \text{y} \quad \lambda = [\alpha(1)]^{-1}$$

Los multiplicadores de largo plazo se obtienen directamente de la estimación, al observar que son precisamente los coeficientes de x_{jt} de acuerdo a la relación: $\theta_j = \lambda \sum_{i=0}^n \beta_{ji}$

Nótese que en este caso la estimación debe realizarse por variables instrumentales para obtener estimadores consistentes de los parámetros, ya que y_t , la variable dependiente, está incluida entre los regresores a través del término Δy_t . Por lo mismo, la varianza se obtiene de acuerdo al procedimiento habitual de

estimación por variables instrumentales.

El modelo de Bårdsen

Se obtiene de sustraer al MRD y_{t-1} de ambos lados de la ecuación

y de sumar y restar del lado derecho $\sum_{j=1}^p \sum_{i=0}^{n-1} \beta_{ji}^* x_{jt-i-1}$:

$$\Delta y_t = \alpha_0 + \sum_{i=1}^{m-1} \alpha_i^* \Delta y_{t-i} + \sum_{j=1}^p \sum_{i=0}^{n-1} \beta_{ji}^* \Delta x_{jt-i} + \alpha_m^* y_{t-m} + \sum_{j=1}^p \beta_{jn}^* x_{jt-n} + \varepsilon_t$$

donde:

$$\varepsilon_t \sim iid(0, \sigma^2)$$

$$\alpha_i^* = \sum_{h=1}^i \alpha_h - 1 = \sum_{h=1}^i \eta_h \quad i = 1, 2, \dots, m$$

$$\beta_{ji}^* = \sum_{h=0}^i \beta_{jh} \quad i = 1, 2, \dots, n$$

Como se observa en las definiciones de α^* y β^* , los coeficientes de esta formulación miden el impacto acumulado de los ajustes a desequilibrios pasados, es decir, la suma acumulada de los efectos de los términos de corrección de error en el MCE.

Por otra parte, los multiplicadores de largo plazo se obtienen como:

$$\theta_j = \beta_{jn}^* / -\alpha_m^*$$

La varianza del estimador del multiplicador se calcula utilizando la fórmula:

$$\begin{aligned} \text{var}(\hat{\theta}_j) &= (-1/\hat{\alpha}_m^*)^2 \text{var}(\hat{\beta}_{jn}^*) + \left[\hat{\beta}_{jn}^* / (\hat{\alpha}_m^*)^2 \right]^2 \text{var}(\hat{\alpha}_m^*) + \\ &+ 2(-1/\hat{\alpha}_m^*) \left[\hat{\beta}_{jn}^* / (\hat{\alpha}_m^*)^2 \right] \text{cov}(\hat{\beta}_{jn}^*, \hat{\alpha}_m^*) \end{aligned}$$

Parece conveniente ejemplificar las formulaciones anteriores para

el caso más sencillo en que $p=n=m=1$. Bajo estas condiciones los modelos serían:

$$\text{MRD: } y_t = \alpha_0 + \alpha_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

con $\theta_1 = (\beta_0 + \beta_1)/(1 - \alpha_1)$ el multiplicador de largo plazo.

$$\text{MCE: } \Delta y_t = \nu_0 + \nu_1 (y_{t-1} - \theta_1 x_{t-1}) + \nu_2 \Delta x_t + \varepsilon_t$$

con $\nu_0 = \alpha_0$; $\nu_1 = \alpha_1 - 1$; $\nu_2 = \beta_0$; $\theta_1 =$ el multiplicador de largo plazo.

Su forma equivalente es:

$$\Delta y_t = \nu_0 + \nu_1 (y_{t-1} - x_{t-1}) + \nu_2 \Delta x_t + \nu_3 x_{t-1} + \varepsilon_t$$

con $\nu_0 = \alpha_0$; $\nu_1 = \alpha_1 - 1$;
 $\nu_2 = \beta_0$; $\nu_3 = \nu_1 (1 - \theta_1) = (\alpha_1 - 1)(1 - \theta_1) = \alpha_1 - 1 + \beta_0 + \beta_1$
 $\theta_1 = (\nu_1 - \nu_3)/\nu_1$ el multiplicador de largo plazo

$$\text{Bewley: } y_t = \mu_0 + \mu_1 \Delta y_t + \mu_2 x_t + \mu_3 \Delta x_t + u_t$$

con $\mu_0 = \lambda \alpha_0$; $\mu_1 = -\lambda \alpha_1$; $\mu_2 = \lambda(\beta_0 + \beta_1)$; $\mu_3 = -\lambda \beta_1$; $\lambda = (1 - \alpha_1)^{-1}$

$u_t = \lambda \varepsilon_t$; $\theta_1 = \mu_2$ el multiplicador de largo plazo

$$\text{Bårdsen: } \Delta y_t = \omega_0 + \omega_1 y_{t-1} + \omega_2 \Delta x_t + \omega_3 x_{t-1} + \varepsilon_t$$

con $\omega_0 = \alpha_0$; $\omega_1 = \alpha_1 - 1$; $\omega_2 = \beta_0$; $\omega_3 = \beta_1 + \beta_0$

$\theta_1 = -\omega_3/\omega_1$ el multiplicador de largo plazo.

IV. REGRESIONES ESPUREAS

Las regresiones espúreas o "sin sentido" se obtienen cuando, aún no existiendo relación alguna entre las variables involucradas en el modelo propuesto, la correlación entre ellas es alta, de forma que las pruebas estadísticas y el R^2 indican, equivocadamente, que el modelo es estadísticamente correcto. Este problema, señalado tempranamente por Yule (1926) y desarrollado más recientemente por Granger y Newbold (1974), ocurre de forma muy frecuente cuando las variables son integradas, ya que es posible que posean una tendencia temporal similar, que es el origen de la alta correlación observada. La razón por la cual las pruebas estadísticas estándar no son aplicables es que cuando las series son no estacionarias los estadísticos *t-Student* o *F* divergen en distribución al aumentar el tamaño de la muestra, invalidando así el uso de las tablas habituales. Esto implica, además, que al crecer la muestra la tasa de rechazo aumentará también, con lo cual el rechazo de la hipótesis de no relación entre las variables se vuelve más y más frecuente al aumentar el número de observaciones disponibles.

Granger y Newbold plantearon, en el artículo mencionado, un ejemplo que ilustra esta situación. Sean los procesos generadores de información de dos variables x_t y y_t los siguientes:

$$x_t = x_{t-1} + u_t \quad u_t \sim iid(0, \sigma_u^2) \quad (1)$$

$$y_t = y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma_\varepsilon^2) \quad (2)$$

$$E(u_t \varepsilon_s) = 0 \quad \forall t, s$$

Dado el planteo anterior, resulta que x_t y y_t son caminos aleatorios no correlacionados. Siendo así, al especificar la regresión: $y_t = \beta_0 + \beta_1 x_t + e_t$ sería esperable que la hipótesis nula $H_0: \beta_1 = 0$ no se rechace, es decir, que $\hat{\beta}_1$ sea estadísticamente no significativo. Sin embargo, dado que las series son integradas de orden 1, esto no ocurrirá generalmente: bajo la hipótesis nula, se

propone $y_t = \beta_0 + e_t$, con e_t un ruido blanco, lo cual implica que y_t es, a su vez, un ruido blanco; bajo la alternativa, se plantea un modelo falso para la variable. En este segundo caso, puede ocurrir además, que la no estacionariedad de x_t se proyecte sobre y_t . De esta forma, el resultado más probable es, entonces, que se rechazará incorrectamente la no relación entre las variables, simplemente porque no es posible sustentar que y_t sea un ruido blanco. Esto sucede porque, en realidad, se están contrastando dos hipótesis conjuntamente: la no relación entre las variables y la estacionariedad de y_t . Al ser falsa la segunda proposición se rechaza la *totalidad* de la hipótesis nula.

Con el modelo planteado arriba se realizaron simulaciones de Monte Carlo (Banerjee et al., 1993) que permiten reproducir las distribuciones de los estadísticos de prueba. De ello se deducen algunos puntos de interés:

(1) Si bien la forma de la distribución se parece a una *t-Student*, la media del estadístico no es 0 ni la varianza es unitaria.

(2) La desviación estándar es tan grande que la probabilidad de rechazar H_0 siendo cierta es muy alta: para obtener un nivel de significancia en la prueba del 5% sería necesario que el valor crítico de tablas fuese ± 14.5 en vez de 1.96.

(3) La desviación estándar del estadístico aumenta con el tamaño muestral, de forma que el problema es mayor cuanto mayor es la muestra.

(4) La esperanza del estimador de β_1 es significativamente distinta de 0 sólo en muestras grandes, pero el sesgo no cambia demasiado al variar el tamaño de la muestra. La desviación estándar, a su vez, no disminuye con aumentos en el número de observaciones, contrariamente a lo que ocurre al trabajar con series estacionarias.

Por otra parte, las simulaciones realizadas por Yule en 1926 mostraban que la distribución de la correlación muestral (R) en el

caso de variables integradas de primer orden, tiene una gran concentración de frecuencia en las colas, de modo que los valores de R lejanos a 0 son mucho más probables de obtener que en el caso de series estacionarias. El resultado es aún peor si las series son integradas de segundo orden, obteniéndose una distribución de frecuencias de R en forma de U , con lo cual lo más probable es obtener correlaciones de ± 1 , independientemente de que esto sea cierto.

Finalmente, Phillips (1986) demostró que el estadístico de Durbin-Watson converge a 0 al aumentar el tamaño de la muestra si las series no están relacionadas. Esto confirmó la observación realizada con anterioridad por algunos autores sobre la obtención de altos valores de R^2 junto con bajos valores del estadístico DW cuando la regresión es espúrea.

En la búsqueda de soluciones distintas a la diferenciación de las series cuando éstas son integradas, se propuso incorporar tendencias determinísticas en los modelos, que capturaran el componente no estacionario de las variables involucradas. Sin embargo, esto no resulta conveniente ya que el estadístico a utilizar para la prueba sobre la significancia del coeficiente de la tendencia no se distribuye *t-Student*, como ya se señaló anteriormente. Con ello, la probabilidad de rechazar erróneamente que dicho parámetro es nulo es, nuevamente, muy alta. Aún más, el problema de la regresión espúrea se mantiene incluso al incorporar una tendencia.

Lo expuesto hasta aquí ilustra la importancia que reviste realizar un análisis exhaustivo de las series involucradas antes de estimar un modelo. Detectar si son no estacionarias y, cuando corresponda, el orden de integración de cada una de ellas, es fundamental para decidir correctamente qué tipo de técnica es adecuada en cada caso. De esta forma, en lo que sigue se discutirán algunos de los métodos para determinar el orden de integración de las series y, posteriormente, cómo trabajar con series no estacionarias.

V. PROCESOS ESTOCÁSTICOS: CARACTERIZACIÓN Y PROPIEDADES

Sea el proceso generador de información representable por:

$$y_t = \rho y_{t-1} + \varepsilon_t \quad \varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$$

Si $|\rho| < 1$ el proceso es, al menos asintóticamente, estacionario débil.

Si $|\rho| = 1$ el proceso posee una raíz unitaria y es integrado de orden 1.

Si $|\rho| > 1$ el proceso es explosivo.

Utilizando el operador de rezagos (L), se puede expresar y_t como:

$$y_t = \varepsilon_t / (1 - \rho L) = \varepsilon_t (1 + \rho L + \rho^2 L^2 + \dots) = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}$$

En cada uno de los casos planteados más arriba, es posible obtener una *forma final* del proceso, en función de su valor inicial y el proceso ε_t , que resulta de utilidad para su caracterización a través de sus primeros momentos poblacionales. Suponiendo que el proceso se extiende desde el momento $t=0$, la fórmula anterior se reduce a la forma final:

$$Y_t = \sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j} + \rho^t Y_0$$

Caso 1: $|\rho| < 1$

Los primeros momentos del proceso son :

$$E(Y_t) = E\left(\sum_{j=0}^{t-1} \rho^j \varepsilon_{t-j} + \rho^t Y_0\right) = \sum_{j=0}^{t-1} \rho^j E(\varepsilon_{t-j}) + \rho^t Y_0 = \rho^t Y_0$$

Así, $E(y_t)$ tiende a 0 cuando t tiende a ∞

$$V(Y_t) = \sum_{j=0}^{t-1} \rho^{2j} \sigma_\varepsilon^2 = \sigma_\varepsilon^2 (1 - \rho^{2t}) / (1 - \rho^2)$$

Así, $V(y_t)$ tiende a $\sigma_\varepsilon^2 / (1 - \rho^2)$ cuando t tiende a ∞

$$\text{Cov}(Y_t, Y_{t-s}) = \rho^s V(Y_{t-s}) = \rho^s \sigma_\varepsilon^2 (1 - \rho^{2(t-s)}) / (1 - \rho^2)$$

Así, $Cov(y_t y_{t-s})$ tiende a $\rho^s \sigma_\varepsilon^2 / (1 - \rho^2)$ cuando t tiende a ∞ y s está dado.

De esta forma, si el valor inicial es nulo ($y_0 = 0$), el proceso es estacionario en media aunque no en varianza. Sin embargo, asintóticamente, dado que ninguno de los momentos depende del punto del tiempo en que se calcula, el proceso estocástico es estacionario débil.

Caso 2: $|\rho| = 1$

A este proceso se lo conoce con el nombre de *camino aleatorio*. Si bien es integrado de primer orden tanto cuando ρ toma un valor positivo como negativo, las características de cada caso son distintas. Se analizará, por tanto, cada proceso separadamente.

2.a: $\rho = 1$

La forma final del proceso es: $y_t = \sum_{j=0}^{t-1} \varepsilon_{t-j} + Y_0$

Sus primeros momentos son:

$$E(y_t) = \sum_{j=0}^{t-1} E(\varepsilon_{t-j}) + y_0 = y_0$$

$$V(y_t) = V(\sum_{j=0}^{t-1} \varepsilon_{t-j}) = t\sigma_\varepsilon^2$$

$$Cov(y_t y_{t-s}) = (t-s)\sigma_\varepsilon^2$$

El camino aleatorio resulta ser estacionario en media, pero no estacionario en varianza, ni siquiera asintóticamente ya que la varianza y la covarianza divergen cuando t tiende a infinito. Además, la forma de la covarianza implica que existe una alta correlación entre observaciones sucesivas, que disminuye al aumentar la distancia s en el tiempo.

Caso 2.b: $\rho = -1$

La forma final del proceso es: $y_t = \sum_{j=0}^{t-1} (-1)^j \varepsilon_{t-j} + (-1)^t y_0$

$$E(y_t) = \sum_{j=0}^{t-1} (-1)^j E(\varepsilon_{t-j}) + (-1)^t y_0 = (-1)^t y_0$$

$$V(y_t) = V(\sum_{j=0}^{t-1} (-1)^j \varepsilon_{t-j}) = \sum_{j=0}^{t-1} (-1)^{2j} \sigma_\varepsilon^2 = t \sigma_\varepsilon^2$$

$$\text{Cov}(y_t, y_{t-s}) = (-1)^s (t-s) \sigma_\varepsilon^2$$

Si $y_0 = 0$, el proceso es estacionario en media, pero no en varianza, ni siquiera asintóticamente. Si $y_0 \neq 0$, sin embargo, no será estacionario en media tampoco, ya que ésta toma valores positivos y negativos de acuerdo a si t es par o impar. La forma de la covarianza determina, en este caso, que el proceso muestre fluctuaciones fuertes, dado el signo opuesto de las sucesivas correlaciones.

Caso 3: $|\rho| > 1$

$$E(y_t) = \rho^t y_0$$

$$V(y_t) = \sigma_\varepsilon^2 (1 - \rho^{2t}) / (1 - \rho^2)$$

$$\text{Cov}(y_t, y_{t-s}) = \rho^s \sigma_\varepsilon^2 (1 - \rho^{2(t-s)}) / (1 - \rho^2)$$

Los primeros momentos dependen del tiempo, con lo cual el proceso no es estacionario débil, con excepción del caso en que $y_0 = 0$, que llevaría a que en pequeñas muestras el proceso fuese estacionario en media. Cuando t tiende a ∞ , por el contrario, todos los momentos divergen, por lo cual el proceso es explosivo.

Un segundo tipo de procesos que interesa analizar son los caminos aleatorios que además poseen una ordenada al origen ("drift") y/o una tendencia determinística, distinguiendo nuevamente entre el caso en que la raíz es $+1$ y -1 .

Caso 4: $\rho = 1 \quad Y_t = \mu + Y_{t-1} + \varepsilon_t$

La forma final es: $y_t = t\mu + \sum_{j=0}^{t-1} \varepsilon_{t-j} + y_0$

$$E(y_t) = y_0 + t\mu$$

$$V(y_t) = t\sigma_\varepsilon^2$$

$$\text{Cov}(y_t, y_{t-s}) = (t-s)\sigma_\varepsilon^2$$

De esta forma, se observa que el proceso posee en realidad una tendencia estocástica, por lo cual no es asintóticamente estacionario en media como en el caso 2.

Caso 5: $\rho = -1$ $y_t = \mu - y_{t-1} + \varepsilon_t$

La forma final es: $y_t = \sum_{j=0}^{t-1} (-1)^j \mu + \sum_{j=0}^{t-1} \varepsilon_{t-j} + (-1)^t y_0$

$$E(y_t) = y_0 \quad \text{si } t \text{ par} \quad E(y_t) = -y_0 + \mu \quad \text{si } t \text{ impar}$$

$$V(y_t) = t\sigma_\varepsilon^2 \quad \text{Cov}(y_t, y_{t-s}) = (-1)^s (t-s)\sigma_\varepsilon^2$$

En este caso, el proceso no es estacionario en media ya que ésta toma dos valores distintos, pero tampoco presenta una tendencia estocástica como el proceso anterior.

Caso 6: $\rho = 1$ $y_t = \mu + vt + y_{t-1} + \varepsilon_t$

La forma final es: $y_t = t\mu + t^2v + \sum_{j=0}^{t-1} \varepsilon_{t-j} + y_0$

$$E(y_t) = y_0 + t\mu + t^2v$$

$$V(y_t) = t\sigma_\varepsilon^2 \quad \text{Cov}(y_t, y_{t-s}) = (t-s)\sigma_\varepsilon^2$$

Este proceso posee una tendencia estocástica cuadrática, no siendo entonces tampoco asintóticamente estacionario en media.

Caso 7: $\rho = -1$ $y_t = \mu + vt - y_{t-1} + \varepsilon_t$

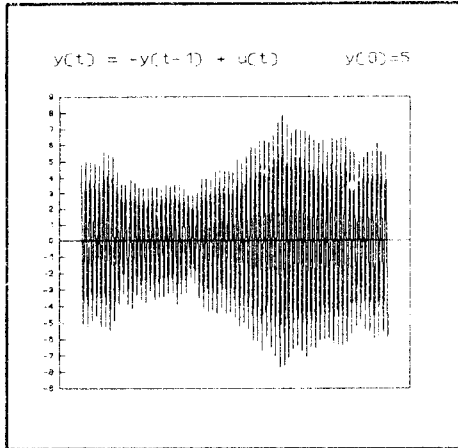
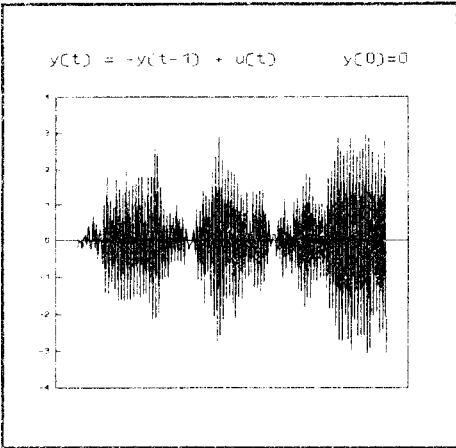
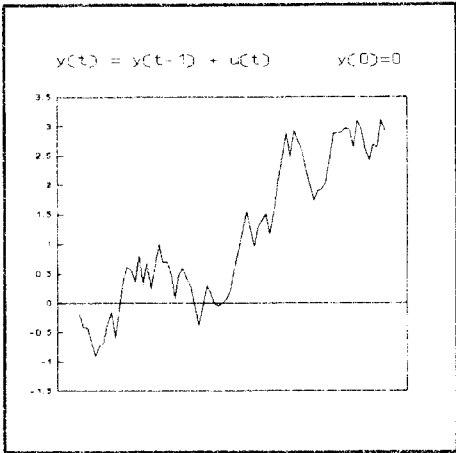
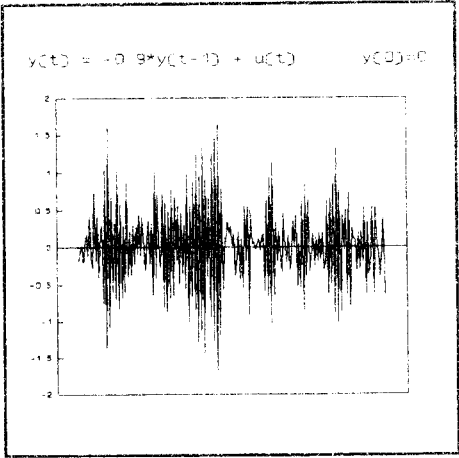
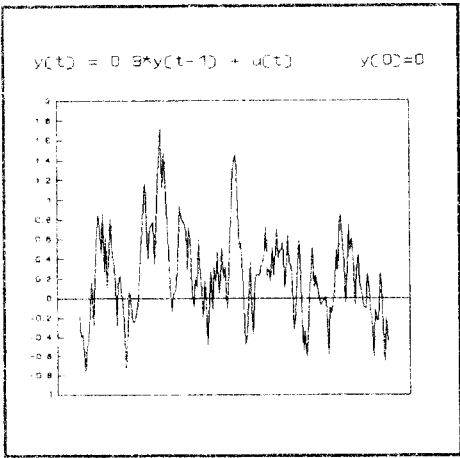
La forma final es: $y_t = \sum_{j=0}^{t-1} (-1)^j (\mu + vt) + \sum_{j=0}^{t-1} \varepsilon_{t-j} + (-1)^t y_0$

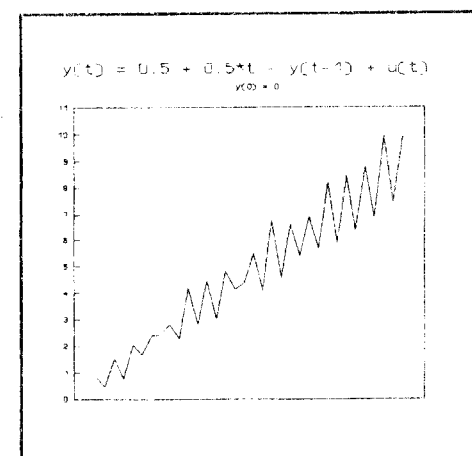
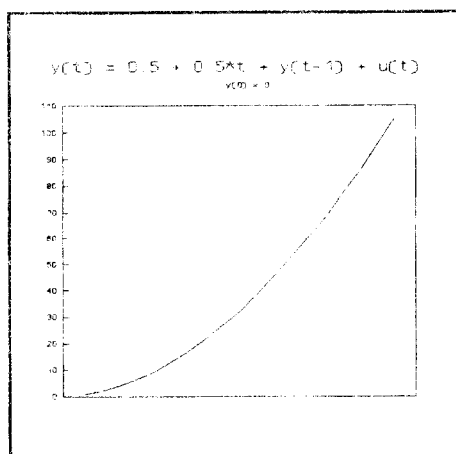
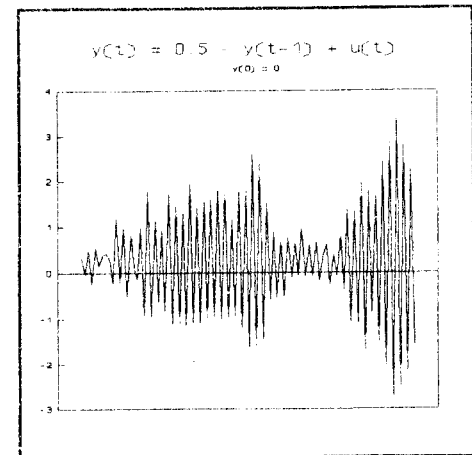
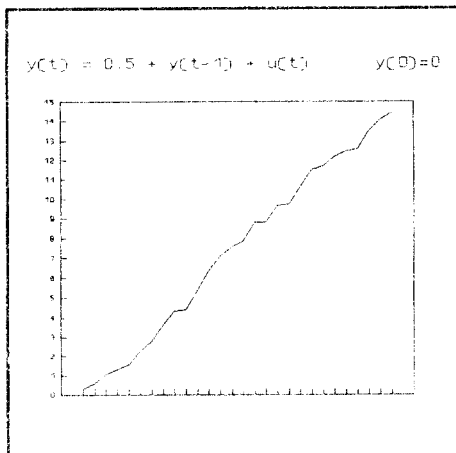
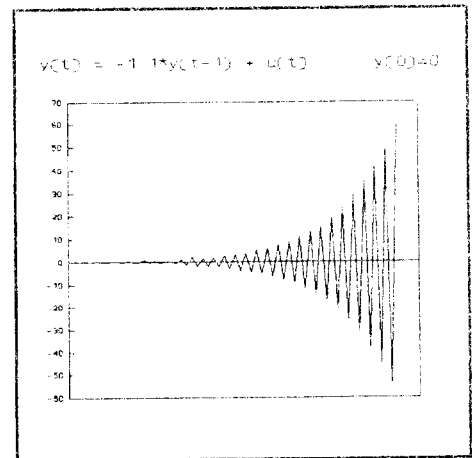
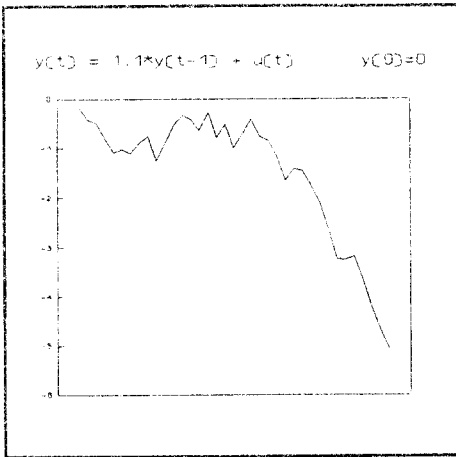
$E(y_t) = y_0$ si t par $E(y_t) = \mu + vt - y_0$ si t impar

$V(y_t) = t\sigma_\varepsilon^2$ $Cov(y_t, y_{t-s}) = (-1)^s (t-s)\sigma_\varepsilon^2$

Al igual que en el caso 5, este proceso no es estacionario en media ya que ésta no toma un valor único, pero su comportamiento es diferente ya que ésta varía de una constante a un valor que crece con t linealmente.

En las gráficas que se presentan a continuación se ilustran algunas de las diferencias analizadas en la sección. Las series $I(0)$ poseen una media constante que actúa como un "atractor", en el sentido que existe una tendencia inherente a la serie a retornar a este valor, siendo la varianza acotada. Las series integradas, por el contrario, no presentan ningún atractor y raramente retomarán un valor específico, incluyendo la media y el valor inicial, ya que si bien oscilan alrededor de un valor, la varianza con que lo hacen crece rápidamente. Finalmente, los valores que una serie estacionaria toma en un momento del tiempo tienen poca relación con sus valores futuros, de forma que a este tipo de procesos se los puede catalogar como de memoria corta. Por su parte, las series integradas son de "memoria larga", en cuanto lo que sucede en un momento tiene un efecto permanente e importante sobre todos los valores futuros.





Procesos casi integrados

La definición de procesos integrados revela que un proceso $I(k)$ es aquel que posee k raíces unitarias. Sin embargo, esta clasificación teórica no es tan clara en la práctica cuando el proceso presenta una raíz menor que 1 pero muy cercana a dicho valor. A estos procesos se les conoce habitualmente como procesos "casi integrados". Su análisis resulta interesante dado que su comportamiento es más parecido al de procesos integrados que al de procesos estacionarios, mostrando así los problemas de potencia (la probabilidad de rechazar la hipótesis nula cuando ésta es falsa) que presentan las pruebas de raíz unitaria en estos casos. Además, permiten observar que no existe una discontinuidad en el comportamiento de la varianza de procesos con raíces menores e iguales a 1, al menos en muestras no demasiado grandes.

Teoría asintótica para procesos integrados

Cuando las series son integradas, la teoría asintótica desarrollada para procesos estacionarios no es aplicable. La principal diferencia resulta de observar que en el primer caso los momentos muestrales convergen a variables aleatorias en vez de a constantes como en el segundo. De igual forma, los teoremas derivados del Teorema Central del Límite no son utilizables y deben ser sustituidos por teoremas del límite referidos a *funcionales*. Varios de los estadísticos utilizados convencionalmente convergerán a funciones de procesos de Wiener, cuya característica más sobresaliente es que la velocidad de convergencia es de un orden de magnitud superior al correspondiente a procesos estacionarios.

V. PRUEBAS DE RAIZ UNITARIA

Sea el proceso estocástico y_t tal que: $y_t = \rho y_{t-1} + \varepsilon_t$. Si $|\rho| < 1$ el proceso es estacionario (al menos asintóticamente), mientras que si $|\rho| > 1$ es explosivo. El caso $|\rho| = 1$ resulta singular: siendo un proceso no estacionario, su primera diferencia resulta integrada de orden 0 (estacionaria). Dada esta característica y el hecho que la existencia de raíces unitarias en los procesos económicos es bastante frecuente, probar la hipótesis $|\rho| = 1$ resulta del mayor interés.

Sin embargo en este caso no es posible utilizar los resultados estándar ya que la distribución del estadístico resultante no es una *t-Student* en pequeña muestra ni *Normal(0,1)* asintóticamente. Aún más, la distribución no es siquiera simétrica, originando así serios problemas de tamaño en la prueba (una alta probabilidad de rechazar la hipótesis cuando ésta es en realidad verdadera).

Una solución posible es utilizar el estadístico Q de Box-Pierce, para los residuos de la regresión de y_t sobre y_{t-1} , que bajo la hipótesis nula resultan ser simplemente $\varepsilon_t = \Delta y_t$. Sin embargo, la potencia de la prueba es muy baja (Dickey y Fuller, 1979).

La solución alternativa se originó en el trabajo de Dickey (1976), y se formalizó en Dickey y Fuller (1979) con el diseño de nuevas tablas de valores críticos para la realización de la prueba, construidas en base a experimentos de Monte Carlo. Los autores tabularon dichos valores para tres modelos distintos y para dos estadísticos distintos, uno equivalente a la *t-Student* y otro al estadístico aplicable en muestras grandes.

La presentación de tres modelos responde al hecho que al realizar las pruebas de raíz unitaria generalmente se ignora la forma específica que tiene el verdadero proceso generador de información (PGI), con lo cual se podría incurrir en los problemas conocidos de inferencia en modelos incorrectamente especificados (para un desarrollo del tema, véase Spanos, 1986). A pesar de ello, los casos presentados recogen únicamente la posible existencia de

tendencia determinística y/o una constante en el proceso, con lo cual resultan insuficientes. Varios autores han tratado de realizar correcciones al procedimiento, orientadas tanto a incorporar nuevas variables al modelo (paramétricas) como a corregir el estadístico a utilizar (no paramétricas) para mejorar el desempeño de la prueba. Una segunda extensión se refiere al análisis de la existencia de raíces unitarias en el componente estacional de una serie, es decir, al estudio de procesos estacionales no estacionarios. Finalmente, otra de las líneas principales de trabajo ha sido el diseño de pruebas que mejoren los resultados en los casos en que existe más de una raíz unitaria en el proceso estocástico.

El estadístico de Dickey-Fuller

Sea el verdadero PGI: $y_t = \rho y_{t-1} + \varepsilon_t$ (1)

con $y_0 = 0$ y $\varepsilon_t \sim iid(0, \sigma^2)$, donde y_0 es el valor inicial.

Los modelos propuestos por estos autores para representar el proceso estocástico son:

a) $y_t = \rho_a y_{t-1} + \varepsilon_t$, con $y_0 = 0$ y $\varepsilon_t \sim iid(0, \sigma^2)$

b) $y_t = \mu_b + \rho_b y_{t-1} + \varepsilon_t$, con $y_0 = 0$ y $\varepsilon_t \sim iid(0, \sigma^2)$

c) $y_t = \mu_c + \nu_c t + \rho_c y_{t-1} + \varepsilon_t$, con $y_0 = 0$ y $\varepsilon_t \sim iid(0, \sigma^2)$

El modelo elegido para aproximar el PGI real dependerá del análisis que se haga de la serie para determinar cuál de ellos es más adecuado. De cualquier forma, la idea en el diseño de los estadísticos de prueba es que éstos sean tales que la existencia de parámetros irrelevantes en el modelo afecte lo menos posible la inferencia. A estas pruebas se les denomina generalmente "similares", en cuanto las distribuciones del estimador de ρ y del estadístico de prueba, así como los valores críticos, no se ven afectadas por la existencia de dichos parámetros irrelevantes.

En el contexto anterior, se desea analizar la existencia de una raíz unitaria, es decir, probar la hipótesis $H_0: |\rho_i| = 1$ contra $H_1: |\rho_i| \neq 1$ para $i = a, b, c$.

En realidad y dado que los procesos explosivos son fácilmente detectables, aún en muestras pequeñas, la hipótesis alternativa relevante será: $H_1: |\rho_i| < 1$. Esto es, la prueba a desarrollar buscará evidencia que sustente que y_t es un proceso $I(1)$ frente a la opción que el proceso sea estacionario asintóticamente o $I(0)$.

Bajo H_0 el estadístico $(\hat{\rho}_i - \rho) / DS(\hat{\rho}_i)$ se distribuye de forma tal que los valores críticos que corresponden son los tabulados en la Tabla 1 del anexo (parte a, b, o c según el modelo) de acuerdo a los estudios de Dickey y Fuller (1979). En muestras grandes es posible utilizar el estadístico $T(\hat{\rho}_i - \rho)$, donde T es el número de observaciones, cuyos valores críticos se reportan en la Tabla 2 (nuevamente, partes a, b, o c) del anexo.

En Dickey y Fuller (1979) se derivan los estadísticos para muestras pequeñas y grandes, así como las distribuciones límite, resultando que, para el modelo (a):

$\hat{\rho} = \left(\sum_{t=1}^T Y_{t-1}^2 \right)^{-1} \sum_{t=1}^T Y_t Y_{t-1}$ es el estimador máximo verosímil de ρ

$\hat{t} = (\hat{\rho} - \rho) S_e^{-1} \left(\sum_{t=2}^T Y_{t-1}^2 \right)^{1/2}$ con $S_e^2 = (t-2)^{-1} \sum_{t=2}^T (Y_t - \hat{\rho} Y_{t-1})^2$

es el estadístico análogo a la t -Student para $H_0: |\rho_i| = 1$. El resultado se generaliza de la siguiente forma:

$\hat{t}_i = (\hat{\rho}_i - 1) (S_{ek}^2 c_i)^{-1/2}$ con $i = a, b, c$ y $k = 1, 2, 3$

de acuerdo al modelo que se proponga. S_{ek}^2 es el estimador de la varianza correspondiente a cada modelo, corregido por los grados de libertad adecuados para que sea insesgado $(T-k-1)$. c_i es el elemento correspondiente de la inversa del producto de la matriz de variables "explicativas" por sí misma $((X'X)^{-1}$ en la notación convencional) que será distinto según los casos a, b, o c y $\hat{\rho}$ es

el estimador máximo verosímil de ρ . Los valores críticos de estos estadísticos fueron obtenidos por los autores a través de simulaciones de Monte Carlo.

Por su parte, la distribución asintótica de normalizaciones adecuadas de $(\hat{\rho}_1 - \rho)$ son:

$$\tau_1^{1/2}(\hat{\rho}_1 - \rho) \sim \text{Normal}(0, \sigma_1^2) \quad \text{para } |\rho| < 1$$

$$|\rho|^{-1}(\rho^2 - 1)^{-1/2}(\hat{\rho}_1 - \rho) \sim \text{Cauchy} \quad \text{para } |\rho| > 1$$

El estadístico para probar $|\rho|=1 - T(\hat{\rho}_1 - 1)$ - es el cociente de dos integrales definidas sobre un proceso de Wiener. A partir de ello, se derivan su distribución límite así como la del estadístico τ_1 en los tres modelos. Este último sin embargo, tendrá una distribución asintótica *Normal* si efectivamente el PGI posee una constante y se plantea el modelo (b) con $\mu \neq 0$, o en el caso en que posee una tendencia determinística y se plantea el modelo (c) con $\beta \neq 0$.

El análisis del desempeño de los estadísticos propuestos muestra que: (i) resultan mucho más potentes que la Q de Box-Pierce; (ii) su validez se mantiene aún con errores con una distribución distinta de la *Normal*; y (iii) en muestras pequeñas los resultados pueden verse afectados si el valor inicial y_0 no es nulo.

Por otra parte, si el PGI es (1), la utilización de los modelo (b) y (c), y por lo tanto los correspondientes $\hat{\rho}_1$ y $\hat{\tau}_1$, originará pérdida de potencia en la prueba respecto al nivel que se obtendría usando $\hat{\rho}_a$ y $\hat{\tau}_a$, dado que estos últimos corresponden a la mejor aproximación posible al proceso, pero siempre serán más potentes que la Q .

Además, mientras los estadísticos $\hat{\rho}_a$ y $\hat{\tau}_a$ poseen similar potencia para cualquier valor de ρ , $\hat{\rho}_b$ es más potente que $\hat{\tau}_b$ si $|\rho| < 1$ y la relación se revierte en caso que $|\rho| > 1$.

Finalmente, los valores iniciales que tome el proceso deben

también ser tenidos en cuenta, como se desprende del análisis siguiente para distintos PGI:

(i) Si el PGI es el planteado en (1), las pruebas realizadas a partir de cualquiera de los tres modelos son similares.

(ii) Si el PGI es (1) pero $y_0 \neq 0$, o bien, posee una constante y el valor inicial es nulo, sólo aquellas realizadas sobre los modelos b y c serán similares.

(iii) Si el PGI posee constante y $y_0 \neq 0$, o bien posee constante y tendencia con un valor inicial nulo, las pruebas similares se obtienen únicamente usando el modelo c .

(iv) Si el PGI posee constante y tendencia y $y_0 \neq 0$, es necesario realizar extensiones para realizar una prueba adecuada.

De lo anterior es posible concluir, en primer lugar, que la elección del modelo no es una trivialidad, ya que afectará la potencia de la prueba, llevando a no rechazar la existencia de una raíz unitaria cuando en realidad el proceso es estacionario. Hay entonces que analizar qué tan plausible resulta tanto la existencia de una tendencia determinística en la serie como de una ordenada al origen distinta de 0. Al observar que la tendencia implica una tendencia cuadrática y la constante una tendencia lineal (de acuerdo a la forma final del proceso derivada en sección anterior), no parece demasiado difícil realizar este análisis. En segundo lugar, si se emplea el modelo b pero no se tiene demasiada evidencia sobre la existencia de una constante distinta de 0, la comparación de los resultados de ambos estadísticos puede ser de utilidad, dados los posibles problemas de potencia señalados más arriba. Finalmente, si el PGI posee constante y/o tendencia y esto se refleja en el modelo especificado, el estadístico de prueba para la hipótesis de existencia de una raíz unitaria se distribuye asintóticamente normal, es decir, las parametrizaciones adecuadas del proceso verdadero preservan los resultados estándares, asintóticamente.

Extensiones a la prueba de Dickey-Fuller

Propuesta de Bhargava

Este autor plantea que la forma que posea el verdadero PGI no es irrelevante para la correcta inferencia a partir de los estadísticos propuestos por Dickey y Fuller (DF) en los modelos (b) y (c), ya que el comportamiento de y_t que se plantea en H_0 y en H_1 no es simétrico.

En el caso (b): $y_t = \mu + \rho y_{t-1} + \varepsilon_t$ (2)

Bajo la hipótesis nula el proceso será un camino aleatorio con constante, es decir, la serie poseería una tendencia estocástica, mientras que bajo la alternativa será un proceso estacionario alrededor de una media no nula, igual a $\mu/(1-\rho)$, y sin tendencia. De esta forma, si el proceso real posee una tendencia determinística, resulta muy probable que no se rechace H_0 aún cuando ésta es falsa, ya que un proceso con tendencia estocástica será una mejor aproximación frente a un proceso sin tendencia. Obsérvese que el uso del modelo c no soluciona el problema, ya que si bien plantea la existencia de una tendencia bajo la hipótesis alternativa, induce a una tendencia cuadrática bajo la nula.

Para salvar este obstáculo, Bhargava (1986) propone la utilización del modelo:

$$y_t = \mu + u_t \quad \text{con } u_t = \rho u_{t-1} + \varepsilon_t \quad \text{y } \varepsilon_t \sim \text{iid} \quad (3)$$

de forma que, bajo $H_0: |\rho|=1$, y_t es un camino aleatorio sin constante: $y_t = \mu + u_t = \mu + u_{t-1} + \varepsilon_t = \mu + y_{t-1} - \mu + \varepsilon_t = y_{t-1} + \varepsilon_t$

Bajo $H_1: |\rho| < 1$, y_t es un proceso estacionario alrededor de una media no nula: $y_t = \mu + u_t = \mu + \rho u_{t-1} + \varepsilon_t = \mu + \rho y_{t-1} - \rho \mu + \varepsilon_t = \mu(1-\rho) + \rho y_{t-1} + \varepsilon_t$

En el caso del modelo (c), la solución es la misma, obteniéndose los siguientes procesos:

(i) bajo H_0 , y_t es un camino aleatorio con constante, es decir

con una tendencia estocástica, ya que:

$$y_t = \mu + \beta t + u_t = \mu + \beta t + u_{t-1} + \varepsilon_t = \mu + \beta t + y_{t-1} - \mu - \beta(t-1) + \varepsilon_t$$

$$\Rightarrow y_t = \beta + y_{t-1} + \varepsilon_t$$

(ii) bajo H_1 , y_t es un proceso estacionario con tendencia determinística:

$$y_t = \mu + \beta t + u_t = \mu + \beta t + \rho u_{t-1} + \varepsilon_t = \mu + \beta t + \rho y_{t-1} - \rho \mu - \rho \beta(t-1) + \varepsilon_t$$

$$\Rightarrow y_t = [\mu(1-\rho) + \beta\rho] + \beta(1-\rho)t + \rho y_{t-1} + \varepsilon_t$$

Para realizar la prueba de hipótesis se plantea el modelo de y_t en función de una constante (y/o tendencia) y utilizando la estimación por Mínimos Cuadrados Ordinarios (MCO) se realiza la prueba de autocorrelación de primer orden en los residuos con el estadístico de Durbin-Watson (DW). Un valor bajo del estadístico de DW $\approx 2(1-\rho)$ indica que la hipótesis $\rho = 1$ no será rechazada, con lo cual la evidencia señalará que y_t es un proceso no estacionario. Sargan y Bhargava (1983) presentan una tabla de valores críticos a utilizar que se reproduce como Tabla 3 en el anexo.

La prueba, denominada habitualmente CRDW, ha sido analizada por otros autores que la han encontrado en algún sentido preferible a la original de Dickey y Fuller: es la prueba invariante (en particular, a la existencia de tendencia) uniformemente más potente frente a una hipótesis alternativa de un proceso AR(1). Ambas, sin embargo, son escasamente potentes cuando el proceso posee raíces menores pero muy próximas a la unidad. Su principal limitación resulta, entonces, de la falta de generalidad de la hipótesis alternativa.

El estadístico de Dickey-Fuller aumentado (ADF)

Como ya se mencionó, si el modelo no es una buena aproximación del PGI real, las pruebas de hipótesis sobre ρ pueden conducir a decisiones equivocadas con facilidad. En particular, la incorrecta

especificación dinámica o la omisión de variables exógenas, se verán reflejadas en un error que no es *iid*.

En el primer caso, se propuso (Fuller, 1976) especificar y_t como un proceso autorregresivo de orden p , que puede ser representado por:

$$y_t = \sum_{i=1}^p \rho_i y_{t-i} + \varepsilon_t$$

y reformulado de acuerdo a:

$$y_t = \rho y_{t-1} + \sum_{i=1}^{p-1} v_i \Delta y_{t-i} + \varepsilon_t$$

De esta forma, los incrementos pasados de y_t estarían capturando la dinámica del proceso y permitiendo que el error sea independiente e idénticamente distribuido. La inclusión de rezagos en exceso no afecta la distribución asintótica del estadístico de prueba bajo la hipótesis nula, ya que la variable y_t es en este caso integrada de orden 1, con lo que sus primeras diferencias Δy_t son variables $I(0)$ para todo valor de t por definición y, por lo tanto, asintóticamente ortogonales a y_{t-1} .

La validez de la prueba es asintótica, con lo cual corresponde usar las mismas tablas de Dickey y Fuller en la línea " $T \rightarrow \infty$ ". Es importante, sin embargo, agregar un número suficiente de rezagos, (dado que p será desconocido), ya que si bien la sobreparametrización le quita eficiencia a la estimación, la no inclusión de rezagos significativos origina inconsistencia, invalidando el uso de los valores de las tablas. Además, siempre es posible realizar análisis de residuos para diferentes especificaciones, hasta obtener aquellas en que los errores se aproximen a un proceso ruido blanco.

Cabe destacar que este procedimiento también resulta útil, en general, para el segundo caso mencionado, cuando se han omitido variables exógenas relevantes que presentan autocorrelación, ya que los incrementos rezagados de la variable de interés capturarían dicha estructura autorregresiva, depurando así al error, que se aproximaría a un ruido blanco.

Formulación de Saïd y Dickey

En la búsqueda de formulaciones más generales para resolver la omisión de variables exógenas, estos autores (Saïd y Dickey, 1984) plantean trabajar con un modelo que presente una estructura de los errores de tipo *ARMA*, como una generalización del caso anterior. De esta forma, el PGI real será aproximado por un modelo del tipo:

$$Y_t = \mu + u_t \quad \text{con:} \quad u_t + \sum_{i=1}^p \alpha_i u_{t-i} = \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j} \quad \varepsilon_t \sim iid \quad (4)$$

Así, los errores siguen un proceso *ARMA(p,q)*, con parámetros desconocidos pero que se supone es estacionario e invertible. Con ello, es posible reformular (4) como:

$$\Delta Y_t = (\rho - 1)Y_{t-1} + \sum_{i=1}^k \alpha_i \Delta Y_{t-i} + v_t \quad (5)$$

donde el residuo será "casi" ruido blanco, gracias a la utilización de un número k suficientemente grande de rezagos. Además, dicho número deberá crecer junto con el número de observaciones, demostrándose que $T^{1/3}$ es una cota superior para la tasa a la que k debe crecer con el tamaño de la muestra. El modelo se estima por *MCO*, siendo el estimador de ρ consistente. La prueba que se realiza es del tipo *t-Student* pero usando las tablas de *DF* para el modelo (a). La extensión para el caso (b) se realiza utilizando el modelo:

$$\Delta Y_t = (\rho - 1)Y_{t-1}^* + \sum_{i=1}^k \alpha_i \Delta Y_{t-i} + v_t \quad (6)$$

con $Y_{t-1}^* = Y_{t-1} - \bar{Y}$ siendo $\bar{Y} = (1/T) \sum_{t=1}^T Y_t$

Para el modelo (c), Saïd (1991) realizó una extensión de la prueba original demostrando que el estadístico resulta proporcional al usado en el caso *AR(1)*, y cuyos valores críticos coinciden con los tabulados por Dickey y Fuller.

Las simulaciones de Monte Carlo realizadas para analizar la potencia de la prueba muestran que el estadístico del tipo *t-Student* tiene un desempeño superior.

El principal problema que presenta esta propuesta es que su

potencia es muy baja cuando el componente *MA* de los errores es negativo. Lo mismo ocurre cuando posee una raíz cercana a 1 ya que tiende a cancelarse con la raíz unitaria de la parte *AR*. Además, de ello, la forma en que se determina el número de rezagos a incluir no siempre es trivial, replanteando la discusión sobre la conveniencia de estrategias de sobre-especificación frente a aquellas de sub-especificación.

Formulación de Phillips y Perron

Estos autores buscan, al igual que los anteriores, modificar las pruebas originales de forma de incorporar aquellos casos en que la estructura del PGI real es más compleja (Phillips, 1987; Perron, 1988; Phillips y Perron, 1988). Su propuesta se orienta a la corrección no paramétrica de los estadísticos de *DF*, que corregiría la autocorrelación presente en los errores por omisión de variables relevantes en el modelo. Así, el efecto de esta estrategia es idéntico al del *ADF*, sólo que en lugar de estimar nuevos parámetros (los de los términos en Δy_{t-1}), se modifica el estadístico a utilizar. Su validez es sólo asintótica y las tablas a utilizar son las de *DF*. Los tres modelos análogos a los de Dickey y Fuller son:

$$a') y_t = \rho_a y_{t-1} + \varepsilon_t ; y_0 = 0 ; \varepsilon_t \text{ no ruido blanco}$$

$$b') y_t = \mu_b + \rho_b y_{t-1} + \varepsilon_t ; y_0 = 0 ; \varepsilon_t \text{ no ruido blanco}$$

$$c') y_t = \mu_c + \nu_c (t-T/2) + \rho_c y_{t-1} + \varepsilon_t ; y_0 = 0 ; \varepsilon_t \text{ no ruido blanco}$$

Los estadísticos propuestos son, respectivamente:

Caso a'

$$Z(\hat{\rho}_a) = T(\hat{\rho}_a - 1) - 1/2(S_{Th}^2 - S_{\varepsilon}^2)/T^{-2} \sum_{t=2}^T y_{t-1}^2$$

$$Z[t(\hat{\rho}_a)] = (S_{\varepsilon}/S_{Th})t(\hat{\rho}_a) - 1/2(S_{Th}^2 - S_{\varepsilon}^2)/\left(S_{Th} T^{-1} (\sum_{t=2}^T y_{t-1}^2)^{1/2}\right)$$

Caso b'

$$Z(\hat{\rho}_b) = T(\hat{\rho}_b - 1) - 1/2(S_{Th}^2 - S_{\epsilon}^2) / T^{-2} \sum_{t=2}^T (Y_{t-1} - \bar{Y})^2$$

$$Z[t(\hat{\rho}_b)] = (S_{\epsilon}/S_{Th})t(\hat{\rho}_b) - 1/2(S_{Th}^2 - S_{\epsilon}^2) / \left(S_{Th} T^{-1} \left(\sum_{t=2}^T (Y_{t-1} - \bar{Y})^2 \right)^{1/2} \right)$$

Caso c'

$$Z(\hat{\rho}_c) = T(\hat{\rho}_c - 1) - (T^6/24)(D_x^{-1})(S_{Th}^2 - S_{\epsilon}^2)$$

$$Z[t(\hat{\rho}_c)] = (S_{\epsilon}/S_{Th})t(\hat{\rho}_c) - (T^3/4\sqrt{3})(D_x^{-1/2}S_{Th}^{-1})(S_{Th}^2 - S_{\epsilon}^2)$$

Siendo: $S_{Th}^2 = 1/T \sum_{t=1}^T \epsilon_t^2 + 2/T \sum_{j=1}^h \sum_{t=j+1}^T \hat{\epsilon}_t \hat{\epsilon}_{t-j}$ un estimador consistente

de la varianza; y h el último rezago de la función de autocorrelación de los residuos. El número de rezagos h debe ser creciente con el de observaciones (T), de forma que la tasa de crecimiento sea de orden $T^{1/4}$. Dado que no es posible garantizar la no negatividad del estimador, se utiliza una corrección al mismo definida como:

$$\bar{S}_{Th}^2 = 1/T \sum_{t=1}^T \epsilon_t^2 + 2/T \sum_{j=1}^h w_h(j) \sum_{t=j+1}^T \hat{\epsilon}_t \hat{\epsilon}_{t-j}$$

El estimador de la varianza es: $S_{\epsilon}^2 = 1/T \sum_{t=1}^T \hat{\epsilon}_t^2$

Este estimador es consistente para cualquier modelo cuando $\mu=0$ en el PGI, mientras que cuando $\mu \neq 0$ sólo lo es en los casos (b) y (c). Finalmente, D_x es el determinante del producto interno de la matriz de datos. Obsérvese que estos estadísticos son idénticos a los de DF cuando $S_{\epsilon}^2 = S_{Th}^2$, es decir, cuando ϵ_t es un ruido blanco.

La comparación del desempeño de estos estadísticos en el caso de errores con estructuras tipo ARMA(p,q) con respecto al de la

prueba propuesta por Saïd y Dickey, indica que si bien tienen mayor potencia, adolecen de serios problemas de tamaño para coeficientes negativos en la parte MA. Por ello, los mismos autores sugieren el uso del estadístico de Saïd y Dickey en este último caso.

Formulación de Khan y Ogaki

Estos autores proponen una prueba (Khan y Ogaki, 1990) con mayor potencia que las de Dickey-Fuller y Phillips-Perron en muestras finitas para el caso en que existen raíces próximas a la unidad. La distribución del estadístico, a diferencia de la mayoría de las pruebas existentes, se basa en el teorema central del límite, en vez de en el teorema central del límite funcional. Esto permite condiciones más generales para la aplicabilidad de la prueba, como es el caso en que el proceso contiene componentes estacionales determinísticos.

$$\text{Sea: } y_t = \rho y_{t-1} + u_t \quad (7)$$

con $\{u_t\}$ una secuencia estacionaria con media cero y y_0 , el valor inicial, una variable aleatoria cualquiera.

$$\text{Considérese la regresión: } y_t = \alpha \Delta y_t + e_t \quad t = 1, 2, \dots, T \quad (8)$$

El estadístico que se propone para la prueba $\rho = 1$ viene dado por:

$$J_T = 2(S_{uT}^2/S_T^2)(\hat{\alpha} - 0.5) \stackrel{d}{\sim} \chi_{(1)}^2$$

$$\text{donde: } S_{uT}^2 = T^{-1} \sum_{t=1}^T (y_t - \hat{\rho} y_{t-1})^2$$

$$S_T^2 = T^{-1} \sum_{\tau=h(T)}^{h(T)} [h(T) - |\tau|/h(T)] \sum_{t=1}^T (y_t - \hat{\rho} y_{t-1})(y_{t-\tau} - \hat{\rho} y_{t-1-\tau})$$

$h(\tau)$ es el número de rezagos en que se trunca la muestra tal que su orden de magnitud es $O(\tau^{1/4})$.

$\hat{\rho}$ y $\hat{\alpha}$ son los estimadores mínimo cuadráticos de las regresiones (7) y (8), respectivamente.

Las propuestas de prueba reseñadas hasta aquí son aquellas que se

han vuelto más populares. Esto se relaciona a su efectividad, tanto en términos del tamaño y la potencia de la prueba como de su robustez frente a distintas formas del PGI.

Extensiones al caso de pruebas de más de un parámetro

En los modelos presentados como (b) y (c), además del coeficiente ρ se incluyen otros parámetros. Para realizar pruebas sobre éstos o bien pruebas conjuntas con ρ , los desarrollos anteriores son aplicables, con las modificaciones pertinentes en los estadísticos, aunque las tablas a utilizar son otras (Dickey y Fuller, 1981) y se presentan en el anexo como Tabla 4 y Tabla 5. La corrección no paramétrica de Phillips y Perron, sin embargo, muestra en estas extensiones problemas importantes de tamaño, aún en muestras grandes, con lo cual deben ser usadas con precaución. Por el contrario, las pruebas ADF parecen ser las más robustas ante la presencia de componentes MA negativos. Finalmente, Hall (1989) mostró que el uso de variables instrumentales en lugar de MCO mejora el desempeño del ADF, utilizando como instrumento para y_{t-1} la variable y_{t-1-k} , donde k es el orden del polinomio autorregresivo de los residuos. A continuación se reproduce el resumen presentado en Banerjee et al (1993) de los diferentes casos:

A. Estadísticos para hipótesis simples en modelos tipo (b') y (c'):

1. Modelo (b') $H_0: \mu_b = 0$

tipo t-Student: $\tau_1 = \hat{\mu}_b / DS(\hat{\mu}_b) = \hat{\mu}_b (c_{11(b)} S_\varepsilon^2)^{-1/2}$

no paramétrico:

$$Z(\tau_1) = (S_\varepsilon / S_{Th}) \tau_1 + \frac{1}{2} S_{Th}^{-1} (S_{Th}^2 - S_\varepsilon^2) T^{1/2} \sum_{t=2}^T y_{t-1} \left(\sum_{t=2}^T (y_{t-1} - \bar{y})^2 \sum_{t=2}^T y_{t-1}^2 \right)^{-1/2}$$

2. Modelo (c') $H_0: \mu_c = 0$

tipo t-Student: $\tau_2 = \hat{\mu}_c / DS(\hat{\mu}_c) = \hat{\mu}_c (C_{11(c)} S_{\epsilon}^2)^{-1/2}$

no paramétrico:

$$Z(\tau_2) = (S_{\epsilon} / S_{Th}) \tau_2 + 1/24 T^3 (S_{Th} D_x^{1/2} G_x)^{-1} (S_{Th}^2 - S_{\epsilon}^2) T^{-3/2} \sum_{t=2}^T Y_{t-1}$$

3. Modelo (c') $H_0: \nu_c = 0$

tipo t-Student: $\tau_3 = \hat{\nu}_c / DS(\hat{\nu}_c) = \hat{\nu}_c (C_{22(c)} S_{\epsilon}^2)^{-1/2}$

no paramétrico:

$$Z(\tau_3) = (S_{\epsilon} / S_{Th}) \tau_3 - \left\{ T^4 (2S_{Th} D_x^{1/2})^{-1} (S_{Th}^2 - S_{\epsilon}^2) \left(\sum_{t=2}^T (Y_{t-1} - \bar{Y})^2 \right)^{-1/2} \right\} \times \\ \times \left\{ 1/2 [T^{-3/2} T^{-5/2}] \sum_{t=2}^T Y_{t-1} - T^{-5/2} \sum_{t=2}^T Y_{t-1} \right\}$$

B. Estadísticos para hipótesis compuestas en modelos tipo b' y c':

1. Modelo (b') $H_0: \mu_b = 0, \rho_b = 1$

tipo F: $\phi_1 = (2S_{\epsilon}^2)^{-1} [(T-1)S_0^2 - (T-3)S_{\epsilon}^2]$

no paramétrico:

$$Z(\phi_1) = (S_{\epsilon}^2 / S_{Th}^2) \phi_1 - 1/2 S_{Th}^{-2} (S_{Th}^2 - S_{\epsilon}^2) \left\{ T(\rho_b - 1) - T^2/4 (S_{Th}^2 - S_{\epsilon}^2) \left(\sum_{t=2}^T (Y_{t-1} - \bar{Y})^2 \right)^{-1} \right\}$$

2. Modelo (c') $H_0: \mu_c = 0, \nu_c = 0, \rho_c = 1$

tipo F: $\phi_2 = (3S_{\epsilon}^2)^{-1} [(T-1)S_0^2 - (T-4)S_{\epsilon}^2]$

no paramétrico:

$$Z(\phi_2) = (S_{\epsilon}^2 / S_{Th}^2) \phi_2 - 1/3 S_{Th}^{-2} (S_{Th}^2 - S_{\epsilon}^2) [T(\rho_c - 1) - (T^6/48) D_x^{-1} (S_{Th}^2 - S_{\epsilon}^2)]$$

3. Modelo (c') $H_0: \nu_c = 0 \quad \rho_c = 1$

tipo F:
$$\phi_3 = (2S_c^2)^{-1} \{ (T-1)[S_0^2 - (\bar{Y}_0 - \bar{Y}_1)^2] - (T-4)S_c^2 \}$$

no paramétrico:

$$Z(\phi_3) = (S_c^2/S_{Th}^2)\phi_3 - (2S_{Th}^2)^{-1}(S_{Th}^2 - S_c^2)[T(\rho_c - 1) - (T^6/48)D_x^{-1}(S_{Th}^2 - S_c^2)]$$

Siendo:

$$S_0^2 = (T - 1)^{-1} \sum_{t=2}^T (Y_t - Y_{t-1})^2$$

$$\bar{Y}_0 = (T - 1)^{-1} \sum_{t=2}^T Y_t$$

$$\bar{Y}_1 = (T - 1)^{-1} \sum_{t=2}^T Y_{t-1}$$

$$G_x = \left(T^{-6} D_x + 1/12 \left(T^{-3/2} \sum_{t=2}^T Y_{t-1} \right)^2 \right)^{1/2}$$

Los valores críticos para los estadísticos $Z(\tau_i)$ son los mismos que los correspondientes a τ_i , para $i = 1, 2, 3$ respectivamente, y están tabulados en la Tabla 4. Lo mismo vale para $Z(\phi_i)$ y ϕ_i , aplicando la Tabla 5. Finalmente, nótese que tanto los estimadores de la varianza como los elementos c_{11} , D_x , G_x varían de acuerdo al modelo que se esté utilizando.

Extensión al caso de más de una raíz unitaria

Las pruebas expuestas hasta aquí están diseñadas para probar la presencia de una raíz unitaria frente a la alternativa de no existencia de raíces unitarias en absoluto. Con ello, su uso en el caso en que existe más de una raíz no es apropiado, ya que carecen de un nivel adecuado de potencia.

Dickey y Pantula (1987) sugieren utilizar un procedimiento secuencial que, comenzando con una hipótesis nula de un número

suficientemente alto de raíces unitarias, vaya reduciendo dicho número cada vez que la hipótesis nula es rechazada. Esto se repite hasta la primera vez en que no se rechaza H_0 .

La prueba es del tipo:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_d = 1 \quad \text{versus} \quad H_1: \rho_i \neq 1 \text{ para algún } i$$

En realidad se realiza una transformación del modelo para obtener una prueba análoga a la anterior y con mayor potencia. En el caso en que se comience con $d=2$, es decir dos raíces unitarias, se tiene que el modelo es:

$$(1 - \rho_1 L)(1 - \rho_2 L)y_t = \varepsilon_t \quad (9)$$

Si se despeja $\Delta^2 y_t$ en el lado derecho de la igualdad y se pasa el resto al otro lado, al que se le suma y resta $(\rho_1 \rho_2 - 1)$, se obtiene:

$$\Delta^2 y_t = (\rho_1 \rho_2 - 1) \Delta y_t + (\rho_1 - 1)(1 - \rho_2) y_{t-1} + \varepsilon_t$$

Definiendo: $\beta_1 = (\rho_1 \rho_2 - 1)$ y $\beta_2 = (\rho_1 - 1)(1 - \rho_2)$

se tiene la siguiente formulación, idéntica a la (9):

$$\Delta^2 y_t = \beta_1 \Delta y_{t-1} + \beta_2 y_{t-1} + \varepsilon_t \quad (10)$$

La prueba de existencia de dos raíces unitarias contra la alternativa de menos de dos raíces igual a 1 (en este caso, una raíz unitaria) sería, en principio:

$$H_0: \beta_1 = \beta_2 = 0 \quad H_1: \beta_1 < 0 \text{ y } \beta_2 = 0$$

Sin embargo, dado que la prueba es parcialmente a una cola, el estadístico F no es muy potente. Por ello, los autores sugieren replantear el modelo teniendo en cuenta que bajo las dos hipótesis $\beta_2 = 0$. Así, se estimaría el modelo:

$$\Delta^2 y_t = \beta_1 \Delta y_{t-1} + \varepsilon_t$$

y se realizaría la prueba:

$$H_0: \beta_1 = 0 \quad H_1: \beta_1 < 0$$

usando las tablas de *DF*. En el caso en que se rechace la hipótesis nula, se procedería a estimar el modelo (10) y a realizar la prueba:

$$H_0: \beta_1 < 0 \text{ y } \beta_2 = 0 \quad H_1: \beta_1 < 0 \text{ y } \beta_2 < 0$$

usando el estadístico tipo *t-Student* y las tablas de *DF*.

Extensión al caso de raíces unitarias en el componente estacional

La existencia de raíces unitarias en frecuencias distintas de cero implica que la serie presenta un patrón estacional que no es estacionario. Este es el caso, por ejemplo, en que se permite que el patrón estacional crezca en el tiempo, de forma que el efecto de las perturbaciones no desaparece y la varianza crece indefinidamente. Su tratamiento, si bien no es idéntico al caso anterior, resulta muy similar.

Una serie integrada estacionalmente de orden (d, D) se define como:

X_t es $SI(d, D)$ sii $\Delta^d S_D(L) X_t = \Delta^d (1 - L^D) X_t = \gamma(L) X_t$ es $I(0)$ con d y D los mínimos valores para los que se obtiene una serie estacionaria.

$\gamma(L)$ se puede particionar dando pie a la posibilidad de una raíz unitaria que no esté en el componente estacional

:

$$\gamma(L) = (1 - \delta_1 L)(1 + \delta_2 L) \cdots (1 + \delta_m L^{m-1})$$

Esto a su vez se puede re-escribir utilizando la aproximación de Taylor de manera que el proceso se pueda modelar, por ejemplo en el caso en que el polinomio $\gamma(L)$ fuese de grado 4 (utilizando

datos trimestrales y buscando estacionalidad por trimestre), a través de:

$$\gamma^*(L)(1-L^4)X_t = \pi_1 Z_{1,t-1} + \pi_2 Z_{2,t-1} + \pi_3 Z_{3,t-2} + \pi_4 Z_{3,t-1} + \varepsilon_t \quad (11)$$

donde:

$$\gamma(L) = -\pi_1 L(1+L+L^2+L^3) + \pi_2 L(1-L+L^2-L^3) + (\pi_3 L + \pi_4) L(1-L^2) + \gamma^*(L)(1-L^4)$$

$$Z_{1,t} = (1+L+L^2+L^3)X_t, \quad Z_{2,t} = -(1-L+L^2-L^3)X_t, \quad Z_{3,t} = -(1-L^2)X_t$$

$\gamma^*(L)$ es un polinomio en el operador rezago con coeficientes que pueden o no ser nulos. A efectos de estimar la ecuación (11) por MCO se supone en general que no lo son, con lo cual será necesario incluir rezagos de la variable dependiente $((1-L^4)X_t)$. Estos, además, capturarían la posible autocorrelación de los residuos por omisión de variables. Las pruebas de raíces unitarias serían las siguientes:

(i) para probar una raíz igual a 1 en la frecuencia 0: $H_0: \pi_1 = 0$

(ii) para probar una raíz unitaria en la frecuencia correspondiente al primer semestre: $H_0: \pi_2 = 0$

(iii) para probar una raíz unitaria en la frecuencia anual:

$$H_0: \pi_3 = \pi_4 = 0$$

Esta última prueba posee una versión en "dos etapas" equivalente que consiste en:

(iii') Probar: $H_0: \pi_4 = 0$ y si no se rechaza la hipótesis nula realizar la prueba: $H_0: \pi_3 = 0$

Sólo si se rechazan todas las hipótesis anteriores se puede pensar en X_t como un proceso estacionario. Los estadísticos a utilizar son tipo *t-Student* y la tabla correspondiente, relacionada a la de *DF*, fue confeccionada por Hylleberg et al (1990), reproduciéndose en el anexo como Tabla 6.

VII. COINTEGRACION

El concepto de cointegración, como ya fue señalado brevemente en las primeras secciones, es la noción estadística equivalente a la idea de equilibrio estable, en el sentido que cuando existe una relación de este tipo entre variables económicas, las desviaciones de la misma no pueden crecer ilimitadamente. De esta forma, si se verifica que un conjunto de variables integradas de primer orden están cointegradas, se asegura la existencia de una relación no espúrea entre ellas que, además, es estacionaria (es decir, de equilibrio en el sentido estadístico).

Cuando una serie es no estacionaria o integrada, los valores que toma en un momento del tiempo son, por definición, la acumulación de todas las perturbaciones o "shocks" pasados, a diferencia de las series estacionarias para las cuales el efecto de las perturbaciones es transitorio. De esta forma, el pasado de la variable determina su futuro y no hay razón para que exista un valor medio al que eventualmente retorne. Por lo tanto, el hecho que una combinación lineal de un conjunto de variables de este tipo sea estacionaria implica, intuitivamente, que la forma en que éstas se "mueven" en el tiempo es similar. Además, dicha relación de cointegración actuaría como un "atractor", en el sentido ya utilizado en este trabajo, para el conjunto de variables bajo análisis. Es interesante señalar que en este planteamiento no es necesario a priori establecer una dirección de causalidad entre las variables ni de exogeneidad de algún subconjunto de ellas. Este tipo de consideraciones será pertinente únicamente cuando se desee establecer un modelo específico.

Para ello, son fundamentales los resultados del *Teorema de Representación de Granger* (Engle y Granger, 1987) para series $I(1)$ y las extensiones posteriores realizadas por Johansen (1991), que permiten probar la equivalencia entre distintas representaciones (modelos) para una misma relación entre series cointegradas, de forma que en estas circunstancias resulte indiferente trabajar válidamente con una u otra, eligiendo el investigador aquella que le sea más útil. Además, de dicho teorema se derivan relaciones de

interés entre las matrices polinómicas involucradas en cada formulación alternativa.

En particular, la equivalencia entre la noción de cointegración y la existencia de un modelo de corrección de error para las variables resulta importante por varios motivos. En primer lugar, la cointegración brinda un sustento estadístico firme y con una interpretación económica clara para esta formulación, al relacionarla tanto con la idea de equilibrio estadístico como con los desajustes al mismo en el corto plazo. Además, la modelización conjunta de los efectos de corto y largo plazo permite resolver en algún sentido el debate sobre la utilización de variables en niveles y en diferencias. De esta forma, el enfoque de series temporales de Box-Jenkins y los modelos econométricos "estructurales" aparecen como casos particulares del MCE. Finalmente, el análisis del orden de integración de las variables y la existencia de cointegración entre ellas permite evitar el problema de las regresiones espúreas³.

Definiciones y propiedades

Definición 1

Sea \mathbf{x} un vector de n variables. Sean d_i con $i=1, \dots, n$ los órdenes de integración de cada variable x_i del vector. Se define el orden de integración de \mathbf{x} como $d = \max(d_i)$:

$\mathbf{x} = (x_1, x_2, \dots, x_n) \sim I(d)$ si $x_i \sim I(d_i)$, $d = \max(d_i)$

Definición 2

Los componentes del vector \mathbf{x} están cointegrados de orden (d, b) si \mathbf{x} es $I(d)$ y existe un vector no nulo β tal que $\beta' \mathbf{x}$ es $I(d-b)$, con

³ Para una comprensión detallada de esta sección es conveniente revisar la bibliografía relacionada a modelos ARMA y VARMA. Un texto recomendable es "Métodos cuantitativos para el análisis de la coyuntura económica", A.Espasa y J.R. Cancelo (eds.), Alianza Editorial, Madrid, 1993.

$d \geq b > 0$. El vector β se denomina *vector de cointegración*.

Propiedad 1

Si el vector x posee 2 componentes, β es único. Si, por el contrario, x posee más de 2 elementos, por ejemplo n , pueden existir hasta $n-1$ vectores de cointegración linealmente independientes. El rango de la matriz β que se puede formar con los distintos vectores de cointegración se denomina *rango de cointegración*. Esto implica que al considerar más de dos variables, es posible observar distintas relaciones de equilibrio entre ellas (Dolado, 1987).

Propiedad 2

Si x_t y y_t están cointegrados, entonces también lo estarán x_t y $(by_{t-k} + \omega_t)$, para $\omega_t \sim I(0)$ y cualquier k . Puede ocurrir, sin embargo, que el vector de cointegración no sea el mismo. Esta propiedad implica que aún cuando las variables se observen con errores (de medida, por ejemplo), es posible identificar que están cointegradas, siempre que dichos errores sean estacionarios (Granger, 1991).

Propiedad 3

Sea $x_{t+h/t}^P$ el predictor óptimo de x_{t+h} , variable $I(1)$, con información hasta el momento t . Sea I_t el conjunto de información disponible en el momento de realizar la predicción. Entonces, x_{t+h} y $x_{t+h/t}^P$ están cointegradas si se cumple alguna de las siguientes condiciones: I_t está bien definido en cuanto incluye todo el pasado de x_t o bien x_t está cointegrada con alguna de las variables incluidas en el conjunto de información. Esto implica que un predictor óptimo tendrá una trayectoria similar a la variable sólo cuando se esté considerando toda la historia de la misma o, en su defecto, variables que tengan una tendencia común con la variable de interés (Granger, 1991).

Propiedad 4

Si x_{t+h} y y_{t+h} están cointegradas con parámetro de cointegración β , sus predictores óptimos (de menor varianza entre los predictores lineales) $x_{t+h/t}^P$ y $y_{t+h/t}^P$, en base a un conjunto de información bien definido, cumplirán la relación de cointegración:

$x_{t+h/t}^p = \beta y_{t+h/t}^p$ cuando h tiende a ∞ . Esto es, los pronósticos óptimos de variables cointegradas mantendrán la relación de equilibrio (Granger, 1991).

Propiedad 5

Si x_t y y_t son variables $I(1)$ cointegradas, debe existir causalidad en el sentido de Granger al menos en alguna dirección, ya que por la propiedad 4 una variable ayuda a predecir a la otra (Granger, 1991).

Teorema de Representación de Granger⁴

Sea $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ tal que $x_i \sim I(1) \forall i$. El teorema de descomposición de Wold asegura que existe una representación del tipo: $(1-L)\mathbf{x}_t = C(L)\varepsilon_t$, con ε_t un proceso estacionario y $C(L)$ tal que $C(1) = \sum_{i=0}^{\infty} C_i < \infty$. Si además los elementos x_i del vector están cointegrados, con rango de cointegración r se cumple que:

1. $C(1)$ tiene rango $n-r$.
2. Existe una representación ARMA del tipo: $A(L)\mathbf{x}_t = d(L)\varepsilon_t$ tal que se cumple que $A(1)$ tiene rango r y $d(L)$ es un polinomio escalar con $d(1)$ finito. Si $d(L) = 1$, se obtiene un modelo VAR.
3. Existen las matrices β y γ de orden $n \times r$ y rango r tales que:
 $\beta' C(1) = 0$; $C(1)\gamma = 0$; $A(1) = \gamma\beta'$
4. Existe una representación de corrección de error del tipo:
 $A^*(L)(1-L)\mathbf{x}_t = -\gamma\mathbf{z}_{t-1} + d(L)\varepsilon_t$; $\mathbf{z}_{t-1} = \beta'\mathbf{x}_{t-1}$ de orden r .
5. El vector \mathbf{z}_t está dado por: $\mathbf{z}_t = K(L)\varepsilon_t$ de forma que:
 $(1-L)\mathbf{z}_t = -\beta'\gamma\mathbf{z}_{t-1} + J(L)\varepsilon_t$. $K(L)$ es una matriz polinómica en el operador rezago de orden $r \times n$ y rango r , dada por $\beta' C^*(L)$ y tal que todos los elementos de $K(1)$ son finitos. Además, el determinante de $\beta'\gamma$ es positivo.
6. Si es posible encontrar una representación VAR, las matrices $A(L)$ y $A^*(L)$ que intervienen en las formulaciones ARMA y MCE serán matrices en polinomios finitos.

⁴ Existen distintas formulaciones de este Teorema. La que se expone en el texto es la realizada en Engle y Granger (1987).

El teorema anterior brinda las condiciones bajo las cuales es posible encontrar representaciones isomórficas para un cierto proceso vectorial. De esta forma, cuando existe una representación de medias móviles invertible (MA), si las variables están cointegradas existirán formulaciones vectoriales autorregresivas (VAR) y de mecanismo de corrección de error (MCE), cumpliéndose además ciertas relaciones entre las matrices involucradas. De igual forma, es posible establecer condiciones bajo las cuales la existencia de una representación MA implica cointegración de las variables involucradas.

La demostración del teorema se omite, dada su extensión y complejidad y el hecho que se encuentra disponible en diversas fuentes. Vale la pena, sin embargo, reflexionar sobre algunas de las implicaciones de los distintos puntos.

La matriz $C(1)$ mide el efecto acumulado de una perturbación en el momento 0 sobre Δx y se denomina *matriz de impacto*. Se deriva el rango de la misma (punto 1) del hecho de que las variables x_t están cointegradas. A partir de este resultado se deriva el punto 3. Este refleja la dualidad existente entre las formulaciones VAR y MA, ya que de él se deriva que $A(1)C(1)=0$ y, por lo tanto, $A(1)$ pertenece al espacio nulo de $C(1)$. Además, la matriz de impacto en la formulación VAR, $A(1)$, es una combinación lineal de los vectores de cointegración β . Dicha combinación lineal se obtiene a partir del vector γ , que es una medida de la velocidad de ajuste del fenómeno a los desequilibrios de corto plazo en la formulación MCE.

Para ilustrar las relaciones expuestas más arriba, se desarrolla a continuación un ejemplo para el caso en que $d=b=1$ y $n=2$. Sean:

$$x_t + \alpha y_t = u_t \quad \text{con} \quad u_t = u_{t-1} + \varepsilon_{1t}, \quad E(\varepsilon_{1t}) = 0, \quad V(\varepsilon_{1t}) = \sigma_{11} \quad (1)$$

$$x_t + \beta y_t = e_t \quad \text{con} \quad e_t = \rho e_{t-1} + \varepsilon_{2t}, \quad |\rho| < 1, \quad E(\varepsilon_{2t}) = 0, \quad (2)$$

$$V(\varepsilon_{2t}) = \sigma_{22}, \quad \text{cov}(\varepsilon_{1t}, \varepsilon_{2t}) = \sigma_{12}$$

De esta forma, resulta que x_t y y_t son variables integradas de primer orden, por (1), y cointegradas con vector de cointegración

(1, β) de acuerdo a (2). El rango de cointegración r es igual a 1.

La representación *MA* del sistema formado por (1) y (2) es:

$$\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = \frac{1}{\beta - \alpha} \begin{pmatrix} \beta & -\alpha(1-L) \\ -1 & (1-L) \end{pmatrix} \cdot \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

de forma que:

$$C(L) = \frac{1}{\beta - \alpha} \begin{pmatrix} \beta & -\alpha(1-L) \\ -1 & (1-L) \end{pmatrix} \quad C(1) = \frac{1}{\beta - \alpha} \begin{pmatrix} \beta & 0 \\ -1 & 0 \end{pmatrix}$$

Por su parte la formulación *VAR* resultante es:

$$\begin{pmatrix} (1-L) & \alpha(1-L) \\ (1-\rho L) & \beta(1-\rho L) \end{pmatrix} \cdot \begin{pmatrix} x_t \\ y_t \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

de forma que:

$$A(L) = \begin{pmatrix} (1-L) & \alpha(1-L) \\ (1-\rho L) & \beta(1-\rho L) \end{pmatrix} = (1-L)C(L)^{-1}, \quad A(1) = \begin{pmatrix} 0 & 0 \\ 1-\rho & \beta(1-\rho) \end{pmatrix}$$

Finalmente, la representación por *MCE* es como sigue:

$$\begin{pmatrix} 1 & \alpha \\ 1 & \beta \end{pmatrix} \cdot \begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = - \begin{pmatrix} 0 \\ 1-\rho \end{pmatrix} \cdot (x_{t-1} + \beta y_{t-1}) + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$$\text{siendo: } \gamma = \begin{pmatrix} 0 \\ 1-\rho \end{pmatrix}, \quad A(1) = -\gamma\beta' = \begin{pmatrix} 0 \\ 1-\rho \end{pmatrix} \cdot (1, \beta), \quad z_t = (1, \beta) \cdot \begin{pmatrix} x_t \\ y_t \end{pmatrix}$$

En este punto conviene sintetizar los argumentos más importantes desarrollados en este texto, para así discutir la estimación e inferencia con variables cointegradas y señalar algunos puntos controvertidos alrededor del tema.

La teoría convencional para el tratamiento estadístico de series temporales se desarrolla bajo el supuesto de que éstas son estacionarias. Es posible también realizar el análisis para series estacionarias alrededor de una tendencia temporal determinística, ya que ésta es estimable y se puede depurar la serie de dicho componente. Por el contrario, cuando se trabaja con series no estacionarias, existen dos obstáculos principales: los estadísticos de prueba ya no poseen distribuciones estándar, con lo cual la inferencia queda invalidada; mientras que el riesgo de trabajar con regresiones espúreas o sin sentido resulta bastante alto. Una de las soluciones de mayor difusión consiste en la diferenciación de las series, tantas veces como sea necesario para transformarlas en procesos estacionarios. Este procedimiento ha sido criticado, sin embargo, al observar que implica pérdida de información de largo plazo que en muchos casos resulta de interés para el investigador. El camino alternativo consiste en la utilización de modelos dinámicos, siendo el modelo de corrección de error una de las formulaciones más populares. La definición del concepto de cointegración ha servido, de alguna manera, para dar cohesión a los diversos desarrollos aislados alrededor del tema: brindó una justificación teórica interesante para diferenciar los modelos estáticos con sentido de aquellos que se denominaron espúreos; llevó a un primer plano el análisis del orden de integración de las series involucradas, destacando la importancia conceptual de distinguir entre tendencias determinísticas y estocásticas y sus implicaciones para la inferencia estadística; finalmente, resultó una condición suficiente para sustentar la equivalencia entre distintas formulaciones dinámicas, como queda demostrado en el teorema de Representación de Granger.

Si un conjunto de variables no estacionarias está cointegrado, es posible entonces plantear un modelo estático (de las variables en niveles, sin rezagos) que tenga sentido. Sin embargo, la distribución de los estimadores no será estándar en muchos casos, impidiendo la realización de pruebas de hipótesis y la predicción. Los modelos dinámicos no están exentos de problemas, especialmente cuando existe más de un vector de cointegración o las variables no son débilmente exógenas. Por último, la precisión de la inferencia

dependerá críticamente de la forma del sistema de ecuaciones que se deriva de las relaciones de cointegración.

Con estas consideraciones, en el resto del capítulo se desarrollarán las pruebas de cointegración, la forma de estimación de una ecuación y de un sistema, así como algunos puntos que hay que tener especialmente en cuenta para realizar una inferencia adecuada.

Pruebas de cointegración

Como ya se vio en otras secciones, si dos variables son integradas de orden 1, en general una combinación lineal de ellas también será $I(1)$. Sólo cuando existe una relación de cointegración entre ellas es posible encontrar una combinación (y en el caso general r) que sea estacionaria. Así, la forma natural de probar si esta relación existe es analizar si en una regresión de una variable sobre la otra, los residuos poseen una raíz unitaria (en cuyo caso son integrados de orden 1 y las variables no estarían cointegradas) o, por el contrario, son estacionarios. Para realizar la prueba se estima, entonces: $y_t = \beta'x_t + u_t$; y se plantea: $u_t = \rho u_{t-1} + \varepsilon_t$, con $E(\varepsilon_t) = 0$ y $V(\varepsilon_t) = \sigma^2$. Se estima el modelo para los residuos utilizando \hat{u}_t y se realiza la prueba de hipótesis:

$$H_0: |\rho| = 1 \quad (\text{variables no cointegradas})$$

$$H_1: |\rho| < 1 \quad (\text{variables cointegradas})$$

Los estadísticos de prueba son los discutidos en la sección V (*DF*, *ADF*, *CRDW*, etc.), manteniéndose la discusión sobre sus ventajas y limitaciones. Las tablas construidas para realizar las pruebas de raíz unitaria, sin embargo, no serían exactamente precisas en este caso debido a que la variable que se utiliza es un estimador de las perturbaciones (\hat{u}_t) que depende de parámetros estimados. Esto origina cambios en los valores críticos, con lo cual es importante que el valor del estadístico obtenido esté bastante distante del valor de tablas. Esto motivó, de alguna manera, propuestas

alternativas.

Una de dichas propuestas consiste en calcular *superficies de respuesta* (MacKinnon, 1991) en función del número de observaciones. La ventaja de este procedimiento es que al definir el valor crítico como una función del tamaño muestral, con una única tabla para los parámetros que la definen se puede hacer inferencia para diversas muestras y distinto número de regresores. La función propuesta es:

$$C(p, T) = \phi_{\infty} + \phi_1 T^{-1} + \phi_2 T^{-2}$$

donde p es el percentil que define la región crítica (1%, 5%, 10%) y T es el número de observaciones. El valor de los parámetros está tabulado (Tabla 7 del anexo) en base a simulaciones de Monte Carlo para un número de regresores que va de 2 a 6. Si la muestra es suficientemente grande como para considerar que T tiende a ∞ , T^{-1} y T^{-2} tienden a cero y el valor crítico es ϕ_{∞} .

Es importante señalar que, teóricamente, el resultado de las pruebas es invariante a la elección del regresando (cuál de las variables se elige como la dependiente en la regresión estática), ya que la distribución del estadístico no se modifica. Sin embargo, en la práctica pueden existir diferencias en el valor que tome el estadístico, conduciendo inclusive a conclusiones distintas sobre la existencia de una relación de cointegración entre las variables, debido a que la validez de las pruebas es asintótica y normalmente se trabaja con muestras chicas. Cuando se trabaja con un vector de orden mayor a dos, el problema anterior se ve agravado por el hecho que al cambiar de regresando se está pasando de una ecuación del sistema a otra y, por lo tanto, de una relación de cointegración a otra.

Estimación de modelos con variables cointegradas

El método de Engle y Granger en dos etapas

En principio, si las variables están cointegradas, es posible estimar la regresión estática, obteniéndose estimadores

superconsistentes de los parámetros (es decir, cuya velocidad de convergencia es de orden T en vez de $T^{1/2}$), de acuerdo al resultado demostrado por Stock (1987). Sin embargo, dado que la distribución de los estimadores no es estándar, para realizar inferencia conviene especificar la forma dinámica asociada a la relación de cointegración, utilizando los resultados del teorema de Representación de Engle y Granger. En base a lo anterior, Engle y Granger (1987) proponen un método de estimación de los parámetros del modelo de corrección de error en dos etapas: en la primera se estima el vector de cointegración a partir de la regresión estática; en la segunda se estima el MCE asociado, sustituyendo el vector de cointegración en el mecanismo de corrección del error por su estimador, obtenido en la primera etapa. Así,:

Etapa 1: con la regresión $y_t = \beta'x_t + u_t$ se obtiene $\hat{\beta}$ por MCO

Etapa 2: se estima el modelo $A^*(L)(1-L)y_t = -\gamma(y_{t-1} - \hat{\beta}'x_{t-1}) + d(L)e_t$ por MCO y se obtienen todos los estimadores restantes.

El método resulta muy atractivo por su simplicidad, dado que únicamente requiere de la aplicación de MCO a dos ecuaciones. Las críticas, sin embargo, se sustentan básicamente en el hecho que aún cuando $\hat{\beta}$ es superconsistente, en pequeñas muestras el sesgo puede ser muy grande debido a la omisión de la estructura dinámica. Además, la estimación de regresiones estáticas, incluso cuando no son espúreas, presenta problemas por varios motivos: se desconoce la distribución de los estimadores; el residuo no es una martingala en diferencias; los regresores pueden no ser débilmente exógenos y además pueden ser procesos autocorrelacionados; y, finalmente, es posible que exista más de un vector de cointegración. Por su parte, la estimación de la forma dinámica asociada, al plantearse en términos de variables estacionarias, sólo presentará dificultades en el caso en que existan múltiples vectores de cointegración o bien si las variables explicativas no cumplen con la propiedad de exogeneidad débil, con lo cual será

necesario estimar un sistema⁵. Así, la estimación dinámica aparece como más robusta que la estimación estática.

A pesar de lo anterior, West (1988) demuestra que existen algunas excepciones en las que los estimadores de los parámetros de variables integradas de orden 1 en una regresión poseen una distribución asintóticamente normal. Dicha excepción tiene lugar cuando en la regresión está presente, además de la tendencia estocástica, un componente no estocástico que la domina. Sea el modelo:

$$y_t = \mathbf{x}'_t \alpha + \mu w_t + \varepsilon_t$$

donde \mathbf{x}_t es un vector de variables estacionarias y w_t es una variable $I(1)$. Entonces, y_t será estacionaria o $I(1)$ dependiendo de si μ es cero o distinta de 0. West demuestra que los estimadores de α y μ son asintóticamente normales cuando se da que $E(\Delta w_t) \neq 0$. Así, bajo esta condición, el estadístico de prueba *t-Student* para $\mu = 0$ tiene una distribución asintótica normal. Un caso particular es cuando $w_t = y_{t-1}$. Si la esperanza de su incremento fuese nula, se estaría en el caso planteado por Dickey y Fuller y la distribución del estadístico sería no estándar, correspondiendo usar las tablas que los autores diseñaron. Obsérvese además que en el caso particular en que $\mathbf{x}'_t \alpha$ es una constante, se obtiene el mismo resultado que se describió para los caminos aleatorios con constante (normalidad asintótica del estadístico).

Dado lo expuesto más arriba, la alternativa más inmediata a la estimación de regresiones estáticas es el planteo directo del MCE, con un vector de unos sustituyendo al vector de cointegración y agregando rezagos suficientes de los regresores para romper la homogeneidad impuesta (como se vio en la sección III). Nótese que

⁵ La definición de exogeneidad débil, fuerte y superexogeneidad se encuentra en Engle et al., 1983. Para una discusión en relación al tema de cointegración, véase Ericsson, 1992.

a partir de la estimación del modelo, se debería probar la existencia misma de cointegración de las variables con una prueba sobre el coeficiente del mecanismo de corrección de error: si éste es cero, no existe una relación de cointegración entre las variables. El estadístico de prueba, si bien no es asintóticamente normal, resulta en una prueba similar cuando los regresores son fuertemente exógenos (Kiviet y Phillips, 1992). En este caso, además, la prueba es más potente que la que resulta de utilizar el estadístico de *ADF* a la ecuación que utiliza los residuos de la regresión estática (Banerjee et al., 1993).

El método de mínimos cuadrados en tres etapas

El método de dos etapas propuesto por Engle y Granger presenta, como ya se mencionó, problemas de sesgo y de distribución no estándar del vector de cointegración cuando las variables no son débilmente exógenas. Para salvar este obstáculo, Engle y Yoo (1991) proponen una modificación que incorpora una tercera etapa y que resulta en la obtención de estimadores equivalentes a los de Máxima Verosimilitud con Información Completa (*MVIC*). En la primera etapa se estima por *MCO* el vector de cointegración β a partir de la regresión estática. El segundo paso consiste en la estimación, nuevamente por *MCO*, del *MCE* utilizando $\tilde{\beta}$ para definir el desequilibrio. Los residuos estimados a partir de dicha regresión se utilizan para corregir el estimador del vector de cointegración de la primera etapa. En el caso en que las variables no sean débilmente exógenas, únicamente habría que considerarlo en la tercera etapa utilizando los residuos de todas las ecuaciones del sistema resultante. Sintetizando:

Sea x_t un vector de N variables débilmente exógenas con respecto a la estimación de los parámetros de la regresión de y sobre x y sea el rango de cointegración $r=1$.

Etapa 1: Estimar por *MCO* $y_t = \beta'x_t + u_t$ para obtener $\tilde{\beta}$.

Etapa 2: Estimar por *MCO* $A^*(L)(1-L)y_t = -\gamma(y_{t-1} - \tilde{\beta}'x_{t-1}) + d(L)\varepsilon_t$

para obtener todos los estimadores restantes: $\tilde{\gamma}$, $\tilde{\varepsilon}_t$.

Etapa 3: Estimar por MCO $\tilde{\varepsilon}_t = b' \mathbf{x}_{t-1}^* + \omega_t$ siendo $\mathbf{x}_{t-1}^* = \tilde{\gamma} \mathbf{x}_{t-1}$. Se obtiene \tilde{b} y se realiza la corrección de $\tilde{\beta}$ de acuerdo a: $\hat{\beta} = \tilde{\beta} + \tilde{b}$.

Obsérvese que en la etapa 2 se está estimando utilizando el término $\gamma \tilde{\beta}' \mathbf{x}_{t-1}$ en lugar de $\gamma \hat{\beta}' \mathbf{x}_{t-1}$, con lo cual se está omitiendo $\gamma(\hat{\beta} - \tilde{\beta})' \mathbf{x}_{t-1} = \gamma \tilde{b}' \mathbf{x}_{t-1}$, y esto queda en el residuo $\tilde{\varepsilon}_t$. Por ello, la regresión de $\tilde{\varepsilon}_t$ sobre $\gamma \mathbf{x}_{t-1}$ brinda un estimador del sesgo de $\tilde{\beta}$.

El estimador $\hat{\beta}$ así obtenido posee una distribución asintótica normal, al igual que el de *MVIC*. Si bien es posible realizar iteraciones con las etapas 2 y 3, se observa que la ganancia en precisión es escasa.

Si las N variables x_{it} no son débilmente exógenas y suponiendo que existe un único vector de cointegración para simplificar, habrá que modificar levemente el procedimiento, debiéndose estimar en la segunda etapa un sistema de $(N+1)$ ecuaciones, una para cada variable involucrada, y obtener $\tilde{\gamma}_i$ y $\tilde{\varepsilon}_{it}$, para $i=0,1,\dots,N$, donde el subíndice 0 corresponde a la ecuación de y sobre \mathbf{x} . El sistema es del tipo:

$$\begin{aligned} A_0^*(L) \Delta Y_t &= -\gamma_0 (Y_{t-1} - \tilde{\beta}' \mathbf{x}_{t-1}) + \varepsilon_{0t} \\ A_1^*(L) \Delta X_{1t} &= -\gamma_1 (Y_{t-1} - \tilde{\beta}' \mathbf{x}_{t-1}) + \varepsilon_{1t} \\ &\vdots \\ &\vdots \\ &\vdots \\ A_N^*(L) \Delta X_{Nt} &= -\gamma_N (Y_{t-1} - \tilde{\beta}' \mathbf{x}_{t-1}) + \varepsilon_{Nt} \end{aligned}$$

Con ello se forma un vector de $(N+1)T \times 1$, $\tilde{\mathbf{E}}$, de residuos ponderados y una matriz $\tilde{\mathbf{X}}^*$ de orden $(N+1)T \times N$ de regresores:

$$\tilde{\mathbf{E}} = (\tilde{\varepsilon}_0/\tilde{\sigma}_0, \tilde{\varepsilon}_1/\tilde{\sigma}_1, \tilde{\varepsilon}_2/\tilde{\sigma}_2, \dots, \tilde{\varepsilon}_N/\tilde{\sigma}_N)'$$

$$\tilde{\mathbf{X}}_{t-1}^* = (\tilde{\gamma}_1 \mathbf{x}_{t-1}/\tilde{\sigma}_1, \dots, \tilde{\gamma}_N \mathbf{x}_{t-1}/\tilde{\sigma}_N)'$$

Así, en la tercera etapa, se estima la regresión: $\tilde{\mathbf{E}}_t = \tilde{\mathbf{b}} \tilde{\mathbf{X}}_{t-1}^* + \omega_t$ y

se obtiene el corrector del sesgo de $\tilde{\beta}$ como:

$$\tilde{b} = (\mathbf{X}^*{}' \mathbf{X}^*)^{-1} \mathbf{X}^*{}' \tilde{\mathbf{E}}$$

El método de mínimos cuadrados modificados

Una segunda opción es la sugerida por Phillips y Hansen (1990), quienes proponen realizar una corrección no paramétrica a los estimadores que elimine los sesgos de segundo orden de la regresión estática (originados en la no exogeneidad débil y la autocorrelación de las variables). Los autores demuestran que los estimadores resultantes son asintóticamente normales y equivalentes a los de *MVIC*. Además, si los regresores son fuertemente exógenos, son equivalentes a los mínimo-cuadráticos de la regresión dinámica.

Sea el modelo:

$$\begin{aligned} y_t &= \beta' \mathbf{x}_t + u_t \\ \mathbf{x}_t &= \mathbf{x}_{t-1} + \varepsilon_t \quad \text{con } \text{cov}(u_t, \varepsilon_t) \neq 0 \end{aligned}$$

Entonces, el estimador anterior es:

$$\hat{\beta} = (\mathbf{x}' \mathbf{x})^{-1} [\mathbf{x}' \mathbf{y}^* - T \hat{\delta}^* \iota_k]$$

con: $\mathbf{y}^* = \mathbf{y} - \hat{\omega}_{12} \hat{\omega}_{22}^{-1} \Delta \mathbf{x}$

$\hat{\omega}_{12}$ y $\hat{\omega}_{21}$ son estimadores consistentes de los correspondientes elementos de la matriz de varianzas y covarianzas de (u_t, ε_t)

$\hat{\delta}^* = \hat{\Delta} (1, -\hat{\omega}_{22}^{-1} \hat{\omega}_{21})'$ es el corrector del sesgo

$$\hat{\Delta} = \sum_{j=0}^{\infty} E[\varepsilon_0(\varepsilon_j, u_j)]$$

$$\iota_k = (1, 1, \dots, 1)' \quad \text{vector de orden } k$$

Para analizar si las variables son débilmente exógenas conviene plantear el sistema completo para observar las distintas razones por las que puede no darse esta situación. Sean dos variables integradas de orden 1 y cointegradas, y_t y x_t . El sistema resultante es:

$$y_t = \beta x_t + u_t \quad (3)$$

$$x_t = x_{t-1} + e_t \quad (4)$$

En el caso más general, las perturbaciones de cada ecuación estarían autocorrelacionadas e interrelacionadas, es decir:

$$\begin{pmatrix} u_t \\ e_t \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \begin{pmatrix} u_{t-1} \\ e_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

lo cual implica que en general las ecuaciones que describen a las perturbaciones serán:

$$u_t = c_{11}u_{t-1} + c_{12}e_{t-1} + \varepsilon_{1t} \quad (5)$$

$$e_t = c_{21}u_{t-1} + c_{22}e_{t-1} + \varepsilon_{2t} \quad (6)$$

Para obtener la forma reducida del sistema, se sustituye u_t por su expresión (5) en la ecuación (3). En ésta se expresan, además, u_{t-1} como $(y_{t-1} - \beta x_{t-1})$ y e_{t-1} como Δx_{t-1} utilizando la ecuación (4), lo cual da origen a:

$$y_t = \beta x_t + c_{11}(y_{t-1} - \beta x_{t-1}) + c_{12}\Delta x_{t-1} + \varepsilon_{1t}$$

Si se resta de ambos lados y_{t-1} y se suma y resta del lado derecho el término βx_{t-1} , se obtiene la expresión asociada al MCE, donde todas las variables son $I(0)$ cuando y y x están cointegradas:

$$\Delta y_t = \beta \Delta x_t + (c_{11} - 1)(y_{t-1} - \beta x_{t-1}) + c_{12}\Delta x_{t-1} + \varepsilon_{1t} \quad (7)$$

La estimación del sistema (3)-(4) por variables instrumentales presentará ventajas sobre la estimación por MCO de la ecuación (7), en general, cuando exista simultaneidad o bien cuando x no sea débilmente exógena para la estimación de los parámetros de interés. Así, es importante definir quiénes son estos parámetros, en primer lugar, y cuáles son las condiciones que aseguran no simultaneidad. Además, para garantizar un comportamiento adecuado de los estimadores en términos de lo discutido más arriba, es importante observar en qué circunstancias se da que x es fuertemente exógena. Para ello, obsérvese que, si se supone que $(\varepsilon_{1t}, \varepsilon_{2t})$ posee una distribución normal multivariada, se tendrá:

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} \right) \Rightarrow E(\varepsilon_{1t}/\varepsilon_{2t}) = \sigma_{12}\sigma_{22}^{-1}\varepsilon_{2t} = \mu\varepsilon_{2t}$$

Así, si $\mu = 0$, o sea, cuando $\sigma_{12} = \text{cov}(\varepsilon_{1t}, \varepsilon_{2t}) = 0$, no hay simultaneidad en el sistema. Esto, sin embargo, no garantiza exogeneidad débil. Para notar esto último, se sustituye en (7) la expresión $(\mu\varepsilon_{2t} + v_t)$ por ε_{1t} , con v_t una perturbación estacionaria. Trabajando de manera conveniente con las relaciones planteadas se llega a la siguiente expresión equivalente a (7):

$$\Delta y_t = \beta^* \Delta x_t + \gamma(y_{t-1} - \beta x_{t-1}) + \delta \Delta x_{t-1} + v_t \quad (8)$$

donde:

$$\beta^* = \beta + \mu$$

$$\gamma = c_{11} - 1 + \mu c_{21}$$

$$\delta = c_{12} - \mu c_{22}$$

$$\sigma^2 = \text{Var}(\Delta y_t) = \text{Var}(v_t) = \text{Var}(\varepsilon_{1t} - \mu\varepsilon_{2t}) = c_{11} + \beta c_{22} + 2\beta c_{21}$$

Los parámetros de interés resultantes son: $\theta = (\beta, \beta^*, \gamma, \delta, \sigma^2)$, donde β es el vector de cointegración.

De esta forma, la estimación de los parámetros de interés no dependerá de la distribución marginal de x_t cuando se cumpla que $c_{21} = 0$, (porque la varianza de y_t dependerá únicamente del valor de c_{11} y c_{22}). Obsérvese que si la covarianza entre las perturbaciones ε_{1t} y ε_{2t} es menor o igual que la expresión $(-1/2\beta)(c_{11} + \beta^2 c_{22})$, la varianza de y_t es no positiva, resultando imposible sostener que los parámetros de la distribución marginal no restringen la estimación de los parámetros de interés. En este caso, la estimación de (7) por MCO dará por resultado estimadores sesgados e ineficientes. Aún más, la distribución asintótica de $\hat{\theta}$ no será estándar.

Por último, podría suceder que aún no siendo nula la covarianza entre las perturbaciones, el pasado de y_t afecte el presente de x_t , originando causalidad en el sentido de Granger, con lo cual x será débilmente exógena pero no se sostendrá la exogeneidad fuerte. En este caso no se verifican problemas de sesgo importante, aunque no será posible predecir correctamente a partir de dichos estimadores.

Resumiendo, si $\mu = 0$ y $c_{21} = 0$, no hay simultaneidad y las x_t son débilmente exógenas, con lo cual la estimación mínimo-cuadrática de la forma reducida (7) es eficiente y los estimadores tendrán distribuciones asintóticas estándar. Si $\mu \neq 0$ pero $c_{21} = 0$, la estimación de (7) no presenta problemas de sesgo ni de distribución de los estimadores. Sin embargo, se gana en eficiencia con la estimación de un sistema por variables instrumentales o SUR según corresponda. Si $c_{21} \neq 0$, será necesario utilizar los estimadores mínimo-cuadráticos modificados, en tres etapas incorporando las modificaciones pertinentes o de MVIC. Finalmente, si no es posible garantizar la no causalidad en el sentido de Granger, a pesar de que c_{21} sea 0, la estimación de (7) no permitirá predecir eficientemente pero no existirán problemas de sesgo ni de distribución.

El método de máxima verosimilitud con información completa

Una forma alternativa de plantear el sistema (3)-(4) que resulta conveniente para lo que sigue, es la que se deriva al realizar sustituciones en la ecuación (4) usando las relaciones (5) y (6):

$$\Delta y_t = \beta^* \Delta x_t + \gamma (y_{t-1} - \beta x_{t-1}) + \delta \Delta x_{t-1} + v_t \quad (9)$$

$$\Delta x_t = c_{21} (y_{t-1} - \beta x_{t-1}) + c_{22} \Delta x_{t-1} + \varepsilon_{2t} \quad (10)$$

En esta formulación queda aún más claro que la condición $c_{21} = 0$ garantiza exogeneidad débil y, en general, también fuerte. Además, se observa que el origen de la conveniencia de la estimación del sistema se puede encontrar en el hecho que el vector de cointegración β aparece en ambas ecuaciones, mientras que c_{21} , a su vez, participa de las dos relaciones al ser parte de la definición de γ .

Dado el sistema (9)-(10), Johansen (1988) propone su estimación por MVIC. La prueba sobre exogeneidad débil consiste en el análisis de la hipótesis $c_{21} = 0$ en relación a $c_{21} \neq 0$, que en este planteo es una prueba común sobre significancia del parámetro en una de las variables de la ecuación (10).

La estimación por MVIC mencionada se desarrolla a partir de la maximización de la función de verosimilitud del sistema VAR asociado al (9)-(10):

$$\Delta z_t = \sum_{i=1}^{k-1} \delta_i \Delta x_{t-i} + \pi z_{t-k} + e_t \quad (11)$$

donde: $z_t = (y_t, x_t)'$; $e_t \sim N(0, \Omega)$ y $\pi = -\gamma\beta'$

Primeramente, se deriva la función de verosimilitud y se obtiene el estimador habitual de Ω : $\hat{\Omega} = 1/T \sum_{t=1}^T \hat{e}_t \hat{e}_t'$.

Para obtener el estimador de π , se procede de la siguiente forma:

1. Se concentra la verosimilitud con respecto a δ_i , de manera de eliminar las variables $I(0)$, es decir, se hace la regresión de Δz_t y z_{t-k} sobre los k rezagos de Δz_t (como si fuese una estimación por variables instrumentales con las variables Δz_{t-1} como instrumentos). Los residuos que se obtienen se denominan:

$$R_{0t} = \Delta z_t - \sum_{i=1}^{k-1} \hat{\delta}_i \Delta z_{t-i} \quad \text{y} \quad R_{kt} = z_{t-k} - \sum_{i=1}^{k-1} \tilde{\delta}_i \Delta z_{t-i}$$

2. Se define $L(\pi) = h - T/2 \ln |\sum_{t=1}^T (R_{0t} - \pi R_{kt})(R_{0t} - \pi R_{kt})'|$ (12)

y se obtienen las matrices de segundos momentos de acuerdo a:

$$S_{ij} = 1/T \sum_{t=1}^T R_{it} R_{jt}' \quad \text{con } i, j = 0, k$$

Sustituyendo en (12) e imponiendo las restricciones $\pi = \gamma\beta'$, se obtiene:

$$L(\gamma, \beta) = m - T/2 \ln |(S_{00} - \gamma\beta'S_{k0} - S_{0k}\beta\gamma' + \gamma\beta'S_{kk}\beta\gamma'| \quad (13)$$

La derivada de (13) con respecto a γ permite obtener la expresión de su estimador como: $\tilde{\gamma} = S_{0k}\beta(\beta'S_{kk}\beta)^{-1}$

3. Sustituyendo γ por $\tilde{\gamma}$ en (13) se obtiene $L(\tilde{\gamma}, \beta)$, expresión que se debe maximizar con respecto a β para obtener su estimador. Ello es equivalente a minimizar el cociente de dos determinantes:

$$|\beta'(S_{kk} - S_{k0}S_{00}^{-1}S_{0k})\beta|/|\beta'S_{kk}\beta| \quad (14)$$

La expresión se puede normalizar de modo que: $|\beta'S_{kk}\beta| = I$ (15)

Así, el problema es equivalente a minimizar el numerador de (14) sujeto a la restricción (15). Esto transforma la obtención de los estimadores en un problema de cálculo de valores característicos que conduce a resolver un sistema de ecuaciones homogéneo del tipo:

$$(\lambda S_{kk} - S_{k0}S_{00}^{-1}S_{0k})\beta' = 0 \quad (16)$$

Para obtener los valores característicos y los vectores asociados, se resuelve el conjunto de ecuaciones que resulta de igualar a cero el determinante de la matriz que multiplica a β' (para eliminar la solución trivial): $|\lambda S_{kk} - S_{k0}S_{00}^{-1}S_{0k}| = 0$. Sea n su orden en el caso general (si el sistema posee dos variables como en el caso tratado $n = 2$). Entonces se obtendrán n valores característicos λ_i que pueden ser ordenados en forma decreciente. Los vectores característicos asociados a los r (en este caso 1) mayores serán los vectores de cointegración: $\beta = (v_1, \dots, v_r)$.

Obsérvese que la restricción (15) determina que β diagonaliza simultáneamente S_{kk} (a la matriz identidad) y $S_{k0}S_{00}^{-1}S_{0k}$ a la matriz Λ de vectores característicos.

4. Una vez obtenido $\hat{\beta}$ se recalculan $\hat{\gamma}$, $\hat{\Omega}$ y $\hat{\pi}$ de acuerdo a:

$$\hat{\gamma} = S_{0k}\hat{\beta} \quad T\hat{\Omega} = S_{00} - S_{0k}\hat{\beta}\hat{\beta}'S_{k0} \quad \hat{\pi} = \hat{\gamma}\hat{\beta}'$$

Johansen demuestra que los estimadores obtenidos son consistentes.

Además:

$$\hat{\Delta}_r = \hat{\beta}'S_{k0}S_{00}^{-1}S_{0k}\hat{\beta} = \hat{\gamma}'S_{00}^{-1}\hat{\gamma} \quad \text{y} \quad \hat{\Delta}_{n-r} = \hat{\rho}'S_{00}^{-1}\hat{\rho}$$

donde $\hat{\rho}$ es la matriz de los $n-r$ vectores característicos restantes.

La estimación por *MVIC* es siempre mejor que la resultante de

estimar formas reducidas, en el sentido que en el caso en que los supuestos de exogeneidad y/o simultaneidad no se cumplan, permite obtener estimadores con propiedades deseables y cuya distribución es asintóticamente normal. Además, a partir de la estimación del sistema es posible someter a prueba la exogeneidad de las variables predeterminadas de forma muy sencilla.

Por otra parte, en el caso en que existan r vectores de cointegración, es posible plantear el sistema y estimarlos directamente, agregando los términos correspondientes a las ecuaciones:

$$A_1^*(L)(1-L)y_t = -\gamma_{01}(y_{t-1} - \beta_1'x_{t-1}) - \dots - \gamma_{0r}(y_{t-1} - \beta_r'x_{t-1}) + d(L)\varepsilon_{0t}$$

$$A_N^*(L)(1-L)x_{Nt} = -\gamma_{N1}(y_{t-1} - \beta_1'x_{t-1}) - \dots - \gamma_{Nr}(y_{t-1} - \beta_r'x_{t-1}) + d(L)\varepsilon_{Nt}$$

Determinación del rango de cointegración

Cuando se tienen $n > 2$ variables en el sistema, será necesario determinar el orden r de vectores de cointegración antes de realizar la estimación, para lo cual Johansen propone dos pruebas de muy fácil implementación.

La primera prueba plantea como hipótesis nula la existencia de r vectores de cointegración como máximo o, lo que es equivalente, de $n-r$ raíces unitarias - $(n-r)$ valores de $\lambda_i \neq 0$ -, y como alternativa que haya n vectores de cointegración (ninguna raíz unitaria, todas las variables del sistema son estacionarias). La estrategia sugerida para determinar el valor de r es comenzar con el valor $r = 0$ (que implica que $\pi = 0$, y por lo tanto el sistema VAR está en diferencias). Cuando la hipótesis nula es rechazada, se procede a formularla para $r = 1$. De esta forma secuencial, una vez que no se rechaza la hipótesis nula se obtiene, en realidad, que se rechazó la hipótesis de $r-1$ vectores de cointegración y no se rechazó la existencia de r vectores (es decir, el estadístico

para H_{r-1} es significativo y el estadístico para H_r no es significativo). El estadístico de prueba propuesto por Johansen, denominado estadístico de traza, es:

$$\eta_r = -T \ln |I - \hat{\Delta}_r| = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i) \quad r = 0, 1, \dots, n-1$$

Este estadístico de razón de verosimilitud se obtiene comparando el valor máximo de la verosimilitud restringida a que existan r vectores de cointegración con el que resulta de maximizar la verosimilitud sin restricciones (cuando existen $n-1$ relaciones de cointegración). De esta forma, su valor mide la importancia relativa de omitir los $n-r$ vectores restantes de la relación estimada, en términos del efecto potencial de sus coeficientes de ajuste al desequilibrio.

Su distribución no se puede derivar analíticamente pero resulta ser adecuadamente aproximada por una chi-cuadrado:

$$\eta_r \approx h \chi_{(c)}^2 \quad \text{con } c = 2m^2, \quad m = n-r \quad \text{y } h = 0.85 - 0.58/c$$

La segunda estrategia de prueba sugerida permite evaluar la significancia de la raíz característica mayor, λ_r . Así, la hipótesis nula plantea que existen r vectores de cointegración frente a la alternativa de que hay $(r+1)$ de estos vectores. Dicho de otra forma, frente a H_0 : "existen $(n-r)$ raíces unitarias", se plantea H_1 : "existen $[n-(r+1)]$ raíces unitarias". En este caso, nuevamente, para determinar que existen r vectores de cointegración será necesario no rechazar la hipótesis nula que lo plantea versus $(r+1)$ y además rechazar que existen $(r-1)$ vectores en favor de r . El estadístico, denominado del máximo eigenvalor (valor característico) o λ -max, es:

$$\zeta_r = -T \ln(1 - \hat{\lambda}_{r+1}) \quad r = 0, 1, \dots, n-1$$

Al igual que en el caso anterior, su distribución es una función de procesos de Wiener multivariados (resultando una generalización del DF), con lo cual los valores críticos pueden ser obtenidos por

simulaciones de Monte Carlo. Estos han sido tabulados por Johansen (1988) y también por Osterwald-Lenum (1992) y se reproducen en las tablas 8 a 12 del anexo.

De forma análoga a las pruebas de raíz unitaria, los valores críticos no son invariantes a la existencia de constante y/o tendencia en el proceso, por lo cual existe más de una versión de cada tabla.

Finalmente, se constata que la potencia de ambas pruebas disminuye con r , es decir, al aumentar el número de raíces unitarias. Se ha encontrado que el estadístico traza posee mayor potencia que el λ -max para detectar el r -ésimo vector de cointegración cuando varias de las restantes $n-r$ raíces están cerca de ser significativas. Por el contrario, si se obtiene una discriminación clara entre valores característicos grandes y pequeños, el estadístico λ -max tendrá mayor potencia (Juselius, 1991).

Pruebas de hipótesis sobre los vectores de cointegración

En muchas circunstancias resulta de interés delimitar de alguna forma el espacio generado por los vectores de cointegración. Esto es equivalente a someter a prueba restricciones lineales sobre los mismos, que tengan algún interés económico específico. Este tipo de pruebas, desarrollado ampliamente en Johansen (1988) y artículos posteriores, involucra, al igual que la determinación del rango de cointegración, la determinación de valores característicos de una matriz de momentos reducida.

El tipo de hipótesis mencionada puede referirse, por ejemplo, a la exclusión de ciertas variables de la relación de cointegración (los parámetros correspondientes en el vector β serían cero) o bien a que algunas variables pertenezcan a esa combinación lineal con parámetros que cumplen ciertas restricciones (de igualdad, de suma igual a una cierta constante, etc.). La forma general para plantear este tipo de hipótesis es:

$$H_0: \beta = H\varphi$$

con H una matriz de $n \times s$, $r \times s \times n$, de rango completo s que determina las restricciones y ϕ una matriz de $s \times r$ de parámetros desconocidos. El número s es el orden del subespacio al que pertenecen los β restringidos, por lo tanto, es igual a n menos el número de restricciones. Así, si $s = n$ en realidad no se están planteando restricciones al vector de cointegración mientras que si $s=r$ se obtiene que el espacio de cointegración está completamente especificado (en cuanto cualquier vector de cointegración será una combinación lineal de los r determinados).

El estadístico de prueba a utilizar se construye bajo el principio de razón de verosimilitud comparando el máximo de la función sin restricciones con el correspondiente a la estimación restringida a las condiciones planteadas en H_0 . De esta forma, la realización de la prueba implica la estimación de acuerdo al procedimiento ya descrito en el apartado de estimación por *MVIC* para obtener los $\hat{\lambda}_1$ sin restricciones (bajo H_1) y un procedimiento análogo para obtener los valores característicos $\hat{\lambda}_1^*$ y sus vectores asociados bajo las restricciones planteadas en H_0 . El sistema a resolver en este último caso es:

$$|\lambda^* H' S_{kk} H - H' S_{k0} S_{00}^{-1} S_{0k} H| = 0 \quad (17)$$

El estadístico bajo el principio de razón de verosimilitud es:

$$\xi_r = T \sum_{i=1}^r \ln[(1-\hat{\lambda}_1^*) / (1-\hat{\lambda}_1)] - \chi_{r(n-s)}^2$$

Obsérvese que la distribución en este caso es conocida, ya que se está trabajando bajo la hipótesis de que existen $n-r$ raíces unitarias en el proceso vectorial y, por lo tanto, se trata de un modelo en el que todas las variables son estacionarias.

Por último, conviene realizar algunas observaciones sobre el papel de la constante en la relación de cointegración y en el modelo de corrección de error: en el primer caso, su interpretación es que la media del proceso es distinta de cero, con lo cual entre las variables z_t habrá que incluir un vector de unos; en el segundo caso, la constante es el valor de la pendiente de la tendencia

lineal del proceso. A pesar de lo anterior, al plantear el MCE no necesariamente se diferenciarán ambos roles. Supóngase que se tiene un proceso multivariado de orden 4 y una relación de cointegración ($n=4, r=1$), y que el MCE está dado por:

$$\Delta z_t = \gamma \beta' z_{t-1} + \mu + \varepsilon_t \quad (18)$$

Si las variables z_{1t} tienen media distinta de cero, la relación de cointegración estará dada por: $\beta' z_t + \beta_0$ y se definen $z_t^* = (z_t, 1)$ y $\beta^* = (\beta, \beta_0)$, de forma que la relación de cointegración es expresable como: $\beta^* z_t^*$ y el modelo (18) se transforma en:

$$\Delta z_t = \gamma \beta^* z_{t-1}^* + \mu + \varepsilon_t \quad (19)$$

Sin embargo, si además existe una tendencia lineal en el componente no estacionario de la serie, se tendrá que $\mu \neq 0$ en (19) y es posible también redefinir μ como $\mu^* = \mu + \gamma \beta_0'$. Entonces, el modelo (19) puede expresarse como:

$$\Delta z_t = \gamma \beta' z_{t-1} + \mu^* + \varepsilon_t \quad (20)$$

De esta forma, μ^* puede descomponerse en dos elementos: la ordenada al origen de la relación de largo plazo y la pendiente de la tendencia lineal en el componente no estacionario.

Para someter a prueba la existencia de una tendencia lineal, será necesario estimar el modelo (20) - que no posee restricciones - y el modelo (19), que está restringido a que $\mu = 0$, es decir, a que $\mu^* = \gamma \beta_0'$.

VIII. COMENTARIOS FINALES

La reseña presentada en este trabajo se refiere únicamente a aquellos aspectos básicos en el estudio de procesos cointegrados. Su propósito es, por lo tanto, introducir al lector en la discusión del tema. Como consecuencia de ello, su alcance se ve limitado en diversas direcciones: algunos temas, como la no estacionariedad estacional, sólo han sido mencionados sin discutir a fondo sus especificidades; se han excluido algunos desarrollos más recientes que tratan, fundamentalmente, con las características y problemas adicionales que plantean los procesos integrados de orden 2 y superiores; se han omitido temas que están siendo actualmente analizados, como las dificultades que presentaría la no estabilidad de los parámetros de cointegración tanto para la estimación como para la inferencia.

Dado que la investigación sobre procesos no estacionarios avanza a gran velocidad, la actualización resulta una tarea fundamental para el analista económico. En este sentido es de esperar que el presente documento sea útil como una referencia básica que permita acceder rápidamente a los nuevos resultados de investigación.

REFERENCIAS BIBLIOGRAFICAS

A. Banerjee, J. Dolado, J.W. Galbraith y D. Hendry (1993) *Co-integration, error-correction, and the econometric analysis of non-stationary data*, Oxford University Press.

A. Bhargava (1986) "On the theory of testing for unit roots in observed time series", *Review of Economic Studies* 53.

D.A. Dickey (1976) "Estimation and hypothesis testing in nonstationary time series", PhD Dissertation, Iowa State University.

_____ y W.A. Fuller (1979) "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American Statistical Association* 74/366.

_____ y _____ (1981) "Likelihood ratio statistics for autoregressive time series with a unit root", *Econometrica* 49.

_____ y S.G. Pantula (1987) "Determining the order of differencing in autoregressive processes", *Journal of Business and Economic Statistics* 15.

J.J. Dolado y T. Jenkinson (1987) "Cointegration: a survey on recent developments", *Applied Economics Discussion Paper* 39, Institute of Economics and Statistics, University of Oxford.

R.F. Engle y C.W.J. Granger (1987) "Cointegration and error - correction: representation, estimation, and testing", *Econometrica* 55/2.

_____, D.F. Hendry y J.F. Richard (1983) "Exogeneity", *Econometrica*, 51/2.

_____ y S. Yoo (1987) "Forecasting and testing in co-integrated systems", *Journal of Econometrics* 35.

_____ y _____ (1991) "Cointegrated economic time series: an overview with new results", en R.F. Engle y C.W.J. Granger (eds.), *Long-run Economic Relationships*, Oxford University Press.

N.R. Ericsson (1992) "Cointegration, exogeneity, and policy analysis: an overview", *Journal of Policy Modeling* 3.

W.A. Fuller (1976) *Introduction to statistical time series*, John Wiley, New York.

C.W.J. Granger (1989) "Linear and non-linear cointegration", *Notes for presentation in the Latin American Meeting of the Econometric Society* (mimeo).

_____ (1991) "Developments in the study of cointegrated economic variables", en R.F. Engle y C.W.J. Granger (eds.), *Long-run Economic Relationships*, Oxford University Press.

_____ (1991) "Some recent generalizations of cointegration and the analysis of long-run relationships", en R.F. Engle y C.W.J. Granger (eds.), *Long-run Economic Relationships*, Oxford University Press.

_____ y P. Newbold (1974) "Spurious regressions in econometrics", *Journal of Econometrics* 2.

A. Hall (1989) "Testing for a unit root in the presence of moving average errors", *Biometrika* 79.

R.I.D. Harris (1992) "Testing for unit roots using the augmented Dickey-Fuller test", *Economics Letters* 38.

D.F. Hendry (1986) "Econometric modelling with cointegrated variables: an overview", *Oxford Bulletin of Economics and Statistics* 48/3.

S. Hylleberg, R.F. Engle, C.W.J. Granger y B.S. Yoo (1990) "Seasonal integration and co-integration", *Journal of Econometrics* 44.

S. Johansen (1988) "Statistical analysis of cointegration vectors", *Journal of Economic Dynamics and Control* 12.

_____ (1991) "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models", *Econometrica* 59.

_____ y K. Juselius (1990) "Maximum likelihood estimation and inference on cointegration - with applications to the demand for money", *Oxford Bulletin of Economics and Statistics* 52/2.

K. Juselius (1991) "Long-run relations in a well defined statistical model for the data generating process: Cointegration analysis of the PPP and the UIP relations", en J. Gruber (ed.), *Econometric Decision Models: New Methods of Modelling and Applications*.

J.A. Khan y M. Ogaki (1990) "A chi-square test for a unit root", *Economics Letters* 34.

J. Kiviet y G.D.A. Phillips (1992) "Exact similar tests for unit roots and cointegration", *Oxford Bulletin of Economics and Statistics* 54.

J.G. MacKinnon (1991) "Critical values for cointegration tests", en R.F. Engle y C.W.J. Granger (eds.), *Long-run Economic Relationships*, Oxford University Press.

M. Osterwald-Lenum (1992) "A note with the quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics", *Oxford Bulletin of Economics and Statistics* 54/3.

P. Perron (1988) "Trends and random walks in macroeconomic time series: further evidence for a new approach", *Journal of Economic Dynamics and Control* 12.

P.C.D. Phillips (1987) "Time series regression with a unit root", *Econometrica* 55.

_____ y B.E. Hansen (1990) "Statistical inference in instrumental variables regressions with I(1) processes", *Review of Economic Studies* 57.

_____ y P. Perron (1988) "Testing for a unit root in time series regression", *Biometrika* 75.

S.E. Saïd (1991) "Unit-roots test for time-series data with a linear trend", *Journal of Econometrics* 47.

_____ y D.A. Dickey (1984) "Testing for unit roots in autoregressive - moving average models of unknown order", *Biometrika* 71.

J.D. Sargan y A. Bhargava (1983) "Testing residuals from least squares regression for being generated by the Gaussian random walk", *Econometrica* 51.

A. Spanos (1986) *Statistical foundations of econometric modelling*, Cambridge University Press.

J.H. Stock (1987) "Asymptotic properties of least-squares estimators of cointegrating vextors", *Econometrica* 55.

K.D. West (1988) "Asymptotic normality, when regressors have a unit root", *Econometrica* 56.

G.U. Yule (1926) "Why do we sometimes get nonsense correlations between time series? A study in sampling and the nature of time series", *Journal of the Royal Statistical Society* 89.

Tabla 1

Distribución del estadístico $(\hat{\rho} - 1)/DS(\hat{\rho})$

$$PGI: y_t = y_{t-1} + u_t \quad u_t \sim iid(0, \sigma^2) \quad y_0 = 0$$

	T	p							
		0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
Caso a									
	25	-2.66	-2.26	-1.95	-1.60	0.92	1.33	1.70	2.16
	50	-2.62	-2.25	-1.95	-1.61	0.91	1.31	1.66	2.08
	100	-2.60	-2.24	-1.95	-1.61	0.90	1.29	1.64	2.03
	250	-2.58	-2.23	-1.95	-1.62	0.89	1.29	1.63	2.01
	500	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00
	∞	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00
Caso b									
	25	-3.75	-3.33	-3.00	-2.63	-0.37	0.00	0.34	0.72
	50	-3.58	-3.22	-2.93	-2.60	-0.40	-0.03	0.29	0.66
	100	-3.51	-3.17	-2.89	-2.58	-0.42	-0.05	0.26	0.63
	250	-3.46	-3.14	-2.88	-2.57	-0.42	-0.06	0.24	0.62
	500	-3.44	-3.13	-2.87	-2.57	-0.43	-0.07	0.24	0.61
	∞	-3.43	-3.12	-2.86	-2.57	-0.44	-0.07	0.23	0.60
Caso c									
	25	-4.38	-3.95	-3.60	-3.24	-1.14	-0.80	-0.50	-0.15
	50	-4.15	-3.80	-3.50	-3.18	-1.19	-0.87	-0.58	-0.24
	100	-4.04	-3.73	-3.45	-3.15	-1.22	-0.90	-0.62	-0.28
	250	-3.99	-3.69	-3.43	-3.13	-1.23	-0.92	-0.64	-0.31
	500	-3.98	-3.68	-3.42	-3.13	-1.24	-0.93	-0.65	-0.32
	∞	-3.96	-3.66	-3.41	-3.12	-1.25	-0.94	-0.66	-0.33

Nota:

p: probabilidad de obtener un valor menor. Por ejemplo, en el caso a, para T=100, la $\text{Prob}[(\hat{\rho}-1)/DS(\hat{\rho}) < 1.64] = 0.975$

Fuente: Banerjee et al. (1993)

Tabla 2

Distribución del estadístico $T(\hat{\rho} - 1)$

$$PGI: Y_t = Y_{t-1} + u_t \quad u_t \sim iid(0, \sigma^2) \quad Y_0 = 0$$

	T	p							
		0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
Caso a									
	25	-11.9	-9.3	-7.3	-5.3	1.01	1.40	1.79	2.28
	50	-12.9	-9.9	-7.7	-5.5	0.97	1.35	1.70	2.16
	100	-13.3	-10.2	-7.9	-5.6	0.95	1.31	1.65	2.09
	250	-13.6	-10.3	-8.0	-5.7	0.93	1.28	1.62	2.04
	500	-13.7	-10.4	-8.0	-5.7	0.93	1.28	1.61	2.04
	∞	-13.8	-10.5	-8.1	-5.7	0.93	1.28	1.60	2.03
Caso b									
	25	-17.2	-14.6	-12.5	-10.2	-0.76	0.01	0.65	1.40
	50	-18.9	-15.7	-13.3	-10.7	-0.81	-0.07	0.53	1.22
	100	-19.8	-16.3	-13.7	-11.0	-0.83	-0.10	0.47	1.14
	250	-20.3	-16.6	-14.0	-11.2	-0.84	-0.12	0.43	1.09
	500	-20.5	-16.8	-14.0	-11.2	-0.84	-0.13	0.42	1.06
	∞	-20.7	-16.9	-14.1	-11.3	-0.85	-0.13	0.41	1.04
Caso c									
	25	-22.5	-19.9	-17.9	-15.6	-3.66	-2.51	-1.53	-0.43
	50	-25.7	-22.4	-19.8	-16.8	-3.71	-2.60	-1.66	-0.65
	100	-27.4	-23.6	-20.7	-17.5	-3.74	-2.62	-1.73	-0.75
	250	-28.4	-24.4	-21.3	-18.0	-3.75	-2.64	-1.78	-0.82
	500	-28.9	-24.8	-21.5	-18.1	-3.76	-2.65	-1.78	-0.84
	∞	-29.5	-25.1	-21.8	-18.3	-3.77	-2.66	-1.79	-0.87

Nota:

p: probabilidad de obtener un valor menor. Por ejemplo, en el caso a, para T=100, la $Prob[T(\hat{\rho}-1) < 1.65] = 0.975$

Fuente: Banerjee et al. (1993)

Tabla 3

Valores críticos CRDW

$p = 0.95$

T	11		21		31		51		101	
	Ri	Rs	Ri	Rs	Ri	Rs	Ri	Rs	Ri	Rs
n										
0*	1.733	1.733	1.069	1.069	0.770	0.770	0.493	0.493	0.259	0.259
1	1.592	2.404	1.022	1.560	0.747	1.156	0.484	0.755	0.257	0.404
3	1.273	3.201	0.925	2.273	0.699	1.736	0.465	1.172	0.252	0.645
5	0.916	3.641	0.827	2.775	0.651	2.185	0.446	1.518	0.247	0.855
7			0.719	3.152	0.601	2.552	0.427	1.818	0.242	1.046
10			0.556	3.553	0.526	2.989	0.399	2.205	0.234	1.308
15							0.350	2.723	0.221	1.692
20									0.209	2.205

$p = 0.99$

T	11		21		31		51		101	
	Ri	Rs	Ri	Rs	Ri	Rs	Ri	Rs	Ri	Rs
n										
0*	2.264	2.264	1.470	1.470	1.081	1.081	0.705	0.705	0.376	0.376
1	2.095	2.843	1.409	1.960	1.050	1.480	0.692	0.988	0.373	0.538
3	1.689	3.468	1.281	2.608	0.985	2.046	0.666	1.416	0.366	0.795
5	1.204	3.781	1.143	3.048	0.919	2.467	0.640	1.759	0.358	1.013
7			1.000	3.364	0.851	2.800	0.613	2.049	0.351	1.208
10			0.771	3.685	0.745	3.186	0.573	2.416	0.340	1.467
15							0.504	2.890	0.322	1.840
20									0.304	2.154

Notas:

- 0 : estos límites corresponden a un camino aleatorio simple.
- Ri/Rs: límite inferior/superior de confianza de R para la primera diferencia de los residuos de la regresión.
- T: número de observaciones.
- n: número de variables exógenas.

Fuente: Sargan y Bhargava (1983)

Tabla 4

Distribución del estadístico de prueba para otros parámetros

$$PGI: Y_t = Y_{t-1} + u_t \quad u_t \sim iid(0, \sigma^2) \quad Y_0 = 0$$

T	P	0.90	0.95	0.975	0.99
---	---	------	------	-------	------

1. Modelo b' $H_0: \mu_b = 0$; Estadístico: τ_1

25	2.20	2.61	2.97	3.41
50	2.18	2.56	2.89	3.28
100	2.17	2.54	2.86	3.22
250	2.16	2.53	2.84	3.19
500	2.16	2.52	2.83	3.18
∞	2.16	2.52	2.83	3.18

2. Modelo c' $H_0: \mu_c = 0$; Estadístico: τ_2

25	2.77	3.20	3.59	4.05
50	2.75	3.14	3.47	3.87
100	2.73	3.11	3.42	3.78
250	2.73	3.09	3.39	3.74
500	2.72	3.08	3.38	3.72
∞	2.72	3.08	3.38	3.71

3. Modelo c' $H_0: \gamma_c = 0$; Estadístico: τ_3

25	2.39	2.85	3.25	3.74
50	2.38	2.81	3.18	3.60
100	2.38	2.79	3.14	3.53
250	2.38	2.79	3.12	3.49
500	2.38	2.78	3.11	3.48
∞	2.38	2.78	3.11	3.46

Nota:

p: probabilidad de obtener un valor menor. Por ejemplo, en el caso 1, para T=100, la $Prob(\tau_1 < 3.22) = 0.99$

Fuente: Banerjee et al. (1993)

Tabla 5

Distribución del estadístico de prueba para otros parámetros

T	p							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
1. Modelo b'								
$H_0: \mu_b = 0, \rho_b = 1$; Estadístico: ϕ_1								
PGI: $y_t = y_{t-1} + u_t$								
25	0.29	0.38	0.49	0.65	4.12	5.18	6.30	7.88
50	0.29	0.39	0.50	0.66	3.94	4.86	5.80	7.06
100	0.29	0.39	0.50	0.67	3.86	4.71	5.57	6.70
250	0.30	0.39	0.51	0.67	3.81	4.63	5.45	6.52
500	0.30	0.39	0.51	0.67	3.79	4.61	5.41	6.47
∞	0.30	0.40	0.51	0.67	3.78	4.59	5.38	6.43
2. Modelo c'								
$H_0: \mu_c = 0, \gamma_c = 0, \rho_c = 1$; Estadístico: ϕ_2								
PGI: $y_t = y_{t-1} + u_t$								
25	0.61	0.75	0.89	1.10	4.67	5.68	6.75	8.21
50	0.62	0.77	0.91	1.12	4.31	5.13	5.94	7.02
100	0.63	0.77	0.92	1.12	4.16	4.88	5.59	6.50
250	0.63	0.77	0.92	1.13	4.07	4.75	5.40	6.22
500	0.63	0.77	0.92	1.13	4.05	4.71	5.35	6.15
∞	0.63	0.77	0.92	1.13	4.03	4.68	5.31	6.09
3. Modelo c'								
$H_0: \gamma_c = 0, \rho_c = 1$; Estadístico: ϕ_3								
PGI: $y_t = \mu_c + y_{t-1} + u_t$								
25	0.74	0.90	1.08	1.33	5.91	7.24	8.65	10.61
50	0.76	0.93	1.11	1.37	5.61	6.73	7.81	9.31
100	0.76	0.94	1.12	1.38	5.47	6.49	7.44	8.73
250	0.76	0.94	1.13	1.39	5.39	6.34	7.25	8.43
500	0.76	0.94	1.13	1.39	5.36	6.30	7.20	8.34
∞	0.77	0.94	1.13	1.39	5.34	6.25	7.16	8.27

Nota:

p: probabilidad de obtener un valor menor. Por ejemplo, en el caso 1, para T=100, la $\text{Prob}(\phi_1 < 6.70) = 0.99$

Fuente: Banerjee et al. (1993)

Tabla 6

Pruebas de raíces unitarias estacionales: valores críticos

$$PGI: \Delta_4 x_t = \varepsilon_t \sim \text{nid}(0,1)$$

Regresiones Auxiliares	T	Percentiles							
		't': π_1				't': π_2			
		0.01	0.025	0.05	0.10	0.01	0.025	0.05	0.10
(a)	48	-2.72	-2.29	-1.95	-1.59	-2.67	-2.27	-1.95	-1.60
	100	-2.60	-2.26	-1.97	-1.61	-2.61	-2.22	-1.92	-1.57
	136	-2.62	-2.25	-1.93	-1.59	-2.60	-2.23	-1.94	-1.61
	200	-2.62	-2.23	-1.94	-1.62	-2.60	-2.24	-1.95	-1.61
(b)	48	-3.66	-3.25	-2.96	-2.62	-2.68	-2.27	-1.95	-1.60
	100	-3.47	-3.14	-2.88	-2.58	-2.61	-2.24	-1.95	-1.60
	136	-3.51	-3.17	-2.89	-2.58	-2.60	-2.21	-1.91	-1.58
	200	-3.48	-3.13	-2.87	-2.57	-2.58	-2.22	-1.92	-1.59
(c)	48	-3.77	-3.39	-3.08	-2.72	-3.75	-3.37	-3.04	-2.69
	100	-3.55	-3.22	-2.95	-2.63	-3.60	-3.22	-2.94	-2.63
	136	-3.56	-3.23	-2.94	-2.62	-3.49	-3.15	-2.90	-2.59
	200	-3.51	-3.18	-2.91	-2.59	-3.50	-3.16	-2.89	-2.60
(d)	48	-4.23	-3.85	-3.56	-3.21	-2.65	-2.24	-1.91	-1.57
	100	-4.07	-3.73	-3.47	-3.16	-2.58	-2.24	-1.94	-1.60
	136	-4.09	-3.75	-3.46	-3.16	-2.65	-2.25	-1.96	-1.63
	200	-4.05	-3.70	-3.44	-3.15	-2.59	-2.25	-1.95	-1.62
(e)	48	-4.46	-4.04	-3.71	-3.37	-3.80	-3.41	-3.08	-2.73
	100	-4.09	-3.80	-3.53	-3.22	-3.60	-3.22	-2.94	-2.63
	136	-4.15	-3.80	-3.52	-3.21	-3.57	-3.18	-2.93	-2.61
	200	-4.05	-3.74	-3.49	-3.18	-3.52	-3.18	-2.91	-2.60

(Continuación)

Pruebas de raíces unitarias estacionales: valores críticos

$$PGI: \Delta_4 x_t = \varepsilon_t \sim \text{nid}(0,1)$$

Regresiones Auxiliares	T	Percentiles							
		0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
't': π_4									
(a)	48	-2.51	-2.11	-1.76	-1.35	1.33	1.72	2.05	2.49
	100	-2.43	-2.01	-1.68	-1.32	1.31	1.67	2.00	2.40
	136	-2.44	-1.99	-1.68	-1.31	1.30	1.66	1.99	2.38
	200	-2.43	-1.98	-1.65	-1.30	1.29	1.67	1.97	2.36
(b)	48	-2.44	-2.06	-1.72	-1.33	1.30	1.68	2.04	2.41
	100	-2.38	-1.99	-1.68	-1.30	1.28	1.65	1.97	2.32
	136	-2.36	-1.98	-1.68	-1.31	1.27	1.65	1.97	2.31
	200	-2.36	-1.98	-1.66	-1.29	1.28	1.65	1.96	2.30
(c)	48	-2.86	-2.37	-1.98	-1.53	1.54	1.96	2.35	2.81
	100	-2.78	-2.32	-1.96	-1.53	1.52	1.93	2.29	2.73
	136	-2.72	-2.31	-1.96	-1.52	1.51	1.92	2.28	2.71
	200	-2.74	-2.33	-1.96	-1.54	1.53	1.95	2.32	2.78
(d)	48	-2.41	-2.05	-1.70	-1.33	1.26	1.64	1.96	2.37
	100	-2.38	-1.97	-1.65	-1.28	1.28	1.65	1.98	2.32
	136	-2.36	-1.97	-1.64	-1.29	1.26	1.62	1.92	2.31
	200	-2.35	-1.97	-1.66	-1.29	1.26	1.64	1.96	2.30
(e)	48	-2.75	-2.26	-1.91	-1.48	1.51	1.97	2.34	2.78
	100	-2.76	-2.32	-1.94	-1.51	1.51	1.92	2.28	2.69
	136	-2.71	-2.78	-1.94	-1.51	1.53	1.96	2.31	2.78
	200	-2.65	-2.27	-1.92	-1.48	1.55	1.97	2.31	2.71

(Continuación)

Pruebas de raíces unitarias estacionales: valores críticos

$$PGI: \Delta_4 x_t = \varepsilon_t \sim \text{nid}(0,1)$$

Regresiones Auxiliares	T	Percentiles							
		0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
		't': π_3				'F': $\pi_3 \cap \pi_4$			
		-----				-----			
(a)	48	-2.66	-2.23	-1.93	-1.52	2.45	3.26	4.04	5.02
	100	-2.55	-2.18	-1.90	-1.53	2.39	3.12	3.89	4.89
	136	-2.58	-2.21	-1.92	-1.56	2.41	3.14	3.86	4.81
	200	-2.58	-2.24	-1.92	-1.55	2.42	3.16	3.92	4.81
(b)	48	-2.64	-2.23	-1.90	-1.52	2.32	3.04	3.78	4.78
	100	-2.61	-2.23	-1.90	-1.54	2.35	3.08	3.81	4.77
	136	-2.53	-2.18	-1.88	-1.53	2.36	3.00	3.70	4.73
	200	-2.57	-2.21	-1.90	-1.53	2.37	3.12	3.86	4.76
(c)	48	-4.31	-3.92	-3.61	-3.24	5.50	6.60	7.68	9.22
	100	-4.06	-3.72	-3.44	-3.14	5.56	6.57	7.72	8.74
	136	-4.06	-3.72	-3.44	-3.11	5.56	6.63	7.66	8.92
	200	-4.00	-3.67	-3.38	-3.07	5.56	6.61	7.53	8.93
(d)	48	-2.68	-2.27	-1.92	-1.52	2.23	2.95	3.70	4.64
	100	-2.56	-2.19	-1.89	-1.54	2.31	2.98	3.71	4.70
	136	-2.56	-2.20	-1.90	-1.52	2.33	3.04	3.69	4.57
	200	-2.58	-2.21	-1.92	-1.56	2.34	3.07	3.76	4.66
(e)	48	-4.46	-4.02	-3.66	-3.28	5.37	6.55	7.70	9.27
	100	-4.12	-3.76	-3.48	-3.14	5.52	6.60	7.52	8.79
	136	-4.05	-3.72	-3.44	-3.12	5.55	6.62	7.59	8.77
	200	-4.04	-3.69	-3.41	-3.10	5.56	6.57	7.56	8.96

Nota:

Las regresiones auxiliares son las siguientes:

- (a): sin constante, sin binarias estacionales y sin tendencia.
- (b): con constante, sin binarias estacionales y sin tendencia.
- (c): con constante, con binarias estacionales y sin tendencia.
- (d): con constante, sin binarias estacionales y con tendencia.
- (e): con constante, con binarias estacionales y con tendencia.

Fuente: Hylleberg et al (1990)

Tabla 7

Pruebas de cointegración

Superficies de respuesta de MacKinnon

n	Modelo	p(%)	ϕ_{∞}	DS	ϕ_1	ϕ_2
1	Sin constante Sin tendencia	1	-2.5658	0.0023	-1.960	-10.04
		5	-1.9393	0.0008	-0.398	0.00
		10	-1.6156	0.0007	-0.181	0.00
1	Constante Sin tendencia	1	-3.4336	0.0024	-5.999	-29.25
		5	-2.8621	0.0011	-2.738	-8.36
		10	-2.5671	0.0009	-1.438	-4.48
1	Constante Tendencia	1	-3.9638	0.0019	-8.353	-47.44
		5	-3.4126	0.0012	-4.039	-17.83
		10	-3.1279	0.0009	-2.418	-7.58
2	Constante Sin tendencia	1	-3.9001	0.0022	-10.534	-30.03
		5	-3.3377	0.0012	-5.967	-8.98
		10	-3.0462	0.0009	-4.069	-5.73
2	Constante Tendencia	1	-4.3266	0.0022	-15.531	-34.03
		5	-3.7809	0.0013	-9.421	-15.06
		10	-3.4959	0.0009	-7.203	-4.01
3	Constante Sin tendencia	1	-4.2981	0.0023	-13.790	-46.37
		5	-3.7429	0.0012	-8.352	-13.41
		10	-3.4518	0.0010	-6.241	-2.79
3	Constante Tendencia	1	-4.6676	0.0022	-18.492	-49.35
		5	-4.1193	0.0011	-12.024	-13.13
		10	-3.8344	0.0009	-9.188	-4.85
4	Constante Sin tendencia	1	-4.6493	0.0023	-17.188	-59.20
		5	-4.1000	0.0012	-10.745	-21.57
		10	-3.8110	0.0009	-8.317	-5.19
4	Constante Tendencia	1	-4.9695	0.0021	-22.504	-50.22
		5	-4.4294	0.0012	-14.501	-19.54
		10	-4.1474	0.0010	-11.165	-9.88
5	Constante Sin tendencia	1	-4.9587	0.0026	-22.140	-37.29
		5	-4.4185	0.0013	-13.641	-21.16
		10	-4.1327	0.0009	-10.638	-5.48
5	Constante Tendencia	1	-5.2497	0.0024	-26.606	-49.56
		5	-4.7154	0.0013	-17.432	-16.50
		10	-4.4345	0.0010	-13.654	-5.77
6	Constante Sin tendencia	2	-5.2400	0.0029	-26.278	-41.65
		5	-4.7048	0.0018	-17.120	-11.17
		10	-4.4242	0.0010	-13.347	0.00
6	Constante Tendencia	1	-5.5127	0.0033	-30.735	-52.50
		5	-4.9767	0.0017	-20.883	-9.05
		10	-4.6999	0.0011	-16.445	0.00

Fuente: Banerjee et al (1993)

Tabla 8

Cuantiles de la distribución asintótica de los estadísticos para la prueba de Rango de Cointegración

Caso 1.

$$PGI: \Delta x_t = \Gamma_1 \Delta x_{t-1} + \dots + \Gamma_{k-1} \Delta x_{t-k+1} + \alpha \beta' x_{t-1} + \psi D_t + \varepsilon_t$$

p-r	99%	97.5%	95%	90%	80%	50%	20%
	λ -max						
1	6.51	4.93	3.84	2.86	1.82	0.58	0.09
2	15.69	13.27	11.44	9.52	7.58	4.83	2.89
3	22.99	20.02	17.89	15.59	13.31	9.71	6.96
4	28.82	26.14	23.80	21.58	18.97	14.94	11.55
5	35.17	32.51	30.04	27.62	24.83	20.16	16.41
6	41.00	38.59	36.36	33.62	30.37	25.54	21.39
7	47.15	44.28	41.51	38.98	36.01	30.65	26.22
8	53.90	50.78	47.99	44.99	41.65	36.18	31.37
9	59.78	56.55	53.69	50.65	47.26	41.36	36.22
10	65.21	61.57	59.06	56.09	52.65	46.69	41.48
11	72.36	68.35	65.30	61.96	58.11	52.06	46.50
	Traza						
1	6.51	4.93	3.84	2.86	1.82	0.58	0.09
2	16.31	14.43	12.53	10.47	8.45	5.42	3.32
3	29.75	26.64	24.31	21.63	18.83	14.30	10.55
4	45.58	42.30	39.89	36.58	33.16	27.10	22.04
5	66.52	62.91	59.46	55.44	51.13	43.79	37.23
6	90.45	86.09	82.49	78.36	73.10	64.24	56.53
7	119.80	114.22	109.99	104.77	98.87	88.53	79.15
8	152.32	146.78	141.20	135.24	128.67	116.97	106.32
9	187.31	181.44	175.77	169.45	161.63	148.61	136.23
10	226.40	219.88	212.67	206.05	198.04	184.15	170.59
11	269.81	261.71	255.27	248.45	239.44	223.51	209.00

(Continuación)

Cuantiles de la distribución asintótica de los estadísticos
para la prueba de Rango de Cointegración

Caso 1.

$$PGI: \Delta x_t = \Gamma_1 \Delta x_{t-1} + \dots + \Gamma_{k-1} \Delta x_{t-k+1} + \alpha \beta' x_{t-1} + \psi D_t + \varepsilon_t$$

p-r	10%	5%	2.5%	1%	Media	Var
			λ -max			
1	0.02	0.01	0.00	0.00	1.09	1.94
2	2.19	1.72	1.41	1.10	5.45	9.48
3	5.77	4.94	4.35	3.77	10.32	16.21
4	10.17	9.09	8.24	7.50	15.49	20.88
5	14.72	13.38	12.38	11.29	20.76	26.39
6	19.53	17.98	16.66	15.33	26.07	31.21
7	24.09	22.51	21.27	19.70	31.22	34.76
8	29.19	27.52	26.09	24.67	36.70	38.86
9	33.92	32.20	30.95	29.12	41.92	43.77
10	38.80	36.86	35.38	33.74	47.18	46.13
11	43.98	41.92	40.26	38.39	52.59	51.62
			Traza			
1	0.02	0.01	0.00	0.00	1.09	1.94
2	2.54	1.97	1.63	1.28	6.09	10.86
3	8.93	7.79	6.91	6.03	14.90	25.80
4	19.75	17.90	16.55	14.83	27.74	44.31
5	34.16	31.62	29.66	27.32	44.41	71.20
6	52.68	49.55	46.67	43.73	64.90	101.33
7	74.48	70.63	67.55	64.67	89.20	140.76
8	100.87	96.81	93.36	89.04	117.71	181.52
9	129.81	125.30	121.36	116.39	149.14	233.42
10	163.48	157.65	153.21	147.15	184.58	279.79
11	201.28	194.77	189.68	183.78	224.28	335.26

Fuente: Osterwald-Lenum (1992)

Tabla 9

Cuantiles de la distribución asintótica de los estadísticos para la prueba de Rango de Cointegración

Caso 2.

$$PGI: \Delta x_t = \Gamma_1 \Delta x_{t-1} + \dots + \Gamma_{k-1} \Delta x_{t-k+1} + \alpha(\beta', \beta_0)(x'_{t-1}, 1)' + \psi D_t + \varepsilon_t$$

p-r	50%	80%	90%	95%	97.5%	99%	Media	Var
<i>λ-max</i>								
1	3.40	5.91	7.52	9.24	10.80	12.97	4.03	7.07
2	8.27	11.54	13.75	15.67	17.63	20.20	8.86	13.08
3	13.47	17.40	19.77	22.00	24.07	26.81	14.02	19.24
4	18.70	22.95	25.56	28.14	30.32	33.24	19.23	23.83
5	23.78	28.76	31.66	34.40	36.90	39.79	24.48	29.26
6	29.08	34.25	37.45	40.30	43.22	46.82	29.72	34.63
7	34.73	40.13	43.25	46.45	48.99	51.91	35.18	38.35
8	39.70	45.53	48.91	52.00	54.71	57.95	40.35	41.98
9	44.97	50.73	54.35	57.42	60.50	63.71	45.55	44.13
10	50.21	56.52	60.25	63.57	66.24	69.94	50.82	49.28
11	55.70	62.38	66.02	69.74	72.64	76.63	56.33	54.99
<i>Traza</i>								
1	3.40	5.91	7.52	9.24	10.80	12.97	4.03	7.07
2	11.25	15.25	17.85	19.96	22.05	24.60	11.91	18.94
3	23.28	28.75	32.00	34.91	37.61	41.07	23.84	37.98
4	38.84	45.65	49.65	53.12	56.06	60.16	39.50	59.42
5	58.46	66.91	71.86	76.07	80.06	84.45	59.16	91.65
6	81.90	91.57	97.18	102.14	106.74	111.01	82.49	126.94
7	109.17	120.35	126.58	131.70	136.49	143.09	109.75	167.91
8	139.83	152.56	159.48	165.58	171.28	177.20	140.57	208.09
9	174.88	198.08	196.37	202.92	208.81	215.74	175.44	257.84
10	212.93	228.08	236.54	244.15	251.30	257.68	213.53	317.24
11	254.84	272.82	282.45	291.40	298.31	307.64	256.15	413.35

Fuente: Osterwald-Lenum (1992)

Tabla 10

Cuantiles de la distribución asintótica de los estadísticos para la prueba de Rango de Cointegración

Caso 3.

$$PGI: \Delta x_t = \Gamma_1 \Delta x_{t-1} + \dots + \Gamma_{k-1} \Delta x_{t-k+1} + \alpha \beta' x_{t-1} + \mu_0 + \psi D_t + \varepsilon_t$$

p-r	50%	80%	90%	95%	97.5%	99%	Media	Var
	λ -max							
1	0.44	1.66	2.69	3.76	4.95	6.65	0.99	2.04
2	6.85	10.04	12.07	14.07	16.05	18.63	7.47	12.42
3	12.34	16.20	18.60	20.97	23.09	25.52	12.88	18.67
4	17.66	21.98	24.73	27.07	28.98	32.24	18.26	23.47
5	23.05	27.85	30.90	33.46	35.71	38.77	23.67	28.82
6	28.45	33.67	36.76	39.37	41.86	45.10	29.06	33.57
7	33.83	39.12	42.32	45.28	47.96	51.57	34.37	37.41
8	39.29	45.05	48.33	51.42	54.29	57.69	39.85	42.90
9	44.58	50.55	53.98	57.12	59.33	62.80	45.10	44.93
10	49.66	55.97	59.62	62.81	65.44	69.09	50.29	49.41
11	54.99	61.55	65.38	68.83	72.11	75.95	55.63	54.92
	Traza							
1	0.44	1.66	2.69	3.76	4.95	6.65	0.99	2.04
2	7.55	11.07	13.33	15.41	17.52	20.04	8.23	14.38
3	18.70	23.64	26.79	29.68	32.56	35.65	19.32	32.43
4	33.60	40.15	43.95	47.21	50.35	54.46	34.24	52.75
5	52.30	60.29	64.84	68.52	71.80	76.07	52.95	79.25
6	75.26	84.57	89.48	94.15	98.33	103.18	75.74	114.65
7	101.22	112.30	118.50	124.24	128.45	133.57	101.91	158.78
8	131.62	143.97	150.53	156.00	161.32	168.36	132.09	201.82
9	165.11	178.90	186.39	192.89	198.82	204.95	165.90	246.45
10	202.58	217.81	225.85	233.13	239.46	247.18	203.39	300.80
11	243.90	260.82	269.96	277.71	284.87	293.44	244.66	379.56

Fuente: Osterwald-Lenum (1992)

Tabla 11
Cuantiles de la distribución asintótica de los estadísticos
para la prueba de Rango de Cointegración

Caso 4.

$$PGI: \Delta x_t = \Gamma_1 \Delta x_{t-1} + \dots + \Gamma_{k-1} \Delta x_{t-k+1} + \alpha(\beta', \beta_1)(x'_{t-1}, t)' + \mu_0 + \psi D_t + \varepsilon_t$$

p-r	50%	80%	90%	95%	97.5%	99%	Media	Var
	<i>λ-max</i>							
1	5.55	8.65	10.49	12.25	14.21	16.26	6.22	10.11
2	10.90	14.70	16.85	18.96	21.14	23.65	11.51	16.38
3	16.24	20.45	23.11	25.54	27.68	30.34	16.82	22.01
4	21.50	26.30	29.12	31.46	33.60	36.65	22.08	27.74
5	26.72	31.72	34.75	37.52	40.01	42.36	27.32	31.36
6	32.01	37.50	40.91	43.97	46.84	49.51	32.68	37.91
7	37.57	43.11	46.32	49.42	51.94	54.71	38.06	39.74
8	42.72	48.56	52.16	55.50	58.08	62.46	43.34	44.83
9	48.17	54.34	57.87	61.29	64.12	67.88	48.74	49.20
10	53.21	59.49	63.18	66.23	69.56	73.73	53.74	52.64
11	58.54	64.97	69.26	72.72	75.72	79.23	59.15	56.97
	<i>Traza</i>							
1	5.55	8.65	10.49	12.25	14.21	16.26	6.22	10.11
2	15.59	20.19	22.76	25.32	27.75	30.45	16.20	24.90
3	29.53	35.56	39.06	42.44	45.42	48.45	30.15	45.68
4	47.17	54.80	59.14	62.99	66.25	70.05	47.79	74.48
5	68.64	77.83	83.20	87.31	91.06	96.58	69.35	106.56
6	94.05	104.73	110.42	114.90	119.29	124.75	94.67	143.33
7	122.87	134.57	141.01	146.76	152.52	158.49	123.51	182.85
8	155.40	169.10	176.67	182.82	187.91	196.08	156.41	234.11
9	192.37	207.25	215.17	222.21	228.05	234.41	193.03	288.30
10	231.59	247.91	256.72	263.42	270.33	279.07	232.25	345.23
11	276.34	294.12	303.13	310.81	318.02	327.45	276.88	416.98

Fuente: Osterwald-Lenum (1992)

Tabla 12
Cuantiles de la distribución asintótica de los estadísticos
para la prueba de Rango de Cointegración

Caso 5.

$$PGI: \Delta x_t = \Gamma_1 \Delta x_{t-1} + \dots + \Gamma_{k-1} \Delta x_{t-k+1} + \alpha \beta' x_{t-1} + \mu_1 t + \mu_0 + \psi D_t + \varepsilon_t$$

p-r	50%	80%	90%	95%	97.5%	99%	Media	Var
				λ -max				
1	0.45	1.61	2.57	3.74	4.85	6.40	0.98	1.96
2	8.84	12.55	14.84	16.87	18.57	21.47	9.46	16.02
3	14.70	18.94	21.53	23.78	26.07	28.83	15.30	21.57
4	19.99	24.81	27.76	30.33	32.56	35.68	20.72	27.29
5	25.78	30.75	33.74	36.41	38.68	41.58	26.27	31.67
6	30.96	36.51	39.50	42.48	45.12	48.17	31.57	36.39
7	36.44	42.07	45.49	48.45	51.46	54.48	37.07	40.22
8	41.68	47.51	51.14	54.25	56.87	60.81	42.33	44.31
9	46.92	53.12	57.01	60.29	62.98	66.91	47.56	49.45
10	52.33	59.01	62.69	66.10	69.41	72.96	53.02	54.08
11	57.76	64.40	68.22	71.68	74.90	78.51	58.43	55.68
				Traza				
1	0.45	1.61	2.57	3.74	4.85	6.40	0.98	1.96
2	9.68	13.56	16.06	18.17	20.13	23.46	10.31	18.34
3	22.66	28.13	31.42	34.55	36.94	40.49	23.31	37.90
4	39.43	46.66	50.74	54.64	57.79	61.24	40.19	64.17
5	60.33	68.66	73.40	77.74	80.94	85.78	60.93	92.36
6	84.53	94.45	100.14	104.94	109.62	114.36	85.13	129.92
7	112.75	124.18	130.84	136.61	141.55	146.99	113.48	173.03
8	144.39	157.11	164.34	170.80	176.43	182.51	145.11	217.53
9	179.72	194.04	201.95	208.97	215.41	222.46	180.57	270.83
10	219.42	235.26	244.12	250.84	256.60	263.94	220.05	332.87
11	262.30	279.31	289.08	295.99	303.98	312.58	262.92	398.78

Fuente: Osterwald-Lenum (1992)