

# CAPITULO 13

## MODELOS DE SERIES TEMPORALES

### 13.1. INTRODUCCION

En los capítulos anteriores hemos tratado de la estimación de modelos econométricos, en los que una lista de variables exógenas se utiliza para explicar el comportamiento de otra variable, endógena. Hemos tratado de la contrastación de hipótesis con dichos modelos, así como de su utilización con fines predictivos y, posteriormente, hemos considerado modelos dinámicos. Este desarrollo convencional del modelo econométrico de relación entre variables se basa en el principio de mínimos cuadrados, tradicional en la Teoría Estadística, que se ha aplicado habitualmente en la estimación de relaciones en todo tipo de ciencias, experimentales o no experimentales.

Pero en ciencias no experimentales como la Economía, el investigador debe incluso tratar de descubrir la especificación del modelo, pues no existe un diseño experimental que haya generado los datos de que dispone. Apenas hemos tratado hasta ahora de la especificación de dichos modelos, excepto por algunas referencias conceptuales, nunca suficientemente precisas, al aspecto de Teoría Económica en debate. En 1970, Box y Jenkins propusieron una metodología rigurosa para la identificación, estimación y diagnóstico de modelos dinámicos para datos de series temporales que, justificadamente, se ha convertido en una herramienta habitual en el análisis de series económicas, y que presentamos en este capítulo. En la primera parte analizamos modelos en los que el comportamiento de una variable se explica utilizando sólo su propio pasado, por lo que se conocen como *modelos univariantes*. Posteriormente, consideramos modelos dinámicos con variables explicativas, que se conocen como *modelos de función de transferencia*.

Conceptualmente, los modelos de función de transferencia no son distintos de los modelos econométricos que hasta ahora hemos considerado, con la diferencia de que los procedimientos de especificación y diagnóstico que vamos a introducir permiten que las propiedades dinámicas de las relaciones entre variables, así como la estructura estocástica del término de error, queden

perfectamente recogidas en los modelos estimados. Los modelos univariantes, por otra parte, son un complemento perfecto al estudio de relaciones entre variables. Un modelo univariante permite caracterizar adecuadamente muchas de las características de una variable económica en su dimensión temporal, y ello es importante, al menos por dos razones: *a)* porque el analista querrá que las variables que aparecen a ambos lados de un modelo de relación tengan en común algunas de sus características más importantes, como indicativo de que la relación que ha especificado es realmente relevante, *b)* porque constituye una referencia, relativamente sencilla de obtener, con la que comparar posibles modelos de relación que posteriormente pudieran estimarse.

En este capítulo vamos a definir y caracterizar una amplia familia de estructuras estocásticas lineales, así como una familia de estadísticos que nos permitan escoger una de tales especificaciones como la más adecuada para representar la estructura estocástica de la variable económica que se está analizando.

## 13.2. PRIMERAS DEFINICIONES

### 13.2.a. Proceso estocástico, ruido blanco, paseo aleatorio

Comenzamos introduciendo algunos conceptos fundamentales sobre los que se sustenta la teoría que se desarrolla en las secciones siguientes:

**Definición 13.1.** Llamamos *proceso estocástico* a una sucesión de variables aleatorias  $\{y_t\}$ ,  $t = -\infty, \dots, -2, -1, 0, 1, 2, \dots, \infty$ .

Aunque el índice que describe la sucesión de variables aleatorias que configura un proceso estocástico no necesita tener una interpretación concreta, la utilización de este concepto de Econometría confiere al índice  $t$  una interpretación como del período al que corresponde la variable aleatoria  $y_t$ . Esta definición es muy general, y las variables aleatorias  $y_t$  no precisan satisfacer ninguna propiedad en particular. Podrían carecer de algunos momentos, como varianzas o esperanzas; incluso su distribución marginal (es decir, no condicional en un conjunto de información como el constituido por los valores de las variables previas  $y_s$ ,  $s < t$ ) podría no existir.

Según que las variables  $y_t$  satisfagan unas u otras propiedades, tenemos un proceso estocástico de uno u otro tipo, como los que comenzamos ya a presentar:

**Definición 13.2.** Se llama *ruido blanco* a una sucesión de variables aleatorias con esperanza cero, igual varianzas, e independientes en el tiempo. En lo sucesivo, denotamos un ruido blanco por  $\{\varepsilon_t\}$ .

**Definición 13.3.** Un *paseo aleatorio* es un proceso estocástico  $\{y_t\}$  cuyas primeras diferencias forman un proceso de ruido blanco, es decir:

$$\nabla y_t = \varepsilon_t$$

o, lo que es lo mismo:

$$y_t = y_{t-1} + \varepsilon_t, \quad t = -\infty, \dots, -2, -1, 0, 1, 2, \dots$$

donde  $\varepsilon_t$  es un ruido blanco.

Una forma de abordar el análisis estadístico de una serie temporal económica consiste en considerar que la serie temporal correspondiente a una variable como el consumo agregado es la *realización* de un proceso estocástico. Bajo este punto de vista, cada dato de la serie es una realización (es decir, una muestra de tamaño 1) de la variable  $C_t$  que compone el proceso estocástico de consumo  $\{C_t\}_{t=1}^{\infty}$ .

### 13.2.b. Estacionariedad

**Definición 13.4.** Un proceso estocástico  $\{y_t\}$  es *estacionario en sentido estricto* si para toda  $m$ -tupla  $(t_1, t_2, \dots, t_m)$  y todo entero  $k$  el vector de variables  $(y_{t_1}, y_{t_2}, \dots, y_{t_m})$  tiene la misma distribución de probabilidad conjunta que el vector  $(y_{t_1+k}, y_{t_2+k}, \dots, y_{t_m+k})$ .

Examinemos esta definición en detalle: ¿Qué ocurre cuando  $m = 1$ ? En tal caso, el conjunto de variables se reduce a una sola ( $y_t$ ) y, de acuerdo con la definición, en un proceso estocástico estacionario este conjunto ha de tener la misma distribución independientemente del valor del índice  $t$ . Es decir, que las *variables aleatorias que componen un proceso estocástico estacionario están idénticamente distribuidas*. En particular, la esperanza y la varianza de las variables  $y_t$  son independientes del tiempo.

Cuando  $m = 2$ , la distribución conjunta del par  $(y_t, y_{t-s})$  en un proceso estacionario debe ser independiente del tiempo. Como consecuencia, la covarianza entre ellas, así como su coeficiente de correlación, debe ser invariante en  $t$ , dependiendo únicamente del valor del retardo,  $s$ . Es decir, en un proceso estocástico estacionario, la covarianza entre  $y_t$  e  $y_{t-2}$  es igual a la covarianza entre  $y_s$  e  $y_{s-2}$ , para todo  $t$  y  $s$ , si bien será generalmente diferente de la covarianza entre  $y_t$  e  $y_{t-1}$  o de la covarianza entre  $y_t$  e  $y_{t-3}$ , que serán a su vez constantes en el tiempo, aunque diferentes entre sí.

Sin embargo, el concepto de estacionariedad en sentido estricto implica el cumplimiento de un número de condiciones que es excesivo para nuestras necesidades prácticas. Generalmente, nos conformamos con un concepto menos exigente, como es el de *estacionariedad en sentido débil* o *de segundo orden*, que se produce cuando todos los momentos de primer y segundo orden del proceso estocástico son invariantes en el tiempo. Estos momentos incluyen la esperanza matemática y la varianza de las variables  $y_t$ , pero también las covarianzas y correlaciones entre diversos retardos a que antes hemos hecho referencia. En lo sucesivo, cuando hablamos de proceso estacionario nos referimos a un proceso estocástico estacionario de segundo orden.

Algunas de estas condiciones son fácilmente contrastables. Por ejemplo, si una serie temporal muestra una tendencia creciente, como cada observación es una realización de la variable aleatoria correspondiente, no podremos

mantener que la esperanza matemática de dichas variables es constante, sino que deberemos aceptar que crece con el tiempo. En otras ocasiones, una serie temporal económica experimenta fluctuaciones de amplitud creciente en el tiempo, en cuyo caso deberemos reconocer que la varianza de las variables  $y_t$  no es constante y el proceso no es estacionario de segundo orden. *El ruido blanco es un ejemplo de proceso estocástico estacionario de segundo orden.*

El paseo aleatorio no es un proceso estacionario, puesto que puede escribirse:

$$y_t = y_{t-1} + \varepsilon_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t = y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t + \dots$$

sin que esté definida su esperanza ni su varianza, ni tampoco la distribución marginal de cada  $y_t$ . Si está definida, sin embargo, su distribución condicional en el pasado:  $\{y_{t-1}, y_{t-2}, \dots\}$ , que será del mismo tipo que la de  $\varepsilon_t$  (es decir, Normal o binomial dependiendo de que ésta lo sea), con igual varianza, y esperanza igual a la de  $\varepsilon_t$ , aumentada en el valor realizado de  $y_{t-1}$ .

A veces se considera un paseo aleatorio al que se le ha impuesto una condición inicial, la de comenzar a partir de un valor numérico  $y_0$ , de modo que puede escribirse:

$$\begin{aligned} y_t &= y_{t-1} + \varepsilon_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t = y_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t + \dots = \\ &= y_0 + \varepsilon_t + \varepsilon_{t-1} + \dots + \varepsilon_2 + \varepsilon_1 \end{aligned}$$

Este proceso, sobre el que hay que insistir que es *distinto del paseo aleatorio* que antes hemos definido, tiene distribución marginal, que será, por ejemplo Normal si la de  $\varepsilon_t$  lo es para todo  $t$ , con esperanza igual a  $y_0$  y varianza igual a  $t\sigma_\varepsilon^2$ , creciente en el tiempo. Como consecuencia, tampoco es estacionario. Su distribución condicional es igual a la del paseo aleatorio.

Unos estadísticos fundamentales en la especificación de modelos univariantes son las *funciones de autocovarianza, de autocorrelación simple y de autocorrelación parcial* que a continuación introducimos:

**Definición 13.5.** La *función de autocovarianza* de un proceso estocástico  $\{y_t\}$  es una función, a la que en lo sucesivo nos referimos como *fac.*, que para cada instante  $t$  y cada número entero  $k$  toma un valor, denotado por  $\gamma_k(t)$ , igual a la covarianza entre  $y_t$  e  $y_{t-k}$ , es decir:

$$\gamma_k(t) = \text{Cov}(y_t, y_{t-k})$$

**Definición 13.6.** La *función de autocorrelación simple* de un proceso estocástico  $\{y_t\}$ , a la que en lo sucesivo nos referimos por *fas.*, es una función que para cada instante  $t$  y cada entero  $k$  toma un valor  $\rho_k(t)$  igual a la correlación entre  $y_t$  e  $y_{t-k}$ :

$$\rho_k(t) = \frac{\text{Cov}(y_t, y_{t-k})}{\sqrt{\text{Var } y_t} \sqrt{\text{Var } y_{t-k}}} = \frac{\gamma_k(t)}{\sqrt{\text{Var } y_t} \sqrt{\text{Var } y_{t-k}}}$$

**Definición 13.7.** La *función de autocorrelación parcial* de un proceso estocástico  $\{y_t\}$ , a la que en lo sucesivo nos referimos como *fap*, es una función que para cada instante  $t$  y cada entero  $k$  toma un valor igual a la correlación entre  $y_t$  e  $y_{t-k}$ , ajustada por el efecto de los retardos intermedios  $y_{t-1}, y_{t-2}, \dots, y_{t-k-1}$ .

El gran interés de un proceso estocástico estacionario reside en que las *fas*, *fas* y *fap* son independientes del tiempo  $t$ , por lo que puede omitirse dicho argumento temporal. Lo que es crucial es que dicha invarianza permite la estimación muestral de tales funciones, del modo que analizamos en la sección siguiente. Asimismo, aun antes de pasar a su estimación, el concepto de estacionariedad nos permite deducir algunas propiedades de estas funciones:

En general, parte de la correlación existente entre  $y_t$  e  $y_{t-2}$  estará producida por el hecho de que ambas están correlacionadas con  $y_{t-1}$ , y eso es lo que trata de corregir la *fap*. El primer valor de la *fap* es la correlación entre  $y_t$  e  $y_{t-1}$ , sin que haya que corregir por ningún retardo intermedio, puesto que no existen. Por eso es que *el primer valor de las fas y fap de cualquier proceso estocástico coinciden*. De manera análoga, como el lector verá fácilmente, *el valor inicial de la fas,  $\rho_0$ , es igual a 1* en todo proceso estacionario, por ser el cociente de la varianza del proceso (constante en el tiempo) por sí misma. Por un razonamiento similar al anterior, concluimos que el valor inicial de la *fap* es asimismo igual a 1 en todo proceso estacionario. Por último, las *fas* y *fap* de un proceso de ruido blanco son iguales a cero, excepto en sus valores iniciales que, como hemos visto, son iguales a 1. Esto no es sino una manifestación de una propiedad más general que consiste en que *las fas y fap de todo proceso estocástico estacionario decrecen rápidamente hacia cero*.

### 13.2.c. Estimación de las funciones de autocorrelación de un proceso estacionario

Ya hemos expuesto cómo la estacionariedad de un proceso permite considerar la estimación de sus *fas* y *fap* que son las herramientas básicas de la especificación de su representación univariante. La estimación de los distintos valores  $\rho_k$  se lleva a cabo de la siguiente forma:

$$r_k = \frac{\frac{1}{T-k} \sum_{k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sqrt{\frac{1}{T} \sum_1^T (y_t - \bar{y})^2} \sqrt{\frac{1}{T-k} \sum_{k+1}^T (y_{t-k} - \bar{y})^2}}$$

Distintas simplificaciones se consiguen en esta expresión cuando el tamaño muestral  $T$  es grande con respecto a  $k$ , pues entonces dividir por  $T$  o por  $T-k$  es prácticamente lo mismo, en cuyo caso todos los cocientes de la forma  $\frac{1}{T}$  y  $\frac{1}{T-k}$  que aparecen en la expresión anterior pueden eliminarse.

Por otra parte, las medias muestrales de  $(y_t - \bar{y})^2$  sobre las observaciones 1, 2, ...,  $T$ , o sobre las observaciones  $k + 1, k + 2, \dots, T$  serán muy similares. Con estas aproximaciones se tiene el estimador:

$$r_k = \frac{\sum_{k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_1^T (y_t - \bar{y})^2}, \quad k = 0, 1, 2, \dots$$

La expresión que acabamos de presentar para la estimación de esta función garantiza que el valor estimado de  $\hat{r}_0$  será siempre igual a 1.

Finalmente, en todo proceso estacionario, la función de autocovarianza es simétrica, es decir,  $\gamma_{-k} = \gamma_k$ , lo que proviene del hecho de que la covarianza entre  $y_t$  e  $y_{t-k}$  es igual a la covarianza entre  $y_t$  e  $y_{t+k}$ . Como consecuencia, la función de autocorrelación es también simétrica, por lo que en el trabajo práctico se analizan únicamente dichas funciones para valores de  $k = 0, 1, 2, \dots$

El primer valor de la fap, que vamos a denotar por  $\phi_{11}$  (luego veremos por qué utilizamos dos subíndices), puede estimarse transformando la serie  $y_t$  en desviaciones con respecto a su media muestral  $\tilde{y}_t = y_t - \bar{y}$ , y a continuación estimando una regresión de  $\tilde{y}_t$  sobre  $\tilde{y}_{t-1}$ .

Ya vimos en el Capítulo 3 que, en el caso de variables con media muestral igual a cero, el estimador MCO de  $\beta$  en la regresión  $y_t = \beta x_t + u_t$  es precisamente igual al coeficiente de correlación parcial entre  $x$  e  $y$ , multiplicado por el cociente de sus desviaciones típicas. En este caso, las variables en la regresión son  $\tilde{y}_t$  e  $\tilde{y}_{t-1}$ , y si el proceso es estacionario, sus desviaciones típicas son iguales. Por tanto, la pendiente estimada de la regresión anterior coincide con el coeficiente de correlación entre  $y_t$  e  $y_{t-1}$ , es decir,  $\phi_{11}$ .

Otra consecuencia de este comentario es que el primer valor de la fap es siempre igual al primer valor de la fas. Intuitivamente, la razón es que al estimar la correlación entre  $y_t$  e  $y_{t-1}$ , no hay que corregirla de ningún valor intermedio, por lo que las dos funciones,  $\phi_{11}$  y  $r_1$ , toman el mismo valor numérico, al igual que ocurre con sus valores teóricos.

El segundo valor de la fap se estima mediante una regresión de  $\tilde{y}_t$  sobre  $\tilde{y}_{t-1}$  e  $\tilde{y}_{t-2}$ <sup>(1)</sup>.

$$\tilde{y}_t = \phi_{21}\tilde{y}_{t-1} + \phi_{22}\tilde{y}_{t-2} + u_t$$

El coeficiente estimado  $\phi_{22}$  mide la correlación entre  $\tilde{y}_t$  e  $\tilde{y}_{t-2}$  una vez que se ha tenido en cuenta el efecto común de  $\tilde{y}_{t-1}$ , incluida como otra variable explicativa adicional. Así, la fap puede estimarse mediante una serie de regresiones, cada una de las cuales contiene como variable explicativa un retardo más que la anterior, y de las que nos vamos quedando en cada caso con los coeficientes estimados en los retardos más altos:  $\phi_{11}, \phi_{22}, \phi_{33}, \dots$ . Otra posibilidad de obtener la fap estimada es mediante fórmulas recursivas.

<sup>(1)</sup> Rigurosamente hablando, las desviaciones con respecto a la media muestral alteran las propiedades estocásticas del proceso, pues para que no cambiasen se debería sustraer una constante. Restar la esperanza poblacional satisface esa condición, pero dicha esperanza es desconocida, y la corrección se hace con la media muestral, que es una variable aleatoria.

utilizando la fas previamente estimada, y utilizando las ecuaciones de Yule-Walker, como veremos en la sección siguiente.

### 13.3. MODELOS AUTORREGRESIVOS

Comenzamos a introducir en esta sección las estructuras estocásticas lineales que trataremos de asociar a una serie temporal de datos económicos. La primera de tales estructuras es la de los modelos autorregresivos.

**Definición 13.8.** Un *proceso autorregresivo* de orden 1, denotado por AR(1), viene definido por:

$$y_t = \phi y_{t-1} + \delta + \varepsilon_t \quad [13.1]$$

donde  $\phi$  y  $\delta$  son constantes y  $\varepsilon_t$  es un ruido blanco.

Si un modelo AR(1) es estacionario, entonces su esperanza y varianza son constantes en el tiempo, y se tiene que:

$$a) \quad E y_t = \phi E y_{t-1} + \delta + E \varepsilon_t = \phi E y_{t-1} + \delta, \text{ pero } E y_t = E y_{t-1}, \text{ por lo que}$$

$$E y_t = \mu_y = \frac{\delta}{1 - \phi}.$$

$$b) \quad \text{Var } y_t = \phi^2 \text{Var } y_{t-1} + \text{Var } \varepsilon_t. \text{ Ahora bien, si el proceso AR(1) es estacionario, entonces } \text{Var } y_t = \text{Var } y_{t-1}, \text{ por lo que } \text{Var } y = \frac{\sigma_\varepsilon^2}{1 - \phi^2}.$$

Por otra parte, mediante sustituciones repetidas, puede verse que:

$$y_t = (\delta + \delta\phi + \delta\phi^2 + \dots) + (\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots) = \frac{\delta}{1 - \phi} + \sum_{s=0}^{\infty} \phi^s \varepsilon_{t-s} \quad [13.2]$$

Esta expresión tendrá sentido sólo si la suma infinita que en ella aparece converge. Dicha suma es aleatoria, puesto que es una combinación lineal de las variables aleatorias  $\varepsilon_t$  y converge si y sólo si lo hace su varianza. Dado que las variables  $\varepsilon_t$  son independientes, se tiene:

$$\text{Var}(y_t) = \text{Var}\left(\sum_0^{\infty} \phi^s \varepsilon_{t-s}\right) = \sum_0^{\infty} \text{Var}(\phi^s \varepsilon_{t-s}) = \sum_0^{\infty} \phi^{2s} \text{Var } \varepsilon_{t-s} = \sum_0^{\infty} \phi^{2s} \sigma_\varepsilon^2 = \frac{\sigma_\varepsilon^2}{1 - \phi^2}$$

por lo que la varianza de la combinación lineal es finita si y sólo si  $|\phi| < 1$ . Sólo si  $|\phi| < 1$  tendrá sentido la expresión [13.2] y, en consecuencia,  $y_t$  podrá expresarse como función de  $\varepsilon_t$  y de todas las variables  $\varepsilon$  anteriores al instante  $t$ , pero de ninguna futura. Sólo entonces es el proceso AR(1) *estacionario*, para lo que es necesario y suficiente que  $|\phi| < 1$ . Es entonces cuando su esperanza y varianza están bien definidas por las expresiones que aquí hemos derivado.

El coeficiente de cada variable  $\varepsilon_{t-k}$  en dicha expresión es  $\phi^k$ . Como el proceso  $\varepsilon_t$  es un ruido blanco, entonces se concluye que:

a)  $E(y_{t-k}\varepsilon_t) = 0$  para todo  $k > 0$ , ya que  $y_{t-k}$  depende de  $\varepsilon_{t-k}$  y valores anteriores del proceso  $\varepsilon_t$ , pero no de sus valores futuros. Por tanto:

$$E(y_{t-k}\varepsilon_t) = E\left[\frac{\delta\varepsilon_t}{1-\phi}\right] + \sum_{s=0}^{\infty} \phi^s E(\varepsilon_{t-k-s}\varepsilon_t) = 0 + 0 = 0, \quad \forall k > 0$$

$$b) E(y_t\varepsilon_{t-k}) = E\left[\frac{\delta\varepsilon_{t-k}}{1-\phi}\right] + \sum_{s=0}^{\infty} \phi^s E(\varepsilon_{t-s}\varepsilon_{t-k}) = \phi^k \sigma_\varepsilon^2, \quad \forall k \geq 0.$$

Por otra parte, si se sustituye  $\delta$  por su expresión equivalente  $(1-\phi)\mu_y$  en el modelo [13.1], se tiene:

$$y_t - \mu_y = \phi(y_{t-1} - \mu_y) + \varepsilon_t, \quad \text{es decir: } \tilde{y}_t = \phi\tilde{y}_{t-1} + \varepsilon_t$$

donde  $\tilde{y}_t$  denota la diferencia entre  $y_t$  y su esperanza poblacional. Dicha diferencia tiene los mismos momentos que la variable  $\tilde{y}_t$ . En particular:  $\sigma_y^2 = \sigma_{\tilde{y}}^2$ , y pueden aplicarse los resultados anteriores a) y b). Con todo ello, la función de autocovarianza de este proceso resulta ser:

$$\gamma_0 = \sigma_y^2 = \frac{\sigma_\varepsilon^2}{1-\phi^2}, \text{ como vimos antes, y:}$$

$$\gamma_1 = E(\tilde{y}_t\tilde{y}_{t-1}) = E(\phi\tilde{y}_{t-1}^2 + \varepsilon_t\tilde{y}_{t-1}) = \phi\sigma_{\tilde{y}}^2 = \phi\gamma_0$$

$$\gamma_2 = E(\tilde{y}_t\tilde{y}_{t-2}) = E(\phi\tilde{y}_{t-1}\tilde{y}_{t-2} + \varepsilon_t\tilde{y}_{t-2}) = \phi\gamma_1 + E(\varepsilon_t\tilde{y}_{t-2}) = \phi^2\gamma_0 + 0 = \phi^2\gamma_0$$

$$\gamma_3 = E(\tilde{y}_t\tilde{y}_{t-3}) = E(\phi\tilde{y}_{t-1}\tilde{y}_{t-3} + \varepsilon_t\tilde{y}_{t-3}) = \phi\gamma_2 + E(\varepsilon_t\tilde{y}_{t-3}) = \phi^3\gamma_0$$

y así sucesivamente, de modo que:  $\rho_0 = 1$ ;  $\rho_1 = \phi$ ;  $\rho_2 = \phi^2$ ;  $\rho_3 = \phi^3$ ; ...;  $\rho_k = \phi^k$  para todo  $k > 0$ , por lo que los valores de la fas son las sucesivas potencias del parámetro  $\phi$ . Además, la condición  $|\phi| < 1$  garantiza que los sucesivos valores de la fas convergen a cero. Esta función puede tener dos aspectos distintos, dependiendo del signo de tal parámetro. Así, se tiene que:

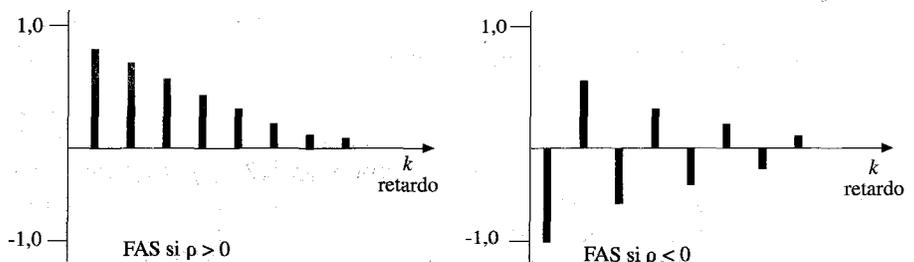


FIGURA 13.1. Función de autocorrelación simple de un proceso AR(1).

De la representación analítica [13.1] del proceso AR(1) se deduce que su fap tiene como primer valor  $\phi_{11} = \phi$ , y como restantes valores  $\phi_{kk} = 0$ . Por ejemplo, el segundo valor de dicha función sería el parámetro  $\phi_{22}$  en la regresión:

$$\tilde{y}_t = \phi_{21}\tilde{y}_{t-1} + \phi_{22}\tilde{y}_{t-2} + \varepsilon_t$$

pero, de acuerdo con el modelo teórico [13.1],  $\phi_{22}$  debe ser cero, y lo mismo ocurre para todo  $\phi_{kk}$  con  $k \geq 2$ .

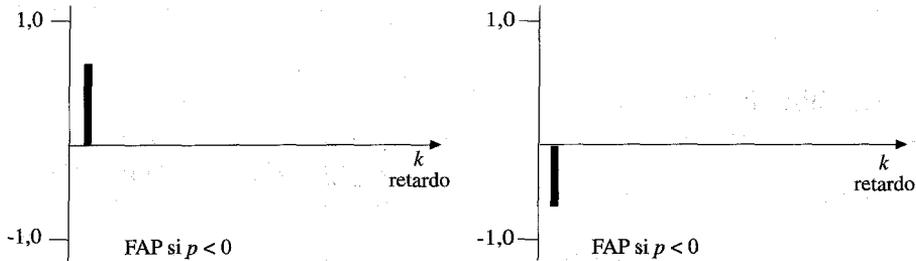


FIGURA 13.2. Función de autocorrelación parcial de un proceso AR(1).

**Definición 13.9.** Un proceso es *autorregresivo de orden 2*, que denotamos AR(2), si responde a la ley estocástica:

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad [13.3]$$

y es estacionario si  $|\phi_2| < 1$ ,  $\phi_1 + \phi_2 < 1$  y  $\phi_2 - \phi_1 < 1$ . Como veremos más adelante, bajo estas condiciones,  $y_t$  puede escribirse como función de  $\varepsilon_t$  y sus valores previos. Calculemos la función de autocorrelación de un proceso AR(2) estacionario: para ello comencemos notando que  $Ey_t = \mu_y = \frac{\delta}{1 - \phi_1 - \phi_2}$ . Si ahora sustituimos  $\delta$  en la ecuación del modelo por  $(1 - \phi_1 - \phi_2)\mu_y$ , se tiene:

$$y_t - \mu_y = \phi_1(y_{t-1} - \mu_y) + \phi_2(y_{t-2} - \mu_y) + \varepsilon_t$$

Multiplicando por  $y_{t-k} - \mu_y$  para  $k = 0, 1, 2$  y tomando esperanzas se tiene que:

$$\begin{aligned} \gamma_0 &= \phi_1\gamma_1 + \phi_2\gamma_2 + \sigma_\varepsilon^2 \\ \gamma_1 &= \phi_1\gamma_0 + \phi_2\gamma_1 \\ \gamma_2 &= \phi_1\gamma_1 + \phi_2\gamma_0 \end{aligned} \quad [13.4]$$

y, en general, multiplicando [13.3] por  $y_{t-k} - \mu_y$  para algún  $k > 0$  y tomando esperanzas se tiene:

$$\gamma_k = \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2}, \quad k \geq 1$$

donde hemos utilizado el hecho de que la función de autocovarianza es simétrica, de modo que  $\gamma_k = \gamma_{-k}$ . El sistema de ecuaciones [13.4] puede resolverse para obtener:

$$\text{Var}(y_t) = \gamma_0 = \frac{(1 - \phi_2)\sigma_\varepsilon^2}{(1 + \phi_2)[(1 - \phi_2)^2 - \phi_1^2]}$$

$$\gamma_1 = \frac{\phi_1 \gamma_0}{1 - \phi_2}$$

$$\gamma_2 = \frac{\phi_2(1 - \phi_2) + \phi_1^2}{1 - \phi_2} \gamma_0$$

que implican:

$$\rho_1 = \frac{\phi_1}{1 - \phi_2}$$

$$\rho_2 = \frac{\phi_1^2}{1 - \phi_2} + \phi_2$$

y en general:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2}, \quad k \geq 1$$

Partiendo de estimadores de  $\sigma_y^2$ ,  $\phi_1$ ,  $\phi_2$  y  $\sigma_\varepsilon^2$ , las ecuaciones [13.4] pueden utilizarse para obtener estimaciones de  $r_0$ ,  $r_1$  y  $r_2$  o, recíprocamente, como es más usual en la práctica, podemos comenzar de los valores estimados en la muestra de  $\sigma_y^2$ ,  $\rho_0$ ,  $\rho_1$  y  $\rho_2$ , y obtener estimaciones de  $\phi_1$ ,  $\phi_2$  y  $\sigma_\varepsilon^2$ . Esta estrategia dual se puede extender a modelos autorregresivos de orden  $p$ ,  $\text{AR}(p)$ , para cualquier  $p > 0$ . El sistema de ecuaciones [13.4] constituye las llamadas ecuaciones de Yule-Walker, que pueden extenderse a procesos  $\text{AR}(p)$ .

Un proceso  $\text{AR}(2)$  con parámetro  $\phi_2$  negativo puede generar raíces complejas en su ecuación característica  $1 - \phi_1 L - \phi_2 L^2 = 0$ . En tal caso,  $y_t$  presentará ciclos de *período*  $T$  dado por:

$$\cos \frac{2\pi}{T} = \frac{\phi_1}{2\sqrt{-\phi_2}}$$

y *factor de amortiguamiento*:  $d = \sqrt{-\phi_2}$ .

La función de autocorrelación de los procesos  $\text{AR}(2)$  también converge exponencialmente a cero. Sin embargo, dicha convergencia puede presentar una gran variedad de aspectos. Puede ser siempre con signo positivo, pero también puede alternar en signo. Estas dos eran las posibilidades en los modelos  $\text{AR}(1)$ . Sin embargo, ahora puede haber también convergencia a cero siguiendo una curva sinusoidal. Esta diversidad, junto con el hecho de que para algunos valores de  $\phi_1$  y  $\phi_2$  la fase es similar a la de los modelos  $\text{AR}(1)$ , hace que sea preciso alguna herramienta adicional para identificar un pro-

ceso AR(2). En cuanto a su fap, es claro que las estimaciones de los parámetros  $\phi_{11}$  y  $\phi_{22}$  en las regresiones

$$\begin{aligned}\tilde{y}_t &= \phi_{11}\tilde{y}_{t-1} + u_t \\ \tilde{y}_t &= \phi_{21}\tilde{y}_{t-1} + \phi_{22}\tilde{y}_{t-2} + \varepsilon_t\end{aligned}$$

serán no nulas, pero, sin embargo, la estimación de  $\phi_{33}$  en

$$\tilde{y}_t = \phi_{31}\tilde{y}_{t-1} + \phi_{32}\tilde{y}_{t-2} + \phi_{33}\tilde{y}_{t-3} + u_t$$

será estadísticamente igual a cero, al igual que  $\phi_{44}$ ,  $\phi_{55}$ , ... Esto implica que la fap de un proceso AR(2) es cero para valores  $k > 2$ . Lo importante es que a pesar de la variedad de formas que puede adoptar la fas dependiendo de los valores de los parámetros  $\phi_1$  y  $\phi_2$ , sin embargo, la propiedad que acabamos de mencionar para la fap es independiente de dichos valores.

Una propiedad similar es válida para todo modelo AR( $p$ ), donde  $p$  es cualquier entero positivo  $p$ : su función de autocorrelación parcial es cero para valores  $k > p$ .

### 13.4. MODELOS DE MEDIAS MOVILES

Analizamos en esta sección una clase diferente de procesos estocásticos, los procesos de medias móviles.

**Definición 13.10.** Se llama *proceso de medias móviles de orden 1*, que denotamos MA(1), a la estructura

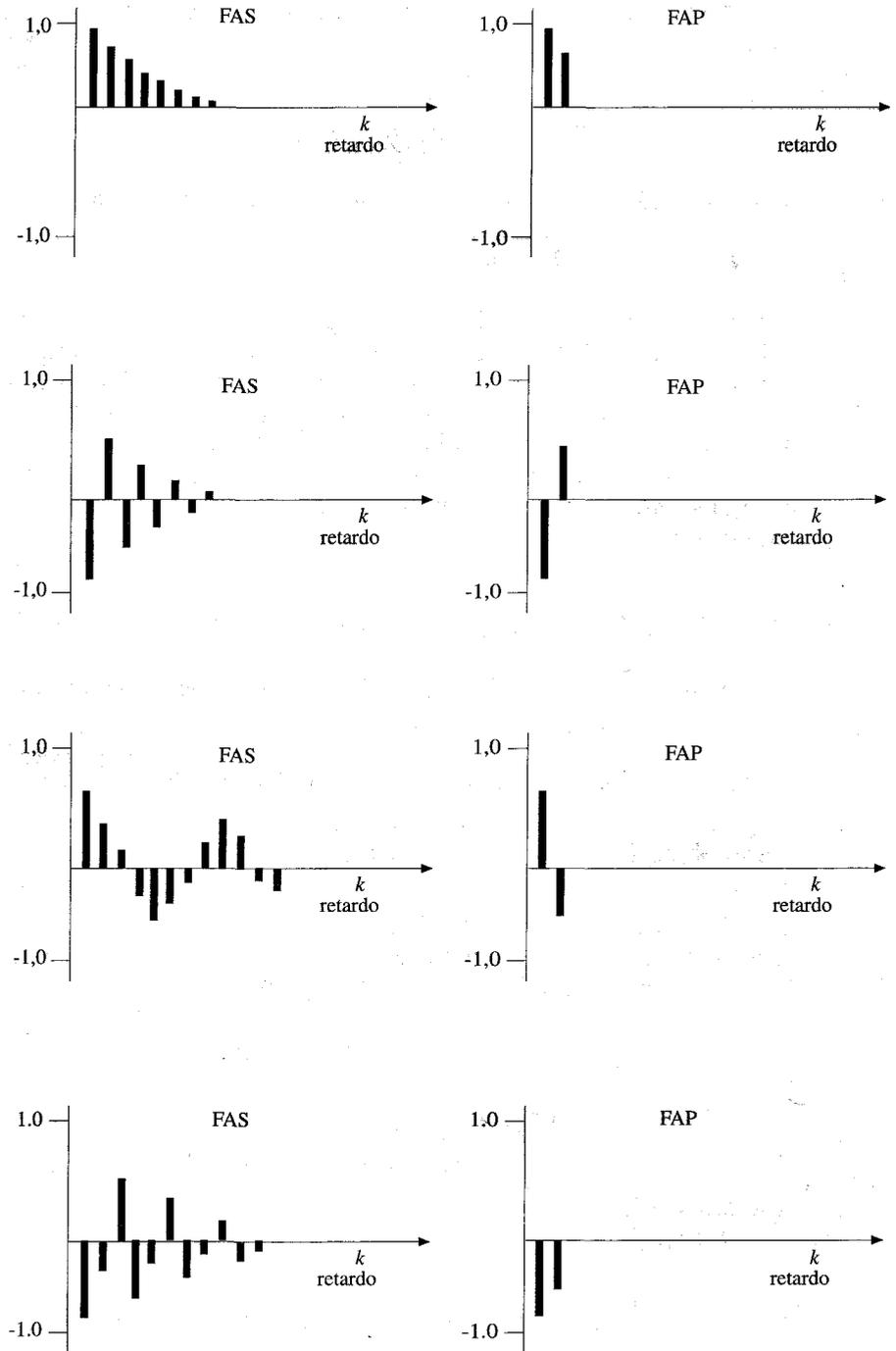
$$y_t = \delta + \varepsilon_t - \theta\varepsilon_{t-1}$$

donde  $\varepsilon_t$  es un ruido blanco. Como primeras propiedades de este proceso se tiene inmediatamente de su definición que  $Ey_t = \delta$  y también que  $\text{Var } y_t = (1 + \theta^2)\sigma_\varepsilon^2$ . En cuanto a la función de autocovarianza, se tiene:

$$\begin{aligned}\gamma_0 &= \sigma_y^2 = (1 + \theta^2)\sigma_\varepsilon^2 \\ \gamma_1 &= -\theta\sigma_\varepsilon^2 \\ \gamma_2 &= 0 \\ \gamma_k &= 0 \text{ para todo } k \geq 2\end{aligned}$$

y por tanto:  $\rho_1 = -\frac{\theta}{1 + \theta^2}$ ;  $\rho_2 = 0$ ,  $\rho_k = 0$  para todo  $k \geq 2$ . La función

$g(\theta) = -\frac{\theta}{1 + \theta^2}$  es monótona decreciente en  $\theta$  y, como consecuencia, puede verse fácilmente que el máximo valor absoluto que puede tomar  $\rho_1$  en un modelo MA(1) es 0,50, y éste es el *único valor no nulo de su función de autocorrelación simple*, siendo negativo si  $\theta > 0$ , y positivo en caso contrario.



**FIGURA 13.3.** Algunos ejemplos de funciones de autocorrelación simple y parcial en modelos AR(2).

A partir del proceso MA(1) puede llevarse a cabo el siguiente esquema «de inversión»:

$$y_t = \delta + \varepsilon_t - \theta\varepsilon_{t-1}, \quad \text{que implica: } \varepsilon_t = y_t - \delta + \theta\varepsilon_{t-1}$$

Del mismo modo:

$$\varepsilon_{t-1} = y_{t-1} - \delta + \theta\varepsilon_{t-2}$$

$$\varepsilon_{t-2} = y_{t-2} - \delta + \theta\varepsilon_{t-3}$$

por lo que se tiene, sustituyendo:

$$\begin{aligned} \varepsilon_t &= y_t - \delta + \theta(y_{t-1} - \delta + \theta\varepsilon_{t-2}) = y_t + \theta y_{t-1} - \delta(1 + \theta) + \theta^2\varepsilon_{t-2} = \\ &= y_t + \theta y_{t-1} - \delta(1 + \theta) + \theta^2(y_{t-2} - \delta + \theta\varepsilon_{t-3}) = \\ &= y_t + \theta y_{t-1} + \theta^2 y_{t-2} - \delta(1 + \theta + \theta^2) + \theta^3\varepsilon_{t-3} \end{aligned}$$

Por tanto:

$$y_t = -\theta y_{t-1} - \theta^2 y_{t-2} + \delta(1 + \theta + \theta^2) - \theta^3\varepsilon_{t-3} + \varepsilon_t \quad [13.5]$$

y el proceso continuaría, eliminando ahora el término  $\varepsilon_{t-3}$ .

Este procedimiento conduce, en el límite, a expresar  $y_t$  como función de su propio pasado, así como del valor contemporáneo del ruido blanco  $\varepsilon_t$ , pero sólo tiene sentido si  $|\theta| < 1$ . De otro modo, se tendría en [13.5] que el pasado de  $y_t$  tiene una gran importancia para determinar su comportamiento actual, y tanta más importancia cuanto más atrás en el tiempo. Cuando  $|\theta| < 1$  se dice que el proceso MA(1) es *invertible* y podemos obtener su representación autorregresiva como límite del procedimiento descrito en [13.5]:

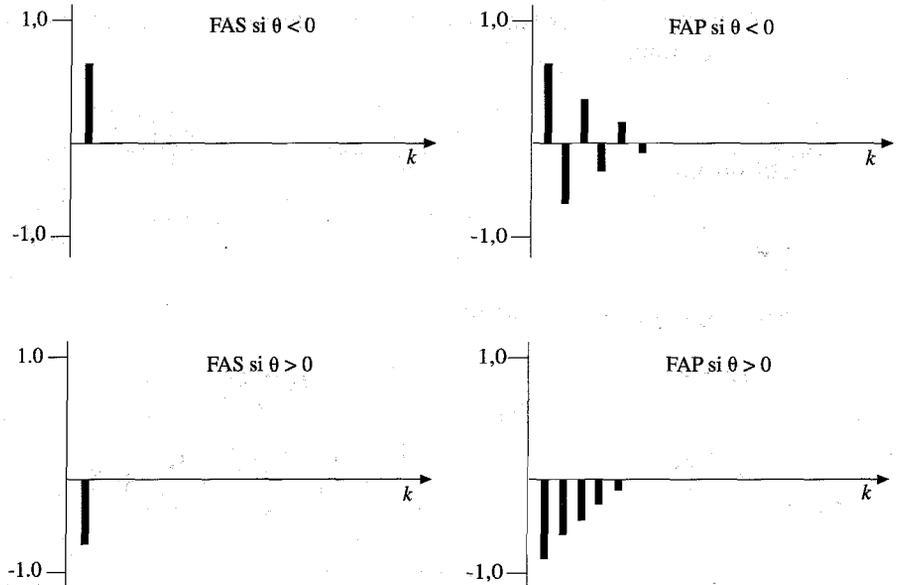
$$y_t = \frac{\delta}{1 - \theta} - \sum_{s=1}^{\infty} \theta^s y_{t-s} + \varepsilon_t$$

El hecho de que al invertir un proceso MA(1) se tengan coeficientes  $\theta^s$  en los retardos de  $y_t$  sugiere, como así es, que la fap de un proceso MA(1) decae exponencialmente hacia cero, quizá alternando en signo.

Si el parámetro  $\theta$  es negativo, entonces la fap converge a cero exponencialmente alternando en signo, y empezando de un valor positivo. Si, en cambio, el parámetro  $\theta$  tiene signo positivo, entonces la convergencia va a ser con todos los valores de la fap. tomando signo negativo. Nótese, por tanto, que un proceso MA(1) no puede generar nunca una fap que sea siempre positiva, pero sí puede generar una función de autocorrelación parcial que es siempre negativa.

**Definición 13.11.** Un proceso de medias móviles de orden 2 es un proceso estocástico que sigue la ley:

$$y_t = \delta + \varepsilon_t + \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2}$$



**FIGURA 13.4.** Ejemplos de funciones de autocorrelación simple y parcial de un proceso MA(1).

Siguiendo un proceso de inversión análogo al que hicimos con el proceso MA(1) se puede probar fácilmente que la fap de este proceso puede tener diversas formas, dependiendo de los signos y los valores relativos de los parámetros  $\theta_1$  y  $\theta_2$ . En cambio, la función de autocovarianza cumple:

$$\begin{aligned}
 * \quad \gamma_0 &= \sigma_y^2 = (1 + \theta_1^2 + \theta_2^2)\sigma_\varepsilon^2 \\
 \gamma_1 &= E[y_t y_{t-1}] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2})(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2} - \theta_2 \varepsilon_{t-3})] = \\
 &= E\varepsilon_t \varepsilon_{t-1} - \theta_1 E\varepsilon_t \varepsilon_{t-2} - \theta_2 E\varepsilon_t \varepsilon_{t-3} - \theta_1 E\varepsilon_{t-1}^2 + \theta_1^2 E\varepsilon_{t-1} \varepsilon_{t-2} + \\
 &+ \theta_1 \theta_2 E\varepsilon_{t-1} \varepsilon_{t-3} - \theta_2 E\varepsilon_{t-1} \varepsilon_{t-2} + \theta_1 \theta_2 E\varepsilon_{t-2}^2 + \theta_2^2 E\varepsilon_{t-2} \varepsilon_{t-3} = \\
 &= -\theta_1(1 - \theta_2)\sigma_\varepsilon^2
 \end{aligned}$$

donde se ha utilizado en repetidas ocasiones la ortogonalidad intertemporal del proceso de ruido blanco  $\varepsilon_t$ . Del mismo modo, se tiene:

$$\gamma_2 = E[y_t y_{t-2}] = E[-\theta_2 \varepsilon_{t-2} \varepsilon_{t-2}] = -\theta_2 \sigma_\varepsilon^2$$

mientras que  $\gamma_k = 0$  para todo  $k > 2$ . Por tanto, la fas es:

$$\begin{aligned}
 \rho_1 &= -\frac{\theta_1(1 - \theta_2)}{1 + \theta_1^2 + \theta_2^2} \\
 \rho_2 &= -\frac{\theta_2}{1 + \theta_1^2 + \theta_2^2} \\
 \rho_k &= 0 \quad \text{para todo } k > 2
 \end{aligned}$$

La identificación de un modelo MA(2) mediante su fap es bastante difícil [al igual que ocurre con las fas de un AR(2)], mientras que es muy sencillo identificar un modelo MA(2) por medio de su fas [igual que identificar un AR(2) mediante su fap]. Además los criterios de identificación son los mismos: la fas de un proceso MA(2) es cero para  $k > 2$ , de igual modo que la fap de un modelo AR(2) es cero para  $k > 2$ . Por otra parte, estos resultados son perfectamente generalizables para modelos de orden superior, AR( $p$ ) y MA( $q$ ), con  $p, q$  enteros positivos cualesquiera.

### 13.5. MODELOS ARMA

Hasta ahora hemos visto modelo de series temporales sencillos, en los que un proceso estocástico tenía una estructura autorregresiva «pura», o una estructura de medias móviles «pura». Sin embargo, en el análisis empírico de series temporales económicas es muy frecuente encontrar representaciones que tienen una componente autorregresiva así como una componente de medias móviles. Estos modelos se denotan como modelos ARMA( $p, q$ ), donde  $p$  y  $q$  denotan, respectivamente, los órdenes de los componentes autorregresivo y de medias móviles.

**Definición 13.12.** La estructura ARMA más sencilla es el modelo ARMA(1, 1):

$$y_t = \delta + \phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

que depende de cuatro parámetros desconocidos:  $\phi$ ,  $\theta$ ,  $\sigma_\varepsilon^2$  y  $\delta$ . El proceso ARMA(1, 1) es estacionario cuando  $|\phi| < 1$ , e invertible cuando  $|\theta| < 1$ . Distintos modelos pertenecientes a esta familia pueden escribirse sin más que variar los órdenes ( $p, q$ ) de los componentes autorregresivo y de medias móviles.

La esperanza del proceso ARMA(1, 1) es  $E(y_t) = \frac{\delta}{1 - \phi}$ . Para determinar su varianza, supongamos que  $\delta = 0$ , lo que sin duda no afecta a la varianza del proceso, para obtener:

$$\begin{aligned} \text{Var}(y_t) = \gamma_0 &= E[(\phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})^2] = \phi^2 E(y_{t-1}^2) + 2\phi E(y_{t-1} \varepsilon_t) - \\ &- 2\theta \phi E(y_{t-1} \varepsilon_{t-1}) + E(\varepsilon_t^2) - 2\theta E(\varepsilon_t \varepsilon_{t-1}) + \theta^2 E(\varepsilon_{t-1}^2) = \\ &= \phi^2 \gamma_0 - 2\phi\theta E(y_{t-1} \varepsilon_{t-1}) + \sigma_\varepsilon^2 + \theta^2 \sigma_\varepsilon^2 \end{aligned}$$

donde se han utilizado las siguientes propiedades: a) Si  $|\phi| < 1$ , el proceso ARMA(1, 1) es estacionario,  $y_{t-1}$  depende de  $\varepsilon_{t-1}$  y anteriores, pero no de sus valores futuros y, en particular,  $y_{t-1}$  es independiente de  $\varepsilon_t$ , y b) por ser ruido blanco,  $E(\varepsilon_t \varepsilon_{t-1}) = 0$ .

Como además,  $E(y_{t-1} \varepsilon_{t-1}) = \sigma_\varepsilon^2$ , se tiene de la expresión anterior que

$$\gamma_0(1 - \phi^2) = \sigma_\varepsilon^2(1 + \theta^2 - 2\phi\theta)$$

y, finalmente:

$$\gamma_0 = \frac{1 + \theta^2 - 2\phi\theta}{1 - \phi^2} \sigma_\varepsilon^2$$

Los distintos valores de la función de autocovarianza del proceso ARMA(1, 1) pueden obtenerse de un modo similar:

$$\gamma_1 = E[y_{t-1}(\phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})] = \phi \gamma_0 - \theta \sigma_\varepsilon^2 = \frac{(1 - \phi\theta)(\phi - \theta)}{1 - \phi^2} \sigma_\varepsilon^2$$

$$\gamma_2 = E[y_{t-2}(\phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1})] = \phi \gamma_1$$

y, en general,  $\gamma_k = \phi \gamma_{k-1}$  para todo  $k \geq 2$ .

Como resultado, la función de autocorrelación del proceso ARMA(1, 1) es:

$$\rho_1 = \frac{(1 - \phi\theta)(\phi - \theta)}{1 + \theta^2 - 2\phi\theta}$$

$$\rho_k = \phi \rho_{k-1} \quad \text{para todo } k \geq 2$$

En consecuencia, la función de autocorrelación de un proceso ARMA(1, 1) comienza del valor  $\rho_1$  que acabamos de mostrar y, a partir de él, decrece a una tasa  $\phi$ . Es decir, dicha función de autocorrelación se comporta, a partir de  $k = 1$ , como la función de autocorrelación de un proceso AR(1). Esta propiedad puede generalizarse: La función de autocorrelación de un proceso ARMA( $p$ ,  $q$ ) se comporta como la función de autocorrelación del proceso AR( $p$ ), a partir de  $k > q$ .

Ello hace que la identificación de estos modelos por inspección de la serie temporal correspondiente no se ajuste a unas normas tan bien definidas como la identificación de modelos AR( $p$ ) o MA( $q$ ). Ello se debe a que tanto la función de autocorrelación como la función de autocorrelación parcial de los modelos ARMA heredan características de ambos componentes. Así puede probarse también que mientras que la función de autocorrelación parcial de un modelo AR(1) es cero para  $k > 1$ , sin embargo, la de un modelo ARMA(1, 1) no tendrá esta característica, pues a ella hay que superponer la de la componente MA(1), que genera, como sabemos, una función de autocorrelación parcial que converge exponencialmente a cero. Las mismas consideraciones pueden hacerse para la función de autocorrelación simple.

Sin embargo, en la práctica esta dificultad no resulta excesivamente importante, pues tampoco se trata de obtener la mejor identificación del modelo en un primer intento. Así, el modo más frecuente de terminar con una especificación ARMA(2, 1), por ejemplo, es haber comenzado con una especificación AR(2), y observar que los residuos obtenidos tras la estimación de tal modelo tienen una estructura MA(1) (o recíprocamente).

Es interesante hacer hincapié en que esta superposición de modelos tiene una fundamentación analítica. Si se ha especificado el modelo

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$$

y tras estimarlo e inspeccionar la función de autocorrelación de los residuos se observa que éstos parecen seguir una estructura MA(1):

$$\hat{u}_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

entonces estos dos modelos pueden unirse para obtener:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

es decir, una estructura ARMA(2, 1). En uno de los ejercicios al final del capítulo se pide al lector que lleve a cabo varios ejemplos similares a éste.

### 13.6. VARIABLES NO ESTACIONARIAS

Al presentar los modelos AR( $p$ ), supusimos que se satisfacían las condiciones necesarias para garantizar que el proceso en estudio era estacionario. Sin embargo, las series de datos económicos suelen caracterizarse por ser claramente no estacionarias. Como ya hemos citado, la observación de una tendencia lineal o cuadrática en el tiempo basta para rechazar el supuesto de estacionariedad.

Cuando esto ocurre en series económicas, es también cierto que si se toman primeras o segundas diferencias de la serie  $\nabla y_t = y_t - y_{t-1}$ , o  $\nabla^2 y_t = y_t - 2y_{t-1} + y_{t-2}$ , se obtienen series transformadas que son estacionarias. Lo que se hace en tales situaciones es trabajar con las series en primeras o segundas diferencias, especificando y estimando un modelo para ellas. Como veremos más adelante, si se ha llevado a cabo un análisis de predicción para estas series, es bastante sencillo traducir las predicciones obtenidas para estas series diferenciadas en predicciones para las series originales, que son, en definitiva, aquellas en cuyo análisis estaba interesado el investigador.

Los procesos no estacionarios que pueden transformarse en estacionarios mediante sus diferencias de orden  $d$  se conocen como *procesos estocásticos no estacionarios, homogéneos de orden  $d$* . Un ejemplo de tales procesos es el paseo aleatorio. Su varianza crece con el tiempo, lo que le hace no estacionario. Sin embargo, su primera diferencia es, por definición:

$$\nabla y_t = y_t - y_{t-1} = \varepsilon_t$$

un ruido blanco y, como tal, es un proceso estocástico estacionario.

Cuando se especifica un modelo ARMA( $p, q$ ) para la variable  $\Delta^d y_t$ , se dice que tenemos un modelo ARIMA( $p, d, q$ ) para  $y_t$ .

Si la evidencia de no estacionariedad proviene de que el gráfico de la serie temporal que se pretende analizar muestra fluctuaciones cuya amplitud cambia para distintos intervalos del período muestral, pensaremos igualmente que el proceso estocástico que generó la serie temporal no es estacionario. En este caso, la no estacionariedad surgiría porque la varianza de las diferentes va-

riables aleatorias que lo integran a lo largo del tiempo no son iguales entre sí.

El procedimiento habitual en esta situación consiste en transformar la variable tomando logaritmos, y pasar a analizar esta variable transformada. Veremos en la sección dedicada a predicción cómo recuperar las predicciones de la serie original a partir de las predicciones obtenidas para la serie en logaritmos. Es importante hacer notar que la transformación logarítmica no va a corregir el problema de heteroscedasticidad a que nos referimos, sino que simplemente lo va a amortiguar, hasta el punto de hacerlo apenas perceptible. En este sentido, esta transformación es conceptualmente diferente de la diferenciación en el caso de no estacionariedad en la esperanza matemática que hemos visto antes. Por otra parte, la transformación logarítmica persigue estabilizar la varianza de la variable, mientras que la diferenciación busca estabilizar su esperanza matemática.

También debe notarse que la transformación Box-Cox que analizamos en el Capítulo 11 incluye a la logarítmica como caso particular y que, en ocasiones, podría considerarse otra transformación de la familia Box-Cox.

### 13.7. ESTACIONARIEDAD E INVERTIBILIDAD

Hay varias razones importantes para pretender que un proceso estocástico con el que se va a efectuar trabajo empírico sea estacionario: En primer lugar, si no es estacionario en media, por ejemplo, ello quiere decir que la esperanza matemática de las variables del proceso cambia con el tiempo, y entonces habría que estimar un número infinito de parámetros. Aún más importante, un proceso puede ser no estacionario porque sus momentos, o su distribución, no existan. Lo mismo ocurriría con la varianza.

Un modelo AR(1) puede desarrollarse:

$$\begin{aligned} y_t &= \phi y_{t-1} + \varepsilon_t = \phi^2 y_{t-2} + \phi \varepsilon_{t-1} + \varepsilon_t = \dots = \\ &= \phi^t y_0 + \phi^{t-1} \varepsilon_1 + \dots + \phi \varepsilon_{t-1} + \varepsilon_t \end{aligned}$$

Vimos anteriormente que el interés de que un proceso AR(1) sea estacionario se basa en que, de lo contrario, con  $|\phi| > 1$  se tendría que la expresión anterior es divergente, como puede apreciarse en el hecho de que los coeficientes de las variables aleatorias en la serie son crecientes sin límite.

Estos comentarios pueden extenderse a cualquier proceso AR( $k$ ). Por ejemplo, un proceso AR(2),  $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$ , es estacionario si al descomponer su *ecuación característica* en factores

$$1 - \phi_1 L - \phi_2 L^2 = (1 - \lambda_1 L)(1 - \lambda_2 L)$$

donde  $L$  denota al operador de retardos, ambos parámetros,  $\lambda_1$  y  $\lambda_2$ , tienen módulo menor que 1. Bastaría, por el contrario, que uno de ellos fuese mayor que 1 en valor absoluto, para que la realización de  $y_t$  dependiese de valores futuros de las variables  $\varepsilon_t$ , y quizá también de sus valores pasados. Ya vimos

en la Sección 13.3 que las condiciones sobre los parámetros para que ello no ocurra son:

$$|\phi_2| < 1, \quad \phi_2 + \phi_1 < 1, \quad \phi_2 - \phi_1 < 1$$

Por otra parte, todo proceso MA de orden finito, por ser una combinación lineal de un número finito de variables aleatorias con distribución Normal, tiene asimismo una distribución Normal, con esperanza cero y una varianza que depende de los coeficientes del modelo, pero que es independiente del tiempo y es, por tanto, siempre un proceso estocástico estacionario.

En el caso de procesos de medias móviles, las condiciones similares a las de estacionariedad son las de invertibilidad, pero tienen una interpretación diferente: Cuando un proceso de medias móviles es invertible, entonces dicho modelo admite una representación autorregresiva, en que los valores pasados de la variable  $y_t$  reciben una ponderación (coeficientes) cada vez menor cuanto más alejados en el tiempo. Esta propiedad de un modelo de medias móviles es fundamental a efectos de predicción, para lo que, como veremos, es necesario invertir primero el proceso MA y obtener su representación AR.

Un proceso ARMA( $p, q$ ) es estacionario si lo es su componente AR, y es invertible si lo es su componente MA.

## 13.8. ESTACIONALIDAD

Las series temporales de datos económicos presentan generalmente características estacionales cuando se observan a una frecuencia inferior a la anual, ya sea mediante datos trimestrales o mensuales. Estas características se deben a que las decisiones tomadas por los agentes económicos en un determinado trimestre del año pueden estar correlacionadas con las decisiones tomadas en el mismo trimestre de otros años. Algo similar ocurre, en general, con datos mensuales.

Tales correlaciones pueden representarse analíticamente por modelos univariantes totalmente análogos a los vistos en las secciones anteriores. Así, la especificación de las características estacionales de una serie pudiera ser del tipo

$$y_t = \phi_{12}y_{t-12} + e_t$$

si además existen correlaciones entre las observaciones de meses consecutivos, se tendrá:

$$e_t = \phi e_{t-1} + \varepsilon_t$$

y, en resumen:

$$(1 - \phi L)(1 - \phi_{12}L^{12})y_t = (y_t - \phi_{12}y_{t-12}) - \phi(y_{t-1} - \phi_{12}y_{t-13}) = \varepsilon_t$$

La especificación del modelo univariante para la *estructura estocástica estacional* de una serie temporal se lleva a cabo de modo idéntico al descrito

en las secciones anteriores para la identificación de un modelo para su estructura estocástica regular, con la salvedad de que para ello se examinan los valores «estacionales» de las funciones de autocorrelación,  $\rho_4, \rho_8, \rho_{12}, \dots$  si los datos son trimestrales, o  $\rho_{12}, \rho_{24}, \rho_{36}, \dots$  si son mensuales.

Así, por ejemplo, una serie puede requerir diferencias de orden estacional si los valores estacionales de su función de autocorrelación no tienden a cero rápidamente. Si trabajando con datos trimestrales se tiene que tanto  $\rho_4$  como  $\rho_8$  y  $\rho_{12}$  son significativamente diferentes de cero, será en general apropiado transformar la variable mediante la diferencia:

$$z_t = \nabla_4 y_t = y_t - y_{t-4}$$

Así, una estructura posible para una serie mensual podría ser:

$$\nabla \nabla_{12} \ln y_t = (1 - \theta L) (1 - \Theta L^{12}) \varepsilon_t$$

que indica que la serie precisó de la transformación logarítmica, así como de una diferencia, tanto de carácter regular como estacional, esta última sugerida, sin duda, porque los valores de la fas para  $k = 12, 24, 36$  no convergían a cero. Tras estas transformaciones, los únicos valores no nulos de la fas eran los correspondientes a  $k = 1$  y  $k = 12$ , por lo que se especificó el modelo descrito.

## 13.9. PREDICCIÓN CON MODELOS ARIMA

Es un resultado conocido en Estadística que cuando se pretende que la predicción  $\hat{y}_{T+k}^T$  elaborada en el instante  $T$  acerca del valor de la variable endógena en el instante  $T+k$  minimice la función de pérdida  $E_T[(y_{T+k} - \hat{y}_{T+k}^T)^2]$ , entonces la predicción debe ser  $\hat{y}_{T+k}^T = E_T y_{T+k}$ , es decir, precisamente la esperanza condicional de la variable  $y_{T+k}$ , calculada sobre la base de la información disponible en el instante  $T$  (véase Problema 13.4). Este es el criterio que utilizaremos en esta sección.

### 13.9.a. Modelos autorregresivos

En un modelo AR(1) se tiene:

$$E_T y_{T+1} = E_T(\delta + \phi y_T + \varepsilon_{T+1}) = \delta + \phi y_T + E_T \varepsilon_{T+1} = \delta + \phi y_T$$

donde hemos utilizado dos resultados:

a) Que  $y_T$  está en el conjunto de información disponible en el instante  $T$ , es decir, que la persona que realiza la predicción observa el valor  $y_T$  previamente al cálculo de la predicción, con lo que  $E_T y_T = y_T$ .

b)  $E_T \varepsilon_{T+1} = 0$  por ser  $\varepsilon_t$  un ruido blanco. De modo similar:

$$\begin{aligned} E_T y_{T+2} &= E_T(\delta + \phi y_{T+1} + \varepsilon_{T+2}) = \delta + \phi E_T y_{T+1} + E_T \varepsilon_{T+2} = \\ &= \delta + \phi E_T y_{T+1} = \delta(1 + \phi) + \phi^2 y_T \\ E_T y_{T+k} &= \phi^k y_T + \delta(1 + \phi + \phi^2 + \dots + \phi^{k-1}) \end{aligned}$$

Notemos que:

$$1 + \phi + \phi^2 + \dots + \phi^{k-1} = \frac{1 - \phi^k}{1 - \phi}$$

y recordando que  $Ey = \delta/(1 - \phi)$  se tiene:

$$\delta(1 + \phi + \phi^2 + \dots + \phi^k) = \left[ \frac{1 - \phi^k}{1 - \phi} \right] (1 - \phi) Ey = (1 - \phi^k) Ey$$

por lo que se tiene finalmente:

$$E_T y_{T+k} = \phi^k y_T + (1 - \phi^k) Ey$$

de modo que la predicción óptima  $k$  periodos hacia el futuro es una combinación lineal convexa de dos términos:

- la última observación obtenida de la variable  $y_t$ ;
- su esperanza matemática.

Según avanzamos hacia el futuro, la última observación recibe una ponderación más pequeña, y la esperanza matemática recibe un peso más importante. Ello refleja el hecho de que, en un proceso estacionario, cuanto más lejos hacia el futuro queremos predecir, mayor será la incertidumbre bajo la que se obtiene la predicción: En un régimen de total incertidumbre, la predicción óptima del valor de una variable aleatoria es igual a su esperanza matemática. Por el contrario, cuando predecimos a horizontes cortos, la información muestral permite mejorar la predicción que se haría si se utilizase únicamente la esperanza matemática de la variable aleatoria. La propiedad característica del proceso AR(1) es que toda la información muestral relevante para la predicción quede resumida en el último valor observado de la variable; en consecuencia, dicho valor es todo lo que se precisa, junto con la esperanza matemática del proceso, para elaborar predicciones.

En un proceso AR(2) se tiene:

$$\begin{aligned} E_T y_{T+1} &= E_T(\delta + \phi_1 y_T + \phi_2 y_{T-1} + \varepsilon_{T+1}) = \delta + \phi_1 y_T + \phi_2 y_{T-1} \\ E_T y_{T+2} &= \delta + \phi_1 E_T y_{T+1} + \phi_2 y_T = \delta(1 + \phi_1) + (\phi_1^2 + \phi_2) y_T + \phi_1 \phi_2 y_{T-1} \\ E_T y_{T+3} &= \delta(1 + \phi_1 + \phi_2 + \phi_1^2) + \phi_1(\phi_1^2 + 2\phi_2) y_T + (\phi_1^2 \phi_2 + \phi_2^2) y_{T-1} \end{aligned}$$

y así sucesivamente. En procesos AR( $p$ ) de orden superior a 1 no existe una forma analítica sencilla para la predicción  $k$  periodos hacia el futuro. A pesar

de ello, es fácil observar una diferencia con respecto a la predicción con modelos AR(1), y es que la información muestral relevante para la predicción se resume en las  $p$  últimas observaciones de la variable a predecir.

### 13.9.b. Modelos de medias móviles

Un proceso MA(1) invertible puede escribirse:

$$y_{T+1} = -\theta y_T - \theta^2 y_{T-1} - \theta^3 y_{T-2} - \dots + \varepsilon_{T+1}$$

por lo que:

$$E_T y_{T+1} = -\theta y_T - \theta^2 y_{T-1} - \theta^3 y_{T-2} + \dots$$

$$\begin{aligned} E_T y_{T+2} &= -\theta E_T y_{T+1} - \theta^2 y_T - \theta^3 y_{T-1} - \theta^4 y_{T-2} - \dots = \\ &= -(-\theta^2 y_T - \theta^3 y_{T-1} - \theta^4 y_{T-2} - \dots) - \theta^2 y_T - \theta^3 y_{T-1} - \theta^4 y_{T-2} - \dots = 0 \end{aligned}$$

$$E_T y_{T+k} = 0 \quad \text{para } k \geq 2$$

es decir, que la predicción de un proceso MA(1) para 2 o más períodos hacia el futuro es cero. Es fácil ver las razones que explican este resultado, puesto que el modelo

$$y_t = \varepsilon_t - \theta \varepsilon_{t-1}$$

implica que

$$y_{T+1} = \varepsilon_{T+1} - \theta \varepsilon_T, \quad y_{T+2} = \varepsilon_{T+2} - \theta \varepsilon_{T+1}$$

y así sucesivamente.

Como  $\varepsilon_t$  es ruido blanco, se tiene  $E_T \varepsilon_{T+1} = 0$ , y más generalmente,  $E_T \varepsilon_{T+k} = 0$  para todo  $k > 0$ . En consecuencia,  $E_T y_{T+2} = 0$ , y también  $E_T y_{T+s} = 0$  para todo  $s \geq 2$ .

Análogamente, puede verse que en un modelo MA(2),  $E_T y_{T+k} = 0$  para  $k \geq 3$ . Más generalmente, en un modelo MA( $q$ ),  $E_T y_{T+s} = 0$  para todo  $s > q$ .

### 13.9.c. El modelo ARMA(1, 1)

Las expresiones analíticas correspondientes a las predicciones de valores futuros del modelo  $y_t = \delta + \phi y_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$ , son:

$$E_T y_{T+1} = \delta + \phi y_T - \theta \varepsilon_T$$

$$E_T y_{T+2} = \delta + \phi E_T y_{T+1} = (1 + \phi)\delta + \phi^2 y_T - \phi \theta \varepsilon_T$$

y, finalmente:

$$E_T y_{T+k} = (1 + \phi + \phi^2 + \dots + \phi^{k-1})\delta + \phi^k y_T - \phi^{k-1} \theta \varepsilon_T$$

donde puede verse que, como ocurre para todo proceso estacionario:

$$\lim_{k \rightarrow \infty} E_T y_{T+k} = \frac{\delta}{1 - \phi} = E(y_t)$$

la predicción converge a la esperanza matemática del proceso cuando el horizonte de predicción tiende a infinito.

### 13.9.d. Error de predicción

El error de predicción es la diferencia entre la realización de la variable aleatoria y la predicción hecha para dicho valor. Por supuesto, el error cometido en la predicción de  $y_{T+k}$  depende del periodo en que dicha predicción se hizo.

Denotamos por  $e_T(k)$  el error de predicción  $k$  periodos hacia adelante, elaborada en el período  $T$ , es decir:  $e_T(k) = y_{T+k} - E_T y_{T+k}$ . En todos estos modelos, la esperanza del error de predicción es cero, precisamente, por ser el predictor óptimo la esperanza condicional de la variable a predecir. En efecto, se tiene:

$$E(e_T(k)) = E(y_{T+k} - E_T y_{T+k}) = E y_{T+k} - E[E_T y_{T+k}] = E y_{T+k} - E y_{T+k} = 0$$

Siendo el error de predicción  $e_T(k)$  una variable aleatoria de esperanza cero, su varianza nos da una medida de su tamaño. En este sentido, una de las variables de más interés en el análisis de predicción es, por tanto, *la varianza del error de predicción*. Dicha varianza (más exactamente, su raíz cuadrada) es la cantidad que se utiliza para construir los intervalos de confianza de las predicciones puntuales obtenidas del modo descrito en los párrafos anteriores.

De acuerdo con dichas expresiones para las predicciones óptimas se tienen los siguientes errores y varianzas de error de predicción (que el lector debe asegurarse de entender):

a) Errores de predicción en un modelo AR(1):

$$e_T(1) = y_{T+1} - E_T y_{T+1} = (\delta + \phi y_T + \varepsilon_{T+1}) - (\delta + \phi y_T) = \varepsilon_{T+1}$$

$$e_T(2) = \varepsilon_{T+2} + \phi \varepsilon_{T+1}$$

.....

$$e_T(k) = \varepsilon_{T+k} + \phi \varepsilon_{T+k-1} + \phi^2 \varepsilon_{T+k-2} + \dots + \phi^{k-1} \varepsilon_{T+1}$$

Varianza del error de predicción:

$$\text{Var}(e_T(1)) = \sigma_\varepsilon^2$$

$$\text{Var}(e_T(2)) = \sigma_\varepsilon^2(1 + \phi^2)$$

$$\text{Var}(e_T(3)) = \sigma_\varepsilon^2(1 + \phi^2 + \phi^4)$$

$$\text{Var}(e_T(k)) = \sigma_\varepsilon^2(1 + \phi^2 + \phi^4 + \dots + \phi^{2(k-1)}) = \sigma_\varepsilon^2 \left[ \frac{1 - \phi^{2k}}{1 - \phi^2} \right]$$

b) Errores de predicción en un modelo AR(2):

$$e_T(1) = \varepsilon_{T+1}$$

$$e_T(2) = \varepsilon_{T+2} + \phi_1 \varepsilon_{T+1}$$

$$e_T(3) = \varepsilon_{T+3} + \phi_1 \varepsilon_{T+2} + (\phi_1^2 + \phi_2) \varepsilon_{T+1}$$

Varianza del error de predicción:

$$\text{Var}(e_T(1)) = \sigma_\varepsilon^2$$

$$\text{Var}(e_T(2)) = \sigma_\varepsilon^2(1 + \phi_1^2)$$

$$\text{Var}(e_T(3)) = \sigma_\varepsilon^2[1 + \phi_1^2 + (\phi_1^2 + \phi_2)^2]$$

c) Errores de predicción en un modelo MA(1):

$$e_T(1) = \varepsilon_{T+1}$$

$$e_T(2) = \varepsilon_{T+2} - \theta \varepsilon_{T+1}$$

$$e_T(k) = \varepsilon_{T+k} - \theta \varepsilon_{T+k-1} \quad \text{para todo } k > 1$$

Varianza del error de predicción:

$$\text{Var}(e_T(1)) = \sigma_\varepsilon^2$$

$$\text{Var}(e_T(2)) = \sigma_\varepsilon^2(1 + \theta^2)$$

$$\text{Var}(e_T(k)) = \sigma_\varepsilon^2(1 + \theta^2) \quad \text{para todo } k > 1$$

d) Errores de predicción en un modelo MA(2):

$$e_T(1) = \varepsilon_{T+1}$$

$$e_T(2) = \varepsilon_{T+2} - \theta_1 \varepsilon_{T+1}$$

$$e_T(3) = \varepsilon_{T+3} - \theta_1 \varepsilon_{T+2} - \theta_2 \varepsilon_{T+1}$$

$$e_T(k) = \varepsilon_{T+k} - \theta_1 \varepsilon_{T+k-1} - \theta_2 \varepsilon_{T+k-2} \quad \text{para todo } k > 2$$

Varianza del error de predicción:

$$\text{Var}(e_T(1)) = \sigma_\varepsilon^2$$

$$\text{Var}(e_T(2)) = \sigma_\varepsilon^2(1 + \theta_1^2)$$

$$\text{Var}(e_T(3)) = \sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2)$$

$$\text{Var}(e_T(k)) = \sigma_\varepsilon^2(1 + \theta_1^2 + \theta_2^2) \quad \text{para todo } k > 2$$

e) Errores de predicción en un modelo ARMA(1, 1):

$$e_T(1) = \varepsilon_{T+1}$$

$$e_T(2) = \varepsilon_{T+2} + (\phi - \theta)\varepsilon_{T+1}$$

$$e_T(3) = \varepsilon_{T+3} + (\phi - \theta)\varepsilon_{T+2} + \phi(\phi - \theta)\varepsilon_{T+1}$$

$$e_T(k) = \varepsilon_{T+k} + (\phi - \theta)\varepsilon_{T+k-1} + \phi(\phi - \theta)\varepsilon_{T+k-2} + \phi^2(\phi - \theta)\varepsilon_{T+k-3} + \dots + \phi^{k-2}(\phi - \theta)\varepsilon_{T+1}$$

Varianza del error de predicción:

$$\text{Var}(e_T(1)) = \sigma_\varepsilon^2$$

$$\text{Var}(e_T(2)) = [1 + (\phi - \theta)^2]\sigma_\varepsilon^2$$

$$\text{Var}(e_T(3)) = [1 + (\phi - \theta)^2(1 + \phi^2)]\sigma_\varepsilon^2$$

$$\text{Var}(e_T(k)) = \left[ 1 + (\phi - \theta)^2 \frac{1 - \phi^{2(k-1)}}{1 - \phi^2} \right] \sigma_\varepsilon^2$$

que, al igual que en los modelos AR(*p*) y MA(*q*), tiende a la varianza del proceso ARMA(1, 1) cuando el horizonte de predicción *k* tiende a infinito.

Obsérvese que las amplitudes del intervalo de confianza para las sucesivas predicciones elaboradas con un modelo AR crecen continuamente con el horizonte de predicción, mientras que para un modelo MA(*q*) permanecen constantes a partir de *q* períodos hacia el futuro.

El lector puede notar asimismo que el error de predicción un período hacia el futuro es el mismo,  $\varepsilon_{T+1}$ , independientemente de cual sea el verdadero modelo de la serie temporal en estudio. Este es un resultado lógico, pues dicho error es la componente del valor futuro de la serie que no puede predecirse sobre la base de la información muestral, lo que en la literatura de series temporales se conoce como *innovación en la serie*.

Ello no quiere decir, sin embargo, que el error de predicción de una serie temporal un período hacia el futuro vaya a ser independiente del modelo utilizado para obtener dicha predicción. El error de predicción, por definición, es la diferencia entre la realización de la serie y su *predicción óptima*, obtenida con el *mejor* modelo que pueda elaborarse sobre la base de la información muestral. En la práctica, estas diferencias se reflejan en diferentes valores estimados del parámetro  $\sigma_\varepsilon^2$  según se ajuste un modelo u otro. Dicho de otro modo, si no se utiliza el *mejor* modelo para la serie, entonces el error de predicción incluirá, junto con la innovación  $\varepsilon_t$ , un error de especificación  $\xi_t$  y, agregadamente, tendrán una varianza superior a la de la auténtica innovación.

### 13.9.e. Intervalos de confianza para las predicciones

La varianza del error de predicción puede utilizarse para obtener *intervalos de confianza de las predicciones* elaboradas, mediante la expresión:

$$E_T y_{T+k} \pm \lambda_\alpha \hat{\sigma}_{e_T(k)}$$

donde, si se supone que la innovación  $\varepsilon_t$  sigue una distribución Normal, el parámetro  $\lambda_\alpha$  se obtendrá de las tablas de dicha distribución, al nivel de confianza  $\alpha$  elegido. Así, por ejemplo, para obtener en el instante  $T$  un intervalo del 95 por 100 de confianza para el valor del proceso  $y_t$  en el instante  $T+k$ , supuesto que la innovación  $\varepsilon_{T+k}$  siga una distribución Normal, basta sumar y restar 1,96 veces el valor estimado de  $\sigma_{e_T(k)}$  a la predicción  $E_T y_{T+k}$ . Como es habitual, este intervalo será tan sólo una aproximación, por haber estimado el parámetro  $\hat{\sigma}_e$ .

### 13.9.f. Predicción de una serie en diferencias

Si se ha estimado un modelo ARIMA con un número no nulo de diferencias, entonces será preciso recuperar las predicciones de la serie original a partir de las predicciones elaboradas para la serie en diferencias. Ello puede hacerse del siguiente modo: Supongamos que  $y_t$  denota la serie en cuyo análisis estamos interesados, y que se ha especificado y estimado un modelo univariante para la serie de primeras diferencias:  $z_t = \nabla y_t$ . Entonces, es claro que:

$$E_T z_{T+k} = E_T y_{T+k} - E_T y_{T+k-1}$$

por lo que:

$$\begin{aligned} E_T y_{T+k} &= E_T z_{T+k} + E_T y_{T+k-1} = \dots = \\ &= E_T z_{T+k} + E_T z_{T+k-1} + E_T z_{T+k-2} + \dots + E_T z_{T+1} + y_T \end{aligned}$$

y, consecuentemente, para obtener las predicciones de la serie  $y_t$  basta añadir a su última observación muestral las predicciones elaboradas para sus incrementos,  $z_{T+k}$ .

El lector puede comprobar, como se pide en uno de los ejercicios al final del capítulo, que la recuperación de las predicciones de una serie  $y_t$  a partir de las predicciones elaboradas para la serie de segundas diferencias  $\nabla^2 y_t$  puede llevarse a cabo de un modo similar al que acabamos de describir.

Por último cabe citar que si el modelo univariante se ha estimado para la transformación logarítmica de la variable original  $z_t = \ln(y_t)$ , entonces el modo de recuperar las predicciones de los valores futuros de  $y_t$  es:

$$E_T y_{T+k} = \exp \left\{ E_T z_{T+k} + \frac{1}{2} \text{Var } e_T(k) \right\}$$

Dependiendo del tamaño de la varianza del error de predicción, la expresión anterior supondrá una diferencia significativa o no con respecto a la simple alternativa de hallar la función exponencial de los valores  $E_T z_{T+k}$ .

Sin embargo, los límites inferior y superior de los intervalos de confianza para  $E_T y_{T+k}$  deben hallarse calculando los valores de la función exponencial en los límites inferior y superior del intervalo de  $E_T z_{T+k}$ . En consecuencia, el intervalo de confianza que se obtenga para la predicción  $E_T y_{T+k}$  no será simétrico alrededor de dicha predicción.

Nótese, por otra parte, que las variaciones en las predicciones de  $z_{T+k}$  para valores sucesivos de  $k$  pueden interpretarse como variaciones porcentuales previstas en la variable original  $y_{t+k}$ .

## 13.10. ESTIMACION DE MODELOS ARIMA

### 13.10.a. Estimación de modelos autorregresivos

Un modelo autorregresivo presenta una diferencia notable con respecto a los modelos econométricos que hasta ahora hemos considerado. Las variables explicativas son ahora aleatorias, ya que son retardos de la variable  $y_t$ , que es aleatoria. Como hemos visto en el Capítulo 9, puede probarse que, en tales condiciones, el estimador MCO tiene buenas propiedades y, en particular, es un estimador consistente siempre que las variables explicativas  $x_{it}$  satisfagan la condición  $E(x_{i,t-s}u_t) = 0$ .

Si el término de error no tiene autocorrelación y si el modelo es estacionario, esta propiedad se satisface. En efecto, los valores  $x_{i,t-s}$  de la condición anterior son ahora retardos de  $y_t$ , mientras que  $u_t$  es ahora  $\varepsilon_t$ . Como hemos visto en dicho capítulo, bajo el supuesto de estacionariedad, la variable  $y_t$  depende de  $\varepsilon_t$  y de sus valores anteriores, pero de ningún valor futuro de  $\varepsilon_t$ . Por tanto, las esperanzas  $E(y_{t-s}\varepsilon_t)$  son cero para todo  $s > 0$ .

En tal caso, la estimación consistente del modelo autorregresivo

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

puede llevarse a cabo por MCO. Por tanto, todo depende de que la especificación sea correcta.

Si, por el contrario, el término de error del modelo tuviese autocorrelación, entonces la condición de ortogonalidad no se cumpliría y el estimador MCO dejaría de ser apropiado. En efecto, si  $\varepsilon_t$  no fuese ruido blanco sino que obedeciese, por ejemplo, al modelo  $\varepsilon_t = \phi \varepsilon_{t-1} + \zeta_t$ , con  $\zeta_t$  ruido blanco, entonces se tendría que: a)  $y_{t-1}$  estaría correlacionado con  $\varepsilon_{t-1}$ , a través del modelo univariante elaborado, y b)  $\varepsilon_t$  también estaría correlacionado con  $\varepsilon_{t-1}$ , a través de su modelo de autocorrelación. En estas condiciones,  $E(y_{t-1}\varepsilon_t) \neq 0$ , ya que ambos están correlacionados con  $\varepsilon_{t-1}$ , contradiciendo la condición de ortogonalidad necesaria para justificar el uso del estimador MCO.

La autocorrelación del término de error de un modelo univariante es un

indicio evidente de mala especificación de dicho modelo. Una especificación correcta debe generar un término de error con estructura de ruido blanco.

### 13.10.b. Estimación de modelos de medias móviles

Como ejemplo de la estimación de modelos de medias móviles, discutiremos las cuestiones importantes en el contexto de un modelo MA(2):  $y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$ . Para ello, si suponemos que la innovación  $\varepsilon_t$  sigue una distribución Normal  $(0, \sigma_\varepsilon^2)$ , se tiene la siguiente expresión para el logaritmo de la función de verosimilitud:

$$\log L(\theta_1, \theta_2, \sigma_\varepsilon^2) = \text{Constante} - \frac{T}{2} \log \sigma_\varepsilon^2 - \frac{\text{SR}(\theta_1, \theta_2)}{2\sigma_\varepsilon^2}$$

donde  $\text{SR}(\theta_1, \theta_2)$  denota la suma residual:  $\text{SR}(\theta_1, \theta_2) = \sum_3^T \varepsilon_t^2$ , de modo que, condicional en el valor del parámetro  $\sigma_\varepsilon^2$ , la maximización de la función de verosimilitud coincide con la minimización de la suma residual. Se trataría, por tanto, de minimizar la suma de cuadrados de los residuos:

$$\text{mín } \text{SR}(\theta_1, \theta_2) = \sum_3^T \varepsilon_t^2 = \sum_3^T (y_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2})^2$$

donde, como puede verse, se están ignorando los dos primeros períodos de la muestra, pues no se dispone para ellos de observaciones acerca del ruido blanco  $\varepsilon_t$ . La solución a este problema de minimización proporciona los valores estimados de los parámetros del modelo y, finalmente, la varianza  $\sigma_\varepsilon^2$  se estima mediante:

$$\sigma_\varepsilon^2 = \frac{\text{SR}(\hat{\theta}_1, \hat{\theta}_2)}{T - p - q}$$

donde  $p + q$  es el número de parámetros estimados en el modelo univariante. en este caso 2.

La minimización de la suma residual anterior podría llevarse a cabo mediante una red de búsqueda. Este procedimiento sería muy sencillo en el caso de un proceso MA(1), puesto que el espacio paramétrico admisible sería  $(-1, 1)$ , y bastaría hacer una partición de este intervalo, generar la serie de residuos  $\{\hat{\varepsilon}_t\}$  tomando como dadas las observaciones  $y_1$  y  $y_2$ , y evaluando la función  $\text{SR}(\theta)$ . La certeza de haber hallado un mínimo global en vez de un mínimo local, aumenta si la partición que se lleva a cabo es suficientemente fina, o si se lleva a cabo una exploración bastante exhaustiva del espacio paramétrico. En el caso de un proceso MA(2), el proceso de búsqueda es más complejo, pues el espacio paramétrico tiene dos dimensiones. Sin embargo, dicho espacio queda limitado por las condiciones de invertibilidad del proceso que vimos anteriormente.

Como alternativa, dados unos valores iniciales de los parámetros  $\theta_1$  y  $\theta_2$ , la expresión

$$\varepsilon_t = y_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad [13.6]$$

podrá utilizarse para generar «observaciones» de  $\varepsilon_t$  para  $t = 3, 4, \dots, T$ . Claramente, los valores numéricos de los residuos  $\hat{\varepsilon}_t \geq 3$  varían con los valores de los parámetros  $\theta_1$  y  $\theta_2$ , y se trata de encontrar el par de valores  $(\theta_1, \theta_2)$  para el que la suma de cuadrados de los residuos es mínima. Es imposible utilizar todos los pares de valores paramétricos posibles y calcular para cada uno de ellos la suma residual correspondiente, por lo que es preciso recurrir a un procedimiento analítico, que pasamos a describir.

Denotamos por  $(\theta_1^0, \theta_2^0)$  los valores iniciales de los parámetros y por  $\varepsilon_t^0$ ,  $t \geq 3$  los residuos obtenidos con el par de valores  $(\theta_1^0, \theta_2^0)$ . El verdadero valor del residuo  $\varepsilon_t$  es, sin embargo, el correspondiente a los verdaderos valores de los parámetros  $\theta_1$  y  $\theta_2$ , y se puede llevar a cabo la siguiente aproximación por medio del desarrollo en serie de Taylor:

$$\begin{aligned} \varepsilon_t = & \varepsilon_t^0 + (\theta_1 - \theta_1^0) \left[ \frac{\partial \varepsilon_t}{\partial \theta_1} \right]_{\theta = \theta^0} + (\theta_2 - \theta_2^0) \left[ \frac{\partial \varepsilon_t}{\partial \theta_2} \right]_{\theta = \theta^0} + \\ & + \frac{1}{2} (\theta_1 - \theta_1^0)^2 \left[ \frac{\partial^2 \varepsilon_t}{\partial \theta_1^2} \right]_{\theta = \theta^0} + \frac{1}{2} (\theta_1 - \theta_1^0)(\theta_2 - \theta_2^0) \left[ \frac{\partial^2 \varepsilon_t}{\partial \theta_1 \partial \theta_2} \right]_{\theta = \theta^0} + \\ & + \frac{1}{2} (\theta_2 - \theta_2^0)^2 \left[ \frac{\partial^2 \varepsilon_t}{\partial \theta_2^2} \right]_{\theta = \theta^0} + \dots \end{aligned} \tag{13.7}$$

donde ya hemos despreciado los términos de orden superior a 2, en la confianza de que tanto las potencias  $(\theta_1 - \theta_1^0)^3$  como las derivadas parciales de esos órdenes serán pequeñas en valor absoluto. Si por idénticas razones despreciamos asimismo los términos de segundo orden que aparecen en [13.7], se llega a

$$\varepsilon_t = \varepsilon_t^0 + (\theta_1 - \theta_1^0) \left[ \frac{\partial \varepsilon_t}{\partial \theta_1} \right]_{\theta = \theta^0} + (\theta_2 - \theta_2^0) \left[ \frac{\partial \varepsilon_t}{\partial \theta_2} \right]_{\theta = \theta^0}$$

y puesto que en este modelo

$$\frac{\partial \varepsilon_t}{\partial \theta_1} = \varepsilon_{t-1} \quad \text{y} \quad \frac{\partial \varepsilon_t}{\partial \theta_2} = \varepsilon_{t-2}$$

se tiene:

$$\varepsilon_t^0 - \theta_1^0 \varepsilon_{t-1}^0 - \theta_2^0 \varepsilon_{t-2}^0 = -\theta_1 \varepsilon_{t-1}^0 - \theta_2 \varepsilon_{t-2}^0 + \varepsilon_t$$

que muestra que se pueden obtener estimaciones de los parámetros  $\theta_1$  y  $\theta_2$  sin más que estimar el modelo lineal:

$$w_t = \theta_1 x_{1t} + \theta_2 x_{2t} + \varepsilon_t \tag{13.8}$$

donde  $w_t = \varepsilon_t^0 - \theta_1^0 \varepsilon_{t-1}^0 - \theta_2^0 \varepsilon_{t-2}^0$ ,  $x_{1t} = -\varepsilon_{t-1}^0$ ,  $x_{2t} = -\varepsilon_{t-2}^0$ .

El lector debe notar que este procedimiento no es sino la aplicación del algoritmo Gauss-Newton a la minimización de la suma residual condicional.

La estimación del modelo [13.8] requiere de la previa construcción de estas variables. Los valores  $\theta_i^0$ ,  $i = 1, 2$  que en ellas aparecen son los valores iniciales de los parámetros que el investigador debe seleccionar. Pongamos, por ejemplo, que  $\theta_i^0 = 0,10$ , aunque más adelante daremos unas reglas más rigurosas para la elección de estos valores iniciales. Por otra parte, la relación [13.6] puede utilizarse para generar «datos» para la variable  $\hat{\varepsilon}_t^0$ , para  $t = 3, 4, \dots, T$ , utilizando los valores iniciales de los parámetros. La estimación del modelo [13.8] se llevaría a cabo con  $T - 3$  observaciones.

El procedimiento descrito puede y debe iterarse, utilizando como valores  $(\theta_1^0, \theta_2^0)$  en una segunda etapa las estimaciones obtenidas en la primera etapa. El esquema iterativo podría terminar cuando las diferencias entre los valores paramétricos iniciales y finales de una etapa del proceso son pequeñas. En tal caso, se dice que el proceso de estimación numérica que hemos descrito ha convergido (supuestamente a los verdaderos valores de los parámetros). Como posible criterio de convergencia, podría decidirse detener el proceso si las variaciones en todos los parámetros son inferiores a 0,001, o al 5 por 100 de su valor inicial.

Un criterio de convergencia alternativo consistiría en detener el proceso si la variación producida en la suma residual (recordemos que ésta es la función objetivo del problema de optimización) es pequeña. De nuevo, habría que definir lo que se entiende por «pequeña», pero podría ser una diferencia inferior al 1 por 100. Un criterio más exigente sería el de aceptar la convergencia del proceso sólo cuando se cumplen simultáneamente los dos criterios que acabamos de citar.

Por supuesto, también existe la posibilidad de que el proceso anterior no converja. Ello puede deberse al hecho de que el procedimiento de estimación se basa en una aproximación lineal a la función que hace depender  $\varepsilon_t$  de los parámetros del modelo, y no en la verdadera relación entre ambos. Otra posibilidad es que los valores iniciales escogidos para los parámetros no hayan sido muy adecuados por lo que, cuando esto ocurre, debe repetirse el procedimiento con otros valores iniciales. Finalmente, la convergencia puede no producirse por una mala especificación del modelo, es decir, porque el modelo univariante que se está tratando de estimar no sea el modelo lineal que mejor representa la estructura del proceso estocástico que generó la serie temporal que está siendo objeto de análisis.

El procedimiento que acabamos de describir es válido, con las modificaciones obvias, para la estimación de cualquier modelo univariante. Al estimar el modelo

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

se ignoran los primeros  $q$  residuos, lo que equivale a suponerlos iguales a su esperanza, cero. Por ello, este procedimiento de estimación se conoce como *mínimos cuadrados condicionados* al supuesto de que los primeros residuos son iguales a cero. Si la muestra consta de pocas observaciones, el número

de grados de libertad (número de observaciones menos número de parámetros a estimar) puede ser muy pequeño, e ignorar las primeras observaciones en la estimación será una mala aproximación.

### 13.10.c. Obtención de valores iniciales para los parámetros del modelo

Para llevar a cabo el procedimiento analítico de estimación que hemos descrito en los párrafos anteriores es preciso haber identificado previamente un modelo para la serie temporal en estudio. Esta identificación se habrá llevado a cabo mediante el análisis de sus funciones de autocorrelación y de autocorrelación parcial. Los valores numéricos de las estimaciones muestrales de estas funciones pueden utilizarse para obtener estimaciones iniciales de los parámetros del modelo.

Por ejemplo, supongamos que el examen de la función de autocorrelación parcial de una serie sugiere que el proceso que la ha generado tiene una estructura autorregresiva pura. En tal caso, las ecuaciones de Yule-Walker permiten obtener valores de los parámetros  $\phi_i$  a partir de los valores estimados de la función de autocorrelación. Nótese que si las estimaciones muestrales de la función de autocorrelación fuesen muy próximas a los verdaderos valores de dicha función, entonces la solución al sistema de Yule-Walker proporcionaría los valores de los parámetros del modelo, sin necesidad de utilizar el proceso analítico antes descrito. Sin embargo, la estimación de la función de autocorrelación no suele tener tan buenas propiedades, y las primeras iteraciones del procedimiento anterior producen, en general, variaciones significativas.

Si, por el contrario, se ha especificado un modelo MA(1), entonces, recordando que  $\rho_1 = -\frac{\theta}{1 + \theta^2}$ , esta ecuación puede invertirse para obtener el valor de  $\theta$ . La ecuación que así resulta es de segundo grado y tiene, por tanto, dos soluciones; en general, una de las raíces es mayor, en valor absoluto, y la otra menor que 1. Para garantizar la invertibilidad del proceso, escogemos siempre la raíz inferior a la unidad en valor absoluto.

Para procesos de orden superior, la obtención de valores iniciales de los parámetros es más compleja, pero puede intentarse un procedimiento similar al descrito.

Una alternativa consistiría en obtener la representación autorregresiva del proceso MA, y estimarla. Este método tiene la gran ventaja de su sencillez. Sin embargo, ya hemos visto que la representación AR de cualquier proceso MA de orden finito es de orden infinito. Por tanto, habría que truncar la autorregresión. Ello no es un problema excesivo, puesto que los coeficientes en los sucesivos retardos son potencias de los coeficientes  $\phi$ , que son menores que 1 en valores absoluto. Pero además, una vez truncada la regresión, los coeficientes restantes están sujetos a restricciones no lineales, que habría que imponer si se pretende obtener un estimador eficiente. Para ilustrar esta

problemática, consideremos el modelo MA(1), cuya representación autorregresiva es:

$$y_t = \varepsilon_t - \theta y_{t-1} - \theta^2 y_{t-2} - \theta^3 y_{t-3} - \dots$$

Si el parámetro  $\theta$  fuese suficientemente pequeño como para que

$$y_t = -\theta y_{t-1} + u_t$$

fuese una buena aproximación, entonces el coeficiente de  $y_{t-1}$  sería una estimación bastante aproximada del parámetro  $\theta$ . Hay que hacer notar, sin embargo, que tal regresión no dejaría de ser una mala especificación de un modelo que es, realmente, una autorregresión de orden infinito. De este modo, el término de error  $u_t$  incorporaría los retardos omitidos, y se produce el problema de correlación entre el regresor  $y_{t-1}$  y  $u_t$  al que antes hicimos mención. También es cierto que la trascendencia de tal correlación depende de la magnitud de los coeficientes de los retardos omitidos, es decir, de la calidad de la aproximación anterior.

Dicha aproximación puede mejorarse, y con ello aminorar el sesgo producido por la correlación entre  $y_{t-1}$  y  $u_t$ , si se incluye algún otro retardo, por ejemplo:

$$y_t = -\theta y_{t-1} - \theta^2 y_{t-2} + u_t$$

pero como vemos, ya aparece una restricción no lineal entre los coeficientes de  $y_{t-1}$  e  $y_{t-2}$ .

### 13.11. DIAGNOSTICO DEL MODELO

Tras estimar un modelo ARIMA, es *esencial* llevar a cabo un análisis de los coeficientes y residuos del modelo, con el objetivo de detectar posibles indicios de mala especificación. Respecto a los coeficientes, éstos deben de satisfacer siempre las condiciones de estacionariedad e invertibilidad. Cuando se estiman términos AR(2) o MA(2), deben obtenerse siempre sus raíces, y factorizar dichos términos cuando las raíces sean reales, por la razón que en seguida veremos. También debe hacerse hincapié en que los coeficientes estimados sólo serán relevantes si el algoritmo numérico de estimación ha convergido. Si no lo ha hecho, puede concederse un número mayor de iteraciones, pero puede también ser un indicio de ausencia de estacionariedad o de invertibilidad y, por tanto, de un deficiente ajuste del modelo especificado a las series analizadas.

Asimismo, deben examinarse cuidadosamente los residuos resultantes hasta que se consiga, en la medida de lo posible, eliminar toda duda acerca de que éstos obedecen un proceso de ruido blanco. Esta verificación es crucial, pues sobre tal supuesto se habrá diseñado la estrategia de estimación y predicción, como hemos expuesto en las secciones anteriores. Cualquier evidencia en

contra de la hipótesis de ruido blanco para los residuos constituye un indicio de mala especificación del modelo.

### 13.11.a. Análisis de residuos

Ya vimos en la Sección 13.5 cómo unos residuos MA(1) obtenidos en un modelo AR(2) sugieren un modelo más correcto ARMA(2, 1). Como otro ejemplo, supongamos que se ha especificado y estimado el modelo:

$$y_t = \delta + \phi_1 y_{t-1} + u_t \quad [13.9]$$

cuyos residuos parecen obedecer el modelo:

$$u_t = \phi_2 u_{t-1} + \varepsilon_t \quad [13.10]$$

donde, a diferencia de  $u_t$ ,  $\varepsilon_t$  es ruido blanco. Una simple sustitución del modelo [13.10] en [13.9] muestra que el verdadero modelo de  $y_t$  es:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \varepsilon_t$$

donde  $\beta_0 = \delta(1 - \phi_2)$ ,  $\beta_1 = \phi_1 + \phi_2$  y  $\beta_2 = -\phi_1 \phi_2$ . Es decir, el proceso  $y_t$  tiene una estructura AR(2) aunque, por error, se había especificado un modelo AR(1).

Sin embargo, si el término de error del modelo [13.9] tuviese una estructura de media móvil:  $u_t = \varepsilon_t + \theta \varepsilon_{t-1}$ , entonces la sustitución directa de esta estructura en [13.9] muestra que, en realidad,  $y_t$  sigue una estructura ARMA(1, 1).

Supuesto que se ha identificado correctamente una estructura AR, la elección del orden se basa en contrastes de significación llevados a cabo con los valores estimados de la función de autocorrelación parcial. Cuando el verdadero modelo es AR( $p$ ), la distribución de los valores estimados de la fap es, aproximadamente:  $\hat{\phi}_{jj} \sim N\left(0, \frac{1}{T}\right)$  para  $j > p$ , por lo que si  $\hat{\phi}_{jj}$  está en el

intervalo  $\pm \frac{2}{\sqrt{T}}$  para todos los valores de  $j$  superiores a un cierto  $p$ , entonces no se rechazará la hipótesis de que el orden del modelo AR es menor o igual que  $p$ .

Un procedimiento similar permite identificar el orden de un proceso MA( $q$ ). Como sabemos, un proceso MA( $q$ ) se caracteriza porque su función de autocorrelación simple teórica es cero para valores  $k > q$ . Se trata, por tanto, de efectuar contrastes de hipótesis de significación estadística acerca de los valores estimados de dicha función. En un proceso de ruido blanco, las estimaciones  $r_k$  se distribuyen asintóticamente  $N\left(0, \frac{1}{T}\right)$ . Así, se acepta la no

significación de las estimaciones  $r_k$  si caen dentro del intervalo  $\pm \frac{2}{\sqrt{T}}$ . Si ello

ocurre para todo  $k$  mayor que un cierto  $q$ , pensamos que el proceso es MA( $q$ ). Si ocurre para todo  $k$ , entonces se mantiene el supuesto de que la serie temporal de datos proviene de un proceso de ruido blanco. Si el proceso no es ruido blanco, entonces el cociente  $\frac{1}{\sqrt{T}}$  puede no ser una aproximación suficientemente buena a la desviación típica de  $r_k$ . Bartlett [1946] ha probado que si  $\rho_k$  es distinto de cero para  $k \leq q$  e igual a cero para  $k > q$ , entonces la varianza de  $r_k$  es aproximadamente igual a:

$$\text{Var}(r_k) = \frac{1 + 2(\rho_1^2 + \rho_2^2 + \dots + \rho_q^2)}{T}$$

En particular, para contrastar la hipótesis nula de que el proceso estocástico subyacente tiene estructura de ruido blanco, como ocurre cuando aplicamos este procedimiento a los residuos de un modelo estimado, la expresión anterior se reduce a:

$$\text{Var}(r_k) = \frac{1}{T}$$

que es la justificación para utilizar  $\pm \frac{2}{\sqrt{T}}$  como intervalo de confianza del 95 por 100 para los valores estimados de la función de autocorrelación. Como acabamos de decir, tal práctica sólo está justificada si la hipótesis nula que se contrasta es la de ruido blanco. Si, por el contrario, ya se ha aceptado previamente que la serie en análisis proviene de un proceso con una cierta estructura estocástica, y se está considerando la posibilidad de añadir más estructura, entonces la práctica anterior sólo puede entenderse como una aproximación sencilla al contraste dado por la expresión anterior para la  $\text{Var}(r_k)$ .

La fórmula de Bartlett es únicamente una aproximación a la verdadera varianza de los valores estimados de la función de autocorrelación. Dicha aproximación tiende a sobrestimar el valor de la varianza, generando así unos intervalos de confianza de un tamaño mayor del que realmente debieran tener. En consecuencia, la utilización de la fórmula de Bartlett tiende a mantener la hipótesis nula de no significación de  $r_k$  más a menudo de lo que debiera y, con ello, a no detectar estructura estocástica en casos en que dicha estructura existe.

Conviene, por tanto, prestar especial atención a la posible detección de alguna regularidad en la función de autocorrelación que la asemeje a alguna de las vistas en este capítulo, *incluso* si sus valores no son significativos. En especial, conviene ser exigente con las estimaciones de los primeros valores de dicha función, y utilizar para ellos un intervalo de confianza de aproximadamente  $\pm 1,25$  o  $\pm 1,5$  desviaciones típicas, en vez de las habituales dos desviaciones típicas. En definitiva,  $\frac{1}{T}$  no es sino una *cota superior* para la varianza de dichos valores.

En ocasiones, para formalizar el contraste de la hipótesis nula  $H_0$ : «los residuos del modelo son ruido blanco», se resume toda su estructura de autocorrelación en el estadístico de Box-Pierce:

$$Q = T(r_1^2 + r_2^2 + \dots + r_k^2)$$

que se distribuye como una chi-cuadrado con  $k - p - q$  grados de libertad, donde  $p$  y  $q$  son el número de parámetros AR y MA estimados. Con el objeto de ganar potencia, se ha propuesto también el estadístico modificado por Ljung y Box:

$$Q' = T(T + 2) \sum_1^k \frac{r_j^2}{T - j}$$

que tiene distribución chi-cuadrado con  $k$  grados de libertad.

De igual modo, el gráfico de residuos puede ilustrar la presencia de posibles tendencias que no apareciesen obvias en la fas. Recordemos que, generalmente, en situaciones de no estacionariedad, la fas decrece hacia cero sólo lentamente, pero, en ocasiones, los residuos muestran una clara tendencia y, sin embargo, sólo los valores iniciales de la fas están fuera del intervalo de confianza del 95 por 100.

Cuando se analiza una variable estacional, debe prestarse especial atención a los primeros valores de las fas y fap, pero también a los primeros valores de orden estacional. Así, con una serie trimestral, el analista debe observar detenidamente los valores 1, 2, 3, 4 y 8 de ambas funciones, a los valores 1, 2, 3, 4, 5, 6, 12 y 24 si trabaja con una serie mensual. Todo ello sin menoscabo de examinar asimismo el resto de los valores.

### 13.11.b. Sobreparametrización y sobrediferenciación

Dos aspectos de la identificación de un modelo univariante están muy relacionados entre sí: a) la posible *sobreparametrización*, por existencia de factores comunes, y b) la posible *sobrediferenciación* del proceso. En primer lugar, debe notarse que los modelos:

$$\nabla^2 y_t = (1 - 0,97L)\varepsilon_t \quad [13.11]$$

$$\nabla y_t = \varepsilon_t \quad [13.12]$$

son prácticamente indistinguibles, puesto que como  $\nabla = (1 - 1,0L)$ , existe un factor común en ambos miembros de [13.11]. De modo similar, son modelos aproximados:

$$(1 - 0,95L)y_t = (1 - 0,20L)\varepsilon_t \quad [13.13]$$

$$\nabla y_t = (1 - 0,20L)\varepsilon_t \quad [13.14]$$

puesto que un término AR(1) con coeficiente positivo y próximo a 1 sugiere la no estacionariedad de la variable  $y$ , por ello, la conveniencia de diferenciarla. Por último, los modelos

$$(1 - 1,20L + 0,35L^2)\nabla y_t = (1 - 0,70L)\varepsilon_t \quad [13.15]$$

$$(1 - 0,50L)\nabla y_t = \varepsilon_t \quad [13.16]$$

son indistinguibles, por cuanto que la factorización del polinomio AR(2):  $(1 - 1,20L + 0,35L^2) = (1 - 0,70L)(1 - 0,50L)$ , muestra la existencia de un factor común en ambos miembros de [13.5] que no era evidente a simple vista. Como norma general, el analista siempre debe obtener las raíces del polinomio cuando estima un término AR(2).

En ocasiones, no resulta evidente cuál es el orden correcto de diferenciación de una serie. A este respecto, no hay una doctrina clara: algunos investigadores prefieren no diferenciar por miedo a eliminar información relevante; otros temen que no diferenciar deje en la serie aspectos de no estacionariedad que sesguen el proceso de elaboración del modelo. Esta última estrategia parece, sin embargo, dar mucho mejores resultados. El analista no debe temer introducir una diferencia adicional en la variable, especialmente si incorpora simultáneamente un término de media móvil. Por ejemplo, pasaría del modelo:

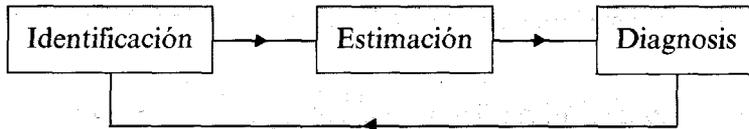
$$(1 + \phi_1 L + \phi_2 L^2)\nabla y_t = \varepsilon_t \quad [13.17]$$

Sobre cuyos residuos existen sospechas de no estacionariedad, al modelo:

$$(1 + \phi_1 L + \phi_2 L^2)\nabla^2 y_t = (1 - \theta L)\varepsilon_t \quad [13.18]$$

y si el parámetro  $\theta$  se aproxima mucho a 1, puede volverse al modelo [13.7]. Sin embargo, siempre que el valor de  $\theta$  no produzca un modelo claramente no invertible, conviene mantener la especificación [13.8], pues la combinación de diferencia y término de media móvil suele producir buenos resultados predictivos.

Cuando se han detectado factores comunes, no debe simplemente mantenerse el modelo tras la simplificación pertinente, sino que debe volverse a estimar, al igual que ocurre con cualquier evidencia de mala especificación que se obtuviese de los residuos o cuando se decide introducir una diferencia adicional. En definitiva, el proceso de elaboración de un modelo univariante tiene varias etapas:



que sólo finalizan cuando la diagnosis de un modelo estimado no sugiere indicación de mala especificación.

### 13.11.c. Valores influyentes y anomalías. Análisis de intervención

Otro aspecto importante en la elaboración de modelos univariantes lo constituye la posible existencia de valores influyentes. Estos son valores de la serie que se analiza que, por su magnitud, *distorsionan* el proceso iterativo de construcción del modelo. Piénsese que, en general, el modelo se especifica para la variable diferenciada al menos una vez, puesto que prácticamente ninguna variable stock, y tan sólo pocas variables flujo o precios, son estacionarias. Por consiguiente, un aumento o disminución mensual, trimestral o anual, según la frecuencia de observación de la serie, de importante magnitud, pueden resultar valores influyentes.

Hay dos posibles estrategias frente a este tipo de valores: Por una parte, pueden tratarse de igual modo que el resto de las observaciones de la serie. Por otro lado, puede *estimarse* su influencia para, en cierto modo, descontar su posible efecto en la elaboración del modelo univariante.

En algunos casos, una intervención de política económica tiene un importante efecto sobre la variación en la oferta monetaria o los tipos de interés que podría hacer dichos valores influyentes. Precisamente porque se sabe que han sido forzados por razones exógenas al proceso generador de la serie, que estamos tratando de elaborar, requieren un trato especial, por lo que debe estimarse su efecto. Algo análogo debe hacerse con el impacto de condiciones meteorológicas especialmente adversas sobre la cosecha de un producto agrícola, cuya producción se pretende modelizar, o con la ocurrencia de una huelga en un sector industrial cuya actividad productiva se quiere prever.

Por tanto, hay muchas instancias en que la estimación de tales efectos por separado del resto del modelo está justificada. A esta estrategia se le denomina *Análisis de intervención*. Cuando se desconoce la causa de un valor influyente, no procede, en general, llevar a cabo su *intervención*. Sin embargo, puede intervenir dicho valor con objeto de que no distorsione la especificación de la estructura estocástica de la serie, para comprobar posteriormente si la estructura que se ha identificado se mantiene al no intervenir el valor influyente, aunque haya que estimar nuevamente el modelo.

Las *intervenciones* efectuadas sobre un valor influyente son, generalmente, de dos tipos. Si  $y_{t_0}$  es un valor influyente, definimos una *variable impulso* en  $t_0$  como:

$$\zeta_{t_0}^i = \begin{cases} 1 & \text{si } t = t_0 \\ 0 & \text{si } t \neq t_0 \end{cases}$$

y una *variable escalón* en  $t_0$  como:

$$\xi_{t_0}^s = \begin{cases} 0 & \text{si } t < t_0 \\ 1 & \text{si } t \geq t_0 \end{cases}$$

El análisis de intervención consiste en introducir una de estas variables en el modelo univariante, y estimar su coeficiente. Se incluye una variable impulso cuando el valor influyente se ha producido tan sólo en  $t_0$ . Una variable escalón es más adecuada cuando los valores de la serie (o de sus tasas de cambio) son sistemáticamente mayores o menores después de  $t_0$  que antes de dicho instante o, dicho de otro modo, cuando se ha producido un cambio permanente en el valor medio de la variable, a diferencia del impulso, que corresponde a un cambio meramente transitorio. Cuando la media ha aumentado, el coeficiente asociado a  $\xi_{t_0}^s$  será positivo, siendo negativo en caso contrario.

### 13.12. MODELOS DE FUNCION DE TRANSFERENCIA

La representación ARIMA univariante de una serie temporal correspondiente a una variable  $Y$  puede generalizarse para incorporar otras variables  $X$  como explicativas. El modelo resultante se conoce como *función de transferencia*, con las variables  $X$  como *input* e  $Y$  como *output*. Centramos nuestra presentación en el caso de un solo input, pues la extensión a múltiples inputs, tanto en términos de especificación como de estimación, es inmediata.

Conviene especificar un modelo de función de transferencia cuando:

1. Se espera que la relación entre input y output, tanto a través de sus componentes regulares como estacionales, tenga una *estructura dinámica* suficientemente rica como para que la representación econométrica habitual requiriese de un número elevado de parámetros.
2. Se cree que una representación adecuada de la estructura estocástica del *término de error* resultante de la relación de  $X$  hacia  $Y$  precisa de una modelización ARIMA, pues no estaría suficientemente recogida por los sencillos esquemas de autocorrelación utilizados en el modelo lineal general.

Un modelo de función de transferencia tiene varias *ventajas* adicionales:

1. Permite una representación *parsimoniosa*, es decir, basada en un número reducido de coeficientes, de relaciones dinámicas, incluso si son muy complicadas.
2. Dispone de una estrategia sencilla y gradual para la especificación de un modelo dinámico de relación que capture adecuadamente el efecto del input sobre el output.
3. Proporciona instrumentos para comprobar que se utilizan como input y output variables que tienen análogas características de estacionariedad, de modo que los residuos del modelo son estacionarios.

Esta propiedad, crucial para justificar el análisis de inferencia que con el modelo estimado pudiera llevarse a cabo, ha generado una considerable cantidad de estudios recientes, destinados a generar procedimientos de contraste de estacionariedad, así como de especificación, ambos en el contexto del modelo econométrico lineal, que serán objeto de análisis en el Capítulo 14.

La representación genérica del modelo de función de transferencia con un input es:

$$\begin{aligned} Y_t &= v(L)X_t + N_t = (v_0 + v_1L + v_2L^2 + \dots)L^b X_t + N_t = \\ &= \frac{\omega(L)}{\delta(L)} X_{t-b} + N_t = \delta^{-1}(L)\omega(L)X_{t-b} + N_t = \quad [13.19] \\ &= \frac{\omega_0 - \omega_1L - \dots - \omega_sL^s}{1 - \delta_1L - \dots - \delta_rL^r} X_{t-b} + N_t \end{aligned}$$

donde, al igual que en secciones previas,  $L$  denota el operador de retardos. Los polinomios  $\delta(L)$  y  $\omega(L)$  se denominan *autorregresivo* y de *medias móviles*, respectivamente, mientras que  $N_t$  es la *perturbación* del modelo. El polinomio  $v(L)$  se denomina *función de respuesta al impulso*, pues sus sucesivos coeficientes describen el efecto que, a través del tiempo, tendría sobre  $Y$  un impulso (es decir, un cambio puramente transitorio) en la variable  $X$ . El parámetro  $b$ , ( $b \geq 0$ ) que aparece como subíndice en la variable  $X$  se denomina *tiempo muerto* de la relación, y denota el número de períodos que deben transcurrir para que la variación en  $X$  comience a dejarse sentir sobre  $Y$ .

Por muy rico que sea el efecto de  $X$  sobre  $Y$ , tanto en intensidad temporal como en duración, es posible recogerlo adecuadamente por medio del polinomio  $v(L)$ , consiguiendo además que el término de error o perturbación  $N_t$  sea estacionario. Para ello, puede ser preciso, en ocasiones, utilizar un polinomio  $v(L)$  de orden elevado, quizá infinito, pero su descomposición como cociente de dos polinomios de orden finito, incluso muy reducido, como aparece en [13.19], permite la representación parsimoniosa de la relación.

A su vez, la perturbación del modelo admitirá una representación univariante:

$$N_t = \psi(L)a_t = (1 + \psi_1L + \psi_2L^2 + \dots)a_t = \frac{\theta(L)\Theta(L)}{\phi(L)\Phi(L)} a_t$$

donde las mayúsculas hacen referencia a la naturaleza estacional de los polinomios respectivos, y donde los polinomios autorregresivos  $\phi(L)$  y  $\Phi(L)$  pueden contener raíces unitarias, aunque no raíces de módulo superior a la unidad, y  $a_t$  es un ruido blanco gaussiano, es decir, con distribución Normal.

La acumulación sucesiva de los coeficientes de la función de respuesta al impulso genera la *función de respuesta al escalón*, que proporciona el efecto gradual que sobre  $Y$  tendría una desviación *permanente* en el valor de  $X$ . Si la perturbación  $N_t$  es estacionaria, la función de respuesta al escalón es acotada; su límite se denomina *ganancia* de la relación, y mide el efecto a

largo plazo que sobre  $Y_t$  tiene una variación en  $X_t$ . Puede calcularse mediante  $g = \omega(1)/\delta(1)$ .

Un instrumento básico en la especificación del modelo de función de transferencia es la *función de correlación cruzada* (en lo sucesivo, fcc), similar a las funciones de autocorrelación, y definida como cociente entre la covarianza entre  $X$  e  $Y$ , a distintos retardos y en ambas direcciones, por el producto de sus desviaciones típicas:

$$\rho_{xy}(k) = \frac{\gamma_{xy}(k)}{s_x s_y}$$

donde:

$$\gamma_{xy}(k) = \frac{1}{n} \sum_{t=1}^{T-k} (x_t - \bar{x})(y_{t+k} - \bar{y}), \quad k = 0, 1, 2, \dots$$

$$\gamma_{xy}(k) = \frac{1}{n} \sum_{t=k+1}^T (x_{t-k} - \bar{x})(y_t - \bar{y}), \quad k = 0, -1, -2, \dots$$

A diferencia de las fas y fap de una serie, no es simétrica, y su valor central  $\rho_{xy}(0)$  es el coeficiente de correlación habitual entre ambas variables.

Si las dos variables  $X$  e  $Y$  no estuviesen correlacionadas entre sí y una de ellas fuese ruido blanco, por ejemplo  $Y_t = a_t$ , se tendría:

$$\text{Cov}[\rho_{xa}(k), \rho_{xa}(k+s)] \cong \frac{\rho_{xx}(s)}{T-k}$$

$$\text{Var}[\rho_{xa}(k)] \cong \frac{1}{T-k}$$

de modo que:

$$\text{Corr}[\rho_{xa}(k), \rho_{xa}(k+s)] \cong \rho_{xx}(s)$$

En tales condiciones, el inverso del número de grados de libertad da la varianza de cada valor de la fcc, que reproduciría la estructura de la fas de  $X_t$ . Ello ocurrirá *a pesar* de la ausencia de correlación cruzada entre ambas variables. Por consiguiente, sólo puede esperarse una fcc con valores nulos fuera de la correlación contemporánea  $\rho_{xy}(0)$ , si *ambas series* fuesen ruido blanco.

Las expresiones anteriores muestran que los sucesivos valores de la fcc están generalmente correlacionados entre sí, lo que puede dificultar la identificación de la estructura dinámica de la relación entre  $X$  e  $Y$ . Lo que acabamos de ver es que la *fcc tenderá a sugerir más relación entre las variables de la que realmente existe*.

### 13.12.a. Identificación del modelo de función de transferencia

La identificación del modelo de transferencia consiste en obtener valores aproximados de los coeficientes de la función de respuesta al impulso  $v(L)$ , de modo que puedan utilizarse para inferir los órdenes  $r$  y  $s$  de los polinomios  $\delta(L)$  y  $\omega(L)$  en [13.19], así como el tiempo muerto  $b$ . Identificados  $r$  y  $s$ , los coeficientes de  $v(L)$  pueden utilizarse asimismo para obtener preestimaciones de los coeficientes en  $\delta(L)$  y  $\omega(L)$ .

Si escribimos [13.19] en la forma

$$\begin{aligned} (1 - \delta_1 L - \delta_2 L^2 - \dots - \delta_r L^r)(v_0 + v_1 L + v_2 L^2 + \dots) = \\ = (\omega_0 - \omega_1 L - \omega_2 L^2 - \dots - \omega_s L^s) L^b \end{aligned} \quad [13.20]$$

se tiene el sistema:

$$\begin{aligned} v_j &= 0 && \text{para } j < b \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} + \omega_0 && \text{para } j = b \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} - \omega_{j-b} && \text{para } j = b + 1, \dots, b + s \\ v_j &= \delta_1 v_{j-1} + \delta_2 v_{j-2} + \dots + \delta_r v_{j-r} && \text{para } j > b + s \end{aligned} \quad [13.21]$$

de modo que los coeficientes de la función de respuesta al impulso tienen la siguiente estructura:

a) Los primeros  $b$  coeficientes son nulos, lo que permite identificar el tiempo muerto  $b$ :  $v_0 = v_1 = \dots = v_{b-1} = 0$ .

b) Desde el coeficiente  $v_b$  hasta el coeficiente  $v_{b+s}$  no se aprecia ninguna regla de formación.

c) Los coeficientes  $v_j$  para  $j \geq b + s + 1$  se comportan de acuerdo con la ecuación en diferencias de orden  $r$  que aparece en [13.21], tomando como valores iniciales  $v_{b+s}$ ,  $v_{b+s-1}$ ,  $\dots$ ,  $v_{b+s-r+1}$ .

Por ejemplo, en la función de transferencia:

$$Y_t = \frac{\omega_0}{1 - \delta_1 L - \delta_2 L^2} X_t$$

todos los coeficientes de la función de respuesta al impulso obedecen a una ecuación en diferencias de orden 2; lamentablemente, ello puede generar múltiples configuraciones, aunque una típica sería sinusoidal.

Por el contrario, en el modelo

$$Y_t = \frac{\omega_0 - \omega_1 L - \omega_2 L^2}{1 - \delta_1 L} X_{t-1}$$

se tendría un primer coeficiente  $v_0$  nulo, seguido de  $v_1$  y  $v_2$ , que no obedecerían a ninguna regla, mientras que los sucesivos responderían a una ecuación en diferencias de orden 1, por lo que decaerían exponencialmente, aunque quizá alternando en signo.

Si las variables que se pretende relacionar no son estacionarias, su función de correlación cruzada, así como sus fas y fap, no decaerá rápidamente hacia cero. En tal caso, es preciso transformarlas mediante diferencias para lograr estacionariedad, lo que denotamos en lo sucesivo por  $y_t = \nabla^d Y_t$ ,  $x_t = \nabla^d X_t$ ,  $n_t = \nabla^d N_t$ , que serían estacionarias. Si es preciso tomar asimismo diferencias de orden estacional, así como quizá tomar logaritmos para atenuar la heteroscedasticidad, todo ello estaría incorporado en nuestra notación:  $y_t$ ,  $x_t$ ,  $n_t$ . La función de transferencia transformada resulta:

$$y_t = v_0 x_t + v_1 x_{t-1} + v_2 x_{t-2} + \dots + n_t \quad [13.22]$$

que conserva los mismos coeficientes de la original. Supongamos que esta función de transferencia es tal que  $v_j = 0$  para  $j > q$ . Si premultiplicamos en [13.22] por  $x_{t-k}$  para sucesivos valores  $k \geq 0$  y tomamos esperanzas se tiene:

$$\begin{pmatrix} \gamma_{xy}(0) \\ \gamma_{xy}(1) \\ \dots \\ \gamma_{xy}(q) \end{pmatrix} = \begin{pmatrix} \gamma_{xx}(0) & \gamma_{xx}(1) & \dots & \gamma_{xx}(q) \\ \gamma_{xx}(1) & \gamma_{xx}(0) & \dots & \gamma_{xx}(q-1) \\ \dots & \dots & \dots & \dots \\ \gamma_{xx}(q) & \gamma_{xx}(q-1) & \dots & \gamma_{xx}(0) \end{pmatrix} \cdot \begin{pmatrix} v_0 \\ v_1 \\ \dots \\ v_q \end{pmatrix} \quad [13.23]$$

de modo que si se dispusiese de las estimaciones muestrales de la fcc entre  $x$  e  $y$ , lo que siempre es posible, podríamos resolver el sistema [13.23] para obtener estimaciones de los coeficientes de la función de respuesta al impulso  $v_0, v_1, \dots, v_n$ .

### 13.12.b. Identificación con preblanqueo

El proceso de identificación de la función de transferencia se simplifica notablemente si se filtran previamente ambas variables de manera adecuada. Supongamos que la variable  $x_t$ , siendo estacionaria, admite la representación ARMA:

$$\phi_x(L)\theta_x^{-1}(L)x_t = \alpha_t \quad [13.24]$$

donde  $\alpha_t$  es ruido blanco y  $\phi_x(L)$  y  $\theta_x(L)$  son los polinomios autorregresivos y de medias móviles de dicho modelo ARMA. Una vez estimado [13.24], obtendríamos los residuos  $\hat{\alpha}_t$ , que no son sino el resultado de *filtrar* la variable  $x_t$  por su modelo ARMA. Supongamos que filtramos el output  $y_t$  con el mismo modelo ARMA del input que acabamos de estimar, obteniendo:

$$\beta_t = \phi_x(L)\theta_x^{-1}(L)y_t$$

Las variables así filtradas (no cabe esperar que  $b_t$  sea ruido blanco) satisfacen la relación:

$$\beta_t = v(L)\alpha_t + \varepsilon_t \quad [13.25]$$

donde:

$$\varepsilon_t = \phi_x(L)\theta_x^{-1}(L)n_t$$

puediendo observarse que la función de respuesta al impulso en [13.25] es idéntica a la original. Si multiplicamos por  $\alpha_{t-k}$  y tomamos esperanzas, se tiene:

$$\gamma_{\alpha\beta}(k) = v_k \sigma_\alpha^2$$

donde  $\gamma_{\alpha\beta}(k)$  denota la covarianza entre  $\alpha_t$  y  $\beta_t$  en el retardo  $k$ , por lo que:

$$v_k = \frac{\gamma_{\alpha\beta}(k)}{\sigma_\alpha^2} = \frac{\rho_{\alpha\beta}(k)}{\sigma_\alpha} \sigma_\beta, \quad k = 0, 1, 2, \dots \quad [13.26]$$

de modo que *tras preblanquear input y output con el modelo del input, la función de correlación cruzada es proporcional a la función de respuesta al impulso*. El efecto de preblanquear es transformar el sistema de ecuaciones [13.23] en un conjunto de ecuaciones como [13.26], que pueden resolverse por separado para obtener preestimaciones de la función de respuesta al impulso a partir de las desviaciones típicas y de la función de correlación cruzada muestral de  $\alpha$  y  $\beta$ .

### 13.12.c. Identificación de un modelo para el ruido

Una vez identificada una función de respuesta al impulso, y habiendo obtenido estimaciones para sus coeficientes  $\hat{v}(L)$ , podemos estimar el ruido:  $\hat{n}_t = y_t - \hat{v}(L)x_t$ . Alternativamente, suele comenzarse especificando para  $n_t$  el modelo univariante de  $y_t$ , lo que sería correcto si no existiese relación entre  $y_t$  y  $x_t$ , en cuyo caso, la función de transferencia sería nula. Al ir estimando coeficientes de  $v(L)$ , el modelo de  $n_t$  se hace más sencillo<sup>(2)</sup>.

Ambas opciones se basan en modelos estimados previamente, por lo que todo lo que hacemos a continuación tiene un valor tan sólo aproximado. En definitiva, ha de ser el rigor del analista el que debe conducir a la especificación de un modelo univariante, como los estudiados en las secciones previas, para los residuos resultantes del modelo de función de transferencia.

<sup>(2)</sup> Existen otros procedimientos analíticos más complejos para identificar un modelo univariante para la perturbación, que el lector interesado puede consultar en Box y Jenkins (1972). En cualquier caso, no es clara cuál es su utilidad práctica, más allá del buen análisis univariante de la serie de perturbaciones.

Al especificar un modelo de transferencia, Box y Jenkins sugieren:

1. Que un modelo con estructuras AR o MA de orden 1 ó 2 (además del posible tiempo muerto) será normalmente suficiente, si bien esto debe tenerse en cuenta tanto para la componente regular como para la estacional.
2. Que, aunque la estimación del modelo de transferencia es eficiente sólo si el modelo especificado es correcto, las estimaciones de los valores  $v_k$  de la función de respuesta al impulso son útiles en la identificación de la estructura dinámica de relación.
3. Que tampoco se gana mucho con tratar de obtener estimaciones eficientes de los  $v_k$ , pues son precisos muchos coeficientes de tal tipo para resumir el cociente de polinomios  $\delta(L)/\omega(L)$ . Recíprocamente, el objetivo último del analista es obtener estimaciones precisas de estos dos polinomios, y las estimaciones que ellas implican para los coeficientes  $v_k$  están altamente correlacionadas entre sí, y tienen una elevada varianza.

### 13.12.d. Estimación de un modelo de función de transferencia

Una vez identificado el modelo de transferencia, sin estacionalidad:

$$y_t = \delta^{-1}(L)\omega(L)x_{t-b} + \phi^{-1}(L)\theta(L)a_t$$

y dados unos valores iniciales  $(x_0, y_0, a_0)$  podemos obtener los residuos sucesivos:  $a_t(b, \delta, \omega, \phi, \theta/x_0, y_0, a_0)$  a partir de  $t_0 = \max\{r, s + b\} + 1$  y, bajo el supuesto de Normalidad, tratar de minimizar su suma de cuadrados.

Como ejemplo, consideremos la función de transferencia:

$$y_t = \frac{\omega_0 + \omega_1 L + \omega_2 L^2}{1 - \delta L} x_t + \frac{1}{1 - \rho L} a_t$$

en la que  $\max\{r, s + b\} + 1 = 3$ , y donde  $n_t = a_t/(1 - \rho L)$ , en la que podemos hacer:

$$\begin{aligned} n_3 &= y_3 - \delta y_2 - \omega_0 x_3 - \omega_1 x_2 - \omega_2 x_1 + \delta n_2 \\ n_4 &= y_4 - \delta y_3 - \omega_0 x_4 - \omega_1 x_3 - \omega_2 x_2 + \delta n_3 \\ &\dots \end{aligned}$$

comenzando de  $n_2 = 0$  para después obtener las innovaciones  $a_t$  a partir de  $t = 4$ , por medio de:

$$\begin{aligned} a_4 &= n_4 - \rho n_3 \\ a_5 &= n_5 - \rho n_4 \\ &\dots \end{aligned}$$

y formar, por último, la suma residual:

$$S^2(b, \delta, \omega, \rho) \equiv \sum_4^T a_t^2(b, \delta, \omega, \rho/x_0, y_0, a_0)$$

donde las primeras innovaciones se han tomado iguales a su esperanza incondicional, que es cero.

La expresión genérica de cada  $a_t$  es:

$$a_t = y_t - \delta y_{t-1} - \omega_0 x_t - \omega_1 x_{t-1} - \omega_2 x_{t-2} + \delta n_{t-1} - \rho [y_{t-1} - \delta y_{t-2} - \omega_0 x_{t-1} - \omega_1 x_{t-2} - \omega_2 x_{t-3} + \delta n_{t-2}]$$

con vector gradiente:

$$\frac{\partial a_t}{\partial \delta} = -y_{t-1} + \rho y_{t-2} + n_{t-1} - \rho n_{t-2}$$

$$\frac{\partial a_t}{\partial \omega_0} = -x_t + \rho x_{t-1}$$

$$\frac{\partial a_t}{\partial \omega_1} = -x_{t-1} + \rho x_{t-2}$$

$$\frac{\partial a_t}{\partial \omega_2} = -x_{t-2} + \rho x_{t-3}$$

$$\frac{\partial a_t}{\partial \rho} = -(y_{t-1} - \delta y_{t-2} - \omega_0 x_{t-1} - \omega_1 x_{t-2} - \omega_2 x_{t-3} + \delta n_{t-2})$$

por lo que podemos comenzar un algoritmo iterativo del tipo Gauss-Newton a partir de estimaciones iniciales  $\hat{\boldsymbol{\theta}} = (\hat{\delta}, \hat{\omega}_0, \hat{\omega}_1, \hat{\omega}_2, \hat{\rho})$  del modo que se discutió en el Capítulo 12. Como allí se vio, un modo de proceder consiste en obtener series temporales para cada una de las componentes del gradiente, así como para los propios residuos  $a_t$ , utilizando las estimaciones iniciales de los parámetros, para estimar la regresión:

$$\hat{a}_t = (\hat{\delta} - \delta) \frac{\partial a_t(\hat{\boldsymbol{\theta}})}{\partial \delta} + (\hat{\omega}_0 - \omega_0) \frac{\partial a_t(\hat{\boldsymbol{\theta}})}{\partial \omega_0} + (\hat{\omega}_1 - \omega_1) \frac{\partial a_t(\hat{\boldsymbol{\theta}})}{\partial \omega_1} + (\hat{\omega}_2 - \omega_2) \frac{\partial a_t(\hat{\boldsymbol{\theta}})}{\partial \omega_2} + (\hat{\rho} - \rho) \frac{\partial a_t(\hat{\boldsymbol{\theta}})}{\partial \rho} + a_t$$

que se repite, tomando en cada iteración las últimas estimaciones como iniciales, hasta lograr la convergencia. La aparición del vector  $\boldsymbol{\theta}$  hace referencia a que las componentes del gradiente de la función  $a_t$  están evaluadas en las preestimaciones. Este procedimiento está sujeto a todas las consideraciones que en el Capítulo 12 se hicieron en relación con la convergencia de un algoritmo numérico. Lograda la convergencia, la matriz de covarianzas de las estimaciones se aproxima por el producto  $\hat{\sigma}_a^2 (\mathbf{V}\mathbf{a}\mathbf{V}')^{-1}$ , donde  $\hat{\sigma}_a^2$  se estima mediante el cociente del valor alcanzado por la suma residual en la

última iteración, y el número de observaciones, o de grados de libertad, si se prefiere.

### 13.12.e. Diagnóstico del modelo de función de transferencia

Uno de los análisis que debe hacerse de toda función de transferencia estimada es la búsqueda de posibles factores comunes que pudiesen quedar enmascarados en los polinomios AR y MA que la componen. Nótese que el modelo de función de transferencia, al igual que todo modelo lineal, es único excepto por operadores factoriales, posiblemente dinámicos. Es decir, si el «verdadero» modelo de transferencia es:

$$Y_t = \delta^{-1}(L)\omega(L)X_{t-b} + \phi^{-1}(L)\theta(L)a_t$$

también es válido el modelo:

$$D(L)Y_t = D(L)\delta^{-1}(L)\omega(L)X_{t-b} + D(L)\phi^{-1}(L)\theta(L)a_t$$

donde  $D(L)$  es un polinomio cualquiera en el operador de retardos, aunque el analista preferirá el primero por su sencillez; en la práctica, el menor número de parámetros se traducirá en una mayor eficiencia en la estimación, así como en una ausencia de correlaciones entre sus valores estimados. El objetivo del investigador ha de ser siempre lograr un modelo tan sencillo como sea posible, comenzando siempre por especificaciones sencillas, para complicarlas posteriormente, sólo si es preciso. En todo caso, las correlaciones entre los parámetros estimados deben ser reducidas, pues valores elevados indican sobreparametrización.

Supongamos que el modelo correcto es  $y_t = v(L)x_t + \psi(L)a_t$ , pero el investigador, incorrectamente, especifica  $y_t = v_0(L)x_t + \psi_0(L)a_{0t}$ , que, una vez estimado, genera unos residuos que denotamos por  $a_{0t}$ . Estos quedan definidos por:

$$a_{0t} = \psi_0^{-1}(L)[v(L) - v_0(L)]x_t + \psi_0^{-1}(L)\psi(L)a_t$$

donde puede apreciarse que la mala especificación puede hacer que: *a)* los residuos tengan autocorrelación, y *b)* estén correlacionados con las variables  $x_t$ , y por tanto, con el ruido blanco  $\alpha_t$  que genera las  $x_t$ .

**Caso 1:** El modelo de transferencia se especifica correctamente, pero el modelo del ruido incorrectamente:  $v_0(L) = v(L)$ , pero  $\psi_0(L) \neq \psi(L)$ . En tal caso:  $a_{0t} = \psi_0^{-1}(L)\psi(L)a_t$ , por lo que: *a)* la función de correlación cruzada entre los residuos y el input del modelo debe ser no significativamente distinta de cero, pero *b)* los residuos presentan autocorrelación.

**Caso 2:** El modelo de transferencia se especifica incorrectamente. En este caso se tiene  $a_{0t} = \psi_0^{-1}(L)[v(L) - v_0(L)]x_t + a_t$ , por lo que se tiene tanto

autocorrelación de los residuos, como una función de correlación cruzada significativamente distinta de cero con el input.

La conveniencia de preblanquear el input antes de obtener su función de correlación cruzada con los residuos proviene de que, como vimos al comienzo de esta sección, los sucesivos valores de dicha función de correlación cruzada están correlacionados, lo que no ocurre con la función de correlación cruzada entre el input preblanqueado y los residuos, si éstos son efectivamente ruido blanco.

Pero además, este mismo análisis sugiere el tipo de modificaciones que deben introducirse en la función de transferencia previamente estimada para obtener una mejor especificación. En efecto, considerando el modelo preblanqueado:  $\beta_t = v(L)\alpha_t + \varepsilon_t$  y si denotamos por  $\varepsilon_{0t}$  los residuos del modelo «incorrecto», se tiene:

$$\varepsilon_{0t} = [v(L) - v_0(L)]\alpha_t + \varepsilon_t$$

se tiene:

$$v_k - v_{0k} = \rho_{\alpha, \varepsilon_0}(k) \frac{\sigma_{\varepsilon_0}}{\sigma_\alpha}, \quad k = 0, 1, 2, \dots$$

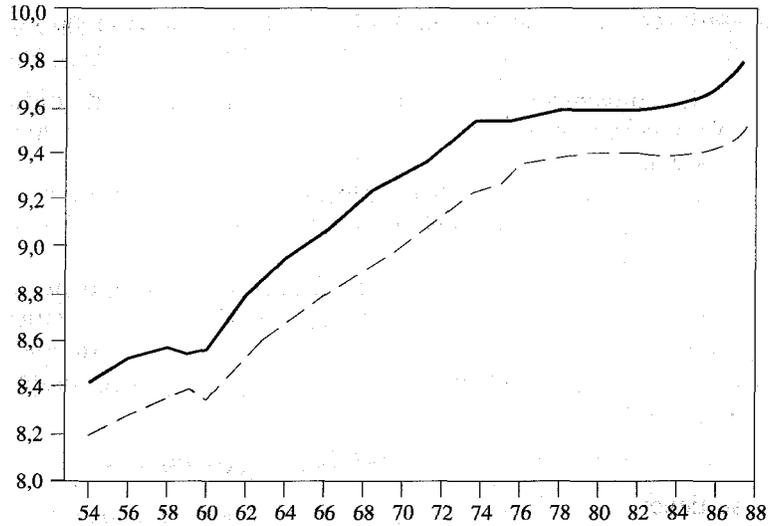
de modo que la función de correlación cruzada entre los residuos y el input preblanqueado mide la discrepancia entre la función de respuesta al impulso «verdadera» y la estimada.

Por último, puesto que es necesario para una correcta especificación de la función de transferencia que los residuos estén libres de autocorrelación, puede utilizarse los estadísticos de Box-Pierce o de Ljung-Box introducidos en la Sección 13.11. En este caso,  $Q_g = T \sum_0^g \hat{r}_\alpha^2(k)$  se distribuye como una  $\chi_{g+1-(r+s-1)}^2$ , donde  $r$  y  $s$  son los órdenes de los polinomios AR y MA de la función de transferencia, y no dependen del modelo del ruido.

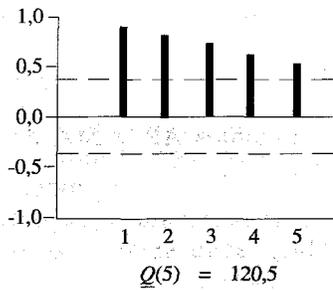
### 13.13. ALGUNOS EJEMPLOS

**Ejemplo 13.1.** Consideremos nuevamente las series de datos anuales de Consumo y PIB español, en pesetas constantes de 1980, de la Sección 7.6. La Figura 13.5 presenta ambas series, con evidencia obvia de no estacionariedad, ya que no podría mantenerse que la media de la serie fuese estable en el tiempo. En el análisis de series temporales, la transformación logarítmica es utilizada frecuentemente para amortiguar la heteroscedasticidad que suele aparecer cuando la varianza de la variable crece con la magnitud de ésta al avanzar en el período muestral.

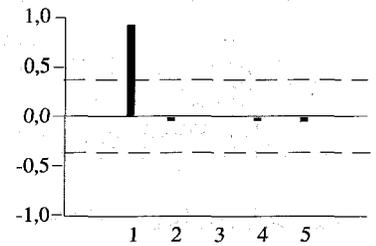
Las funciones de autocorrelación simple (fas) de las variables transformadas (Figuras 13.6 y 13.8) son ejemplos típicos de ausencia de estacionariedad, amortiguándose sus valores con excesiva lentitud. Las funciones de autocorrelación parcial (fap, Figuras 13.7 y 13.9) presentan asimismo clara evidencia de ausencia de estacionariedad, con tan sólo su primer valor significativo, y muy próximo a 1.



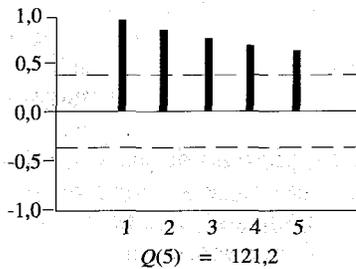
**FIGURA 13.5.** Consumo [---] y PIB [—]. Transformación logarítmica sobre las series en pesetas constantes base 1980, en miles de millones.



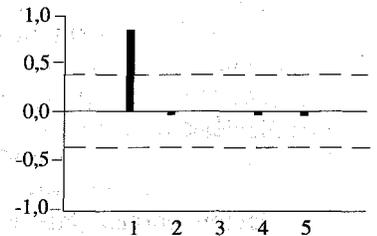
**FIGURA 13.6.** fas ln (PIB).



**FIGURA 13.7.** fap ln (PIB).



**FIGURA 13.8.** fas ln (Consumo).



**FIGURA 13.9.** fap ln (Consumo).

Las series de diferencias del logaritmo (Figura 13.10) tienen la ventaja de poder interpretarse como las tasas de variación anual del Consumo y del PIB. Sus fas y fap en las Figuras 13.11 a 13.14 no presentan clara evidencia de ausencia de estacionariedad. Sin embargo, es importante observar que las series diferenciadas tienen medias muestrales positivas y significativas, con estadísticos  $t$  por encima de 8. La Figura 13.10 muestra la evidencia de tal media, junto con el hecho de que las series deambulan, sin oscilar alrededor de sus medias muestrales, como cabría esperar de una serie estacionaria.

Consideremos inicialmente la serie anual del Consumo en pesetas constantes de 1980. Si, a pesar de la evidencia apuntada sobre las primeras diferencias de su logaritmo, intentamos modelizar esta transformación de la variable, el primer valor de la fap, junto con el rápido amortiguamiento de su fas, sugieren un modelo ARIMA(1, 1, 0) para el logaritmo del Consumo. Otra alternativa consistiría en tomar una diferencia adicional de la variable, que aparece, junto con igual transformada del PIB, en la Figura 13.15. Las fas y fap (Figuras 13.6 a 13.19) sugieren ahora una posible estructura de media móvil de orden 1, por lo que deberíamos probar un ARIMA(0, 2, 1). Hay que notar que la duda acerca de una posible diferencia adicional debe ir acompañada siempre de un término MA(1) que pueda cancelarse (si se estima próximo a 1) con la diferencia adicional que estaba en duda; por ello, el modelo alternativo que proponemos tiene bastante sentido. Su evidencia en las fas y fap no es totalmente clara, pero hay que tener en cuenta que estas dos series tienen un número no muy elevado de observaciones. Por otra parte, la segunda diferencia elimina el problema de la media residual significativa.

La estimación de ambos modelos resultó en:

$$(1 - 0,857 L) \nabla \ln C_t = a_t \\ (0,088)$$

$$T = 33, \quad \hat{\sigma}_a = 2,75 \%, \quad \bar{a} = 0,0061 \\ (0,0046)$$

$$\nabla^2 \ln C_t = (1 - 0,545 L) a_t \\ (0,149)$$

$$T = 33, \quad \hat{\sigma}^2 = 2,61 \%, \quad \bar{a} = -0,0045 \\ (0,0261)$$

El modelo en primeras diferencias presenta tres dificultades:

1. El elevado coeficiente del término AR en el primer modelo sugiere la necesidad de una diferencia adicional, puesto que un término  $(1 - \rho L)$  con  $\rho$  próximo a 1 es esencialmente igual a  $(1 - L)$ .
2. Lo que es más importante, su gráfico de residuos (Figura 13.20) presenta indicios claros de no estacionariedad, con rachas largas de valores sistemáticamente por encima o por debajo de su media.

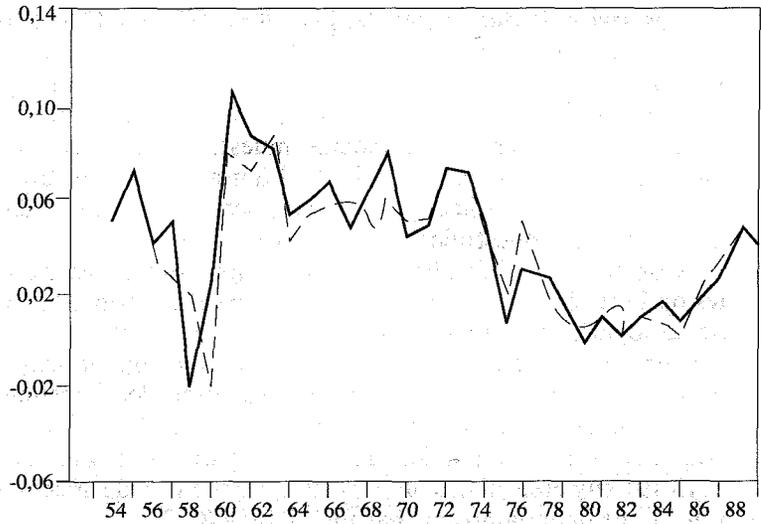


FIGURA 13.10. Consumo [---] y PIB [—]. Tasas de variación logarítmica.

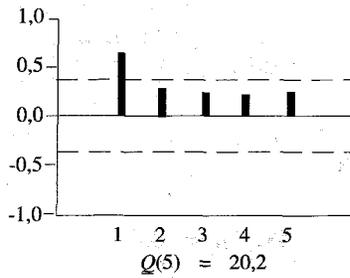


FIGURA 13.11. fas  $\nabla \ln(\text{PIB})$ .

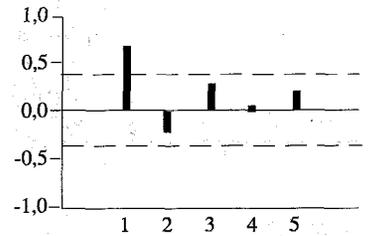


FIGURA 13.12. fap  $\nabla \ln(\text{PIB})$ .

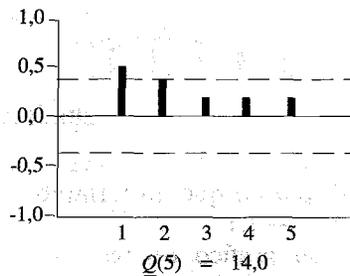


FIGURA 13.13. fas  $\nabla \ln(\text{Consumo})$ .

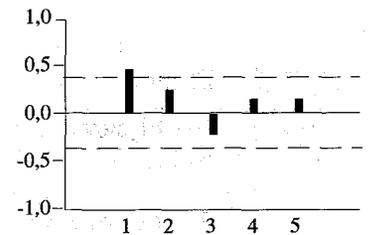


FIGURA 13.14. fap  $\nabla \ln(\text{Consumo})$ .

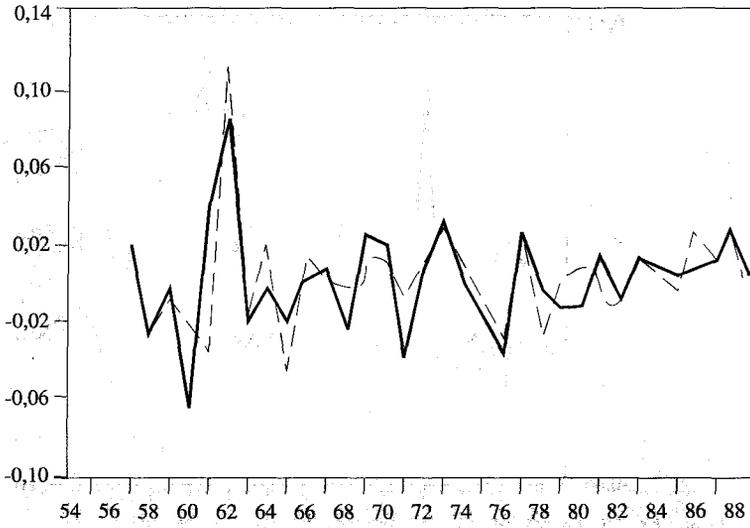


FIGURA 13.15. Consumo [---] y PIB [—]. Segundas diferencias del logaritmo.

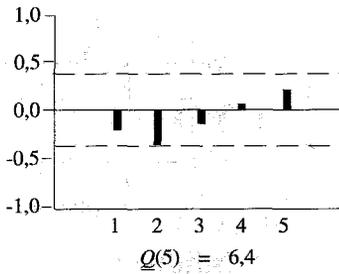


FIGURA 13.16. fas  $\nabla^2 \ln(\text{PIB})$ .

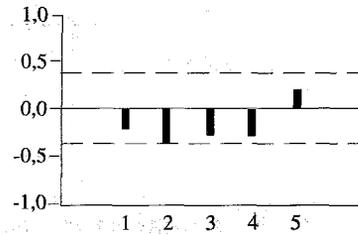


FIGURA 13.17. fap  $\nabla^2 \ln(\text{PIB})$ .

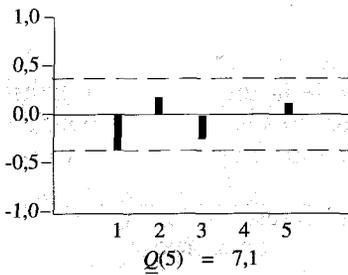


FIGURA 13.18. fas  $\nabla^2 \ln(\text{Consumo})$ .

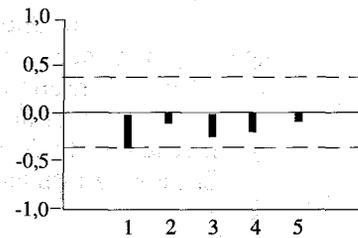


FIGURA 13.19. fap  $\nabla^2 \ln(\text{Consumo})$ .

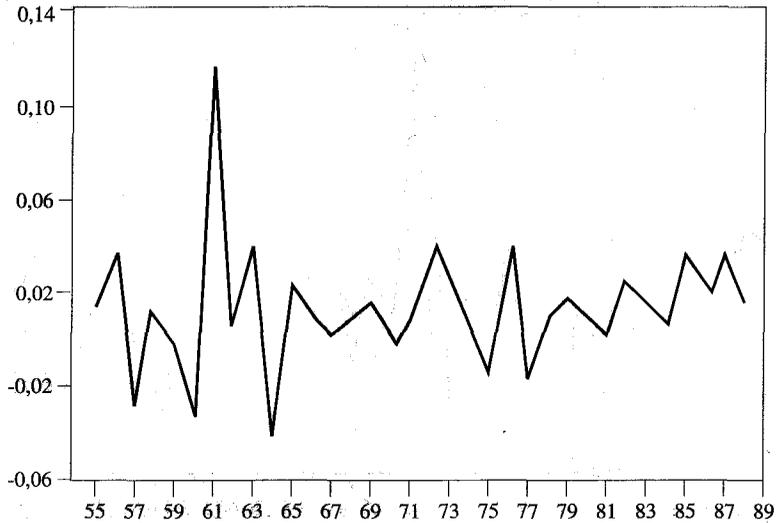


FIGURA 13.20. Residuos Consumo (Modelo en primeras diferencias).

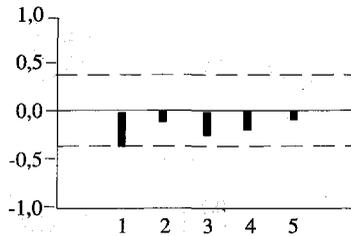


FIGURA 13.21. fas residuos.

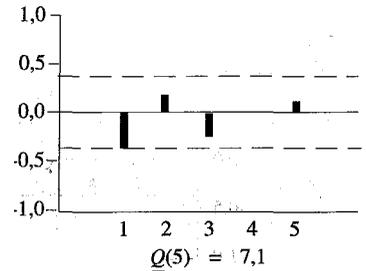


FIGURA 13.22. fap residuos.

3. Consistentemente con esta observación, las fas y fap (Figuras 13.21 y 13.22) de dichos residuos sugieren la existencia de estructura estocástica adicional, aunque en la forma de un término AR(1) que ya está incluido en el modelo.

El modelo en segundas diferencias no presenta ninguno de estos problemas: a) su parámetro MA está lejos de la unidad, lo que habría supuesto la cancelación con una de las diferencias, b) sus residuos (Figura 13.23) parecen estacionarios, y c) no hay indicios de estructura estocástica residual en las fas y fap (Figuras 13.26 y 13.27).

El modelo estimado del PIB resultó ser:

$$(1 + 0,16L + 0,38L^2)\nabla^2 \ln \text{PIB}_t = a_t$$

(0,16)    (0,16)

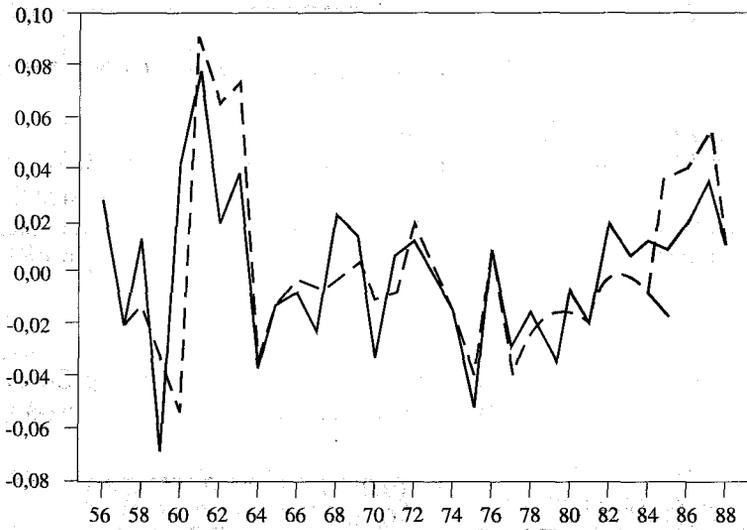


FIGURA 13.23. Residuos Consumo [---] y residuos PIB [—].

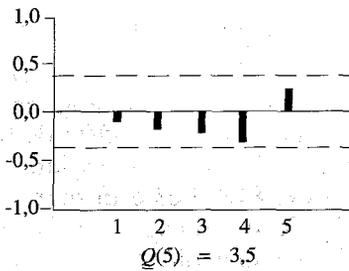


FIGURA 13.24. fas residuos PIB.

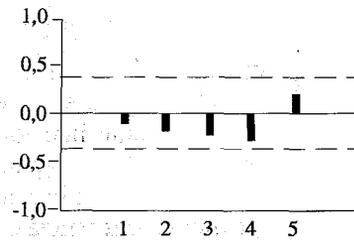


FIGURA 13.25. fap residuos PIB.

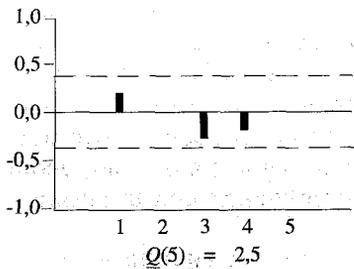


FIGURA 13.26. fas residuos Consumo.

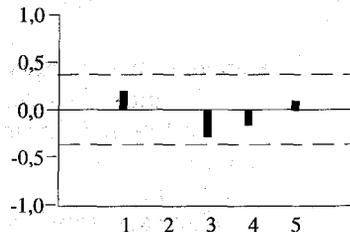


FIGURA 13.27. fap residuos Consumo.

con

$$T = 33, \quad \hat{\sigma}_u = 2,58 \% ; \quad a = -0,0002(0,0045)$$

sin que su gráfico de residuos (Figura 13.23) ni sus fas y fap (Figuras 13.24 y 13.25) mostrasen la necesidad de modificar el modelo. El signo positivo del coeficiente de  $L^2$  en el polinomio AR(2) implica la existencia de una estructura cíclica en esta variable, que recoge la evolución del ciclo económico. De acuerdo con los coeficientes estimados, el período de dicho ciclo es de 4,37 años.

Posteriormente consideramos la posibilidad de establecer una relación dinámica entre ambas variables, tratando de recoger el efecto que variaciones en el PIB (como indicador de Renta) tienen sobre el Consumo. Tras preblanquear ambas variables y examinar la fcc resultante, estimamos el modelo:

$$\nabla \ln C_t = \frac{(0,062)}{0,766} \nabla \ln \text{PIB}_t + (1 - 0,292)a_t$$

$$\frac{(0,067)}{1 - 0,204L} \quad (0,156)$$

con

$$T = 33, \quad \hat{\sigma} = 1,22 \% ; \quad a = 0,0011(0,0021)$$

cuyos residuos, así como sus fas, fap, aparecen en las Figuras 13.28 a 13.30.

Como puede verse, la desviación típica de este modelo de transferencia es sensiblemente inferior a la obtenida con el modelo univariante del Consumo, lo que avala la capacidad explicativa del PIB. Por otra parte, a diferencia de los modelos univariantes, el modelo de transferencia se ha especificado en primeras diferencias. La cancelación de una diferencia en ambas variables sugiere la existencia de una «tendencia común» en ambas, lo que será objeto de análisis posterior en el próximo capítulo.

**Ejemplo 13.2.** Consideremos ahora la serie trimestral del número de ocupados en España, tomado de la Encuesta de Población Activa (EPA) para el período (1976, IV) a (1992, III) (Figura 13.31). Las fas y fap (Figuras 13.32 y 13.33) muestran una evidencia obvia de ausencia de estacionariedad. Las fas y fap de la primera diferencia de la variable (Figuras 13.34 y 13.35) sugieren su no estacionariedad estacional, puesto que los valores  $r_4, r_8, r_{12}$  son elevados. Los primeros valores de la fas hasta  $r_7$  son muy altos, lo que sugiere tomar asimismo una diferencia regular adicional, acompañada de un término MA(1) que puede llegar a simplificarse con la misma, si se estima en torno a la unidad.

Con esta transformación, las fas y fap, que aquí no se muestran, sugirieron un posible modelo AR(2) estacional, si bien no de modo obvio. Al estimarse la estructura ARIMA(0, 2, 1)  $\times$  ARIMA<sub>4</sub>(2, 1, 0) se detectó un valor atípico

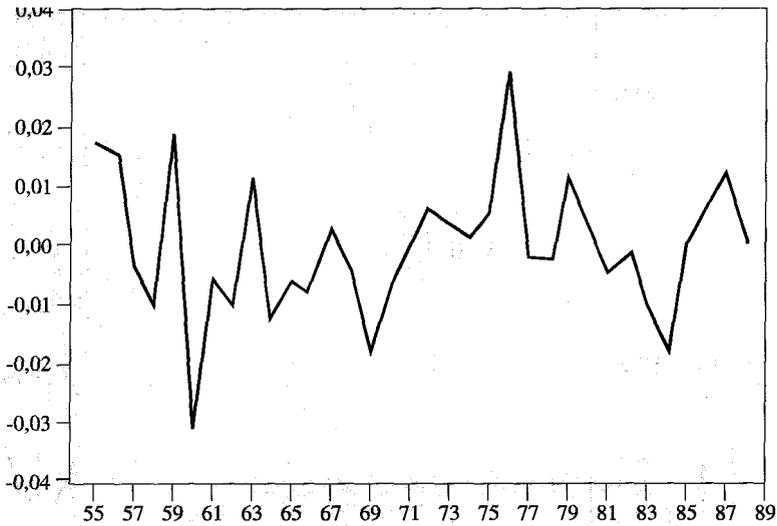


FIGURA 13.28. Residuos función transferencia (Consumo frente a PIB).

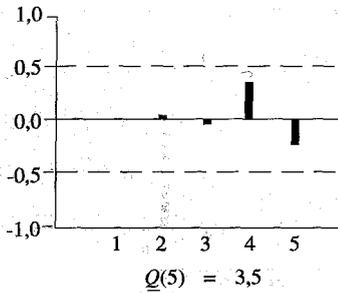


FIGURA 13.29. fas residuos.

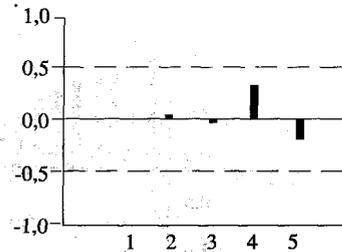


FIGURA 13.30. fap residuos.

elevado en el segundo trimestre de 1987, en que se llevó a cabo un cambio de metodología en la encuesta. Tras incorporar una intervención del tipo escalón en dicho trimestre, se llegó al modelo:

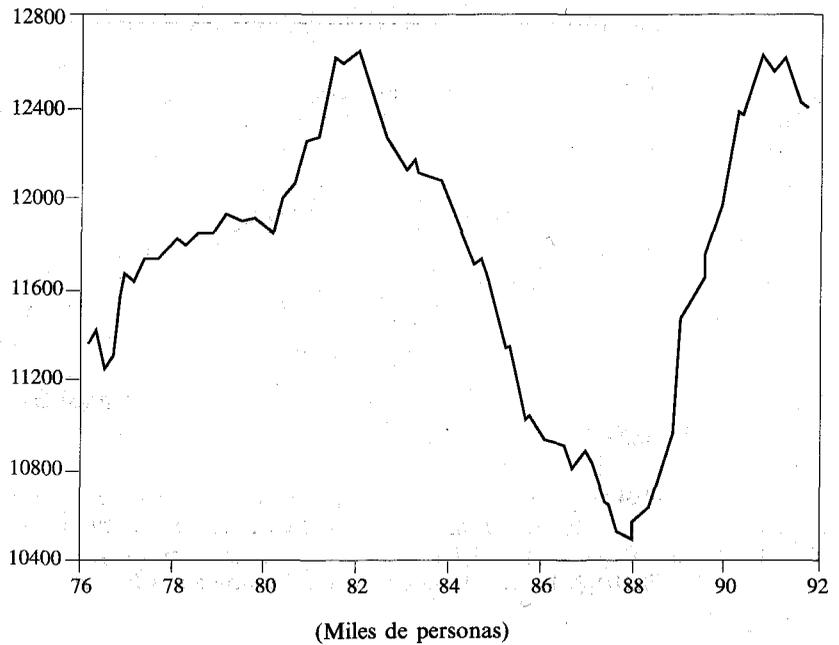
$$(1 + 0,62L^4 + 0,46L^8)\nabla^2\nabla_4 \text{OCUP}_t = 259,01 \xi_{87,2}^s + (1 - 0,53)a_t$$

(0,08)
(0,07)
(35,9)
(0,09)

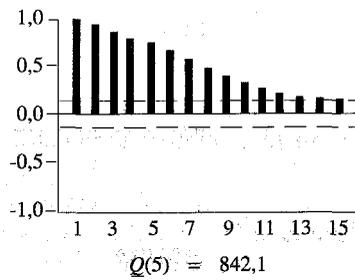
con

$$T = 115, \quad \hat{\sigma} = 56,3; \quad a = -6,3(5,0)$$

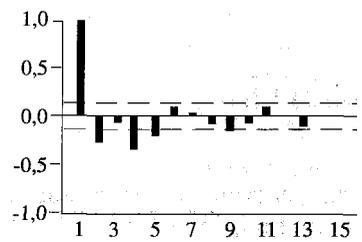
sin que los residuos, ya sea directamente a través de la Figura 13.36 o de sus fas y fap (Figuras 13.37 y 13.38), mostrasen indicios de mala especificación.



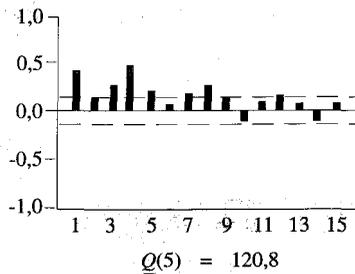
**FIGURA 13.31.** Ocupados EPA.



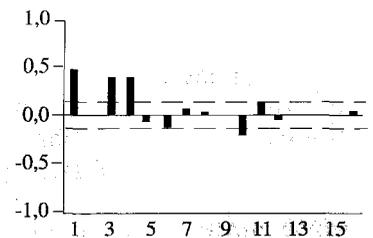
**FIGURA 13.32.** fas ocupados.



**FIGURA 13.33.** fap ocupados.



**FIGURA 13.34.** fas  $\nabla$  ocupados.



**FIGURA 13.35.** fap  $\nabla$  ocupados.

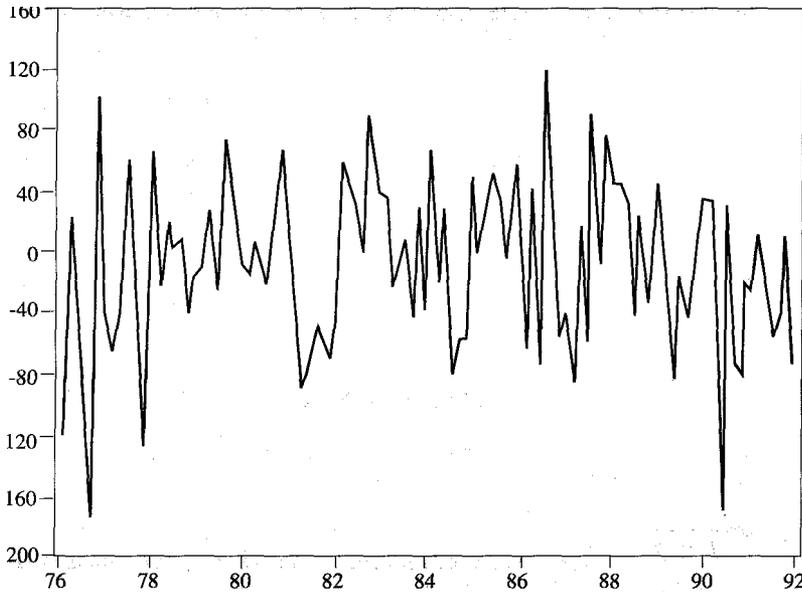


FIGURA 13.36. Residuos del modelo de ocupados.

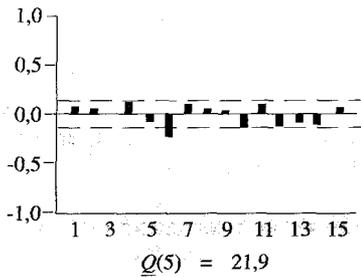


FIGURA 13.37. fas residuos ocupados.



FIGURA 13.38. fap residuos ocupados.

El signo positivo del coeficiente de  $L^8$  en el polinomio AR(2) estacional sugiere nuevamente una estructura cíclica en esta variable, en coherencia con la idea de que el número de ocupados evoluciona con variables del ciclo económico, como el PIB. De acuerdo con nuestras estimaciones, el período del ciclo en ocupados es de 5,75 años, algo por encima del estimado con observaciones anuales para el PIB.

## PROBLEMAS

**Problema 13.1.** Supongamos que se ha especificado y estimado un modelo ARIMA( $p, d, q$ ) con  $d = 2$ , del que se han obtenido unas predicciones para los valo-

res futuros de  $z_t$ , donde  $z_t = \Delta^2 y_t$ . Probar que las predicciones correspondientes a la variable original  $y_t$  pueden recuperarse mediante la fórmula:

$$E_T y_{T+k} = y_T + k(y_T - y_{T-1}) + kE_T z_{T+1} + (k-1)E_T z_{T+2} + \dots + E_T z_{T+k}$$

**Problema 13.2.** Probar que las predicciones de los valores futuros de una variable  $y_t$  con estructura ARIMA(1, 1, 0) obedecen la relación:

$$\begin{aligned} E_T y_{T+1} &= \delta + (1 + \phi)y_T - \phi y_{T-1} \\ E_T y_{T+2} &= \delta(2 + \phi) + (1 + \phi + \phi^2)y_T - \phi(1 + \phi)y_{T-1} \end{aligned}$$

y, en general:

$$E_T y_{T+k} = E_T y_{T+k-1} + \phi E_T z_{T+k-1} + \delta$$

donde  $z_t = \Delta y_t$ . Utilizando este último resultado, probar que, según tiende el horizonte de predicción  $k$  a infinito, se espera que este proceso incremente a una tasa constante de  $\frac{\delta}{1 - \phi}$ .

**Problema 13.3.** Obtener las expresiones analíticas para las predicciones 1, 2, 3 y  $k$  periodos hacia adelante en un proceso ARIMA(1, 1, 1).

**Problema 13.4.** Demostrar que el valor de  $\hat{y}_{T+1}^T$  que minimiza la función objetivo  $E_T[(y_{T+1} - \hat{y}_{T+1}^T)^2]$  es  $\hat{y}_{T+1}^T = E_T y_{T+1}$ .

**Problema 13.5.** Describir cómo se llevaría a cabo la estimación de un modelo ARMA(2, 1).

Explicar la utilización de la expansión en serie de Taylor en este proceso, las variables auxiliares a construir, y el modelo lineal a estimar.

**Problema 13.6.** Demostrar que la función de autocorrelación de un proceso MA( $q$ ) viene dada por:

$$\rho_k = 0 \begin{cases} \frac{-\theta_s + \theta_1 \theta_{s+1} + \dots + \theta_{q-s} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{si } k = 1, 2, \dots, q \\ 0 & \text{si } k > q \end{cases}$$

**Problema 13.7.** Hallar la función de autocorrelación de un proceso ARMA(2, 1) como función de los parámetros de dicho proceso univariante.

**Problema 13.8.** Probar que las mismas expresiones para las predicciones de modelos univariantes que aparecen en la Sección 13.9 pueden obtenerse si se obtiene primero la «representación de medias móviles» del modelo univariante:

$$y_t = \varepsilon_t - \psi_1 \varepsilon_{t-1} - \psi_2 \varepsilon_{t-2} - \psi_3 \varepsilon_{t-3} - \dots$$

**Problema 13.9.** Probar que si en la estimación de un modelo MA(1) para la variable  $y_t$  los residuos  $u_t$  presentan evidencia de seguir, a su vez, un proceso MA(1), el modelo que debería especificarse para la variable  $y_t$  debería ser un modelo MA(2). ¿Qué re-

lación existirá entre los parámetros de los modelos previos para  $y_t$  y  $u_t$  y el modelo especificado finalmente para  $y_t$ ?

**Problema 13.10.** Explicar en detalle cómo puede utilizarse la función de autocorrelación simple para estimar los parámetros de un modelo AR(2), así como la varianza de su término de error.

**Problema 13.11.** Explicar la utilización de la función de autocorrelación simple para estimar el parámetro de un modelo MA(1). ¿Cuál es el rango de valores admisibles de dicho parámetro?

**Problema 13.12.** Probar que si  $y_t = \delta + \rho y_{t-1} + u_t$  con  $|\rho| > 1$ , entonces  $y_t$  admite la representación:

$$y_t = -\frac{\delta}{\rho - 1} - \sum_1^{\infty} \left(\frac{1}{\rho^j}\right) u_{t+j}$$

**Problema 13.13.** Explicar cómo estimar el modelo  $y_t = \varepsilon_t - \theta \varepsilon_{t-1}$  por mínimos cuadrados condicionados (o máxima verosimilitud condicionada).

**Problema 13.14.** El desarrollo de la Sección 13.10 acerca de la estimación de modelos univariantes es tan sólo una aproximación, puesto que ignora las derivadas  $\frac{\theta \varepsilon_{t-1}}{\theta \theta_i}$ ,  $\frac{\theta \varepsilon_{t-2}}{\theta \theta_i}$ ,  $i = 1, 2$ , en la expresión [13.7].

Probar que si se tienen en cuenta estas derivadas, se llega a la relación:

$$\varepsilon_t = \varepsilon_t^0 + (\theta_1 - \theta_1^0) \frac{y_{t-1}}{(1 - \theta_1^0 L - \theta_2^0 L^2)^2} + (\theta_2 - \theta_2^0) \frac{y_{t-2}}{(1 - \theta_1^0 L - \theta_2^0 L^2)^2}$$

¿Cómo podría llevarse a cabo la estimación del modelo MA(2) con esta formulación más completa?

**Problema 13.15.** Probar que la estimación del modelo ARMA(1, 1), de acuerdo con el desarrollo de la Sección 13.10, se reduce a la estimación del modelo:

$$\varepsilon_t^0 + \phi^0 y_{t-1} - \theta^0 \varepsilon_{t-1}^0 = \phi y_{t-1} - \theta \varepsilon_{t-1}^0 + \varepsilon_t$$

**Problema 13.16.** Probar que para llevar a cabo la estimación de un modelo ARMA(1, 1), de acuerdo con el Problema 13.14, sería preciso estimar el modelo:

$$w_t = \phi x_{1t} + \theta x_{2t} + u_t$$

donde:

$$\begin{aligned} x_{1t} &= \theta^0 x_{1,t-1} + y_{t-1} \\ x_{2t} &= 2\theta^0 x_{2,t-1} - (\theta^0)^2 x_{2,t-2} - y_{t-1} + \phi^0 y_{t-2} \\ w_t &= \varepsilon_t^0 + \phi^0 x_{1t} + \theta^0 x_{2t} \\ u_t &= \varepsilon_t \end{aligned}$$

**Problema 13.17.** Obtener las expresiones para los errores de predicción y sus varianzas, que aparecen en la Sección 13.9.

**Problema 13.18.** Explique cómo obtendría sus previsiones a partir de modelos univariantes que han precisado de dos diferencias de orden regular (no estacional).

**Problema 13.19.** En el análisis empírico de la relación entre dos series temporales estacionarias,  $y_t = v(B)x_t + N_t$ , se han obtenido los siguientes resultados numéricos:

$$y_t = (1 - 0,3L)a_{yt}$$

$$(1 - 0,3L)x_t = a_{xt}$$

$$a_{yt} = \frac{2L - 2,6L^2 + 0,6L^3}{1 - 0,3L} a_{xt} + \xi_t$$

donde  $\xi_t = (1 - 0,5L)a_t$ ,  $\hat{\sigma}_\beta = 0,04$ ,  $\hat{\sigma}_{ax} = 0,03$ ,  $\hat{\sigma}_a = 0,001$  y  $\bar{a} = 0,0$ .

a) Calcule la función de respuesta al impulso, la función de correlación cruzada y la ganancia correspondiente a  $v(L)$ .

b) De haber conocido previamente la función de correlación cruzada que ha obtenido en a), ¿qué función de transferencia había especificado y qué preestimaciones habría dado a sus parámetros?

c) ¿Qué proceso seguirá  $N_t$ ? Calcule una medida de la bondad de ajuste del modelo de función de transferencia.

d) A la vista de los resultados anteriores, ¿relacionaría las variables  $y_t$  y  $x_t$  en niveles? ¿Cómo cambiaría la especificación del modelo para lograr una mayor eficiencia en la estimación?

**Problema 13.20.** Tras estimar los siguientes modelos univariantes con observaciones anuales del IPC y los ALP:

a)  $\nabla \text{IPC}_t = (1 - 0,45L)a_t$ .

b)  $(1 - 0,80L - 0,60L^2)\nabla \text{ALP}_t = (1 - 0,60L)a_t$ .

Se utilizó el modelo b) para preblanquear tanto la serie del IPC como la de los ALP. A continuación se estimó la función de correlación cruzada entre las series transformadas, obteniéndose:

Correlación entre  $\text{IPC}_t^*$  y  $\text{ALP}_{t-k}^*$ :

$k =$	0	1	2	3	4	5	6	7	8	9	10
	-0,03	0,80	0,83	0,60	0,35	0,20	0,10	-0,02	0,01	0,03	0,02

Especificar posibles modelos de función de transferencia entre IPC y ALP (incluido el modelo del ruido).

**Problema 13.21.** Con objeto de prever el nivel de la Formación Bruta de Capital Fijo ( $I_t$ ), se ha identificado el siguiente modelo de transferencia, que incluye como input la tasa de variación del stock nominal de ALP ( $M_t$ ):

$$\ln I_t = (\omega_0 - \omega_1 L)\nabla \ln M_t + N_t$$

$$\nabla N_t = (1 - \phi L)^{-1} a_t$$

1. Describa, con todo detalle, las expresiones analíticas que conducirían a la estimación de máxima verosimilitud de dicho modelo, mediante el algoritmo de Gauss-Newton.

Si se define la siguiente transformación de variables:

$$\begin{aligned} X_{1t} &= (1 - 0,1L)\nabla^2 \ln M_t \\ X_{2t} &= (0,1L - 1)\nabla^2 \ln M_{t-1} \\ X_{3t} &= \nabla \ln I_{t-1} - (0,5 + 0,2L)\nabla^2 \ln M_{t-1} \end{aligned}$$

se obtienen, a partir de la información muestral, los siguientes valores:

$$\begin{aligned} n &= 100; & x_1 &= x_2 = x_3 = x_4 = a_0 = 0; \\ \Sigma x_{1t}^2 &= 2; & \Sigma x_{2t}^2 &= 1; & \Sigma x_{3t}^2 &= 2; \\ \Sigma x_{1t}x_{2t} &= \Sigma x_{2t}x_{3t} = 0; & \Sigma x_{1t}x_{3t} &= 1; \\ \Sigma x_{1t}a_{0t} &= 2,5; & \Sigma x_{2t}a_{0t} &= -0,5; & \Sigma x_{3t}a_{0t} &= 2; \\ \Sigma a_{0t}^2 &= 4,75 \end{aligned}$$

2. Obtenga estimaciones consistentes y eficientes de los parámetros  $\omega_0$ ,  $\omega_1$ ,  $\phi$ ,  $\sigma_a^2$  y  $R^2$ .

3. Contraste la hipótesis de Superneutralidad Monetaria:  $g = 0$ .

4. Discuta el efecto que sobre  $\ln I_t$  tendrá un cambio permanente unitario sobre el nivel de  $\ln M_t$ .

**Problema 13.22.** a) Explique cómo pueden utilizarse los dos primeros valores estimados de la función de autocorrelación simple para obtener estimaciones de  $\phi_1$  y  $\phi_2$  en el modelo de AR(2):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

donde  $\varepsilon_t$  es ruido blanco.

b) Obtenga la expresión analítica de  $\sigma_y^2$  como función de los parámetros  $\phi_1$ ,  $\phi_2$  y  $\sigma_\varepsilon^2$  en el modelo anterior.

c) Obtenga las expresiones analíticas que sirven para la predicción, varios períodos hacia el futuro, en un modelo ARMA(1, 1) con término constante. Halle una expresión genérica para  $E_T y_{T+k}$ ,  $k > 0$ .

d) Suponga que para analizar la serie temporal  $y_t$  se ha estimado un modelo ARIMA( $p, d, q$ ) con  $d = 2$ . De dicho modelo se han obtenido predicciones:  $E_T z_{T+1}$ ,  $E_T z_{T+2}$ ,  $E_T z_{T+3}$ , donde  $z_t = \nabla^2 y_t$ . Obtenga expresiones para  $E_T y_{T+1}$ ,  $E_T y_{T+2}$ ,  $E_T y_{T+3}$  como función únicamente de  $E_T z_{T+i}$ ,  $i = 1, 2, 3$  y de la información muestral acerca de la variable  $y_t$ .

e) Obtenga los cuatro primeros términos de la representación AR de un modelo MA(2) invertible.

f) ¿Por qué es interesante que un modelo MA( $q$ ) sea invertible? ¿Cuáles son las condiciones de invertibilidad de un modelo MA(2)? ¿Cuáles son las condiciones de estacionariedad de dicho modelo?

**Problema 13.23.** Para llevar a cabo el seguimiento mensual de las Disponibilidades Líquidas (M3) en España, se ha especificado el modelo:

$$\Delta^2 y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

donde  $y_t$  denota el logaritmo de M3 y se han obtenido las estimaciones:

$$\hat{\theta}_1 = 0,60; \quad \hat{\theta}_2 = 0,30$$

a) Utilizando este modelo, obtener previsiones para el logaritmo de M3 para los meses de junio a diciembre de 1988, si se dispone de la siguiente historia acerca de la realización de la serie y su predicción hecha con un mes de antelación:

Mes	Valor real de M3	Predicción hecha el mes anterior
5/87	22.913	22.881
6/87	22.434	22.615
7/87	23.852	23.450
8/87	23.686	23.782
9/87	23.932	23.901
10/87	23.753	23.842
11/87	23.747	23.810
12/87	24.781	24.805
1/88	24.255	24.310
2/88	24.357	24.342
3/88	24.918	24.902
4/88	24.936	24.950
5/88	25.115	25.164

b) ¿Cómo cambiaría su respuesta si el modelo especificado hubiese sido

$$\nabla \nabla^{12} y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

con las mismas estimaciones y datos anteriores?

**Problema 13.24.** Para estimar el modelo ARIMA siguiente:

$$(1 - \phi_1 L) \nabla Z_t = (1 - \theta_1 L) a_t$$

se dispone de la siguiente información muestral:

$t$	$\nabla Z_t$
1986	1
1987	0
1988	-1
1989	0
1990	-1
1991	1

tomándose como condiciones iniciales:

$$\begin{aligned} a_{1985} &= 0 \\ \nabla Z_{1984} &= \nabla Z_{1985} = 0 \\ \left( \frac{\partial a_t}{\partial \beta} \right)_{\beta_{0,1984}} &= \left( \frac{\partial a_t}{\partial \beta} \right)_{\beta_{0,1985}} = 0 \\ \beta_0 &= (\phi_1 = 0,8, \theta_1 = 0,5)' \end{aligned}$$

1. Complete la primera iteración del algoritmo de Gauss-Newton. Calcule la matriz de varianzas y covarianzas de las estimaciones y su  $R^2$ .
2. Calcule previsiones para  $t = 1993$  y  $1994$  y los riesgos asociados.

**Problema 13.25.** Derive las condiciones de optimalidad necesarias para obtener el estimador de máxima verosimilitud del modelo:

$$\begin{aligned} y_t &= \mathbf{x}_t' \beta + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, T \end{aligned}$$

**Problema 13.26.** Pruebe que el modelo

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \eta_t \\ \beta_t &= \beta_{t-1} + \zeta_t \end{aligned}$$

donde  $\varepsilon_t, \eta_t, \zeta_t$  son ruidos blancos independientes entre sí, puede formularse como una ARIMA(0, 2, 2).

**Problema 13.27.** Igualando las funciones de autocovarianza de ambos procesos, halle el valor del parámetro  $\theta$  de modo que el proceso estocástico  $z_t = \eta_t + \varepsilon_t - \varepsilon_{t-1}$ , donde  $\eta_t$  es i. i. d.,  $N(0, \sigma_\eta^2)$  y  $\varepsilon_t$  es i. i. d.,  $N(0, \sigma_\varepsilon^2)$ , admita la representación MA(1):  $z_t = \xi_t - \theta \xi_{t-1}$ , donde  $\xi_t$  es i. i. d.,  $N(0, \sigma_\xi^2)$ . Halle asimismo la varianza de  $\xi_t$ .

**Problema 13.28.** Describa la aplicación del algoritmo Gauss-Newton a la estimación del modelo de función de transferencia:

$$\begin{aligned} y_t &= \frac{\omega_0 + \omega_1 L + \omega_2 L^2}{1 - \delta L} x_t + u_t \\ u_t &= \frac{1}{(1 - \delta L)(1 - \rho L)} a_t \end{aligned}$$

**Problema 13.29.** Suponga que se ha estimado un modelo ARIMA(0, 2, 1) con parámetro 0,30 para representar la serie trimestral de población activa española, cuyos valores durante el año 1992 fueron:

Trimestre	(1992, I)	(1992, II)	(1992, III)	(1992, IV)
Población activa (Miles de personas)	14.000	14.150	14.200	14.350

Elabore previsiones para la población activa durante los cuatro trimestres de 1993, sabiendo que la previsión que se había hecho en (1992, III) para (1992, IV) era de 14.500.

**Problema 13.30.** Tras estimar el modelo econométrico  $y_t = \mathbf{x}'_t \boldsymbol{\beta} + u_t$  con veinte años de observaciones trimestrales, por mínimos cuadrados, se calcularon las funciones de autocorrelación simple y parcial de los residuos, obteniéndose:

$k =$	0	1	2	3	4	5	6	7	8
<i>fas.</i>	1	0,50	0,20	0,04	-0,15	-0,21	-0,14	0,12	0,10
<i>fap.</i>	1	0,50	0,45	-0,15	0,18	0,12	-0,21	0,10	0,05

- a) Qué indican estas funciones acerca de la estructura estocástica del término de error? Obtenga estimaciones para los parámetros de dicha estructura.
- b) ¿Cómo debería estimarse eficientemente el modelo anterior con esta información adicional?
- c) ¿Qué relación existe entre el estimador que se acaba de proponer y el estimador de máxima verosimilitud?
- d) ¿Cómo puede calcularse una estimación para la varianza  $\sigma_u^2$ ?

**Problema 13.31.** Tras estimar el modelo  $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$  se ha identificado una estructura ARMA(1, 1) para los residuos MCO:

$$\hat{u}_t = \rho \hat{u}_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1}$$

y, como consecuencia, se decide estimar simultáneamente el sistema:

$$\begin{cases} y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t \\ u_t = \rho u_{t-1} + \varepsilon_t - \theta \varepsilon_{t-1} \end{cases}$$

Describa, en todo detalle, cómo podría obtener estimadores eficientes de  $\beta_0, \beta_1, \beta_2, \rho$  y  $\theta$ . ¿Cómo utilizaría sus series temporales para poner en práctica dicho procedimiento? ¿Podría pensarse que un procedimiento que estimase primero los coeficientes  $\beta$  y después  $\rho$  y  $\theta$  conjuntamente pudiera ser totalmente eficiente? ¿Cuál sería la matriz de covarianzas de sus estimaciones en el procedimiento eficiente que ha propuesto?

**Problema 13.32.** Explique los detalles de la utilización práctica del método de «scoring» para la estimación del modelo

$$y_t = \phi y_{t-1} + u_t$$

donde  $u_t$  es  $i. i. d.$ ,  $N(0, \sigma_u^2)$ . ¿Cuál será la matriz de covarianzas de los estimadores  $\hat{\phi}$  y  $\hat{\sigma}_u^2$ ?

# CAPITULO 14

## REGRESION CON VARIABLES NO ESTACIONARIAS

### 14.1. INTRODUCCION

#### 14.1.a. Primeras definiciones

La observación de un agregado económico como la renta de un país o la cantidad de dinero en circulación muestra una *tendencia creciente* a lo largo del tiempo. Otras variables económicas, como los tipos de interés o la tasa de inflación, quizá no tienen tal tendencia, pero presentan una clara inclinación a permanecer durante largos períodos de tiempo por encima o por debajo de su valor central en la muestra. Decimos entonces que la variable *deambula* alrededor de dicho valor central. Ambas son situaciones que reflejan la potencial *no estacionariedad* de la variable objeto de análisis.

Desde el punto de vista estadístico, la presencia de una tendencia temporal, lineal o no, de naturaleza determinista, no representa una dificultad importante. Siendo determinista, y supuesta estable, puede estimarse con una razonable precisión para, posteriormente, ser sustraída de la serie temporal original. Así se obtiene el *componente puramente estocástico* de la variable económica que se pretende analizar. Cuando realizamos tales ejercicios, observamos con frecuencia que el componente estocástico de la variable *deambula* alrededor de su valor central, es decir, nos hemos reducido al segundo de los casos antes mencionados.

Tales rachas de valores sistemáticamente por encima o por debajo de la media, se deben a la *existencia de una raíz unitaria* en la estructura estocástica de la variable. Si existe dicha raíz en su representación autorregresiva<sup>(1)</sup>:  $A(L)y_t = \varepsilon_t$ , ésta puede descomponerse:

$$A(L)y_t = A^*(L)(1 - L)y_t = A^*(L)\Delta y_t$$

<sup>(1)</sup> Que será teóricamente de orden infinito si el modelo ARIMA de  $y_t$  tiene un término invertible de medias móviles. Siguiendo con la convención habitual, denotamos por  $\Delta y_t$  la primera diferencia de la variable, que en el capítulo anterior denotamos por  $\nabla y_t$ .

Si no hay más raíces unitarias en el polinomio  $A^*(L)$ , entonces la variable diferencia  $\Delta y_t$  es estacionaria, y oscilará de modo puramente aleatorio y, en particular, sin deambular, alrededor de su valor central.

Veamos por qué suele hablarse de las raíces unitarias como de *tendencias estocásticas*. El caso más sencillo de un proceso AR(1), cuando existe una raíz unitaria, puede representarse:

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

que conduce a:

$$y_t = y_0 + ct + e_t$$

donde  $e_t = \sum_1^t \varepsilon_s$ , supuesto que el valor inicial de la variable  $y_0$  sea una constante dada, aunque desconocida. En consecuencia, el proceso  $y_t$  admite una tendencia lineal, si bien en una representación en que el término de error es la acumulación temporal de un ruido blanco y, por consiguiente, *con una varianza creciente en el tiempo, sin límite*. Ello sugiere el tipo de comportamiento que cabe esperar de tal estructura estocástica, que definimos a continuación:

**Definición 14.1.** Una variable es *integrada de orden  $d$*  si su diferencia de orden  $d$  admite una representación ARMA estacionaria e invertible. Diremos en tal caso que  $x_t$  es  $I(d)$ .

Por ejemplo, una variable cuya representación autorregresiva tiene una única raíz unitaria es  $I(1)$ . La suma de variables  $I(d)$  e  $I(d')$  con  $d < d'$  es  $I(d')$ ; si  $x_t$  es  $I(d)$ , entonces  $ax_t + b$  es también  $I(d)$ , siempre que  $a \neq 0$ .

Una variable  $I(0)$  es estacionaria y: *a)* tiene varianza finita, *b)* dicha varianza es independiente del tiempo, *c)* una perturbación sobre su nivel tiene un efecto transitorio, pero no permanente, *d)* los valores sucesivos de su función de autocorrelación decrecen rápidamente hacia cero; además, *e)* no presentará intervalos en que deambule, de modo que el número de periodos que transcurren entre dos cruces consecutivos con su valor medio debe ser reducido.

Por el contrario, si  $x_t$  es  $I(d)$  con  $d \geq 1$ : *a)* su varianza aumenta generalmente al transcurrir el tiempo, tendiendo a infinito, *b)* una perturbación sobre su nivel tiene efectos permanentes, *c)* los valores de su función de autocorrelación decrecen sólo muy lentamente hacia cero, *d)* el número esperado de periodos que deben transcurrir entre dos cruces consecutivos con su valor central muestral es infinito.

Consideremos, por simplicidad, que el modelo AR(1):

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 2, \dots, T$$

evoluciona a partir de un valor inicial  $y_1$  dado, determinista, y que  $\varepsilon_t \sim N(0, \sigma^2)$  para todo  $t$ , independiente en el tiempo. En dicho modelo, el estimador MCO

de  $\rho$  es *siempre consistente*, cualquiera que sea el verdadero valor de  $\rho$  (Rubin, 1950).

Las dificultades aparecen cuando se pretende contrastar la hipótesis de existencia de raíz unitaria en  $y_t$ :  $H_0: \rho = 1$ , es decir, de no estacionariedad de  $y_t$ , mediante el estimador MCO  $\hat{\rho}$  porque, bajo  $H_0$ , el estadístico  $(\hat{\rho} - 1)/\sqrt{\text{Var}(\hat{\rho})}$  no tiene la distribución «t» que suele asociársele. Sólo sigue una distribución  $t$  cuando el proceso  $y_t$  es estacionario, es decir, cuando  $|\rho| < 1$ .

En efecto, mientras que en el caso estacionario ( $|\rho| < 1$ ) el estadístico  $t$  de Student converge a una Normal(0, 1) cuando el número de grados de libertad tiende a infinito (White, 1958):

$$\sqrt{\frac{\sum_2^T y_{t-1}^2}{\hat{\sigma}^2}} (\hat{\rho} - \rho) \xrightarrow{d} N(0, 1)$$

en el caso explosivo ( $|\rho| > 1$ ) no se tiene tal resultado, si bien puede afirmarse que:

$$|\rho|^n \left( \frac{\hat{\rho} - \rho}{\rho^2 - 1} \right) \xrightarrow{d} \text{Cauchy}$$

El caso más interesante de existencia de raíz unitaria,  $|\rho| = 1$ , es más complejo y se tiene (véase Fuller, 1976) que  $T(\hat{\rho} - 1)$  converge en distribución a un cociente de integrales de Wiener, que no estudiamos en este texto. A modo de ilustración, es interesante apuntar que  $\text{Prob}(|\hat{\rho}| < 1/|\rho| = 1)$  tiende a 0,68 cuando el tamaño muestral tiende a infinito.

Puede, por tanto, caracterizarse la distribución de probabilidad asintótica del estimador MCO del modelo AR(1) en todos los casos, excepto cuando  $|\rho| = 1$ , que representa, por consiguiente, una discontinuidad. Tal discontinuidad se produce, sin embargo, estrictamente en el límite, ya que ligeramente por encima o por debajo de tal punto del espacio paramétrico la distribución asintótica pertenece a una determinada clase, Normal o Cauchy.

### 14.1.b. Regresión entre procesos no estacionarios

Existen muchas situaciones en Economía en que variables que son no estacionarias muestran una relación estable a través del tiempo, lo que sugiere la existencia de una relación de equilibrio a largo plazo entre ellas. Por ejemplo, para distintos países, el Consumo agregado  $C_t$  y la Renta disponible  $Y_t$  son variables  $I(1)$ ; sin embargo, examinado a largo plazo, el Consumo viene a ser una proporción bastante estable de la Renta, por lo que existe una constante  $\beta$  para la que la diferencia  $C_t - \beta Y_t$  es estacionaria.

La teoría de la Paridad del Poder Adquisitivo propone que, en ausencia de costes de transacción u otras fricciones, bienes análogos deberían tener el mismo precio efectivo en distintos países. Así, si  $P_t$  y  $P_t^*$  denotan el nivel de precios en España y Estados Unidos respectivamente y  $E_t$  denota el tipo

de cambio (ptas./\$), debería tenerse para bienes análogos que  $P_t = E_t P_t^*$  y, tomando logaritmos:

$$p_t = e_t + p_t^*$$

que debería cumplirse exactamente, donde las letras minúsculas denotan los logaritmos de sus análogos mayores. Sin embargo, fricciones como las indicadas pueden impedir tal exactitud, por lo que una versión más débil de la teoría propone que la diferencia

$$z_t = p_t - e_t - p_t^*$$

sea estacionaria.

Todas estas son situaciones en que un conjunto de variables económicas están sujetas a una relación estable a largo plazo que, siendo lineal, puede representarse por un modelo de regresión:  $y_t = \mathbf{x}_t' \boldsymbol{\beta} + u_t$ . Si las variables  $\mathbf{x}_t$  e  $y_t$  son no estacionarias, la regresión anterior sólo puede reflejar una relación de equilibrio si las posibles tendencias de las variables  $\mathbf{x}_t$  e  $y_t$  evolucionan conjuntamente, de modo que los residuos del modelo sean estacionarios, a pesar de que las variables que en él aparecen no lo sean.

Los comentarios de la sección anterior sugieren que los procedimientos de inferencia habituales dejan de ser válidos en un modelo autorregresivo cuando la variable que en él aparece no es estacionaria. Pero esta cuestión es aún más importante de lo que a primera vista puede parecer, por cuanto que los problemas pueden surgir también en modelos de regresión, cuando las variables que en él se incluyen no sean estacionarias. Un trabajo pionero en este área es el de Granger y Newbold (1974), en el que los autores generaron datos para dos variables,  $x_t$  e  $y_t$ , con estructura de camino aleatorio cuyas respectivas innovaciones eran independientes. Es decir, no existía ninguna relación entre las series temporales así generadas para ambas variables, a pesar de lo cual, al estimar por MCO un modelo de regresión lineal:

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, 2, \dots, T \quad [14.1]$$

estos autores observaron con mucha frecuencia que el estadístico  $t$  del coeficiente  $\beta$  conducía a rechazar la hipótesis nula de ausencia de relación entre ambas variables:  $H_0: \beta = 0$ , lo que comenzó a conocerse con el problema de *regresión espúrea*.

Más recientemente, se ha desarrollado la teoría estadística de regresión entre variables no estacionarias, que ha permitido explicar este tipo de resultados. En particular, ahora se sabe (Phillips, 1987) que en una regresión como [14.1], con variables no estacionarias, el estadístico «del tipo  $t$  de Student» que se obtiene dividiendo  $\hat{\beta} - \beta_0$ , donde  $\beta_0$  es el valor de  $\beta$  que se contrasta, por la desviación típica de  $\hat{\beta}$ , no se distribuye como una  $t$  de Student. En realidad, en algunas ocasiones, los estadísticos «tipo  $t$ » tienden a infinito con el tamaño muestral, por lo que al aumentar éste tenderá a

rechazarse la hipótesis nula de ausencia de relación en [14.1]. Esto constituye una posible explicación al problema de la regresión espúrea.

Una situación así, en que los procedimientos de inferencia habituales están sesgados hacia aceptar la existencia de relación entre variables no estacionarias, viene generalmente acompañada de un  $R^2$  aceptable e incluso, en ocasiones, muy elevado, junto con un estadístico Durbin-Watson muy bajo, por razones que luego veremos.

Es fácil entender que el problema es importante, porque una variable explicativa  $x_t$  en el modelo [14.1] que no sea estacionaria no satisfará la propiedad  $plim(T^{-1}\Sigma_1^T x_t^2) = \text{constante}$ , que es uno de los supuestos que justifican la utilización de los procedimientos de inferencia habituales, basados en el estimador MCO.

De este modo, la mayor parte de los resultados estadísticos habituales y, en particular, aquellos relacionados con las propiedades del estimador mínimo-cuadrático en modelos lineales dependen, precisamente, de supuestos, no siempre explícitos, acerca de la estacionariedad del término de error del modelo correspondiente. Tras puntualizar la importancia que tiene tal supuesto para llevar a cabo el análisis de inferencia estadística en un modelo de regresión, la segunda sección de este capítulo se dedica a la contrastación de la existencia de raíz unitaria en la serie temporal correspondiente a una variable. La Sección 14.7 aplica dichos resultados al análisis de estacionariedad de los residuos de un modelo de regresión, al objeto de garantizar que la utilización estadística de las estimaciones obtenidas es totalmente rigurosa. El capítulo termina con la discusión de algunas aplicaciones.

En este último contexto de existencia de raíz unitaria, los trabajos de Dickey y Fuller (1979) y (1981) han resultado enormemente influyentes. En ellos se establece una metodología sencilla para el contraste de existencia de raíz unitaria en un proceso estocástico  $y_t$ , que examinamos a continuación.

## 14.2. CONTRASTES DE RAZ UNITARIA DE DICKEY Y FULLER

Comenzamos pasando revista en esta sección a los contrastes de raíz unitaria en un proceso AR(1), para generalizarlos a otros tipos de estructuras estocásticas en secciones posteriores. Examinamos sucesivamente tres versiones del modelo, según cuáles sean sus componentes deterministas:

A) Modelo sin componentes deterministas:

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.2]$$

con  $y_1$  fijo y  $\varepsilon_t \sim N(0, \sigma^2)$ , independiente en el tiempo. Si no se rechaza la hipótesis nula, se dice que el proceso  $y_t$  es *estacionario en diferencias*.

B) Incluyendo un término constante en el modelo AR(1):

$$y_t = \alpha + \rho y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.3]$$

y manteniendo el supuesto de que  $y_1$  es fijo y  $\varepsilon_t \sim N(0, \sigma^2)$ , independiente en el tiempo. Si no se rechaza la hipótesis nula conjunta:  $H_0: \alpha = 0, \rho = 1$ , se tiene nuevamente que el proceso  $y_t$  es *estacionario en diferencias*.

C) Incluyendo, además de la constante, una tendencia lineal:

$$y_t = \alpha + \beta t + \rho y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.4]$$

con  $y_1$  fijo y  $\varepsilon_t \sim N(0, \sigma^2)$ , independiente en el tiempo. Si rechazando la hipótesis nula concluimos que  $\rho < 1$  y  $\beta \neq 0$ , se dice que el proceso  $y_t$  es *estacionario alrededor de una tendencia lineal*.

En la práctica, se ha popularizado la utilización de los contrastes sobre el modelo AR especificado en primeras diferencias. Para ello, basta observar que la contrastación de la hipótesis nula  $H_0: \rho = 1$  en el modelo [14.4] equivale a la contrastación de la hipótesis  $H_0: v = 0$  en el modelo:

$$\Delta y_t = \alpha + \beta t + v y_{t-1} + \varepsilon_t \quad [14.5]$$

mientras que las variantes A) y B) se obtienen al imponer en [14.5] algunas de las restricciones  $\alpha = 0, \beta = 0$ .

Para contrastar la hipótesis nula  $H_0: v = 0$  en el modelo

$$\Delta y_t = v y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.6]$$

con  $y_1$  fijo y  $\varepsilon_t \sim N(0, \sigma^2)$ , independiente en el tiempo, puede utilizarse el estadístico  $t$  construido a partir del estimador MCO de  $v$ :

$$\frac{\hat{v}}{[\text{Var}(\hat{v})]^{1/2}} = \frac{\hat{v}}{\hat{\sigma}} \left( \sum_2^T y_{t-1}^2 \right)^{1/2}$$

pero, como ya hemos mencionado, su distribución bajo  $H_0$  ya no es del tipo « $t$  de Student», y hay que comparar su valor numérico, que denotamos por  $\tau$ , con los que aparecen en la Tabla A.11 al final de este texto [Tabla 8.5.2 de Fuller (1976), pág. 373].

En cualquiera de estos tres modelos existe una forma alternativa de llevar a cabo el contraste de la hipótesis nula de existencia de raíz unitaria, comparando los valores del producto  $T(\hat{\rho} - 1)$  con los de la Tabla A.12 [Tabla 8.5.1 de Fuller (1976), pág. 371]. Si el valor numérico del estadístico excede al de las tablas, rechazamos la hipótesis nula  $H_0: \hat{\rho} = 1$ , mientras que la mantenemos en otro caso.

Como resumen de dichas tablas, que aparecen al final de este libro, si queremos contrastar la hipótesis nula de existencia de raíz unitaria, utilizaremos los valores críticos que se muestran en las Tablas 14.1 y 14.2.

**TABLA 14.1.** Contraste de raíz unitaria

Valor crítico $c$ tal que $P(\tau < c) = x$ para $x = 0,01; 0,025; 0,05$						
	$T = 25$			$T = 100$		
Nivel de significación: $x$	0,01	0,025	0,05	0,01	0,025	0,05
Modelo [14.6]	-2,66	-2,26	-1,95	-2,60	-2,24	-1,95
Modelo [14.7]	-3,75	-3,33	-3,00	-3,51	-3,17	-2,89
Modelo [14.8]	-4,38	-3,95	-3,60	-4,04	-3,73	-3,45

Por ejemplo, si se ha estimado el modelo [14.3] con 25 observaciones y se aplica un contraste de hipótesis convencional, se rechazaría la hipótesis nula de existencia de raíz unitaria si el estadístico  $t$  del retardo  $y_{t-1}$  resultase inferior a  $-1,71$ , que es el valor crítico para el contraste de una cola al 95 por 100 de dicha distribución. Sin embargo, de acuerdo con la Tabla 14.1, debe rechazarse  $H_0$  sólo si el estadístico  $t$  estuviese por debajo de  $-3,0$ , es decir, mucho menos a menudo de lo que sugiere el análisis tradicional.

Si el modelo incluye una constante

$$\Delta y_t = \alpha + \nu y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.7]$$

puede contrastarse la hipótesis nula  $H_0: \alpha = 0$  mediante el estadístico  $t$  de dicho coeficiente, que denotamos como  $\tau_{\alpha\mu}$ , que debe compararse con los valores críticos de la Tabla A.13 al final de este texto (Tabla I en Dickey y Fuller, 1981). Si no se rechaza la hipótesis nula, estamos en el modelo [14.6]. Por otra parte, el contraste de la hipótesis conjunta  $H_0: \alpha = 0, \nu = 0$ , es decir, que  $y_t$  es un paseo aleatorio, puede llevarse a cabo a partir del estadístico  $F$

**TABLA 14.2.** Contraste de raíz unitaria

Valor crítico $c$ tal que $P[T(\hat{\rho} - 1) < c] = x$ para $x = 0,01; 0,025; 0,05$						
	$T = 25$			$T = 100$		
Nivel de significación: $x$	0,01	0,025	0,05	0,01	0,025	0,05
Modelo [14.6]	-11,9	-9,3	-7,3	-13,3	-10,2	-7,9
Modelo [14.7]	-17,2	-14,6	-12,5	-19,8	-16,3	-13,7
Modelo [14.8]	-22,5	-19,9	-17,9	-27,4	-23,6	-20,7

correspondiente,  $\Phi_1$ , sólo que comparando esta vez con la Tabla A.16 [la Tabla IV en Dickey y Fuller (1981)].

En el caso del modelo con constante y tendencia:

$$\Delta y_t = \alpha + \beta t + \nu y_{t-1} + \varepsilon_t, \quad t = 2, 3, \dots, T \quad [14.8]$$

con  $y_1$  fijo y  $\varepsilon_t \sim N(0, \sigma^2)$ , independiente en el tiempo, la significación de la constante  $H_0: \alpha = 0$  puede contrastarse mediante el estadístico  $t$  del estimador MCO de dicho parámetro,  $\tau_{\alpha}$ , utilizando la Tabla A.14 [Tabla II de Dickey y Fuller (1981)].

La significación estadística de la pendiente determinista  $H_0: \beta = 0$  puede contrastarse mediante el estadístico  $t$  del estimador MCO de dicho parámetro,  $\tau_{\beta}$ , utilizando la Tabla A.15 [Tabla III de Dickey y Fuller (1981)]. Si no se rechaza la hipótesis nula, estamos en la especificación [14.7] anteriormente discutida.

Por último, la existencia de una raíz unitaria puede contrastarse en una versión restringida:  $H_0: \alpha = 0, \beta = 0, \nu = 0$  (es decir,  $y_t$  es paseo aleatorio), utilizando el estadístico  $F$  correspondiente,  $\Phi_2$ , y la Tabla A.17 [Tabla V de Dickey y Fuller (1981)], mientras que en su versión menos restrictiva:  $H_0: \beta = 0, \nu = 0$  ( $y_t$  es paseo aleatorio con deriva), puede contrastarse utilizando el estadístico  $F$  correspondiente,  $\Phi_3$  —diferente del anterior—, y la Tabla A.18 [Tabla VI de Dickey y Fuller (1981)]. La utilización de estas tablas, construidas por Dickey y Fuller, en vez de las habituales, puede representar diferencias importantes. Así, para este contraste conjunto, los valores críticos de Dickey-Fuller para el estadístico  $F$  al 5 por 100 son 7,24; 6,73 y 6,49, según que el tamaño muestral empleado sea  $T = 25, 50$  ó  $100$ . Por otra parte, los valores de la tablas habituales  $F_{2,22}$ ,  $F_{2,47}$  y  $F_{2,97}$  son 3,44; 3,20 y 3,10, por lo que su utilización conduciría a rechazar la hipótesis nula de existencia de raíz unitaria *demasiado frecuentemente*.

Al efectuar estos contrastes de existencia raíz unitaria, es decir, de significación del parámetro  $\nu$  en los modelos [14.2], [14.3] y [14.4], hay que tener en cuenta que puesto que la hipótesis alternativa es de estacionariedad, es decir,  $\rho < 1$ , el contraste debe ser de una cola. En segundo lugar, que es fundamental utilizar las tablas a las que nos hemos venido refiriendo, porque los estadísticos «de tipo  $t$ » y «de tipo  $F$ » para el contraste de las hipótesis individuales y conjuntas que hemos mencionado no tienen las distribuciones  $t$  o  $F$  que habitualmente les asignaríamos. Lamentablemente, ya hemos visto que los valores críticos para estos contrastes dependen tanto del tamaño muestral como de la especificación del modelo en que se llevan a cabo, por lo que hay que poner especial cuidado en utilizar la tabla adecuada en cada caso.

Es importante observar que los valores de los estadísticos de Dickey-Fuller que aparecen en estas tablas están obtenidos bajo los supuestos: *a*) el proceso sobre el que se contrasta la existencia de una raíz unitaria tiene una estructura AR(1), y *b*) no existe autocorrelación en su término de error. Por otra parte, las distribuciones empíricas que se presentan en las tablas al final del texto, y se resumen en nuestra Tabla 14.1, no cambian al introducir en el modelo retardos de la variable en diferencias:  $\Delta y_{t-1}$ ,  $\Delta y_{t-2}$ , etc., lo que da lugar al

contraste conocido como *Dickey-Fuller ampliado* (que denotamos en lo sucesivo por DFA) en el que se estima una regresión que en el caso del modelo [14.3] anterior sería:

$$\Delta y_t = \alpha + \rho y_{t-1} + \delta_1 \Delta y_{t-1} + \delta_2 \Delta y_{t-2} + \delta_3 \Delta y_{t-3} + \delta_4 \Delta y_{t-4} + u_t$$

y se compara el estadístico  $t$  obtenido para  $\rho$  con las tablas de Dickey y Fuller (las mismas de antes). El contraste DFA se utiliza para eliminar, en lo posible, la autocorrelación del término de error, pues, como hemos dicho, éste es un supuesto bajo el que fueron obtenidas las distribuciones empíricas del estadístico  $t$ . Esto es especialmente importante cuando se trabaja con datos trimestrales, por ejemplo, en cuyo caso la inclusión de cuatro diferencias retardadas de la variable endógena producirá, generalmente, un residuo prácticamente libre de correlación.

Al contrastar la posible existencia de una raíz unitaria, debe escogerse uno de los modelos [14.2], [14.3] o [14.4] como representación más adecuada de la variable en estudio, pues, como hemos visto, ello condiciona el modo de llevar a cabo el contraste. Tal elección debe hacerse tras un examen preliminar de la serie temporal correspondiente, que nos sugiera si hay evidencia de constante (si la media muestral de  $y_t$  es distinta de cero), posible tendencia determinista (si la media muestral de  $\Delta y_t$  es distinta de cero), etc.

### 14.3. CONTRASTACION EN MODELOS AUTORREGRESIVOS DE ORDEN SUPERIOR

El análisis que hasta ahora hemos llevado a cabo se refiere al caso en que el analista considera que la variable endógena obedece a un modelo autorregresivo del primer orden. Supongamos, alternativamente, que el modelo adecuado es de segundo orden:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t, \quad t = 1, 2, \dots, T$$

y supongamos que su polinomio característico acepta dos raíces reales, de modo que puede descomponerse:  $1 - \phi_1 L - \phi_2 L^2 = (1 - \lambda_1 L)(1 - \lambda_2 L)$ , donde  $\lambda_1 + \lambda_2 = \phi_1$  y  $\lambda_1 \lambda_2 = -\phi_2$ . Para que el proceso sea estacionario, es preciso que las dos raíces,  $\lambda_1$  y  $\lambda_2$ , sean inferiores a la unidad en valor absoluto. Vamos a examinar el modo de contrastar una posible situación de no estacionariedad, en que una de las raíces,  $\lambda_1$ , es igual a 1, siendo el valor absoluto de  $\lambda_2$  inferior a la unidad. En función de los coeficientes estimados, debemos contrastar  $H_0: \phi_1 + \phi_2 = 1$ , condicionado a  $|\phi_2| < 1$ , puesto que  $\lambda_1 = 1$  implica que  $\lambda_2 = -\phi_2$ , es decir:  $1 - \phi_2 = \phi_1$ . Para ello, lo más sencillo es escribir el modelo en la forma:

$$y_t = (\phi_1 + \phi_2)y_{t-1} - \phi_2(y_{t-1} - y_{t-2}) + u_t, \quad t = 1, 2, \dots, T$$

o, equivalentemente:

$$\Delta y_t = (\phi_1 + \phi_2 - 1)y_{t-1} - \phi_2 \Delta y_{t-1} + u_t, \quad t = 1, 2, \dots, T$$

en el que el contraste de significación del coeficiente de  $y_{t-1}$  equivale al contraste de existencia de una raíz unitaria (pero no más) en el modelo AR(2) original.

También podríamos añadir a este modelo una constante, así como una tendencia determinista. Dependiendo de cómo se formulara finalmente el modelo, el estadístico  $t$  de  $y_{t-1}$  se compararía con los valores críticos de una u otra de las filas en la Tabla 14.1 (o Tablas A.11 y A.12).

Más generalmente, sean  $\lambda_j$ ,  $j = 1, \dots, p$  los valores propios del polinomio característico de  $y_t$  del modelo:

$$y_t = \alpha_0 + \beta t + \sum_{i=1}^p \rho_i y_{t-i} + \varepsilon_t \quad [14.9]$$

es decir, las raíces de la ecuación  $1 - \rho_1 L - \rho_2 L^2 - \dots - \rho_p L^p$ , y supongamos que se quiere contrastar la existencia de exactamente una raíz unitaria:  $H_0: \lambda_1 = 1, |\lambda_i| < 1, i = 2, \dots, p$ . Para ello, se lleva a cabo un contraste del tipo DFA, estimando por MCO el modelo:

$$\Delta y_t = \alpha_0 + \beta t + \gamma_1 y_{t-1} + \sum_{i=1}^{p-1} \gamma_{2i} \Delta y_{t-i} + \varepsilon_t \quad [14.10]$$

que es una reparametrización del anterior, donde el orden  $p$  se ha escogido de modo que  $\varepsilon_t$  sea ruido blanco.

## 14.4. CONTRASTACION EN MODELOS CON ESTRUCTURA MA

Puesto que todo modelo ARIMA( $p, 1, q$ ) puede aproximarse por una estructura AR finita, Said y Dickey (1984) propusieron utilizar un contraste del tipo DFA escogiendo  $s$  aproximadamente igual a la raíz cúbica del tamaño muestral y aplicando los contrastes anteriores a un modelo AR( $s$ ).

Si se conocieran los órdenes  $p$  y  $q$  de los polinomios autorregresivo y de medias móviles  $\phi(L)$  y  $\theta(L)$  del proceso y se pretendiera contrastar la hipótesis ARIMA( $p, 1, q$ ):  $H_0: \phi(L)\Delta y_t = \theta(L)\varepsilon_t$  frente a la hipótesis ARIMA( $p, 0, q$ ):  $H_1: \phi(L)(1 - \rho L)y_t = \theta(L)\varepsilon_t$ , donde se supone que el polinomio  $\phi(L)$  no tiene raíces unitarias, puede estimarse el modelo mediante una iteración del algoritmo Gauss-Newton a partir de estimaciones iniciales consistentes. El estadístico  $t$  del parámetro  $\rho$  resultante tiene la misma distribución que el estadístico  $\tau$  de la Tabla A.11 [Tabla 8.5.2 en Fuller (1976), pág. 373]. Si se utilizase la variable  $y_t$  transformada en diferencias respecto a su media muestral, la distribución del estadístico  $t$  sería la misma que la de  $\tau_\mu$  en la misma tabla.

Molinas (1986) y Schwert (1987), (1989) han mostrado que la potencia de los tests de Dickey-Fuller pueden verse afectados en muestras finitas, incluso

si son largas, por la estructura del verdadero modelo generador de los datos. Por ejemplo, en un modelo ARIMA(0, 1, 1) con parámetro MA próximo a 1, los estadísticos DFA que hemos propuesto en las secciones previas tienen valores críticos muy por debajo de los valores tabulados por Dickey y Fuller, lo que, de no ser tenido en cuenta, haría que ambos contrastes sugirieran la estacionariedad de las series con excesiva frecuencia. La idea es que, como es sabido, un término MA(1) con coeficiente elevado aproxima una diferencia (es decir, una raíz unitaria) de la variable, por lo que la compensación aproximada de ambas hace que tal raíz unitaria pudiera no detectarse. En cambio, el estadístico de Said-Dickey que antes propusimos parece funcionar bien incluso en tales situaciones, con independencia del valor del parámetro de medias móviles.

## 14.5. CONTRASTE DE $k$ RAICES UNITARIAS

Algunas variables económicas pueden contener más de una raíz unitaria, como sugiere el análisis del tipo Box-Jenkins de las mismas. Por ello, es interesante disponer de estrategias de contraste de varias raíces unitarias. Aunque los contrastes que hemos presentado en la sección anterior pueden utilizarse para las sucesivas diferencias de una variable, Dickey y Pantula (1987) sugieren una secuencia de contrastes, comenzando del máximo número posible ( $k$ ) de raíces unitarias, descendiendo el valor de  $k$  cada vez que se rechaza  $H_0$ , y parando en cuanto que no se rechace. Ello tiene en cuenta que contrastamos cada vez la presencia de una raíz unitaria y que, para aplicar los estadísticos descritos en las secciones anteriores, es preciso que el modelo especificado bajo la hipótesis alternativa sea estacionario.

Por consiguiente, si  $k = 3$ , comenzaríamos estimando el modelo  $\Delta^3 y_t = \xi_0 + \xi_1 \Delta^2 y_{t-1} + \varepsilon_t$  para obtener el estadístico  $t$  de  $\xi_1$ . Compararemos dicho estadístico con los valores críticos en Fuller (1976) y rechazamos  $H_3$ : la variable tiene *exactamente* tres raíces unitarias, si  $t \leq \tau_\mu$ . A continuación estimaríamos el modelo  $\Delta^3 y_t = \xi'_0 + \xi'_1 \Delta^2 y_{t-1} + \xi'_2 \Delta y_{t-1} + \varepsilon_t$  para obtener los estadísticos  $t$  de  $\xi'_1$  y  $\xi'_2$ . Rechazaremos  $H_2$ : *exactamente* dos raíces unitarias, si además de  $\hat{t}(\xi'_1) \leq \tau_\mu$  se tiene  $\hat{t}(\xi'_2) \leq \tau_\mu$ . Finalmente, estimaríamos la regresión  $\Delta^3 y_t = \xi''_0 + \xi''_1 \Delta^2 y_{t-1} + \xi''_2 \Delta y_{t-1} + \xi''_3 y_{t-1} + \varepsilon_t$ , rechazando  $H_1$ : *exactamente* una raíz unitaria, si  $t(\xi''_i) \leq \tau_\mu$  para  $i = 1, 2, 3$ .

## 14.6. INTEGRACION Y ESTACIONALIDAD

Un proceso  $y_t$  es *integrable de órdenes* ( $d, D$ ), lo que se representa por  $y_t \sim I(d, D)$ , si  $\Delta^d \Delta_s^D y_t$  admite una representación ARMA estacionaria, invertible, donde  $\Delta_s$  denota la diferencia de orden estacional. En nuestra descripción suponemos que se trabaja con datos trimestrales, por lo que  $s = 4$  y  $\Delta_4 = 1 - L^4$ .

### 14.6.a. Raíz unitaria estacional

Para contrastar  $H_0: I(0, 1)$  frente a  $I(0, 0)$ , es decir:  $y_t$  tiene una raíz unitaria estacional pero ninguna regular, Dickey, Hasza y Fuller (1984) (DHF) han mostrado que la distribución límite del estadístico no varía al eliminar las medias estacionales de modelos autorregresivos, por lo que se sugiere comenzar con una regresión de  $y_t$  sobre variables ficticias estacionales, junto con una constante y una tendencia, con el objeto de eliminar la estacionalidad determinista, y utilizar los residuos  $\hat{u}_t$  de dicha regresión en:

$$\Delta_4 \hat{u}_t = \theta_0 + \sum_1^p \theta_i \Delta_4 \hat{u}_{t-i} + \varepsilon_t$$

definiendo después la variable filtrada:

$$z_t = \hat{\theta}(L)\hat{u}_t = (1 - \hat{\theta}_1 L - \hat{\theta}_2 L^2 - \dots - \hat{\theta}_p L^p)\hat{u}_t$$

y estimando, finalmente:

$$\Delta_4 \hat{u}_t = \xi_0 z_{t-4} + \sum_1^p \xi_i \Delta_4 \hat{u}_{t-1} + \varepsilon_t$$

El estadístico  $t$  del parámetro  $\xi_0$ , denotado por  $\tau_{\mu 4}$ , se compara con los valores críticos:

**TABLA 14.3.** Valores críticos para el contraste DHF

Nivel de significación	$T = 40$	$T = 60$	$T = 80$	$T = 200$
0,01	-5,01	-4,85	-4,78	-4,67
0,05	-4,21	-4,14	-4,11	-4,06

La limitación de este tipo de contraste es que no considera todas las raíces unitarias que puede haber en un proceso estacional. En efecto, con datos trimestrales, el operador  $\Delta_4$  puede descomponerse:  $(1 - L^4) = (1 - L)(1 + L)(1 - iL)(1 + iL)$ , por lo que el proceso estacional puede tener cuatro raíces diferentes, todas ellas de módulo unidad: 1,  $-1$ ,  $i$ ,  $-i$ . La *segunda* raíz, al actuar sobre  $y_t$  genera la ecuación en diferencias:  $y_t = -y_{t-1}$ , es decir:  $y_t = y_{t-2}$ , por lo que corresponde a un ciclo que se completa en dos períodos. por lo que, con datos trimestrales, generará dos ciclos por año. La *primera* raíz corresponde a la ecuación  $y_t = y_{t-1}$ , un ciclo de un período. Por último, es sencillo comprobar que las dos raíces complejas generan la condición:  $y_t = y_{t-4}$  y por ello un ciclo de cuatro períodos, es decir, un ciclo por año. Estas dos raíces se consideran conjuntamente en el contraste de raíces unitarias estacionales.

Engle, Granger, Hylleberg y Yoo (1987) propusieron el siguiente procedimiento:

- a) Obtener el estimador  $\hat{\theta}(L)$  como en el contraste DHF.
- b) Calcular las series filtradas:

$$\begin{aligned} z_{1t} &= \hat{\theta}(L)(1 + L + L^2 + L^3)\hat{u}_t \\ z_{2t} &= -\hat{\theta}(L)(1 - L + L^2 - L^3)\hat{u}_t \\ z_{3t} &= -\hat{\theta}(L)(1 - L^2)\hat{u}_t \end{aligned}$$

- c) Estimar:

$$\Delta_4 \hat{u}_t = \pi_1 z_{1t-1} + \pi_2 z_{2t-1} + \pi_3 z_{3t-2} + \sum_{i=1}^p \pi_{4i} \Delta_4 \hat{u}_{t-i} + \varepsilon_t$$

para analizar la significatividad estadística de los coeficientes  $\pi_1, \pi_2, \pi_3$ , mediante sus estadísticos  $t$ . Al interpretar los resultados, debe observarse que el polinomio que define  $z_{1t}$  es el resultado de extraer de  $\Delta_4$  el término  $1 - L$ ; por consiguiente, incluir  $z_{1t}$  como regresor es similar a incorporar  $y_t$  como variable explicativa en [14.5], [14.6] y [14.7] al contrastar la existencia de la raíz  $1 - L$ . Por otra parte, el polinomio que define  $z_{2t}$ , es el resultado de extraer  $1 + L$ , mientras que el que define  $z_{3t}$ , resulta de extraer las dos raíces complejas, es decir,  $1 + L^2$ .

Así, si  $y_t$  es  $I(0, 0)$ , los tres estadísticos deben ser conjuntamente significativos; si únicamente el estadístico  $t$  de  $\pi_1$  es no significativo, entonces  $y_t$  es  $I(1, 0)$ , mientras que si alguno de los estadísticos  $t$  de  $\pi_2$  o  $\pi_3$  es no significativo, entonces  $y_t$  es  $I(0, 1)$ , pues sólo una raíz unitaria estacional podría explicar tal ausencia de significación. Los valores críticos aparecen en la referencia citada.

### 14.6.b. Raíz unitaria regular, junto con raíz unitaria estacional

Hasza y Fuller (1982) (HF) propusieron contrastar la hipótesis nula  $H_0: y_t \sim I(1, 1)$ , estimando un modelo autorregresivo para la variable doblemente diferenciada:

$$\Delta \Delta_4 \hat{u}_t = \psi_0 + \sum_{i=1}^p \psi_i \Delta \Delta_4 \hat{u}_{t-i} + \varepsilon_t$$

para, posteriormente, obtener las variables filtradas:

$$z_{4t} = \hat{\psi}(L)\Delta_4 \hat{u}_t = (1 - \hat{\psi}_1 L - \dots - \hat{\psi}_p L^p)\Delta_4 \hat{u}_t \quad y \quad z_{5t} = \hat{\psi}(L)\Delta u_t$$

y, finalmente, estimar el modelo:

$$\Delta \Delta_4 \hat{u}_t = \phi_1 z_{4t-1} + \phi_2 z_{5t-4} + \sum_{i=1}^p \phi_{3i} \Delta \Delta_4 \hat{u}_{t-i} + \varepsilon_t$$

Existe evidencia a favor de la hipótesis nula  $H_0: y_t \sim I(1, 1)$  frente a la alternativa compuesta  $H_1: y_t \sim I(0, 0), y_t \sim I(0, 1), y_t \sim I(1, 0)$  cuando se

TABLA 14.4. Valores críticos del contraste de Hasza y Fuller (1982)

$T = 100$			
<i>Nivel de significación</i>	$H_0: \phi_1 = 0$	$H_0: \phi_2 = 0$	$H_0: \phi_1 = \phi_2 = 0$
Constante y variables ficticias			
0,01	-2,61	-2,79	5,98
0,05	-1,95	-2,05	3,81
Sin constantes ni variables ficticias			
0,01	-2,48	-2,59	4,99
0,05	-1,87	-1,93	3,20

tiene:  $\phi_1 = \phi_2 = 0$ , lo que puede contrastarse mediante el estadístico  $F$  correspondiente. Los estadísticos  $t$  de  $\phi_1$  y  $\phi_2$  pueden utilizarse por separado para contrastar la misma  $H_0$  frente a  $H_1: y_t \sim I(0, 1)$ , si:  $\phi_1 = 0, \phi_2 < 0$ , o a  $H_1: y_t \sim I(1, 0)$ , si:  $\phi_1 < 0, \phi_2 = 0$ , respectivamente.

Los valores críticos para estos contrastes son los que se muestran en la Tabla 14.4.

Alternativamente, si denotamos por  $Z_1, Z_2$  y  $Z_3$  los filtros

$$Z_1 = (1 + L + L^2 + L^3), Z_2 = -(1 - L + L^2 - L^3), Z_3 = -(1 - L^2)$$

de modo que, por ejemplo:  $Z_3 y_t = -y_t + y_{t-2}$ , el contraste de Hylleberg, Engle, Granger y Yoo (1990) (HEGY) de raíz unitaria estacional se basa en la regresión:

$$\Delta \Delta_4 y_t = \alpha_0 + \alpha_1 (Q_{1t} - Q_{4t}) + \alpha_2 (Q_{2t} - Q_{4t}) + \alpha_3 (Q_{3t} - Q_{4t}) + \pi_1 Z_1 y_{t-1} + \pi_2 Z_2 y_{t-1} + \pi_3 Z_3 y_{t-1} + \pi_4 Z_3 y_{t-2} + \sum_{i=1}^p \phi_i \Delta \Delta_4 y_{t-i} + \varepsilon_t$$

aceptándose la existencia de las raíces unitarias  $1, -1, i$  y  $-i$  si los coeficientes  $\pi_i$  son no significativos para  $i = 1, 2, 3$  y  $4$ , respectivamente. La hipótesis nula es  $y_t \sim I(1, 1)$  y se consideran como alternativas  $y_t \sim I(2, 0)$  e  $y_t \sim I(1, 0)$ . Por ejemplo, la significación conjunta de  $\pi_2, \pi_3$  y  $\pi_4$  sugiere que  $y_t \sim I(2, 0)$ , mientras que si, además,  $\pi_1$  también es significativa, entonces  $y_t \sim I(1, 0)$ . También en este contraste puede sustituirse  $y_t$  por  $\hat{u}_t$ .

Los valores críticos para estos contrastes (Hylleberg, Engle, Granger y Yoo, 1990) son los que se muestran en la Tabla 14.5.

## 14.7. ESTACIONARIEDAD Y COINTEGRACION

En las secciones anteriores hemos presentado diversos procedimientos para contrastar la hipótesis de existencia de raíces unitarias en una variable

**TABLA 14.5.** Valores críticos de Hylleberg, Engle, Granger y Yoo (1990) para 100 observaciones

	$\pi_1 = 0$	$\pi_2 = 0$	$\pi_3 = 0$	$\pi_4 = 0$	$\pi_3 = \pi_4 = 0$
Con constante					
0,01	-3,47	-2,61	-2,61	-2,38	4,77
0,05	-2,88	-1,95	-1,90	-1,68	3,08
Con constante y variables ficticias					
0,01	-3,55	-3,60	-2,56	-2,78	8,74
0,05	-2,95	-2,94	-1,89	-1,96	6,57
Con tendencia, constante y variables ficticias					
0,01	-4,09	-3,60	-4,12	-2,76	8,79
0,05	-3,53	-2,94	-3,48	-1,94	6,60

económica. Supongamos dos variables económicas como Consumo y Renta Disponible, entre las que creemos que existe una relación de equilibrio estable a largo plazo. Cabe esperar que, bajo tal supuesto, los residuos de la regresión que explica el gasto en Consumo de las familias mediante su Renta Disponible sean estacionarios, a pesar de que ninguna de las dos variables del modelo lo sean. Ello sería debido a que su evolución temporal es, en gran medida, común. Esta es la idea de *cointegración*.

Supongamos que las dos variables  $x_t$  e  $y_t$  que componen el vector  $\mathbf{z}_t = (x_t, y_t)$  son  $I(d)$ , pero que existe una constante  $a$  tal que la diferencia  $x_t - ay_t$  es  $I(d - b)$  para un  $b > 0$ . En tal caso, se dice que  $x_t$  e  $y_t$  están cointegradas de orden  $(d, b)$ ; por ejemplo, si  $x_t$  e  $y_t$  son  $I(1)$ , pero  $x_t - ay_t$  es estacionaria para una determinada constante  $a$ , entonces  $x_t$  e  $y_t$  son *cointegradas* de orden  $(1, 1)$ , lo que denotamos  $CI(1, 1)$ . El vector  $\alpha = (1, -a)$  se denomina *vector de cointegración*, y el producto  $\alpha' \mathbf{z}_t$  es estacionario. Es claro que el vector de cointegración está definido de forma única excepto por un factor, puesto que  $\alpha^* = (\lambda, -a\lambda)$  es asimismo un vector de cointegración entre  $x_t$  e  $y_t$ . Más generalmente,  $\mathbf{z}_t$  puede ser un vector de dimensión  $k$ :  $\mathbf{z}_t = (x_{1t}, x_{2t}, \dots, x_{kt})$ , en cuyo caso puede existir más de un vector de cointegración; el número máximo de tales vectores es  $k - 1$ .

En definitiva, si las componentes del vector  $\mathbf{z}_t$  están ligadas entre sí por una condición lineal de equilibrio estable a largo plazo, existiría un vector  $\alpha$  tal que  $\alpha' \mathbf{z}_t = 0$ . Período a período, pueden existir desviaciones respecto del equilibrio a largo plazo, por lo que se tendrá  $\alpha' \mathbf{z}_t \neq 0$ , pero a través del tiempo, el comportamiento de las variables en  $\mathbf{z}_t$  será tal que la relación  $\alpha' \mathbf{z}_t$  tienda nuevamente hacia cero. Por ejemplo, si la propensión marginal a consumir en una economía es de 0,85 y si la Renta Disponible  $Y_t^d$  es el único determinante del gasto en Consumo  $C_t$ , se tendría que  $C_t = 0,85 Y_t^d$  para

todo  $t$ , siendo la diferencia entre ambos miembros estacionaria. En tales condiciones, ambas variables, que quizá fuesen  $I(1)$ , estarían cointegradas.

La importancia estadística del concepto de cointegración estriba en que cuando las variables no estacionarias que aparecen a ambos lados de un modelo de regresión están cointegradas, entonces la estimación de MCO continúa teniendo buenas propiedades. Si, por ejemplo, las variables  $x_t$  e  $y_t$  son  $CI(1, 1)$ , entonces al estimar por MCO el modelo

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, 2, \dots, T$$

la estimación de  $\beta$  coincidirá con la constante de cointegración y los residuos serán estacionarios. El estimador MCO es consistente a pesar de ser las dos variables no estacionarias; más aún, el estimador MCO resulta ser *superconsistente*, puesto que converge al verdadero valor de los coeficientes  $\alpha$  y  $\beta$  a una tasa  $1/T$ , en vez de a la tasa habitual<sup>(2)</sup>,  $1/\sqrt{T}$ .

La superconsistencia del estimador MCO en la regresión estática se debe a que, para valores del coeficiente  $\beta$  distintos de la constante de cointegración  $a$ , el residuo no es estacionario, y su varianza tiende a infinito mientras que, para  $\beta = a$ , su varianza es finita. Como el estimador MCO «busca» en el espacio paramétrico tratando de minimizar la varianza residual, no es sorprendente que escoja el valor  $a$  para el coeficiente  $\beta$  con más facilidad que si las variables no estuviesen cointegradas.

El concepto de cointegración responde a la idea de que un modelo que pretende explicar una variable  $y_t$ , que sea  $I(1)$  utilizando variables explicativas  $x_t$ , todas  $I(0)$ , es decir, estacionarias, no puede dar un buen resultado estadístico, como tampoco podría darlo un modelo que pretendiera explicar una variable  $I(2)$  mediante un conjunto de variables explicativas que fuesen  $I(1)$  o  $I(0)$ . Por el contrario, puede aspirar a obtener un buen modelo para explicar una variable  $y_t$  que fuese  $I(1)$  utilizando como explicativas  $x_t$  y  $z_t$ , ambas  $I(2)$ , siempre y cuando éstas fuesen  $CI(2, 1)$  y su combinación lineal de cointegración  $x_t - az_t$  estuviese, a su vez, cointegrada con  $y_t$ .

### 14.7.a. Cointegración y el modelo de corrección de error (MCE)

Supongamos que dos variables  $x_t$  e  $y_t$  son  $C(1, 1)$ . Un *modelo general de corrección de error* entre ambas es del tipo:

$$\begin{aligned} \Delta y_t &= -\gamma_1 w_{t-1} + A(L)\Delta y_{t-1} + B(L)\Delta x_{t-1} + u_{1t} \\ \Delta x_t &= -\gamma_2 w_{t-1} + C(L)\Delta y_{t-1} + D(L)\Delta x_{t-1} + u_{2t} \end{aligned} \quad [14.11]$$

<sup>(2)</sup> Ello quiere decir que la diferencia  $\hat{\beta} - \beta$  converge asintóticamente a una Normal, incluso cuando se le multiplica por  $T$ , lo que no aceptan otros resultados sobre distribuciones límite. Por ejemplo, en el contexto habitual de regresión con variables estacionarias se tiene que  $\sqrt{T}(\hat{\beta} - \beta)$  converge a una Normal, por lo que  $T(\hat{\beta} - \beta)$  no tiene una distribución asintótica bien definida.

donde  $w_t = y_t - ax_t$  representa la combinación lineal de cointegración entre  $x_t$  e  $y_t$  y donde  $u_{it}$ ,  $i = 1, 2$  son términos de error estacionarios, aunque pueden tener autocorrelación.

En este modelo puede observarse que, por estar cointegradas  $y_t$  y  $x_t$ , entonces tanto  $\Delta y_t$  como  $\Delta x_t$  e  $y_{t-1} - ax_{t-1}$  son  $I(0)$ , y como también lo son las diferencias que aparecen en [14.11], se tiene que las propiedades asintóticas habituales del estimador MCO son válidas para los coeficientes, si bien no los son para la constante de cointegración  $a$ . En particular, si  $u_t$  está libre de autocorrelación, entonces el estimador MCO es consistente. También es Normal en muestras finitas si  $u_t$  también lo es.

En general, se dice que un vector de variables  $z_t$  admite una *representación de corrección de error (MCE)* si puede expresarse como:

$$A(L)\Delta z_t = -\gamma'w_{t-1} + u_t, \quad \gamma \neq 0$$

siendo  $u_t$  un *vector de perturbaciones estacionario*,  $A(0) = I_k$ , teniendo la ecuación característica  $|A(L)| = 0$  todas sus raíces fuera del círculo unidad. En esta representación,  $w_t$  es un vector de relaciones de cointegración entre las variables de  $z_t$ :  $w_t = \alpha'z_t$ . En el modelo [14.11],  $z_t = (x_t, y_t)$ .

Enlazando con nuestros comentarios al comienzo del capítulo,  $w_{t-1}$  representa el margen por el que las condiciones de equilibrio estable entre las variables que configuran el vector  $z_t$  dejaron de cumplirse en el período  $t - 1$  y, aparte de su propio pasado, estas desviaciones son las únicas variables explicativas de la variación que experimenta el vector  $z_t$  en el instante  $t$ . El ajuste hacia el equilibrio será más o menos gradual, en función del número de retardos que aparezcan en [14.11], así como de sus coeficientes.

Si existe entre  $y_t$  y  $x_t$  una relación del tipo [14.11], entonces ambas variables están cointegradas con constante  $a$ , ya que todas las variables que aparecen en [14.11] aparte de dicha diferencia son estacionarias. El teorema siguiente prueba el recíproco: si su conjunto de variables está cointegrado, existe una representación suya en la forma de modelo de corrección de error.

**Teorema de representación de Granger.** Si existen  $r$  relaciones de cointegración entre las variables que componen el vector  $z_t$  con  $d = 1$ ,  $b = 1$ , entonces:

a) Existe una *representación MCE* en función del vector  $w_t = \alpha'z_t$ , de dimensión  $r \times 1$ :

$$A^*(L)\Delta z_t = -\gamma'w_{t-1} + d(L)\varepsilon_t, \quad A^*(0) = I_k \quad [14.12]$$

donde las filas de  $\alpha$ , de dimensión  $r \times k$ , contienen los coeficientes de cada relación de cointegración.

b) Además, existe una representación ARMA vectorial:

$$A(L)z_t = d(L)\varepsilon_t \quad [14.13]$$

siendo  $d(L)$  el polinomio (escalar) de [14.12], rango  $A(1) = r$  y  $A(0) = I_k$ . Si el polinomio  $d(L) = 1$ , entonces el vector  $z_t$  acepta una representación vectorial autorregresiva:  $A(L)z_t = \varepsilon_t$ , pero la matriz  $A(1)$  es singular, pues es de rango igual al número de vectores de cointegración.

Un sistema bivalente cointegrado admite la representación:

$$\begin{aligned}\Delta x_t &= b(x_{t-1} - ay_{t-1}) + \text{retardos}(\Delta x_t, \Delta y_t) + u_t \\ \Delta y_t &= c(x_{t-1} - ay_{t-1}) + \text{retardos}(\Delta x_t, \Delta y_t) + v_t\end{aligned}$$

donde al menos uno de los coeficientes  $b$  o  $c$  es no nulo, y  $u_t$  es estacionario, aunque puede estar autocorrelacionado. En consecuencia, tal sistema debe tener una ordenación causal al menos en una dirección, puesto que  $w_t$  es una función de ambas variables. Sin embargo, los retardos de  $\Delta x_t$  e  $\Delta y_t$  pueden aparecer o no en las ecuaciones anteriores, por lo que puede tenerse todo tipo de estructuras de causalidad. En particular, si el término de corrección de error aparece en ambas ecuaciones, entonces ninguna variable es débilmente exógena con respecto a los parámetros de la otra ecuación, debido a las restricciones entre los coeficientes de las distintas ecuaciones.

Si dos variables están cointegradas  $CI(1, 1)$ , entonces existe un modelo del tipo anterior, pero no sabemos la dirección de causalidad. Esto es importante, porque se tendrán distintas estimaciones de los coeficientes del vector de cointegración según qué dirección de la relación se estime. Por ejemplo, si  $x_t$  e  $y_t$  son  $CI(1, 1)$ , entonces el producto de las estimaciones de  $a$  y  $1/a$  en las regresiones de  $y_t$  sobre  $x_t$  y de ésta sobre la primera no es igual a 1 sino a  $R^2$ .

### 14.7.b. Estimación de modelos de corrección de error

Engle y Granger (1987) sugieren una estimación en dos etapas: inicialmente, se estima el vector de cointegración a partir de una regresión entre los valores contemporáneos de las variables que componen el vector  $z_t$ . El residuo resultante es estacionario, de modo que el estimador MCO de dicho modelo es consistente. Condicionando en tal estimación del vector de cointegración  $\alpha$ , se construye el vector de residuos  $w_t$  y se incorpora al modelo MCE, que pasa a estimarse por MCO, una vez que se han incorporado al mismo suficientes retardos de  $\Delta x_t$  y  $\Delta y_t$  como para que el término de error sea estacionario<sup>(3)</sup>.

La desviación típica del vector de cointegración en la primera etapa es inconsistente, porque la distribución del estimador MCO *no es Normal*. Ello implica que hay que escoger las variables a incluir en la regresión de largo

<sup>(3)</sup> Debe prestarse atención al hecho de que otros procedimientos alternativos de estimar  $\alpha$  no producen estimaciones consistentes: por ejemplo, una regresión de  $\Delta y_t$  sobre  $\Delta x_t$  no genera estimaciones consistentes. Un resultado similar se obtiene si, para evitar la casi inevitable autocorrelación en la regresión de cointegración, se utilizan transformaciones del tipo Cochrane-Orcutt.

plazo a priori, puesto que no se puede decidir su inclusión sobre la base de los contrastes habituales de significación. Por supuesto, uno de los criterios fundamentales para la inclusión de variables en dicha regresión es que estén cointegradas. Banerjee y colaboradores (1986) han probado que el sesgo en muestras finitas en el estimador MCO de los coeficientes de largo plazo, es decir, el vector de cointegración, puede ser apreciable y depende: a) del tamaño relativo de las varianzas de las ecuaciones de largo y corto plazo, y b) está inversamente relacionado con el  $R^2$  de la regresión estática de cointegración.

Alternativamente, consideremos un modelo específico de corrección de error:

$$\Delta y_t = \beta_1 + \beta_2 \Delta x_{t-1} + \beta_3 \Delta y_{t-1} + \beta_4 (y_{t-1} - a x_{t-1}) + u_t \quad [14.14]$$

y veamos cómo puede estimarse simultáneamente el coeficiente de cointegración  $\beta$  y el vector  $\beta = (\beta_1, \beta_2, \beta_4)$  del MCE. A primera vista, parece que éste debiera ser un problema de estimación no lineal sujeto a restricciones, pero si hacemos el cambio de parámetros:

$$\gamma_1 = \beta_1 \quad ; \quad \gamma_2 = \beta_2 \quad ; \quad \gamma_3 = -a\beta_4 \quad ; \quad \gamma_4 = \beta_4$$

basta estimar por MCO una regresión sin restricciones de  $\Delta y_t$  sobre: una constante,  $\Delta x_t$ ,  $\Delta x_{t-1}$  e  $\Delta y_{t-1}$ , para obtener estimadores de los parámetros  $\gamma_i$ ,  $i = 1, 2, 3, 4$ . La respuesta a largo plazo de  $y_t$  a  $x_t$  queda estimada mediante  $a = -\hat{\gamma}_3/\hat{\gamma}_4$ , aunque no tendremos una desviación típica para este parámetro. La estimación por MCNL nos habría proporcionado la desviación típica de  $a$ . Lo que ocurre es que, en presencia de no estacionariedad, la distribución de este estimador no es la habitual, por lo que no hubiéramos podido utilizar dicha desviación típica estimada.

### 14.7.c. Contrastes de cointegración

Un contraste de cointegración puede entenderse como equivalente a un contraste de raíz unitaria en los residuos de la ecuación de cointegración, ya que si las variables  $y_t$  y  $x_t$ , siendo ambas  $I(1)$ , estuviesen cointegradas, entonces los residuos del modelo  $y_t = \alpha + x_t' \beta + u_t$  deberían ser estacionarios, no teniendo por tanto ninguna raíz unitaria.

Veamos dos contrastes propuestos por Engle y Granger (1987):

1. El estadístico DW de la ecuación de cointegración (Sargan y Barghava, 1983) conduce a rechazar la hipótesis nula de no cointegración, es decir, la existencia de raíz unitaria en los residuos, si el estadístico DW es significativamente mayor que cero. Este contraste, que parece bastante potente, se basa en que  $DW = 2(1 - \rho)$ , y los valores críticos para un tamaño muestral  $T = 100$  son los que se muestran en la Tabla 14.6.

**TABLA 14.6.** Contraste de cointegración. Estadístico Durbin-Watson de la regresión de largo plazo o de cointegración

Nivel de significación	0,01	0,05	0,10
Valor crítico	0,511	0,386	0,322

Una dificultad con este contraste es que dichos valores críticos son función de la especificación adoptada para el modelo verdadero. El contraste se realiza mediante la existencia de cotas superior e inferior —como en un contraste DW de autocorrelación—, pero dichas cotas se separan entre sí mucho en cuanto aumenta el número de variables explicativas.

Una gran ventaja de este contraste es su invariancia frente a la posible inclusión de constantes y tendencias en el modelo, sin que por ello varíen sus valores críticos.

2. El segundo contraste, del tipo DF o DFA sobre la regresión de cointegración, consiste en estimar por MCO:  $\Delta \hat{u}_t = \gamma_1 \hat{u}_{t-1} + \sum_1^p \gamma_{2i} \Delta \hat{u}_{t-i} + \varepsilon_t$ , donde  $\hat{u}_t$  son los residuos MCO de la ecuación de cointegración, y el número de retardos  $p$  se escoge suficientemente grande como para que  $\varepsilon_t$  sea ruido blanco.

Los valores críticos con los que comparar el estadístico  $t$  del coeficiente  $\gamma_1$  aparecen en Engle y Granger, y son diferentes de los de Fuller (1976). La razón para esta disparidad es que  $\hat{u}_t$  es una variable «generada» mediante un procedimiento que ha escogido  $\hat{\alpha}$  y  $\hat{\beta}$  de modo que se minimice la varianza residual, es decir, la varianza de  $\hat{u}_t$  y, por tanto, de modo que parezca tan estacionario como sea posible. En consecuencia, si utilizamos los valores críticos de los contrastes DF o DFA para este caso, rechazaríamos la hipótesis nula de no estacionariedad con demasiada frecuencia. Para evitar este sesgo, los valores críticos deben aumentarse ligeramente.

Otro factor que influye sobre los valores críticos apropiados es el número de variables explicativas, pues una combinación lineal de una larga lista de variables explicativas no estacionarias puede parecer estacionaria, incluso si las variables no están cointegradas. Engle y Granger (1987) presentaron los valores críticos correctos *para una sola variable explicativa*, lo que fue generalizado en Engle y Yoo (1987) hasta cinco variables explicativas.

**TABLA 14.7.** Contrastes de cointegración Dickey-Fuller y Dickey-Fuller ampliado. Valores críticos.  $T = 100$  observaciones

	0,01			0,05			0,10		
	$N = 2$	$N = 3$	$N = 4$	$N = 2$	$N = 3$	$N = 4$	$N = 2$	$N = 3$	$N = 4$
DF	-4,07	-4,45	-4,75	-3,37	-3,93	-4,22	-3,03	-3,59	-3,89
DFA	-3,77	-4,22	-4,61	-3,17	-3,62	-4,02	-2,84	-3,32	-3,71

Ambos contrastes están diseñados para el caso en que tanto la hipótesis nula como la alternativa consideran modelos de primer orden, y presentan problemas para valores próximos a 1 del coeficiente  $\rho$  del modelo AR(1) de los residuos, es decir, para valores muy bajos del estadístico DW. A tales niveles, la potencia, tanto del contraste basado en el estadístico DW como de tipo DFA, es muy baja.

3. Otros contrastes se basan en la propiedad de que las variables cointegradas aceptan una representación en la forma de MCE y, por ello, como ya hemos visto, al menos una variable, digamos  $x_t$ , *causa* a la otra  $y_t$ , en el sentido de que sus retardos  $x_{t-s}$ ,  $s > 1$ , contienen información relevante acerca de  $y_t$  que no está contenida en el pasado de esta última variable. Es decir, los retardos  $x_{t-s}$  deben aparecer en la regresión de  $y_t$ , incluso si se incluyen retardos  $y_{t-s}$ . Por consiguiente, podemos especificar:

$$\begin{aligned} \Delta x_t &= a_1 + b_1 \hat{u}_{t-1} + \varepsilon_{1t} \\ \Delta y_t &= a_2 + b_2 \hat{u}_{t-1} + c_2 \Delta x_t + \varepsilon_{2t} \end{aligned}$$

donde  $\hat{u}_t$  son los residuos mínimo-cuadráticos de la regresión de cointegración de  $y_t$  sobre  $x_t$ .

El modelo de corrección de error implica que  $b_1$  y  $b_2$  deben ser conjuntamente significativos. El estadístico para el contraste de tal hipótesis es la suma de cuadrados de los respectivos estadísticos «tipo  $t$ »,  $\tau_{b_1}^2 + \tau_{b_2}^2$ . Si se rechaza la hipótesis nula  $H_0: b_1 = b_2 = 0$ , entonces  $x_t$  e  $y_t$  satisfacen un mecanismo de corrección de error y están cointegrados. Este es el contraste conocido como *VAR restringido* (RVAR). Un *test ampliado* (ARVAR) de esta hipótesis consiste en incluir retardos de  $\Delta x_t$  y  $\Delta y_t$  como variables explicativas de ambas ecuaciones.

**TABLA 14.8.** Contrastes de cointegración. Valores críticos

N = 2			
	0,01	0,05	0,10
RVAR	18,3	13,6	11,0
ARVAR	15,8	11,8	9,7
UVAR	23,4	18,6	16,0
AUVAR	22,6	17,9	15,5

Alternativamente, puede utilizarse un contraste de tipo *VAR sin restricciones* (UVAR), mediante la estimación del sistema:

$$\begin{aligned} \Delta x_t &= a_1 + b_1 y_{t-1} + c_1 x_{t-1} + \varepsilon_{1t} \\ \Delta y_t &= a_2 + b_2 y_{t-1} + c_2 x_{t-1} + d_2 \Delta x_t + \varepsilon_{2t} \end{aligned}$$

y contrastar la hipótesis nula conjunta  $H_0: b_1 = c_1 = b_2 = c_2 = 0$  mediante el estadístico  $2(F_1 + F_2)$ , siendo  $F_1$  el estadístico  $F$  para el contraste de la hipótesis conjunta  $H_0: b_1 = c_1 = 0$  y  $F_2$  el estadístico  $F$  para el contraste de  $H_0: b_2 = c_2 = 0$ . Si se rechaza dicha hipótesis,  $\Delta x_t$  e  $\Delta y_t$  dependen de sus niveles y, potencialmente, siguen un esquema MCE. Por último, puede construirse un *test ampliado* (AUVAR) incluyendo retardos adicionales de  $\Delta x_t$  e  $\Delta y_t$  como variables explicativas. Para este contraste ampliado se utiliza el mismo estadístico:  $2(F_1 + F_2)$  anterior.

#### 14.7.d. Contrastes de cointegración estacional

Los polinomios en el operador de retardos que utilizamos en la Sección 14.6 para la contrastación de la hipótesis de raíz unitaria bajo estacionalidad, pueden utilizarse asimismo para contrastar la cointegración de variables. Así, con datos trimestrales: a) existe *cointegración en el ciclo de un periodo* si existe un vector  $\alpha_1$  tal que la combinación lineal  $u_t = \alpha_1' Z_1 z_t$  es estacionaria, b) existe *cointegración en el ciclo de dos periodos* si existe un vector  $\alpha_2$  tal que la combinación lineal  $v_t = \alpha_2' Z_2 z_t$  es estacionaria, y c) existe *cointegración en el ciclo anual* si existen dos vectores  $\alpha_3$  y  $\alpha_4$  tal que la combinación lineal  $w_t = (\alpha_3' + \alpha_4' L) Z_3 z_t$  es estacionaria.

Existe una versión estacional del modelo de corrección de error, que permite la incorporación al mismo de las combinaciones lineales de cointegración que puedan existir en las diversas raíces unitarias que ya hemos examinado. Teniendo en cuenta que todas las variables incluidas en el modelo deben ser  $I(0)$ , la forma general de dicho modelo es:

$$\begin{aligned} \Delta\Delta_4 y_t = & \sum_{i=1}^m \beta_i \Delta\Delta_4 y_{t-i} + \sum_{i=1}^n \delta_i \Delta\Delta_4 x_{t-i} + \\ & + \gamma_1 (Z_1 y_{t-1} - \alpha_{12} Z_1 x_{t-1}) + \gamma_2 (Z_2 y_{t-1} - \alpha_{22} Z_2 x_{t-1}) + \\ & + (\gamma_3 + \gamma_4 L) (Z_3 y_{t-1} - \alpha_{32} Z_3 x_{t-1} - \alpha_{41} Z_3 y_{t-2} - \alpha_{42} Z_3 x_{t-2}) + u_t \end{aligned} \quad [14.15]$$

donde se han incorporado suficientes retardos de  $\Delta\Delta_4 y_t$  y  $\Delta\Delta_4 x_t$  como para que  $u_{1t}$  sea ruido blanco. Permitimos que los vectores de cointegración de cada raíz unitaria (los  $\alpha_{ij}$ ) sean diferentes, así como también sus efectos sobre la corrección que experimenta en el periodo  $t$  la variable dependiente (los  $\gamma_i$ ).

La estimación en dos etapas de este modelo de corrección de error estacional consiste en estimar los coeficientes  $\alpha_{ij}$ ,  $i = 1, 2, 3, 4$ ,  $j = 1, 2$  en cada regresión de cointegración, para luego incorporarlos a [14.15] y estimar por MCO. El número de retardos que se habrá incorporado a cada suma ha de ser suficiente como para eliminar la autocorrelación del término de error.

Hay dos hipótesis de interés que podrían imponerse al estimar el MCE estacional [14.15]:

1. Que se produce cointegración en todos los ciclos con igual parámetro:  $\alpha_{12} = \alpha_{22} = \alpha_{32} = \alpha$  y  $\alpha_{41} = \alpha_{42} = 0$ , con lo que el MCE se reduce al expresado en las variables originales:

$$\Delta_4 y_t = \sum_{i=1}^m \beta_i \Delta_4 y_{t-i} + \sum_{i=1}^n \delta_i \Delta_4 x_{t-i} + \sum_{i=1}^3 \gamma_i Z_i (y_{t-1} - \alpha x_{t-1}) + \gamma_4 Z_3 (y_{t-2} - \alpha x_{t-2}) + u_t$$

2. Que existe cointegración en el ciclo de un período en las variables ajustadas estacionalmente, y en los demás ciclos con el mismo parámetro:  $\alpha_{12} = \alpha$ ;  $\alpha_{22} = \alpha_{32} = \alpha_s$ ;  $\alpha_{41} = \alpha_{42} = 0$ .

## 14.8. APLICACIONES DEL CONCEPTO DE COINTEGRACION

### 14.8.a. Una síntesis

Antes de examinar algunas aplicaciones de la teoría que hemos desarrollado en este capítulo, resumamos algunas ideas esenciales en relación a la importancia de la cointegración en la especificación y estimación de modelos de regresión:

1. No pueden mezclarse variables de distintos órdenes de integración en una regresión estática y esperar que los resultados estadísticos estén justificados rigurosamente.

2. Si se pretende generalizar una regresión estática incluyendo retardos para obtener su versión dinámica, es *esencial* incorporar retardos de todas las variables que entran en la regresión estática o, al menos, de todas las variables que parezcan relevantes en dicha regresión. Esto es lo que, de hecho, sugiere un MCE, por lo que no puede resultar sorprendente la estrecha conexión que existe entre MCE y cointegración.

3. Cuando se especifica una regresión dinámica, la propia riqueza de la estructura dinámica puede hacer que los residuos aparenten ser estacionarios en una muestra finita cuando en realidad las variables que aparecen en el modelo no estén cointegradas, en cuyo caso los procesos de inferencia habitual no serían válidos.

4. En la especificación de una relación dinámica entre variables económicas puede adoptarse una estrategia de seleccionar variables cointegradas, especificar su relación de largo plazo o regresión de cointegración, y estimar ésta, así como su modelo de corrección de error. Alternativamente, puede procederse especificando un modelo dinámico muy general en el que ir imponiendo restricciones para llegar a un modelo dinámico sencillo.

Teóricamente, en muestras infinitas, ambos enfoques deberían producir el mismo resultado. Sin embargo, en muestras finitas esto no es así; además de errores muestrales, al estimar un modelo dinámico general, puede que se incluyan variables de diverso orden de integración, con las dificultades que hemos expuesto en secciones anteriores, tanto en términos de cuál sea la distribución correcta para las estimaciones numéricas obtenidas y sus desviaciones típicas, así como para la estimación del efecto a largo plazo de las variables explicativas sobre la variable dependiente.

### 14.8.b. La eficiencia de un mercado financiero

Hay varias formas en que la cointegración de variables puede utilizarse para contrastar *la eficiencia de los mercados financieros*: la noción de eficiencia de mercado, fundamentada en la idea de que el precio actual resume toda la información disponible acerca del precio futuro, equivale a decir que la proyección del precio futuro  $p_{t+1}$  sobre el conjunto de información hoy disponible  $\Omega_t$  es simplemente igual a  $p_t$ , sin que ninguna otra información resulte relevante, es decir:

$$p_{t+1} = p_t + u_{t+1}$$

con  $E_t u_{t+1} = 0$ , es decir, que  $u_t$  es ruido blanco y  $p_t$  un paseo aleatorio.

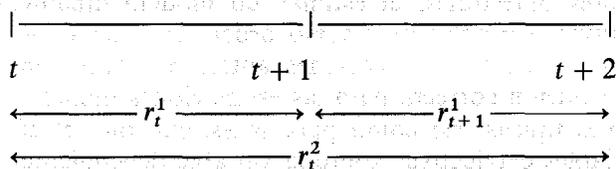
#### a) Cointegración entre precios de activos

El modelo de corrección de error implica que el cambio en  $y_t$ ,  $\Delta y_t$ , se debe a un cambio contemporáneo en otra variable  $x_t$ ,  $\Delta x_t$ , junto con el efecto a largo plazo que proviene del desajuste producido en el período anterior con respecto al equilibrio a largo plazo,  $y_{t-1} - ax_{t-1}$ . Por tanto, aun dejando aparte  $\Delta y_t$ , que era desconocido, el modelo sugiere que existía información conocida con anterioridad (en los valores de  $x_{t-1}$  e  $y_{t-1}$ ) que podía ayudar a prever el cambio  $\Delta x_t$ .

Pero si  $x_t \equiv p_t$  representa el precio de un activo cuyo mercado se supone eficiente, tal situación no es posible, puesto que la variación en su precio,  $\Delta p_t$ , debe ser impredecible a partir del conjunto de información  $\Omega_{t-1}$ ; ello es debido a que el precio pasado,  $p_{t-1}$ , incorpora toda la información relevante acerca del precio futuro, desconocido,  $p_{t+s}$ ,  $s \geq 0$ . En definitiva, todo ello nos sirve para afirmar que si  $x_t$  e  $y_t$  denotan los precios determinados en dos mercados financieros eficientes, por ejemplo, dos tipos de cambio, o dos tipos de interés forward, *no pueden estar cointegrados*.

#### b) El tipo forward como predictor del tipo spot futuro

Por otra parte, en muchos mercados financieros se intercambia un activo emitido a distintos plazos, por lo que podemos deducir tipos forward implícitos. En tal situación, la idea de eficiencia se traduce en el supuesto de que  $r_t^1$  y  $r_t^2$ , los rendimientos actuales de mercado a 1 y 2 períodos de horizonte de madurez, resumen toda la información acerca de  $r_{t+1}^1$ , el tipo a un periodo que resultará vigente el próximo período:



Los tipos  $r_t^1$  y  $r_t^2$  determinan el tipo forward implícito  $F_t^1$  definido como  $F_t^1 = (2r_t^2 - r_t^1)$ , siendo entonces la hipótesis de eficiencia la propiedad de que la proyección del tipo «spot» futuro  $r_{t+1}^1$  sobre  $\Omega_t$  sea igual a  $F_t^1$ , es decir:

$$r_{t+1}^1 = F_t^1 + u_{t+1} = (2r_t^2 - r_t^1) + u_{t+1}$$

con  $u_t$  ruido blanco o, lo que es lo mismo:  $E_t r_{t+1}^1 = F_t^1 = 2r_t^2 - r_t^1$ .

Por esta razón, un buen número de trabajos se han dedicado a la contrastación de la hipótesis conjunta  $H_0: a = 0, b = 1$  en el modelo:

$$r_{t+1}^1 = a + bF_t^1 + u_{t+1} \quad [14.16]$$

lo que se conoce en la literatura como que *el tipo forward es un predictor insesgado del tipo spot futuro*. La eficiencia del mercado precisa de una condición adicional: que el término de error  $u_t$  del modelo sea **ruido blanco**, pues de otro modo, la proyección de  $r_{t+1}^1$  sobre  $\Omega_t$  incluiría, además de  $F_t^1$ , valores retardados de ambas variables.

Desde el punto de vista de la cointegración de variables, afirmar que el tipo forward es un predictor insesgado del tipo spot futuro o que el modelo anterior es válido implica no sólo que ambas variables están cointegradas, sino también que su constante de cointegración es igual a 1 y que el término de error  $u_t$  es ruido blanco. Todo ello, conjuntamente, puede interpretarse como un contraste de eficiencia del mercado.

Respecto a la puesta en práctica de este contraste, cabe hacer dos observaciones:

1. La observación de la no estacionariedad de los tipos spot y forward ha llevado a especificar modelos del tipo:

$$r_{t+1}^1 - r_t^1 = a + b(F_t^1 - r_t^1) + u_{t+1} \quad [14.17]$$

en la confianza de que ambas variables transformadas resultasen estacionarias.

2. El contraste puede formularse de otro modo alternativo: si  $r_{t-1}^1$  y  $F_t^1$  están cointegradas, se tiene la representación en modelo de corrección de error:

$$r_{t+1}^1 - r_t^1 = A(F_t^1 - F_{t-1}^1) + B(r_t^1 - cF_{t-1}^1) + \varepsilon_{t+1} \quad [14.18]$$

donde  $c$  es la constante de cointegración. Sólo si  $A = 1$  y  $B = -1$  podrían compatibilizarse [14.17] y [14.18], supuesto que  $a = 0$  y  $b = 1$ . Para ver este punto de otro modo, escribamos [14.18] como:

$$(1 - (1 + B)L)r_{t+1}^1 = -(A + Bc)F_t^1 + AF_{t-1}^1 + \varepsilon_{t+1} \quad [14.19]$$

que conduce a:

$$r_{t+1}^1 = AF_t^1 + B(A - c)\sum_{s=0}^{\infty} (1 + B)^s F_{t-1-s}^1 + \sum_{s=0}^{\infty} (1 + B)^s \varepsilon_{t-s+1} \quad [14.20]$$

supuesto que  $B < 0$ , pero esta expresión sólo podría ser compatible con [14.16] si  $B = -1$  y  $A = c = 1$ . De este modo, los retardos de  $F_t^1$  en la regresión  $r_{t+1}^1$  sobre  $F_t^1$  tendrían coeficientes igual a cero.

### 14.8.c. La cointegración del Consumo y el PIB españoles

Consideremos nuevamente los datos de Consumo y PIB españoles para el periodo 1954-1988, en pesetas constantes de 1980. Tras tomar logaritmos en ambas variables, estimamos modelos de regresión que nos permitieran contrastar la existencia de raíces unitarias en ambas variables. La Tabla 14.9 contiene los estadísticos  $t$  correspondientes al contraste de la hipótesis  $H_0: I(2)$  frente a  $I(1)$ , mientras que en la Tabla 14.10 se recogen los estadísticos  $t$  para el contraste de  $H_0: I(1)$  frente a  $I(0)$ . En cada caso, se presentan las estimaciones de los modelos [14.2] (primera y segunda filas de cada tabla) y [14.3] (tercera y cuarta filas); no se presenta el modelo [14.4] porque la variable de tendencia no resultó significativa en ninguno de los casos en que se incluyó. En ambos contrastes, la constante resultó significativa, por lo que son las dos últimas filas las que deberían utilizarse<sup>(4)</sup>.

Aunque nuestro tamaño muestral ( $T = 35$ ) no coincide con los considerados en Fuller (1976), la comparación de los estadísticos  $t$  de la Tabla 14.9 con los valores críticos de la Tabla 14.1 permite mantener la hipótesis nula de dos raíces unitarias, frente a la alternativa de tan sólo una raíz unitaria. A la vista de estos resultados, el contraste de una raíz unitaria frente a estacionariedad ya es redundante, puesto que no hemos rechazado la existencia de dos raíces unitarias. Sin embargo, presentamos los resultados como ilustración al lector. Así, puede apreciarse en las Tablas 14.9 y 14.10 que, mientras el contraste  $I(1)$  frente a  $I(0)$  tiene una conclusión contundente, el contraste  $I(2)$  frente a  $I(1)$  es mucho más marginal, si bien puede mantenerse que ambas variables son  $I(2)$ .

**TABLA 14.9.** Contrastes de raíz unitaria  $I(2)$  frente a  $I(1)$

<i>Variable dependiente: <math>\Delta^2 y_t</math></i>		
<i>Variables independientes</i>	<i>Consumo</i>	<i>PIB</i>
$\Delta y_{t-1}$	-1,71	-1,57
$\Delta y_{t-1} + \Delta^2 y_{t-1}$	-1,29	-1,54
Constante + $\Delta y_{t-1}$	-3,34	-2,93
Constante + $\Delta y_{t-1} + \Delta^2 y_{t-1}$	-2,46	-2,92

<sup>(4)</sup> En el primer contraste, el retardo de la segunda diferencia no resultó significativo, por lo que en la Tabla 14.9 es la tercera fila la que debería tomarse, mientras que en el segundo contraste, la primera diferencia retardada resultó significativa, por lo que es la última fila de la Tabla 14.10 la que debería ser utilizada. Debe observarse, sin embargo, que ambos modelos no son totalmente consistentes entre sí: la presencia de una constante en el contraste de  $I(2)$  frente a  $I(1)$  debería conducir a una tendencia en el modelo utilizado en el contraste de  $I(1)$  frente a  $I(0)$ .

TABLA 14.10. Contrastes de raíz unitaria  $I(1)$  frente a  $I(0)$ 

Variable dependiente: $\Delta y_t$		
Variables independientes	Consumo	PIB
$y_{t-1}$	8,30	7,98
$y_{t-1} + \Delta y_{t-1}$	2,62	2,26
Constante + $y_{t-1}$	-2,13	-2,56
Constante + $y_{t-1} + \Delta y_{t-1}$	-1,41	-1,54

La regresión de cointegración entre Consumo y PIB condujo a:

$$\ln C_t = 0,0408 + 0,969 \ln Y_t + \hat{u}_t$$

(0,80) (0,01)

$$\bar{R}^2 = 0,997 \quad DW = 0,36$$

y el estadístico DF para los residuos<sup>(5)</sup> fue de  $-1,86$ , por lo que no se rechaza la hipótesis de que dichos residuos no son estacionarios. Sin embargo, al contrastar la presencia de dos raíces unitarias en los residuos frente a la alternativa de una sola raíz, se obtuvo un estadístico DF de  $-6,66$ , por lo que se rechaza dicha hipótesis nula. Los residuos tienen una sola raíz unitaria, y el Consumo y el PIB son variables  $CI(2, 1)$ .

Una evidencia similar se obtiene mediante la regresión en diferencias:

$$\Delta \ln C_t = 0,638 \cdot 10^{-2} + 0,818 \Delta \ln Y_t + \hat{u}_t$$

(0,41  $10^{-2}$ ) (0,08)

$$\bar{R}^2 = 0,762 \quad DW = 2,39$$

con un estadístico DF para los residuos (incluyendo una constante en dicho contraste) de  $-6,79$ , por lo que estos últimos residuos son estacionarios.

<sup>(5)</sup> Los retardos que se incluyeron al realizar un contraste DFA resultaron no significativos, por lo que nos limitamos al contraste DF.