

CAPÍTULO 9

Raíces Unitarias y Tendencias Estocásticas (Modelos ARIMA)

Índice

9.1. Introducción	225
9.2. Modelos ARIMA	225
9.3. Modelos de Marcha Aleatoria	227
9.4. Ejemplos de modelos ARIMA	230
9.5. Pruebas de Hipótesis para Raíces Unitarias	236
9.6. Prueba Dickey-Fuller	238
9.7. Valores críticos de la prueba Dickey-Fuller	239
9.8. Prueba Dickey-Fuller Aumentada	241
9.9. La prueba KPSS	247
9.10. La prueba PP	249
9.11. Pruebas de No linealidad en Series de Tiempo	250
9.12. Modelos híbridos ARIMA - Red Neuronal	252
9.13. Combinación de pronósticos	254

9.1. Introducción

En este capítulo se presenta un modelo alternativo al de componentes con errores SARIMA, conocido como modelo ARIMA-SARIMA ó modelo de Box-Jenkins. El modelo ARIMA es una alternativa al modelo con tendencia, $Y_t = T_t + \varepsilon_t$, y el modelo SARIMA una alternativa al modelo con tendencia y estacionalidad, $Y_t = T_t + S_t + \varepsilon_t$. Primero se introducen los modelos ARIMA, luego las pruebas de raíz unitaria de Dickey-Fuller para detectar estos modelos. En el capítulo siguiente se introducen los modelos SARIMA y las pruebas de raíz unitaria estacional para detectarlos.

9.2. Modelos ARIMA

Observación. Si en $Y_t = a + bt + \varepsilon_t$, $\varepsilon_t \sim RB(0, \sigma^2)$ se toma $\Delta Y_t = (1 - L)Y_t = Y_t - Y_{t-1}$, se obtiene

$$\begin{aligned} Y_t - Y_{t-1} &= a + bt + \varepsilon_t - (a + b(t-1) + \varepsilon_{t-1}) \\ &= b + \varepsilon_t - \varepsilon_{t-1} = b + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad \theta_1 = -1, \end{aligned}$$

luego, llamando

$$\begin{aligned} W_t &= \Delta Y_t \\ \eta_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1}, \\ \therefore W_t &= b + \eta_t. \end{aligned}$$

Nótese que η_t es MA(1) y por tanto W_t es un MA(1) con media diferente de cero, estacionario en covarianza. Luego, diferenciar una serie no estacionaria, con tendencia lineal, puede producir una serie estacionaria en covarianza. Esta idea es básica en los modelos ARIMA. Pero requiere algunos conceptos para formalizarla.

Definición 9.2.1 (Raíz Unitaria Autoregresiva). Suponga un modelo ARMA(p, q)

$$\Phi_p(L)Y_t = \Theta_q(L)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2),$$

tal que una de las p raíces de la ecuación $\Phi_p(z) = 0$ es $z = 1$. Entonces se dice que Y_t tiene una raíz unitaria autorregresiva.

En este caso el polinomio $\Phi_p(z)$ factoriza como $\Phi_p(z) = (1 - z)\Phi_{p-1}(z)$ donde $\Phi_{p-1}(z)$ es un polinomio de grado $p - 1$ y

$$\Phi_{p-1}(L)(1 - L)Y_t = \Theta_q(L)\varepsilon_t,$$

luego si $Y_t \sim ARMA(p, q)$ con una raíz unitaria entonces

$$(1 - L)Y_t = \Delta Y_t \sim ARMA(p - 1, q).$$

Ejemplo 9.2.1. Si $\Phi(L) = (1 - 0.2L)(1 - L)$, $\Theta(L) = 1 - 0.3L$, entonces

$$\begin{aligned} (1 - 0.2L)(1 - L)Y_t &= (1 - 0.2L)\varepsilon_t \\ (1 - 1.2L + 0.2L^2)Y_t &= \varepsilon_t - 0.3\varepsilon_{t-1} \\ Y_t - 1.2Y_{t-1} + 0.2Y_{t-2} &= \varepsilon_t - 0.3\varepsilon_{t-1}. \end{aligned}$$

Entonces $Y_t \sim ARMA(2, 1)$, y no es estacionario en covarianza pues $1 - 1.2z + 0.2z^2 = 0$ tiene raíces $z_1 = 1$, $z_2 = 5$ y sólo una cumple $|z| > 1$. Pero $W_t = \Delta Y_t = Y_t - Y_{t-1}$ cumple $(1 - 0.2L)W_t = (1 - 0.3)\varepsilon_t \sim ARMA(1, 1)$ es estacionario en covarianza e invertible.

Definición 9.2.2 (Proceso ARIMA(p,1,q)). Una serie de tiempo $(Y_t, t \in \mathbb{Z})$ sigue un modelo $ARIMA(p, 1, q)$, con tendencia, si $(1 - L)Y_t$ sigue un proceso $ARMA(p, q)$ estacionario en covarianza, con media diferente de cero. Es decir, si

$$\Phi_p(L)[(1 - L)Y_t - \mu] = \Theta_q(L)\varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2). \quad (9.1)$$

Nótese que si $Y_t \sim ARIMA(p, 1, q)$ entonces $\Delta Y_t = W_t$, con W_t estacionario. Si Y_0 se asume dada, entonces $Y_t = Y_0 + \sum_{j=1}^t \Delta Y_j$. Por lo que la operación de suma acumulada puede considerarse el inverso de la operación de diferencia. Un $ARIMA(p, 1, q)$ es una suma acumulada de una serie $ARMA(p, q)$ estacionaria en covarianza.

Una serie de tiempo sigue un modelo $ARIMA(p, d, q)$, $d = 1, 2, \dots$ si

$$\Delta^d Y_t = (1 - L)^d Y_t$$

es un proceso $ARMA(p, q)$ estacionario en covarianza. En la práctica los casos más importantes son $d = 0, 1$, ver Diebold [1999, pág. 210].

9.3. Modelos de Marcha Aleatoria

Las Marchas Aleatorias (*Random Walks*) son modelos estocásticos muy estudiados. Una lista de al menos 15 áreas de aplicaciones de las marchas aleatorias, incluyendo el arte (escultura), aparece en Wikipedia ⁽¹⁾.

La marcha aleatoria sin tendencia se define como

$$Y_t = Y_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z}, \quad \varepsilon_t \sim iid(0, \sigma^2). \quad (9.2)$$

Según como se tome el conjunto de valores de Y_t , E , y se defina la distribución de ε_t , se obtienen muchos tipos de marchas aleatorias unidimensionales. Las marchas aleatorias no acotadas tienen $E = \mathbb{R}, \mathbb{Z}$. Las acotadas tienen $E = \mathbb{R}^+, \mathbb{Z}^+$, $E = \{d_1, \dots, d_k\}$, $E = [a, b]$, etc.

También se definen marchas aleatorias multi-dimensionales, $E = \mathbb{R}^d$, $E = \mathbb{Z}^d$, $d = 2, 3$. Las marchas aleatorias para $d \geq 3$ tienen características radicalmente diferentes de las unidimensionales.

Nótese que la marcha aleatoria en (9.2) es un proceso $Y_t \sim AR(1)$, de media cero, con coeficiente $\varphi = 1$. Las marchas aleatorias son ejemplos básicos de modelos ARIMA por lo que se las denotará: $Y_t \sim ARIMA(0, 1, 0)$.

La marcha aleatoria sin tendencia cumple, para $t \geq 0$, y Y_0 dado, no aleatorio,

$$Y_t = Y_0 + \sum_{i=0}^t \varepsilon_i, \quad (9.3a)$$

$$E(Y_t) = Y_0, \quad (9.3b)$$

$$Var(Y_t) = t\sigma^2. \quad (9.3c)$$

La marcha aleatoria con tendencia se define como un proceso $AR(1)$ de media diferente de cero, con coeficiente $\varphi = 1$, es decir

$$Y_t = \mu + Y_{t-1} + \varepsilon_t, \quad \mu \in \mathbb{R}. \quad (9.4)$$

La marcha aleatoria con tendencia cumple

$$Y_t = t\mu + Y_0 + \sum_{i=0}^t \varepsilon_i, \quad (9.5a)$$

¹https://en.wikipedia.org/wiki/Random_walk

$$E(Y_t) = t\mu + Y_0, \quad (9.5b)$$

$$Var(Y_t) = t\sigma^2. \quad (9.5c)$$

Un ejemplo de trayectorias de marchas aleatorias sin y con tendencia se muestran a continuación en la Figura 9.1.

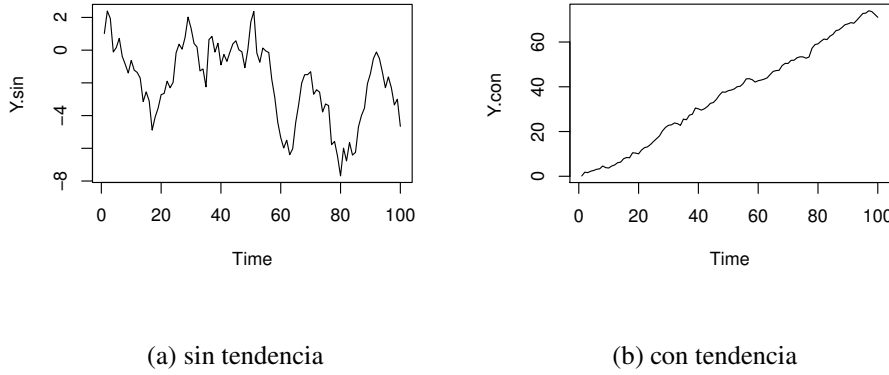


Figura 9.1: Trayectorias de Marchas Aleatorias

Obsérvese que la marcha aleatoria con tendencia Y_t , tiene una característica similar a una serie con tendencia lineal de la forma $X_t = a + bt + \varepsilon_t$, $\varepsilon_t \sim iid(0, \sigma^2)$, porque en promedio crecen en cada período $[t-1, t]$, una cantidad constante, $E(Y_t - Y_{t-1}) = \mu$ y $E(X_t - X_{t-1}) = b$.

Suponga $Y_t = \mu + \varphi Y_{t-1} + \varepsilon_t$, $t = 1, 2, \dots, T$, un proceso AR(1) con media diferente de cero. Los pronósticos a j pasos a partir de T son

$$\hat{Y}_T(j) = \mu + \varphi^j Y_T, \quad j = 1, 2, \dots, h.$$

En el caso $\varphi = 1$, se obtiene una marcha aleatoria con tendencia. En este caso $\hat{Y}_T(j) = \mu + Y_T$, $j \geq 1$, lo cual se interpreta como que una marcha aleatoria con tendencia no puede pronosticarse. Su mejor pronóstico es el último valor conocido.

Finalmente, a partir de $Y_t = Y_0 + \mu t + \sum_{j=1}^t \varepsilon_j$, para $k = 1, 2, \dots$,

$$Cov(Y_t, Y_{t+k}) = E((Y_t - E(Y_t))(Y_{t+k} - E(Y_{t+k})))$$

$$\begin{aligned}
&= \mathbb{E}\left(\sum_{j=1}^t \varepsilon_j \sum_{i=1}^{t+k} \varepsilon_i\right) = \sum_{j=1}^t \sum_{i=1}^{t+k} \mathbb{E}(\varepsilon_j \varepsilon_i) \\
&= \sigma^2 t,
\end{aligned}$$

Como $\text{Var}(Y_t) = \sigma^2 t$ entonces

$$\text{Corr}(Y_t, Y_{t+k}) = \frac{\sigma^2 t}{\sqrt{\sigma^2 t \sigma^2 (t+k)}} = \frac{1}{\sqrt{1+k/t}}. \quad (9.6)$$

Para t grande tal que k/t sea muy pequeño, se puede aproximar a cero y obtener que las autocorrelaciones en marchas aleatorias con tendencia son casi uno, $\rho(k) \approx 1$. Como la tendencia no determina el valor de la autocorrelación, la conclusión también es válida en el caso de marchas aleatorias sin tendencia. Las marchas aleatorias son series con fuertes autocorrelaciones.

Ejemplo 9.3.1. Si $Y_t \sim \text{ARMA}(1, 1)$ con $E(Y_t) = 0$ entonces

$$(1 - \varphi L)Y_t = (1 + \theta L)\varepsilon_t$$

y se puede comprobar que las autocorrelaciones de Y_t satisfacen:

$$\begin{aligned}
\rho(1) &= \frac{(1 + \varphi\theta)(\varphi + \theta)}{1 + \theta^2 + 2\varphi\theta} \\
\rho(k) &= \rho(k-1), \quad k \geq 2.
\end{aligned}$$

Si $\varphi \rightarrow 1$ obtenemos un $\text{ARIMA}(0, 1, 1) = \text{IMA}(1)$ y

$$\rho(1) \rightarrow \frac{(1 + \theta)(1 + \theta)}{1 + \theta^2 + 2\theta} = 1.$$

Por lo tanto $\rho(k) \equiv 1, \forall k \geq 1$.

Los procesos $\text{ARIMA}(p, 1, q)$ se comportan como marchas aleatorias en ciertos aspectos. Las marchas aleatorias son muy autocorrelacionadas. Los procesos ARIMA también muestran una fuerte autocorrelación. La fac muestral es muy persistente y cercana a uno.

Definición 9.3.1. Si la serie Y_t no estacionaria en covarianza es tal que ΔY_t es estacionario, se dice que Y_t es integrada de orden 1, $I(1)$. Si $\Delta^2 Y_t = \Delta(\Delta Y_t) = (1 - L)^2 Y_t$ es estacionario en covarianza se dice que integrada de orden 2, $I(2)$. La definición de proceso integrado de orden $d \geq 1$ es similar.

Luego un $ARIMA(p, 1, q)$ es un proceso $I(1)$ y un $ARMA(p, q)$ es un $I(0)$.

Definición 9.3.2. Una serie $(Y_t, t \in \mathbb{Z})$ se dice que tiene tendencia estocástica si es integrada de orden $d = 1$.

La marcha aleatoria con tendencia (9.4) es un ejemplo de tendencia estocástica, en contraposición a una serie con tendencia lineal $T_t = a + bt$, la cual se dice que tiene tendencia determinística. De igual manera, un proceso $ARIMA(p, 1, q)$ es un proceso con tendencia estocástica.

Una serie con tendencia estocástica presenta períodos de duración aleatoria durante los cuales crece linealmente.

La implementación de los modelos ARIMA asume series con tendencia estocástica. La metodología consiste en diferenciar la serie $d = 1, 2$ veces, hasta obtener otra serie estacionaria en covarianza, identificable por un modelo ARMA, y estimar los parámetros de este modelo.

9.4. Ejemplos de modelos ARIMA

A continuación se muestran algunos ejemplos de estimación de modelos ARIMA con la función `arima()`. En la sección siguiente se trata las pruebas de hipótesis para detectar la presencia de raíces unitarias, equivalente a detectar un ARIMA.

Ejemplo 9.4.1. *Análisis de la Tasa de cambio: USD/YEN, el precio de 1 US Dólar en Yen, ver Diebold [1999, pág. 221]. La serie es mensual desde 1973-01 hasta 1996-07. Por ejemplo, $Y_t = 107$ significa que en el mes t se pagaban 107 Yen por 1 USD. El análisis siguiente es con el logaritmo de esta tasa. En lo que sigue y denota el logaritmo de la tasa y dy la serie diferenciada $dy = \text{diff}(y, 1, 1)$.*

La Figura 9.2 (b) sugiere un proceso integrado $I(1)$ debido a que la fac es muy persistente; y las Figuras (d) y (e) sugieren un proceso estacionario para la serie diferenciada, por lo que un modelo ARMA puede ser apropiado. Utilizando la función $z = \text{autoarmafit}(dy)$, de la librería `timsac`, con la serie diferenciada se obtiene un modelo $ARMA(3,2)$.

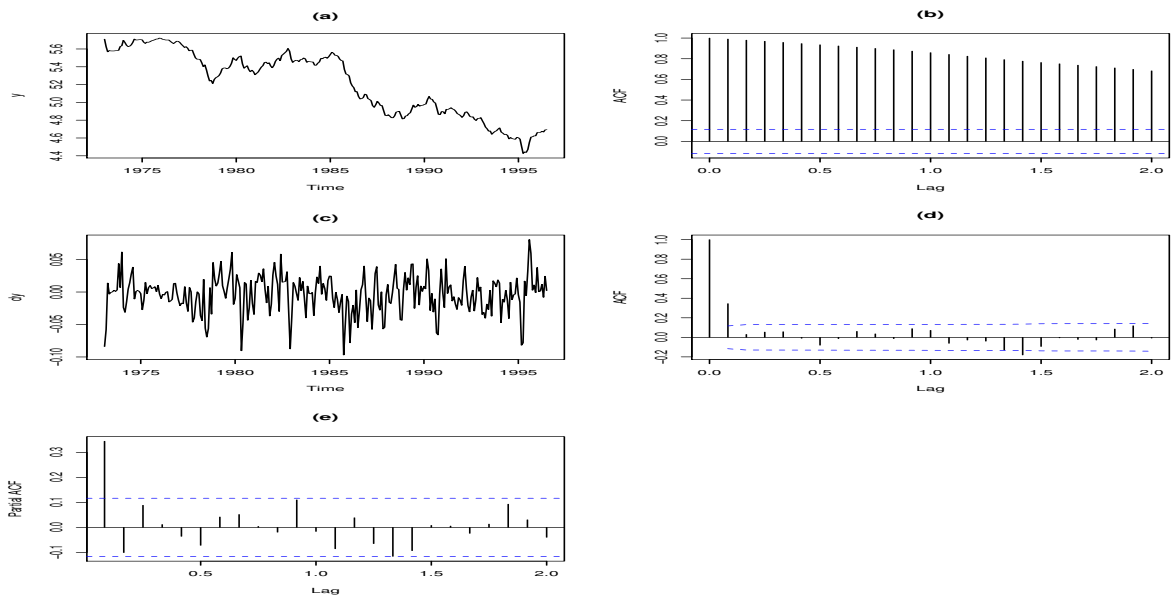


Figura 9.2: (a) : Serie log(Precio) del Yen en USD, (b) Su fac, (c): La serie diferenciada, (d),(e): fac y facp de la serie diferenciada

Utilizando la prueba Ljung-Box, se comprueba que los residuos de este modelo son Ruido Blanco. Este modelo es válido para pronosticar el logaritmo de la tasa de cambio, a corto plazo.

Hay que anotar que no es lo mismo estimar el modelo ARIMA(3,1,2) con la variable $\log(Y_t)$, que estimar el modelo ARMA(3,2) con la variable $\Delta \log(Y_t)$.

Los pronósticos a 19 períodos se muestran en la Figura 9.3, comparados con los pronósticos de un modelo con tendencia lineal y error estructural AR(2).

Ejemplo 9.4.2. Considerando Y_t la serie del logaritmo de la tasa de cambio USD/-Libra, es decir, el log del precio de 1 libra esterlina en usd, mensual, en el período 01/80-12/88, con 478 observaciones.

La trayectoria parece una marcha aleatoria sin tendencia, como se observa en la Figura 9.6. La primera diferencia parece estacionaria, casi ruido blanco. La fac de la primera diferencia confirma que es ruido blanco, obteniendo un modelo de marcha aleatoria con tendencia, dado por

$$(1 - L)Y_t = \mu + \varepsilon_t, \quad \mu = -0.0009.$$

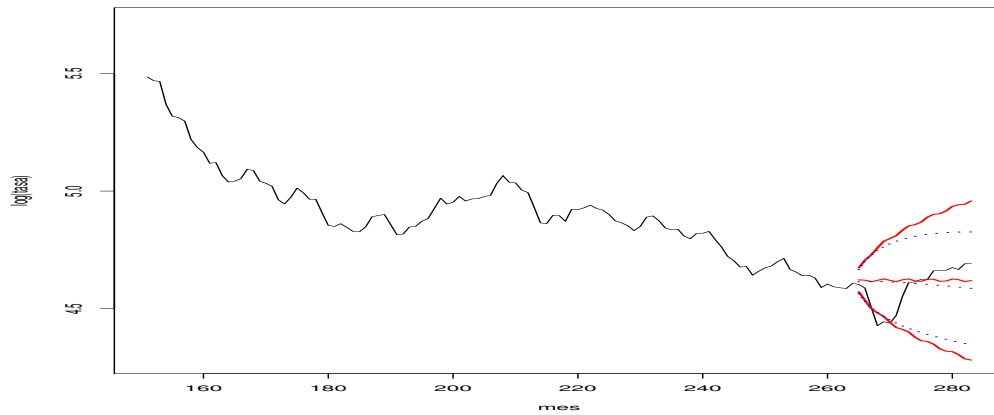
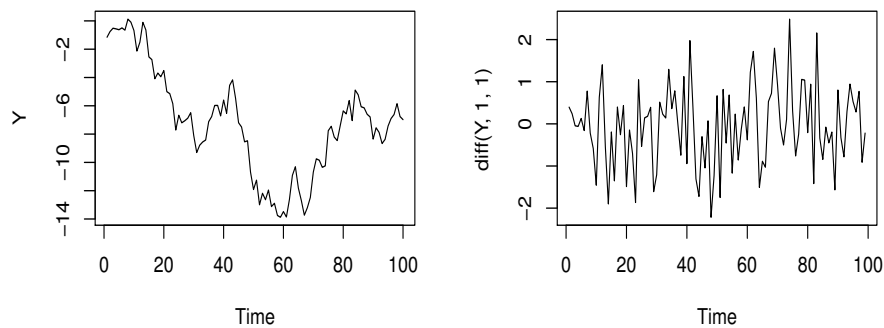


Figura 9.3: Pronósticos usd/yen con $ARIMA(3,1,2)$ (continua) y Tendencia Lineal+ $AR(2)$ (punteada)



(a) Serie USD/Libra

(b) Primera Diferencia de la Serie USD/Libra

Figura 9.4: Serie USD por Libra

Este modelo se denomina Modelo de Black-Scholes ó Modelo Log-Normal para el precio de un activo.

Definición 9.4.1. *La hipótesis de mercados eficientes asume que los retornos del precio de las acciones son $ARIMA(0,1,0)$.*

En ⁽²⁾: “En finanzas, la hipótesis del mercado eficiente (EFM) afirma que los precios de los activos que se negocian en un mercado financiero eficiente reflejan toda la información existente y se ajustan total y rápidamente a los nuevos datos que puedan surgir...Es por esta razón, que los precios de las acciones siguen una trayectoria incierta, es decir, el mejor pronóstico que existe para el precio de mañana es el precio de hoy. Esto es lo que se denomina Teoría de la marcha aleatoria...

Crash, burbujas y depresiones son compatibles con la hipótesis siempre que esta conducta no sea predecible...

Hay economistas, matemáticos y especuladores que no creen en la hipótesis del mercado eficiente. Piensan que hay razones poderosas para que existan ineficiencias: la lenta difusión de la información, el poder desigual de los distintos participantes en los mercados y la existencia de inversores profesionales que, al menos aparentemente, logran altos rendimientos...

Joseph Stiglitz, George Akerlof y Michael Spence ganaron el Premio Nobel de Economía por sus investigaciones sobre los mercados con información asimétrica y las inconsistencias, ineficiencias y fallos de los mercados financieros, derivados de la existencia de información asimétricas.”

En ⁽³⁾ : “Quienes está a favor de la EFM argumentan que no tiene sentido buscar acciones infravaloradas o tratar de predecir tendencias en el mercado a través de análisis fundamentales o técnicos”.

En un período de tiempo el precio puede aumentar por efecto de la demanda. Pero llega un momento en el cual la demanda se satura y los precios empiezan a oscilar, o a decrecer. Hay períodos de duración aleatoria en los que aparecen tendencias lineales, que se asumen tendencias estocásticas, y por tanto, se pueden describir por un modelo ARIMA(0,1,0).

Sin embargo, se dan casos en los que se encuentra un ARIMA(0,1,1), denominado IMA(1), ver, por ejemplo, Box et al. [1976], pag. 103.

Ejemplo 9.4.3. *En Diebold [1999, pág. 12] se considere la serie del Producto Interno Bruto en EUA entre 1869-1933, con el período 1934-1993 para comparación. Se examinan dos posibles modelos para esta serie.*

²https://es.wikipedia.org/wiki/Hipótesis_del_mercado_eficiente

³<https://www.investopedia.com/terms/e/efficientmarkethypothesis.asp>

1. Modelo con tendencia determinística lineal y errores $AR(2)$

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

$$\varepsilon_t = \varphi_1 \varepsilon_{t-1} + \varphi_2 \varepsilon_{t-2} + \eta_t, \quad \eta_t \sim RB(0, \sigma^2).$$

2. Modelo ARIMA

$$Y_t \sim ARIMA(1, 1, 0) \text{ con tendencia}$$

$$(1 - \varphi L)(1 - L)Y_t = \delta + \varepsilon_t.$$

Entre 1932 y 1933 hubo una grave recesión en EUA y los pronósticos con base en 1933 se hacen en una posición por debajo de la tendencia El modelo 1 produce

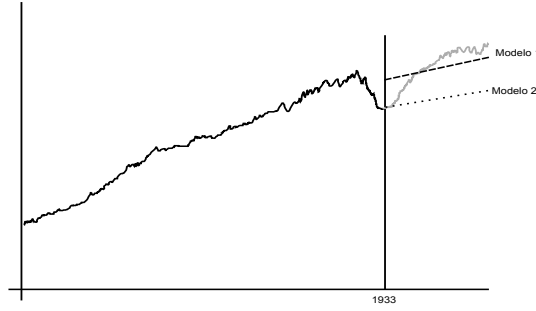


Figura 9.5: Pronósticos de la Serie PNB-USA.

mejores pronósticos que siguen la tendencia, en cambio, el modelo 2 no, pues subestima los valores.

Ejemplo 9.4.4. Considerando Y_t la serie del empleo en Canadá de periodicidad trimestral en el período 01/1962-04/1993, desestacionalizada.

Se propuso un modelo $AR(2)$. Pero usando la función `auto.arima`, se obtiene $Y_t \sim ARIMA(1, 1, 0)$ sin intercepto, es decir,

$$(1 - \varphi L)(1 - L)Y_t = \varepsilon_t, \quad \varepsilon_t \sim RB(0, \sigma^2), \quad (9.7)$$

con $\hat{\varphi} = 0.4598$, $s.e. = 0.0758$, $\hat{\sigma} = 2.068$, $BIC = 491.22$, $AIC = 485.41$, $MAPE = 1.0453$.

Si se ajusta un modelo $AR(2)$ a la serie, es decir,

$$Y_t = \mu + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t, \quad (9.8)$$

se obtiene $\hat{\varphi}_1 = 1.4505$, $\hat{\varphi}_2 = -0.4763$, $\hat{\sigma} = 2.022$, $\hat{\mu} = 97.498$, $AIC = 493.57$, $BIC = 505.22$, $MAPE = 1.0648$.

```
# programa en R
library(forecast) #
# Canadian employment index, seasonally adjusted, 1961:1-1994:4
# 136 obs
E = read.table("CAEMP.DAT", header = TRUE)
y = ts(E$caemp, frequency = 4, start = c(1961,01), end = c(1994,04))
f1 = arima(y, order=c(2,0,0))
f2 = arima(y, order=c(1,1,0))
```

La conclusión es que sería preferible utilizar el modelo (9.7) por tener un mejor ajuste. Pero puede utilizarse otro recurso.

La prueba Diebold-Mariano permite discriminar entre dos modelos de acuerdo a su capacidad predictiva dentro de la muestra.

La prueba de hipótesis Diebold-Mariano, en la función `dm.test` de la librería `forecast` compara las medias de los errores de pronóstico a h pasos, dentro de la muestra, para dos modelos.

La hipótesis nula es que las medias son iguales y la alterna que no lo son. Es posible utilizar hipótesis de una cola, con la especificación de que el primer modelo tiene menor error medio que el segundo.

En el Ejemplo anterior, se obtiene el valor-p para la prueba de dos colas: $p\text{-value} = 0.8412$. Luego no rechaza la hipótesis nula y ambos modelos tienen la misma capacidad predictiva.

```
#### programa en R
accuracy(f1)
accuracy(f2)
dm.test(residuals(f1), residuals(f2), h=1)

-----resultado
data: residuals(f1) residuals(f2)
DM = -0.2004, Forecast horizon = 1,
```

```
Loss function power = 2, p-value = 0.8412
alternative hypothesis: two.sided
```

9.5. Pruebas de Hipótesis para Raíces Unitarias

La fac y facp muestrales en presencia de Raíces Unitarias

En un proceso Y_t con raíz unitaria, por ejemplo, $ARIMA(p, 1, q)$, la fac muestral converge a cero con mucha lentitud, la facp muestral presenta un valor cercano a 1 en el rezago $k = 1$ y los demás valores son aproximadamente cero. Por ejemplo la Figura 9.6 siguiente se muestra la fac y la facp típicas de un proceso con raíz unitaria.

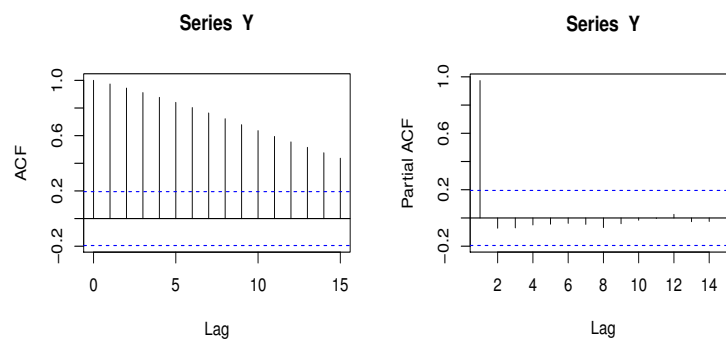


Figura 9.6: fac y facp muestrales de proceso con raíz unitaria

Necesidad de pruebas para detectar Raíces Unitarias

Hay que tener precaución con relación a tomar la decisión de diferenciar la serie para buscar un ARMA cuando se observan las características anteriores. Y es que es necesario aplicar una prueba de hipótesis para determinar si existe una raíz unitaria en lugar de una tendencia lineal determinística ó, incluso, un $AR(p)$ muy persistente. Las siguientes observaciones son pertinentes.

1. Diebold [1999, pág. 221]: “Si no hay raíz unitaria es conveniente utilizar modelos de niveles (componentes determinísticas) y es adecuando diferenciar

sólo en el caso de raíz unitaria; si la diferenciación es inadecuada puede ser dañina, incluso asintóticamente”.

2. Cuando una serie tiene raíz unitaria la serie es no estacionaria y los estimadores MCO no se distribuyen Normal.
3. Soares and Medeiros [2008, pág. 4]: “La mayoría de los artículos sobre pronósticos de demanda de energía toman diferencias sin hacer una prueba previa de raíz unitaria. Esto es un error grande cuando la tendencia es determinística, tomar diferencias introduce un componente MA no invertible el cual causa problemas serios de estimación.”

Pruebas de raíz unitaria en R

Existen en la literatura de series de tiempo varias pruebas para detectar raíz unitaria en una serie de tiempo. Y varias librerías en R que las implementan.

1. Prueba Dickey-Fuller aumentada. En las librerías:
 - `forecast`, función: `ndiffs`
 - `fUnitRoots`, función: `adfTest`
 - `tseries`, función: `adf.test`
 - `aTSA`, función: `adf.test`
 - `urca`, función: `ur.df`
2. Prueba KPSS, en las librería `aTSA`, función: `kpss.test`
3. Prueba PP, en las librería `aTSA`, función: `pp.test`

Estimadores MCO en caso de raíz unitaria en modelos AR(1)

La discusión en esta sección es para un $Y_t \sim ARIMA(p, 1, q)$. Suponga un proceso $AR(1)$, $Y_t = \varphi Y_{t-1} + \varepsilon_t$, en el cual $\varphi = 1$ pero φ se estima usando MCO. Este estimador $\hat{\varphi}$, tiene dos propiedades, ver Diebold [1999]

1. Superconsistencia

a) En el caso $\varphi = 1$ el estimador de MCO $\hat{\varphi}_T$ tiene la propiedad de que

$$T(\hat{\varphi}_T - 1) \xrightarrow{d} Z, \quad T \rightarrow \infty, \quad (9.9)$$

donde Z es una variable aleatoria no degenerada, y el símbolo \xrightarrow{d} denota la convergencia en distribución.

b) En el caso $|\varphi| < 1$, el estimador de MCO, $\hat{\varphi}_T$ cumple

$$\sqrt{T}(\hat{\varphi}_T - \varphi) \xrightarrow{d} Z \quad T \rightarrow \infty, \quad (9.10)$$

donde Z es variable aleatoria no degenerada.

como $\sqrt{T} < T$, se dice que la convergencia en el caso (9.9) es más rápida que en el caso (9.10), y a esto se denomina superconsistencia, es decir, el estimador de MCO, $\hat{\varphi}_T$ de una raíz unitaria es superconsistente.

2. Sesgo de $\hat{\varphi}$

Si $\hat{\varphi}_T$ es el estimador de MCO de φ entonces $E(\hat{\varphi}_T) < \varphi$ y el sesgo es mayor cuando $\varphi = 1$. El sesgo es $\varphi - \hat{\varphi}_T$ y crece si se considera tendencia en la serie. Aunque $\hat{\varphi}_T$ converge a φ , cuando $T \rightarrow \infty$, el sesgo puede ser apreciable en muestras no muy grandes.

9.6. Prueba Dickey-Fuller

La prueba Dickey-Fuller se basa en asumir que la serie se puede aproximar por un proceso AR(1) con tres variantes: media cero, media diferente de cero y tendencia lineal. Inicialmente se asume que Y_t sigue un modelo AR(1) de media cero y se procede a transformar este modelo de la siguiente manera.

$$\begin{aligned} Y_t &= \varphi_1 Y_{t-1} + \varepsilon_t, \\ \Leftrightarrow Y_t - Y_{t-1} &= (\varphi_1 - 1)Y_{t-1} + \varepsilon_t, \\ \Leftrightarrow \Delta Y_t &= \rho Y_{t-1} + \varepsilon_t. \end{aligned}$$

con $\rho = \varphi_1 - 1$. La existencia de una raíz unitaria $\varphi_1 = 1$ equivale a $\rho = 0$.

1. Prueba DF para el caso 1: $Y_t \sim AR(1)$ con media cero. Entonces

$$\Delta Y_t = \rho Y_{t-1} + \varepsilon_t. \quad (9.11)$$

La hipótesis nula es $H_0 : \rho = 0$ versus la alterna $H_a : \rho < 0$. El estadístico de la prueba se denota por τ y su distribución bajo H_0 permite calcular los valores críticos, por ejemplo, $\tau_{0.05}$ para un nivel de significación de 5 %.

El criterio de rechazo es $\hat{\tau} < \tau_{0.05}$, con $\hat{\tau}$ el valor calculado del estadístico. El valor p es $\mathbb{P}(\tau < \hat{\tau} | H_0)$. Se rechaza la nula a un nivel de significación dado si el valor p es menor que tal nivel. En R la prueba DF se implementa, para este caso, con la librería `urca`, por medio de la función `ur.df(y, type="none", lags=0)`.

2. Prueba DF para el caso 2: $Y_t \sim AR(1)$ con media diferente de cero. Entonces se modifica (9.11) así:

$$\Delta Y_t = \alpha + \rho Y_{t-1} + \varepsilon_t, \quad (9.12)$$

con la misma hipótesis del caso 1, $H_0 : \rho = 0$, $H_a : \rho < 0$. En R se programa la prueba para este caso con la instrucción: `ur.df(y, type="drift", lags=0)`.

3. Prueba DF para el caso 3: suponiendo que $Y_t \sim AR(1)$ con media diferente de cero y con tendencia lineal, entonces

$$\Delta Y_t = \alpha + \beta t + \rho Y_{t-1} + \varepsilon_t, \quad (9.13)$$

con la misma hipótesis del caso 1, $H_0 : \rho = 0$, $H_a : \rho < 0$. En R se programa la prueba para este caso con la instrucción: `ur.df(y, type="trend", lags=0)`.

9.7. Valores críticos de la prueba Dickey-Fuller

Con base en la Tabla siguiente, tomada de Fuller [2009]

Como $\hat{\tau}$ es mayor que cualquiera de los valores críticos, no rechaza la nula de raíz unitaria. Note que el valor p es en este caso 0.2924.

Recordamos la definición de error en las pruebas de hipótesis. El error tipo I se denomina “falso positivo”. Es el evento de rechazar H_0 siendo cierta. La probabilidad del error tipo I es α , el nivel de significación.

Tabla 9.1: Valores críticos de la prueba Dickey-Fuller

Sample size	Without trend		With trend	
	1 %	5 %	1 %	5 %
T = 25	-3.75	-3.00	-4.38	-3.60
T = 50	-3.58	-2.93	-4.15	-3.50
T = 100	-3.51	-2.89	-4.04	-3.45
T = 250	-3.46	-2.88	-3.99	-3.43
T = 500	-3.44	-2.87	-3.98	-3.42
T = ∞	-3.43	-2.86	-3.96	-3.41

Tabla 9.2: Ejemplo de resultado de la prueba Dickey-Fuller, con $n = 50$ observaciones

Estadístico Prueba	1 %	5 %	10 %
$\hat{\tau} = -1.987$	-3.580	-2.930	-2.600

		H_0 real	
		V	F
H_0 según prueba	V	OK	error II
	F	error I	OK

Tabla 9.3: Tabla de decisiones en una prueba de hipótesis

Recordamos la definición de potencia de una prueba, $1 - \beta$ donde $\beta = \text{Prob}(\text{error II})$. El error tipo II se denomina “falso negativo”. Es el evento de no rechazar H_0 siendo falsa. Luego la potencia es la probabilidad de rechazar H_0 siendo falsa.

Una prueba con baja potencia significa que es poco capaz de detectar la hipótesis alterna cuando esta es cierta.

Fijado el nivel α y el tamaño de la muestra, la potencia puede ser baja ó alta.

Las pruebas DF se sabe que tienen baja potencia. Por tanto, no rechazar la hipótesis nula de raíz unitaria no necesariamente significa que se deba asumir que existe. Como señala Diebold [1999, pág 220]: “Las pruebas de raíz unitaria tienen problemas de potencia y tamaño de muestra. De potencia porque las hipótesis alternas son muy cercanas a la hipótesis nula”. La presencia de cambios estructurales en la serie se sabe que disminuye la potencia de las pruebas DF, pero no se desarrollará este tema.

9.8. Prueba Dickey-Fuller Aumentada

La prueba aumentada de Dickey-Fuller es preferible a la prueba básica, pero no es solamente una prueba sino que requiere una estrategia de análisis para su aplicación. Como señalan Elder and Kennedy [2001, pag. 139]:

“ Un ingrediente crucial en esta prueba, que no se reconoce bien en los libros de texto, es que se requiere una estrategia de prueba, en oposición al simple cálculo del estadístico de la prueba. Esta estrategia es necesaria para determinar si un intercepto, un intercepto más una tendencia con el tiempo, o ninguna de las dos anteriores debería incluirse al correr la regresión para la prueba de raíz unitaria. Incluir demasiados regresores puede generar una pérdida de potencia, mientras que no incluir suficientes puede generar resultados sesgados hacia el no rechazo de la hipótesis nula...incluir intercepto, ó intercepto más tendencia, es necesario para permitir una representación de la hipótesis alterna que pueda competir contra la hipótesis nula de raíz unitaria.”

La estrategia para la prueba Dickey-Fuller Aumentada consiste en asumir que $Y_t \sim AR(p)$, y determinar cuál de los tres casos considerados anteriormente, con $p = 1$, determina una mejor aproximación a la serie original. Nótese que los casos a examinar dependen del orden autorregresivo p , por lo que la búsqueda para, por ejemplo, $p = 1, 2, 3$, requiere examinar $3 \times 3 = 9$ casos. La decisión se toma con base en el menor AIC.

Caso 1. Suponiendo que $Y_t \sim AR(p)$ con media cero, caso indicado por “none”, entonces se realizan las transformaciones siguientes. Defina

$$\rho_1 = \sum_{j=1}^p \varphi_j, \quad \rho_i = - \sum_{j=i}^p \varphi_j, \quad i = 2, \dots, p,$$

entonces

$$Y_t = \sum_{j=1}^p \varphi_j Y_{t-j} + \varepsilon_t, \quad t = 1, 2, \dots, T, \quad (9.14)$$

con $\varepsilon_t \sim iidN(0, \sigma^2)$, se transforma en (ver Diebold [1999, pág. 218])

$$\Delta Y_t = (\rho_1 - 1)Y_{t-1} + \sum_{j=2}^p \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \quad (9.15)$$

Si hay una raíz unitaria se cumple

$$\sum_{j=1}^p \varphi_j = 1 \Leftrightarrow \rho_1 = 1.$$

En este caso el modelo (9.15) equivale a

$$\Delta Y_t = \sum_{j=1}^{p-1} \rho_{j+1} \Delta Y_{t-j} + \varepsilon_t, \quad (9.16)$$

es decir un $AR(p-1)$ en la variable $Z_t = \Delta Y_t$.

Defina $\rho = \rho_1 - 1$, la hipótesis nula es $H_0 : \rho = 0$ y la alterna $H_a : \rho < 0$. No rechazar la nula significa que se detecta la raíz unitaria.

El estadístico DF, $\hat{\tau}$ tiene distribución asintótica tal que, si H_0 es cierta

$$\hat{\tau} \overset{a}{\sim} \frac{\int_0^1 W_t dW_t}{\int_0^1 W_t^2 dt} \quad (9.17)$$

donde el símbolo $\overset{a}{\sim}$ significa: “distribuido asintóticamente”, y W_t es el proceso estocástico Wiener estándar, ver ?, Cap. 6. Es la misma distribución que el estadístico DF en los demás casos considerados. Un hecho relevante es que la distribución de $\hat{\tau}$ bajo H_0 no depende de los parámetros φ_j, σ .

La hipótesis nula se rechaza si $\hat{\tau} < \tau_{0.05}$, y se concluye que no existe raíz unitaria. Es una prueba de cola izquierda.

“Así, los resultados del proceso $AR(1)$ se generalizan asintóticamente en forma directa a procesos de orden superior”. Diebold [1999, pág. 128]

Caso 2 Suponiendo $Y_t \sim AR(p)$ con media diferente de cero, caso indicado por “drift”. Con la misma notación del Caso 1 anterior, se realizan las transformaciones siguientes.

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

$$Y_t = \alpha + \varphi_1 Y_{t-1} + \sum_{j=2}^p \rho_j (Y_{t-j+1} - Y_{t-j}) + \varepsilon_t$$

$$\Delta Y_t = \alpha + \rho Y_{t-1} + \sum_{j=2}^p \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \quad (9.18)$$

donde $\alpha = \mu(1 - \sum_{j=1}^p \varphi_j)$, con los ρ_j definidos como en el Caso 1 anterior. La hipótesis nula es $H_0 : \rho = 0$ (existe raíz unitaria) y la alterna $H_a : \rho < 0$. Bajo $H_0 : \rho = 0$ el término α se anula. La distribución asintótica del estadístico DF es igual a la del caso $AR(1)$ con media, (9.12). La hipótesis nula se rechaza si $\hat{\tau} < \tau_{0.05}$, y se concluye que no existe raíz unitaria.

Caso 3 Suponiendo $Y_t \sim AR(p)$ con tendencia lineal, caso indicado por “trend”. En este caso se define

$$Y_t = a + bt + \sum_{j=1}^p \varphi_j (Y_{t-j} - a - b(t-j)) + \varepsilon_t \quad (9.19)$$

que se puede reordenar de la siguiente forma

$$\Delta Y_t = k_1 + k_2 t + \rho Y_{t-1} + \sum_{j=2}^p \rho_j \Delta Y_{t-j+1} + \varepsilon_t. \quad (9.20)$$

donde

$$\begin{aligned} k_1 &= a \left(1 - \sum_{i=1}^p \varphi_i \right) + b \sum_{i=1}^p i \varphi_i, \\ k_2 &= b \left(1 - \sum_{i=1}^p \varphi_i \right), \\ \rho_1 &= \sum_{i=1}^p \varphi_i, \quad \rho = \rho_1 - 1. \end{aligned}$$

La hipótesis nula es $H_0 : \rho = 0$ (existe raíz unitaria) y la alterna $H_a : \rho < 0$. Bajo $H_0 : \rho = 0$ se tiene que $k_2 = 0$, $k_1 = b \sum_{i=1}^p i \varphi_i$ y el estadístico DF tiene la misma distribución asintótica del estadístico en el caso $AR(1)$ con tendencia lineal, (9.13). La hipótesis nula se rechaza si $\hat{\tau} < \tau_{0.05}$, y se concluye que no existe raíz unitaria.

Ejemplo 9.8.1. Consideramos la serie de la tasa de cambio USD-Pound, libra inglesa, mensual, entre 1970-01 y 1997-08, como se muestra en la Figura 9.7. Se aplican los tres casos de la prueba ADF para existencia de raíz unitaria, con la librería `aTSA`. Los resultados conducen a no rechazar la nula, por lo que se detecta raíz unitaria, y es factible ajustar un modelo ARIMA.

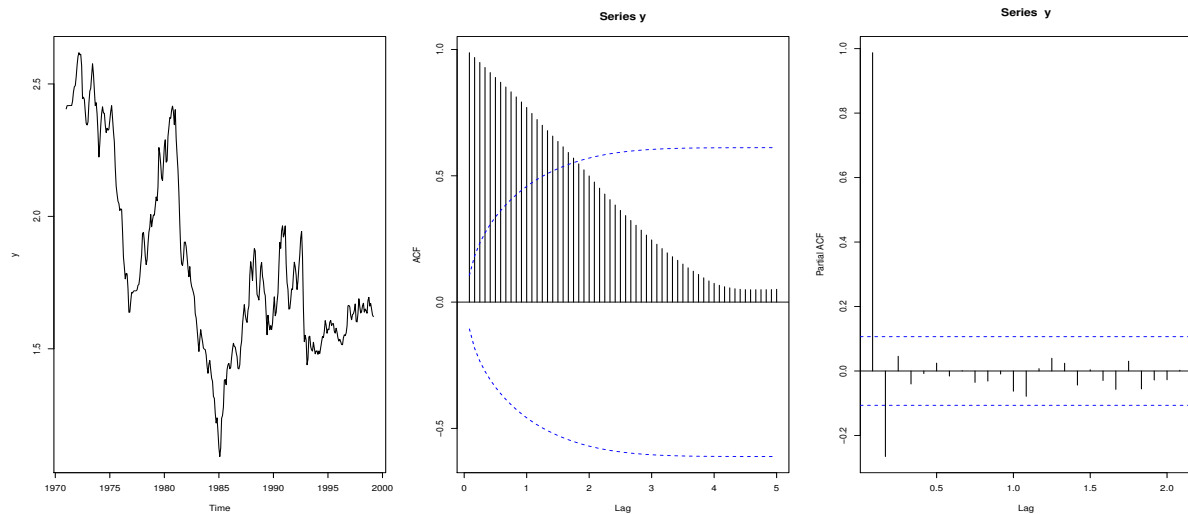


Figura 9.7: fac y facp muestrales de la serie USD/Pound

```
require(aTSA)
aTSA::adf.test(y)
#-----resultados
Augmented Dickey-Fuller Test
alternative: stationary
```

Type 1: no drift no trend

	lag	ADF	p.value
[1,]	0	-1.26	0.229
[2,]	1	-1.04	0.306
[3,]	2	-1.10	0.287
[4,]	3	-1.06	0.302
[5,]	4	-1.07	0.295
[6,]	5	-1.08	0.291

Type 2: with drift no trend

	lag	ADF	p.value
[1,]	0	-1.69	0.451
[2,]	1	-2.15	0.268
[3,]	2	-1.96	0.345
[4,]	3	-2.09	0.293

```

[5,]    4 -2.04    0.313
[6,]    5 -2.02    0.321
Type 3: with drift and trend
      lag    ADF p.value
[1,]    0 -1.57    0.758
[2,]    1 -2.46    0.382
[3,]    2 -2.10    0.533
[4,]    3 -2.35    0.429
[5,]    4 -2.25    0.472
[6,]    5 -2.20    0.494
-----

```

Note: in fact, `p.value = 0.01` means `p.value <= 0.01`

Ejemplo 9.8.2. Análisis detallado de la prueba ADF. A continuación se presenta una implementación de la estrategia de análisis mencionada en Elder and Kennedy [2001, pag. 139]. El objetivo es determinar cuál caso de los tres posibles con modelos $AR(p)$ es el que mejor aproxima la serie para así lograr una mayor potencia en la prueba.

Para esto se utiliza la librería `dynlm` que extiende la función de regresión lineal `lm()` al caso de series de tiempo, permitiendo incluir valores rezagados Y_{t-j} , con el comando `L(y, j)` y tendencia lineal $a + bt$ con el comando `trend(y)`, como variables predictoras. Por ejemplo, para el modelo de la ecuación (9.20), en la pag. 243, con $p = 3$

$$\begin{aligned}\Delta Y_t &= k_1 + k_2 t + \rho Y_{t-1} + \sum_{j=2}^3 \rho_j \Delta Y_{t-j+1} + \varepsilon_t, \\ &= k_1 + k_2 t + \rho Y_{t-1} + \rho_2 \Delta Y_{t-1} + \rho_3 \Delta Y_{t-2} + \varepsilon_t\end{aligned}$$

se programa, con `dy` la variable ΔY_t , y observando que

$$\begin{aligned}\Delta Y_{t-1} &= \Delta(L(Y_t)) = Y_{t-1} - Y_{t-2} = \\ L(Y_t - Y_{t-1}) &= L(\Delta(Y_t)).\end{aligned}$$

```

dy = diff(y, 1, 1)
reg.trend = dynlm(dy ~ trend(y) + L(y, 1) + L(dy, 1) + L(dy, 2)).

```

A continuación se programan las regresiones para $p = 1, 2$, excluyendo el caso “none” de media cero.

```
## estrategia de regresiones para la df aumentada
require(dynlm)

reg.drift1 = dynlm(dy ~ L(y,1))
reg.drift2 = dynlm(dy ~ L(y,1) + L(dy, 1))

reg.trend1 = dynlm(dy ~ trend(y) + L(y,1))
reg.trend2 = dynlm(dy ~ trend(y) + L(y,1) + L(dy, 1))

(c(AIC(reg.drift1),AIC(reg.drift2)))
(c(AIC(reg.trend1),AIC(reg.trend2)))

-----resultados de los valores AIC para los 4 modelos
-1202.376 -1240.019
-1203.215 -1245.147
```

Se detecta el caso `reg.trend1`, que corresponde a $p = 2$ como el de menor AIC.

Y la correspondiente prueba DF aumentada se programa con la función `ur.df()` de la librería `urca`

```
#-----programa la prueba ADF con p = 2 y trend

df.trend = ur.df(y = y, lags = 2, type = 'trend' )
summary(df.trend)
#-----resultados de la prueba DF aumentada
Test regression trend
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
Value of test-statistic is: -2.58 2.9804 3.448
Critical values for test statistics:
      1pct   5pct 10pct
tau3  -3.98 -3.42 -3.13
phi2   6.15  4.71  4.05
```


phi3 8.34 6.30 5.36

A partir de los resultados se concluye que no se rechaza la hipótesis de raíz unitaria ya que $\hat{\tau} > \tau_{0.05}$, el valor observado de la prueba -2.58 es mayor que el valor crítico al nivel de 5 %, -3.42, Luego, es válido diferenciar la serie. En el Ejemplo 9.4.1 se ajustó el modelo ARIMA(3,1,2) que produce buenos pronósticos del logaritmo de la tasa.

9.9. La prueba KPSS

La prueba KPSS (Kwiatkowski-Phillips-Schmidt-Shin), Kwiatkowski et al. [1992], aproxima la serie $Y_t, t = 1, 2, \dots, T$ con tres tipos de marchas aleatorias, en lugar de las tres transformaciones Dickey-Fuller de procesos AR(p) en las pruebas DF aumentadas.

$$Y_t = U_t + \epsilon_t, \text{ (none)} \quad (9.21a)$$

$$Y_t = \mu + U_t + \epsilon_t, \text{ (drift)} \quad (9.21b)$$

$$Y_t = \mu + \alpha t + U_t + \epsilon_t, \text{ (trend)} \quad (9.21c)$$

donde U_t es una marcha aleatoria sin tendencia de la forma $U_t = U_{t-1} + a_t$, con $a_t \sim iid N(0, \sigma^2)$. Y $\epsilon_t \sim iid N(0, \sigma_\epsilon^2)$, independiente de $\{a_t\}$.

La hipótesis nula es $H_0 : \sigma^2 = 0$, versus $H_a : \sigma^2 > 0$. No rechazar la nula equivale a asumir que la serie Y_t es estacionaria. Excepto por una tendencia lineal, en el caso “trend”. Entonces, la existencia de raíz unitaria es la hipótesis alterna.

El estadístico de la prueba es

$$K = \frac{1}{T^2} \frac{\sum_{j=1}^T S_j^2}{\hat{\sigma}^2} \quad (9.22)$$

donde $S_j^2 = \sum_{t=1}^j \hat{\epsilon}_t$, y $\hat{\sigma}^2$ es el estimador Newey-West de la varianza, con base en l rezagos, ver ⁽⁴⁾. Si el valor del estadístico es mayor que el valor crítico se rechaza la hipótesis nula. Es una prueba de cola derecha.

La prueba KPSS se puede programar en la librería aTSA con la función `kpss.test`. El parámetro `lag.short` es una variable lógica para el número de rezagos l en la

⁴https://en.wikipedia.org/wiki/Newey-West_estimator

fórmula del estimador Newey-West. Asumimos el valor por defecto `lag.short = TRUE`. También está en la librería `urca` en la función `ur.kpss`.

Ejemplo 9.9.1. *Considerando la serie de la tasa de cambio USD-Pound, libra inglesa, mensual, entre 1970-01 y 1997-08, como se muestra en la Figura 9.7. Se aplican los tres casos de la prueba KPSS para existencia de raíz unitaria. Los resultados conducen a no rechazar la nula, por lo que no se detecta raíz unitaria, y no sería conveniente utilizar un modelo ARIMA.*

```
#-----prueba KPSS
require(aTSA)
kpss.test(y, lag.short = TRUE, output = TRUE)
#-----resultados
KPSS Unit Root Test
alternative: nonstationary

Type 1: no drift no trend
lag  stat p.value
  4 0.0542    0.1
-----
Type 2: with drift no trend
lag  stat p.value
  4 0.0543    0.1
-----
Type 1: with drift and trend
lag  stat p.value
  4 0.0563    0.1
-----
Note: p.value = 0.01 means p.value <= 0.01
      : p.value = 0.10 means p.value >= 0.10
```

9.10. La prueba PP

La prueba PP, Phillips and Perron [1988], aproxima la serie Y_t con tres tipos de regresiones AR(1), similar a las pruebas DF básicas, pero sin realizar las transformaciones Dickey-Fuller.

$$Y_t = \rho Y_{t-1} + \epsilon_t, \text{ (none)} \quad (9.23a)$$

$$Y_t = \mu + \rho Y_{t-1} + \epsilon_t, \text{ (drift)} \quad (9.23b)$$

$$Y_t = \mu + \alpha t + \rho Y_{t-1} + \epsilon_t, \text{ (trend)} \quad (9.23c)$$

con $\epsilon_t \sim iidN(0, \sigma^2)$. La hipótesis nula es $H_0 : \rho = 1$, versus $H_0 : \rho < 1$. Luego, no rechazar la nula indica que existe raíz unitaria. El estadístico de la prueba es similar al estadístico Dickey-Fuller, $\tau = (\rho - 1)/\sigma$, pero la desviación σ se calcular con el estimador Newey-West de la varianza, con base en l rezagos.

La prueba PP es puede programar en la librería aTSA con la función `pp.test`. El parámetro `lag.short` es una variable lógica, como en la prueba KPSS, y se usará el valor por defecto TRUE. Los resultados muestran evidencia de existencia de raíz unitaria en los tres casos.

```
require(aTSA)
#-----prueba Phillips-Perron para la serie USD/Pound
aTSA::pp.test(y, lag.short = TRUE, output = TRUE)
#-----resultado
Phillips-Perron Unit Root Test
alternative: stationary

Type 1: no drift no trend
lag  Z_rho p.value
   5 -0.624   0.553
-----
Type 2: with drift no trend
lag  Z_rho p.value
   5 -5.88   0.41
-----
Type 3: with drift and trend
```

```
lag Z_rho p.value
5 -8.99 0.505
```

Note: p-value = 0.01 means p.value <= 0.01

9.11. Pruebas de No linealidad en Series de Tiempo

Las definiciones siguientes están en Priestley [1988, Chap. 3]. A partir de la representación de Wold, 8.4.1, pag. 198, todo proceso estacionario en covarianza X_t , sin componentes determinísticos satisface una identidad, en media cuadrática, de la forma (8.33)

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j},$$

donde $\varepsilon_t \sim iid(0, \sigma^2)$. Esta es la expresión general para una serie lineal causal.

Una definición general de serie de tiempo Y_t no lineal, causal, es que existe una función $g()$ tal que

$$Y_t = g(\varepsilon_t, \varepsilon_{t-2}, \dots). \quad (9.24)$$

Se trata de definir una expresión para este supuesto. Por medio de un argumento con base en la expansión de Taylor de la función $g()$ alrededor del punto $(0, 0, \dots)$ se puede extender la representación en el caso lineal a una más general, conocida como expansión de Volterra

$$Y_t = \psi_0 + \sum_{j=1}^{\infty} \psi_j \varepsilon_{t-j} + \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \psi_{j,k} \varepsilon_{t-j} \varepsilon_{t-k-j} + \dots + \varepsilon_t. \quad (9.25)$$

La expansión anterior muestra que en el conjunto de series que tengan la forma (9.25) se tienen efectos aleatorios mucho más complejos que en el caso lineal. Ejemplos simples de series no lineales son

$$Y_t = \alpha + \beta Y_{t-1} + \gamma \varepsilon_t \varepsilon_{t-1}. \quad (9.26)$$

$$Y_t = \alpha + \beta Y_{t-1} + \gamma Y_{t-1} \varepsilon_{t-1} + \varepsilon_t. \quad (9.27)$$

$$Y_t = 2.0 + 0.2 \varepsilon_{t-1} + 0.5 \varepsilon_{t-1} \varepsilon_{t-2} + 1.3 \varepsilon_{t-1} \varepsilon_{t-3} + \varepsilon_t. \quad (9.28)$$

En todos los casos se asume $\epsilon_t \sim iid N(0, \sigma^2)$. Interesa implementar algunas pruebas para detectar no linealidad.

La prueba Teräsvirta asume que la serie Y_t sigue un modelo no lineal tipo red neuronal autorregresiva $Y_t \sim NNAR(p, 0, q)$, con p nodos de entrada y q nodos en la capa oculta, definida en (5.23), más una componente AR(p) estacionaria en covarianza

$$\Phi_p(L)Y_t = \sum_{j=1}^q \alpha_j f \left(\beta_{0,j} + \sum_{k=1}^p \beta_{j,k} Y_{t-k} \right) + \varepsilon_t,$$

donde $f(x) = (1 + e^x)^{-1}$. La hipótesis nula es

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_q = 0. \quad (9.29)$$

$$H_a : \text{no } (H_0) \quad (9.30)$$

El estadístico es del tipo Multiplicador de Lagrange, ML, ⁽⁵⁾, pero es difícil de describir, ver Teräsvirta et al. [1993]. Se puede realizar con la librería `tseries`, con la función `terasvirta.test(x)`. El ejemplo siguiente está tomado de las ayudas de esta función.

```
#-----
require(tseries)
n= 300;x=double(n);
#---genera una serie no lineal: H0 es falsa
x[1] <- 0.0
for(i in (2:n)) {
x[i] <- 0.4*x[i-1] + tanh(x[i-1]) + rnorm(1, sd=0.5)}
x <- as.ts(x)
plot(x)
terasvirta.test(x)
#---resultado
Teraesvirta Neural Network Test
data:  x
X-squared = 26.269, df = 2, p-value = 1.976e-06
```

Una prueba similar es la prueba White, Lee et al. [1993], en la librería `tseries`.

⁵https://en.wikipedia.org/wiki/Score_test

```
#-----
require(tseries)
White Neural Network Test
data:  x
X-squared = 29.2, df = 2, p-value = 4.565e-07
```

La librería `nonlinearAnalysis` implementa en la función `nonlinearityTest` seis pruebas de no linealidad, incluyendo las pruebas Teraesvirta y White.

9.12. Modelos híbridos ARIMA - Red Neuronal

Una aplicación de las prueba de no linealidad es la propuesta de modelos híbridos ARIMA-NNAR, introducida por Zhang [2003].

En este trabajo Zhang [2003], se propone que si $Y_t \sim ARIMA(p, 1, q)$, definido en (9.1)

$$\begin{aligned}\Phi_p(L)(1-L)Y_t &= \mu + \Theta_q(L)\varepsilon_t, \\ \varepsilon_t &\sim RB(0, \sigma^2), \quad t = 1, 2, \dots, T,\end{aligned}$$

entonces la serie de residuos ARIMA $\hat{\varepsilon}_t$, podría tener una dinámica no lineal, que no se revela con las pruebas Ljung-Box de incorrelación. Por lo que es necesario utilizar pruebas de no linealidad previamente. En caso de no linealidad, se podría modelar con una red neuronal, por ejemplo una NNAR(p,0,q).

$$\hat{\varepsilon}_t \sim NNAR(p, 0, q),$$

de tal forma que los pronósticos se pueden combinar como pronósticos ARIMA más pronósticos NNAR,

$$\hat{Y}_{T+j} + \hat{\varepsilon}_{T+j}, j = 1, 2, \dots, h.$$

Ejemplo 9.12.1. *El código siguiente implementa un modelo ARIMA para la serie de la tasa de cambio USD/Pound, y los residuos los ajusta con una red neuronal auto regresiva NNAR. Los pronosticos a 30 meses se muestran en la Figura 9.8.*

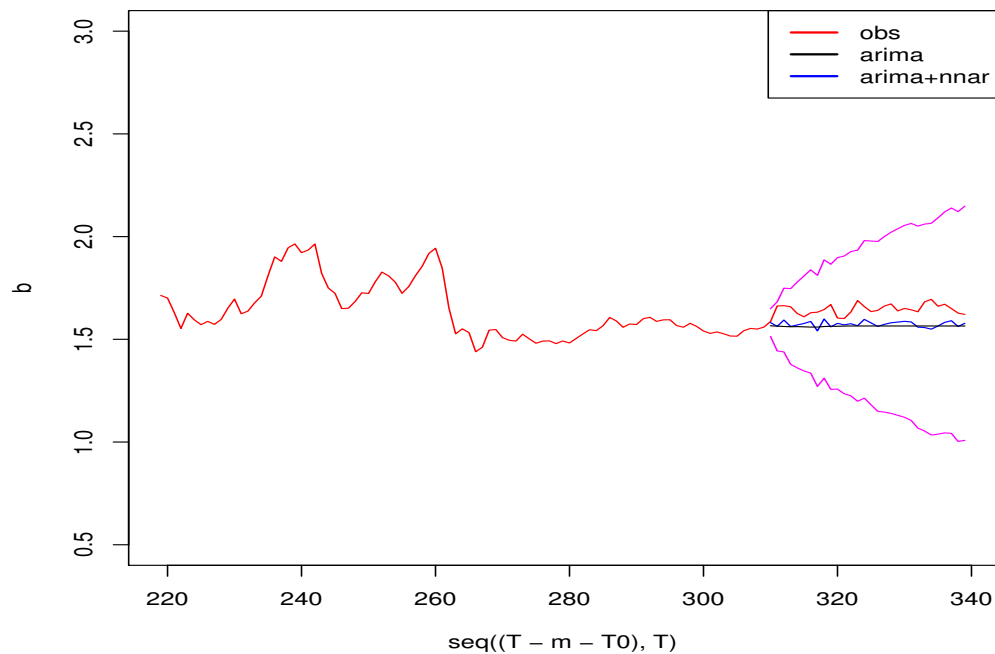


Figura 9.8: Pronósticos con modelo híbrido ARIMA-NNAR de la serie USD/Pound

```
#-----
auto.arima(yi)
ARIMA(0,1,1)(0,0,1)[12]
mod2 = stats::arima(yi,order=c(0,1,1),
seasonal=list(order=c(0,0,1),period=12))
require(lmtest)
coeftest(mod2)
#-----residuos
at=ts(mod2$residuals,frequency=12)
#-----pruebas de no linealidad
terasvirta.test(at)
white.test(at)
#-----grafica
ts.plot(at)
#-----ajuste NNAR
```

```

mod3=nnetar(at)
print(mod3)
#-----pronosticos
pr.nnar = forecast(mod3,h=m)$mean
pr.nnar = ts(pr.nnar,frequency=12,start=c(1996,10))
pr.arima = predict(mod2,m)$pred
sde = predict(mod2,m)$se
pr.arima = ts(pr.arima,frequency=12,start=c(1996,10))
pr = pr.nnar+pr.arima

```

9.13. Combinación de pronósticos

De la página de Rob Hyndman ⁽⁶⁾: “It has been well-known since at least 1969, when Bates and Granger wrote their famous paper on “The Combination of Forecasts”, that combining forecasts often leads to better forecast accuracy. ”

La librería `forecastHybrid` realiza pronósticos con base en los modelos `auto.arima`, `ets`, `nnetar`, `stlm`, `tbats`, y los combina como una suma ponderada para producir un solo pronóstico.

Ejemplo 9.13.1. *El código siguiente implementa una combinación de pronósticos a 30 meses para la serie de la tasa de cambio USD/Pound. Los pronosticos a 30 meses se muestran en la Figura 9.9.*

```

require(forecastHybrid)
mod1 = hybridModel(yi, weights="equal")
#-----pronosticos
B=forecast(mod1,h=m)
Bf = B$pointForecasts
colnames(Bf)
[1] "auto.arima" "ets"          "thetam"      "nnetar"
[5] "stlm"       "tbats"
p1 = apply(Bf,1,mean)
#-----graficar

```

⁶<https://robjhyndman.com/hyndsight/forecast-combinations/>


```

T0 = 90
b = c(yi[(T-m-T0):(T-m)],yf)
T1 = length(yi)
tt = seq((T-m+1),T)
par(mfrow=c(1,1))
plot(seq((T-m-T0),T),b,
type='l', col=2,ylim=c(0.5,3.0))
lines(tt,p1, col='blue')
lines(tt, p1+1.64*sde, col='magenta')
lines(tt, p1-1.64*sde, col='magenta')

```

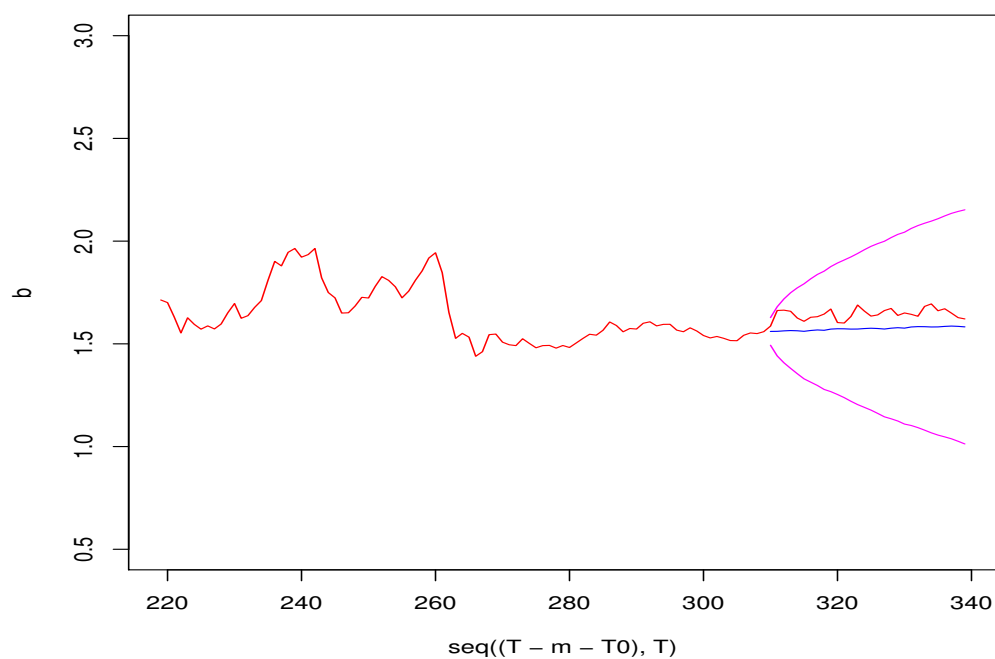


Figura 9.9: combinacion de pronósticos con 6 modelos para la serie USD/Pound