UNIVERSIDAD NACIONAL DE COLOMBIA

PREGRADO EN ESTADISTICA

Departamento de Estadística
Facultad de Ciencias

Series de Tiempo Univariadas

Profesor:
Norman Diego Giraldo Gómez
Trabajo 3 de Series de Tiempo
Modelos ARFIMA-SARFIMA
Grupo 8

Integrantes:

Ricardo William Salazar Espinal	C.C. 1017219472
Alejandro Velásquez Rendón	C.C. 1000913570

Medellín, Colombia

Medellín, Febrero de 2025

Índice

Índice de figuras

Índice de cuadros

1 Enunciado General

El objetivo de este trabajo es encontrar un modelo ARFIMA-SARFIMA para los rendimientos $X_n = log(I_n/I_{n-1})$, donde I_n es el valor de un índice tiempo n.

No se hacen pruebas para detectar memoria larga en la serie. En su lugar, si al realizar los procedimientos de estimación, el modelo ajusta, entonces se confirma la memoria larga.

2 Punto 1

Identifique un posible modelo SARFIMA $(p, q)(p_s, q_s)[fq]$, de período fq, con la función `auto.arima`. Reporte los órdenes (p, q, p_s, q_s). Con esta información estime el modelo SARFIMA con la función `arfima`. Reporte la tabla de valores estimados y valores p. Reporte la prueba Ljung-Box para los residuos.

3 Solución Punto 1

En primer lugar, se analizará si la serie es estacionaria en covarianza, para concluir sobre esto se usará la prueba de hipótesis KPSS, implementada en la librería `aTSA` con el comando `kpss.test`, los resultados se ven a continuación.

```
KPSS Unit Root Test
alternative: nonstationary

Type 1: no drift no trend
 lag stat p.value
   6 1.03     0.1
-----
 Type 2: with drift no trend
 lag  stat p.value
   6 0.436  0.0615
-----
 Type 1: with drift and trend
 lag  stat p.value
   6 0.224    0.01
-----------
Note: p.value = 0.01 means p.value <= 0.01
    : p.value = 0.10 means p.value >= 0.10
```

Si tomamos la serie como tipo 1, no se rechaza la hipótesis nula con un valor P de 0.1 (mayor a un nivel de significancia de $\alpha = 0.05$), o sea que la serie X_n es estacionaria en covarianza. Por tanto, no hay problema para ejecutar el objetivo del trabajo.

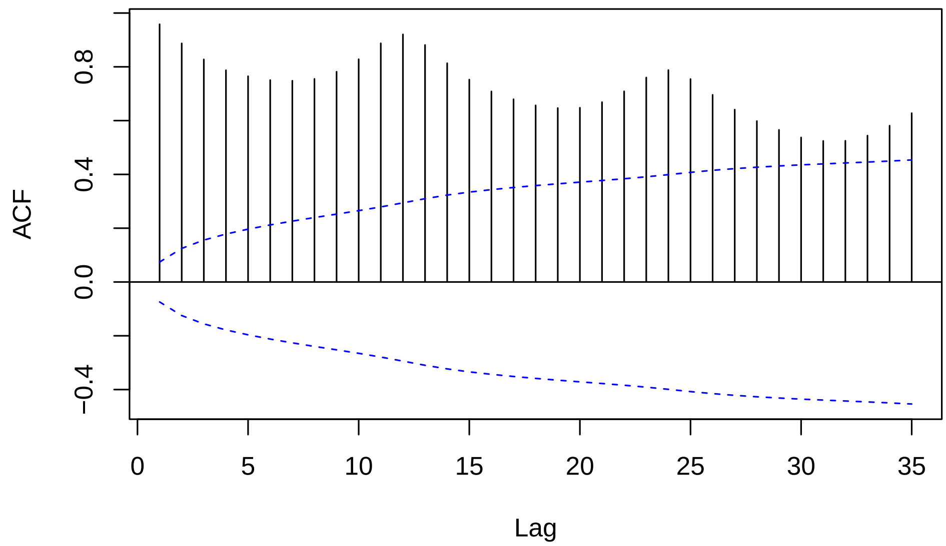

Figura 1: Función FAC de la serie

3.1 Periodo de la serie

Ahora que se sabe que la serie es estacionaria, se procede a buscar el periodo fq del posible modelo SARFIMA, para esto la idea general es buscar los picos relevantes de la función FAC (que se puede ver en la figura 1) de la serie. En este caso se obtiene un valor de $fq = 12$, luego, como mínimo se obtiene que el modelo es de la forma:

$$X_n \sim SARFIMA(p, d, q)(p_s, d_s, q_s)[12]$$

Resta hallar los valores de los hiperparámetros del modelo, para abordar este problema se hará uso de la función `auto.arima` del R, dando como parámetro a $fq = 12$, y asumiendo, como ya se sabe, que la serie X_n es estacionaria en covarianza, y que se quiere hayar un modelo con componente estacional, la función en cuestión nos arroja un resultado de `ARIMA(1, 0, 2)(2, 0, 2)[12]`, es decir que nuestro modelo aparentemente tiene la forma:

$$X_n \sim SARFIMA(1, d, 2)(2, d_s, 2)[12]$$

Sin embargo, al ajustar este modelo, se encontraban problemas en la significancia de los coeficientes estimados del modelo, es decir, algunos de los coeficientes de orden mayor del modelo no daban significativos o no se ajustaban bien a la serie muestral, el orden de los hiperparámetros se fue cambiando hasta llegar a un modelo adecuado que se explora a continuación.

Adicional a esto, una vez ajustado el modelo, se estima la diferenciación de la serie y la diferenciación estacional d y d_s, sin embargo, para confirmar memoria larga, solo será de interés d.

```
Note: only one starting point.  Only one mode can be found -- this is now the default be
Beginning the fits with 1 starting values.
```

Parámetro	**Estimación**	**Error estándar**	**Estadístico z**	**p-valor**
ar1	1.4636	0.3572	4.0974	4.18×10^{-5}
ar2	-0.5286	0.2708	-1.9519	0.0509
ma1	0.2138	0.3233	0.6614	0.5083
ma2	0.3793	0.1657	2.2892	0.0221
sar1	0.9199	0.0280	32.8961	$< 2.22 \times 10^{-16}$
sma1	-0.7614	0.0320	-23.8049	$< 2.22 \times 10^{-16}$
d	0.1958	0.1356	1.4440	0.1487
d_s	-0.2039	0.0614	-3.3206	0.0009

Tabla 1: Estimaciones del modelo SARFIMA

En la tabla 1, se observan los coeficientes estimados de nuestro modelo ajustado, en la última columna de la misma se ve que algunos no son significativos, como el ma1, sin embargo, se confirmará que el modelo ajusta bien, además, los coeficientes de orden mayor (ar2, ma2, etc.) sí son significativos (en el caso de ar2, su valor p es uno frontera).

También se ve que el valor de d está en el intervalo deseado ($0 < d < \frac{1}{2}$), lo que indicaría en primera instancia que el modelo es estacionario en covarianza y tiene presencia de memoria larga, aunque el parámetro aparece como que no es significativo, sin embargo, se modelará y se evaluará la calidad del ajuste.

Respecto a d_s, este aparece como un valor negativo y significativamente distinto de 0, sin embargo, en la mayoría de casos se asume que $d_s = 0$, pues la parte estacional se modela en términos de los otros parámetros (p_s y q_s), y como un valor negativo en este caso no suele ser interpretable, se va a considerar como que es 0 y se hará el respectivo ajuste.

Resta confirmar si nuestros residuos son ruido blanco o no, para esto se hará una prueba formal, la prueba en cuestión es la prueba de Ljung-Box para ruido blanco, el juego de hipótesis es el siguiente:

$$\begin{cases} H_0 : Z_n \sim RB(0, \sigma^2) \\ H_1 : no(H_0) \end{cases}$$

Y la conclusión se obtiene gracias a la función `Box.test` de R, cuya salida se imprime a continuación:

```
	Box-Ljung test
```

```
data:  residuals
X-squared = 80.981, df = 48, p-value = 0.002044
```

Nuevamente, a un nivel de significancia de $\alpha = 0.05$, se rechaza H_0 (VP=0.002044), es decir que los residuos de nuestro modelo no son Ruido Blanco. Este es un supuesto fundamental de los modelos SARFIMA, sin embargo, se ajustará el modelo propuesto por fines académicos, además, el valor p, si bien es bajo, no lo es tanto comparado con los resultados de la prueba para otros modelos que se ensayaron.

4 Punto 2

Verifique el modelo comparando en una gráfica el peridograma suavizado de la serie X_n versus la densidad espectral estimada correspondiente al modelo identificado. Use el procedimiento del trabajo No 2 para decidir si el modelo ajustó la serie.

5 Solución Punto 2

5.1 Densidad Espectral Estimada

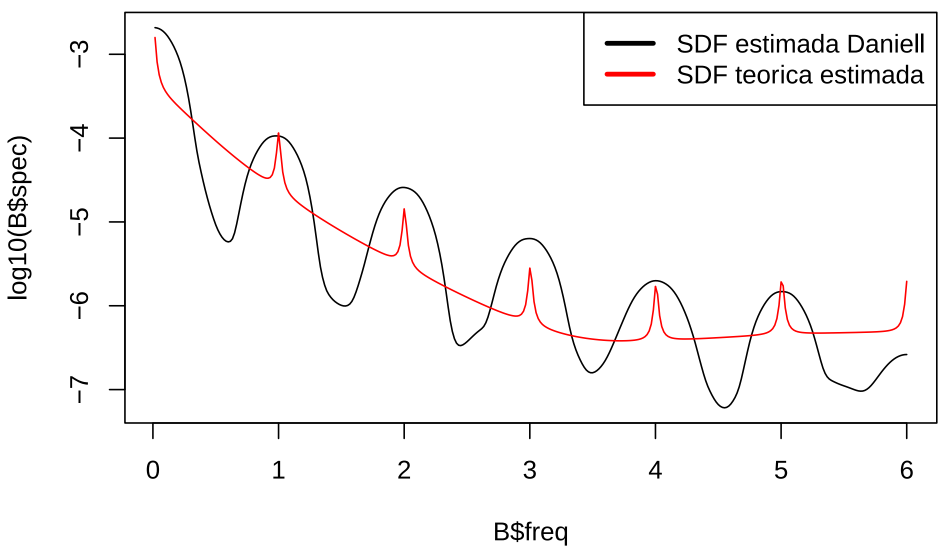

Figura 2: Espectro Observado vs. Teórico

La gráfica 2 muestra la comparación entre la densidad espectral de potencia (SDF) estimada mediante el suavizamiento de Daniell (línea negra) y la SDF teórica estimada a partir del ajuste de un modelo SARFIMA (línea roja). Se observa que la SDF estimada mediante Daniell captura las principales estructuras espectrales de los datos, evidenciando picos en frecuencias específicas. Sin embargo, la SDF teórica presenta una representación más suavizada, con menor variabilidad, lo que sugiere que el modelo SARFIMA capta la tendencia general de la estructura espectral pero no reproduce completamente las fluctuaciones observadas en la estimación empírica.

5.2 Periodogramas Acumulados y prueba KS aleatorizada

Ahora se hará el análisis de los periodogramas acumulados para la serie observada y la serie teórica modelada mediante *SARFIMA*.

Lo primero que se observa es que ambas curvas (azul y negra) siguen un patrón muy similar, lo que sugiere que el modelo SARMA representa razonablemente bien el comportamiento espectral acumulado de la serie observada.

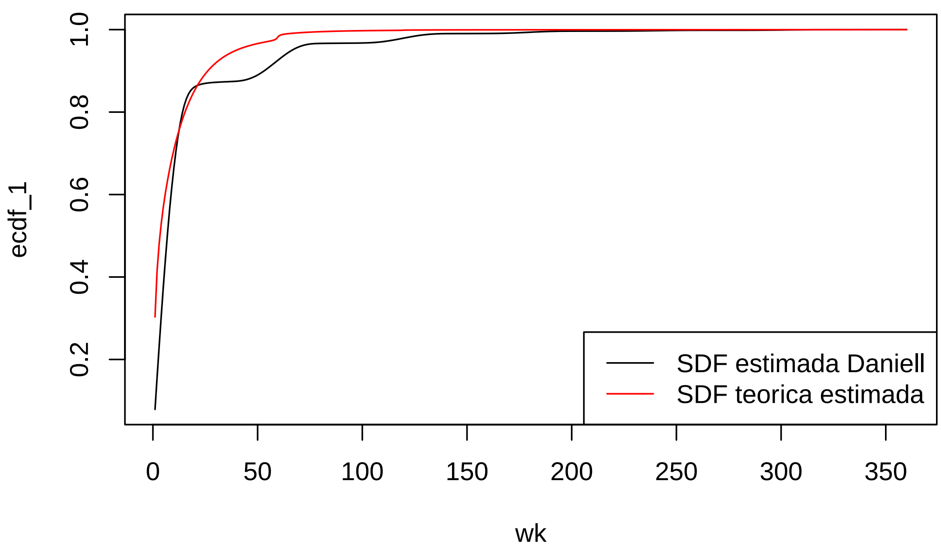

Figura 3: Periodogramas acumulados

A pesar de esto, hay pequeñas discrepancias entre las curvas, especialmente en algunas secciones (por ejemplo, entre los índices 50 y 100). Esto podría indicar posibles Variaciones en la dinámica de las frecuencias que el modelo SARMA no captura perfectamente.

De todas maneras, los dos periodogramas acumulados son muy parecidos, por lo que se puede esperar conseguir un buena ajuste.

A continuación se comprobará la igualdad entre las densidades espectrales de ambas series (la observada vs la teórica), mediante la prueba de homegeneidad aleatorizada usando el estadístico de Kolgomorov-Smirnov (K-S). El juego de hipótesis es:

$$\begin{cases} H_o : I_{X_n}(w_k) = I_X(w_k),\ k = 1, ..., \lfloor \frac{n}{2} \rfloor \\ H_1 : no(H_0) \end{cases}$$

Asumiendo claro, que X representa al modelo teórico.

En este caso lo que se busca es obtener un valor p alto para no rechazar la hipótesis nula, de esta manera se podría concluir que el modelo SARFIMA propuesto es una buena aproximación de la serie observada.

```
[1] 0.97551
```

La prueba de Kolmogorov-Smirnov aplicada a los espectrogramas acumulados del modelo SARFIMA y de la serie observada respalda la hipótesis nula de que no existen diferencias significativas entre sus distribuciones acumuladas. El valor p alto de 0.97551 indica que la distancia máxima entre las distribuciones acumuladas observada (línea roja en el histograma) y las simuladas es consistente con las distribuciones generadas por el modelo.

En términos prácticos, esto significa que el modelo SARFIMA es una muy buena aproximación de las características espectrales de la serie observada. Las discrepancias locales identificadas previamente en los gráficos de espectrogramas acumulados son mínimas y no tienen un impacto estadísticamente significativo.

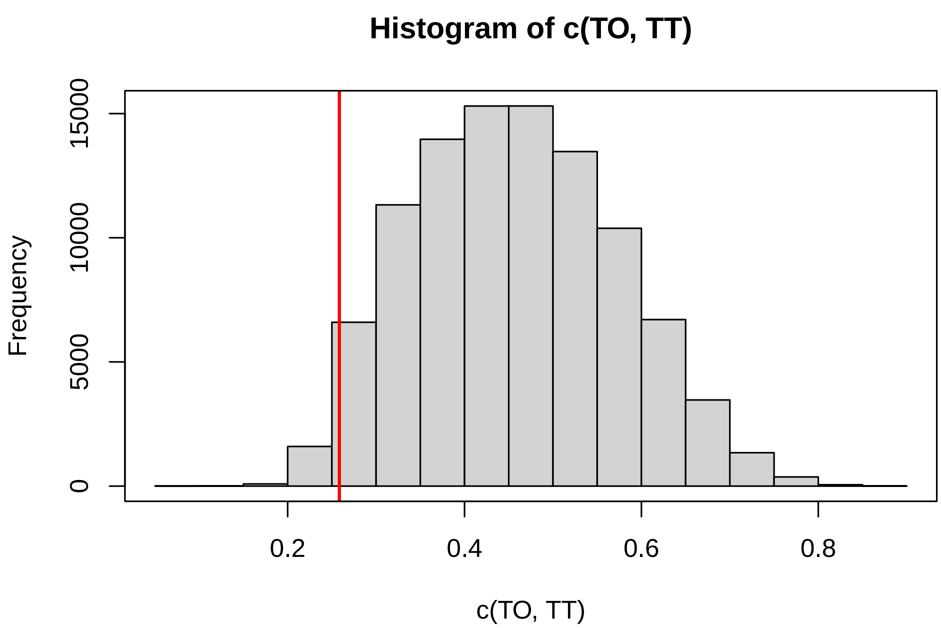

Figura 4: Histograma

5.3 Conclusión

Este análisis confirma que el modelo SARFIMA propuesto puede considerarse adecuado para describir el comportamiento espectral de la serie temporal observada, esto a pesar de que habíamos observado que algunos coeficientes no eran significativos, incluso el d, sin embargo, se pudo ajustar un modelo teórico de buena manera, y, como habíamos planteado, hay presencia de memoria larga en la serie.

6 Punto 3

Estime la volatilidad de los rendimientos del índice, X_n, mediante EWMA, Adicionalmente, pronostique el índice $X_{T+j}, j = 1, 2, ..., m$ para $m = 2 \times p$, con p período. Y recalcule la volatilidad con EWMA para el vector de los X_n ampliado con los de estos pronósticos. Reporte una gráfica de la volatilidad y su pronóstico.

7 Solución Punto 3

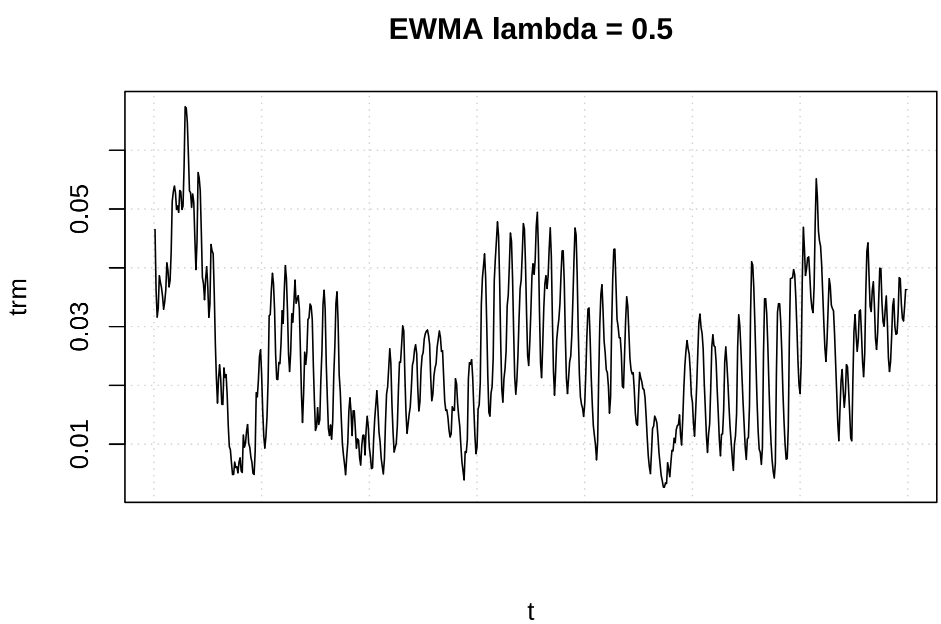

Figura 5: Volatilidad EWMA para Xn

En la figura 5, se ve la volatilidad EWMA de X_n, basada únicamente en sus datos históricos, inicialmente se ve una disminución seguida de periodos con oscilaciones más marcadas, parece que se estabiliza en valores cerca al final.

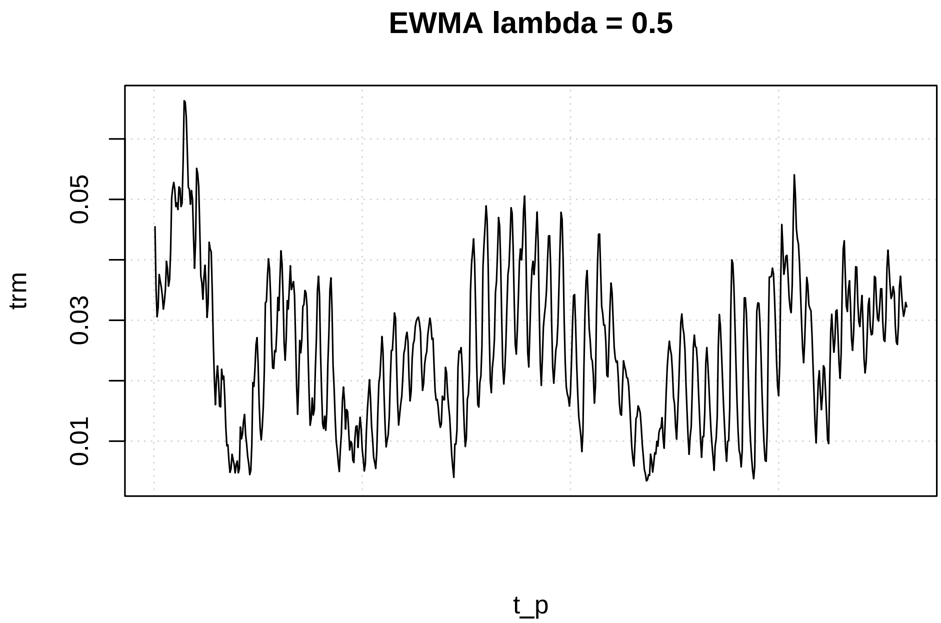

Figura 6: Volatilidad EWMA para Xn con datos extendidos

En la figura 6 se recalcula y se muestra la volatilidad incluyendo los valores de los 24 pronósticos, esta se ve muy similar a la primera gráfica de volatilidad que se vio, con diferencias hacia el final, que refleja la influencia de los pronósticos en el cálculo de la misma.

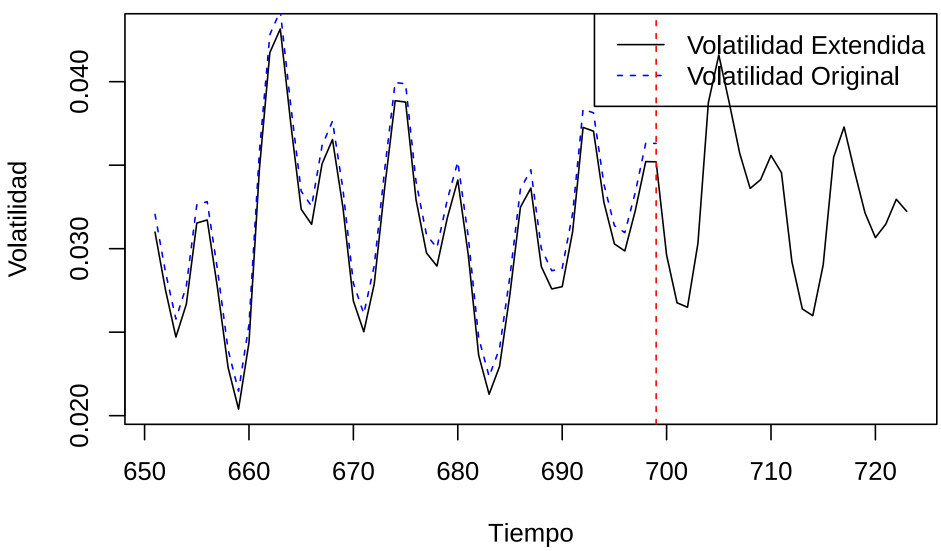

Figura 7: Volatilidad EWMA con Pronósticos

Finalmente, en la gráfica 7 se observan ambas volatilidades EWMA comparadas, haciendo zoom al punto donde se empiezan a introducir los pronósticos.

8 Apéndice - Código Usado

8.1 Llamado de librerías

```
library(rlist)
library(TSA)
library(forecast)
library(arfima)
library(lmtest)
library(astsa)
library(longmemo)
library(fracdiff)
library(fTrading)
library(rugarch)
```

8.2 Punto 1

```
# Lectura base de datos
Cn = list.load("Cm.rdata")
In = Cn$serie8
Xn = diff(log(In),1,1)

aTSA::kpss.test(Xn) #Prueba kpss

# Cálculo del periodo
Rk.est <- TSA::acf(Xn, 35, ci.type = 'ma', drop.lag.0=TRUE,main='') #FAC

k.mx = which(Rk.est$acf[-c(1,2)]==max(Rk.est$acf[-c(1,2)]))
p = k.mx + 1

k3 = kernel("modified.daniell", c(11,7,3))
B = spec.pgram(Xn, k3, taper = 0, log = "yes", ci = 0.8)
abline(v = 1/7, col = "red", lty = 2, lwd = 2)
abline(v = 1/7 * 2, col = "red", lty = 2, lwd = 2)
abline(v = 1/7 * 3, col = "red", lty = 2, lwd = 2)

Xn = ts(Xn, frequency = 12)

# Posible modelo
mod <- forecast::auto.arima(Xn, stationary = TRUE)
```

```r
# Modelo
mod2 <- arfima::arfima(Xn, order = c(2,0,2),
                       seasonal = list(order = c(1,0,1), period = 12))
d.est<- abs(mod2$modes[[1]]$dfrac) #d

# Prueba Ljung-Box
residuals <- mod2$modes[[1]]$residuals
Box.test(residuals, lag = 48, type = "Ljung-Box")
```

8.3 Punto 2

```r
theta = NULL
for(j in 1:4){
  theta = c(theta, mod2$modes[[1]][[j]])}

# Coeficientes
ar = theta[1:2]
ma = theta[3:4]
ars = theta[5]
mas = theta[6]

# Función de usuario
gen.sarma.coef = function(ar,ma,ars,mas,p){

require(polynom)

period = p
bs <- polynomial(c(rep(0,period),1))
b1 <- polynomial(c(0,1))

  arpoly <- polynomial(c(1,-ar))
  mapoly <- polynomial(c(1,ma))
  sarpoly <- polynomial(c(1,-ars))
  smapoly <- polynomial(c(1,mas))

  fullarpoly <- arpoly*predict(sarpoly,bs)
  fullmapoly <- mapoly*predict(smapoly,bs)

  mo <- list()
  mo$ar <- -coef(fullarpoly)[-1]
  mo$ma <- coef(fullmapoly)[-1]

return(mo)}
```

```r
mo = gen.sarma.coef(ar=ar,ma=ma, ars=ars, mas=mas, p=12)

k3 = kernel("modified.daniell", c(11,7,3))
B = spec.pgram(Xn, k3, taper = 0, log = "dB", ci = 0.8, main = "(a)",
               xlab = "frequency")



#Espectro Observado vs. Teórico
V2 = specARIMA(eta = c(H = 0.5 + d.est, phi = mo$ar, psi = mo$ma),
               m = 2*length(B$freq)+1, p = length(mo$ar), q =length(mo$ma))
vm2 = lm(log10(B$spec) ~ log10(V2$spec))
a2 = vm2$coef[1]
b2 = vm2$coef[2]

plot(B$freq, log10(B$spec), type = "l", lwd = 1)
lines(B$freq, a2+b2*log10(V2$spec), col = "red", lty = 1, lwd = 1)
legend("topright", c("SDF estimada Daniell", "SDF teorica estimada"),
       lty = c(1,1), lwd = c(3,3), col = c("black","red"))

# Periodogramas acumulados
ecdf_1 <- cumsum(B$spec)/sum(B$spec)
ecdf_2 <- cumsum(V2$spec)/sum(V2$spec)
wk = seq(1,length(ecdf_1))

plot(wk,ecdf_1,type='l')
lines(wk,ecdf_2,col='red')
legend("bottomright",
       c("SDF estimada Daniell", "SDF teorica estimada"),
       lty = c(1,1), lwd = c(1,1), col = c("black", "red"))

# Prueba K-S
Z <- c(B$spec,V2$spec)
n <- length(B$spec)
m <- length(V2$spec)
N <- length(Z)

TS <- function(Iwk1,Iwk2){
ecdf_1 <- cumsum(Iwk1)/sum(Iwk1)
ecdf_2 <- cumsum(Iwk2)/sum(Iwk2)
abs_dif <-  abs(ecdf_1- ecdf_2)
distancia_ks <- max(abs_dif)
return(distancia_ks)
}
```

```
TO <- TS(B$spec,V2$spec)
TT <- vector()
K = 100000

for(i in 1:K){
  set.seed(i)
  Z.pi <- sample(Z, N, replace = FALSE)
  TT[i] <- TS(Z.pi[1:n], Z.pi[(n+1):(n+m)])
}

pvalue = mean(TT>TO)
pvalue

# Histograma
hist(c(TO,TT))
abline(v = TO, lwd = 2, col = "red")
box()
```

8.4 Punto 3

```
# Volatilidad EWMA para Xn
par(mfrow = c(1,1))
dx <- Xn
rx <- (dx - mean(dx))^2

sigma2 <- emaTA(rx, lambda = 0.5, startup = 30)
sigmax <- sqrt(sigma2)
t <- seq_along(Xn)

plot(t, sigmax , xaxt = "n", panel.first = grid(),
       type = 'l', ylab = 'trm', main = paste("EWMA", "lambda = 0.5"))

# Volatilidad EWMA con pronósticos
m = 2 * 12
Xn_extended <- c(Xn, pred[[1]]$Forecast)

dx_p <- Xn_extended
rx_p <- (dx_p - mean(dx_p))^2

sigma2_p <- emaTA(rx_p, lambda = 0.5, startup = 30)
sigmax_p <- sqrt(sigma2_p)
t_p <- seq_along(Xn_extended)
```

```r
plot(t_p, sigmax_p , xaxt = "n", panel.first = grid(),
       type = 'l', ylab = 'trm', main = paste("EWMA", "lambda = 0.5"))

# Las 2 volatilidaes
m0 <- 48
T <- length(Xn)
start_idx <-T-m0
end_idx <-T+m

# Graficar solo la volatilidad extendida con algunos datos previos
plot(seq(start_idx, end_idx), sigmax_p[start_idx:end_idx], type = "l",
     ylab = "Volatilidad", xlab = "Tiempo",
     main = "", col = "black")

lines(seq(start_idx, T), sigmax[start_idx:T], col = "blue", lty = 2)

legend("topright", legend = c("Volatilidad Extendida", "Volatilidad Original"),
       col = c("black", "blue"), lty = c(1, 2))
abline(v = length(Xn), col = "red", lty = 2)
```

Referencias

Gomez, N.D.G. *Introduccion a Series de Tiempo con aplicaciones en R*. Universidad Nacional de Colombia.

Luque-Calvo, P.L. (2017). *Escribir un Trabajo Fin de Estudios con R Markdown*. Disponible en http://destio.us.es/calvo.