

## Capítulo 6

# PROCESOS INTEGRADOS y DE MEMORIA LARGA

**Norbert Wiener** (1894-1964)

Matemático genial nacido en EEUU de padres judíos rusos. Se doctoró en Harvard a los 18 años y tras ampliar estudios en Cambridge con Russell y en Göttingen con Hilbert fue profesor de matemáticas en el MIT en Boston desde 1919 hasta su jubilación en 1959. Descubrió el movimiento browniano y desarrolló la teoría de predicción de procesos estacionarios y el análisis espectral. Inventor de la Cibernética, sus aportaciones han sido muy influyentes en muchos campos científicos.

### 6.1 Introducción

En este Capítulo comenzamos el estudio de los procesos no estacionarios. Un proceso puede ser no estacionario en la media, en la varianza, en las autocorrelaciones, o en otras características de la distribución de las variables. Cuando el nivel de la serie no es estable en el tiempo, pudiendo en particular tener tendencia creciente o decreciente, diremos que la serie es no estacionaria en la media. Cuando la variabilidad o las autocorrelaciones se modifica con el tiempo diremos que es no estacionario en la varianza o en las autocorrelaciones. Finalmente, si la distribución de la variable en cada instante varía con el tiempo, manteniendo o no la media y las covarianzas, diremos que la serie no es estacionaria en distribución. Los procesos no estacionarios más importantes son los procesos integrados, que son procesos no estacionarios únicamente en la media, y que tienen la propiedad de convertirse en estacionarios tomando diferencias. En este capítulo presentaremos con detalle estos procesos.

Una clase de procesos estacionarios que tienen características similares a algunos procesos integrados son los procesos de memoria larga, que revisaremos también brevemente en este capítulo.

### 6.2 Procesos Integrados

La mayoría de las series reales no son estacionarias, y su nivel medio varía con el tiempo. Por ejemplo, la figura 6.1 presenta una serie no estacionaria. En el capítulo 1 las series de las figuras 1.3 a 1.9 (exceptuando la 1.6) corresponden a series no estacionarias.

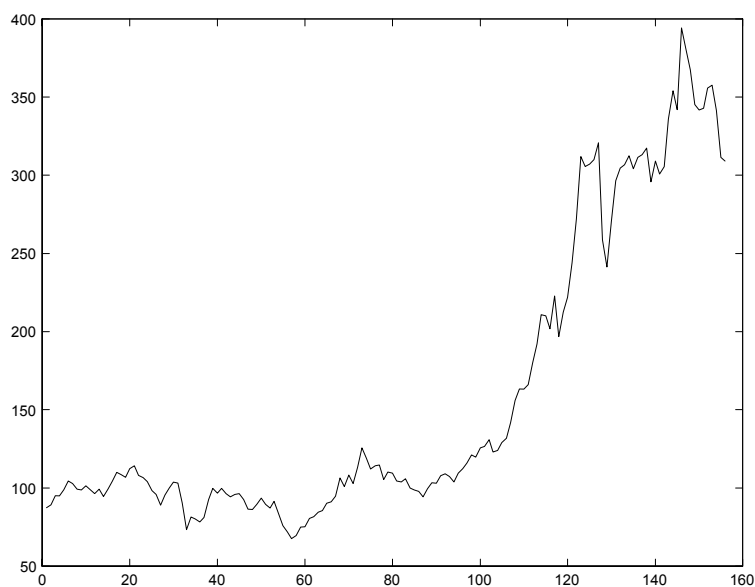


Figura 6.1: Índice general de la bolsa de Madrid, enero de 1988 a mayo de 2003.

La serie de la figura 6.1 no es estacionaria porque ni el nivel ni la variabilidad es constante a lo largo del tiempo. Un tipo de no estacionariedad frecuente en muchas series es que la variabilidad de la serie depende del nivel, y que la serie tenga variabilidad constante al expresarla en logaritmos. La figura 6.2 presenta esta serie en logaritmos y ahora la variabilidad es aproximadamente constante en el tiempo, con lo que se hemos eliminado la no estacionariedad en la varianza. Sin embargo, la serie no es estacionaria en la media ya que su nivel varía a lo largo del tiempo.



Figura 6.2: Logaritmo del índice general de la bolsa de Madrid, enero de 1988 a mayo de 2003.

Es frecuente que las series no estacionarias en la media aunque no tengan nivel constante si presentan

alguna característica estable en el tiempo: por ejemplo, su crecimiento puede variar alrededor de un valor medio constante. Entonces, la serie de crecimientos,  $\omega_t$ , que se obtiene mediante la operación

$$\omega_t = z_t - z_{t-1} = \nabla z_t, \quad (6.1)$$

será estacionaria. La serie  $\omega_t$  definida por la ecuación (6.1) se denomina la primera diferencia de  $z_t$ , y si  $\omega_t$  es estacionaria y  $z_t$  no lo es, diremos que el proceso  $z_t$  es integrado de orden uno, o también, con incrementos estacionarios. Por ejemplo, la figura 6.3 representa la primera diferencia de la serie de la figura 6.2 y ahora parece estacionaria. Para analizar que característica de esta serie es estable en el tiempo vamos a comprobar que la primera diferencia de la serie en logaritmos del índice general representa la serie de rendimientos obtenidos. En efecto, como:

$$\nabla \log z_t = \log \frac{z_t}{z_{t-1}} = \log\left(1 + \frac{z_t - z_{t-1}}{z_{t-1}}\right) \approx \frac{\nabla z_t}{z_{t-1}}$$

la primera diferencia del logaritmo son los crecimientos relativos, que en una serie de precios son los rendimientos. La figura 6.2 muestra que estos rendimientos mensuales son estables en el tiempo.

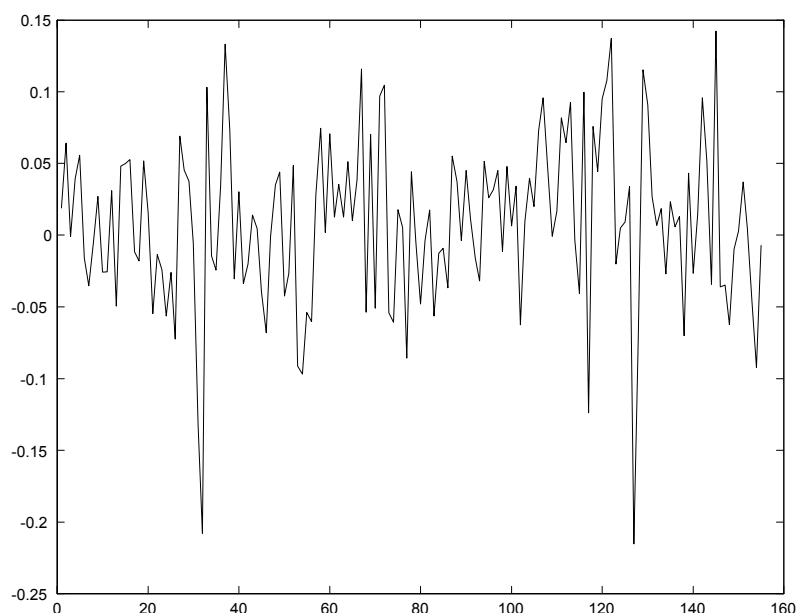


Figura 6.3: Rendimientos mensuales obtenidos en la bolsa de Madrid de acuerdo al índice general en el periodo 1988 a 2000.

Es posible que los crecimientos de la serie no sean estacionarios, pero si lo sean las variaciones del crecimiento. Es decir, la primera diferencia no conduce a una serie estacionaria pero la segunda sí. Entonces, la serie

$$y_t = \omega_t - \omega_{t-1} = z_t - 2z_{t-1} + z_{t-2} = \nabla^2 z_t \quad (6.2)$$

que representa los crecimientos de la series  $\omega_t$  será estacionaria. Esta serie  $y_t$  se denomina proceso «segunda diferencia» de la serie original,  $z_t$ , y diremos que  $z_t$  es integrada de orden dos, o con incrementos de segundo orden estacionarios. Por ejemplo la serie de la figura 6.4 no es estacionaria y su primera diferencia, indicada en la figura 6.5 tampoco lo es. Sin embargo, la segunda diferencia de esta serie, que se muestra en la figura 6.6, si tiene un nivel estable. Diremos que esta serie es integrada de orden dos.

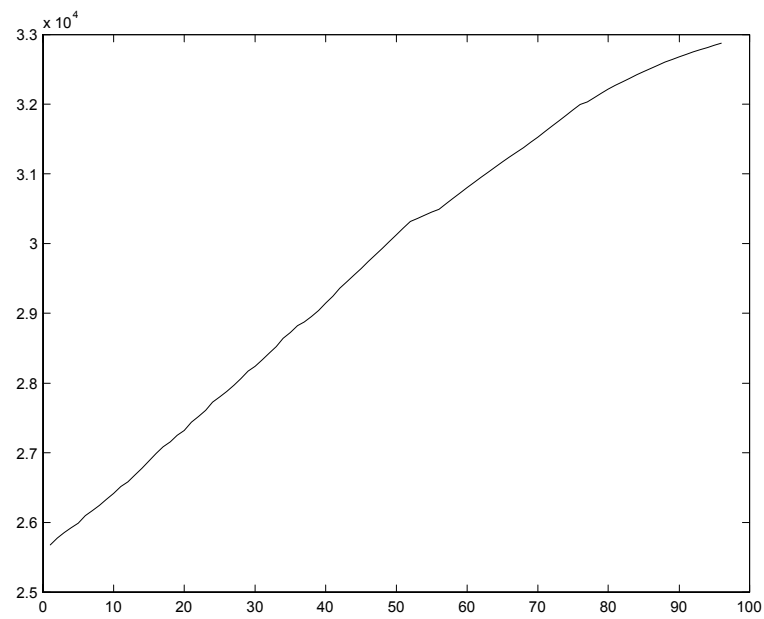


Figura 6.4: Población mayor de 16 años en España entre el primer cuatrimestre de 1977 al cuarto de 2000

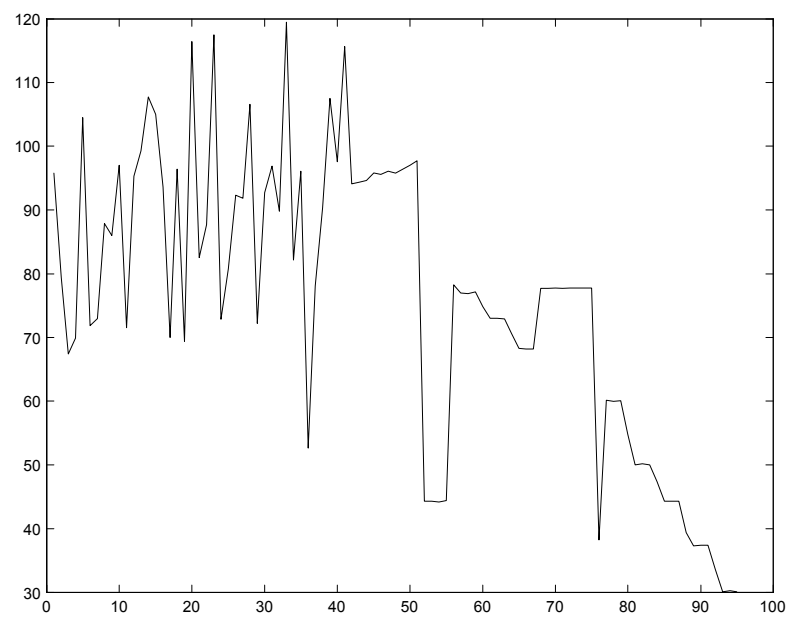


Figura 6.5: Primera diferencia de la serie de población mayor de 16 años.

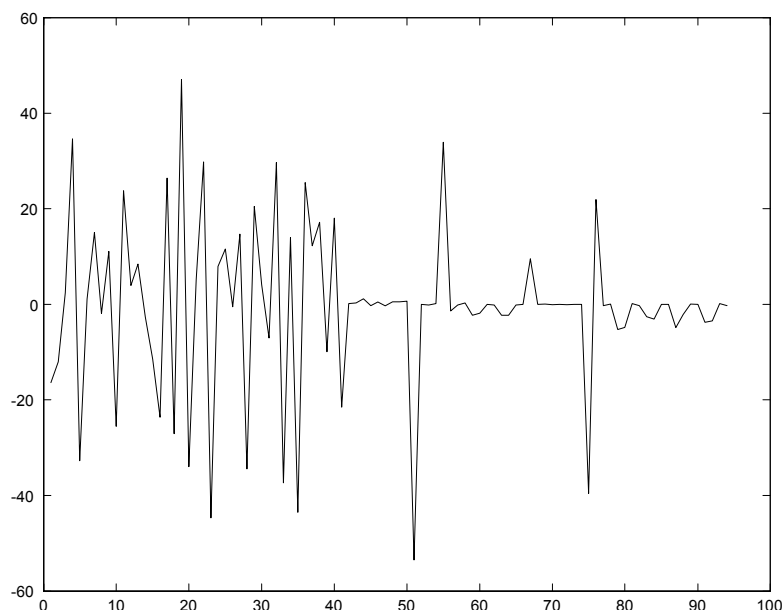


Figura 6.6: Segunda diferencia de la serie de la población mayor de 16 años en España

Generalizando, diremos que un proceso es *integrado de orden*  $h \geq 0$ , y lo representaremos por  $I(h)$ , cuando al diferenciarlo  $h$  veces se obtiene un proceso estacionario. Un proceso estacionario es, por tanto, siempre  $I(0)$ . En la práctica la mayoría de las series no estacionarias que son integradas tienen un orden  $h \leq 3$ .

### 6.3 El Paseo aleatorio

Hemos visto que los procesos MA finitos son siempre estacionarios y que los AR lo son si las raíces de  $\phi(B) = 0$  están fuera del círculo unidad. Consideremos el AR(1):

$$z_t = c + \phi z_{t-1} + a_t. \quad (6.3)$$

Si  $|\phi| > 1$  el proceso es explosivo, como muestra la figura 6.7, que presenta una simulación de este proceso con  $c = 0$  y  $\phi = 1.03$ . Si  $|\phi| = 1$  el proceso no es estacionario, pero tampoco es explosivo, y pertenece a la clase de procesos integrados de orden uno, ya que su primera diferencia,  $z_t - z_{t-1} = c + a_t$ , sí es un proceso estacionario. Este proceso se denomina paseo aleatorio.

Una característica importante que diferencia los procesos estacionarios y los no estacionarios es el papel de las constantes. En un proceso estacionario la constante no es importante, y en los capítulos anteriores hemos restado a las observaciones su media y trabajado con procesos de media cero. La forma del proceso y sus propiedades básicas son las mismas si tiene media cero o distinta de cero. Sin embargo, en un proceso no estacionario las constantes, si existen, son muy importantes y representan alguna propiedad determinista del proceso. Por ejemplo, la figura 6.8 presenta dos simulaciones del modelo (6.3). En la primera  $c = 0$ , se dice que el proceso no tiene deriva, y el nivel de la serie oscila en el tiempo. En la segunda  $c = 1$ , se dice que el proceso tiene una deriva igual a la unidad, y el proceso muestra una tendencia lineal de pendiente  $c$  en la serie. En efecto, el nivel de la serie es el del periodo anterior más  $c$ , lo que supone un crecimiento lineal determinista.

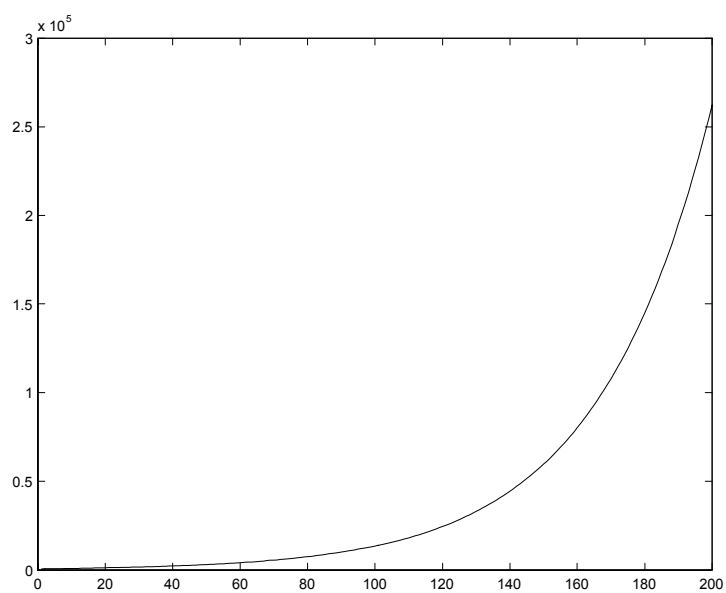


Figura 6.7: Simulación de un AR con parámetro 1.03

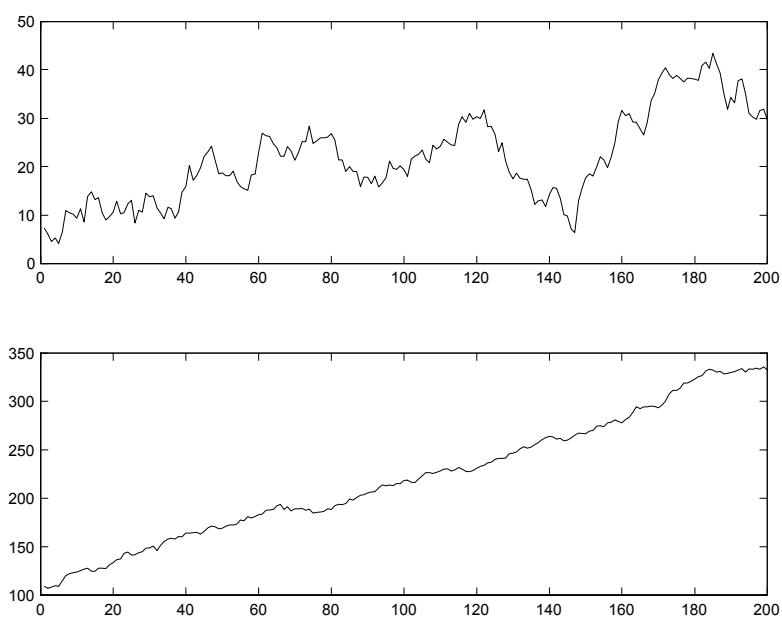


Figura 6.8: Dos paseos aleatorios simulados. El primero no tiene constante, el segundo tiene  $c=1$ .

Utilizando el *operador diferencia*  $\nabla = 1 - B$ , el paseo aleatorio se escribe

$$\nabla z_t = c + a_t. \quad (6.4)$$

Para calcular la función de autocorrelación de este proceso supondremos que comienza en  $t = 0$ . Entonces, sustituyendo sucesivamente  $z_t$  por  $z_{t-1}$  tendremos:

$$z_t = ct + a_t + a_{t-1} + a_{t-2} + \dots + a_1$$

y la media del proceso será  $ct$ . Su varianza será:

$$Var(z_t) = E(a_t + a_{t-1} + a_{t-2} + \dots + a_1)^2 = \sigma^2 t$$

y vemos que tanto la media como la varianza aumentan con el tiempo y, tienden a infinito con  $t$ . Escribiendo la expresión del proceso para  $t + k$ :

$$z_{t+k} = c(t+k) + a_{t+k} + \dots + a_t + \dots + a_1$$

y la autocovarianza será:

$$Cov(t, t+k) = E[(z_t - ct)(z_{t+k} - c(t+k))] = \sigma^2 t$$

Observemos que la autocovarianzas también son variables en el tiempo. Esto es esperable en procesos no estacionarios, donde las autocovarianzas no son sólo función del retardo, sino que dependen del instante en que se calcula. Es fácil comprobar que

$$Cov(t, t-k) = \sigma^2(t-k) \neq Cov(t, t+k).$$

La función de autocorrelación es:

$$\rho(t, t+k) = \frac{t}{\sqrt{t(t+k)}} = \left(1 + \frac{k}{t}\right)^{-1/2},$$

y si  $t$  es grande, los coeficientes de la función de autocorrelación serán próximos a uno y *decrecerán aproximadamente de forma lineal con  $k$* . Tomando logaritmos tenemos que

$$\log \rho(t, t+k) = -\frac{1}{2} \log\left(1 + \frac{k}{t}\right) \approx -\frac{k}{2t}$$

y los logaritmos de las correlaciones decrecerán linealmente con  $k$ . Este decrecimiento lineal se verá también sobre los coeficientes ya que para  $k/t$  pequeño, la función  $(1 + k/t)^{-1/2}$  puede aproximarse como

$$\rho(t, t+k) \simeq 1 - \frac{k}{2t}. \quad (6.5)$$

Observemos que esta función puede considerarse como un caso límite de la de un AR(1) cuando  $\phi$  se aproxima a la unidad. La figura 6.9 presenta un gráfico de la función de autocorrelación de una simulación de 300 observaciones que siguen un paseo aleatorio. Se observa claramente el decrecimiento lineal de la *fas*.

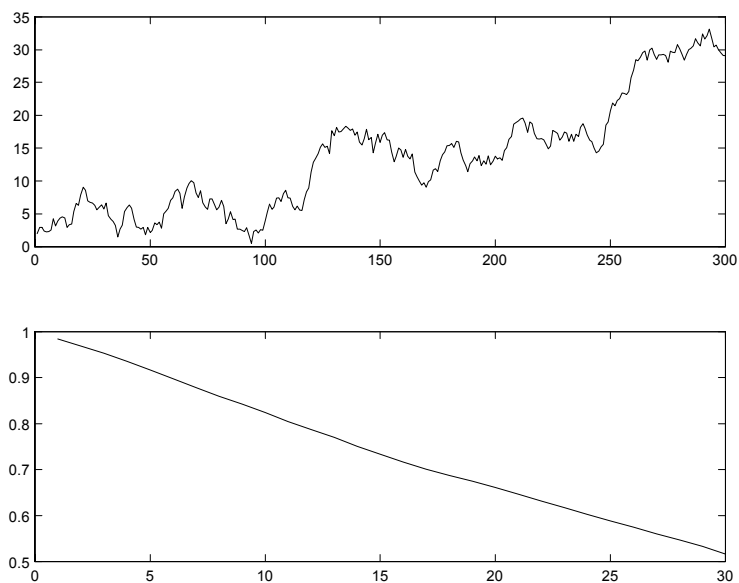


Figura 6.9: Generación de 300 observaciones que siguen un paseo aleatorio (gráfico superior) y su fas calculada en esta muestra (gráfico inferior).

### Ejemplo 6.1

Vamos a calcular la función de autocorrelación muestral de la serie de la figura 6.1, la serie mensual de evolución del índice general de la bolsa de Madrid entre enero de 1988 y mayo de 2003. La figura 6.10 presenta su función de autocorrelación. Se observa que las correlaciones decaen linealmente, indicando un proceso no estacionario.

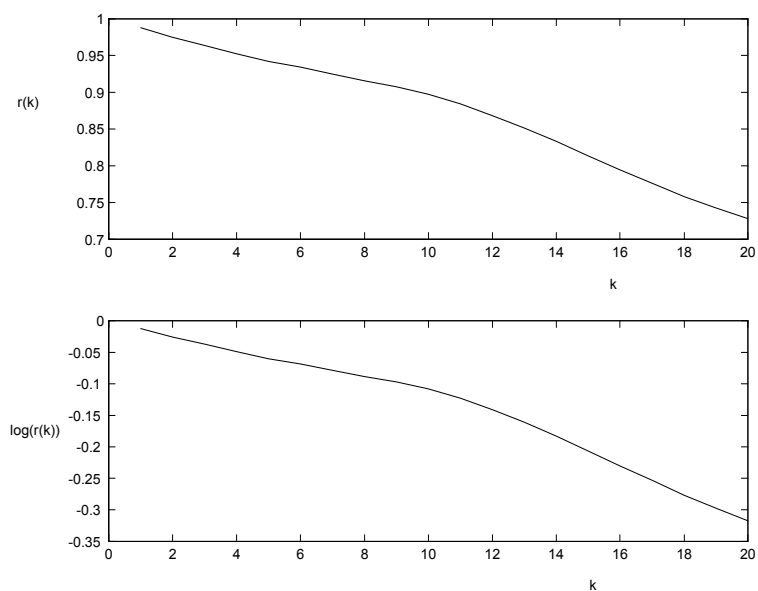


Figura 6.10: Función de autocorrelación simple y su logaritmo como función del retardo para los datos de la bolsa de Madrid, enero 1988 a mayo 2003.



## 6.4 El proceso de alisado exponencial simple

Un proceso no estacionario, que se ha utilizado mucho para la previsión, es:

$$\nabla z_t = c + (1 - \theta B) a_t \quad (6.6)$$

que es un caso límite del ARMA(1, 1) cuando el coeficiente autorregresivo es la unidad. Cuando hacemos  $\theta = 0$  obtenemos el paseo aleatorio. Este proceso es capaz de representar distintas forma de crecimiento lineal de una serie. Por ejemplo, consideremos el proceso de crecimiento lineal determinista:

$$z_t = b + ct + a_t \quad (6.7)$$

donde  $a_t$  es ruido blanco. Este proceso es integrado de orden uno o  $I(1)$ , ya que tomando la primera diferencia de la serie  $z_t$ :

$$\begin{aligned} \omega_t &= z_t - z_{t-1} = b + ct + a_t - (b + c(t-1) + a_{t-1}) = \\ &= c + a_t - a_{t-1} \end{aligned}$$

como  $c$  es constante y  $a_t - a_{t-1}$ , es un proceso estacionario, (al ser la diferencia de dos procesos estacionarios, véase la sección 3.4), el proceso  $\omega_t$  será también estacionario. Por tanto,  $z_t$  definido por (6.7), es integrado de orden uno, y su primera diferencia seguirá el modelo (6.6) pero con un parámetro MA(1) igual a la unidad. El proceso será no invertible, y esta propiedad nos indicará que tenemos un componente lineal determinista en el modelo. Observemos que al diferenciar la constante que queda es la pendiente del crecimiento determinista lineal.

Consideremos ahora una serie con tendencia lineal determinista pero con estructura más flexible que el anterior. Podemos escribir el proceso como en el capítulo 2 con una tendencia lineal pero con componentes variables, como:

$$\begin{aligned} z_t &= \mu_t + v_t \\ \mu_t &= \mu_{t-1} + c + \varepsilon_t \end{aligned} \quad (6.8)$$

tomando diferencias en la serie observada, tenemos que

$$\nabla z_t = c + \varepsilon_t + \nabla v_t$$

y como la suma un proceso MA(1) y un ruido blanco es un MA(1), obtenemos de nuevo el proceso (6.6), pero ahora con un parámetro  $\theta$  menor que la unidad, como en la representación (6.6).

Es interesante que el lector entienda la diferencia entre los dos modelos presentados. El modelo (6.7) es determinista puro, ya que en cada instante,  $t$ , el valor esperado de la serie está determinado, es  $b + ct$ . Además, el valor esperado de la serie en  $z_t$  conocido el valor en  $z_{t-1}$  sigue siendo el mismo,  $b + ct$ . En este modelo conocer los valores previos no modifica las predicciones. El modelo (6.8) es más general, porque aunque tiene también en cada instante un crecimiento lineal igual a  $c$ , el nivel de la serie en cada instante no está determinado desde el principio y puede variar aleatoriamente. Esto se traduce en que el valor esperado de la serie en  $z_t$  conocido el valor en  $z_{t-1}$  es  $z_{t-1} + c$ , y depende del valor observado. El modelo (6.7) es muy raro en la práctica, mientras que el (6.8) si aparece con frecuencia en la práctica.

Suponiendo para simplificar  $c = 0$  y  $|\theta| < 1$  e invirtiendo la parte MA, este proceso se escribe:

$$z_t = (1 - \theta)z_{t-1} + \theta(1 - \theta)z_{t-2} + \theta^2(1 - \theta)z_{t-3} + \dots + a_t \quad (6.9)$$

En el caso habitual donde  $0 < \theta < 1$ , todos los coeficientes de la representación AR( $\infty$ ) de la ecuación (6.9) son positivos y además suman la unidad. En efecto, estos coeficientes forman una progresión geométrica indefinida con primer término  $(1 - \theta)$  y razón  $\theta$ , con lo que su suma es 1. Por tanto, la ecuación (6.9) indica que el valor observado en cada instante es una media ponderada de los valores de la serie en instantes anteriores, con coeficientes que decrecen geométricamente con el retardo. Si  $\theta$  es próximo a la unidad la media se calcula con muchos coeficientes, pero cada uno de ellos con poco peso. Por ejemplo, si  $\theta = .9$  los pesos

de los retardos son .1, .09, .081, ..y los pesos decrecen muy lentamente y son similares para observaciones próximas. Sin embargo, si  $\theta$  es próximo a cero, la media se calcula ponderado sólo los últimos valores observados. Por ejemplo, si  $\theta = .1$  los pesos de los retardos son .9, .09, .009,...y practicamente sólo se tienen en cuenta los dos últimos retardos con pesos muy distintos. Este proceso se conoce como proceso de alisado exponencial simple y fue introducido de forma intuitiva en el capítulo 2 como el proceso de alisado simple. Cuando  $c \neq 0$  el modelo esta relacionado con el alisado doble de Holt, que vimos también en el capítulo 2.

Para calcular la *fas* del proceso supondremos que empezó en  $t = 0$ , con  $z_0 = a_0 = 0$  y  $c = 0$ . Entonces:

$$\begin{aligned} z_1 &= a_1 \\ z_2 &= z_1 + a_2 - \theta a_1 = a_2 + (1 - \theta) a_1 \\ z_3 &= z_2 + a_3 - \theta a_2 = a_3 + (1 - \theta) a_2 + (1 - \theta) a_1 \\ &\dots\dots\dots \\ z_t &= a_t + (1 - \theta) a_{t-1} + \dots + (1 - \theta) a_1 \end{aligned}$$

la varianza será:

$$Var(z_t) = \sigma^2(1 + (t - 1)(1 - \theta)^2)$$

y como el proceso es no estacionario depende del tiempo. Las covarianzas se calculan como:

$$Cov(t, t + k) = \sigma^2(1 - \theta)(1 + (t - 1)(1 - \theta))$$

y dan lugar a la función de autocorrelación:

$$\rho(t, t + k) = \frac{(1 - \theta)(1 + (t - 1)(1 - \theta))}{\sqrt{(1 + (t - 1)(1 - \theta)^2)(1 + (t + k - 1)(1 - \theta)^2)}}.$$

Suponiendo que  $t$  es grande, de manera que  $t - 1 \simeq t$ , y que  $\theta$  no esta muy próximo a uno, de manera que  $1 + t(1 - \theta)^2 \simeq t(1 - \theta)^2$ , entonces:

$$\rho(t, t + k) = \frac{t}{\sqrt{t(t + k)}}$$

que es la misma función de autocorrelación obtenida para el paseo aleatorio: coeficientes próximos a uno que se *amortiguan de forma aproximadamente lineal* al aumentar el retardo. Sin embargo, si  $\theta$  es próximo a uno, los coeficientes no tienen necesariamente que ser grandes, aunque siempre debemos observar el decrecimiento lineal. Por ejemplo, la figura 6.11 muestra una serie generada con  $\theta = .95$ . Se observa que aunque las autocorrelaciones ahora no son nunca próximas a uno, muestran la característica esperada de decrecer linealmente.

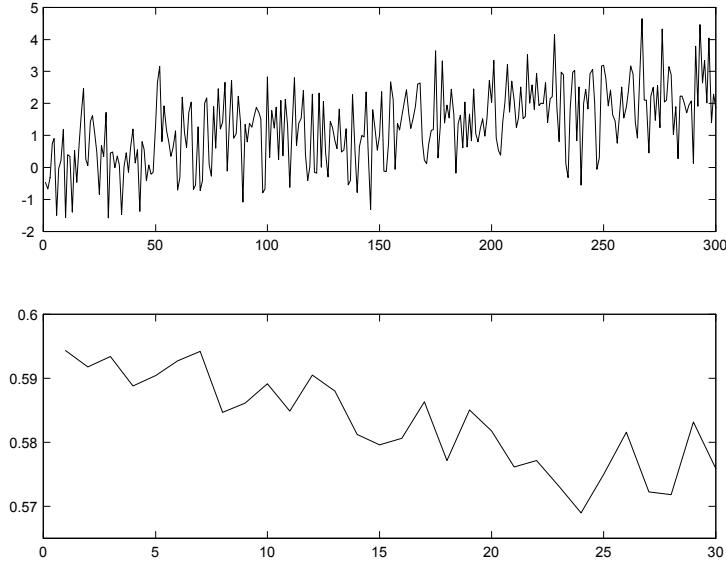


Figura 6.11: Generación de 300 observaciones que siguen un proceso IMA (1,1) con parámetro  $\theta = .95$  (gráfico superior) y su fas calculada en esta muestra (gráfico inferior).

## 6.5 Procesos Integrados ARIMA

Los dos procesos anteriores se han obtenido admitiendo que la raíz de la parte AR de los procesos AR(1) y ARMA(1,1) es la unidad, con lo que estos procesos se convierten en no estacionarios. Esta idea puede generalizarse para cualquier proceso ARMA, permitiendo una o varias raíces unidad en el operador AR. Se obtienen entonces procesos del tipo:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d z_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) a_t$$

que llamaremos procesos ARIMA  $(p, d, q)$ . En esta notación  $p$  es el orden de la parte autorregresiva estacionaria,  $d$  es el número de raíces unitarias (orden de integración del proceso) y  $q$  es el orden de la parte media móvil. Utilizando el operador diferencia,  $\nabla = 1 - B$ , el proceso anterior suele escribirse:

$$\phi_p(B) \nabla^d z_t = c + \theta_q(B) a_t \quad (6.10)$$

El nombre ARIMA proviene de las iniciales en inglés de los procesos autorregresivos integrados de media móvil (**a**utoregressive **i**ntegrated **m**oving **a**verage), donde «integrado» indica que, llamando  $\omega_t = \nabla^d z_t$ , al proceso estacionario,  $z_t$  se obtiene como suma (integración) de  $\omega_t$ . En efecto, si

$$\omega_t = (1 - B) z_t$$

como:

$$(1 - B)^{-1} = 1 + B + B^2 + B^3 + \dots$$

resulta:

$$z_t = (1 - B)^{-1} \omega_t = \sum_{j=-\infty}^t \omega_j.$$

Hemos visto en las secciones anteriores dos ejemplos de procesos ARIMA: el paseo aleatorio es el modelo ARIMA (0,1,0) y el de alisado exponencial simple el ARIMA(0, 1, 1) o IMA(1, 1). Ambos se caracterizan

porque la función de autocorrelación simple tiene coeficientes que decrecen lentamente. Todos los procesos ARIMA no estacionarios tienen esta propiedad general. Para demostrarlo, recordemos que el correlograma de un ARMA(p, q) satisface para  $k > q$  la ecuación:

$$\phi_p(B)\rho_k = 0 \quad k > q$$

cuya solución es del tipo:

$$\rho_k = \sum_1^p A_i G_i^k$$

donde  $G_i^{-1}$  son las raíces de  $\phi_p(B) = 0$  y  $|G_i| < 1$ . Si una de estas raíces  $G_i$  es muy próxima a la unidad, escribiendo  $G_i = 1 - \varepsilon$ , siendo  $\varepsilon$  muy pequeño, para  $k$  grande los términos  $A_j G_j^k$  debidos a las otras raíces serán cero (ya que  $G_j^k \rightarrow 0$ ) y tendremos que, aproximadamente:

$$\rho_k = A_i(1 - \varepsilon)^k \simeq A_i(1 - k\varepsilon) \quad \text{para } k \text{ grande}$$

En consecuencia, la *fas* tendrá coeficientes positivos que se amortiguarán aproximadamente de forma lineal y pueden ser distintos de cero incluso para valores altos de  $k$ . Esta propiedad de persistencia de valores positivos en el correlograma (aunque sean pequeños) y de decrecimiento lineal caracteriza a los procesos no estacionarios. La figura 6.12 muestra una realización de la serie  $(1 - .4B)\nabla z_t = (1 - .8B)a_t$  y su función de autocorrelación observada. Se observa de nuevo el decrecimiento lineal de los coeficientes de autocorrelación.

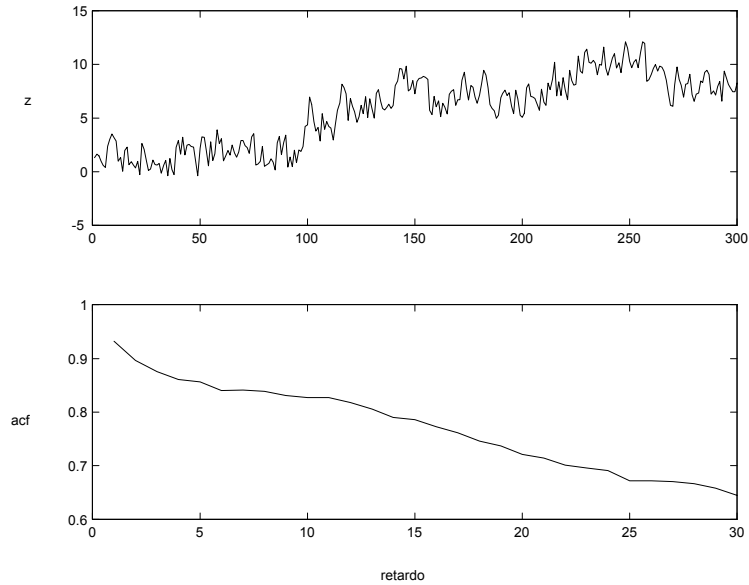


Figura 6.12: Simulación del proceso  $(1 - .4B)\nabla z_t = (1 - .8B)a_t$  y su función de autocorrelación

## 6.6 Procesos integrados y tendencias

No todas las series no estacionarias son integradas, pero los procesos integrados cubren muchos casos interesantes que nos encontramos en la práctica. Es fácil demostrar que cualquier proceso que sea la suma de una tendencia polinómica y un proceso estacionario será integrado. Hemos visto esta propiedad para series generadas como suma de una tendencia lineal y ruido blanco, pero el resultado es válido para procesos generados como suma de una tendencia polinómica de orden  $h$  y un proceso estacionario cualquiera,  $u_t$ :

$$z_t = \mu_t + u_t \quad (6.11)$$

donde

$$\mu_t = a + bt + ct^2 + \dots + dt^h$$

Si diferenciamos  $h$  veces este proceso se obtiene una serie estacionaria con una parte de media móvil no invertible. Por ejemplo, supongamos una tendencia cuadrática. Entonces tomando dos diferencias en el proceso, resulta:

$$\nabla^2 z_t = c + \nabla^2 u_t$$

y el proceso  $\eta = \nabla^2 u_t$  tendrá un componente de media móvil no invertible. Por ejemplo, si  $u_t$  es un AR(1),  $(1 - \phi B)u_t = a_t$ , tendremos que

$$\eta = (1 - \phi B)^{-1} \nabla^2 a_t$$

y  $\eta$  será un proceso ARMA(1,2) no invertible.

Este ejemplo indica que si una serie tiene una tendencia polinómica determinista y se modela como un proceso ARIMA, obtendremos: (1) componentes de media móvil no invertibles; (2) la serie estacionaria tendrá una media distinta de cero y la constante es el coeficiente de orden mayor en la tendencia determinista. De nuevo nos encontramos con la importancia de las constantes en los procesos integrados, ya que representan efectos deterministas.

Estas series con estructura determinista no son frecuentes en la práctica. Son más frecuentes modelos del tipo (6.11) donde se supone que  $u_t$  es ruido blanco y el nivel  $\mu_t$  se supone variable en el tiempo de forma estocástica, como

$$\mu_t = \mu_{t-1} + \beta_t + \varepsilon_t$$

y

$$\beta_t = \beta_{t-1} + v_t$$

donde los procesos  $\varepsilon_t$  y  $v_t$  son ruidos blancos independientes. En este modelo el nivel crece de forma lineal, pero con una pendiente,  $\beta_t$ , que cambia en el tiempo. Estos modelos se denominan a veces estructurales, porque describen el comportamiento de la media del proceso. En esta representación si la varianza de  $v_t$  es pequeña la serie tiene una tendencia casi constante. Si tomamos una diferencia en la serie original, tenemos que

$$\nabla z_t = \beta_t + \varepsilon_t + \nabla u_t$$

y tomando una segunda diferencia

$$\nabla^2 z_t = v_t + \nabla \varepsilon_t + \nabla^2 u_t = (1 - \theta_1 B - \theta_2 B^2) a_t \quad (6.12)$$

ya que si sumamos procesos de media móvil se obtiene un nuevo proceso de media móvil igual al orden máximo de los sumandos. Dependiendo de las varianzas de los tres ruidos blancos podemos tener distintas situaciones. Si la varianza de  $u_t$  es mucho menor que las de los otros ruidos el término  $\nabla^2 u_t$  puede despreciarse en el segundo miembro de (6.12) y la suma será un MA(1). Análogamente, si el término  $v_t$  es el dominante y despreciamos los otros dos la serie seguirá aproximadamente el modelo  $\nabla^2 z_t = v_t$ . Por otro lado, si la varianza de  $v_t$  es cero entonces  $\beta_t = \beta_{t-1} = \beta$  y volvemos al modelo estudiado en la sección 6.4 que tiene una tendencia determinista.

En resumen, los modelos ARIMA incorporaran tendencias estocásticas y, como caso límite, determinísticas. En este caso la serie diferenciada tendrá una media distinta de cero y, en el caso particular de estructuras de crecimiento generadas por una ecuación determinista más ruido blanco, tendremos términos de media móvil no invertibles.

## 6.7 Procesos de memoria larga

Existe una clase de procesos estacionarios que puede confundirse fácilmente con los procesos no estacionarios integrados. Estos son los procesos de memoria larga, que se caracterizan por tener muchos coeficientes de autocorrelación con coeficientes pequeños y que se amortiguan lentamente. Esta propiedad se ha observado en ciertas series meteorológicas o ambientales estacionarias. El lento decrecimiento de las autocorrelaciones

podría llevar a modelar estas series suponiendo que son integradas de orden uno, y tomando una diferencia, en lugar de como procesos estacionarios aunque de memoria muy larga. Los procesos de memoria larga ocupan un lugar intermedio entre los procesos ARMA, donde las correlaciones decrecen rápidamente a cero, y los procesos integrados ARIMA, donde decrecen linealmente.

El proceso de memoria larga más simple se obtiene generalizando el paseo aleatorio con

$$(1 - B)^d z_t = a_t \quad (6.13)$$

donde ahora el parámetro  $d$  en lugar de ser un número entero es un número real. En el caso  $-.5 < d < .5$  se obtiene un proceso estacionario pero que, como vamos a ver, tiene la propiedad de que sus autocorrelaciones decrecen muy lentamente. Para definir el operador  $(1 - B)^d$  cuando  $d$  no es entero, partiremos de la expresión del desarrollo del binomio de Newton. Cuando  $d$  es un número entero positivo, se verifica que:

$$(1 - B)^d = \sum_{i=0}^d \binom{d}{i} (1)^{d-i} (-B)^i = \sum_{i=0}^d \alpha_i B^i \quad (6.14)$$

donde los coeficientes de esta suma se definen por

$$\alpha_i = \binom{d}{i} (-1)^i = \frac{d!}{(d-i)!i!} (-1)^i = \frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)} (-1)^i$$

siendo  $\Gamma(a+1) = a\Gamma(a)$  la función gamma (véase el apéndice 6A) que para valores enteros coincide con la función factorial. Para generalizar esta definición cuando  $d$  sea un número real, podemos aprovechar que la función gamma esta definida para cualquier número real, y escribir para  $d$  real:

$$(1 - B)^d = \sum_{i=0}^{\infty} \alpha_i B^i \quad (6.15)$$

donde los coeficientes de esta expansión infinita viene dados por

$$\alpha_i = \frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)} (-1)^i$$

y puede demostrarse (véase el apéndice 6.1) que cuando  $d$  es entero esta definición es coherente con (6.14), ya que todos los términos mayores de  $d$  en este sumatorio infinito son cero, con lo que para  $d$  entero (6.15) se reduce a (6.14). Observemos que para  $i = 0$  resulta  $\alpha_0 = 1$ , y utilizando las propiedades de la función gamma se demuestra que:

$$\alpha_i = \prod_{0 < j \leq i} \frac{j-1-d}{j}, \quad \text{para } i = 1, 2, \dots$$

Utilizando esta expresión podemos escribir el proceso (6.13) mediante una representación  $\text{AR}(\infty)$  desarrollando el operador  $(1 - B)^d$ , para obtener:

$$z_t = - \sum_{i=0}^{\infty} \alpha_i z_{t-i} + a_t \quad (6.16)$$

Puede demostrarse (véase Beran, 1994) que este proceso admite también una representación  $\text{MA}(\infty)$  de la forma

$$z_t = \sum_{i=0}^{\infty} \psi_i a_{t-i}$$

donde los coeficientes verifican  $\sum \psi_i^2 < \infty$  y se obtienen con

$$\psi_i = \frac{\Gamma(d+i)}{\Gamma(i+1)\Gamma(d)}$$

Se demuestra que la función de autocorrelación del proceso de larga memoria (6.13) es

$$\rho(k) = \frac{\Gamma(d+k)\Gamma(1-d)}{\Gamma(k-d+1)\Gamma(d)}$$

que puede aproximarse, para  $k$  grande, por

$$\rho(k) \simeq \frac{\Gamma(1-d)}{\Gamma(d)} |k|^{2d-1}$$

y tomando logaritmos

$$\log \rho(k) \simeq a + (2d-1) \log |k|$$

Por tanto, si representamos  $\log \rho(k)$  frente a  $\log |k|$  obtenemos una recta con pendiente  $2d-1$ . Esta propiedad diferencia un proceso de memoria larga de un proceso no estacionario donde hemos visto que  $\log \rho(k)$  decrece linealmente con  $k$  y no con  $\log k$ .

La figura 6.13 representa los coeficientes  $-\alpha_i$  de la representación AR dada por (6.16) para varios valores de  $d$  y la figura 6.14 la función de autocorrelación para estos valores de  $d$ . Se observa que los coeficientes AR son en seguida muy pequeños y que la función de autocorrelación tiene muchos coeficientes no nulos que decrecen lentamente. Estas propiedades pueden fácilmente llevar a confundir este proceso con  $\text{AR}(p)$  de orden alto, y si  $d$  es alto, próximo a .5, los altos valores y el lento decrecimiento de la función de autocorrelación pueden sugerir un proceso integrado de orden uno.

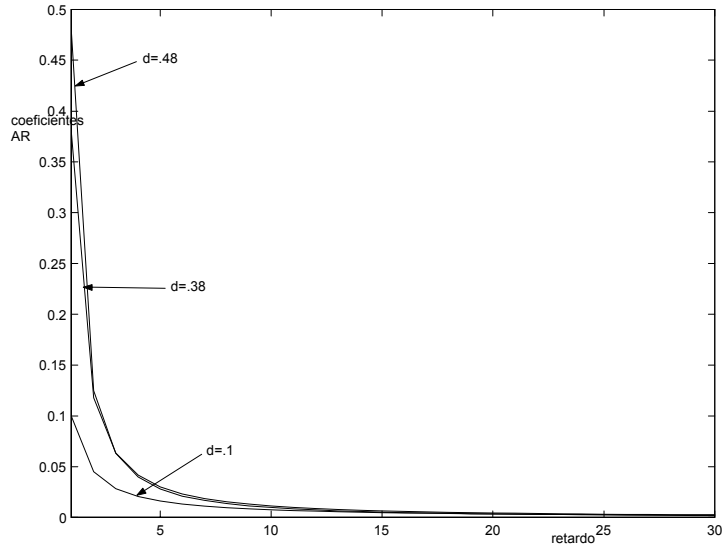


Figura 6.13: Coeficientes AR del proceso de memoria larga  $\nabla^d z_t = a_t$  con  $d < .5$ .

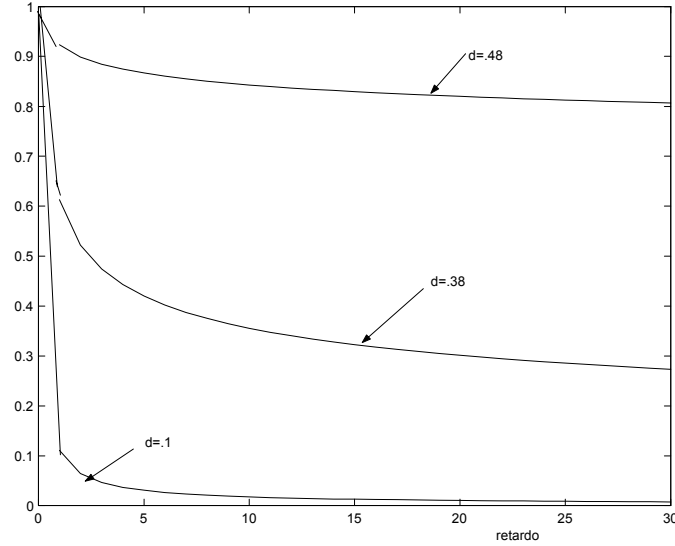


Figura 6.14: Función de autocorrelación del proceso  $\nabla^d z_t = a_t$  con  $d < .5$

Generalizando lo anterior, podemos considerar procesos estacionarios ARMA pero que incluyan además memoria larga, definiendo

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d z_t = c + (1 - \theta_1 B - \dots - \theta_q B^q) a_t$$

donde  $d$  no es necesariamente entero. Si  $d < 0.5$  el proceso es estacionario, mientras que si  $d \geq 0.5$  el proceso es no estacionario. Estos modelos se denominan ARFIMA  $(p, d, q)$ , **autorregresivos fraccionalmente integrados de media móvil**. Si el proceso es estacionario tiene la propiedad de que para retardos altos, cuando la estructura ARMA desaparezca, mostrarán la característica de memoria larga. Si no es estacionario, al tomar una diferencia se convertirán en estacionarios, pero de memoria larga.

Los procesos de memoria larga pueden aparecer al agregar series de memoria corta bajo determinadas condiciones. Granger (1980) demostró que si sumamos  $N$  procesos AR(1) independientes pero cuyos parámetros provienen de una distribución común, si  $N$  es muy grande como límite obtenemos un proceso de memoria larga. Recordemos que, como vimos en la sección 5.9, la agregación de procesos AR independientes conduce a procesos ARMA, y este es el resultado esperable con pocos sumandos. El resultado de Granger indica que en el límite tendremos un proceso de memoria larga y sugiere que estos procesos pueden aproximarse por un ARMA con órdenes altos y similares de  $p$  y  $q$ , lo que se comprueba en la práctica. Con series no muy largas es difícil diferenciar un modelo de memoria larga de un ARMA. Sin embargo, si disponemos de series con muchas observaciones, como suele ocurrir con datos climatológicos y financieros, pueden proporcionar en algunos casos mejores aproximaciones a las series observadas que los procesos ARMA de memoria corta.

### Ejemplo 6.2

La figura 6.15 presenta la serie anual de nivel mínimo de río Nilo entre los años 622 y 1284. Su función de autocorrelación se presenta en la figura 6.16, y se observa que las correlaciones decrecen muy lentamente. La figura 6.17 presenta el logaritmo de las autocorrelaciones frente al retardo, y se observa que la relación es aproximadamente lineal. Esta figura incluye también los datos de la bolsa de Madrid, y se observa cuando  $k$  es grande que los coeficientes para esta serie, que sigue un modelo ARIMA de memoria corta, decrecen más rápidamente de lo que correspondería a un proceso de memoria larga. Como hemos visto en la figura 6.10 si representamos el logaritmo de las autocorrelaciones como función del retardo si se obtiene un decrecimiento lineal.

El gráfico muestra también que para diferenciar estos modelos de los ARIMA necesitamos muchos datos.



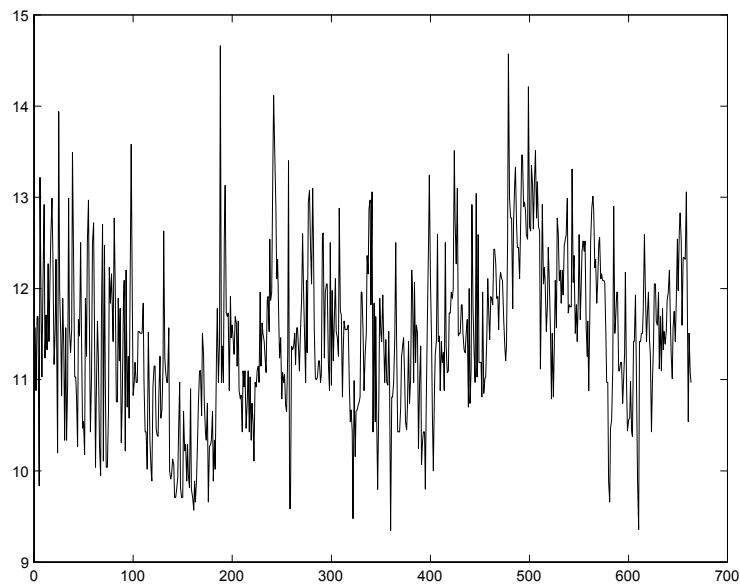


Figura 6.15: Mínimos anuales del nivel del río Nilo desde el año 622 hasta el año 1284

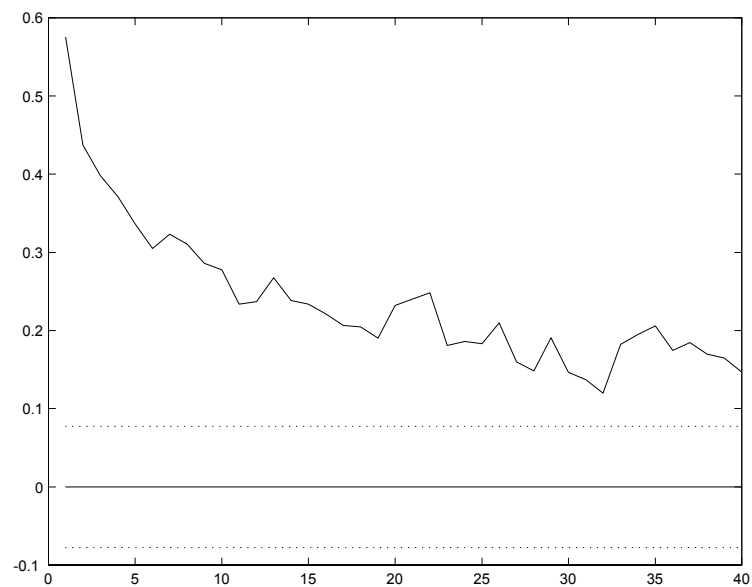


Figura 6.16: Función de autocorrelación para la serie del nivel mínimo del Nilo.

### Ejercicios 6

- 6.1 Demostrar que la suma y la diferencia de dos procesos estacionarios es estacionaria
- 6.2 Demostrar que el modelo  $z_t = a + bt + ct^2 + a_t$ , donde  $a_t$  es un proceso de ruido blanco se convierte en un proceso estacionario y no invertible al tomar dos diferencias.
- 6.3 Demostrar que las autocorrelaciones de un paseo aleatorio pueden aproximarse por  $\rho(t, t+k) \simeq 1 - \frac{k}{2t}$ .
- 6.4 Simular un proceso ARIMA (0,1,1) con valores del parámetro  $\theta = .4, .7$  y  $.9$  y estudiar el decrecimiento de la función de autocorrelación del proceso

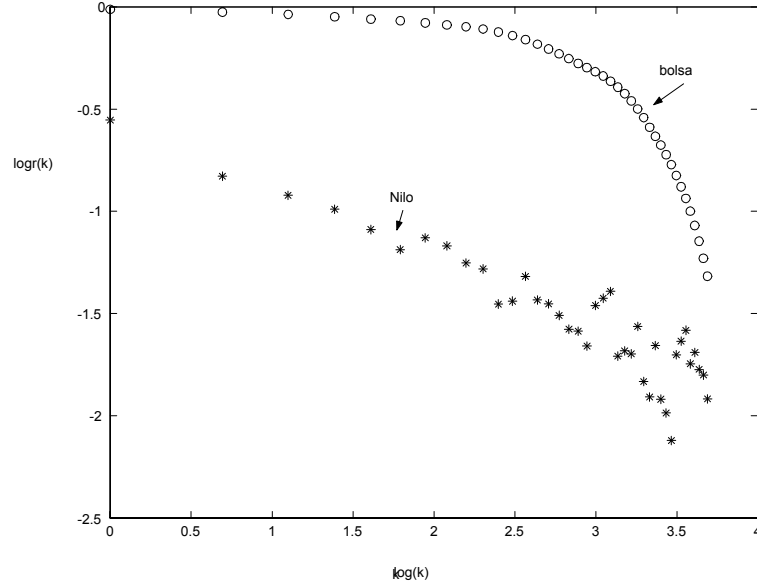


Figura 6.17: Relación entre el logaritmo del coeficiente de autocorrelación y el logaritmo del retardo para los datos del Nilo y los de la Bolsa.

6.5 Simular el proceso  $\nabla^2 z_t = (1 - .8B)a_t$  y estudiar el decrecimiento de la función de autocorrelación del proceso.

### Apendice 6.1

#### La función Gamma y los procesos de memoria larga

La función gamma se define por

$$\Gamma(p) = \int_0^{\infty} x^{p-1} e^{-x} dx$$

para  $p > 0$ . Integrando por partes se obtiene

$$\Gamma(p) = [-x^{p-1} e^{-x}]_0^{\infty} + (p-1) \int_0^{\infty} x^{p-2} e^{-x} dx = (p-1)\Gamma(p-1)$$

Como  $\Gamma(1) = 1$  la función gamma coincide con la función factorial para  $p$  entero. Para cualquier valor positivo y real de  $p$  esta función generaliza la función factorial, ya que siempre podemos escribir  $p = n + r$  siendo  $n$  entero:

$$\Gamma(n+r) = (n+r-1)(n+r-2)\dots(1+r)\Gamma(1+r)$$

Esta propiedad hace que la función gamma se tabule entre 1 y 2 porque podemos calcularla para cualquier  $p$  real aplicando esta relación y sólo necesitamos el valor de  $\Gamma(1+r)$ , que corresponderá a un argumento de la función entre 1 y 2. El valor de la función en ese intervalo es siempre próxima a la unidad, tomando el valor mínimo para  $1+r = 1.4616$  donde entonces la función vale  $\Gamma(1.4616) = .8856$ .

La función gamma puede también definirse para  $p$  negativo, mediante:

$$\Gamma(p+1) = \lim_{n \rightarrow \infty} \frac{n^n n!}{(p+1)(p+2)\dots(p+n)}$$

y se demuestra que este límite existe y que para valores positivos de  $p$  volvemos al caso anterior y el límite coincide con la integral de definición para  $p > 0$ . Además, la función así definida sigue cumpliendo la ley recurrente  $\Gamma(p) = (p-1)\Gamma(p-1)$ . Se demuestra también que esta función se hace infinita para cero y los valores enteros negativos del argumento,  $p = 0, -1, -2, -3$ .

Estos resultados llevan a definir el operador diferencia fraccional por la suma infinita:

$$(1 - B)^d = \sum_{i=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)} (-1)^i B^i$$

si  $d$  es entero cuando  $i > d$  entonces  $\Gamma(d-i+1)$  corresponde a la gamma con argumento entero negativo que, según hemos visto, es infinita. Por tanto, todos los coeficientes serán cero para  $i > d$  y obtendremos la expresión habitual del desarrollo de  $(1-B)^d$  cuando  $d$  es entero mediante el binomio de Newton. Observemos que como

$$\frac{\Gamma(d+1)}{\Gamma(d-i+1)} = d(d-1)(d-2)\dots(d-i+1)$$

e igualmente

$$\frac{\Gamma(i-d)}{\Gamma(-d)} = (i-d-1)(i-d-2)\dots(1-d)(-d)$$

podemos escribir los términos del desarrollo general de  $(1-B)^d$  mediante:

$$\frac{\Gamma(d+1)}{\Gamma(i+1)\Gamma(d-i+1)} (-1)^i = \frac{\Gamma(i-d)}{\Gamma(i+1)\Gamma(-d)}$$

que es la forma habitual de expresar los coeficientes del desarrollo cuando  $p$  no es entero.