

## Detección de raíces unitarias y cointegración mediante métodos de subespacios

### Subspace-Based Methods to Determine Unit Roots and Cointegrating Ranks

ALFREDO GARCÍA-HIERNAUX<sup>1,a</sup>, JOSÉ CASALS<sup>2,b</sup>, MIGUEL JEREZ<sup>2,c</sup>

<sup>1</sup>DEPARTAMENTO DE ESTADÍSTICA, UNIVERSIDAD CARLOS III DE MADRID, ESPAÑA

<sup>2</sup>DEPARTAMENTO DE ECONOMÍA CUANTITATIVA, UNIVERSIDAD COMPLUTENSE DE MADRID,  
ESPAÑA

---

#### Resumen

Proponemos un nuevo procedimiento para detectar raíces unitarias basado en métodos de subespacios. Su planteamiento comporta tres aspectos fundamentales. Primero, la misma metodología se puede aplicar a series individuales o a vectores de series temporales. Segundo, utiliza una familia flexible de criterios de información, cuyas funciones de pérdida se pueden adaptar a las propiedades estadísticas de los datos. Finalmente, no requiere especificar un proceso estocástico para las series analizadas. Se demuestra la consistencia del método y los ejercicios de simulación revelan buenas propiedades en muestras finitas. Además, su aplicación práctica se ilustra mediante el análisis de varias series reales.

**Palabras clave:** modelos estado espacio, análisis de series temporales, criterios de información, función de pérdida.

#### Abstract

We propose a new procedure to detect unit roots based on subspace methods. It has three main original aspects. First, the same method can be applied to single or multiple time series. Second, it uses a flexible family of information criteria, which loss functions can be adapted to the statistical properties of the data. Last, it does not require the specification of a stochastic process for the series analyzed. This procedure is consistent and a simulation exercise shows that it has good finite sample properties. Its application is illustrated with the analysis of several real time series.

**Key words:** State space models, Time series analysis, Information criteria, Loss function.

---

<sup>a</sup>Profesor ayudante doctor. E-mail: aghierna@est-econ.uc3m.es

<sup>b</sup>Profesor asociado. E-mail: jcasalsc@cajamadrid.es

<sup>c</sup>Profesor titular. E-mail: mjerez@ccee.ucm.es

## 1. Introducción

Este trabajo desarrolla un procedimiento para detectar raíces unitarias en series temporales, que tiene su origen en los métodos aplicados habitualmente a sistemas físicos. En ingeniería aeronáutica y de telecomunicaciones existe una importante tradición de análisis de este tipo de series, utilizando instrumentos que no se han extendido completamente a otros campos. Estas técnicas se conocen como de *identificación de sistemas*. El trabajo pionero de Ho & Kalman (1966) sentó las bases de la estrategia de modelización en esta área que se centra en obtener, a partir de los datos, un modelo en forma de espacio de los estados. Este problema se aborda generalmente con un enfoque geométrico-algebraico, aunque también cabe destacar aproximaciones estadísticas como las de Akaike (1976) o Larimore (1990).

En el campo de la ingeniería, desde los años noventa, la identificación de sistemas ha estado dominada por los *métodos de subespacios*; véase Favoreel et al. (2000) para una perspectiva general. Muchos de estos métodos parten de la idea de que, si se observaran los estados de un sistema, las matrices que lo caracterizan podrían estimarse eficientemente por mínimos cuadrados ordinarios. Puesto que en general no ocurre así, el problema de identificación se reduce a aproximar los estados a partir de los datos observables, sin tener información a priori de la estructura del sistema.

En comparación con las técnicas convencionales de análisis estadístico de series temporales, estos métodos tienen tres ventajas fundamentales:

1. Permiten identificar un modelo lineal general a partir de los datos, sin información a priori sobre la estructura del proceso.
2. El tratamiento de los casos univariante y multivariante es conceptualmente el mismo.
3. Se trata de métodos computacionalmente eficientes, apoyados en técnicas numéricas robustas del álgebra lineal. Además no son iterativos y, por tanto, carecen de problemas de convergencia.

La aplicación de estos métodos a series temporales que no proceden de sistemas físicos presenta importantes desafíos. Por ejemplo, las series temporales en ingeniería suelen ser estacionarias y típicamente carecen de componentes estacionales, por lo que los métodos de subespacios disponibles carecen de instrumentos adecuados para acomodar estas características de los datos que son muy comunes, por ejemplo, en series temporales económicas.

En este trabajo se aborda una de estas cuestiones: la detección de raíces unitarias. Para ello, partimos del algoritmo CCA (*Canonical Correlation Analysis*) de Larimore (1983). Este método resulta especialmente conveniente ya que, debido a su elección de matrices de ponderación, el algoritmo calcula eficientemente los coeficientes de correlación canónica (CCC) entre los subespacios de información pasada y futura (Bauer & Ljung 2002). Esta propiedad, además de facilitar la

interpretación estadística del procedimiento, constituye el fundamento del método que proponemos. La idea básica es que los CCC correspondientes a raíces unitarias convergen a sus verdaderos valores a velocidad exponencial, mientras que los demás CCC exhiben una convergencia sustancialmente más lenta. Esta propiedad, conocida como “superconsistencia”, permite distinguir ambos tipos de correlaciones.

La superconsistencia de los CCC fue inicialmente aprovechada por Poskitt (2000) para el análisis de raíces unitarias y relaciones de cointegración. Posteriormente, Bauer & Wagner (2002), en adelante BW, refinan este planteamiento ligándolo con los métodos de subespacios. Concretamente, estos autores proponen dos procedimientos para estimar el rango de cointegración basados en la detección de las raíces unitarias de un proceso: el denominado *eigenvalue test sequence*, que presenta en muestras cortas una fuerte dependencia de la correcta estimación del orden del sistema en un paso previo, y el llamado *combined test sequence* cuyo comportamiento, aunque más satisfactorio, está fuertemente vinculado a una hipótesis nula inicial. De hecho BW sugieren en las conclusiones de su trabajo que debería refinarse el criterio que determina esta hipótesis inicial. Efectivamente:

1. El criterio de información que estos autores proponen para obtener la condición inicial no tiene en cuenta la dimensión del subespacio de información pasada que se ha considerado para calcular las matrices que definen el problema de regresión. Nuestros resultados indican que la consideración de este coeficiente mejora la capacidad discriminante del criterio.
2. Estos autores describen la forma matemática de la función de penalización de su criterio de información como “un tanto heurística”. Basándonos en los trabajos de Hurvich et al. (1990) o Bengtsson & Cavanaugh (2006) proponemos un ajuste de esta función mediante métodos de simulación.

En consecuencia, el objetivo de nuestro trabajo es obtener un criterio de información:

1. cuyo uso conduzca a una estimación consistente del número de raíces unitarias,
2. que presente resultados aceptables en tamaño y potencia en muestras cortas, y
3. que pueda utilizarse como hipótesis nula para el uso del método *combined test sequence* o como estimador preliminar del número de raíces unitarias en el análisis masivo de procesos univariantes o multivariantes.

La estructura del artículo es la siguiente. La sección 2 introduce los métodos de subespacios y los liga con el cálculo de los CCC que se usarán posteriormente. En la sección 3 se presenta la metodología propuesta y se analizan sus propiedades en muestras finitas mediante técnicas de simulación. La sección 4 ilustra la aplicación práctica del método mediante el análisis de varias series reales. Finalmente, la sección 5 resume las principales conclusiones.

## 2. Introducción a los métodos de subespacios

Consideramos un sistema lineal de coeficientes fijos, cuya dinámica puede describirse mediante el siguiente modelo en espacio de los estados en forma de innovaciones:

$$x_{t+1} = \Phi x_t + E\psi_t \quad (1)$$

$$z_t = Hx_t + \psi_t \quad (2)$$

donde  $x_t \in \mathbb{R}^n$  es el vector de estado,  $z_t \in \mathbb{R}^m$  es el vector de *outputs* observables y  $\psi_t \in \mathbb{R}^m$  es una secuencia de perturbaciones  $iidN(0, Q)$ , donde la matriz  $Q$  es semidefinida positiva.  $\Phi \in \mathbb{R}^{n \times n}$ ,  $E \in \mathbb{R}^{n \times m}$  y  $H \in \mathbb{R}^{m \times n}$  son las matrices paramétricas del sistema en el que se omite, sin pérdida de generalidad, un vector de variables exógenas. Esta formulación tiene la ventaja de que numerosos modelos de series temporales, incluyendo estructuras VAR y VARMA, pueden representarse directamente en esta forma (Aoki 1990, Terceiro 1990). Además, cualquier modelo en una forma de espacio de los estados más general, por ejemplo, con una matriz que multiplica el vector de perturbaciones en la ecuación de observación o con distintas fuentes de error (una en cada ecuación), puede, bajo condiciones no muy restrictivas, transformarse en una forma de innovaciones equivalente (véase Casals et al. 1999).

Los métodos de subespacios parten de una representación matricial de las ecuaciones de estado (1) y de observación (2). Así, (1) puede desarrollarse recurrentemente como

$$x_t = \Phi^t x_0 + \sum_{i=0}^{t-1} \Phi^i E \psi_{t-i-1} \quad (3)$$

y sustituyendo a su vez (3) en la ecuación de observación (2) se obtiene

$$z_t = H\Phi^t x_0 + H \sum_{i=0}^{t-1} \Phi^i E \psi_{t-i-1} + \psi_t \quad (4)$$

que es una representación del vector de variables endógenas,  $z_t$ , que depende exclusivamente del vector de estado inicial,  $x_0$ , y de la historia presente y pasada de las innovaciones,  $\psi_t$ . La ecuación (4) puede escribirse en forma matricial como

$$Z_k = O_k X_k + V_k \Psi_k \quad (5)$$

donde los datos se organizan en  $Z_k$ , conocida como matriz *Output Block Hankel* (Van Overschee & De Moor 1994). Además, el subíndice  $k$  es la dimensión del subespacio de información o espacio de filas de  $Z_k$ . Los elementos de la ecuación (5) se estructuran de la siguiente forma:

1. Matrices *Output* e *Input Block Hankel*:

$$\begin{bmatrix} Z_p \\ Z_f \end{bmatrix} = \begin{bmatrix} Z_{0:i-1} \\ Z_{i:2i-1} \end{bmatrix} = \begin{bmatrix} z_0 & z_1 & \dots & z_{T-2i} \\ z_1 & z_2 & \dots & z_{T-2i+1} \\ \vdots & \vdots & \ddots & \vdots \\ z_{i-1} & z_i & \dots & z_{T-i-1} \\ z_i & z_{i+1} & \dots & z_{T-i} \\ z_{i+1} & z_{i+2} & \dots & z_{T-i+1} \\ \vdots & \vdots & \ddots & \vdots \\ z_{2i-1} & z_{2i} & \dots & z_{T-1} \end{bmatrix} \quad (6)$$

donde  $z_0$  es un vector de  $m$  variables en el instante 0 y  $k = p, f$ . La matriz  $\Psi_k$  tiene la misma estructura que  $Z_k$ , pero agrupando las secuencias de las perturbaciones  $\psi_t$ . Los bloques de información, pasada y futura, de la matriz *Output Block Hankel* se corresponden con las dos particiones de la matriz (6). La elección de la dimensión de los subespacios de información pasada y futura,  $p$  y  $f$  respectivamente, se estudia, entre otros, en Viberg (1995), Chui (1997) o Bauer (1998), siendo habitual tomar  $p = f = i$ . En los algoritmos utilizados en este trabajo (véase García-Hiernaux 2005) se define  $i = \max(4, h_t)$ , siendo  $h_t$  el entero más cercano al logaritmo del tamaño muestral, que se basa en una regla heurística fundamentada en la experiencia. Casals (1997) discute con mayor profundidad los posibles valores de  $i$ .

## 2. Matriz de observabilidad:

$$O_i = \begin{bmatrix} H \\ H\Phi \\ H\Phi^2 \\ \vdots \\ H\Phi^{i-1} \end{bmatrix} \in \mathbb{R}^{im \times n} \quad (7)$$

## 3. Secuencia de estados:

$$X_t = (x_t \quad x_{t+1} \quad x_{t+2} \quad \dots \quad x_{t+T-2i})$$

para  $t = 0, 1, \dots, T-1$ .

A partir de esta definición, las secuencias de estados pasados y futuros, que se inician en  $t = 0$  y  $t = i$  respectivamente, pueden expresarse como  $X_p = X_0$  y  $X_f = X_i$ .

4. Matriz *Toeplitz* triangular inferior por bloques:

$$V_i = \begin{bmatrix} I_m & 0 & 0 & \dots & 0 \\ HE & I_m & 0 & \dots & 0 \\ H\Phi E & HE & I_m & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ H\Phi^{i-2}E & H\Phi^{i-3}E & H\Phi^{i-4}E & \dots & I_m \end{bmatrix} \in \mathbb{R}^{im \times im} \quad (8)$$

Debido a la linealidad del sistema, la secuencia de estados futuros puede escribirse como  $X_f = AZ_p$ , donde  $A$  es una matriz de rango  $n$ . Por tanto, seleccionando adecuadamente los subíndices temporales en (5) y sustituyendo se obtiene:

$$Z_f = O_i AZ_p + V_i \Psi_f \quad (9)$$

Los algoritmos de subespacios resuelven los modelos de regresión con restricción de rango dados en (9), estimando las matrices  $O_i$ ,  $A$  y  $V_i$ , para después recuperar las matrices paramétricas del sistema formado por (1) y (2). Para ello, usan la descomposición en valores singulares (SVD) (Eckart & Young 1936) de  $W_1 Z_f W_2$ , donde  $W_1$  y  $W_2$  son dos matrices de ponderación. Cuando  $W_1 = (Z_f Z_f')^{-\frac{1}{2}}$  y  $W_2 = Z_p' (Z_p Z_p')^{-1} Z_p$ , los valores singulares (VS) resultantes de la SVD coinciden con los CCC entre el subespacio de información futura,  $Z_f$ , y el de información pasada,  $Z_p$ .

### 3. Detección de raíces unitarias

El criterio planteado por BW, en lugar de usar los CCC entre  $z_t$  y  $z_{t-1}$  como sugiere Poskitt (2000), utiliza los CCC entre los bloques de información pasada  $Z_p$  y de información futura  $Z_f$ . Estos autores evalúan una transformación de los  $j$  primeros VS calculados,  $\hat{\sigma}_j$ , comparándola con una función de penalización  $C(T)$  que depende del tamaño muestral:

$$1 - \hat{\sigma}_j^2 - C(T) < 0, \quad \forall j \leq i \quad (10)$$

El VS que cumple esta desigualdad puede considerarse igual a la unidad, correspondiendo a una raíz unitaria del sistema. La función  $C(T)$  representa una cota superior de forma que, si  $1 - \hat{\sigma}_j^2$  se sitúa por debajo, entonces el criterio devolverá al menos  $j$  raíces unitarias, ya que los VS cumplen  $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \dots \geq \hat{\sigma}_i \geq 0$ . Cabe destacar que el análisis de correlaciones canónicas no puede detectar directamente series  $I(2)$ , pues en estos  $\sigma_1 = 1$  pero  $\sigma_2 < 1$ . No obstante, esto no causa un problema en la práctica, puesto que al diferenciar la serie y analizar los CCC de la variable transformada se obtendrá la segunda raíz igual a la unidad. Aún con esto, este trabajo se centra únicamente en procesos formados por series  $I(0)$  o  $I(1)$ . La función de pérdida que especifican BW para la ecuación (10) es

$$C(T) = \frac{(\log T)^2}{T} \quad (11)$$

que supone, en la práctica, un decaimiento de la penalización a medida que la muestra crece.

Este criterio plantea dos problemas fundamentales: a) su uso genera estimaciones consistentes del número de raíces unitarias pero el rendimiento en muestras cortas no es muy satisfactorio, y b) no depende de la dimensión del subespacio de información pasada,  $i$ . La primera cuestión es muy relevante, ya que multitud de series económicas, por ejemplo distintas series anuales de contabilidad nacional,

no disponen de una muestra suficientemente larga. La segunda cuestión omite que cambios en la dimensión del subespacio de información pasada pueden producir variaciones importantes en los valores calculados de los CCC. Básicamente, cuanto mayor es la dimensión de los subespacios de información, más elevados son los CCC, al igual que un coeficiente de determinación crece cuando aumenta la cantidad de variables explicativas.

Para resolver estas limitaciones, planteamos un criterio de decisión que: a) está basado tanto en la idea de BW como en un estudio de simulación y b) es aplicable a sistemas con múltiples raíces unitarias. La idea clave es obtener una penalización que refleje el límite entre los  $\hat{\sigma}_j$  que se consideren iguales a la unidad y los inferiores, a partir de la propiedad de convergencia a tasas diferentes a su verdadero valor. Para ello, se definen dos condiciones que debe cumplir esta función.

**Condición 3.1:** Debe depender del número de observaciones  $T$ , de la dimensión del subespacio de información pasada  $i$  y del número de raíces unitarias que se desea evaluar  $\bar{d}$ .

**Condición 3.2:** Debe cumplir  $\lim_{T \rightarrow \infty} G_l(T, i, \bar{d}) = 0$ , donde  $G_l$  representa las posibles funciones de pérdida, para asegurar la consistencia del criterio (ver proposición 1 más adelante).

La forma más general de la familia de criterios propuestos puede formularse como:

$$f(\hat{\sigma}_j) - G_l(T, i, \bar{d}) \leq 0 \quad (12)$$

donde  $f(\hat{\sigma}_j)$  es una transformación que mide la distancia entre el  $j$ -ésimo VS y la unidad. Esta distancia debe converger a cero para VS correspondientes a raíces unitarias; así, en este trabajo se usan las expresiones  $1 - \hat{\sigma}_j^2$  y  $1 - \hat{\sigma}_j$ , aunque existen multitud de  $f(\hat{\sigma}_j)$  válidas.

**Proposición 1.** *Bajo estas condiciones y siendo  $d$  el número de series  $I(1)$  que forman un proceso, entonces se cumple que:*

$$\lim_{T \rightarrow \infty} P[f(\hat{\sigma}_{k,T}) - G_l(T, i, k) \leq \epsilon] = 1, \quad \forall \epsilon \geq 0 \text{ y } \forall k \leq d$$

y por tanto el uso del criterio formulado conduce a estimaciones consistentes del número de raíces unitarias del proceso, i.e.  $\hat{d} \xrightarrow{P} d$  ( $\hat{d}$  converge en probabilidad a  $d$ ).

**Demostración.** Por simplicidad, se obvia el subíndice temporal en los VS definiendo  $\hat{\sigma}_{j,T} = \hat{\sigma}_j$ . Dado que  $\hat{\sigma}_j \xrightarrow{P} \sigma_j$ , se tiene que  $\hat{\sigma}_k \xrightarrow{P} 1, \forall k \leq d$ , como además  $f(\cdot)$  es continua y por tanto se cumplen las condiciones del teorema de Slutsky, aplicando este teorema se obtiene  $f(\hat{\sigma}_k) \xrightarrow{P} f(1) = 0$ . Consecuentemente, cuando  $T \rightarrow \infty, P[|f(\hat{\sigma}_k) - f(1)| \leq \delta] = P[|f(\hat{\sigma}_k)| \leq \delta] \rightarrow 1, \forall \delta \geq 0$ .

Partiendo del criterio propuesto se tiene que:  $\forall \epsilon \geq 0 \text{ y } \forall k \leq d$ ,

$$\begin{aligned} P[f(\hat{\sigma}_k) - G_l(T, i, k) \leq \epsilon] &= P[f(\hat{\sigma}_k) \leq \epsilon + G_l(T, i, k)] \\ &\geq P[|f(\hat{\sigma}_k)| \leq \epsilon + G_l(T, i, k)] \end{aligned} \quad (13)$$

tomando límites cuando  $T \rightarrow \infty$  y dado que  $\lim_{T \rightarrow \infty} G_l(T, i, k) = 0$ , por la condición 3.2, se obtiene que:

$$\lim_{T \rightarrow \infty} P \left[ |f(\hat{\sigma}_k)| \leq \epsilon + \frac{\epsilon}{2} \right] = \lim_{T \rightarrow \infty} P[|f(\hat{\sigma}_k)| \leq \delta] = 1, \quad \forall \delta \geq 0 \quad (14)$$

si se define  $\delta = \epsilon + \frac{\epsilon}{2}$ .  $\square$

### 3.1. Procesos univariantes

Para dar una forma funcional concreta a  $G_l(T, i, \bar{d})$  que, además de estimar consistentemente  $\hat{d}$ , también genere buenos resultados, en términos de tamaño y potencia, en muestras cortas, partimos del trabajo de Bengtsson & Cavanaugh (2006), quienes proponen funciones de penalización estimadas vía Monte Carlo para mejorar el rendimiento del criterio AIC cuando se dispone de pocos datos. Con un objetivo similar, estudiamos la distribución empírica de los  $\hat{\sigma}_1$  correspondientes a raíces no unitarias. El modo de estimar las funciones de penalización es el siguiente:

1. Se escoge un CCC generador de las simulaciones,  $\bar{\sigma}$ . Este debe ser inferior a la unidad, aunque lo suficientemente cercano para reflejar adecuadamente el límite del comportamiento de los CCC que no corresponden a raíces unitarias (por ejemplo,  $\bar{\sigma} = 0.975$ ).
2. Se realizan 1000 replicaciones de distintos modelos caracterizados por  $\bar{\sigma}$ , desde  $T = 20, 21, \dots, 500$ , obteniéndose para cada instante temporal y cada repetición una secuencia de CCC:  $\hat{\sigma}_1 \geq \hat{\sigma}_2 \geq \dots \geq \hat{\sigma}_i \geq 0$ .
3. Se aplica  $f(\cdot)$  a las 1000 series que representan la evolución temporal de  $\hat{\sigma}_1$ , consiguiendo una distribución empírica de  $f(\hat{\sigma}_1)$  para cada  $T$ .
4. Se selecciona un determinado percentil  $\tau$  que resume la información de esta distribución empírica en una sola serie que denominamos  $y_{1,t}$ .
5. Por último, se especifica y estima un modelo en función de  $T, i$  y  $\bar{d}$ , que se ajuste correctamente a  $y_{1,t}$  y que cumpla las condiciones (3.1) y (3.2).

Es importante destacar que los modelos con los que se simulan los procesos no son relevantes aquí, puesto que muchos modelos distintos implican un mismo  $\sigma_1$ . Lo realmente importante es el valor asignado a  $\bar{\sigma}$ . La principal desventaja de esta metodología es que estos valores son elegidos de una forma un tanto heurística. La tabla 1 muestra algunos de ellos para el caso univariante.

En este caso, donde se evalúa si  $d > 0$ , se sugiere la siguiente forma funcional para la penalización:

$$G_l(T, i, 0) = \hat{\alpha}_{1,0} T^{\hat{\beta}_{1,0}} i^{\hat{\beta}_{2,0}} \quad (15)$$

ya que: a) la ecuación (15) cumple las condiciones (3.1) y (3.2) para las ternas  $(\hat{\alpha}_{1,0}, \hat{\beta}_{1,0}, \hat{\beta}_{2,0})$  y b) las estimaciones de los modelos del tipo  $\log y_{1,t} = \alpha_{1,0} + \beta_{1,0} \log T + \beta_{2,0} \log i + \varepsilon_t$ , donde  $\varepsilon_t$  corresponde a un término de error e  $i$  hace

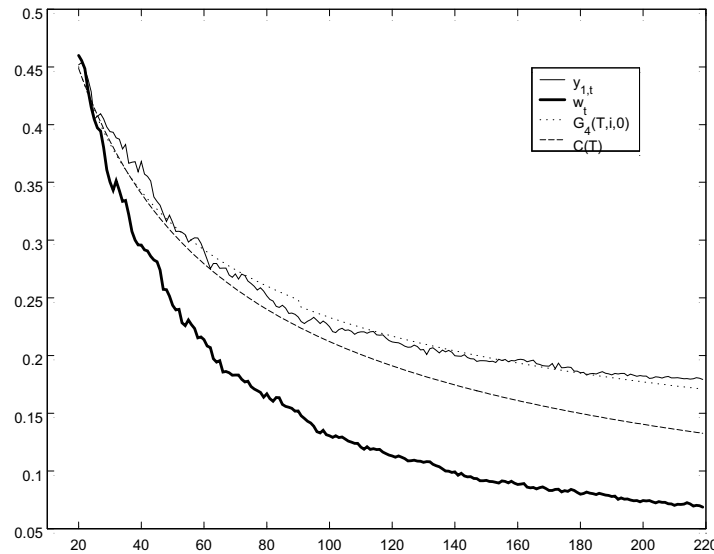


FIGURA 1: Valores singulares (VS) simulados y funciones de penalización (a).  $y_{1,t}$  corresponde al primer VS de un proceso estacionario,  $w_t$  corresponde al primer VS de un proceso no estacionario.  $G_4(T, i, 0)$  es una función de penalización propuesta para evaluar si  $d > 0$  y  $C(T)$  es la función de penalización utilizada por Bauer y Wagner (2002).  $y_{1,t}$  y  $w_t$  reflejan sensiblemente la diferencia entre las tasas de convergencia de los VS a sus valores asintóticos.

referencia a la dimensión del subespacio de información, devuelven un coeficiente de determinación muy aceptable, superior al 90%.

La ecuación (15) supone una penalización distinta para cada terna de parámetros estimados  $(\hat{\alpha}_{1,0}, \hat{\beta}_{1,0}, \hat{\beta}_{2,0})$ , que varían dependiendo del valor del CCC generador de las simulaciones ( $\bar{\sigma}$ ) y del percentil elegido ( $\tau$ ), indispensables para definir la variable  $y_{1,t}$ . Por ello, se realizan múltiples simulaciones alterando estos dos valores. La tabla 1 muestra las ternas de parámetros estimados para (15), para diferentes funciones  $f(\cdot)$  y valores de  $\tau$  y  $\bar{\sigma}$ . Además, incluye la función de penalización propuesta por BW. Como cabía esperar, los resultados indican que diferencias en los valores que definen la distribución empírica conducen a distintas funciones de penalización. Estas se denotan por  $G_l$ , con  $l = 1, 2, \dots, 6$ .

Si  $y_{1,t}$  representa suficientemente bien la frontera entre los CCC iguales y los inferiores a la unidad, entonces la función de pérdida  $G_4$  (que se ajusta correctamente a  $y_{1,t}$ ) es válida para evaluar si  $d > 0$ . Como se obtienen distintas funciones de penalización  $G_l(T, i, 0)$  consistentes, se discrimina en función de sus propiedades en muestras finitas. Para ello, se define tamaño y potencia del criterio, respectivamente, como la probabilidad de rechazar un número concreto de raíces unitarias, cuando existen y la probabilidad de rechazarlo cuando no existen. De este modo, se eligen las funciones de penalización que cumplen alguna de las dos condiciones siguientes:

TABLA 1: Funciones de penalización estimadas para contrastar  $d > 0$ .

Distribución empírica			Valores estimados			Función de penalización
$f(\hat{\sigma}_1) = y_{1,t}$	$\bar{\sigma}$	$\tau$	$\hat{\alpha}_{1,0}$	$\hat{\beta}_{1,0}$	$\hat{\beta}_{2,0}$	
$1 - \hat{\sigma}_1^2$	0.975	95	1.09	-0.51	-0.17	$G_1(T, i, 0)$
$1 - \hat{\sigma}_1^2$	0.975	90	0.60	-0.50	-0.10	$G_2(T, i, 0)$
$1 - \hat{\sigma}_1^2$	0.950	80	0.10	-0.36	-0.10	$G_3(T, i, 0)$
$1 - \hat{\sigma}_1^2$	0.950	90	0.52	-0.40	-0.10	$G_4(T, i, 0)$
$1 - \hat{\sigma}_1$	0.950	90	0.10	-0.44	-0.05	$G_5(T, i, 0)$
$1 - \hat{\sigma}_1$	0.975	95	0.65	-0.56	-0.16	$G_6(T, i, 0)$
$(\log T)^2/T^*$	-	-	0.98	-0.52	-0.10	$C(T)$

\* Corresponde a la función de penalización de Bauer & Wagner (2002).

**Condición 3.3:** Presentar el mínimo tamaño.

**Condición 3.4:** Minimizar el tamaño, sujeto a una potencia mínima para un número de observaciones específico.

Para el caso univariante, en el que se trata de decidir si  $d > 0$ , exigimos que la potencia mínima en la segunda condición sea igual al 50% para  $T = 50$ . Para seleccionar los criterios definitivos se usa, como una aproximación al tamaño del criterio en un punto, la probabilidad de rechazar la existencia de una raíz unitaria en un paseo aleatorio,  $\nabla z_t = a_t$ . Asimismo, se aproxima la función potencia en un punto como la probabilidad de que  $\hat{d} = d$  en un modelo estacionario autorregresivo persistente del tipo  $(1 - 0.9B)z_t = a_t$ . En ambos casos  $a_t \sim iidN(0, 1)$ . Los resultados de las 1000 repeticiones del ejercicio de simulación se presentan en las tablas 2 y 3.

TABLA 2: Tamaño de los criterios para  $d > 0$ , a partir del proceso  $\phi(B)z_t = a_t$ .

T	ADF <sup>†</sup>	$C^\ddagger$	$G_1$	$G_2$	$G_3$	$G_4$	$G_5$	$G_6$
50	0.050	0.047	0.040	0.121	0.112	0.047	0.030	0.044
100	0.049	0.016	0.014	0.065	0.042	0.011	0.005	0.016
500	0.061	0.004	0.003	0.004	0.003	0.000	0.000	0.003

<sup>†</sup> Test aumentado de Dickey & Fuller (1979) al 95% con un retardo.

<sup>‡</sup> Corresponde a la función de penalización de Bauer & Wagner (2002).

$\phi(B) = 1 - B$ .

TABLA 3: Potencia de los criterios para  $d > 0$ , a partir del proceso  $\phi(B)z_t = a_t$ .

T	ADF <sup>†</sup>	$C^\ddagger$	$G_1$	$G_2$	$G_3$	$G_4$	$G_5$	$G_6$
50	0.114	0.246	0.208	0.505	0.464	0.246	0.188	0.230
100	0.341	0.451	0.420	0.762	0.628	0.376	0.290	0.451
500	1.000	1.000	1.000	1.000	1.000	0.994	0.981	1.000

<sup>†</sup> Test aumentado de Dickey & Fuller (1979) al 95% con un retardo.

<sup>‡</sup> Corresponde a la función de penalización de Bauer & Wagner (2002).

$\phi(B) = 1 - 0.9B$ .

Las tablas 2 y 3 confirman las propiedades en muestras finitas de todos los criterios presentados. Como puede verse, varios criterios batan al test ADF en

potencia y tamaño. A la vista de los resultados, se elige como función de penalización que cumple la condición (3.3) a  $G_5$ , que pasa a denominarse  $G_a(T, i, 0)$ . Además, se selecciona la función de pérdida  $G_2$ , en adelante  $G_b(T, i, 0)$ , que cumple la condición (3.4).

### 3.2. Procesos multivariantes

La principal ventaja de la metodología propuesta se encuentra en el tratamiento de modelos multivariantes. En primer lugar, tan solo unos pocos autores como Phillips & Durlauf (1986), Abuaf & Jorion (1990), Flôres et al. (1999) o BW tratan la identificación de raíces unitarias en vectores de series. En segundo lugar, este paso abre las puertas al análisis de cointegración.

Como se ha visto, las funciones de penalización estimadas no solo dependen del tamaño muestral y de la dimensión del subespacio de información pasada, sino también del mínimo número de raíces unitarias que se desee evaluar. Esto implica que cada  $\hat{\sigma}_j$  candidato a ser igual a la unidad, debe compararse con una función de penalización diferente. El criterio desarrollado para procesos univariantes estima si la serie analizada es estacionaria o si, por el contrario,  $d > 0$ . En este apartado se desarrollan los criterios para decidir secuencialmente si  $d > 1$ ,  $d > 2$ ,  $d > 3$  y  $d > 4$ . Para estimar las funciones de penalización se actúa como en el apartado anterior, pero añadiendo una nueva serie no estacionaria en cada caso.

La figura 2 compara las distribuciones empíricas  $y_{1,t}$  y  $y_{2,t}$  correspondientes a los pares  $(\tau = 90, \bar{\sigma} = 1.0)$  y  $(\tau = 90, \bar{\sigma} = 0.95)$ , de un sistema bivalente. Obsérvese que el primer VS, que corresponde a una raíz unitaria del proceso bivalente, se mantiene muy por debajo de  $G_b(T, i, 0)$ , penalización utilizada en el criterio con mayor potencia de los elegidos para evaluar si  $d > 0$ . En efecto, si comparamos un proceso multivariante cualquiera ( $A_t$ ) con  $d$  raíces unitarias, con un mismo sistema ( $B_t$ ) que incorpora, además, una serie no estacionaria (el proceso  $B_t$  tendría  $d + 1$  raíces iguales a la unidad), entonces  $\hat{\sigma}_j(B_t) \geq \hat{\sigma}_j(A_t)$  para  $j = 1, \dots, d + 1$ . Por tanto, aprovechando la ordenación de los VS, es posible usar secuencialmente los criterios expuestos, para decidir cuántos de ellos son iguales a la unidad.

La tabla 4 presenta distintas ternas de parámetros estimados para funciones de pérdida del tipo  $G_l(T, i, 1) = \hat{\alpha}_{1,1} T^{\hat{\beta}_{1,1}} i^{\hat{\beta}_{2,1}}$ . Puede observarse en la figura 2 que el modelo especificado para representar las distribuciones empíricas que se usan para contrastar  $d > 0$  también se ajusta correctamente a las que se utilizan para decidir si  $d > 1$ , salvo, quizá, en las primeras observaciones.

TABLA 4: Funciones de penalización estimadas para  $d > 1$ .

Distribución empírica			Valores estimados			Función de penalización
$f(\hat{\sigma}_2) = y_{2,t}$	$\bar{\sigma}$	$\tau$	$\hat{\alpha}_{1,1}$	$\hat{\beta}_{1,1}$	$\hat{\beta}_{2,1}$	
$1 - \hat{\sigma}_2^2$	0.95	75	0.43	-0.39	-0.07	$G_1(T, i, 1)$
$1 - \hat{\sigma}_2^2$	0.95	80	0.50	-0.39	-0.07	$G_2(T, i, 1)$
$1 - \hat{\sigma}_2^2$	0.95	85	0.57	-0.39	-0.07	$G_3(T, i, 1)$
$1 - \hat{\sigma}_2^2$	0.95	90	0.67	-0.39	-0.06	$G_4(T, i, 1)$

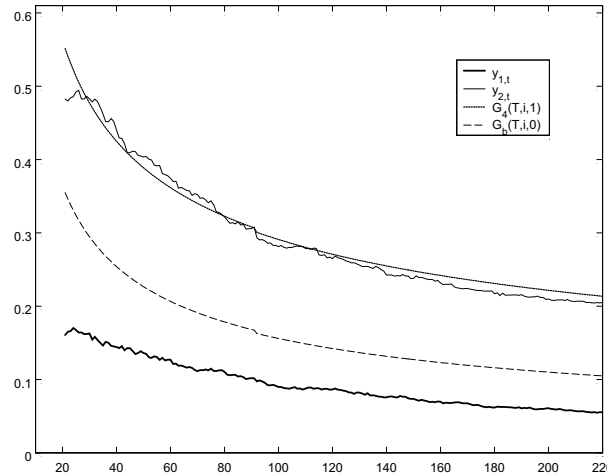


FIGURA 2: Valores singulares (VS) simulados y funciones de penalización (b).  $y_{1,t}$  corresponde al primer VS de un sistema bivalente no estacionario (cuyos procesos univariantes son  $I(1)$  e  $I(0)$ ).  $y_{2,t}$  corresponde al segundo VS del mismo sistema,  $G_4(T, i, 1)$  es una de las funciones de penalización propuestas para evaluar si  $d > 1$  y  $G_b(T, i, 0)$  es la función de penalización usada para evaluar si  $d > 0$ .

Para discriminar entre las posibles funciones de penalización propuestas, se utilizan las condiciones (3.3) y (3.4), establecidas anteriormente. Los resultados de las simulaciones realizadas para evaluar el tamaño y la potencia de los criterios formulados para  $d > 1$  se exponen en las tablas 5 y 6. En ellos se refleja el tamaño y la potencia del criterio definidos anteriormente, utilizando la función de penalización  $G_b(T, i, 0)$  para decidir si  $d > 0$ .

TABLA 5: Tamaño de los criterios para  $d > 1$ , a partir del proceso  $\Phi(B)z_t = a_t$ .

T	$G_1$	$G_2$	$G_3$	$G_4$
50	0.292	0.221	0.163	0.099
100	0.146	0.109	0.076	0.047
500	0.003	0.002	0.001	0.001

$\Phi(B) = \text{diag}(\nabla, \nabla)$

TABLA 6: Potencia de los criterios para  $d > 1$ , a partir del proceso  $\Phi(B)z_t = a_t$ .

T	$G_1$	$G_2$	$G_3$	$G_4$
50	0.452	0.373	0.293	0.204
100	0.516	0.428	0.337	0.239
500	0.993	0.989	0.968	0.947

$\Phi(B) = \text{diag}(\nabla, (1 - .9B))$

Para el análisis de procesos bivariantes, se escogen las funciones de penalización  $G_4$  y  $G_1$  de la tabla 4. Estas se denominan, de aquí en adelante,  $G_a(T, i, 1)$  y  $G_b(T, i, 1)$ , respectivamente. La primera minimiza el tamaño y la segunda lo hace

dada un potencia mínima del 45% para  $T = 50$ . Se decide disminuir la potencia mínima exigida con respecto de  $d > 0$ , al advertir que las propiedades en muestras finitas de los criterios se degradan ligeramente cuando aumenta  $d$ .

El método para obtener las funciones de pérdida para procesos trivariantes, tetravariantes y pentavariantes es análogo al expuesto hasta ahora. Sin embargo, el modelo especificado para representarlas adecuadamente es distinto. De hecho, para mejorar el ajuste a las distribuciones empíricas decidimos modelizar dos submuestras, proponiendo una ecuación específica para cada una de ellas. Los modelos ajustados son los siguientes:

$$\begin{aligned} T < T_{\bar{d}}^* & \quad y_{\bar{d}+1,t} = \alpha_{2,\bar{d}} + \delta_{1,\bar{d}}T + \delta_{2,\bar{d}}T^2 + \delta_{3,\bar{d}}T^3 + \delta_{4,\bar{d}}i + \varepsilon'_{\bar{d},t} \\ T \geq T_{\bar{d}}^* & \quad \log y_{\bar{d}+1,t} = \alpha_{1,\bar{d}} + \beta_{1,\bar{d}}\log T + \beta_{2,\bar{d}}\log i + \varepsilon''_{\bar{d},t} \end{aligned} \quad (16)$$

donde  $T_{\bar{d}}^*$  se refiere a la primera realización de la segunda submuestra y  $\bar{d} = 2, 3, 4$ .

Las estimaciones de las dos funciones de penalización que cumplen las condiciones (3.3) y (3.4) y los parámetros estimados de los modelos propuestos en (16), se ofrecen en la tabla 7. De nuevo, se rebaja la potencia mínima exigida para  $T = 50$ , que hace referencia a la condición (3.4): 40% para  $d > 2, 3, 4$ . Obsérvese en esta tabla cómo aumentar la dimensión del subespacio de información pasada  $i$  afecta positivamente a  $\hat{\sigma}_j$  y, por tanto, negativamente a  $f(\hat{\sigma}_j) \forall j$ .

Los procesos generadores y las simulaciones realizadas para valorar el tamaño y la potencia de los criterios que usan las funciones de penalización de la tabla 7 se presentan en las tablas 8 y 9. Los resultados revelan un deterioro en las propiedades de los criterios cuando aumenta el número de raíces unitarias. En este punto se escoge no estimar funciones de pérdida para decidir si  $d > 5$ , por varias razones: a) hay pocos modelos para más de cinco series temporales, b) el uso de estos criterios para procesos con una dimensión mayor pero con un número, a priori, no superior a cinco raíces unitarias es también válido y c) el deterioro que muestran los criterios con el aumento de  $d$  cuando aumenta la dimensión del sistema no motiva la búsqueda de funciones de penalización para procesos con un mayor número de raíces iguales a la unidad, aunque esto es perfectamente factible.

En suma, la función de penalización que se propone tiene la forma siguiente:

$$G_l(T, i, \bar{d}) = \begin{cases} \text{si } \bar{d} = 0, 1 & \hat{\alpha}_{1,\bar{d}}T^{\hat{\beta}_{1,\bar{d}}}i^{\hat{\beta}_{2,\bar{d}}}, \quad \forall T, i \\ \text{si } \bar{d} = 2, 3, 4 & \begin{cases} \hat{\alpha}_{2,\bar{d}} + \hat{\delta}_{1,\bar{d}}T + \hat{\delta}_{2,\bar{d}}T^2 + \hat{\delta}_{3,\bar{d}}T^3 + \hat{\delta}_{4,\bar{d}}i, & T < T_{\bar{d}}^* \\ \hat{\alpha}_{1,\bar{d}}T^{\hat{\beta}_{1,\bar{d}}}i^{\hat{\beta}_{2,\bar{d}}}, & T \geq T_{\bar{d}}^* \end{cases} \end{cases}$$

donde  $l = a$ , si se está interesado únicamente en minimizar el tamaño del criterio o  $l = b$ , si, además, se exige una potencia mínima. Los valores de todos los parámetros para  $\bar{d} = 0, 1, \dots, 4$  se encuentran en las tablas 1, 4 y 7.

TABLA 7: Funciones de penalización propuestas para  $d > 2$ ,  $d > 3$  y  $d > 4$ .

Distribución empírica				$T < T^*$					$T \geq T^*$			Función de penalización
$j$	$f(\hat{\sigma}_j)$	$\bar{\sigma}$	$\tau$	$\hat{\alpha}_{2,\bar{d}}$	$\hat{\delta}_{1,\bar{d}}$	$\hat{\delta}_{2,\bar{d}}$	$\hat{\delta}_{3,\bar{d}}$	$\hat{\delta}_{4,\bar{d}}$	$\hat{\alpha}_{1,\bar{d}}$	$\hat{\beta}_{1,\bar{d}}$	$\hat{\beta}_{2,\bar{d}}$	
3	$1 - \hat{\sigma}_j^2$	0.95	90	-0.305	0.040	$-6.5 \times 10^{-4}$	$3.3 \times 10^{-6}$	—	0.786	-0.328	-0.226	$G_a(T, i, 2)$
3	$1 - \hat{\sigma}_j^2$	0.95	60	-0.353	0.036	$-5.9 \times 10^{-4}$	$3.0 \times 10^{-6}$	—	0.188	-0.285	-0.172	$G_b(T, i, 2)$
4	$1 - \hat{\sigma}_j^2$	0.95	90	-0.635	0.044	$-6.0 \times 10^{-4}$	$2.7 \times 10^{-6}$	—	1.589	-0.437	-0.365	$G_a(T, i, 3)$
4	$1 - \hat{\sigma}_j^2$	0.95	60	-0.621	0.041	$-5.6 \times 10^{-4}$	$2.5 \times 10^{-6}$	—	1.557	-0.469	-0.418	$G_b(T, i, 3)$
5	$1 - \hat{\sigma}_j^2$	0.95	90	-0.317	0.032	$-3.2 \times 10^{-4}$	$1.1 \times 10^{-6}$	-0.076	1.313	-0.383	-0.280	$G_a(T, i, 4)$
5	$1 - \hat{\sigma}_j^2$	0.95	60	-0.366	0.030	$-3.0 \times 10^{-4}$	$1.0 \times 10^{-6}$	-0.063	1.131	-0.361	-0.378	$G_b(T, i, 4)$

$j = \bar{d} + 1$ ;

Si  $j = 3, 4$ , entonces  $T^* = 88$ ;

Si  $j = 5$ , entonces  $T^* = 121$ .

TABLA 8: Tamaño de los criterios para  $d > 2$ ,  $d > 3$  y  $d > 4$  a partir de los procesos  $\Phi_k(B)z_t = a_t$ .

T	$d > 2$		$d > 3$		$d > 4$	
	$G_a$	$G_b$	$G_a$	$G_b$	$G_a$	$G_b$
50	0.103	0.399	0.097	0.397	0.118	0.400
100	0.081	0.331	0.079	0.336	0.066	0.374
500	0.002	0.004	0.003	0.054	0.010	0.034

$\Phi_1(B) = \text{diag}(\nabla, \nabla, \nabla)$ ;  $\Phi_2(B) = \text{diag}(\nabla, \nabla, \nabla, \nabla)$ ;  
 $\Phi_3(B) = \text{diag}(\nabla, \nabla, \nabla, \nabla, \nabla)$

TABLA 9: Potencia de los criterios para  $d > 2$ ,  $d > 3$  y  $d > 4$ , a partir del proceso  $\Phi_k(B)z_t = a_t$ .

T	$d > 2$		$d > 3$		$d > 4$	
	$G_a$	$G_b$	$G_a$	$G_b$	$G_a$	$G_b$
50	0.166	0.402	0.161	0.398	0.110	0.408
100	0.854	0.487	0.741	0.460	0.144	0.438
500	0.940	0.951	0.920	0.994	0.458	0.844

$\Phi_1(B) = \text{diag}(\nabla, \nabla, (1 - 0.9B))$ ;  $\Phi_2(B) = \text{diag}(\nabla, \nabla, \nabla, (1 - 0.9B))$ ;  
 $\Phi_3(B) = \text{diag}(\nabla, \nabla, \nabla, \nabla, (1 - 0.9B))$

## 4. Ejercicios con datos reales

En esta sección se muestran los resultados obtenidos con dos conjuntos de series típicas en la literatura econométrica. Estos han sido escogidos porque en el primer caso el consenso es que no existe cointegración, mientras que en el segundo caso los estudios realizados evidencian que sí la hay.

### 4.1. Series de precios de la harina

El vector de series analizadas  $Y_t$  representa los índices mensuales de precios de la harina en las ciudades de Buffalo, Minneapolis y Kansas City entre los meses de agosto de 1972 y noviembre de 1980. Fueron presentadas por Tiao & Tsay (1989) para ilustrar su metodología de especificación y más tarde analizadas por Grubb (1992), Lütkepohl & Poskitt (1996) o Casals (1997).

En primer lugar, se analiza el orden de integración de cada una de las series por separado con las dos funciones de penalización obtenidas:  $G_a(T, i, d)$  y  $G_b(T, i, d)$ . Los resultados de la tabla 10 señalan la existencia de al menos una raíz unitaria en cada una de las series. Además, cuando las series se diferencian una vez, el VS más grande, correspondiente a la serie de precios de Minneapolis, resulta igual a 0.257, muy inferior a la unidad. Esto hace que el criterio descarte para cualquiera de las series un grado de integración de orden 2. Por tanto, la conclusión es que cada una de las series presenta una, y solo una, raíz unitaria. Este es el resultado más común de la literatura (Grubb 1992, Lütkepohl & Poskitt 1996, Casals 1997).

Una vez decidido que cada serie es  $I(1)$ , es posible estudiar la existencia de cointegración determinando el orden de integración del vector de las tres series. Los resultados de este análisis se presentan en la tabla 11. Se observa que el criterio

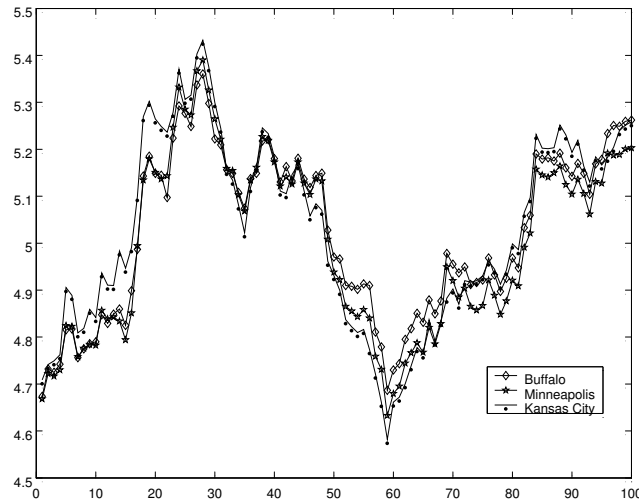


FIGURA 3: Logaritmo de los Precios de la harina en las diferentes ciudades.

TABLA 10: Estudio del orden de integración de los precios de la harina.

$j$	$\hat{\sigma}_j$	$f(\hat{\sigma}_j) - G_a(T, i, j - 1)$	$f(\hat{\sigma}_j) - G_b(T, i, j - 1)$	$\hat{d}_{G_a}$	$\hat{d}_{G_b}$
<i>Buffalo</i>					
1	0.9901	$0.0099 - 0.1344 < 0$	$0.0197 - 0.1551 < 0$	1	1
2	0.0168	$0.9997 - 0.2945 > 0$	$0.9997 - 0.2279 > 0$	1	1
<i>Minnesota</i>					
1	0.9902	$0.0098 - 0.1344 < 0$	$0.0195 - 0.1551 < 0$	1	1
2	0.0197	$0.9996 - 0.2945 > 0$	$0.9996 - 0.2279 > 0$	1	1
<i>Kansas City</i>					
1	0.9901	$0.0099 - 0.1344 < 0$	$0.0197 - 0.1551 < 0$	1	1
2	0.0219	$0.9995 - 0.2945 > 0$	$0.9995 - 0.2279 > 0$	1	1

$\hat{\sigma}_j$  se corresponde al  $j$ -ésimo valor singular.  $G_a(T, i, d)$  y  $G_b(T, i, d)$  son las funciones de pérdida propuestas.  $\hat{d}_{G_a}$  y  $\hat{d}_{G_b}$  es el número de raíces unitarias encontrado al analizar los  $j$  primeros valores singulares.

propuesto encuentra tres raíces unitarias en el sistema, independientemente de la función de pérdida especificada. Es decir, el proceso tiene tres tendencias, cada una perteneciente a una de las series que conforman el sistema, descartando la existencia de cointegración. Otros autores, como Tiao & Tsay (1989), con los llamados modelos de componentes escalares, o Lütkepohl & Poskitt (1996), con un análisis tradicional, también han rechazado la hipótesis de cointegración.

### 4.2. Series de tipos de interés de EEUU

En este caso se analiza el orden de integración de diversas series de tipos de interés de la economía de Estados Unidos entre agosto de 1985 y enero de 2003. Martín Manjón & Treadway (1997) estudian la relación existente entre estas series, encontrando que los tipos con un vencimiento inferior o igual a dos años presentan

TABLA 11: Estudio del orden de integración del proceso  $Y_t$ .

$j$	$\hat{\sigma}_j$	$f(\hat{\sigma}_j) - G_a(T, i, j - 1)$	$f(\hat{\sigma}_j) - G_b(T, i, j - 1)$	$\hat{d}_{G_a}$	$\hat{d}_{G_b}$
1	0.9912	0.0088 - 0.1344 < 0	0.0175 - 0.1551 < 0	1	1
2	0.9420	0.1126 - 0.2945 < 0	0.1126 - 0.2279 < 0	2	2
3	0.8837	0.2191 - 0.3370 < 0	0.2191 - 0.2460 < 0	3	3
4	0.5117	0.7382 - 0.3649 > 0	0.7382 - 0.2788 > 0	3	3

$\hat{\sigma}_j$  se corresponde al  $j$ -ésimo valor singular.  $G_a(T, i, d)$  y  $G_b(T, i, d)$  son las funciones de pérdida propuestas.  $\hat{d}_{G_a}$  y  $\hat{d}_{G_b}$  es el número de raíces unitarias encontrado al analizar los  $j$  primeros valores singulares.

una tendencia común a largo plazo con el tipo objetivo de la Reserva Federal (Fed). En este trabajo se analizan el tipo objetivo de la Fed, conjuntamente con el tipo de interés efectivo y los tipos a tres y seis meses del mercado secundario, proceso denotado por  $Z_t$ . El análisis de cada serie por separado concluye que todas ellas

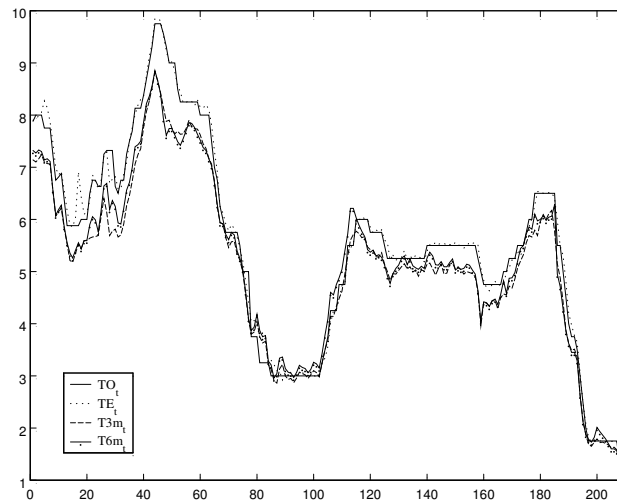


FIGURA 4: Tipos de interés a corto plazo.  $TO_t$  es el tipo objetivo,  $TE_t$  es el tipo efectivo y  $T3m_t$  y  $T6m_t$  son los tipos a 3 y 6 meses, respectivamente. La muestra analizada se extiende desde agosto de 1985 hasta enero de 2003.

son  $I(1)$ , independientemente de la función de pérdida usada. Al diferenciar las series, ninguna de ellas genera un VS suficientemente elevado como para que el criterio encuentre una segunda raíz unitaria.

La tabla 12 resume los resultados del estudio del vector de series. En este caso, el criterio revela la existencia de una única raíz unitaria. Por tanto, la tendencia que presenta cada serie es común. De este modo, se revela la existencia de tres relaciones de cointegración, una por cada raíz unitaria que desaparece en el análisis conjunto de los datos. Este es el mismo resultado que obtienen Martín Manjón & Treadway (1997) con un análisis de cointegración (en modelos bivariantes) basado en contrastes de no invertibilidad sobre la posible combinación lineal estacionaria diferenciada.

TABLA 12: Estudio del orden de integración del proceso  $Z_t$ .

$j$	$\hat{\sigma}_j$	$f(\hat{\sigma}_j) - G_a(T, i, j - 1)$	$f(\hat{\sigma}_j) - G_b(T, i, j - 1)$	$\hat{d}_{G_a}$	$\hat{d}_{G_b}$
1	0.9952	0.0048 - 0.0970 < 0	0.0095 - 0.1070 < 0	1	1
2	0.8797	0.2261 - 0.2205 > 0	0.2261 - 0.1797 > 0	1	1
3	0.7324	0.4636 - 0.2648 > 0	0.4636 - 0.1991 > 0	1	1
4	0.4028	0.8377 - 0.2642 > 0	0.8377 - 0.1968 > 0	1	1

$\hat{\sigma}_j$  se corresponde al  $j$ -ésimo valor singular.  $G_a(T, i, d)$  y  $G_b(T, i, d)$  son las funciones de pérdida propuestas.  $\hat{d}_{G_a}$  y  $\hat{d}_{G_b}$  es el número de raíces unitarias encontrado al analizar los  $j$  primeros valores singulares.

## 5. Conclusiones

En este trabajo se propone una metodología de detección de raíces unitarias basada en métodos de subespacios y en el análisis de correlaciones canónicas. El método se basa en dos nuevos criterios de información que permiten estimar consistentemente el número de raíces unitarias, tanto en el marco univariante como en el multivariante.

Las funciones de penalización subyacentes a los criterios propuestos se ajustan mediante técnicas de simulación. Esta idea sugiere de forma natural la posibilidad de adaptar el criterio de información a muestras con propiedades estadísticas especiales. Por ejemplo, se sabe que los rendimientos financieros observados en alta frecuencia suelen presentar distribuciones con colas más gruesas que las de una normal y, ocasionalmente, una asimetría apreciable. En este contexto, nuestra metodología permite plantearse el diseño de un criterio optimizado para este tipo de muestras.

El trabajo incluye ejercicios de simulación que analizan las propiedades de estos criterios en muestras finitas. Este análisis indica que los procedimientos propuestos tienen una capacidad apreciable para estimar el número de raíces unitarias y el rango de cointegración en distintas situaciones. Además, a diferencia de otros procedimientos, como ocurre en el método de Johansen (1991), la metodología presentada no precisa ajustar un modelo a los datos, con la consiguiente disminución de riesgo de error.

## Agradecimientos

Los autores agradecen la financiación mediante un proyecto del Ministerio de Educación y Ciencia Español (ref. SEJ2005-07388).

*Recibido: octubre de 2006*

*Aceptado: abril de 2007*

## Referencias

- Abuaf, N. & Jorion, P. (1990), 'Purchasing Power Parity in the Long Run', *The Journal of Finance* **45**, 157–174.
- Akaike, H. (1976), *Canonical Correlation Analysis of Time Series and the Use of an Information Criterion*, Academic Press.
- Aoki, M. (1990), *State Space Modelling of Time Series*, Springer Verlag, New York.
- Bauer, D. (1998), Some Asymptotic Theory for the Estimation of Linear Systems Using Maximum Likelihood Methods or Subspace Algorithms, Ph. D. Tesis, TU Wien.
- Bauer, D. & Ljung, L. (2002), 'Some Facts About the Choice of the Weighting Matrices in Larimore Type of Subspace Algorithms', *Automatica* **38**, 763–773.
- Bauer, D. & Wagner, M. (2002), 'Estimating Cointegrated Systems Using Subspace Algorithms', *Journal of Econometrics* **111**, 47–84.
- Bengtsson, T. & Cavanaugh, J. (2006), 'An Improved Akaike Information Criterion for State-Space Model Selection', *Computational Statistics and Data Analysis* **50**(10), 2635–2654.
- Casals, J. (1997), Métodos de subespacios en econometría, Tesis doctoral, Universidad Complutense de Madrid.
- Casals, J., Sotoca, S. & Jerez, M. (1999), 'A Fast Stable Method to Compute the Likelihood of Time Invariant State Space Models', *Economics Letters* **65**(3), 329–337.
- Chui, N. L. C. (1997), Subspace Methods and Informative Experiments for System Identification, Ph. D. Tesis, Pembroke College Cambridge.
- Dickey, D. A. & Fuller, W. A. (1979), 'Distribution of the Estimators for Autoregressive Time Series with Unit Root', *Journal of the American Statistical Association* **74**, 427–431.
- Eckart, C. & Young, G. (1936), 'The Approximation of One Matrix by Another of Lower Rank', *Psychometrika* **1**, 211–218.
- Favoreel, W., De Moor, B. & Van Overschee, P. (2000), 'Subspace State Space System Identification for Industrial Processes', *Journal of Process Control* **10**, 149–155.
- Flôres, R. G., Jorion, P., Preumont, P. Y. & Szafarz, A. (1999), 'Multivariate Unit Roots Test of PPP Hypothesis', *Journal of Empirical Finance* **6**, 335–353.
- García-Hiernaux, A. (2005), Identificación de modelos para series temporales mediante métodos de subespacios, Tesis doctoral, Universidad Complutense de Madrid.

- Grubb, H. (1992), 'A Multivariate Time Series Analysis of Some Flour Price Data', *Applied Statistics* **41**, 95–107.
- Ho, B. & Kalman, R. (1966), 'Effective Construction of Linear State-Variable Models from Input-Output Functions', *Regelungstechnik* **14**, 545–548.
- Hurvich, C., Shumway, R. & Tsai, C. (1990), 'Improved Estimators of Kullback-Leibler in Information for Autoregressive Model Selection in Small Samples', *Biometrika* **77**, 709–719.
- Johansen, S. (1991), 'Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models', *Econometrica* **59**, 1551–1580.
- Larimore, W. E. (1983), System Identification, Reduced-Order Filtering and Modeling Via Canonical Variate Analysis, in 'Proceedings of the American Control Conference, ACC', San Francisco, pp. 445–451.
- Larimore, W. E. (1990), Canonical Variate Analysis in Identification, Filtering and Adaptive Control, in 'Proceedings of the 29th Conference on Decision and Control,' Hawaii, pp. 596–604.
- Lütkepohl, H. & Poskitt, D. S. (1996), 'Specification of Echelon Form VARMA Models', *Journal of Business and Economic Statistics* **14**(1), 69–79.
- Martín Manjón, R. & Treadway, A. (1997), 'The Fed Controls Only One of the Two Interest Rates in the U.S. Economy', *Documento de Trabajo del ICAE* (9716).
- Phillips, P. B. C. & Durlauf, N. S. (1986), 'Multiple Time Series Regression with Integrated Process', *Review of Economic Studies* **53**, 473–495.
- Poskitt, D. S. (2000), 'Strongly Consistent Determination of Cointegrating Rank Via Canonical Correlations', *Journal of Business and Economic Statistics* **18**, 77–90.
- Terceiro, J. (1990), *Estimation of Dynamics Econometric Models with Errors in Variables*, Springer-Verlag, Berlin.
- Tiao, G. C. & Tsay, R. S. (1989), 'Model Specification in Multivariate Time Series', *Journal of the Royal Statistical Society, B Series* **51**(2), 157–213.
- Van Overschee, P. & De Moor, B. (1994), 'N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems', *Automatica* **30**(1), 75–93.
- Viberg, M. (1995), 'Subspace-Based Methods for the Identification of the Linear Time-Invariant Systems', *Automatica* **31**(12), 1835–1852.