

10

C A P Í T U L O

RAÍCES UNITARIAS, TENDENCIAS ESTOCÁSTICAS, MODELOS ARIMA DE PRONÓSTICO Y SUAVIZAMIENTO

Durante el desarrollo de este libro hemos manejado la falta de estacionariedad (por ejemplo, la presencia de tendencia), con componentes determinísticas. Ahora describiremos un método estocástico alternativo. La tendencia estocástica es importante porque a veces proporciona una buena descripción de ciertas series temporales comerciales, económicas y financieras, y porque tiene varias propiedades e implicaciones especiales. Por ejemplo, si supiéramos a ciencia cierta que una serie tiene una tendencia estocástica, podríamos diferenciar la serie y a continuación, ajustar a la diferencia un modelo estacionario de tendencia.¹ La estrategia de diferenciar para lograr la estacionariedad contrasta con el método descrito en los capítulos anteriores, con el cual trabajamos en niveles e incluimos tendencias determinísticas. En la práctica, en ocasiones resulta muy difícil decidir si se modela mejor una tendencia como determinística o como estocástica, y tal decisión es parte importante de la ciencia y el arte de elaborar modelos de pronóstico.

1. TENDENCIAS ESTOCÁSTICAS Y PRONÓSTICO

Consideremos un proceso ARMA (p, q),

$$\Phi(L)y_t = \Theta(L)\varepsilon_t$$

con todas las raíces autorregresivas sobre o fuera del círculo unitario, a lo más una raíz autorregresiva sobre el círculo unitario y todas las raíces de medias móviles fuera de ese círculo. Se dice que y tiene una **raíz unitaria autorregresiva** o simplemente una **raíz unitaria** si una de las p raíces del polinomio en el operador asociado al proceso autorregresivo es 1, en cuyo caso podemos sacar como factor común un polinomio en el operador rezago asociado al proceso autorregresivo, como sigue:

1. Hablaremos de modelar en “diferencias”, contrario a modelar en “niveles”. También emplearemos “diferencias” y “cambios” como sinónimos.

$$\Phi(L) = \Phi'(L)(1-L)$$

en donde $\Phi'(L)$ es de grado $p - 1$. Entonces, en realidad y es un proceso ARMA($p - 1, q$) en diferencias, porque

$$\Phi'(L)(1-L)y_t = \Theta(L)\varepsilon_t$$

es tan solo

$$\Phi'(L)\Delta y_t = \Theta(L)\varepsilon_t$$

Obsérvese que y no es estacionaria en covarianza (una de las raíces del polinomio en su operador autorregresivo rezago asociado al proceso está en el círculo unitario), porque la estacionariedad en covarianza requiere que todas las raíces se encuentren fuera del círculo unitario. Sin embargo, Δy es un proceso ARMA($p - 1, q$) estacionario en covarianza estacionaria e invertible.

Recordemos que en cálculo se puede “deshacer” una integral determinando la derivada. Por analogía, se dice que una serie no estacionaria es integrada si su no estacionariedad se “deshace” al diferenciar. Si sólo se requiere una diferenciación (como con la serie y de arriba), se dice que la serie es integrada de orden 1, o $I(1)$, para abreviar. En forma más general, si se requieren d diferencias, la serie es $I(d)$. El orden de integración es igual a la cantidad de raíces unitarias autorregresivas. En la práctica, los procesos $I(0)$ e $I(1)$ son los casos más importantes; por ello, hemos restringido la descripción anterior para tener en cuenta a lo más una raíz unitaria.²

Con el objeto de adquirir un sentido del comportamiento de los procesos $I(1)$, veamos un ejemplo sencillo y trascendente, la **caminata aleatoria**, que no es más que un proceso AR(1) con un parámetro igual a la unidad,

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

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

La caminata aleatoria no es estacionaria en covarianza, porque el coeficiente AR(1) no es menor que uno. No presenta **reversión a la media**; a diferencia de un AR(1) estacionario, la caminata va hacia arriba y hacia abajo en forma aleatoria, como su nombre lo indica, sin una tendencia a regresar a determinado punto. Aunque el comportamiento de la caminata aleatoria es incierto, su primera diferencia es una serie con buen comportamiento: es un ruido blanco con media cero.

Como ilustración veamos en la figura 10.1.³ una realización de una caminata aleatoria de longitud 300 y también su primera diferencia. La diferencia (cambio), de la caminata aleatoria es un ruido blanco, que vibra al azar. En contraste, el nivel de esa caminata, que es la suma acumulada de los cambios de los ruidos blancos, varía siempre y sin rumbo fijo.

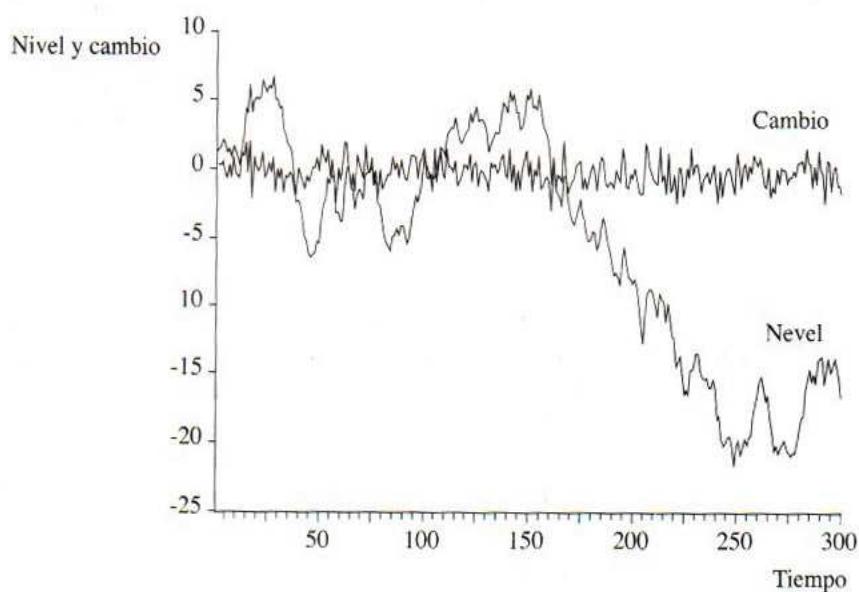
Ahora veamos una caminata aleatoria con deriva,

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

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

2. A veces se presentan series $I(2)$ y son extremadamente raros los órdenes de integración mayores que 2.

3. Esta caminata aleatoria fue simulada en una computadora con $y_1 = 1$ e innovaciones $N(0,1)$.

FIGURA 10.1 Caminata aleatoria: nivel y cambio.

Obsérvese que la caminata aleatoria con deriva es un modelo de tendencia, porque en promedio crece cada periodo el valor de la deriva, δ . Así, el parámetro de la deriva desempeña el mismo papel que el de la pendiente en nuestro modelo anterior de tendencia lineal determinística. A la caminata aleatoria con deriva se le llama modelo de **tendencia estocástica**, porque la tendencia está impulsada por choques estocásticos, en contraste con las **tendencias deterministas** descritas en el capítulo 4.

De la misma manera en que la caminata aleatoria no tiene un nivel particular hacia el que regrese, también la caminata aleatoria con deriva tampoco tiene esa tendencia particular. Por ejemplo, si un choque hace bajar el valor de una caminata aleatoria, no existe una tendencia que necesariamente la haga subir; se espera que permanezca baja en forma permanente. De igual manera, si un choque mueve el valor de una caminata aleatoria con deriva por debajo de la tendencia proyectada actual, no hay forma de que a regrese a ella; la tendencia comienza otra vez a partir del nuevo lugar de la serie. Por ello, los choques a las caminatas aleatorias tienen efectos totalmente permanentes. Un choque unitario mueve para siempre la trayectoria futura esperada de la serie en una unidad, sin la presencia de la deriva.

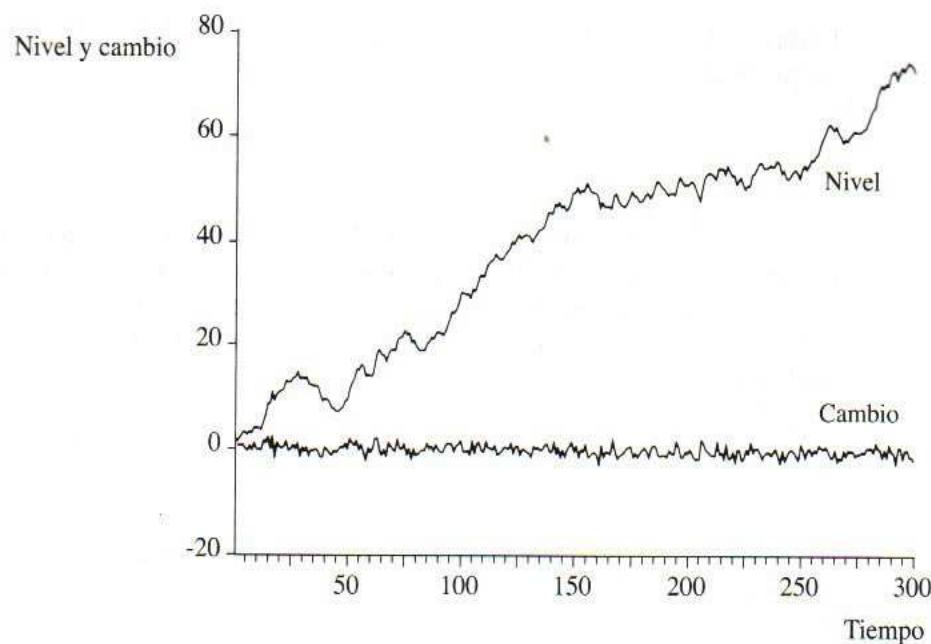
Por ejemplo, en la figura 10.2 mostramos una realización de una caminata aleatoria con deriva, en niveles y en diferencias (cambios). Como antes, el tamaño de la muestra es de 300 y $y_1 = 1$. Las innovaciones son ruido blanco $N(0,1)$ y la deriva es $\delta = 0.3$ por periodo, de modo que las diferencias son ruidos blancos con un promedio igual a 0.3. Resulta difícil percibir el promedio distinto de cero en la diferencia debido a la tendencia estocástica en el nivel, que es la suma acumulada de ruidos blancos $N(0.3, 1)$, donde domina la escala.

Ahora estudiaremos con detalle las propiedades de las caminatas aleatorias. La caminata aleatoria es

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

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

Si suponemos que el proceso se inició en el tiempo 0 con el valor y_0 , se puede representar como

FIGURA 10.2 Caminata aleatoria con deriva: nivel y cambio.

$$y_t = y_0 + \sum_{i=1}^t \varepsilon_i$$

De inmediato,

$$E(y_t) = y_0$$

y

$$\text{var}(y_t) = t \sigma^2$$

En particular obsérvese que

$$\lim_{t \rightarrow \infty} \text{var}(y_t) = \infty$$

y la varianza crece en forma continua, en lugar de converger a cierta varianza finita no condicional.

Ahora veamos la caminata aleatoria con deriva. El proceso es

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

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

Al suponer que el proceso comenzó en el tiempo 0 con valor y_0 , se tiene

$$y_t = t\delta + y_0 + \sum_{i=1}^t \varepsilon_i$$

De inmediato,

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

y

$$\lim \text{var}(y_t) = t \sigma^2$$

Entonces, como en el caso de la caminata aleatoria sin deriva, aquella con deriva también tiene la propiedad de que

$$\lim_{t \rightarrow \infty} \text{var}(y_t) = \infty$$

Así como el ruido blanco es el proceso $I(0)$ más sencillo, la caminata aleatoria es el proceso $I(1)$ más simple. Y así como se puede construir el proceso $I(0)$ con dinámica más rica que el ruido blanco si este último se transforma, así también se pueden obtener los procesos $I(1)$ con dinámica más rica que la de la caminata aleatoria si ésta se transforma. De inmediato llegamos al modelo ARIMA($p, 1, q$),

$$\Phi(L)(1-L)y_t = c + \Theta(L)\varepsilon_t$$

o bien,

$$(1-L)y_t = c\Phi^{-1}(1) + \Phi^{-1}(L)\Theta(L)\varepsilon_t$$

en donde

$$\Phi(L) = 1 - \Phi_1 L - \dots - \Phi_p L^p$$

$$\Theta(L) = 1 - \Theta_1 L - \dots - \Theta_q L^q$$

y todas las raíces de ambos polinomios en el operador rezago están fuera del círculo unitario. ARIMA significa autorregresivo, integrado de medias móviles (por sus siglas en inglés). El proceso ARIMA($p, 1, q$) es sólo un proceso ARMA(p, q) estacionario e invertible en primeras diferencias.

Con más generalidad, podemos trabajar con el **modelo ARIMA(p, d, q)**,

$$\Phi(L)(1-L)^d y_t = c + \Theta(L)\varepsilon_t$$

o sea,

$$(1-L)^d y_t = c\Phi^{-1}(1) + \Phi^{-1}(L)\Theta(L)\varepsilon_t$$

en el que

$$\Phi(L) = 1 - \Phi_1 L - \dots - \Phi_p L^p$$

$$\Theta(L) = 1 - \Theta_1 L - \dots - \Theta_q L^q$$

y todas las raíces de ambos polinomios en el operador rezago están fuera del círculo unitario. El proceso ARIMA(p, d, q) es uno ARMA(p, q), estacionario e invertible, diferenciado d veces. En la práctica, $d = 0$ y $d = 1$ son los casos más importantes. Cuando $d = 0$, y es estacionario en covarianza, o $I(0)$ con media $c\Phi^{-1}(1)$. Cuando $d = 1$, y es $I(1)$ con deriva o tendencia lineal estocástica, $c\Phi^{-1}(1)$ por periodo.

Sucede que los procesos ARIMA($p, 1, q$) más complicados se comportan como caminatas aleatorias en ciertos aspectos clave. En primer lugar, los procesos ARIMA($p, 1, q$) se hacen adecuadamente estacionarios al diferenciarse. En segundo lugar, los choques a los procesos

ARIMA ($p, 1, q$) tienen efectos permanentes⁴. En tercer lugar, varianza de esos procesos crece con el tiempo. Las propiedades especiales de la serie $I(1)$, asociadas con el hecho de que las innovaciones tienen efectos permanentes, tienen implicaciones importantes en los pronósticos. En lo que respecta al pronóstico puntual, la permanencia de los choques significa que los pronósticos óptimos, aún a horizontes muy lejanos, no se revierten por completo hacia una media o una tendencia. Y en lo que respecta a los intervalos y densidades de pronóstico del hecho de que la varianza de un proceso $I(1)$ tienda a infinito con el paso del tiempo significa que la incertidumbre asociada con nuestros pronósticos, traducida en la longitud de los intervalos y en la dispersión de la densidad, aumenta sin límite conforme crece el horizonte de pronóstico.⁵ Veámos cómo funciona todo esto en el contexto de una caminata aleatoria sin deriva, que no es más que un proceso AR (1) con coeficiente unitario. Recordamos que para el proceso AR (1),

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

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

el pronóstico óptimo es

$$y_{T+h,T} = \Phi^h y_T$$

Así, en el caso de la caminata aleatoria en que $j = 1$, el pronóstico óptimo simplemente es igual al valor actual, sin depender del horizonte. Esto aclara cómo afecta a los pronósticos la permanencia de los choques a los procesos de caminata aleatoria: cualquier choque hoy mueve hacia arriba o hacia abajo la serie, también mueve hacia arriba o hacia abajo el pronóstico óptimo a cualquier horizonte. En particular, los efectos de los choques no se desvanecen al alargarse el horizonte de pronóstico, porque no hay media hacia la cual se pueda esperar se revierta la serie.

En la figura 10.3 ilustramos las diferencias importantes en pronósticos a partir de modelos de tendencia determinística y estocásticos, para el producto nacional bruto (PNB) per cápita en Estados Unidos. Se muestra para el periodo 1869-1933, seguido por los pronósticos para 1934-1993 según los modelos de mejor ajuste con tendencias determinística y estocástica, hechos en 1933. El modelo de mejor ajuste con tendencia determinista es uno AR(2) en niveles con tendencia lineal y el modelo estocástico de mejor ajuste es uno AR(1) en diferencias (es decir, un ARIMA(1, 1, 0)), con deriva.⁶ Como 1932 y 1933 fueron años de severa recesión, los pronósticos surgen a partir de una posición muy por debajo de la tendencia. El pronóstico basado en el modelo de tendencia determinista se regresa con rapidez a la tendencia, en agudo contraste con el del modelo de tendencia estocástica, que permanece siempre bajo. Sucedé que el pronóstico con el modelo de tendencia determinística es mucho mejor en este caso, como se observa en la figura 10.4, donde se incluye la realización.

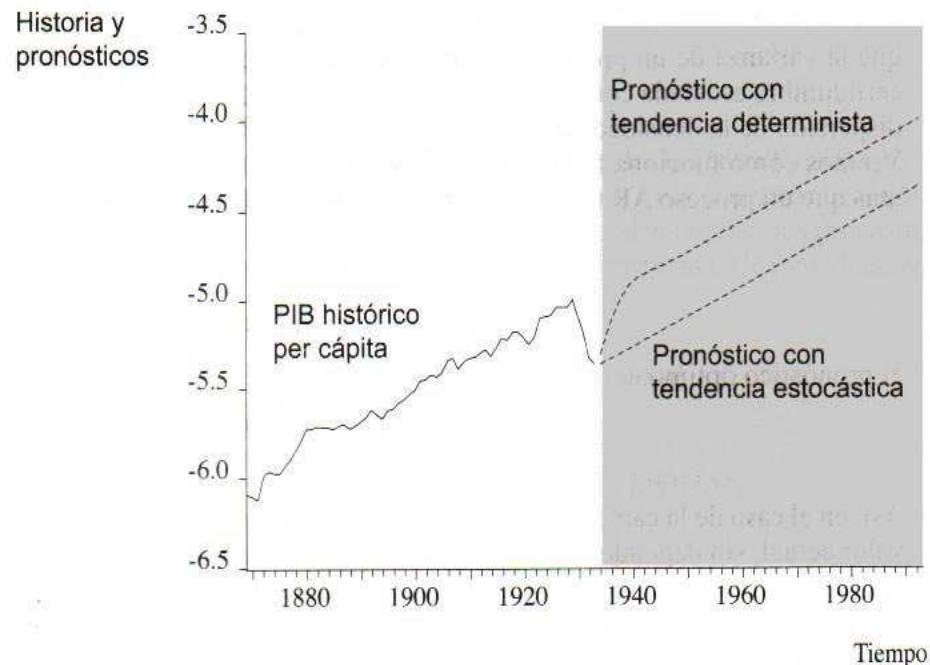
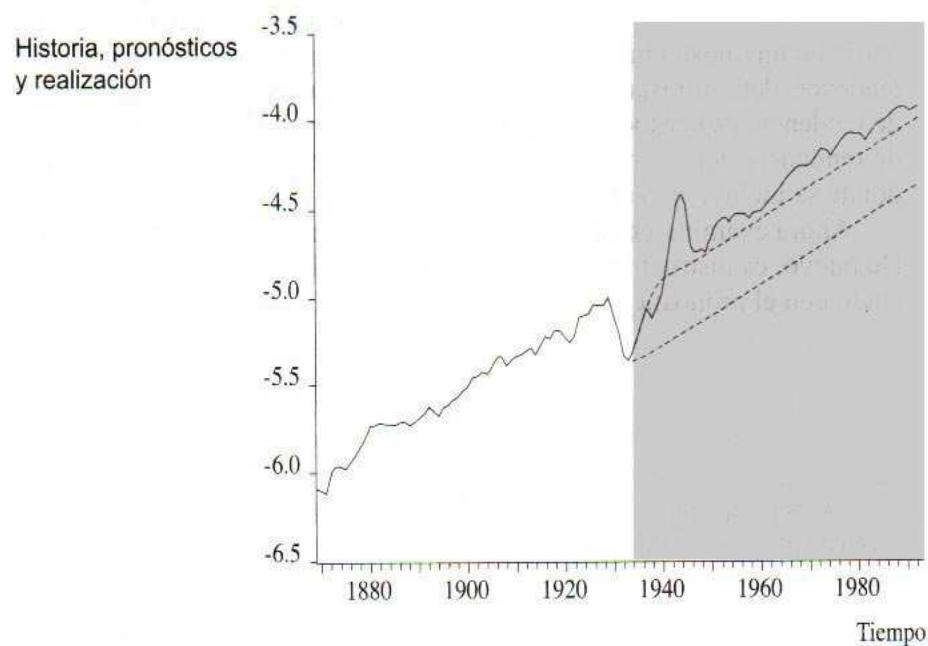
Ahora examinaremos los intervalos y las densidades de pronóstico a partir de modelos $I(1)$. De nuevo, es instructivo recordar la caminata aleatoria sin deriva. Recuérdese que el error asociado con el pronóstico óptimo de un proceso AR(1) es

$$e_{T+h,T} = (y_{T+h,T} - \hat{y}_{T+h,T}) = e_{T+h} + \varphi \varepsilon_{T+h-1} + \dots + \varphi^{h-1} \varepsilon_{T+1}$$

4. Sin embargo, a diferencia de las caminatas aleatorias, el efecto a largo plazo de un choque unitario en un proceso ARIMA($p, 1, q$) puede ser mayor o menor que la unidad, dependiendo de los parámetros de ese proceso.

5. Esto es cierto aun en el caso de no considerar la incertidumbre en la estimación de los parámetros.

6. Obsérvese bien que las dos líneas punteadas de la figura 10.3 son dos pronósticos puntuales distintos y no un intervalo de pronóstico.

FIGURA 10.3 PNB per cápita en Estados Unidos: historia y dos pronósticos.**FIGURA 10.4** PNB per cápita en Estados Unidos: historia, dos pronósticos y realización.

cuya varianza es

$$\sigma_h^2 = \sigma^2 \sum_{i=0}^{h-1} \varphi^{2i}$$

Así, en el caso de la caminata aleatoria, el error es la suma de h innovaciones de ruido blanco,

$$\epsilon_{T+h,T} = \sum_{i=0}^{h-1} \epsilon_{T+h-i}$$

cuya varianza es $h\sigma^2$. La varianza del error de pronóstico es proporcional a h y por consiguiente, crece sin límite al aumentar h . Un intervalo de pronóstico a h etapas al 95% para cualquier horizonte en el futuro es entonces, $y_T \pm 1.96\sigma\sqrt{h}$, y una densidad de pronóstico de densidad a h etapas es $N(y_T, h\sigma^2)$.

Hasta ahora hemos ilustrado con claridad la construcción de pronósticos puntuales y de intervalos y densidades para una caminata aleatoria sin deriva. Los pronósticos para modelos I (1) más complicados se elaboran en forma parecida. Los pronósticos puntuales en niveles de los procesos ARIMA($p, 1, q$), por ejemplo, se obtienen reconociendo que esos son procesos ARMA en diferencias, y sabemos cómo pronosticar procesos ARMA. En consecuencia, se pronostican los cambios, se acumulan los pronósticos de los cambios y se suman al nivel actual para producir

$$\hat{y}_{T+h,T} = y_T + (\Delta y)_{T+1,T} + \dots + (\Delta y)_{T+h,T}$$

2. RAÍCES UNITARIAS: ESTIMACIÓN Y PRUEBA

Regresión de mínimos cuadrados con raíces unitarias

Las propiedades de los estimadores de mínimos cuadrados en modelos con raíces unitarias resultan interesantes para nosotros porque tienen implicaciones en los pronósticos. Como ejemplo usaremos una caminata aleatoria, pero los resultados se aplican a los procesos ARIMA($p, 1, q$) en general. Supongamos que y es una caminata aleatoria tal que

$$y_t = y_t + \epsilon_t$$

pero no sabemos que el coeficiente autorregresivo es 1, así que estimamos el modelo AR(1),

$$y_t = \Phi y_{t-1} + \epsilon_t$$

Existen dos propiedades clave y contradictorias del estimador de mínimos cuadrados: la superconsistencia y el sesgo.

Primero explicaremos la superconsistencia. En el caso de la raíz unitaria $\varphi = 1$, la diferencia entre el estimador mínimo cuadrático $\hat{\varphi}_{LS}$ y 1 disminuye con rapidez a medida que aumenta el tamaño muestral; de hecho, se contrae como $1/T$. Así, $T(\hat{\varphi}_{LS} - 1)$ converge hacia una variable aleatoria no degenerada. En contraste, en el caso de estacionariedad en covarianza de $|\varphi| < 1$, la diferencia entre $\hat{\varphi}_{LS}$ y φ disminuye con más lentitud, como $1/\sqrt{T}$, de modo que $\sqrt{T}(\hat{\varphi}_{LS} - \varphi)$ converge hacia una variable aleatoria no degenerada. A la convergencia extrarrápida en el caso de la raíz

* N.T.: El subíndice corresponde a las siglas en inglés del método de mínimos cuadrados (Least Square: LS).

unitaria se le llama **superconsistencia** y se dice que el estimador de mínimos cuadrados de una raíz unitaria es superconsistente.

En el caso del sesgo, se puede demostrar que el estimador de mínimos cuadrados, $\hat{\phi}_{LS}$, tiene sesgo hacia abajo, de modo que si el valor real de ϕ es ϕ^* , el valor esperado de $\hat{\phi}_{LS}$ es menor que ϕ^* .⁷ En igualdad de circunstancias, cuanto mayor es el valor real de ϕ , mayor es el sesgo, así que el sesgo es mayor en el caso de raíz unitaria. El sesgo también crece si en la regresión se incluye una ordenada al origen y todavía más, si se incluye una tendencia. El sesgo disminuye a medida que aumenta el tamaño de la muestra, porque el estimador converge hacia el verdadero valor poblacional, pero en muestras del tamaño que nos concierne, el sesgo puede ser apreciable.

La superconsistencia y el sesgo tienen efectos opuestos en lo que respecta a los pronósticos. La superconsistencia es útil, significa que la incertidumbre en la estimación de los parámetros disminuye con frecuente rapidez, conforme crece el tamaño de la muestra. En contraste, el sesgo es perjudicial porque los estimadores muy sesgados pueden conducir a pronósticos malos. La superconsistencia asociada con las raíces unitarias garantiza que el sesgo disminuye con rapidez al aumentar el tamaño de la muestra, sin embargo puede tener mucha importancia en muestras pequeñas.

Efectos de las raíces unitarias sobre las funciones de autocorrelación y de autocorrelación parcial muestrales

Si una serie tiene una raíz unitaria, su función de autocorrelación poblacional no está bien definida, porque su varianza es infinita. Sin embargo, la función autocorrelación *muestral* puede calcularse mecánicamente en la forma normal, porque los programas de cómputo no saben o no les importa si los datos que se les alimenta tienen raíz unitaria. La función de autocorrelación muestral tenderá a amortiguar con extrema lentitud; hablando sin mucho rigor, se dice que no se amortigua. La razón es que, al igual que la caminata aleatoria no regresa a determinada media poblacional, cualquier trayecto dado de una muestra tenderá a errar por arriba y por debajo de su media muestral durante largo tiempo, originando autocorrelaciones muestrales positivas muy grandes, incluso para órdenes grandes. En contraste, la función de autocorrelación parcial muestral de un proceso de raíz unitaria tenderá a ser muy grande y cercano a uno en el rezago 1, después disminuirá y decaerá con rapidez.

Si las propiedades de las autocorrelaciones y de las autocorrelaciones parciales muestrales de procesos con raíz unitaria parecen bastante extrañas, las de las autocorrelaciones y autocorrelaciones parciales de las *diferencias* de los procesos de raíz unitaria son mucho más familiares. Esto se debe a que, por definición, la primera diferencia de un proceso con una raíz unitaria, es estacionaria en covarianza estacionaria e invertible.

En las figuras 10.5 y 10.6 se presentan las propiedades de las autocorrelaciones y autocorrelaciones parciales en niveles y en diferencias de procesos con raíz unitaria. En la figura 10.5 se ve el correlograma de nuestra caminata aleatoria simulada. Las autocorrelaciones muestrales no se amortiguan y la autocorrelación parcial muestral es gigantesca en el rezago 1, pero después son diminutas. En la figura 10.6 se ve el correlograma de las primeras diferencias de la caminata aleatoria. Todas las autocorrelaciones y las autocorrelaciones parciales muestrales casi no difieren de cero, como era de esperarse, porque la primera diferencia de una caminata aleatoria es ruido blanco.

7. El sesgo en el estimador de mínimos cuadrados para los casos de raíz unitaria y raíz casi unitaria fue estudiado por Dickey (1976) y Fuller (1976) y a veces se le llama el sesgo de Dickey-Fuller.

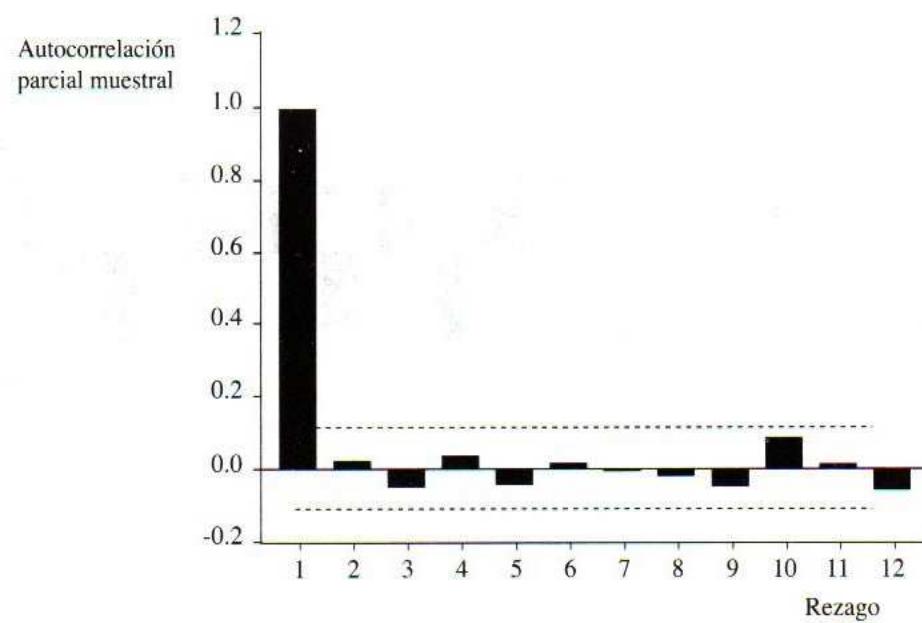
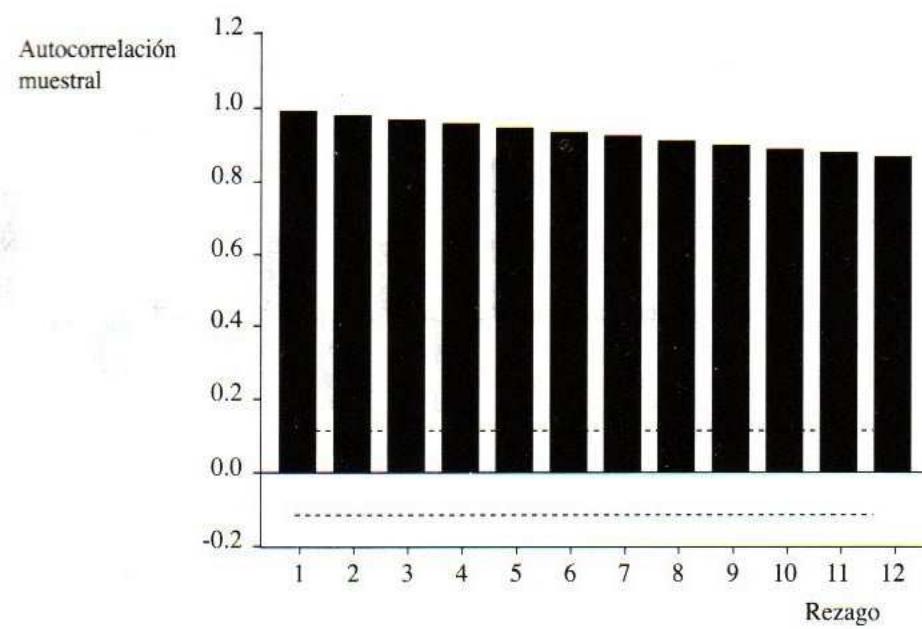
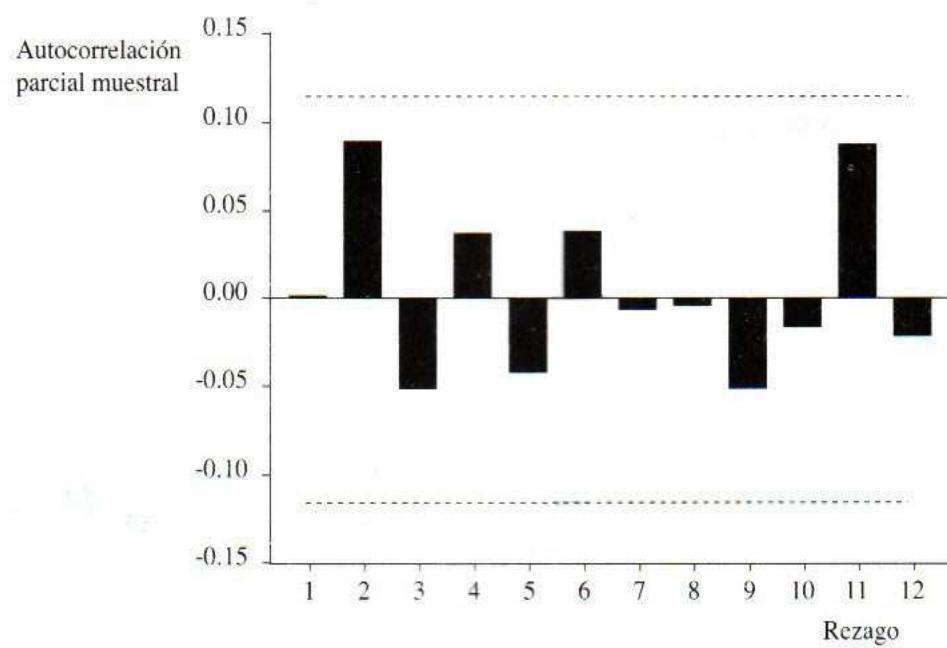
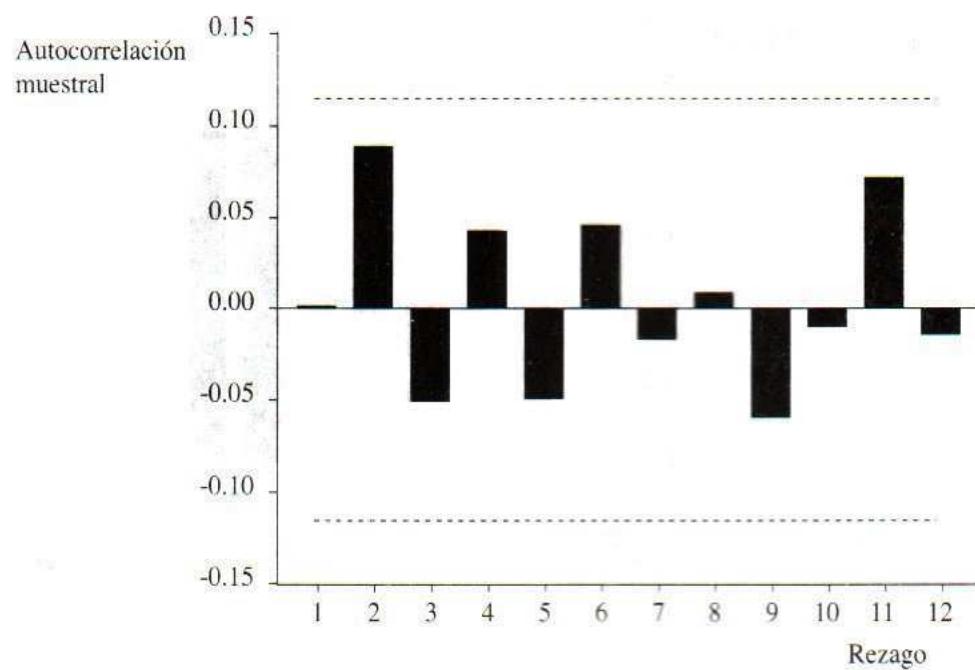
FIGURA 10.5 Caminata aleatoria en niveles: correlogramas.

FIGURA 10.6 Caminata aleatoria en primeras diferencias: correlograma.

Pruebas de raíz unitaria

A la luz de las propiedades especiales de las series con raíces unitarias, a veces resulta interesante investigar su presencia, considerando que es conveniente imponerlas, diferenciando los datos, si parecen estar presentes. Comencemos con el proceso simple AR(1),

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

$$\varepsilon_t \stackrel{\text{idd}}{\sim} N(0, \sigma^2)$$

Podemos hacer la regresión de y_t sobre y_{t-1} , y después aplicar la prueba t normal para investigar si $\varphi = 1$,

$$\tau = \frac{\varphi - 1}{s / \sqrt{\sum_{t=2}^T y_{t-1}^2}}$$

en la que s es el error estándar de la regresión. Obsérvese que el estadístico $\hat{\tau}$ no es el estadístico t calculado en forma automática por los programas de regresión; el estadístico t estándar es para probar la hipótesis nula de que un coeficiente es cero, mientras que $\hat{\tau}$ es el estadístico t para un coeficiente unitario. Sin embargo, con un sencillo truco se logra que los programas normales impriman τ en forma automática. Tan solo se reordena la autorregresión de primer orden como sigue:

$$y_t - y_{t-1} = (\varphi - 1) y_{t-1} + \varepsilon_t$$

Así, $\hat{\tau}$ es el estadístico τ estándar en una regresión de la primera diferencia de y sobre el primer retraso de y .

Un resultado clave es que en el caso de la raíz unitaria, $\hat{\tau}$ no tiene la distribución t , sino una distribución especial, llamada **distribución de Dickey-Fuller**, en honor de los dos estadísticos que la estudiaron con amplitud en las décadas de 1970 y 1980. Fuller (1976), presenta tablas de los puntos porcentuales de la distribución, el estadístico de Dickey-Fuller bajo la hipótesis nula de raíz unitaria: $\hat{\tau}$. Como sólo tomamos en cuenta raíces sobre o fuera del círculo unitario, la prueba unilateral es adecuada.

Con lo anterior hemos mostrado cómo probar la hipótesis nula de una caminata aleatoria sin derivar contra la alternativa de un proceso AR(1) con media cero y estacionario en covarianza. Ahora veamos el caso para una **media m distinta de cero bajo la hipótesis alternativa**, que tiene gran importancia porque los datos en los negocios y la economía casi nunca pueden considerarse con media cero. Bajo la hipótesis alternativa, el proceso AR(1) estacionario en covarianza se convierte en un proceso en *desviaciones respecto a la media*.

$$(y_t - \mu) = \varphi(y_{t-1} - \mu) + \varepsilon_t$$

que podemos reordenar como sigue:

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

en donde $\alpha = \mu(1 - \varphi)$. Si conocieramos μ , podríamos tan solo centrar los datos y proceder como antes. Naturalmente, en la práctica se debe estimar μ junto con los demás parámetros. Aunque a

no aparece bajo la hipótesis nula de raíz unitaria, $\varphi = 1$, está presente bajo la hipótesis alternativa, de modo que se debe incluir una ordenada al origen en la regresión. La distribución del estadístico de Dickey-Fuller correspondiente, $\hat{\tau}_\mu$, se ha tabulado bajo la hipótesis nula de $(\alpha, \varphi) = (0, 1)$, las tablas se pueden encontrar en el libro de Fuller (1976).

Por último, **permitamos la tendencia lineal determinista bajo la hipótesis alternativa** escribiendo el proceso AR(1) en forma de desviaciones respecto a una tendencia lineal,

$$(y_t - a - b \text{TIEMPO}_t) = \varphi(y_{t-1} - a - b \text{TIEMPO}_{t-1}) + \varepsilon_t$$

o sea,

$$y_t = \alpha + \beta \text{TIEMPO}_t + \varphi y_{t-1} + \varepsilon_t$$

en la que $\alpha = a(1 - \varphi) + b\varphi$ y $\beta = b(1 - \varphi)$. Bajo la hipótesis de raíz unitaria, $\varphi = 1$, se tiene una caminata aleatoria con deriva,

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

lo cual a su vez, es una tendencia estocástica, pero bajo la hipótesis alternativa de tendencia determinista la ordenada al origen y la tendencia se deben incluir en la regresión. La caminata aleatoria con deriva es una hipótesis nula que con frecuencia surge en las aplicaciones económicas; las desviaciones estacionarias respecto a la tendencia lineal son una alternativa natural. Fuller (1976) tabuló la distribución del estadístico $\hat{\tau}_t$ de Dickey-Fuller, que toma en cuenta la tendencia lineal bajo la hipótesis alternativa.

Ahora generalizaremos la prueba para una dinámica autorregresiva de orden superior. En el proceso AR(p),

$$y_t = \sum_{j=1}^p \varphi_j y_{t-j} + \varepsilon_t$$

el cual se puede escribir como

$$y_t = \rho_1 y_{t-1} + \sum_{j=2}^p \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t$$

en la que $p > 2$, $\rho_1 = -\sum_{j=1}^{p-1} \varphi_j$ y $\rho_i = \sum_{j=i}^p \varphi_j$, $i = 2, \dots, p$. Si hay una raíz unitaria, entonces $\rho_1 = 1$ y y es tan solo un proceso AR($p - 1$) en primeras diferencias. El estadístico de Dickey-Fuller para la hipótesis nula de $\rho_1 = 1$ tiene la misma distribución asintótica que la de $\hat{\tau}$. Así, los resultados del proceso AR(1) se generalizan (asintóticamente) en forma directa a procesos de orden superior.

Para permitir una media distinta de cero en el AR(p), se escribe

$$(y_t - \mu) + \sum_{j=1}^p \varphi_j (y_{t-j} - \mu) = \varepsilon_t$$

o bien,

$$y_t = \alpha + \rho_1 y_{t-1} + \sum_{j=2}^p \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t$$

en donde $\alpha = \mu(1 + \sum_{j=1}^{p-1} \varphi_j)$ y los demás parámetros son como se definieron arriba. Bajo la hipótesis nula de raíz unitaria, la ordenada al origen se anula, porque en ese caso $\sum_{j=1}^{p-1} \varphi_j = -1$. La distribu-

ción del estadístico de Dickey-Fuller para probar $\rho_1 = 1$ en esta regresión es asintóticamente idéntica a la de t

Por último, a fin de considerar una tendencia lineal bajo la hipótesis alternativa, se escribe

$$(y_t - a - b \text{TIEMPO}_t) + \sum_{j=2}^p \varphi_j (y_{t-j} - a - b \text{TIEMPO}_{t-j}) = \varepsilon_t$$

que se reordena como

$$y_t = k_1 + k_2 \text{TIEMPO}_t + \rho_1 y_{t-1} + \sum_{j=2}^p \rho_j (y_{t-j+1} - y_{t-j}) + \varepsilon_t$$

siendo

$$k_1 = a \left(1 + \sum_{i=1}^p \varphi_i \right) - b \sum_{i=1}^p i \varphi_i$$

y

$$k_2 = b \text{TIEMPO}_t \left(1 + \sum_{i=1}^p \varphi_i \right)$$

Bajo la hipótesis nula,

$$k_1 = -b \sum_{i=1}^p i \varphi_i \quad y \quad k_2 = 0$$

El estadístico de Dickey-Fuller para la hipótesis en donde $r1 = 1$, tiene la distribución de t en forma asintótica.

Ahora pasemos a las representaciones ARIMA generales. Hemos visto que la prueba original de Dickey-Fuller para una raíz unitaria en modelos AR(1) se generaliza con facilidad con el fin de probar una raíz unitaria en el caso AR(p), $r < \infty$, tan solo aumentando la regresión en la prueba, con rezagos de las primeras diferencias esto es la llamada prueba aumentada de Dickey-Fuller o regresión aumentada de Dickey-Fuller. Sin embargo, la situación se complica más en el caso de ARMA(p, q), porque la autorregresión correspondiente es de orden infinito. Se han sugerido varias pruebas, y de ellas la más común es aproximar la autorregresión infinita con una regresión aumentada de Dickey-Fuller de orden finito. Se deja que la cantidad de rezagos aumente a medida que lo hace el tamaño de la muestra, pero que lo haga con menor rapidez. Hall (1994), muestra que bajo ciertas condiciones, la distribución asintótica nula del estadístico de Dickey-Fuller con el aumento en el orden del rezago seleccionado a través del criterio de información de Schwarz, es igual a que si se conociera el orden real, de modo que ese criterio proporciona una guía útil para seleccionar el orden del rezago en la regresión aumentada de Dickey-Fuller. No obstante, Ng y Perron (1995), argumentan que la prueba t estándar proporciona una referencia más confiable. Se necesita investigar más, pero parece que, a diferencia de cuando se seleccionan órdenes de rezago para modelos de pronóstico, podría ser mejor usar penalizaciones menos fuertes para los grados de libertad, como las asociadas con la prueba t o con el criterio de información de Akaike, que seleccionar aumentos en el orden del rezago en las regresiones de Dickey-Fuller.

Bajo la hipótesis alternativa de media cero, media distinta de cero o bien, a tendencia lineal, se puede escribir

$$y_t = \rho_1 y_{t-1} + \sum_{j=2}^{k-1} \rho_j (y_{t-j+1} - y_{t-j}) + \epsilon_t$$

$$y_t = \alpha + \rho_1 y_{t-1} + \sum_{j=2}^{k-1} \rho_j (y_{t-j+1} - y_{t-j}) + \epsilon_t$$

o bien,

$$y_t = k_1 + k_2 \text{TIEMPO}_t + \rho_1 y_{t-1} + \sum_{j=2}^{k-1} \rho_j (y_{t-j+1} - y_{t-j}) + \epsilon_t$$

en donde se han incluido $k - 1$ rezagos. El estadístico de Dickey-Fuller sobre y_{t-1} continúa teniendo las distribuciones asintóticas $\hat{\tau}$, $\hat{\tau}_{\text{m}}$ y $\hat{\tau}_{\text{tt}}$ bajo la hipótesis nula de raíz unitaria. Para seleccionar la cantidad $k - 1$ de rezagos se puede aplicar el criterio de información de Schwarz o el de Akaike, así como el estadístico t , a los diversos rezagos de Dy , que tienen la distribución normal estándar en muestras grandes, independientemente de si la hipótesis de raíz unitaria es verdadera o falsa.

Recientemente se han propuesto otras pruebas, que en ciertos casos han tenido mejor potencia que la de Dickey-Fuller.⁸ Siempre sucede que las pruebas de raíz unitaria están plagadas de problemas de potencia y tamaño; los de potencia se deben a que las hipótesis alternativas relevantes son por lo regular, muy cercanas a la hipótesis nula, y los de tamaño tienen su origen en que deben incluirse infinitos rezagos, cosa que no es posible en la práctica.

Así, aunque a veces son útiles las pruebas de raíz unitaria, no son definitivas y por tanto son el final de la historia respecto a la decisión de si especificar o no modelos en niveles o en diferencias. Por ejemplo, el no poder rechazar una raíz unitaria no necesariamente significa que se deba imponer; es probable que sea baja la potencia de las pruebas de raíz unitaria, cuando las hipótesis alternativas son cercanas a la hipótesis nula, que son las opciones relevantes. Por otro lado, a veces será deseable imponer una raíz unitaria aun cuando la raíz real sea menor que uno, pero muy próxima a uno, porque el sesgo de Dickey-Fuller afecta a la estimación en niveles. Necesitamos usar introspección y teoría, además de pruebas formales, para guiar la difícil decisión de imponer raíces unitarias y comparar la eficiencia del pronóstico con distintos modelos con y sin raíces unitarias.

Una cuestión clave para los pronosticadores es la determinación de los costos comparativos por especificar mal los modelos de pronóstico, en niveles en lugar de en diferencias, como función del tamaño de muestra del horizonte de pronóstico del valor real de la raíz dominante, etc. En relación con esto necesitamos aprender más acerca de la eficacia del pronóstico cuando se usan reglas como la de "imponer una raíz unitaria a menos que la prueba de Dickey-Fuller se rechace al nivel de 5%".⁹ Campbell y Perron (1991), avanzan en esa dirección, pero todavía falta mucho por hacer.

En ciertos aspectos, lo más relevante de la teoría de la raíz unitaria para pronosticar se refiere a la estimación y no a la prueba. Es importante que los pronosticadores entiendan los efectos de

8. Véanse Elliot, Rothenberg y Stock (1996), Dickey y Gonzalez-Farias (1992) y las comparaciones en Pantula, Gonzalez-Farias y Fuller (1994).

9. De nueva cuenta vemos que aparece el principio de contracción: las reglas que imponen a la raíz unitaria a menos que haya fuerte evidencia en su contra, pueden conducir a mejores modelos de pronóstico que los modelos no restringidos, aun cuando la restricción de raíz unitaria sea falsa. Y claro, si la restricción es cierta, es útil imponerla, y más a la luz del sesgo de Dickey-Fuller, que es un grave problema en la estimación en niveles.

las raíces unitarias sobre la consistencia y el sesgo en muestras pequeñas. Su comprensión, por ejemplo, conduce al punto de vista de que al menos *asintóticamente*, es mejor estimar modelos de pronóstico en niveles con tendencias incluidas, porque entonces se obtiene una aproximación muy buena a la dinámica de los datos, independientemente del estado real del mundo, con o sin raíz unitaria. Si no hay raíz unitaria, es claro que conviene trabajar en niveles y si la hay, la máxima raíz estimada convergerá apropiadamente y con gran rapidez a la unidad. Por otro lado, es adecuado diferenciar sólo en el caso de la raíz unitaria; si la diferenciación es inadecuada puede ser peligrosa, incluso asintóticamente.

3. APPLICACIÓN: MODELADO Y PRONÓSTICO DE LA TASA DEL TIPO DE CAMBIO YEN/DÓLAR

A fin de aplicar e ilustrar lo que hemos aprendido, vamos a modelar y pronosticar la tasa del tipo de cambio yen/dólar. Por comodidad, llamaremos y a la serie yen/dólar, $\ln y$ al logaritmo del nivel y y $\Delta \ln y$ al cambio del logaritmo de nivel. Contamos con datos de fin de mes desde 1973.01 hasta 1996.07; graficamos $\ln y$ en la parte superior de la figura 10.7 y $\Delta \ln y$ en la inferior.¹⁰ La gráfica de $\ln y$ se ve muy persistente, quizás tenga una raíz unitaria. Contrariamente, $\Delta \ln y$ se ve bastante estacionaria y de hecho muy cercana al ruido blanco. Las figuras 10.8 y 10.9, que muestran los correlogramas $\ln y$, de $\Delta \ln y$, confirman lo percibido en las gráficas. Las autocorrelaciones muestrales de $\ln y$ son todas muy grandes y no se amortiguan, y la primera autocorrelación parcial muestral es muy grande, mientras que todas las demás difieren muy poco de cero. Sin embargo, el correlograma de $\Delta \ln y$ se ve muy distinto. Las funciones de autocorrelación y de autocorrelación parcial muestrales se amortiguan con rapidez; de hecho, después del rezago 1 todas son muy cercanas de cero. Todo lo anterior sugiere que $\ln y$ es $I(1)$.

Ahora ajustaremos modelos de pronóstico. Basaremos todo el análisis y modelado en $\ln y$, 1994.12 a 1973.01, dejando como reserva 1996.07 a 1995.01 para pronóstico fuera de la muestra. Comenzaremos con el ajuste de modelos de tendencia determinista, haremos la regresión de $\ln y$ sobre una ordenada al origen y una tendencia en el tiempo, tomando en cuenta dinámicas en las perturbaciones hasta del proceso ARMA(3,3). En las tablas 10.1 y 10.2 se muestran los valores de los criterios de Akaike y de Schwarz para todas las combinaciones ARMA(p, q). El criterio de Akaike selecciona el modelo ARMA(3,1), mientras que el de Schwarz, el AR(2). Procederemos con el modelo más parsimonioso, el seleccionado con el criterio de Schwarz. Los resultados de la estimación se presentan en la tabla 10.3 y la gráfica de residuales en la figura 10.10. Obsérvese que la raíz inversa dominante está muy cercana a 1 (0.96), mientras que la segunda raíz inversa es positiva, pero mucho menor (0.35).

Los pronósticos fuera de muestra aparecen en las figuras 10.11 a 10.13. La figura 10.11 muestra la historia de 1990.01 a 1994.12, y los pronósticos puntuales y el intervalo de pronóstico de 1995.01 a 1996.07. Aunque la dinámica estimada que persiste implica una reversión muy lenta a la tendencia, sucede que los valores al final de la muestra del $\ln y$ en 1994, se aproximan a la tendencia estimada. Así, con muy buena aproximación, el pronóstico tan solo extraña la tendencia ajustada. En la figura 10.12 se muestra la historia junto con un pronóstico de horizonte a largo plazo (hasta 2020.12), con el fin de ilustrar que los intervalos de confianza se aplanan,

10. A partir de aquí trabajaremos con el logaritmo de la tasa de tipo de cambio, porque el cambio en el logaritmo tiene la cómoda interpretación de ser una aproximación al cambio porcentual o tasa. Así, cuando mencionemos el nivel de la tasa de tipo de cambio, nos referiremos al logaritmo del nivel ($\ln y$), y cuando mencionemos el cambio, nos referiremos al cambio del logaritmo de la tasa de tipo de cambio ($\Delta \ln y$).

FIGURA 10.7 Logaritmo de la tasa del tipo de cambio (arriba) y cambio en el logaritmo de la tasa del tipo de cambio (abajo).

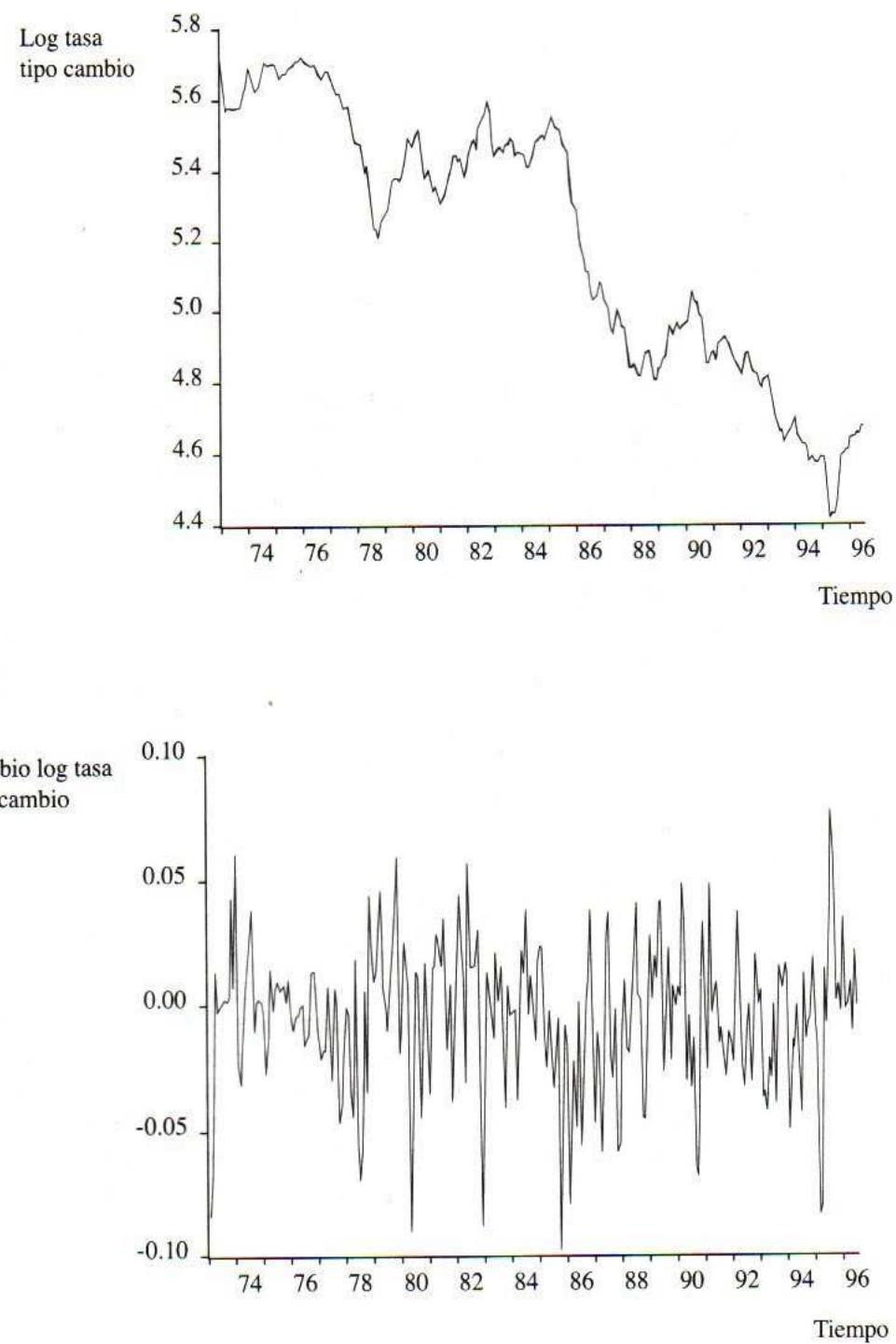


FIGURA 10.8 Logaritmo de la tasa del tipo de cambio. Autocorrelaciones y autocorrelaciones parciales muestrales.

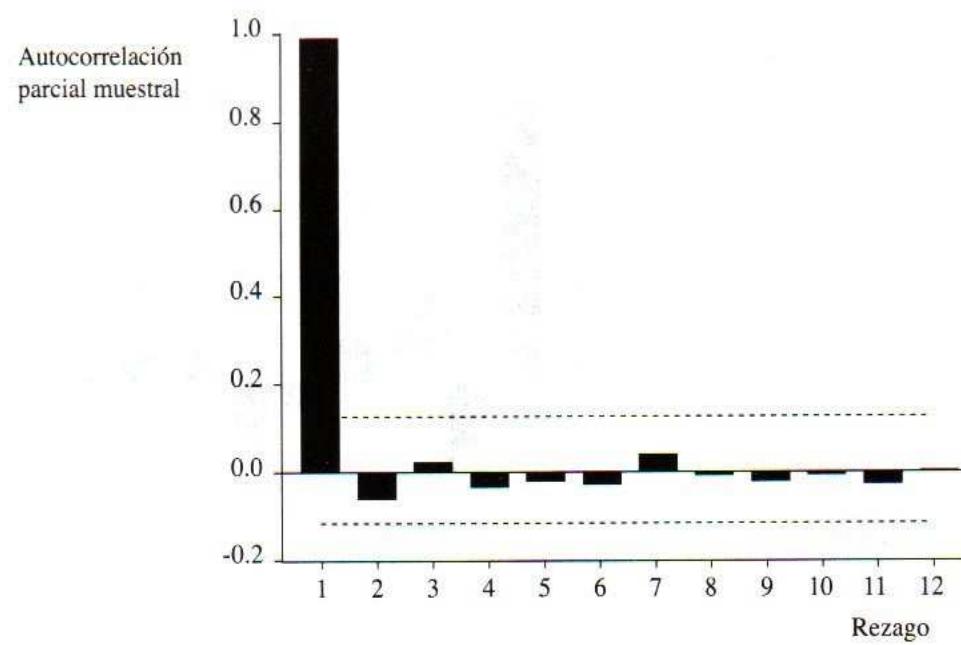
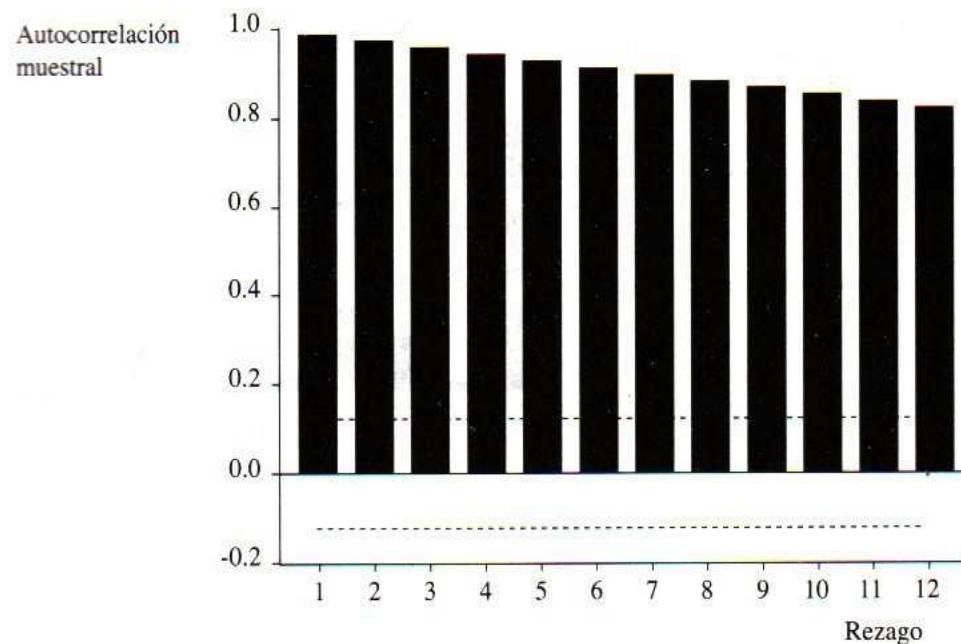


FIGURA 10.9 Primeras diferencias del logaritmo de la tasa del tipo de cambio. Autocorrelaciones y autocorrelaciones parciales muestrales.

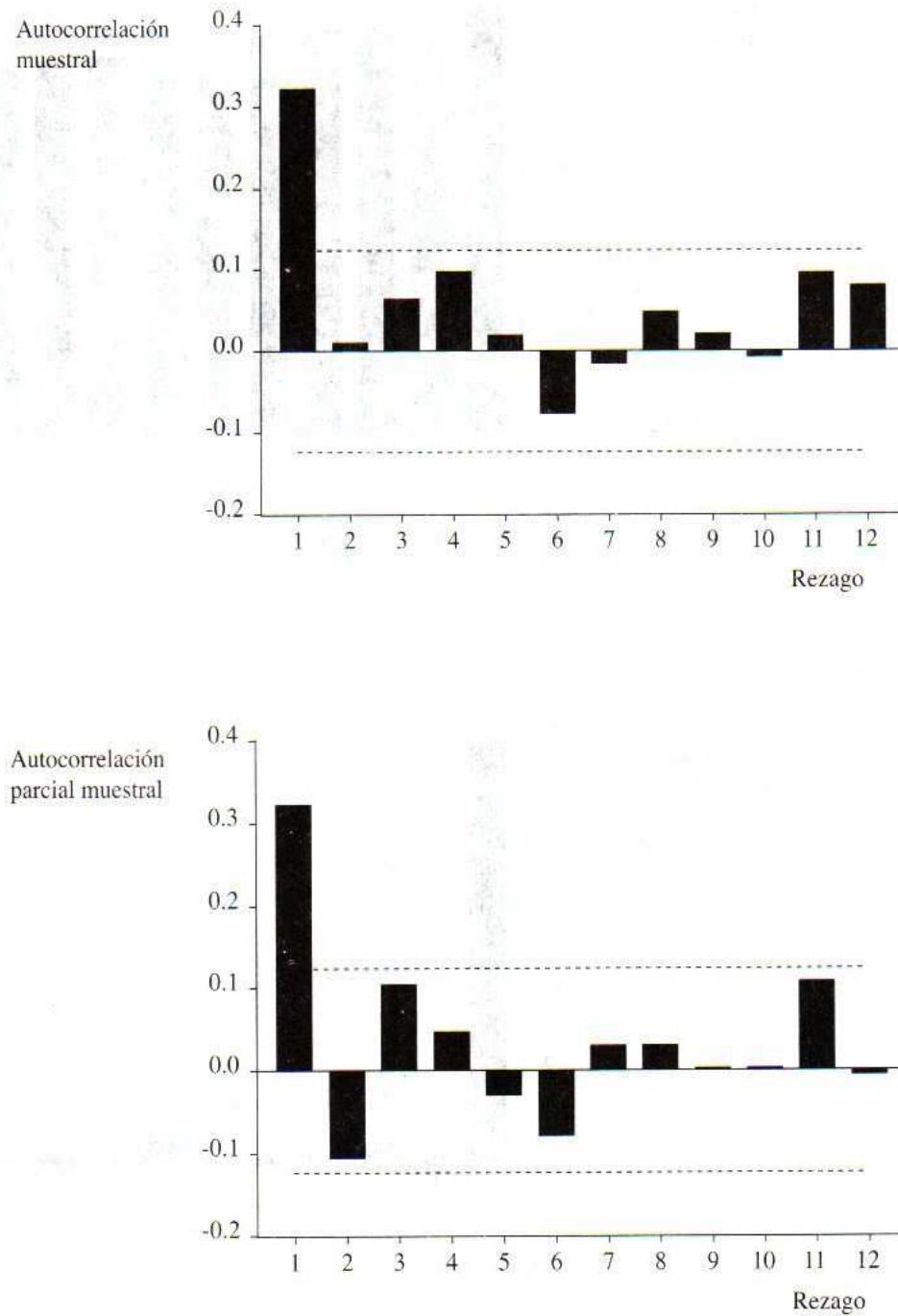


TABLA 10.1 Logaritmo de la tasa del tipo de cambio: valores del CIA para diversos modelos ARMA

		Orden del MA			
		0	1	2	3
Orden del AR	0		-5.171	-5.953	-6.428
	1	-7.171	-7.300	-7.293	-7.287
	2	-7.319	-7.314	-7.320	-7.317
	3	-7.322	-7.323	-7.316	-7.308

TABLA 10.2 Logaritmo de la tasa del tipo de cambio: valores del CIS para diversos modelos ARMA

		Orden del MA			
		0	1	2	3
Orden del AR	0		-5.130	-5.899	-6.360
	1	-7.131	-7.211	-7.225	-7.205
	2	-7.265	-7.246	-7.238	-7.221
	3	-7.253	-7.241	-7.220	-7.199

TABLA 10.3 Logaritmo de la tasa del tipo de cambio: modelo de tendencia determinista con mejor ajuste.

LS//Variable dependiente LYEN

Muestra (ajustada): 1973.03 1994.12

Observaciones incluidas: 262 después de ajustar los extremos

Convergencia alcanzada después de tres iteraciones

Variable	Coeficiente	Error estándar	Estadístico <i>t</i>	Prob.
C	5.904705	0.136665	43.20570	0.0000
TIEMPO	-0.004732	0.000781	-6.057722	0.0000
AR(1)	1.305829	0.05587	22.67561	0.0000
AR(2)	-0.334210	0.057656	-5.796676	0.0000
R cuadrada	0.994468	Media de la variable dependiente		5.253984
R cuadrada ajustada	0.994404	DE de la variable dependiente		0.341563
EE de la regresión	0.025551	Criterio de información de Akaike		-7.319015
Suma del cuadrado de los residuales	0.168435	Criterio de Schwarz		-7.264536
Logaritmo de la verosimilitud	591.0291	Estadístico f		15461.07
Estadístico Durbin-Watson	1.964687	Prob. (estadístico f)		0.000000

Raíces invertidas AR

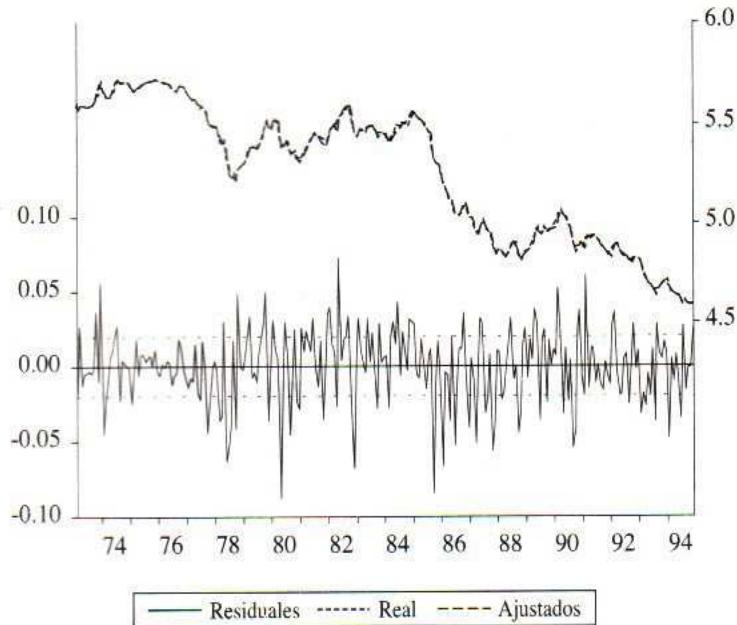
.96

.35

finalmente, a más o menos dos errores estándar. Por último, la figura 10.13 presenta la historia y el pronóstico junto con la realización, que en su mayoría está dentro del intervalo de 95% de confianza.

Con base en lo que sugieren de los correlogramas realizaremos una prueba formal de raíz unitaria tomando en cuenta la tendencia bajo la hipótesis alternativa. En la tabla 10.4 se muestran los resultados con tres rezagos.¹¹ No existe evidencia en contra de la raíz unitaria por consiguiente, trataremos de modelar $\Delta \ln y$. Hacemos la regresión de $\Delta \ln y$ sobre una ordenada al origen y consideraremos la dinámica ARMA(3,3). Los valores de los criterios de Akaike y Schwarz se presentan en las tablas 10.5 y 10.6, respectivamente. El criterio de Akaike selecciona un modelo ARMA(3,2) y el de Schwarz uno AR(1). Obsérvese que los modelos de $\ln y$ y $\Delta \ln y$ seleccionados por el criterio de Schwarz son consistentes entre sí bajo la hipótesis de raíz unitaria: un AR(2) con una raíz unitaria en niveles que equivalen a un AR(1) en diferencias, en contraste con los modelos seleccionados con el criterio de Akaike. Por esta razón y naturalmente, de acuerdo con las consideraciones normales de parsimonia, procederemos con el AR(1) elegido por el criterio de Schwarz. Los resultados de la regresión aparecen en la tabla 10.7 y en la figura 10.14; obsérvese el pequeño, pero significativo, coeficiente de 0.32.¹²

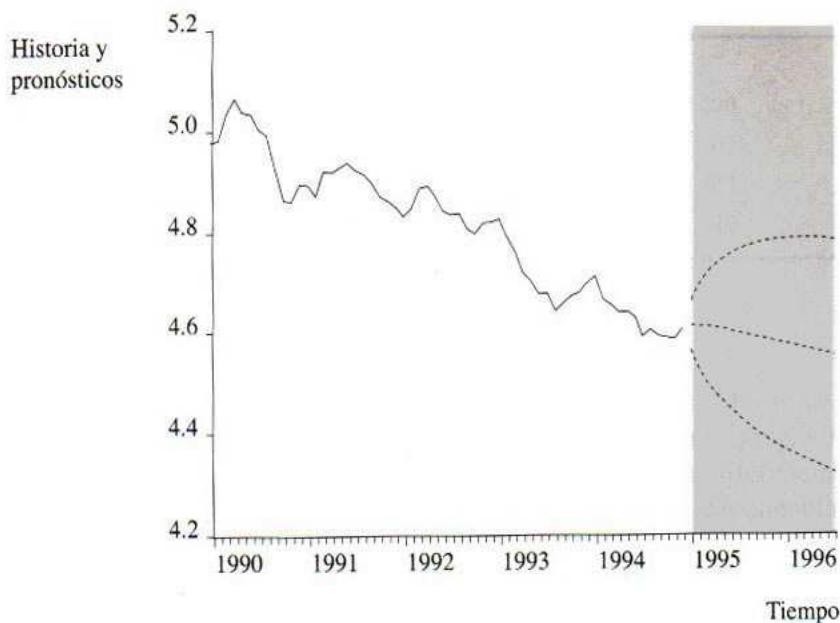
FIGURA 10.10 Logaritmo de la tasa del tipo de cambio: gráfica de los residuales del modelo de tendencia determinista con mejor ajuste.



11. Hemos considerado varios aumentos en el orden de los rezagos, y los resultados fueron consistentes no se puede rechazar la hipótesis de raíz unitaria. El criterio de Schwarz señaló un aumento en el orden del rezago, mientras que el de Akaike y la prueba t indicaron tres aumentos en el orden del rezago.

12. De hecho, el modelo ARMA(3,2) seleccionado por el criterio de información de Akaike es muy cercano a un AR(1), porque las dos raíces estimadas de modelo MA casi se anulan con dos de las estimadas del AR, lo cual dejaría un modelo AR(1).

FIGURA 10.11 Logaritmo de la tasa del tipo de cambio: historia y pronóstico del AR(2) en niveles con tendencia lineal.



Los resultados del pronóstico fuera de la muestra aparecen en las figuras 10.15 a 10.17. En la primera se presenta la historia y el pronóstico. El pronóstico se parece mucho —casi idéntico— al basado en el modelo de tendencia determinista que examinamos antes; lo anterior se debe a que los modelos de tendencia estocástica y de tendencia determinista son de hecho, muy próximos entre sí en este caso; aun cuando no impongamos una raíz unitaria, obtenemos una raíz

FIGURA 10.12 Logaritmo de la tasa del tipo de cambio: historia y pronóstico a largo plazo del AR(2) en niveles con tendencia lineal.

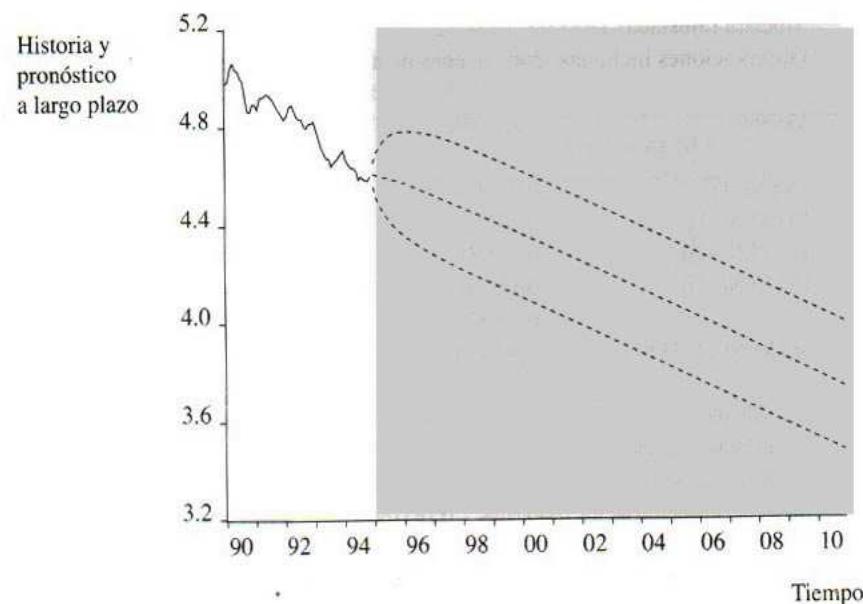


FIGURA 10.13 Logaritmo de la tasa del tipo de cambio: historia, realización y pronóstico del AR(2) en niveles con tendencia lineal.

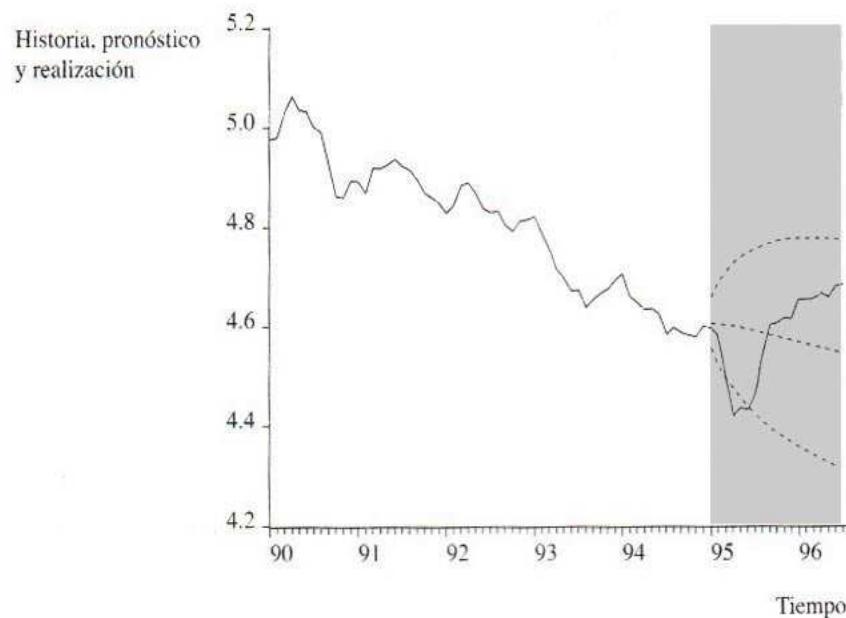


TABLA 10.4 Logaritmo de la tasa del tipo de cambio: prueba aumentada de Dickey-Fuller para raíz unitaria.

Dickey-Fuller aumentada	-2.498863	1% valor crítico	-3.9966
Estadístico de prueba		5% valor crítico	-3.4284
		10% valor critico	-3.1373

Ecuación de prueba aumentada de Dickey-Fuller
 LS//Variable dependiente es D(LYEN)
 Muestra (ajustada) 1973.05 1994.12
 Observaciones incluidas: 260 después de ajustar los extremos

Variable	Coeficiente	Error estándar	Estadístico <i>t</i>	Prob.
LYEN(-1)	-0.029423	0.011775	-2.498863	0.0131
D(LYEN(-1))	0.362319	0.061785	5.864226	0.0000
D(LYEN(-2))	-0.114269	0.064897	-1.760781	0.0795
D(LYEN(-3))	0.118386	0.061020	1.940116	0.0535
C	0.170875	0.068474	2.495486	0.0132
@TREND(1973:01)	-0.000139	5.27E-05	-2.639758	0.0088
R cuadrada	0.142362	Media de la variable dependiente	-0.003749	
R cuadrada ajustada	0.125479	DE de la variable dependiente	0.027103	
EE de la regresión	0.025345	Criterio de información de Akaike	-7.327517	
Suma del cuadrado de los residuales	0.163166	Criterio de Schwarz	-7.245348	
Logaritmo de la verosimilitud	589.6532	Estadístico f	8.432417	
Estadístico Durbin-Watson	2.010829	Prob. (estadístico f)	0.000000	

TABLA 10.5 Diferencia del logaritmo de la tasa del tipo de cambio yen/dólar:
valores del CIA para diversos modelos ARMA.

		Orden del MA				
		0	1	2	3	4
Orden del AR	0		-7.298	-7.290	-7.283	
	1	-7.308	-7.307	-7.307	-7.302	
	2	-7.312	-7.314	-7.307	-7.299	
	3	-7.316	-7.309	-7.340	-7.336	

dominante estimada que se acerca mucho a la unidad. En la figura 10.16 se ve la historia y un pronóstico a largo plazo. Éste revela una diferencia secundaria y una principal entre los pronósticos a partir de modelos de tendencia determinista y estocásticos. La diferencia pequeña es que para después del año 2010, el pronóstico puntual a partir del modelo de tendencia determinista es un poco menor, lo cual refleja el hecho de que la pendiente de la tendencia estimada es algo más negativa para el modelo de tendencia determinista que para el de tendencia estocástica. Sin embargo, en términos estadísticos, los pronósticos puntuales no se distinguen. La gran diferencia concierne a los intervalos de pronóstico: los del modelo de tendencia estocástica se amplían en forma continua a medida que el horizonte se alarga, mientras que los modelos de tendencia determinista no lo hacen. Por último, en la figura 10.17 se muestran la historia y el pronóstico, junto con la realización para 19953.01-1996.07.

Si se compara el AR(2) con la tendencia en niveles (el modelo en niveles seleccionado por el criterio de Schwarz), con y el AR(1) en diferencias (el modelo en diferencias seleccionado por el criterio de Schwarz), parece que el modelo en diferencias se ve favorecido por tener un valor menor del criterio de información de Schwarz. El AR(1) en diferencias sólo se ajusta un poco menos bien que el AR(2) en niveles; recuérdese que el AR(2) en niveles tiene una raíz unitaria cercana a la unidad, y ahorra un grado de libertad.¹³ Además, las consideraciones económicas y

TABLA 10.6 Diferencia del logaritmo de la tasa del tipo de cambio:
valores del CIS para diversos modelos ARMA.

		Orden de MA				
		0	1	2	3	4
Orden de AR	0		-7.270	-7.249	-7.228	
	1	-7.281	-7.266	-7.252	-7.234	
	2	-7.271	-7.259	-7.238	-7.217	
	3	-7.261	-7.241	-7.258	-7.240	

13. Precaución: En cierto sentido, el modelo AR(1) en diferencias puede no ahorrar el grado de libertad, ya que la decisión de imponer una raíz unitaria se basó en una estimación anterior (la prueba de Dickey-Fuller aumentada), lo cual no se tiene en cuenta al determinar el criterio de información de Schwarz para el AR(1) en diferencias.

TABLA 10.7 Diferencia del logaritmo de la tasa del tipo de cambio: modelo de tendencia estocástica con mejor ajuste.

LS/Variable dependiente es DLYEN

Muestra (ajustada): 1973.03 1994.12

Observaciones incluidas: 262 después de ajustar los extremos

Convergencia alcanzada después de tres iteraciones

Variable	Coeficiente	Error estándar	Estadístico <i>t</i>	Prob.
C	-0.003697	0.002350	-1.573440	0.1168
AR(1)	0.321870	0.057767	5.571863	0.0000
R cuadrada	0.106669	Media de la variable dependiente		-0.003888
R cuadrada ajustada	0.103233	DE de la variable dependiente		0.027227
EE de la regresión	0.025784	Criterio de información de Akaike		-7.308418
Suma del cuadrado de los residuales	0.172848	Criterio de Schwarz		-7.281179
Logaritmo de la verosimilitud	587.6409	Estadístico f		31.04566
Estadístico Durbin-Watson	1.948933	Prob. (estadístico f)		0.000000
Raíces invertidas AR	.32			

FIGURA 10.14 Diferencia del logaritmo de la tasa del tipo de cambio: gráfica de los residuales de modelo de tendencia estocástica con mejor ajuste.

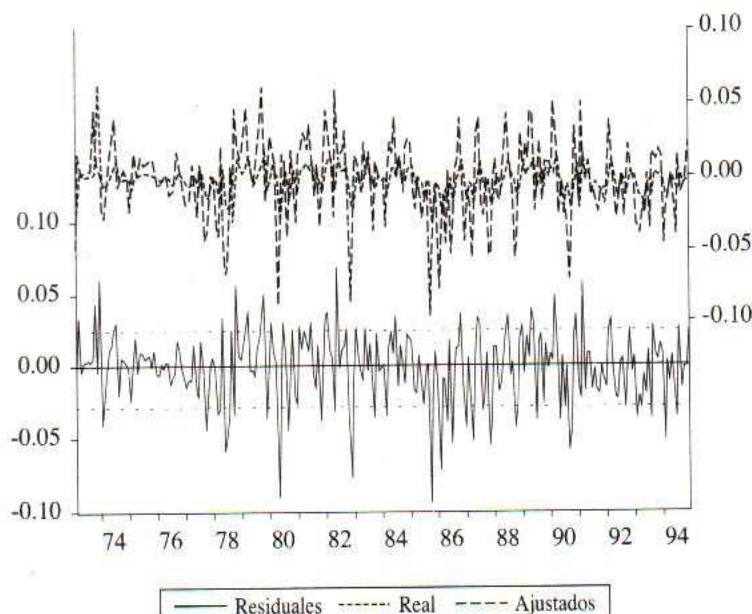


FIGURA 10.15 Logaritmo de la tasa del tipo de cambio: historia y pronóstico del AR(1) en diferencias con ordenada al origen.

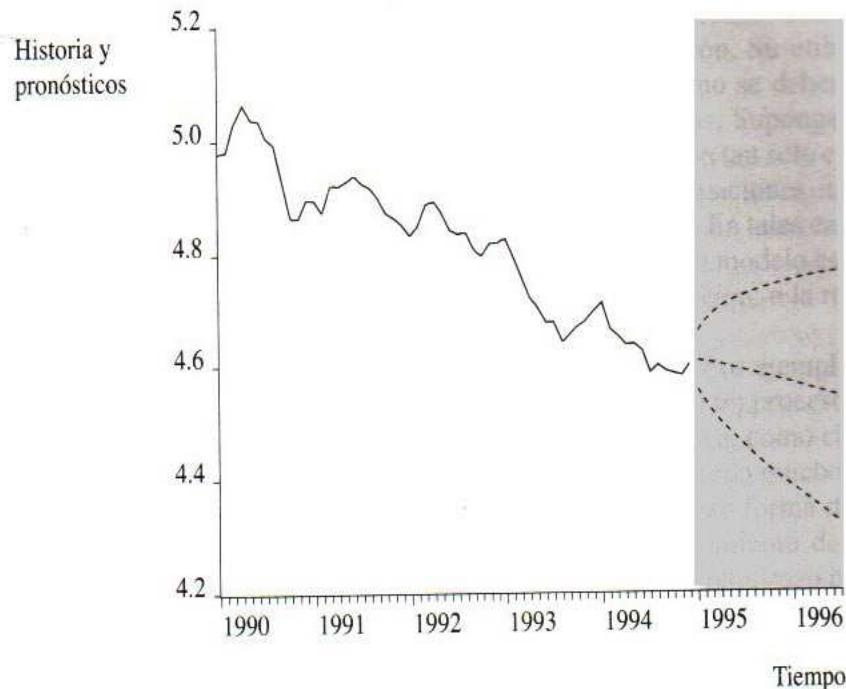


FIGURA 10.16 Logaritmo de la tasa del tipo de cambio: historia y pronóstico a largo plazo del AR(1) en diferencias con ordenada al origen.

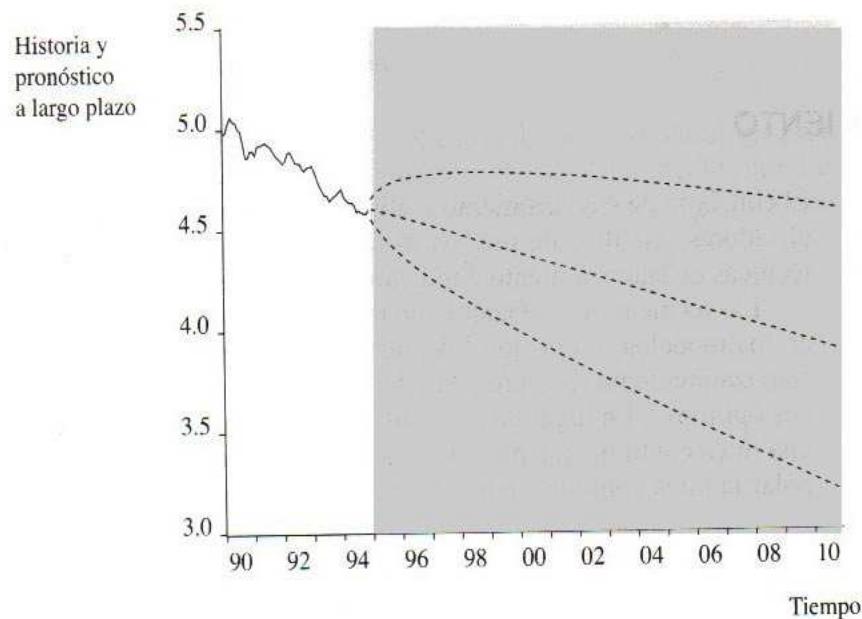
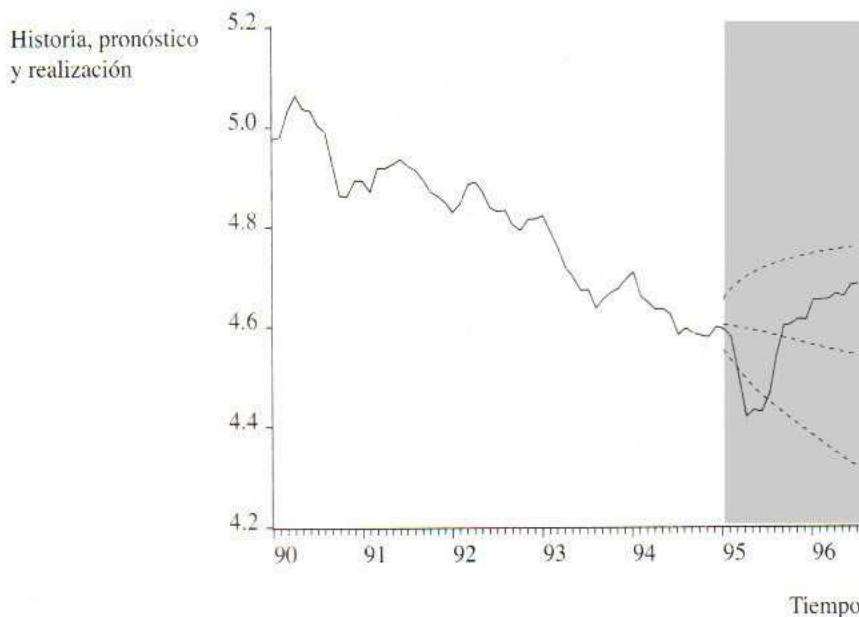


FIGURA 10.17 Logaritmo de la tasa del tipo de cambio: historia, realización y pronóstico del AR(1) en diferencias con ordenada al origen.



financieras sugieren que las tasas del tipo de cambio se deben acercar a las caminatas aleatorias porque si el cambio fuera predecible, uno se podría hacer muy rico con muy poco esfuerzo, y ya el hecho de hacerlo así eliminaría la oportunidad.¹⁴

Irónicamente, a pesar de los argumentos a favor del modelo de tendencia estocástica para $\ln y$, el de tendencia determinista trabaja un poco mejor para pronósticos fuera de la muestra, en especial en este conjunto de datos. El error cuadrático medio del pronóstico con el modelo de tendencia determinista es 0.0107, mientras que para el de tendencia estocástica es de 0.0109. En el capítulo 12 aprenderemos cómo evaluar el significado de esas diferencias de manera precisa; con toda seguridad, la diferencia que nos ocupa en este momento no es significativa.

4. SUAVIZAMIENTO

El concepto de suavizamiento o alisamiento de series temporales lo presentamos al estudiar los alisadores sencillos de medias móviles para estimar la tendencia.¹⁵ Ahora explicaremos otras técnicas de suavizamiento e indicaremos cómo se pueden emplear para hacer pronósticos.

Las técnicas de suavizamiento tradicionales se diferencian de los métodos modernos basados en los modelos, y que son los que se han utilizado en este libro. Por ejemplo, las técnicas de suavizamiento no requieren “modelos de mejor ajuste” y por lo general, no producen “pronósticos óptimos”. En lugar de ello, tan solo son una forma de decirle a la computadora cómo trazar una línea continua que pase por los datos, exactamente como lo haríamos con un lápiz, y extraer la línea continua en una forma razonable y repetible.

14. En lo que respecta a la tendencia (la deriva), ella puede ayudar como aproximación local, pero téngase cuidado con las extrapolaciones muy largas. Véase la sección de los Problemas y extensiones al final de este capítulo.

15. Consulte la sección Problemas y extensiones del capítulo 4.

Al usar tales técnicas no intentamos determinar el modelo que se ajuste mejor a los datos; lo que hacemos es forzar a que un modelo preespecificado siga los datos. Por esa razón a algunas personas no les gustan las técnicas de suavizamiento, pero este prejuicio es ingenuo y refleja una comprensión somera de los aspectos clave de los pronósticos aplicados: esas técnicas durante muchos años se han empleado con provecho y por una buena razón. Su utilidad es mayor en casos en los que los métodos basados en modelo no se pueden o no se deben usar. En primer lugar, las muestras disponibles de datos a veces son muy pequeñas. Supongamos que se debe producir un pronóstico basado en una muestra de datos históricos con tan solo cuatro observaciones. Este caso parece extremo, y lo es, pero se da en ciertas aplicaciones importantes, como cuando se pronostican las ventas de un producto recién introducido. En tales casos, los grados de libertad disponibles están tan limitados que hacen que el valor de un modelo estimado sea dudoso. En contraste, las técnicas de suavizamiento no requieren estimación, o la requieren en grado mínimo.

En segundo lugar, a veces es inmensa la tarea de pronosticar. Por ejemplo, supóngase que cada semana se deben pronosticar los precios de 10 000 insumos en un proceso de manufactura. De nuevo esos casos son extremos, pero sí se presentan en la práctica, como el recordar cuántas partes hay en un avión grande. En esas situaciones, aun si se cuenta con muchos datos históricos (y desde luego, cuando no se cuenta con ellos), simplemente no hay forma de proporcionar el cuidado y la dedicación que requieren la estimación y el mantenimiento de 10 000 modelos distintos de pronóstico. A diferencia, las técnicas de suavizamiento requieren poca atención; son un ejemplo de lo que se llega a llamar métodos de pronóstico “automáticos”, empleados para pronosticar datos voluminosos y de alta frecuencia.

Por último, en ciertas condiciones las técnicas de suavizado *sí* producen pronósticos óptimos que resultan estar íntimamente relacionadas con la presencia de raíces unitarias en la serie que se pronostica. Ésta es la razón por la cual las presentamos en este momento. Además, los métodos más complicados también sólo bajo ciertas condiciones producen pronósticos óptimos, como en la especificación correcta del modelo de pronóstico. Como lo hemos venido haciendo, es importante resaltar que todos nuestros modelos son aproximaciones y con seguridad, malos. Cualquier procedimiento con un buen récord en la práctica es digno de una seria consideración, y las técnicas de suavizamiento lo son en las situaciones descritas.

Repaso del suavizamiento con medias móviles

Como el precursor de las técnicas más complicadas de suavizamiento que pronto presentaremos, recordaremos el funcionamiento de los sencillos alisados con medias móviles. Los datos originales los representamos como $\{y_t\}_{t=1}^T$, y los alisados como $\{\bar{y}_t\}$. Entonces, la media móvil bilateral es

$$\bar{y}_t = (2m+1)^{-1} \sum_{i=-m}^{m} y_{t-i}$$

la media móvil unilateral es

$$\bar{y}_t = (m+1)^{-1} \sum_{i=0}^m y_{t-i}$$

y la media móvil unilateral ponderada es

$$\bar{y}_t = \sum_{i=0}^m w_i y_{t-i}$$

La media móvil unilateral estándar corresponde a una media móvil unilateral ponderada con todos los factores de ponderación iguales a $(m+1)^{-1}$. El usuario debe elegir m , el parámetro de alisamiento; cuanto mayor sea, más alisamiento se obtiene.

Sucede que las medias móviles unilaterales ponderadas son muy útiles en la práctica. La estructura unilateral se traduce en que en cualquier tiempo t sólo se necesitan datos actuales y pasados para calcular el valor suavizado en el tiempo t , y esto quiere decir que se puede llevar a cabo en tiempo real. Además, la ponderación permite flexibilidad en la forma de descontar el pasado. Por ejemplo, a veces se desea descontar con más peso el pasado distante que el reciente. El alisamiento exponencial, el cual veremos a continuación, es una forma especialmente útil y cómoda de implementar esa media móvil.

Suavizamiento exponencial

El **suavizamiento exponencial**, conocido también como **suavizamiento exponencial simple** o **suavizamiento exponencial único**, es lo que se llama una **media móvil exponencialmente ponderada**, las razones las veremos pronto. El marco básico es sencillo. Imagine el lector que una serie c_0 es una caminata aleatoria,

$$c_{0,t} = c_{0,t-1} + \eta_t$$

$$\eta_t \sim WN(0, \sigma_\eta^2)$$

en cuyo caso el nivel de c_0 se mueve aleatoriamente de arriba hacia abajo, y el mejor pronóstico de cualquier valor futuro es simplemente el valor actual. Sin embargo, suponga que no se ve la c_0 , sino que se ve la y , que es c_0 más el ruido blanco,¹⁶

$$y_t = c_{0,t-1} + \eta_t$$

en donde ϵ no está correlacionado con η ni hacia adelante ni hacia atrás. Entonces, el pronóstico óptimo de cualquier y futura es sólo aquel pronóstico óptimo de c_0 futura, que es la c_0 actual más nuestro pronóstico óptimo de ϵ futura, que a su vez es 0. El problema, naturalmente, es que no conocemos la c_0 actual, el “nivel local” actual; pero, sí conocemos las y actual y pasadas, que deben contener información acerca de la c_0 actual. Cuando el proceso generador de datos se escribe en la forma anterior, el suavizamiento exponencial define al estimado óptimo de c_0 y en consecuencia, al pronóstico óptimo de cualquier valor futuro de y , sobre la base de las y actual y pasadas. Cuando el proceso generador de datos no se escribe en la forma de arriba, el pronóstico de suavizamiento exponencial no será óptimo, pero es de esperarse que no estará muy alejado.

Siguiendo la costumbre, enunciaremos el procedimiento de suavizamiento exponencial como un algoritmo para convertir la serie observada, $\{y_t\}_{t=1}^T$, en una serie suavizada, $\{\bar{y}_t\}_{t=1}^T$, y los pronósticos, $\hat{y}_{T+h,T}$:

16. Se puede imaginar que el ruido blanco agregado es un error de medición.

1. Inicia en $t = 1$: $\bar{y}_1 = y_1$.
2. Actualizar: $\bar{y}_t = \alpha y_t + (1 - \alpha) \bar{y}_{t-1}$, $t = 2, \dots, T$.
3. Pronosticar: $\hat{y}_{T+h,T} = \bar{y}_T$.

En cuanto al nivel de c_0 , llamamos \bar{y}_t al estimador del *nivel* en el tiempo t . El parámetro α de suavizamiento está en el intervalo unitario, $\alpha \in [0,1]$. Cuanto menor es α , el nivel estimado es más liso. Cuando α tiende a cero, la serie alisada tiende a ser constante, y cuando α tiende a 1, la misma serie tiende a una interpolación punto a punto. En forma característica, cuanto más observaciones se tengan por unidad de tiempo conforme al calendario, se necesitará más alisamiento; en consecuencia, se alisan más los datos semanales (52 observaciones por año), que los datos trimestrales (4 observaciones por año). Sin embargo, nada sustituye a un método de tanteos en el que intervengan varios valores del parámetro de suavizamiento.

En un principio, no es obvio que el algoritmo descrito produzca una media móvil unilateral con factores de ponderación declinando exponencialmente. Para convencerlo a usted, comenzaremos con la recursión básica

$$\bar{y}_t = \alpha y_t + (1 - \alpha) \bar{y}_{t-1}$$

y sustituimos las \bar{y} hacia atrás, con lo que se obtiene

$$y_t = \sum_{j=0}^{t-1} w_j y_{t-j}$$

siendo

$$w_j = \alpha(1-\alpha)^j$$

Por ejemplo, supongamos que $\alpha = 0.5$. Entonces

$$w_0 = 0.5(1-0.5) = 0.25$$

$$w_1 = 0.5(1-0.5)^1 = 0.5$$

$$w_2 = 0.5(1-0.5)^2 = 0.125$$

y así sucesivamente. Así, los factores de ponderación de las medias móviles declinan en forma exponencial, como lo afirmamos.

Obsérvese que el suavizamiento exponencial tiene una estructura recursiva que puede ser muy conveniente cuando los datos son voluminosos. En cualquier tiempo t , el nuevo estimador del nivel \bar{y}_t en ese tiempo es función sólo del estimador \bar{y}_{t-1} calculado antes, y de la nueva observación, y_t . En esta forma no hay necesidad de volver a alisar todo el conjunto de datos a medida que llegan nuevos.

Suavizamiento de Holt-Winters

Ahora imaginemos que no sólo tenemos un nivel local con evolución lenta, sino también una tendencia con pendiente local de evolución lenta,

$$y_t = c_{0t} + c_{1t} \text{TIEMPO} + \epsilon_t$$

$$c_{1t} = c_{1,t-1} + v_t$$

$$c_{0t} = c_{0,t-1} + \eta_t$$

donde todas las perturbaciones son ortogonales, tanto hacia adelante como hacia atrás. Entonces, el algoritmo óptimo de alisamiento, llamado alisamiento de Holt-Winters en honor de los investigadores que lo desarrollaron en las décadas de 1950 y 1960, es

1. Iniciar en $t = 2$:

$$\bar{y}_2 = y_2$$

$$F_2 = y_2 - \bar{y}_1$$

2. Actualizar:

$$\bar{y}_t = \alpha y_t + (1-\alpha)(\bar{y}_{t-1} + F_{t-1}), \quad 0 < \alpha < 1$$

$$F_t = \beta y_t + (\bar{y}_t - \bar{y}_{t-1}) + (1-\beta)F_{t-1}, \quad 0 < \beta < 1$$

3. Pronosticar: $\hat{y}_{T+h|T} = \bar{y}_T + hF_T$

\bar{y}_t es el nivel estimado o suavizado en el tiempo t , y F_t es la pendiente estimada en el tiempo t . El parámetro α controla el alisamiento de nivel y β el de la pendiente. El pronóstico a h etapas tan solo toma el nivel estimado en el tiempo T y lo aumenta h veces la pendiente estimada en el tiempo T .

Nuevamente, aunque hemos presentado el proceso generador de datos para el que el suavizamiento de Holt-Winters produce pronósticos óptimos, cuando lo aplicamos no suponemos que los datos sean realmente generados por ese proceso. Sin embargo, cabe esperar que el proceso real de generación de datos se parezca a aquel para el cual el suavizado Holt-Winters es óptimo, en cuyo caso los pronósticos Holt-Winters se pueden acercar a lo óptimo.

Suavizamiento Holt-Winters con estacionalidad

Podemos ampliar el algoritmo de suavizamiento de Holt-Winters para considerar la estacionalidad de periodo s . El algoritmo se transforma en:

1. Iniciar en $t = s$:

$$\bar{y}_s = \frac{1}{s} \sum_{t=1}^s y_t$$

$$F_s = 0$$

$$G_j = \frac{y_j}{\left(\frac{1}{s} \sum_{t=1}^s y_t \right)}, \quad j = 1, 2, \dots, s$$

2. Actualizar:

$$\bar{y}_t = \alpha(y_t - G_{t-s}) + (1-\alpha)(\bar{y}_{t-s} - F_{t-s}), \quad 0 < \alpha < 1$$

$$F_t = \beta(\bar{y}_t - \bar{y}_{t-s}) + (1-\beta)F_{t-s}, \quad 0 < \beta < 1$$

$$G_t = \gamma(\bar{y}_t - \bar{y}_s) + (1-\gamma)G_{t-s}, \quad 0 < \gamma < 1$$

$$t = s+1, \dots, T$$

3. Pronosticar:

$$\hat{y}_{T+h,T} = \bar{y}_T + hF_T + G_{T+h}, \quad h=1, 2, \dots, s,$$

$$\hat{y}_{T+h,T} = \bar{y}_T + hF_T + G_{T+h+2}, \quad h=s+1, s+2, \dots, 2s,$$

etc.

Lo único nuevo está en la recursión para la estacionalidad, con el parámetro de alisamiento γ .

Pronóstico con técnicas de suavizamiento

Sin importar qué técnica de suavizamiento se use, el paradigma básico es el mismo. Se sustituyen los datos en un algoritmo que alise los datos y permita generar pronósticos puntuales. Se puede demostrar que esos pronósticos son óptimos para ciertos procesos generadores de datos, como lo indicamos para el suavizamiento exponencial simple y el de Holt-Winters sin estacionalidad. En la práctica no se conoce si el proceso real de generación de datos se aproxima a aquel por el cual la técnica de suavizamiento adoptada óptima; en lugar de ello, tan solo se tiene la esperanza que así sea. Éste es el principal contraste con el método basado en un modelo, en el cual normalmente se pasa uno largo tiempo tratando de encontrar una especificación “buena”.

El carácter de “unitalla” del método de suavizamiento tiene su costo, porque con seguridad una talla *no* le ajusta a todo, pero a veces tiene ventajas porque no necesitan estimarse parámetros, o sólo unos cuantos. En ocasiones tan solo se establecen los valores de los parámetros de alisamiento con base en el conocimiento de las propiedades de la serie manejada y en otras, se seleccionan los valores de parámetro que producen los mejores pronósticos a h etapas bajo la función de pérdida relevante. Por ejemplo, bajo la pérdida de error cuadrático a 1 etapa, si el tamaño de muestra es lo bastante grande como para ameritar entretenernos en la estimación de los parámetros de suavizamiento, los podemos estimar como sigue:

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_{t=m+1}^T (y_t - y_{t-1,t})^2$$

en donde m es un entero lo suficientemente grande, de manera que los valores iniciales del algoritmo de suavizamiento tiene poco efecto.

Para finalizar la sección, haremos notar que las técnicas de alisamiento, tal como se implementan normalmente, sólo producen pronósticos puntuales. Pueden producir pronósticos puntuales óptimos en ciertos procesos especiales de generación de datos, pero en general no se supone que dichos procesos sean verdaderos. En lugar de ello, las técnicas de suavizamiento se usan como “cajas negras” para producir pronósticos puntuales y no se intenta sacar provecho de la estructura estocástica de los datos para definir un modelo de mejor ajuste que se pueda usar para construir intervalos de densidad de pronósticos, además de los puntuales.

5. TASA DEL TIPO DE CAMBIO (CONTINUACIÓN)

En seguida pronosticaremos la tasa del tipo de cambio yen/dólar con un procedimiento de alisamiento. En los modelos ARIMA(p, d, q) que se describieron antes, siempre se tomó en cuenta una tendencia (determinista o estocástica). Para poder seguir comparando, aplicaremos un suavizamiento Holt-Winters que permite tendencia localmente lineal. Los resultados de la estimación aparecen

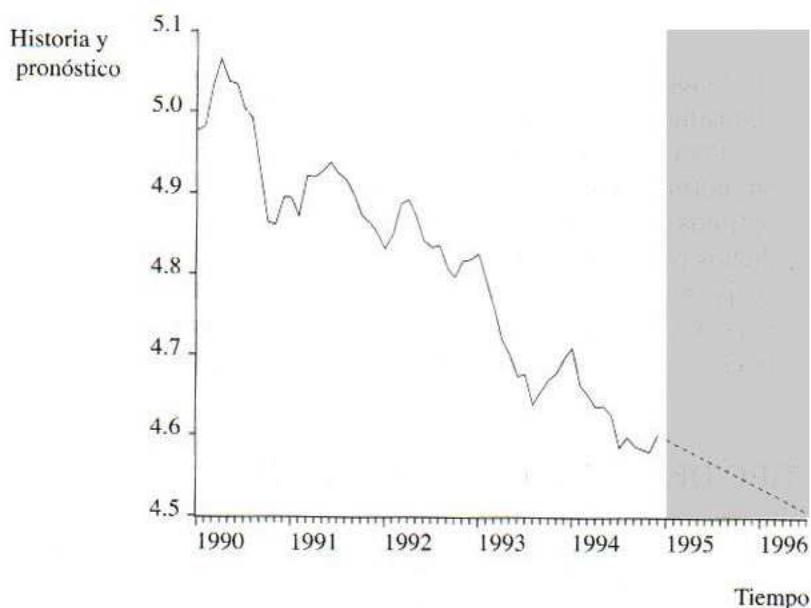
TABLA 10.8 Logaritmo de la tasa del tipo de cambio: alisamiento Holt-Winters

Muestra: 1973.01 1994.12
 Observaciones incluidas: 264
 Método: Holt-Winters, no estacional
 Serie original: LYEN
 Serie pronosticada: LYENSM

Parámetros:	Alfa	1.000000
	Beta	0.090000
	Suma de residuales al cuadrado	0.202421
	Raíz del error cuadrático medio	0. 027690
Niveles al final de periodo:	Media	4.606969
	Tendencia	-0.005193

en la tabla 10.8. El estimador de α es grande, de modo que el nivel local estimado se mueve bastante cerca de la serie. Por otro lado, el estimador de β es pequeño, por lo que la pendiente de la tendencia local es mucho menos adaptable.

El pronóstico Holt-Winters es tan solo la línea de tendencia que comienza en el nivel estimado al final del periodo, con la pendiente estimada al final del periodo. Como la pendiente estimada de la tendencia al final de la muestra tiene mayor valor absoluto que las pendientes de las tendencias correspondientes en los modelos de tendencias determinista y estocástica, esperamos que los pronósticos puntuales Holt-Winters disminuyan con un poco más de rapidez que los basados en los modelos ARIMA. La figura 10.18 refleja la historia y el pronóstico fuera de la muestra. No aparecen los intervalos de confianza porque no los producen las técnicas de suavizamiento. El pronóstico se ve semejante al de los modelos ARIMA, excepto que decrece un poco más rápido, lo que se ve

FIGURA 10.18 Logaritmo de la tasa del tipo cambio: historia y pronóstico mediante el suavizamiento Holt-Winters.

mejor en el pronóstico a largo plazo de la figura 10.19. Por último, en la figura 10.20 se ve también la realización. Para pronosticar fuera de la muestra, el de Holt-Winters da los peores resultados de todos los métodos que se probaron en este capítulo. El error cuadrático medio del pronóstico es 0.0135.

PROBLEMAS Y EXTENSIONES

1. (Modelado y pronóstico de la tasa del tipo de cambio marco alemán/dólar) En la base de datos del disco usted encontrará datos mensuales de la tasa del tipo de cambio marco alemán/dólar exactamente para el mismo periodo que el tipo de yen/dólar que examinamos en este capítulo.
 - a) Modele y pronostique la tasa del tipo de cambio marco alemán/dólar, al igual que se hizo en el texto, y discuta sus resultados.
 - b) Vuelva a hacer sus análisis aplicando métodos de pronóstico sin tendencias: un modelo de niveles sin tendencia, uno de primera diferencia sin deriva y un suavizamiento exponencial simple.
 - c) Compare la capacidad de pronóstico de los métodos con y sin tendencia.
 - d) ¿Se siente usted satisfecho con la inclusión de la tendencia en un modelo de pronóstico para la tasa del tipo de cambio? ¿Por qué?
2. (Modelado ARIMA automático) Existen programas informáticos de pronóstico “automático” para implementar las técnicas ARIMA y de alisamiento exponencial de este capítulo y los anteriores, sin intervención humana.
 - a) ¿Cuáles cree que sean las ventajas de esos programas?
 - b) ¿Cuáles serán los costos?
 - c) ¿Cuándo serán más útiles?
 - d) Lea el artículo de Ord y Lowe (1996), que describe la mayoría de los programas de pronóstico automático e informe lo que aprendió. Despues conteste: ¿cómo modificaría, si es el caso, sus respuestas a los incisos a, b y c?
3. (El modelo multiplicativo estacional ARIMA(p, d, q) x (P, D, Q) multiplicativo estacional.) Se tiene el modelo de pronóstico

$$\Phi_s(L)^s \Phi(L) (1-L)^d (1-L^s)^D y_t = \Theta_s(L)^s \Theta(L) \varepsilon_t$$

$$\Phi_s(L) = 1 - \varphi_1 s L^s - \dots - \varphi_p s L^p$$

$$\Phi(L) = 1 - \varphi_1 L - \dots - \varphi_p L^p$$

$$\Theta_s(L) = 1 - \theta_1 s L^s - \dots - \theta_Q s L^Q$$

$$\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$$

- a) El modelo ARIMA(p, d, q) estándar es un caso especial de este modelo más general. ¿En qué caso se presenta? ¿Cuál es el significado de la notación ARIMA(p, d, q) x (P, D, Q)?
- b) El operador $(1 - L^s)$ se llama operador de diferencia estacional. ¿Qué hace cuando opera sobre y_t ? ¿Por qué aparece en forma rutinaria en modelos para datos estacionales?
- c) La aparición de $(1 - L^s)$ en el polinomio del operador rezago asociado a la parte autorregresiva permite la **estacionalidad estocástica**, en contraste con la estacionalidad determinista del capítulo 5, igual que la aparición de $(1 - L)$ produce una tendencia estocástica en contraste con una determinista. Haga comentarios al respecto.
- d) ¿Puede usted citar alguna motivación intuitiva para el modelo? *Sugerencia:* considere un modelo ARIMA(P, D, Q) puramente estacional, con perturbaciones correlacionadas

FIGURA 10.19 Logaritmo de la tasa del tipo de cambio: historia y pronóstico a largo plazo mediante el alisamiento Holt-Winters.

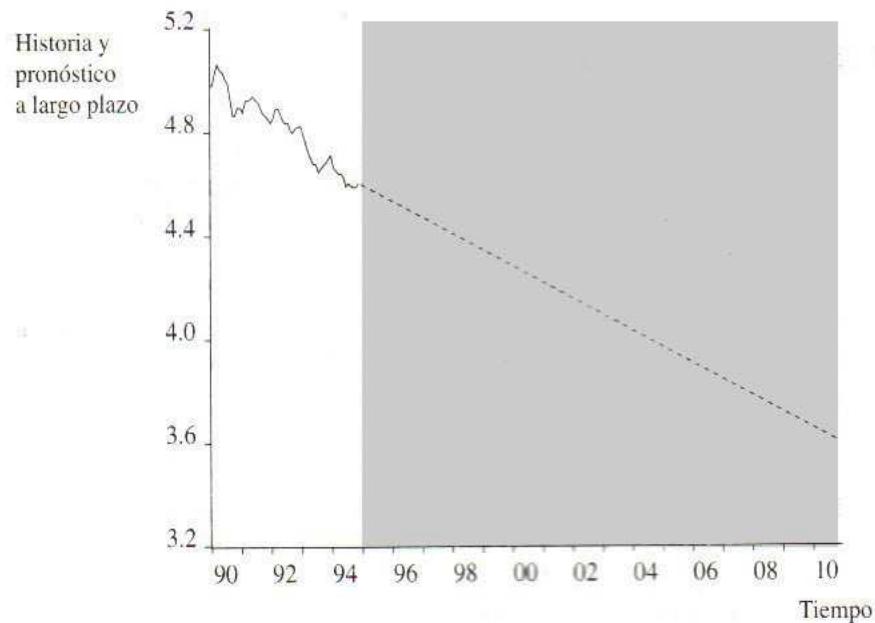
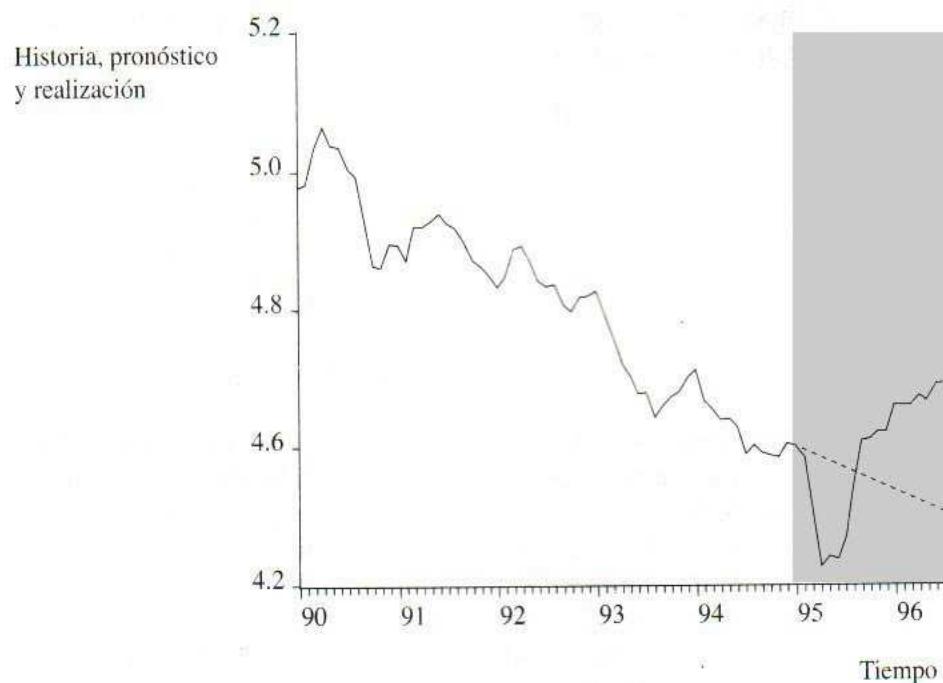


FIGURA 10.20 Logaritmo de la tasa del tipo de cambio: historia, realización y pronóstico mediante el alisamiento Holt-Winters.



- serialmente. ¿Por qué podrían estar serialmente correlacionadas las perturbaciones? ¿En particular, qué sucede si un modelo ARIMA(P, D, Q) tiene perturbaciones ARIMA(p, d, q)?
- La estructura multiplicativa implica restricciones. Por ejemplo, ¿qué se obtiene al multiplicar $\Phi_1(L)$ y $\Phi_2(L)$?
 - ¿Cuáles serán los costos y los beneficios de pronosticar con el modelo ARIMA multiplicativo en comparación con el modelo “estándar”?
4. (La regresión Dickey-Fuller en el caso AR(2)) Se tiene el proceso AR(2),

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \varepsilon_t$$

- a) Demuestre que se puede escribir en la forma

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

siendo

$$\rho_1 = - \sum_{j=1}^p \varphi_j = -(\varphi_1 + \varphi_2)$$

$$\rho_2 = \sum_{j=2}^p \varphi_j = \varphi_2$$

- b) Demuestre que también se puede escribir como una regresión de Δy_t sobre y_{t-1} y Δy_{t-1} .
- c) Compruebe que si $\rho_1 = 1$, el proceso AR(2) es realmente uno AR(1) en primeras diferencias, esto es, el proceso AR(2) tiene una raíz unitaria.
5. (Modelos ARIMA, alisamiento y de contracción.) Considere la posición ventajosa del principio de contracción y describa los beneficios asociados a los pronósticos “óptimos” de los modelos ARIMA ajustados en comparación con los de los pronósticos “ad hoc” a partir de alisamientos.
6. (Suavizamiento Holt-Winters con estacionalidad multiplicativa.) Se tiene un alisamiento estacional Holt-Winters expresado como:

1. Iniciar en $t = s$:

$$\bar{y}_s = \frac{1}{s} \sum_{t=1}^s y_t$$

$$T_s = 0$$

$$F_j = \frac{y_j}{\left(\frac{1}{s} \sum_{t=1}^s y_t \right)}, \quad j = 1, 2, \dots, s$$

2. Actualizar:

$$\bar{y}_t = \alpha \left(\frac{y_t}{F_{t-s}} \right) + (1-\alpha) (\bar{y}_{t-1} + T_{t-1}), \quad 0 < \alpha < 1$$

$$T_t = \beta (\bar{y}_t + \bar{y}_{t-1}) + (1-\beta) T_{t-1}, \quad 0 < \beta < 1$$

$$F_t = \gamma \left(\frac{\bar{y}_t}{\bar{y}} \right) + (1-\gamma) F_{t-s}, \quad 0 < \gamma < 1$$

$t=s+1, \dots, T$

3. Pronosticar:

$$y_{T+h,T} = y_T + h T_F F_{T+h-s}, \quad h=1, 2, \dots, s,$$

$$y_{T+h,T} = \left(y_T + h T_F \right) + F_{T+h-2s}, \quad h=s+1, s+2, \dots, 2s,$$

etc.

- a) El algoritmo de alisamiento estacional Holt-Winters citado en el texto se llama, con más propiedad, alisamiento estacional Holt-Winters con estacionalidad aditiva. El algoritmo citado aquí, en contraste, se llama suavizamiento de Holt-Winters con estacionalidad multiplicativa. ¿En qué difiere este algoritmo del citado en el texto y cuál es el significado, si es que lo hay, de la diferencia?
- b) Evalúe la afirmación de que el método de Holt-Winters con estacionalidad multiplicativa es adecuado cuando el comportamiento estacional presenta variación creciente.
- c) ¿Cómo compara el alisamiento de Holt-Winters con estacionalidad multiplicativa con el de Holt-Winters con estacionalidad aditiva aplicados al logaritmo de los datos originales?
- 7. (Uso de modelos de componentes no observados con tendencia estocástica para implementar las técnicas de alisamiento en un marco probabilista.) En el texto se dijo que las técnicas de alisamiento tal y como se implementan normalmente, se usan como “cajas negras” para construir pronósticos puntuales. No se trata de explotar la estructura estocástica para hacer intervalos o densidades de pronósticos, además de los pronósticos puntuales. Sin embargo, recuerde que los diversos alisamientos producen pronósticos óptimos para procesos específicos de generación de datos especificados como modelos de componentes no observados.
 - a) ¿Para qué procesos generadores de datos es óptimo el suavizamiento exponencial?
 - b) ¿Para qué procesos generadores de datos es óptimo el suavizamiento de Holt-Winters?
 - c) Si supone que el proceso generador de datos para el cual el alisamiento exponencial produce pronósticos óptimos que de hecho es el proceso real de generación de datos, ¿cómo estimaría el modelo de componentes no observados y cómo lo usaría para producir intervalos y de densidades de pronóstico óptimos? (Sugerencia: vea el libro de Harvey, 1989.)
 - d) ¿Cómo interpretaría usted los intervalos y densidades de pronósticos producidos con el método del inciso c, si ya no se supone que sea un modelo determinado para el proceso real de generación de datos?

NOTAS BIBLIOGRÁFICAS Y COMPUTACIONALES

Es de esperarse que las caminatas aleatorias o casi aleatorias, sean buenos modelos de los precios de los activos financieros, y sí lo son. Vea Malkiel (1981). Los modelos ARIMA($p, 1, q$) más generales han encontrado gran aplicación en el comercio, la economía y el gobierno. Beveridge y Nelson (1981), demuestran que el proceso $I(1)$ siempre se puede desglosar en una suma de un componente de caminata aleatoria y uno estacionario en covarianza. Tsay (1984), demuestra que los criterios de información, como el de Schwarz, siguen siendo válidos para seleccionar órdenes del modelo ARIMA, sin importar si está presente una raíz unitaria autorregresiva.

En paralelo con el tratamiento de modelos de componentes no observados con tendencia determinista de Nerlove, Grether y Carvalho (1996, primera edición 1979), Harvey (1989) maneja la especificación, la estimación y el pronóstico con modelos de componentes no observados en conjunción con el filtro Kalman.

Los pronósticos del PNB per cápita de Estados Unidos examinados en el texto y la discusión relacionada, se basan mucho en el artículo de Diebold y Senhadji (1997).

El desarrollo de métodos para eliminar el sesgo Dickey-Fuller de los parámetros los modelos estimados para pronóstico, que podría conducir a mejores pronósticos, es en la actualidad un área de intensa investigación. Véanse entre otros, los artículos de Andrews (1993), Rudebusch (1993) y Fair (1996).

En su influyente libro, Box y Jenkins proponen un proceso iterativo de modelado, que consiste en ciclos repetidos de especificación, estimación, diagnóstico, prueba y verificación del modelo. (La última edición es la de Box, Jenkins y Reinsel, 1994.) Un elemento clave de la estrategia de modelado de Box-Jenkins es la hipótesis de que los datos se apegan a un modelo ARIMA (al que a veces se le llama modelo de Box-Jenkins),

$$\Phi(L)(1-L)^d y_t = \Theta(L)\varepsilon_t$$

Así, aunque y_t sea no estacionaria, se supone que su d -ésima diferencia sigue a un proceso ARMA estacionario e invertible. La adecuación de la táctica de Box-Jenkins, de diferenciar para llegar a la estacionariedad, depende de la existencia de una o más raíces unitarias en el polinomio en el operador rezago autorregresivo, responsable en parte de la gran cantidad de investigaciones sobre las pruebas de raíz unitaria.

Las pruebas de Dickey-Fuller datan de los trabajos de Dickey (1976) y de Fuller (1976). Usando técnicas de simulación MacKinnon (1991), obtiene estimadores muy precisos de los puntos porcentuales de las diversas distribuciones de Dickey-Fuller.

Phillips y Perron (1988), proponen en su artículo alternativas a las pruebas de raíz unitaria, llamadas pruebas Phillips-Perron. La idea básica de esta prueba es estimar una regresión Dickey-Fuller sin aumento

$$x_t = \varphi x_{t-1} + \varepsilon_t$$

y entonces corregir el estadístico Dickey-Fuller para formas generales de correlación serial y/o heteroscedasticidad que pudieran estar presentes en ε_t . Véase en el libro de Hamilton (1994), una discusión detallada de las pruebas Phillips-Perron así como su comparación con las pruebas de Dickey-Fuller aumentadas.

Las técnicas de alisamiento fueron propuestas originalmente como estrategias razonables, si bien ad hoc, de pronóstico; sólo después se formalizaron en términos de pronósticos óptimos para modelos de componentes implícitos no observados de tendencia estocástica. La idea de implementar técnicas de alisamiento en ambientes estocásticos a través de modelos de componentes no observados de tendencia estocástica, es un tema clave en el libro de Harvey (1989), que también contiene referencias a contribuciones anteriores importantes a publicaciones sobre

suavizamiento incluyendo las de Holt (1957) y Winters (1960). El impresionante programa informático Stamp de Koopman *et al.* (1995), se puede emplear para estimar y diagnosticar modelos de componentes no observados de tendencia estocástica y para producir pronósticos.¹⁷ Stamp (por sus siglas en inglés: structural time series analyzer, modeller an predictor) significa analizador, modelador y predictor de series temporales estructurales; a veces, a los modelos de componentes no observados se les llama modelos de serie temporal estructural.

CONCEPTOS A REVISAR

Caminata aleatoria, con y sin deriva	Raíz unitaria
Distribución de Dickey-Fuller	Raíz unitaria autorregresiva
Estacionalidad estocástica	Reversión a la media
Modelo ARIMA(p, d, q)	Suavizamiento exponencial simple
Media móvil exponencialmente ponderada	Suavizamiento exponencial único
Prueba aumentada de Dickey-Fuller	Suavizamiento exponencial
Prueba de raíz unitaria con media distinta de cero bajo la hipótesis alternativa	Suavizamiento Holt-Winters
Prueba de raíz unitaria con tendencia lineal determinista bajo la hipótesis alternativa	Suavizamiento Holt-Winters con estacionalidad
	Tendencia estocástica y determinista

REFERENCIAS Y LECTURAS ADICIONALES

- Andrews, D. W. K., "Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models", *Econometrica*, 61, 1993, 130-165.
- Beveridge, S. y Nelson, C. R., "A New Approach to the Decomposition of Economic Time Series into Permanent and Transient Components with Particular Attention to Measurement of the Business Cycle", *Journal of Monetary Economics*, 7, 1981, 151-174.
- Box, G. E. P., Jenkins, G. W. y Reinsel, G., *Time Series Analysis, Forecasting and Control*, 3a. ed., Prentice Hall, Englewood Cliffs, N. J.: 1994.
- Campbell, J. Y. y Perron, P., "Pitfalls and Opportunities: What Macroeconomists Should Know About Unit Roots", en O. J. Blanchard y S. S. Fischer (editores), *NBER Macroeconomics Annual*, 1991. MIT Press, Cambridge, Mass.: 1991.
- Dickey, D. A., *Estimation and Hypothesis in Nonstationary Time Series*, Disertación doctoral, Universidad del estado de Iowa, 1976.
- Dickey, D. A. y Gonzalez-Farias, G., "A New Maximum-Likelihood Approach to Testing for Unit Roots", manuscrito, Departamento de Estadística, Universidad del estado de North Carolina, 1992.
- Diebold, F. X., Giorgianni, L. e Inoue, A., "STAMP 5.0: A Review", *International Journal of Forecasting*, 12, 1996, 309-315.
- Diebold, F. X. y Rudebusch, G. D., *Business Cycles: Durations, Dynamics, and Forecasting*, 12, 1996, 309-315.
- Diebold, F. X. y Senhadji, A., "Deterministic vs. Stochastic Trend in U.S. GNP, YET Again", *American Economic Review*, 86, 1996, 1291-1298. Reimpreso en Diebold y Rudebusch 1997.
- Elliot, G., Rothenberg, T. J. y Stock, J., "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, 64, 1996, 813-836.

17. Véase un repaso en el artículo de Diebold, Giorgianni e Inoue (1996).

- Fair, R. C., "Computing Median Unbiased Estimates in Macroeconometric Models", *Journal of Business and Economic Statistics*, 12, 1996, 461-470.
- Fuller, W. A., *Introduction to Statistical Time Series*, Wiley, Nueva York, 1976
- Hall, A., "Testing for a Unit Root in Time Series with Pretest Data-Based Model Selection", *Journal of Business and Economic Statistics*, 12, 1994, 461-470.
- Hamilton, J. D., *Time Series Analysis*, Princeton University Press, Princeton N.J., 1994.
- Harvey, A. C., *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge.
- Holt, C. C., "Forecasting Seasonal and Trends by Exponentially Weighted Moving Averages", Memorandum de investigación ONR No. 52, Carnegie Institute of Technology, 1957.
- Koopman, S. J., Harvey, A. C., Doornik, J. A. y Shephard, N., *Stamp 5.0: Structural Time Series Analyzer, Modeller and Predictor*, Chapman and Hall, Londres, 1995.
- MacKinnon, J. G., "Critical Values for Cointegration Tests", en R. F. Engle y C. W. J. Granger (editores), *Long-Run Economic Relationships*, Oxford University Press, Oxford, 1991.
- Malkiel, B. G., *A Random Walk Down Street*, W. W. Norton and Company, Nueva York, 1981.
- Nerlove, M., Grether, D. M. y Carvalho, J. L., *Analysis of Economic Time Series: A Synthesis*, 2a, ed., Academic Press, Nueva York, 1996.
- Ng, S. y Perron, P., "Unit Root Tests in ARMA Models with Data-Dependent Methods for the Selection of the Truncation Lag", *Journal of the American Statistical Association*, 90, 1995, 268-281.
- Ord, K y Lowe, S., "Automatic Forecasting", *American Statistician*, 50, 1996, 88-94.
- Pantula, S. G., Gonzalez-Farias, G. y Fuller, W. A., "A Comparison of Unit Root Test Criteria", *Journal of Business and Economic Statistic*, 12, 1994, 449-459.
- Pillips, P. C. B. y Perron, P., "Testing for a Unit Root in Time Series Regression", *Biometrika*, 75, 1988, 335-346.
- Rudebusch, G. D., "The Uncertain Unit Root in Real GNP", *American Economic Review*, 83, 1993, 264-272. Reimpreso en Diebold y Rudebusch (1997).
- Tsay, R., "Order Selection in Nonstationary Autoregressive Models", *Annals of Statistic*, 12, 1984, 1425-1433.
- Winters, P. R., "Forecasting Sales by Exponentially Weighted Moving Averages", *Management Science*, 6, 1960, 323-342.