Trabajo 2- Estadística 3

Cristina Mercedes Ortega Benavides

18/5/2021

# 1. Introducción

* Definición DANE del índice asignado (ecuación debidamente explicada) dando ejemplo interpretativo de sus valores
* Definición del dominio o clasificación industrial al que está relacionada la serie.
* Resultados alcanzados en el trabajo 1
* Cuáles modelos globales y locales que fueron propuestos (debe dar las ecuaciones teóricas)
* Cuál es el mejor modelo global, qué logró explicar este modelo en relación a los patrones que fueron observados sobre la serie
* Resultó mejor el ajuste y pronóstico global vs. lo local (cuál fue el mejor local entre Holt-Winters y combinación del filtro de la descomposición con loess) y por qué.

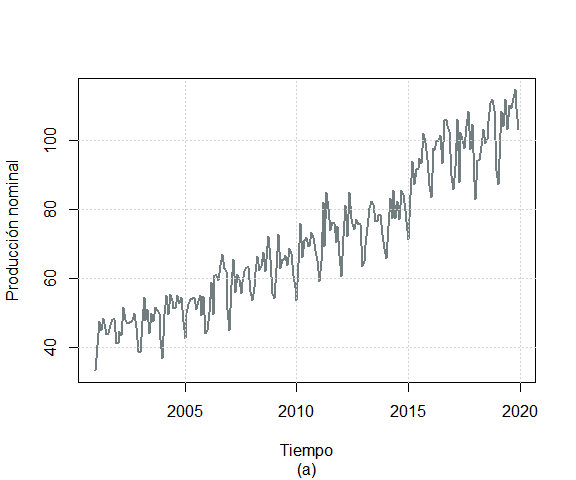
# 2. Análisis descriptivo de la serie, modelo global asignado y sus resultados:

1. Presente y analice brevemente la gráfica de la serie (y su logaritmo natural si la serie es multiplicativa) indicando los patrones observables de tendencia, estacionalidad, varianza, ciclos. Grafique y analice además la ACF de la serie (para el caso multiplicativo sólo presente y analice la ACF del logaritmo natural) y concluya en términos de estacionariedad o no y por qué, contrastando con lo que a partir de la gráfica de la serie se concluye al respecto.

Datos20=read.table("anex-EMMET-dic2019-Fabricacion de otros productos quimicos (1).csv",header=T,sep=";",skip=14,dec=",",colClasses=c(rep("NULL",4),"numeric",rep("NULL",6)))  
  
Datos20=ts(Datos20,freq=12,start=c(2001,1))

# #d8576b

plot(Datos20, lwd=2, xlab="Tiempo", ylab = "Producción nominal", col = '#717D7E')  
grid()  
title(sub= '(a)')



# title(main = "Índice de producción nominal \n del sector manufacturero (Colombia) \n",   
# sub = 'Clase industria: Otros productos químicos')

# Dispositivo JPEG  
jpeg(filename="Grafico3.jpeg", # Nombre del archivo y extension  
 width = 20, # Anchura  
 height = 20, # Altura  
 res= 100, # Resolucion 72ppi es un estandar  
 units = "cm") # Unidades.  
  
plot(Datos20, lwd=2, xlab="Tiempo", ylab = "Producción nominal", col = "#d8576b")  
grid()  
title(sub= '(a)')  
  
# Cerramos el dispositivo   
dev.off()

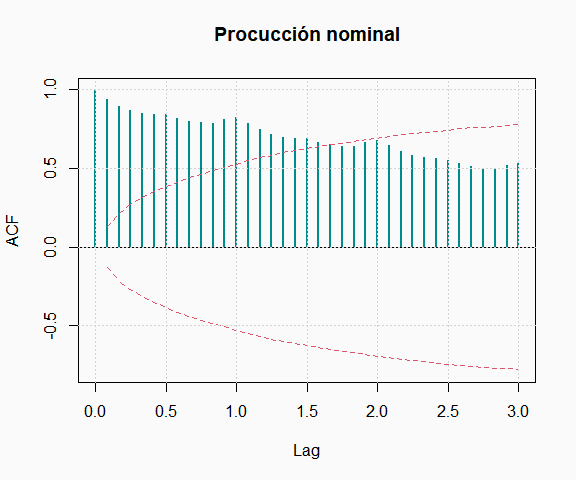
## png   
## 2

require(forecast)

## Loading required package: forecast

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

par(bg='gray98')  
acf(Datos20,lag.max=36,ci.type="ma",col="cyan4", ci.col=2, lwd=2, main = 'Procucción nominal')  
grid()



1. Para el modelo de regresión global señalado en la Tabla 1

* la ecuación teórica con sus supuestos

$$\text{Modelo cuadrático estacional con indicadoras, mes de referencia diciembre}\\
Y\_t=\beta\_0+\beta\_1t+\beta\_2t^2+\sum\_{i=1}^{11}\delta\_iI\_{i,t}+E\_t, ~~~ \{E\_t\}\_{t\in Z^+} \text{un RB} \sim N(0, \sigma^2)$$

* con la estrategia de validación cruzada usando la misma longitud n = 216 de ajuste del trabajo anterior ajuste nuevamente este modelo y reporte los resultados de ajuste

m <- 12 # numero de periodos a pronosticar dentro de la muestra  
n <-length(Datos20)-m # tamaño de la muestra para el ajuste  
t <- 1:n #Indice de tiempo en los periodos de ajuste  
  
# Datos para el ajuste:  
  
yt <- ts(Datos20[t], frequency = 12, start=c(2001, 1))  
  
# Creación de las variables indicadoras para los datos de muestra  
  
mes <- seasonaldummy(yt) #Matriz con las 11 primeras variables Indicadoras mes  
  
#Separando una a una las 11 variables indicadoras  
  
I1 <- mes[,1]  
I2 <- mes[,2]  
I3 <- mes[,3]  
I4 <- mes[,4]  
I5 <- mes[,5]  
I6 <- mes[,6]  
I7 <- mes[,7]  
I8 <- mes[,8]  
I9 <- mes[,9]  
I10 <- mes[,10]  
I11 <- mes[,11]  
  
# Creación de las variables indicadoras para los datos de validación cruzada  
  
tnuevo <- (n+1):length(Datos20)  
ytnuevo <- ts(Datos20[tnuevo], frequency = 12, start = c(2019, 1))  
  
  
mesnuevo <- seasonaldummy(yt, h=12)  
#Separando una a una las 11 indicadoras para los tiempos de pron?stico  
I1n=mesnuevo[,1]  
I2n=mesnuevo[,2]  
I3n=mesnuevo[,3]  
I4n=mesnuevo[,4]  
I5n=mesnuevo[,5]  
I6n=mesnuevo[,6]  
I7n=mesnuevo[,7]  
I8n=mesnuevo[,8]  
I9n=mesnuevo[,9]  
I10n=mesnuevo[,10]  
I11n=mesnuevo[,11]  
  
## Ajuste del Modelo cuadrático estacional con indicadoras  
  
mod1 <- lm(yt~t+I(t^2)+I1+I2+I3+I4+I5+I6+I7+I8+I9+I10+I11)

* tabla de parámetros estimados

$$\begin{array}{| c | c | c | c| c |}
\hline
Parametro&Estimación&Error~Estándar&T\_0&P(|T\_{202}|>|T\_0|)\\
\hline
\beta\_0&35.8242551&1.2250562&29.242949&0.0000000\\
\beta\_1&0.1319807&0.0174478&7.564309&0.0000000\\
\beta\_2&0.0007051&0.0000779&9.055350&0.0000000\\
\delta\_1&-3.1817818&1.3273503&-2.397093&0.0174368\\
\delta\_2&5.3069465&1.3272009&3.998601&0.0000893\\
\delta\_3&12.0220423&1.3270660&9.059114&0.0000000\\
\delta\_4&7.2690612&1.3269454&5.478041&0.0000001\\
\delta\_5&12.1146698&1.3268389&9.130475&0.0000000\\
\delta\_6&9.8255350&1.3267465&7.405736&0.0000000\\
\delta\_7&8.4294343&1.3266681&6.353838&0.0000000\\
\delta\_8&9.0930346&1.3266037&6.854372&0.0000000\\
\delta\_9&12.5163357&1.3265533&9.435230&0.0000000\\
\delta\_{10}&10.5493377&1.3265171&7.952659&0.0000000\\
\delta\_{11}&8.9475962&1.3264952&6.745291&0.0000000\\
\hline
\end{array}$$

summary(mod1)

##   
## Call:  
## lm(formula = yt ~ t + I(t^2) + I1 + I2 + I3 + I4 + I5 + I6 +   
## I7 + I8 + I9 + I10 + I11)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11.2796 -2.4057 -0.1672 2.2769 11.0489   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.582e+01 1.225e+00 29.243 < 2e-16 \*\*\*  
## t 1.320e-01 1.745e-02 7.564 1.35e-12 \*\*\*  
## I(t^2) 7.051e-04 7.787e-05 9.055 < 2e-16 \*\*\*  
## I1 -3.182e+00 1.327e+00 -2.397 0.0174 \*   
## I2 5.307e+00 1.327e+00 3.999 8.93e-05 \*\*\*  
## I3 1.202e+01 1.327e+00 9.059 < 2e-16 \*\*\*  
## I4 7.269e+00 1.327e+00 5.478 1.27e-07 \*\*\*  
## I5 1.211e+01 1.327e+00 9.130 < 2e-16 \*\*\*  
## I6 9.826e+00 1.327e+00 7.406 3.47e-12 \*\*\*  
## I7 8.429e+00 1.327e+00 6.354 1.37e-09 \*\*\*  
## I8 9.093e+00 1.327e+00 6.854 8.50e-11 \*\*\*  
## I9 1.252e+01 1.327e+00 9.435 < 2e-16 \*\*\*  
## I10 1.055e+01 1.327e+00 7.953 1.28e-13 \*\*\*  
## I11 8.948e+00 1.326e+00 6.745 1.57e-10 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.979 on 202 degrees of freedom  
## Multiple R-squared: 0.9589, Adjusted R-squared: 0.9562   
## F-statistic: 362.3 on 13 and 202 DF, p-value: < 2.2e-16

require(kableExtra)

## Loading required package: kableExtra

require(broom)

## Loading required package: broom

## Warning: package 'broom' was built under R version 4.0.4

# tidy(summary(mod1)) %>%   
# kbl() %>%   
# kable\_classic()

* medidas de ajuste

# Calculo del AIC y BIC  
  
#Creando funci?n usuario crit.inf.resid() para calcular C^\*\_n(p)  
crit.inf.resid <- function(residuales,n.par,AIC="TRUE"){  
if(AIC=="TRUE"){  
#Calcula AIC  
CI=log(mean(residuales^2))+2\*n.par/length(residuales)  
}  
if(AIC=="FALSE"){  
#Calcula BIC  
CI=log(mean(residuales^2))+n.par\*log(length(residuales))/length(residuales)  
}  
CI  
}

modelo 1

resmod1.orig <- residuals(mod1) #seudo-residuos en la escala original. Usados solo para calcular AIC y BIC  
  
npar1 <- length(coef(mod1)[coef(mod1)!=0]) #numero parametros modelo 1  
  
AIC1 <- exp(crit.inf.resid(resmod1.orig,n.par=npar1))  
BIC1 <- exp(crit.inf.resid(resmod1.orig ,n.par=npar1, AIC="FALSE"))

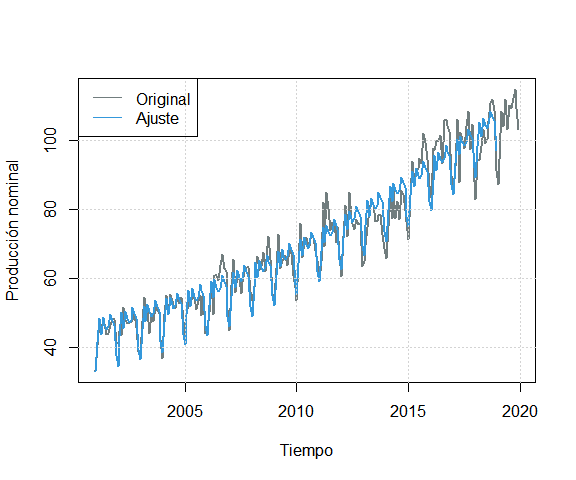
$$\begin{array}{| c | c | c |}
\hline
p&AIC&BIC\\
\hline
14&16.85948&20.98234\\
\hline
\end{array}$$

* gráfico del ajuste

# Valores ajustados de los modelos

mod1\_ajust <- ts(fitted(mod1), start = c(2001,1), frequency = 12)

plot(Datos20, ylab = 'Producción nominal', xlab = 'Tiempo', col = '#717D7E', lwd = 2)  
lines(mod1\_ajust, col='#3498DB', lwd=2)  
legend("topleft", legend = c("Original", "Ajuste"), lty=1, col=c('#717D7E', '#3498DB'))  
grid()



* pronósticos
* Ecuación de pronóstico

$$
\hat{Y}\_{216}(L) \approx 35.8242551 + 0.1319807 (216+L) + 0.0007051 (216+L)^2 \\
- 3.1817818 I\_{1,216+L} + 5.3069465 I\_{2,216+L} + 12.0220423 I\_{3,216+L} + 7.2690612 I\_{4,216+L} + 12.1146698 I\_{5,216+L} \\ + 9.8255350 I\_{6,216+L} + 8.4294343 I\_{7,216+L} + 9.0930346 I\_{8,216+L}+ 12.5163357 I\_{9,216+L}+ 10.5493377 I\_{10,216+L}+ 8.9475962 I\_{11,216+L}
$$

ytpron1 <- predict(mod1, newdata=data.frame(t=tnuevo, I1=I1n, I2=I2n, I3=I3n, I4=I4n, I5=I5n, I6=I6n, I7=I7n, I8=I8n, I9=I9n, I10=I10n, I11=I11n), interval="prediction")  
ytpron1 <- ts(ytpron1,freq=12,start=c(2019,1))  
ytpron1

## fit lwr upr  
## Jan 2019 94.4852 86.27520 102.6952  
## Feb 2019 103.4126 95.19741 111.6279  
## Mar 2019 110.5678 102.34725 118.7884  
## Apr 2019 106.2564 98.03028 114.4825  
## May 2019 111.5449 103.31316 119.7767  
## Jun 2019 109.7001 101.46257 117.9377  
## Jul 2019 108.7498 100.50627 116.9933  
## Aug 2019 109.8606 101.61094 118.1102  
## Sep 2019 113.7324 105.47658 121.9883  
## Oct 2019 112.2154 103.95319 120.4777  
## Nov 2019 111.0651 102.79632 119.3338  
## Dec 2019 102.5703 94.29486 110.8457

* la tabla de pronósticos

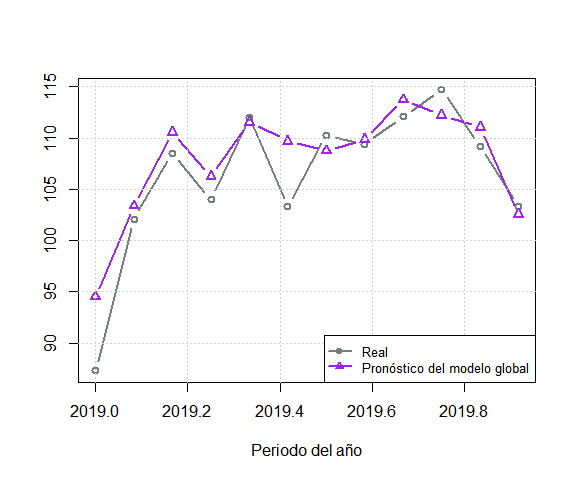
$$\begin{array}{| c | c | c | c| c |}
\hline
Periodo&L&Pronósticos&Lim.Inf&Lim.Sup\\
\hline
Ene~2019&1&94.4852&86.27520&102.6952\\
Feb~2019&2&103.4126&95.19741&111.6279\\
Mar~2019&3&110.5678&102.34725&118.7884\\
Abr~2019&4&106.2564&98.03028&114.4825\\
May~2019&5&111.5449&103.31316&119.7767\\
Jun~2019&6&109.7001&101.46257&117.9377\\
Jul~2019&7&108.7498&100.50627&116.9933\\
Ago~2019&8&109.8606&101.61094&118.1102\\
Sep~2019&9&113.7324&105.47658&121.9883\\
Oct~2019&10&112.2154&103.95319&120.4777\\
Nov~2019&11&111.0651&102.79632&119.3338\\
Dic~2019&12&102.5703&94.29486&110.8457\\
\hline
\end{array}$$

* medidas de cobertura amplitud media de los I.P y medidas MAE, MAPE y RMSE

$$\begin{array}{| c | c | c | c| c |}
\hline
RMSE & MAE & MAPE & Amplitud & Cobertura\\
\hline
3.153708&2.366654&2.338060&16.48278&100\\
\hline
\end{array}$$

* Conclusión breve sobre la calidad del ajuste y de los pronósticos con este modelo
* Gráfica de pronóstico

plot(ytnuevo,lwd=2, col="#717D7E", type="b", pch=1, xlab="Periodo del año", ylab="")  
lines(ytpron1[,1],lwd=2, col="purple", type="b", pch=2)  
grid()  
legend("bottomright",legend=c("Real","Pronóstico del modelo global"),col=c("#717D7E","purple"),lwd=2, pch=1:2, cex=0.8)

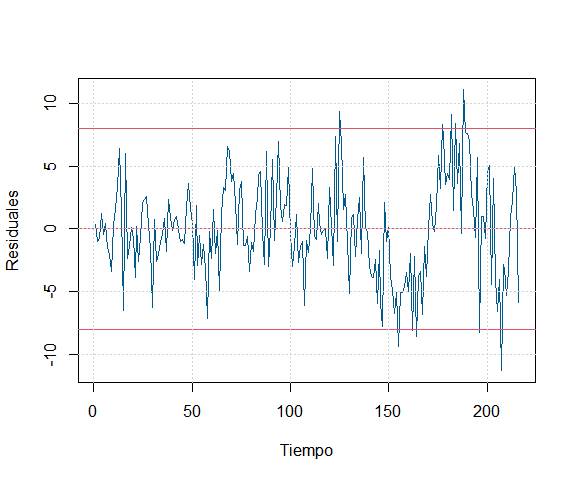


# 3. Evaluación de supuesto de ruido blanco e identificación de procesos estocásticos sobre los errores estructurales del modelo global

1. Validación de supuestos: Guarde los residuos estructurales en la escala en que ajustó la serie. Analice inicialmente las gráficas de estos residuales en términos de los supuestos sobre los errores Et de media constante en cero, varianza constante y determine si hay ciclos evidentes no explicados o rachas en signos ±, qué concluye frente a la existencia de estos patrones.

residuales <- residuals(mod1)

plot(residuales, type = "l", xlab = 'Tiempo', ylab = 'Residuales', col = '#01588A', lwd = 1)  
abline(h=c(-2\*summary(mod1)$sigma, 0, 2\*summary(mod1)$sigma), col=2)  
abline(h = 0, col=2)  
grid()



Realice las Pruebas de incorrelación con:

* Ljung-Box

Prueba de hipótesis

Estadístico de prueba

#DEFINIENDO FUNCION USUARIO PARA TESTES BOX-PIERCE Y LJUNG-BOX  
BP.LB.test=function(serie,maxlag,type="Box"){  
aux=floor(maxlag/6); X.squared=c(rep(NA,aux))  
df=c(rep(NA,aux)); p.value=c(rep(NA,aux))  
for(i in 1:aux){  
test=Box.test(serie,lag=(6\*i),type=type)  
X.squared[i]=test[[1]]; df[i]=test[[2]]  
p.value[i]=test[[3]]  
}  
lag=6\*c(1:aux)  
teste=as.data.frame(cbind(X.squared,df,p.value))  
rownames(teste)=lag; teste  
}  
  
LB\_result <- BP.LB.test(residuals(mod1),maxlag=36,type="Ljung")  
  
  
# cbind("m" = seq(6, 36, 6), LB\_result) %>%   
# kbl(col.names = c('$m$', '$Q\_{LB}$', '$gl$', '$p(χ^2\_{m} > Q\_{LB})$'), row.names = F) %>%   
# kable\_classic(full\_width = F)

xtable::xtable(LB\_result)

## % latex table generated in R 4.0.3 by xtable 1.8-4 package  
## % Thu May 27 12:37:33 2021  
## \begin{table}[ht]  
## \centering  
## \begin{tabular}{rrrr}  
## \hline  
## & X.squared & df & p.value \\   
## \hline  
## 6 & 161.91 & 6.00 & 0.00 \\   
## 12 & 196.24 & 12.00 & 0.00 \\   
## 18 & 216.31 & 18.00 & 0.00 \\   
## 24 & 259.14 & 24.00 & 0.00 \\   
## 30 & 315.97 & 30.00 & 0.00 \\   
## 36 & 385.48 & 36.00 & 0.00 \\   
## \hline  
## \end{tabular}  
## \end{table}

$$\begin{array}{cccc}
\hline
m & Q\_{LB} & gl & p(\chi^2\_{m} > Q\_{LB}) \\
\hline
6 & 161.91 & 6.00 & 0.00 \\
12 & 196.24 & 12.00 & 0.00 \\
18 & 216.31 & 18.00 & 0.00 \\
24 & 259.14 & 24.00 & 0.00 \\
30 & 315.97 & 30.00 & 0.00 \\
36 & 385.48 & 36.00 & 0.00 \\
\hline
\end{array}$$

* Durbin-Watson

require(car)

## Loading required package: car

## Loading required package: carData

#DEFINIENDO FUNCION USUARIO PARA TEST DURBIN-WATSON  
pruebaDW1=function(modelo){  
dwneg=durbinWatsonTest(modelo,max.lag=1,method="normal",alternative="negative")  
dwpos=durbinWatsonTest(modelo,max.lag=1,method="normal",alternative="positive")  
  
res=data.frame(1,dwneg$r,dwneg$dw,dwpos$p,dwneg$p)  
names(res)=c("lag","rho estimado","Estadístico D-W",  
"VP rho>0","VP rho<0")  
res  
}  
  
# pruebaDW1(mod1)   
  
  
# xtable::xtable(pruebaDW1(mod1))

$$
\begin{array}{cccccc}
\hline
& k & \widehat{\rho}(1) & \text{Estadístico } d\_1 & P(DW\_1 < d1) & P(DW\_1 > d1) \\
\hline
& 1 & 0.41 & 1.17 & 0.00 & 1.00 \\
\hline
\end{array}
$$

Modelo Durbin Watson

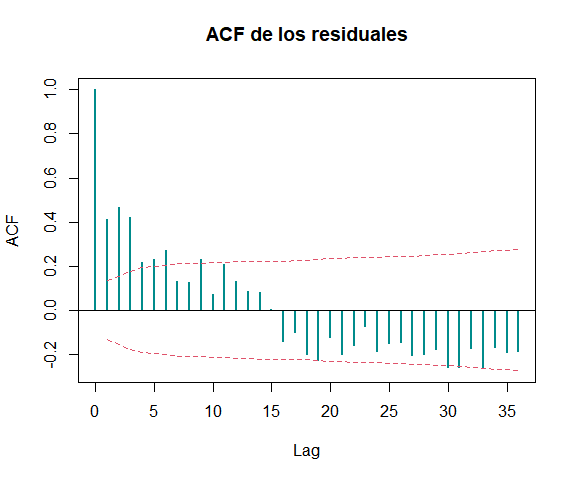
$$\text{Modelo Durbin Watson}\\
Y\_t= \sum\_{j=1}^2 \beta\_jt^j+\sum\_{i=1}^{11}\delta\_iI\_{i,t}+E\_t, \text{ con } E\_t = \phi\_1E\_{t-1} + a\_t$$

\* Como , por lo tanto el test a realizar es:

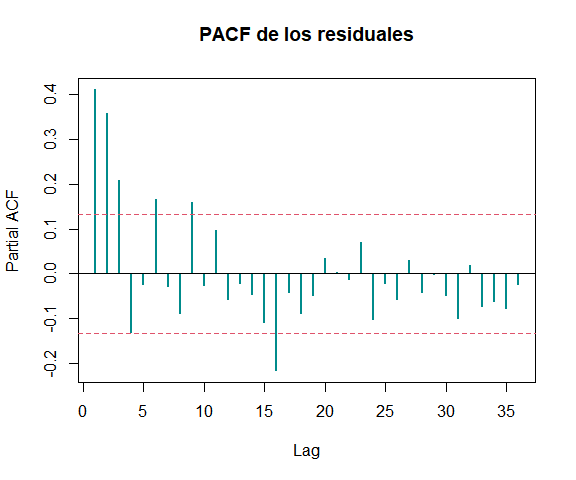
\* El valor p es: , conduce al rechazo de en favor de , por tanto en los dos modelos no es R.B pues presenta autocorrelación de orden 1 positiva.

* gráficas de la ACF y PACF con bandas de Bartlett.

acf(residuales,lag.max=36,ci.type="ma",col="cyan4",ci.col=2, lwd=2, main = 'ACF de los residuales')



pacf(residuales,lag.max=36,col="cyan4",ci.col=2, lwd=2, main = 'PACF de los residuales')



Concluya sobre si los errores estructurales Et son ruido blanco o no y en este último caso, evalúe si por lo menos son estacionarios con media cero.