

Proyecto 3: Series de Tiempo

Clemente Ferrer, Rodrigo Pizarro

PARTE TEÓRICA

1. Se dice que un proceso X_t es un proceso ARUMA(p, d, q) si puede ser descrito por la siguiente ecuación:

$$\Phi(B)\lambda(B)(X_t - \mu) = \Theta(B)\epsilon_t,$$

donde

$$\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p \quad \text{y} \quad \Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

son polinomios con raíces fuera del disco unitario,

$$\lambda(B) = 1 - \lambda_1 B - \dots - \lambda_d B^d$$

es un polinomio con raíces sobre el disco unitario y ϵ_t es un ruido blanco con varianza σ^2 .

- a. Demuestre que $\Phi(B)\lambda(B)X_t = \Phi(1)\lambda(1)\mu + \Theta(B)\epsilon_t$.

Solución. Note que de la definición del proceso ARUMA(p, d, q), se tiene

$$\begin{aligned}\Phi(B)\lambda(B)(X_t - \mu) &= \Theta(B)\epsilon_t \\ \Phi(B)\lambda(B)(X_t) &= \Phi(B)\lambda(B)\mu + \Theta(B)\epsilon_t.\end{aligned}$$

Como μ es constante y no depende de t ,

$$\begin{aligned}\lambda(B)\mu &= (1 - \lambda_1 B - \dots - \lambda_d B^d)\mu \\ &= \mu - \lambda_1 B\mu - \dots - \lambda_d B^d\mu \\ &= \mu - \lambda_1 \mu - \dots - \lambda_d \mu \\ &= \mu(1 - \lambda_1 - \dots - \lambda_d) \\ &= \mu\lambda(1),\end{aligned}$$

y como es analogo para $\Phi(B)$, considerando ahora la constante como $\mu\lambda(1)$, se tendrá que $\Phi(B)\lambda(1)\mu = \Phi(1)\lambda(1)\mu$ es decir

$$\Phi(B)\lambda(B)(X_t) = \Phi(1)\lambda(1)\mu + \Theta(B)\epsilon_t$$

□

- b. Muestre un caso en que $\Phi(1)\lambda(1)\mu = 0$ para $\mu \neq 0$.

Solución. Basta considerar $\lambda(B) = (1 - B)$ y $\Phi(B) = (1 - B - B^2)$.

□

- c. Considere el proceso ARUMA(2, 2, 2) descrito por la ecuación

$$X_t + 0,3X_{t-1} + 0,6X_{t-2} + 0,3X_{t-3} - 0,4X_{t-4} = 9 + \epsilon_t + 0,5\epsilon_{t-1} + 0,9\epsilon_{t-2}.$$

Escriba el proceso en la forma (1).

Solución. Notamos que podemos escribir el proceso como

$$(1 + 0,3B + 0,6B^2 + 0,3B^3 - 0,4B^4)X_t = 9 + (1 + 0,5B + 0,9B^2)\epsilon_t$$

Podemos factorizar el polinomio que acompaña a X_t utilizando el teorema de la raíz racional, ya que si multiplicamos el polinomio por 10, todos los coeficientes son enteros (el polinomio en cuestión quedaría $10 + 3B + 6B^2 + 3B^3 - 4B^4$). Si probamos por inspección todas las raíces racionales dadas por el teorema, hallamos fácilmente que las raíces racionales son 2 y $-1,25$. Si factorizamos ahora el polinomio original mediante división sintética obtenemos que

$$1 + 0,3B + 0,6B^2 + 0,3B^3 - 0,4B^4 = -0,4(B - 2)(B + 1,25)(B^2 + 1)$$

Luego $\Phi(B) = -0,4(B - 2)(B + 1,25) = 1 + 0,3B + 0,4B^2$ y $\lambda(B) = (B^2 + 1)$. Por otra parte, $\Theta(B) = (1 + 0,5B + 0,9B^2)$. Además $\mu\Theta(1)\lambda(1) = 9 \implies \mu = 5$. (pues $\Theta(1)\lambda(1) = 1,8$). Con esto, hemos descrito el proceso dado en la forma (1). □

- d. Definamos el proceso $Y_t = (1 + B^2)(X_t - 5)$. ¿Es Y_t estacionario?

Solución. Note que el proceso en cuestion es un ARMA(2, 2) con polinomio con raíces fuera del disco unitario, por ende, es estacionario. \square

2. Considere una serie de tiempo Y_t que presenta una tendencia lineal. La serie diferenciada ∇Y_t puede ser modelada correctamente por un modelo $W_t = 0,1 + Z_t - 0,5Z_{t-1} - 0,5Z_{t-2}$, donde $Z_t \sim WN(0, 4)$. Proponga un modelo apropiado para Y_t .

Solución. Como Y_t tiene una tendencia lineal y además la diferencia puede ser modelada por W_t , entonces podemos proponer que

$$Y_t = ct + \sum_{i=0}^m c_i Z_{t-i}$$

para algún m a determinar y ciertas constantes $c_i \in \mathbb{R}$. Note que como existe un termino Z_{t-2} , es natural pensar que $m = 1$, pues así al diferenciar aparecerá un término Z_{t-2} . Si $m > 1$, habría un termino Z_{t-3} o con mas dependencia del pasado, lo que no tendría sentido bajo W_t . De este modo, podemos proponer que $Y_t = ct + c_0 Z_t + c_1 Z_{t-1}$. Así, se tendrá

$$\begin{aligned}\nabla Y_t &= ct + c_0 Z_t + c_1 Z_{t-1} - (c(t-1) + c_0 Z_{t-1} + c_1 Z_{t-2}) \\ &= c + c_0 Z_t + (-c_0 + c_1) Z_{t-1} - c_1 Z_{t-2}\end{aligned}$$

Como ∇Y_t está modelada correctamente por W_t , entonces

$$0,1 + Z_t - 0,5Z_{t-1} - 0,5Z_{t-2} = c + c_0 Z_t + (-c_0 + c_1) Z_{t-1} + c_1 Z_{t-2}$$

Igualando coeficientes, obtenemos que $c = 0,1$, $c_0 = 1$ y como $-c_0 + c_1 = -0,5$, entonces $c_1 = 0,5$

Así, concluimos que $Y_t = 0,1t + Z_t + 0,5Z_{t-1}$. \square

3. Considere un proceso AR(1) descrito por la ecuación $X_t = 0,5X_{t-1} + \epsilon_t$ donde $\epsilon_t \sim WN(0, 1)$. Para abordar el problema de una observacion perdida, X_t , las siguientes alternativas han sido propuestas:

- $\hat{X}_t^{(1)} = 0$.
- $\hat{X}_t^{(2)} = (X_{t+1} + X_{t-1}) / 2$.
- $\hat{X}_t^{(3)}$ es el mejor predictor lineal entre X_{t+1} y X_{t-1} .

Compare los tres predictores $\hat{X}_t^{(k)}$, $k = 1, 2, 3$, del valor perdido X_t en función del error cuadrático medio de predicción.

Solución. Para resolver este problema, notemos que nuestro proceso es AR(1) con polinomio asociado $(1 - 0,5B)$. La raíz de este polinomio es $B = 2$ y por ende esta fuera del disco unitario. Con esto en mente, por lo visto en cátedra podemos deducir que el proceso es estacionario. Más aún, por los resultados conocidos utilizados inclusive en proyectos anteriores, se tiene $E(X_t) = 0$ y $C(h) = \frac{4 \cdot 2^{-h}}{3}$, donde hemos usado la

relación para este tipo de procesos dada por $C(h) = \frac{\sigma^2 \phi^h}{1 - \phi^2}$. Calculemos a continuación el error cuadrático medio de cada una de las alternativas:

- Por lo visto en cátedra, lo único que debemos calcular es

$$\begin{aligned}\text{MSE} &= E[(X_t - \hat{X}_t^{(1)})^2] \\ &= E[X_t^2] \\ &= C(0) \\ &= 4/3.\end{aligned}$$

- Para calcular el MSE, utilizaremos el hecho de que $E[X_{t+h}^2] = C(0)$ para todo h y que $E[X_{t+j}X_{t+h}] =$

$C(|h-j|)$ para todo h, j , en conjunto con la linealidad de la esperanza:

$$\begin{aligned} \text{MSE} &= E[(X_t - \hat{X}_t^{(2)})^2] \\ &= E\left[\left(X_t - \left(\frac{X_{t+1} + X_{t-1}}{2}\right)\right)^2\right] \\ &= E\left[X_t^2 + \left(\frac{1}{2}\right)^2 X_{t+1}^2 + \left(\frac{1}{2}\right)^2 X_{t-1}^2 - 2 \cdot \left(\frac{1}{2}\right) X_t X_{t+1} - 2 \cdot \left(\frac{1}{2}\right) X_t X_{t-1} + 2 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) X_{t+1} X_{t-1}\right] \\ &= C(0) + \left(\frac{1}{2}\right)^2 C(0) + \left(\frac{1}{2}\right)^2 C(0) - 2 \left(\frac{1}{2}\right) C(1) - 2 \left(\frac{1}{2}\right) C(1) + 2 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) C(2) \\ &= C(0) \left(1 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2\right) - 2C(1) \left(\frac{1}{2} + \frac{1}{2}\right) + C(2) \cdot 2 \left(\frac{1}{2}\right) \left(\frac{1}{2}\right) \\ &= 5/6, \end{aligned}$$

donde el resultado se obtiene reemplazando $C(0) = 4/3$, $C(1) = 2/3$ y $C(2) = 1/3$. (Escribimos de esa manera la expresión, pues nos ayudará para el siguiente ítem)

- iii) Primero, debemos calcular el mejor predictor lineal, para eso, debemos minimizar la siguiente expresión:
 $\text{MSE}(\alpha, \beta) = E[(X_t - \alpha X_{t+1} - \beta X_{t-1})^2]$
 donde α, β son constantes a determinar. Note que la expresión del ítem dos, es equivalente a la expresión del ítem ii), solo que reemplazando los $1/2$ por α y β , es decir, buscamos minimizar

$$\begin{aligned} \text{MSE}(\alpha, \beta) &= C(0) (1 + \alpha^2 + \beta^2) - 2C(1) (\alpha + \beta) + C(2) \cdot 2\alpha\beta \\ &= \frac{4}{3} (1 + \alpha^2 + \beta^2) - \frac{4}{3} (\alpha + \beta) + \frac{2}{3} \alpha\beta \end{aligned}$$

Para minimizar, calculamos $\nabla \text{MSE} = 0$ y obtenemos dos ecuaciones:

$$\begin{aligned} \frac{2}{3}(4\alpha + \beta - 2) &= 0 \\ \frac{2}{3}(\alpha + 4\beta - 2) &= 0. \end{aligned}$$

Resolviendo, se obtiene que $\alpha = \beta = \frac{2}{5}$. Luego, para calcular el MSE basta reemplazar los valores en la expresión, es decir

$$\begin{aligned} \text{MSE} &= C(0) \left(1 + \left(\frac{2}{5}\right)^2 + \left(\frac{2}{5}\right)^2\right) - 2C(1) \left(\frac{2}{5} + \frac{2}{5}\right) + C(2) \cdot 2 \left(\frac{2}{5}\right) \left(\frac{2}{5}\right) \\ &= 4/5. \end{aligned}$$

En conclusión, el mejor estimador basado en el MSE es el tercero. □

4. Considere un modelo AR(2) descrito por la ecuación $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = Z_t$, donde $Z_t \sim WN(0, \sigma^2)$.
a. Determine el mejor predictor lineal (minimizando el error cuadrático medio de predicción) de X_{t+2} basado en la información $X_s, s \leq t$.

Solución. Como sabemos del libro guía, en general si el proceso tiene media distinto de 0, no siempre es posible escoger el mejor predictor lineal con los métodos estandar y darles una forma explícita sencilla. Para efectos del ejercicio, asumiremos que el proceso tiene media 0, pero el proceso sería similar si este no fuese el caso.

Buscaremos el mejor predictor lineal, cuyos coeficientes los denotaremos como $\hat{\phi}_1, \hat{\phi}_2$. Para ello, procedemos mediante el desarrollo estandar visto en cátedra, definiendo el vector de los parámetros en función

de la matrices y vector con las funciones de covarianza (todos ellos predichos), es decir, recurrimos a la relación $\hat{\Phi}_2 = \hat{T}_2^{-1} \hat{\gamma}_2$, con $T_2 = \begin{pmatrix} C(0) & C(1) \\ C(1) & C(0) \end{pmatrix}$ y $\gamma_2 = (C(1), C(2))^t$.

Esta relación nos permite obtener las relaciones

$$\begin{aligned}\hat{X}_{n+2} &= \hat{\phi}_1 \hat{X}_{n+1} + \hat{\phi}_2 \hat{X}_n \\ \hat{X}_{n+1} &= \hat{\phi}_1 \hat{X}_n + \hat{\phi}_2 \hat{X}_{n-1}.\end{aligned}$$

Si reemplazamos la segunda en la primera, obtenemos lo deseado, es decir

$$\hat{X}_{n+2} = \hat{\phi}_1(\hat{\phi}_1 \hat{X}_n + \hat{\phi}_2 \hat{X}_{n-1}) + \hat{\phi}_2 \hat{X}_n = (\hat{\phi}_1^2 + \hat{\phi}_2) \hat{X}_n + (\hat{\phi}_1 \hat{\phi}_2) \hat{X}_{n-1}$$

□

b. Determine el error cuadrático medio de predicción del predictor en a).

Solución. Por motivos de escritura, omitiremos la notación de \hat{X} dejándolo exclusivamente como X :

$$\begin{aligned}\text{MSE} &= E[(X_{n+2} - (\phi_1^2 + \phi_2)X_n - (\phi_1\phi_2)X_{n-1})^2] \\ &= C(0) + (\phi_1^2 + \phi_2)^2 C(0) + (\phi_1\phi_2)^2 C(0) - 2(\phi_1^2 + \phi_2)C(2) - 2(\phi_1\phi_2)C(3) + 2(\phi_1^2 + \phi_2)(\phi_1\phi_2)C(1) \\ &= C(0)(1 + (\phi_1^2 + \phi_2)^2 + (\phi_1\phi_2)^2) + 2C(1)(\phi_1^2 + \phi_2)(\phi_1\phi_2) - 2(\phi_1^2 + \phi_2)C(2) - 2(\phi_1\phi_2)C(3).\end{aligned}$$

□

5. Calcule la densidad espectral asociada a un proceso AR(2) estacionario. Grafique dicha función para valores fijos de los parámetros del modelo.

Solución. Definimos X_t como un proceso AR(2), es decir,

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = \epsilon_t, \quad \phi_1, \phi_2 \in \mathbb{R}.$$

Donde $\epsilon_t \sim \text{WN}(0, \sigma^2)$. Por la igualdad, se tendrá que el proceso de la izquierda tendrá la misma función de autocovarianza que el de la derecha y como $C_\epsilon(h) = 0$ si $h \neq 0$ y $C_\epsilon(h) = \sigma^2$ si $h = 0$, entonces podemos calcular su función de covarianza como sigue (visto en ayudantía):

$$\begin{aligned}C_\epsilon(h) &= \mathbb{E}[(X_{t+h} - \phi_1 X_{t+h-1} - \phi_2 X_{t+h-2})(X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2})] \\ &= C_x(h)(1 + \phi_1^2 + \phi_2^2) + C_x(h+1)(-\phi_1 + \phi_1\phi_2) + C_x(h-1)(-\phi_1 + \phi_1\phi_2) - \phi_2[C_x(h+2) - C_x(h-2)].\end{aligned}$$

donde C_x es la función de autocovarianza del proceso X_t . Más aún, también vimos en ayudantía que se tiene la siguiente propiedad

$$C_x(h) = \int_{-1/2}^{1/2} f_x(\omega) e^{2\pi i \omega h} d\omega \implies C_x(h+k) = \int_{-1/2}^{1/2} f_x(\omega) e^{2\pi i \omega (h+a)} d\omega = \int_{-1/2}^{1/2} f_x(\omega) e^{2\pi i \omega a} \cdot e^{2\pi i \omega h} d\omega$$

Lo anterior lo utilizaremos por el hecho de que $C_\epsilon(h)$ depende de desplazamientos de $C_x(h)$.

Luego, reemplazando en la función de covarianza del ruido blanco la definición de $C_x(h)$, simplificando, y utilizando la fórmula de Euler se puede obtener

$$\begin{aligned}C_\epsilon(h) &= \int_{-1/2}^{1/2} [(1 + \phi_1^2 + \phi_2^2) + (\phi_1\phi_2 - \phi_1)(e^{2\pi i \omega} + e^{-2\pi i \omega}) - \phi_2(e^{4\pi i \omega} + e^{-4\pi i \omega})] f_x(\omega) e^{2\pi i \omega h} d\omega \\ &= \int_{-1/2}^{1/2} [(1 + \phi_1^2 + \phi_2^2) + 2(-\phi_1 + \phi_1\phi_2) \cos(2\pi \omega) - 2\phi_2 \cos(4\pi \omega)] f_x(\omega) e^{2\pi i \omega h} d\omega\end{aligned}$$

Para concluir, sabemos que la función de densidad espectral del ruido blanco es conocida, y que vale $f_\epsilon(\omega) = \sigma^2$. Utilizando esto en conjunto con la fórmula que dedujimos anteriormente, se puede plantear la siguiente igualdad:

$$\int_{-1/2}^{1/2} \sigma^2 e^{2\pi i \omega h} d\omega = \int_{-1/2}^{1/2} [(1 + \phi_1^2 + \phi_2^2) + 2(\phi_1\phi_2 - \phi_1) \cos(2\pi \omega) - 2\phi_2 \cos(4\pi \omega)] f_x(\omega) e^{2\pi i \omega h} d\omega$$

Por lo tanto, igualando las cantidades de cada integral se obtiene

$$\sigma^2 = [(1 + \phi_1^2 + \phi_2^2) + 2(-\phi_1 + \phi_1\phi_2)\cos(2\pi\omega) - 2\phi_2\cos(4\pi\omega)] f_x(\omega).$$

Finalmente, esto nos permite despejar la densidad espectral del proceso X_t :

$$f_x(\omega) = \frac{\sigma^2}{[(1 + \phi_1^2 + \phi_2^2) + 2(-\phi_1 + \phi_1\phi_2)\cos(2\pi\omega) - 2\phi_2\cos(4\pi\omega)]}.$$

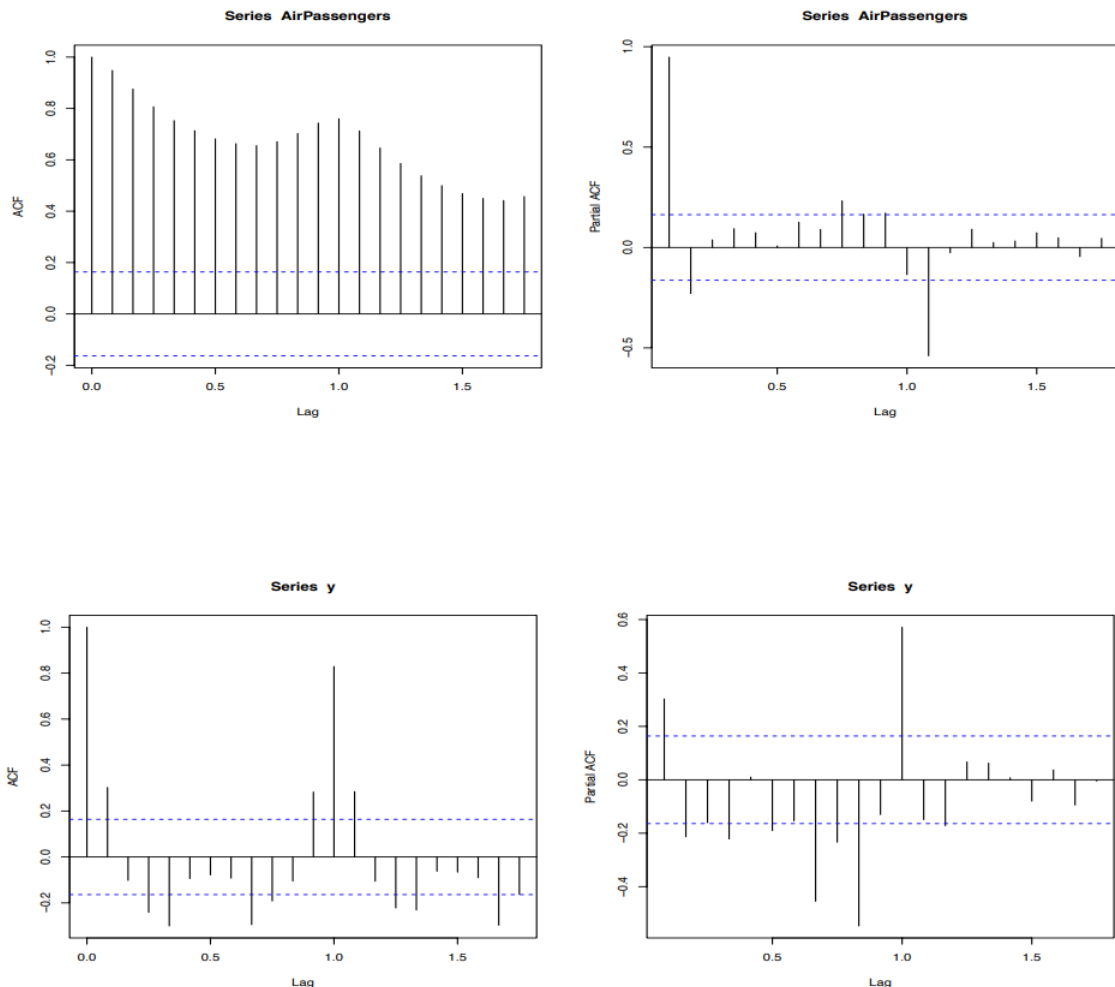
□

PARTE COMPUTACIONAL

6. Para una serie de tiempo que contiene 144 observaciones mensuales ha sido analizada por un especialista. La siguiente tabla muestra la información del análisis realizado.

Modelo	Parámetros estimados	Error estándar	AIC	Test de Box-Ljung para H=20
AR(1)	0,96	0,02	1248	0,04
AR(2)	1,28 - 0,33	0,07 0,07	1413	0,08
MA(1)	0,96	0,02	1618	0,03
MA (2)	1,37 0,99	0,12 0,18	NA	0,02
ARIMA (1, 1, 1)	-0,47 0,86	0,11 0,07	139	0,01
ARIMA (2, 1, 1)	1,09 - 0,48 - 0,84	0,07 0,07 0,04	1378	0,02
ARIMA (1, 1, 0)	0,30	0,07	1401	0,03

Además disponemos de la siguiente información donde en el gráfico y representa la serie diferenciada.



- a. ¿Qué modelo ajustaría Usted? Escriba la ecuación de su modelo elegido. Justifique

Solución. En base los gráficos de ACF y PACF de la serie original, notaríamos que estamos en presencia de un AR(1). Sin embargo, al diferenciar la serie, observamos un lag significativo en el ACF, lo que

conllevaría a considerar el modelo $\text{ARIMA}(1,1,1)$. Lo anterior, adquiere mayor sustento al visualizar los AIC de la tabla, donde el menor lo posee el modelo anterior. Ahora bien, no se conoce el AIC del $\text{MA}(2)$. Sin embargo, el valor obtenido al aplicar el test de Box-Ljung es menor para un $\text{ARIMA}(1,1,1)$. De esta manera, nos quedamos con $\text{ARIMA}(1,1,1)$, cuya ecuación es

$$(1 + 0,47B)(1 - B)X_t = (1 + 0,86B)\epsilon_t$$

□

- b.** Si para el modelo $\text{MA}(2)$ en la tabla anterior se sabe que la verosimilitud evaluada en los parámetros estimados es 1×10^{-10} , calcule el coeficiente de Akaike asociado al modelo.

Solución. Usando la relación vista en clases

$$\text{AIC} = 2k - 2 \ln(L).$$

Reemplazando obtenemos que

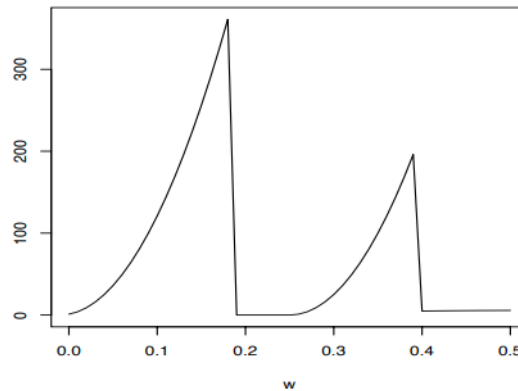
$$\text{AIC} = 2 \cdot 3 - 2 \ln(1 \times 10^{-10}) = 52,05.$$

□

- c.** ¿Qué información adicional le parece razonable conocer? (su respuesta debe ser escrita como máximo en tres líneas).

Solución. Dado que el AIC calculado del $\text{MA}(2)$ fue menor. Este modelo entra directamente a competir con el propuesto $\text{ARIMA}(1,1,1)$. De esta manera, necesitamos un criterio decisivo. Por lo anterior, podríamos conocer el BIC (*Bayesian information criterion*) de ambos modelos para verificar. □

- 7.** Para una serie de tiempo, X_t , se graficó el periodograma $I(w)$ usando 100 observaciones. La figura se muestra a continuación:



- a.** Proponga un modelo de regresión armónica para la serie X_t .

Solución. Dado que vislumbramos dos picos en el periodograma, planteamos el modelo

$$X_t = \mu + \sum_{i=1}^2 \beta_{1i} \cos(2\pi\omega_i t) + \beta_{2i} \sin(2\pi\omega_i t) + \epsilon_t.$$

donde $\omega_1 = 0,2$, $\omega_2 = 0,4$ y $\epsilon \sim \text{WN}(0, \sigma^2)$. □

- b.** ¿Qué método usaría usted para estimar los parámetros del modelo propuesto?

Solución. Escribir el modelo como $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ y luego usar el método de mínimos cuadrados como sigue

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta} \in \mathbb{R}^3} \left\{ (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \right\} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y},$$

$$\hat{\sigma}^2 = \frac{\mathbf{e}^\top \mathbf{e}}{n - k} = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - k},$$

donde k corresponde al número de coeficientes (incluido el intercepto) a estimar. □

- c. ¿Cuáles son los períodos más influyentes en la serie?

Solución. Los periodos más influyentes son los asociados a los valores máximos de ω . En nuestro caso, aproximaremos aquellos con los valores $w_1 = 0,2$ y $w_2 = 0,4$. De esta manera, los periodos más influyentes son

$$T_1 = \frac{1}{\omega_1} = 5 \quad \text{y} \quad T_2 = \frac{1}{\omega_2} = 2,5$$

□

- d. Para $\omega_1 = 0,2$ y $\omega_2 = 0,4$ determine si es posible aseverar con una confianza del 95 % que los valores de la densidad espectral son mayores a 100.

Solución. Para verificar aquello, crearemos el intervalo de confianza unilateral

$$\text{IC}_{(1-\alpha)100\%}(f(\omega)) = \left[\frac{2I(\omega_{j:n})}{\chi^2_{(\alpha,2)}}, \infty \right).$$

En nuestro caso, $I(\omega_1) \approx 350$, $I(\omega_2) \approx 200$ y $\chi^2_{(0,05,2)} \approx 5,99$. Así, los intervalos de confianza para cada ω_j son

$$\begin{aligned} \omega_1 : \quad \text{IC}_{95\%}(f(\omega)) &= \left[\frac{2I(\omega_{j:n})}{\chi^2_{(\alpha,2)}}, \infty \right) = [116,86, \infty). \\ \omega_2 : \quad \text{IC}_{95\%}(f(\omega)) &= \left[\frac{2I(\omega_{j:n})}{\chi^2_{(\alpha,2)}}, \infty \right) = [66,78, \infty). \end{aligned}$$

Por lo tanto, para ω_1 es posible aseverar con una confianza del 95 % que los valores de la densidad espectral son mayores que 100 para ω_1 , mientras que para ω_2 no es factible asegurar lo anterior. □

- e. Para $\omega_1 = 0,2$ ¿es posible aseverar con una significancia del 5 % que la densidad espectral es distinta al valor nominal 200? Justifique su procedimiento planteando adecuadamente sus hipótesis y el método que conduce al resultado.

Solución. Planteamos el test de hipótesis

$$H_0 : f(\omega_1) = 200, \quad H_1 : f(\omega_2) \neq 200.$$

El cual mediante el intervalo de confianza bilateral con 95 % de confianza

$$\text{IC}_{(1-\alpha)100\%}(f(\omega)) = \left[\frac{2I(\omega_{j:n})}{\chi^2_{(\alpha/2,2)}}, \frac{2I(\omega_{j:n})}{\chi^2_{(1-\alpha/2,2)}} \right]$$

nos permitirá decidir.

En nuestro caso $I(\omega_2) \approx 200$ y $\alpha = 0,05$, por ende $\chi^2_{(0,025,2)} = 7,34$ y $\chi^2_{(0,975,2)} = 0,05$. Es decir

$$\text{IC}_{95\%}(f(\omega)) = [95,35, 14000]$$

y concluimos que existe evidencia para aceptar la hipótesis nula $f(\omega_1) = 200$. □

8. Considere la serie de precios de las acciones google usada en ejemplos anteriores.

- a. Determine un modelo ARIMA(p, d, q) apropiado para esta serie y haga un diagnóstico del modelo propuesto y el chequeo de los supuestos.

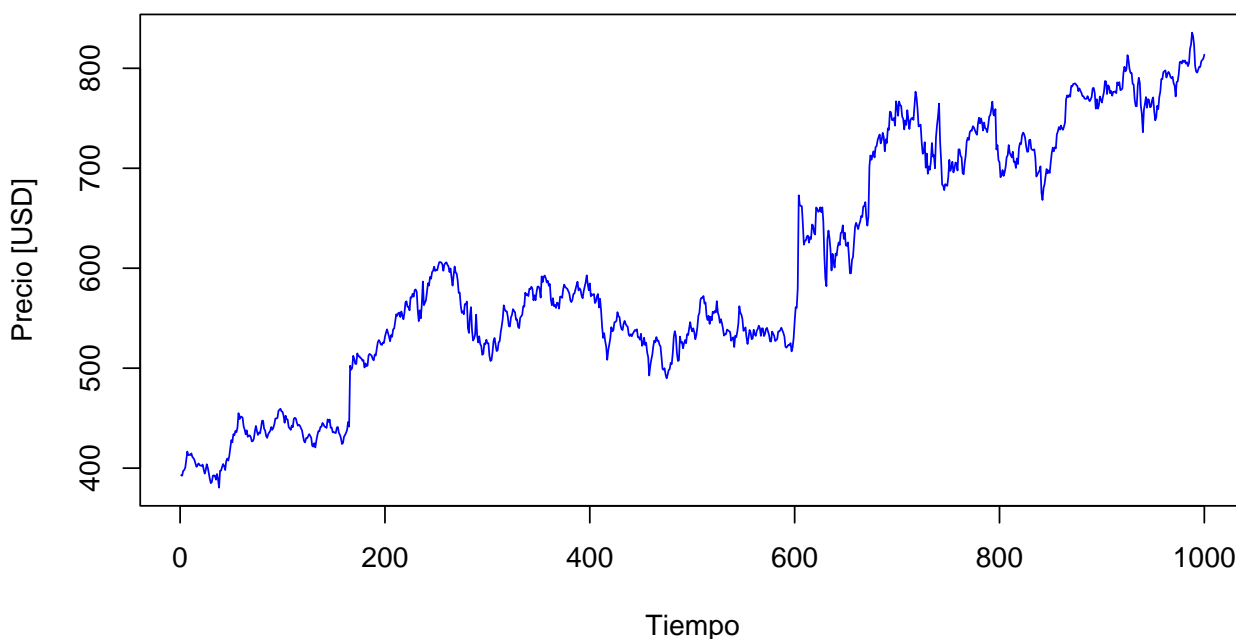
Solución. En primer lugar, para determinar un modelo ARIMA(p, d, q), visualizaremos la serie original

```
library(tidyverse)
library(fpp2)

plot(goog, main = "Precios de las acciones Google", xlab = "Tiempo",
     ylab = "Precio [USD]", col = "blue")
```

OUTPUT

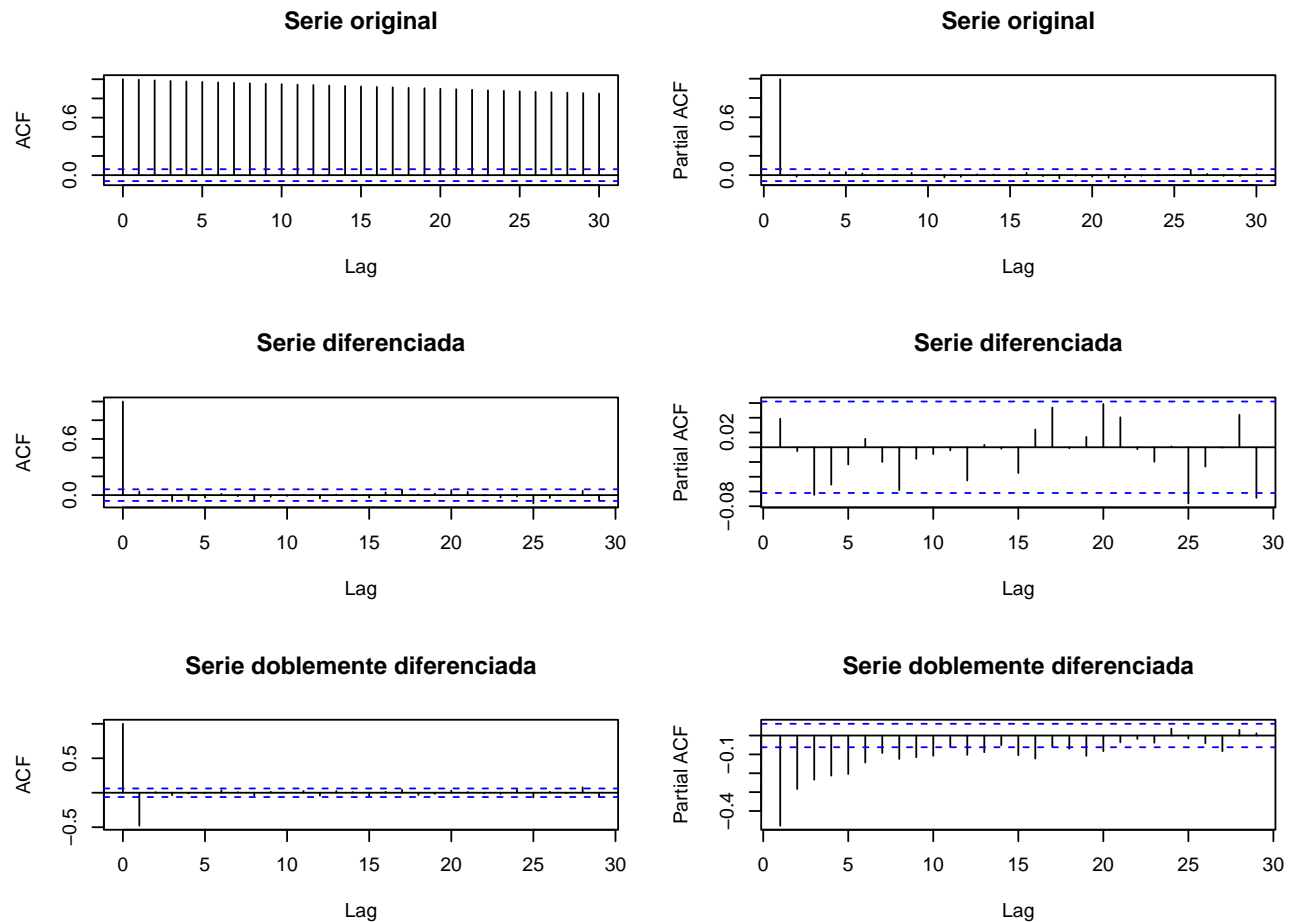
Precios de las acciones Google



Enseguida, notamos una clara tendencia alcista. Por lo que realizar una diferenciación sería un buen punto de partida. Ahora, estimaremos los valores de p , d y q usando el ACF y PACF de la serie original y diferenciada.

```
par(mfrow = c(2,2))
acf(goog, main = "Serie original")
pacf(goog,main = "Serie original")
acf(diff(goog), main = "Serie diferenciada")
pacf(diff(goog), main = "Serie diferenciada")
acf(diff(diff(goog)), main = "Serie doblemente diferenciada")
pacf(diff(diff(goog)), main = "Serie doblemente diferenciada")
```

OUTPUT



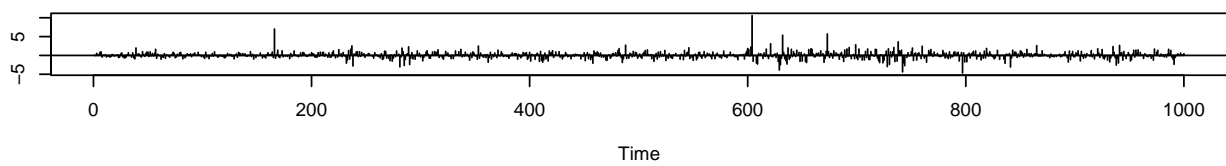
En base a los resultados anteriores, notamos un lag considerable en la PACF de la serie original, lo que denotaría un modelo AR(1). Sin embargo, dado que no existe autocorrelación en la serie diferenciada, no es factible. Ahora bien, en la serie dos veces diferenciada observamos un lag en la ACF para $h = 1$, mientras que en la PACF se aprecian seis lags claros. Es por esto, que escogeremos un modelo ARIMA(6,2,1) para ajustar.

Con el objetivo de chequear nuestra decisión, realizaremos un diagnóstico del modelo analizando la normalidad de los residuos.

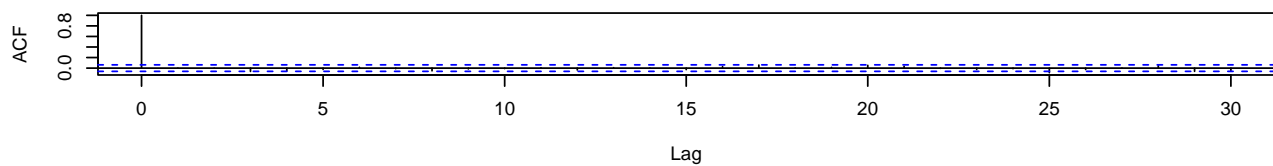
```
fit <- arima(goog, c(6, 2, 1))
tsdiag(fit)
hist(fit$resid, main = "Histograma de residuos")
qqnorm(fit$resid)
shapiro.test(fit$resid)
```

OUTPUT

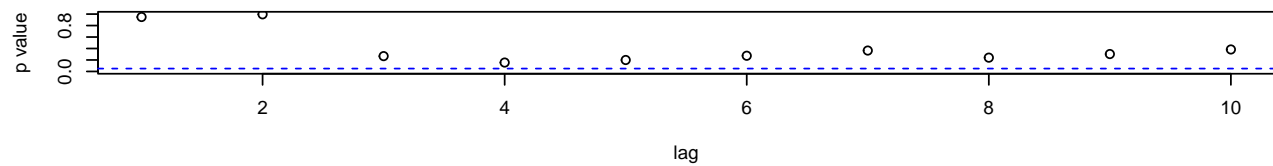
Standardized Residuals



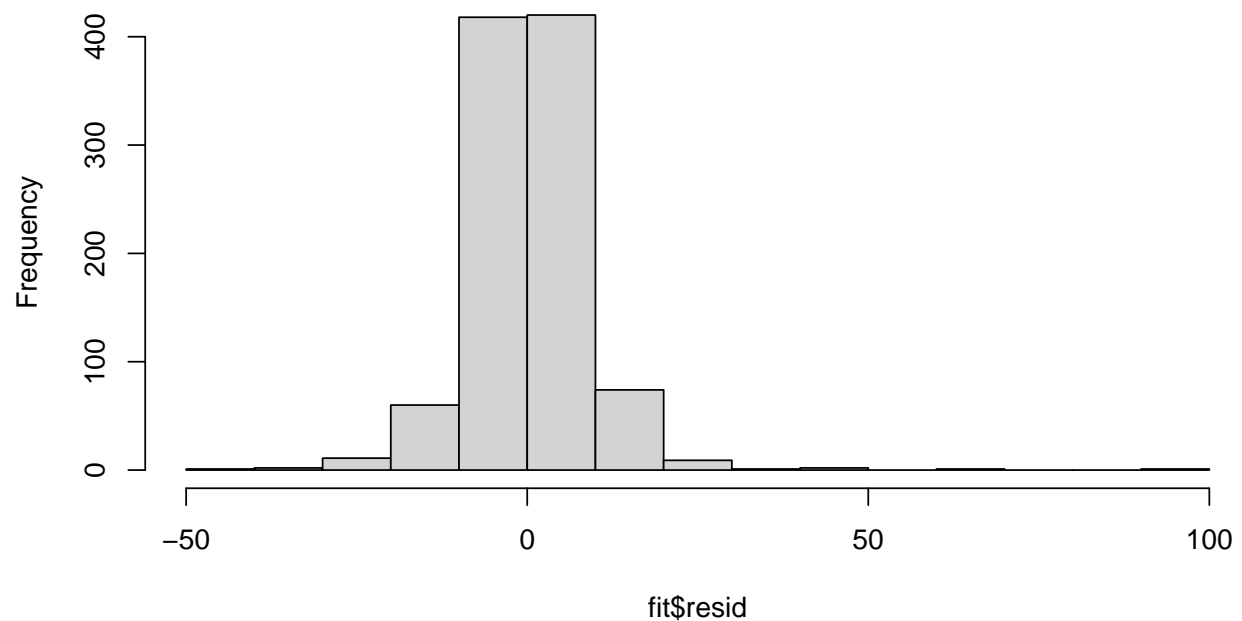
ACF of Residuals

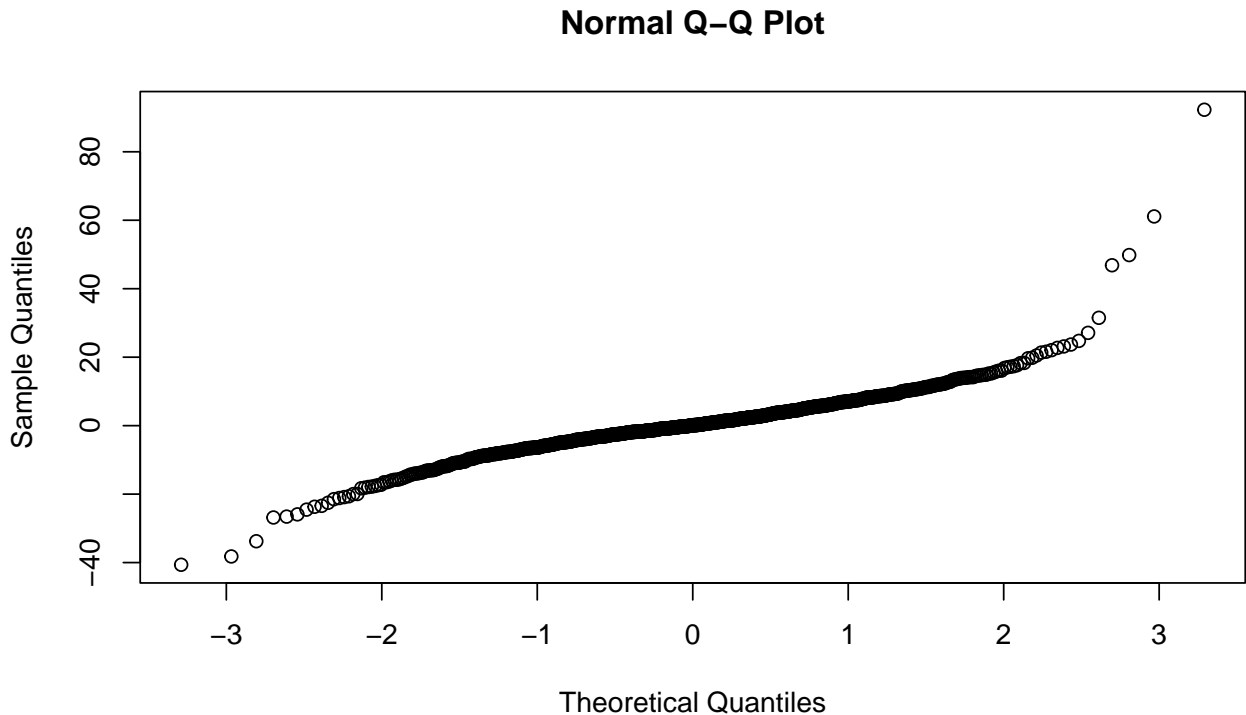


p values for Ljung-Box statistic



Histograma de residuos





Shapiro-Wilk normality test

data: fit\$resid W = 0.88051, p-value <2.2e-16

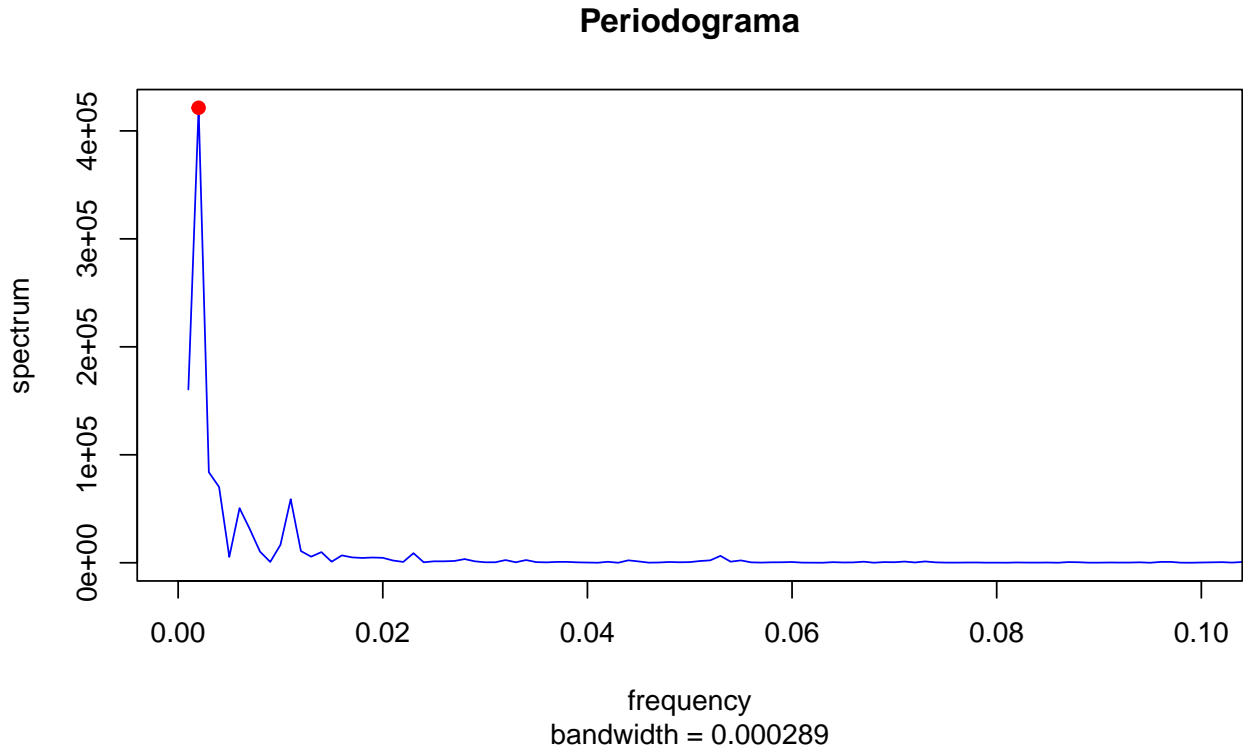
En base a lo anterior, notamos que los residuos tienen cierta normalidad, sin embargo, existen outliers que perturban el ajuste realizado. Lo anterior, se comprueba con el valor p que nos entrega el test de Shapiro-Wilk. Al ser tan bajo, la evidencia para rechazar la hipótesis nula (normalidad de los residuos) es fuerte.

□

- b. Determine si existen períodos significativos en la serie. En tal caso construya un modelo de regresión armónica con tendencia lineal y compárelo que el modelo propuesto en a) usando validación cruzada.
Solución. Verificaremos si existen periodos significativos mediante el periodograma.

```
per <- spec.pgram(goog, taper=0, log = "no", main = "Periodograma", col = "blue",  
xlim = c(0, 0.1))  
sort(per$spec)  
s <- per$spec[per$spec>400000]  
f <- per$freq[per$spec>400000]  
points(f, s, cex = 1, pch = 19, col = "red")
```

OUTPUT



Enseguida vislumbramos un periodo significativo asociado a la frecuencia $\omega = 0,002$, así $T = 1/\omega = 500$. Por ende, podríamos ajustar el siguiente modelo de regresión armónica con tendencia lineal

$$X_t = \mu + \beta_1 t + \beta_2 \cos(2\pi\omega t) + \beta_3 \sin(2\pi\omega t) + \epsilon_t,$$

donde $\epsilon \sim \text{WN}(0, \sigma^2)$. Notemos que los parámetros del modelo se pueden estimar usando mínimos cuadrados para calcular la recta asociada a la regresión lineal $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, donde

$$\begin{aligned} \mathbf{Y} &= (Y_{t_1}, \dots, Y_{t_n})^\top \in \mathbb{R}^n, \\ \boldsymbol{\beta} &= (\mu, \dots, \beta_3)^\top \in \mathbb{R}^4, \\ \mathbf{e} &= (\epsilon_{t_1}, \dots, \epsilon_{t_n})^\top \in \mathbb{R}^n, \\ \mathbf{X} &= \begin{pmatrix} 1 & t_1 & \cos(2\pi\omega t_1) & \sin(2\pi\omega t_1) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & t_n & \cos(2\pi\omega t_n) & \sin(2\pi\omega t_n) \end{pmatrix} \in \mathbb{R}^{n \times 4}. \end{aligned}$$

En particular, las estimaciones están dadas por

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y} \quad \text{y} \quad \hat{\sigma}^2 = \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^\top (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{n - 4},$$

```
omega = 0.002
```

```
y <- goog
int <- rep(1, length(y))
int2 <- 1:length(y)
```

```
matriz = matrix(, nrow = length(y), ncol = 2)
for (t in 1:length(y)) {
```

```
matriz[t, 1] <- cos(2*pi*omega*t)
matriz[t, 2] <- sin(2*pi*omega*t)
}

X <- cbind(int, int2, matriz)

betas <- solve(t(X) %*% X) %*% t(X) %*% y
betas <- round(betas, 2)
var <- (t((y - X %*% betas)) %*% (y - X %*% betas)) / (1000-4)
```

OUTPUT

$$\hat{\mu} = 414.68, \quad \hat{\beta}_1 = 0.37, \quad \hat{\beta}_2 = -39.30, \quad \hat{\beta}_3 = -13.93, \quad \hat{\sigma}^2 = 1.00699$$

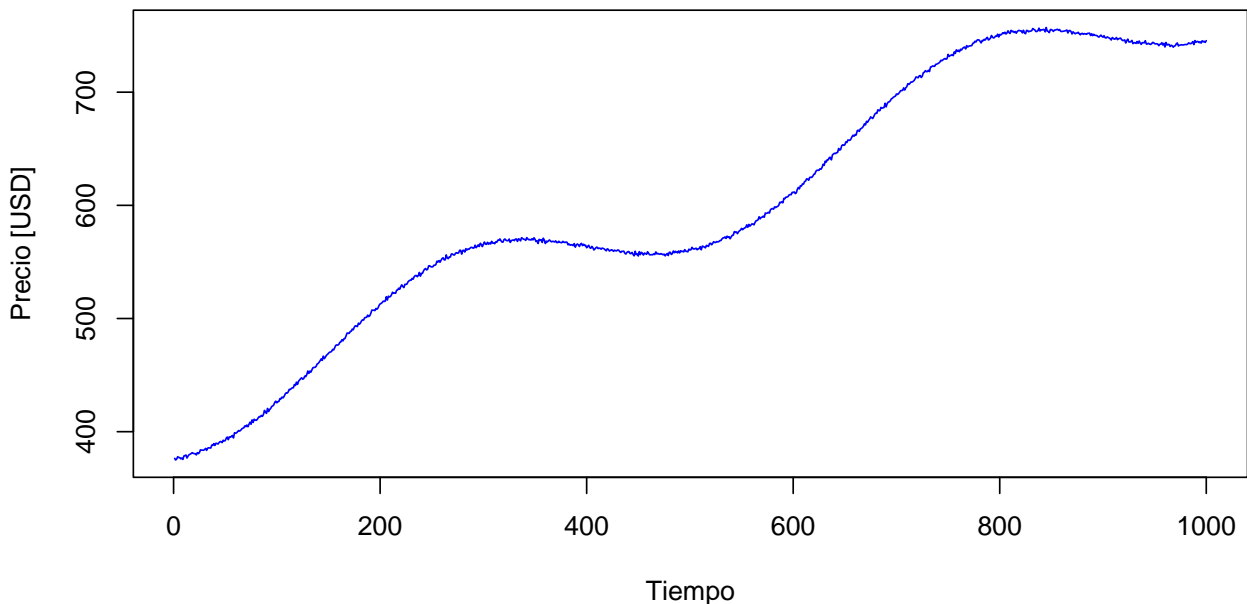
De esta manera, el ajuste queda determinado por

$$X_t = 414,86 + 0,37t - 39,30 \cos(2\pi\omega t) - 13,93 \sin(2\pi\omega t) + \epsilon_t.$$

```
x=rnorm(1000, 0, var)
y=vector(mode="numeric", length=1000)
for (t in 1:1000){
  y[t]= 414.68 + 0.37*t - 39.30*cos(2*pi*omega*t)-13.93*sin(2*pi*omega*t) +x[t]
}
plot.ts(y, main = "Regresión armónica con tendencia lineal", xlab = "Tiempo",
ylab = "Precio [USD]", col = "blue")
```

OUTPUT

Regresión armónica con tendencia lineal



A continuación, pondremos a competir este modelo con el hallado en el ítem anterior. Para esto realizaremos validación cruzada, dividiendo la serie *google* en muestras de entrenamiento y testeo, cuya proporción será de 80 % y 20 % respectivamente. Además, para cuantificar las predicciones, calcularemos el RMSE de ambos ajustes.

```
goog.train <- window(goog, end = 800)
goog.test  <- window(goog, start = 801)

fit <- arima(goog.train, c(6, 2, 1))

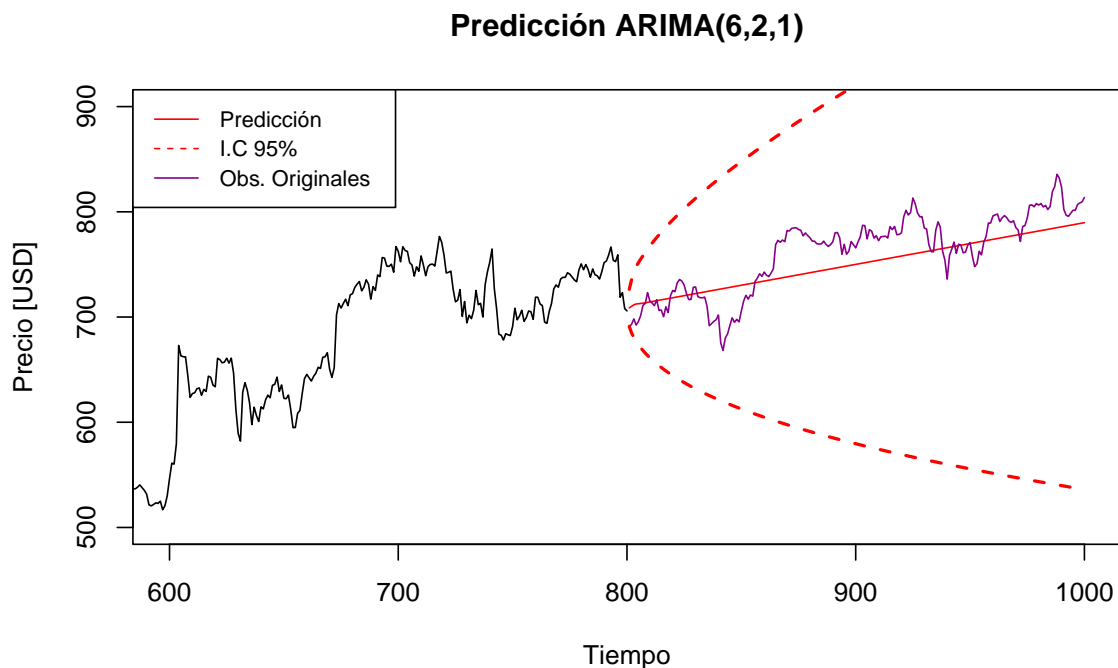
p <- predict(fit, 200)
U<-p$pred+1.96*p$se
L<-p$pred-1.96*p$se

plot(goog.train, xlim = c(600,1000), ylim = c(500,900),
main = "Predicción ARIMA(6,2,1)", xlab = "Tiempo", ylab = "Precio [USD]")
legend("topleft", legend=c("Predicción", "I.C 95%", "Obs. Originales"),
col=c("red", "red", "darkmagenta"), lty=1:2, cex=0.8)
xpred <- 801:1000
lines(xpred, p$pred , type="l", col = "red")
lines(goog.test, col = "darkmagenta")
lines(xpred,U,type = "l", lty=2, lwd = 2, col = "red")
lines(xpred,L,type = "l", lty=2, lwd = 2, col = "red")

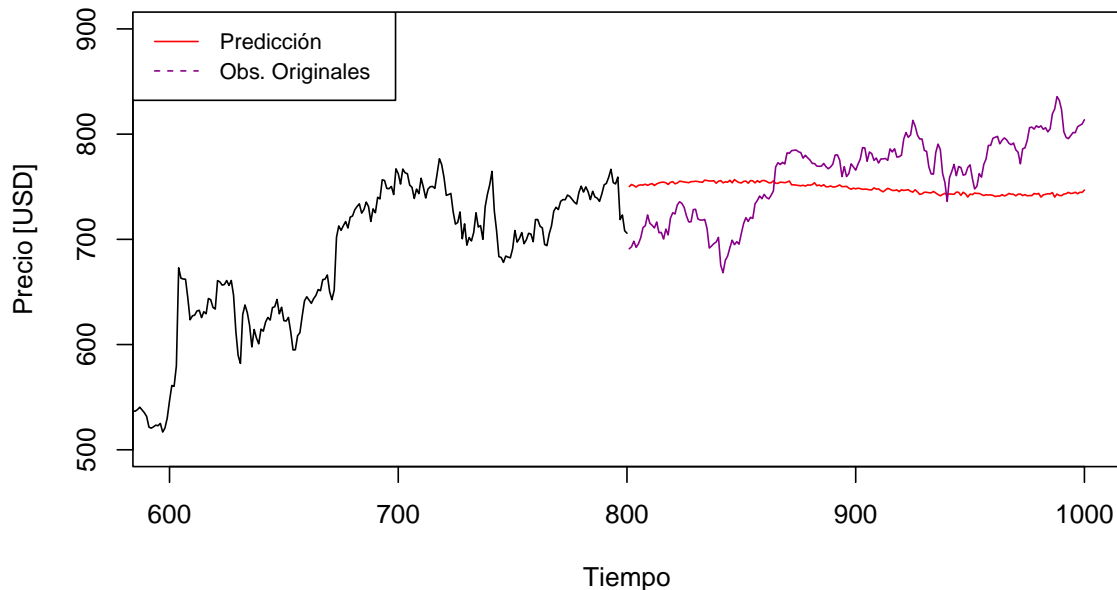
plot(goog.train, xlim = c(600,1000), ylim = c(500,900),
main = "Predicción Modelo Armónico", xlab = "Tiempo", ylab = "Precio [USD]")
legend("topleft", legend=c("Predicción", "Obs. Originales"),
col=c("red", "darkmagenta"), lty=1:2, cex=0.8)
xpred <- 801:1000
lines(xpred, y[801:1000] , type="l", col = "red")
lines(goog.test, col = "darkmagenta")

rmse(goog.test, p$pred)
rmse(goog.test, y[801:1000])
```

OUTPUT



Predicción Modelo Armónico



RMSE ARIMA(6,2,1) = 24.17627, RMSE Regresión Armónica = 42.89448

Notamos que el modelo ARIMA(6,2,1) luce más preciso y su menor valor de RMSE lo respalda. \square

- c. Dentro de la clase de modelos ingenuos (suavizamiento exponencial, modelos de descomposición y Holt-Winters) seleccione un modelo y su ajuste, y determine su desempeño respecto a los modelos propuestos en b).

Solución. Dado que la gráfica de google presenta tendencia, no sería adecuado usar un modelo de suavizamiento simple. Más aún, dado que las acciones bursátiles son sumamente difíciles de predecir, detectar estacionariedad no es trivial. Por ello, para mantener la esencia de un modelo ingenuo, consideraremos un modelo de Holt-Winters, el cual detecta la tendencia a medida que realiza suavizamiento.

Para determinar su desempeño, realizaremos el mismo procedimiento que en el ítem anterior y calcularemos el error cuadrático medio

```
holt.goog <- holt(goog.train, h = 200)
holt.goog.data <- as.data.frame(holt.goog)

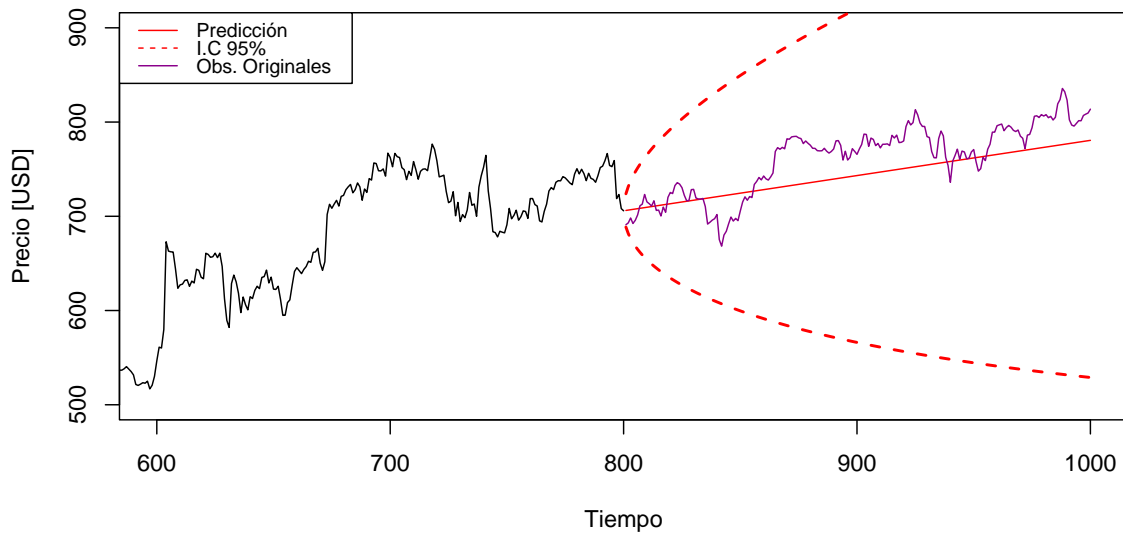
u <- holt.goog.data$`Hi 95`
l <- holt.goog.data$`Lo 95`
pred.holt <- as.numeric(holt.goog$mean)

plot(goog.train, xlim = c(600,1000), ylim = c(500,900),
main = "Predicción Holt-Winters", xlab = "Tiempo", ylab = "Precio [USD]")
lines(xpred, pred.holt , type="l", col = "red")
lines(goog.test, col = "darkmagenta")
lines(xpred, u, lty=2, lwd = 2, col = "red")
lines(xpred, l, lty=2, lwd = 2, col = "red")
legend("topleft", legend=c("Predicción", "I.C 95%", "Obs. Originales"),
col=c("red", "red", "darkmagenta"), lty=1:2, cex=0.8)

rmse(goog.test, pred)
```

OUTPUT

Predicción Holt-Winters



RMSE Holt-Winters = 28.29267

En base a lo anterior, concluimos que el modelo de Holt-Winters presenta menor error que el de regresión armónica, pero es levemente menos preciso que el ARIMA(6,2,1). Lo cual es un asunto no menor, pues mediante un modelo ingenuo obtuvimos predicciones que compiten con los de un ARIMA. Para finalizar el análisis, calcularemos el R^2 de los tres modelos vistos

```
R2 <- function(true, fitted){
  1 - (sum((true - fitted)^2)/sum((true - mean(true))^2))
}

R2(goog.test, p$pred)
R2(goog.test, y[801:1000])
R2(goog.test, pred.holt)
```

OUTPUT

R^2 ARIMA(6,2,1) = 0.5792343, R^2 Regresión Armónica = -0.3245383
 R^2 Holt-Winters = 0.4237516

Lo que ratifica la conclusion entre el modelo ingenuo y el ARIMA. De hecho, este último coeficiente denota que el modelo de regresión armónica actua en contra de la tendencia, por lo que no es adecuado. Todo lo anterior se resume en la siguiente tabla

	ARIMA(6,2,1)	Armónico	Holt-Winters
RMSE	24.18	42.90	28.29
R^2	0.58	-0.32	0.42

□