Punto 6

Para los precios del petróleo crudo ajuste un modelo ARIMA(p,d,q) para la tasa de crecimiento y realice todos los diagnósticos necesarios. Comente.

Como primer paso, se obtienen los datos del petróleo crudo (oil) de la librería astsa, la cual tiene los datos de manera semanal desde el 2000 hasta mediados del 2010, y son los siguientes:

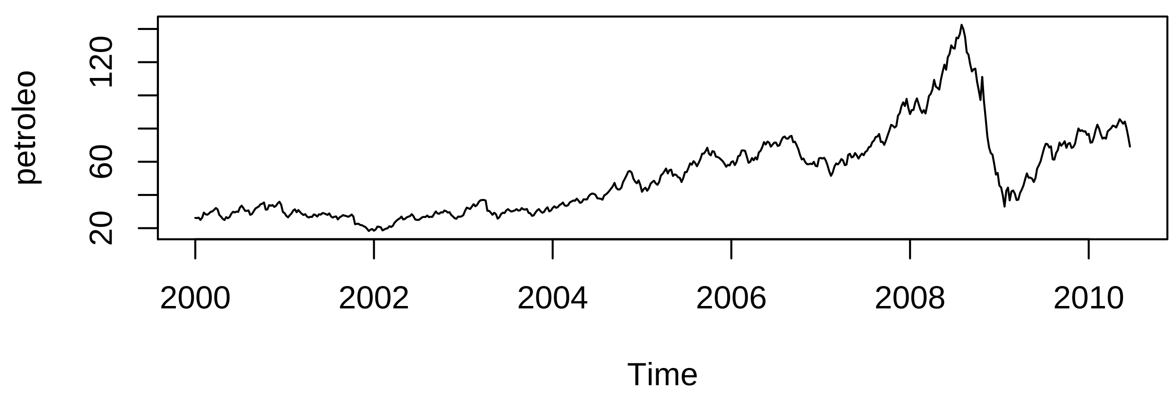
```r
library(forecast)
library(astsa)
library(knitr)
library(stats)
library(portes)
library(LSTS)

petroleo <- oil

head(petroleo)
```

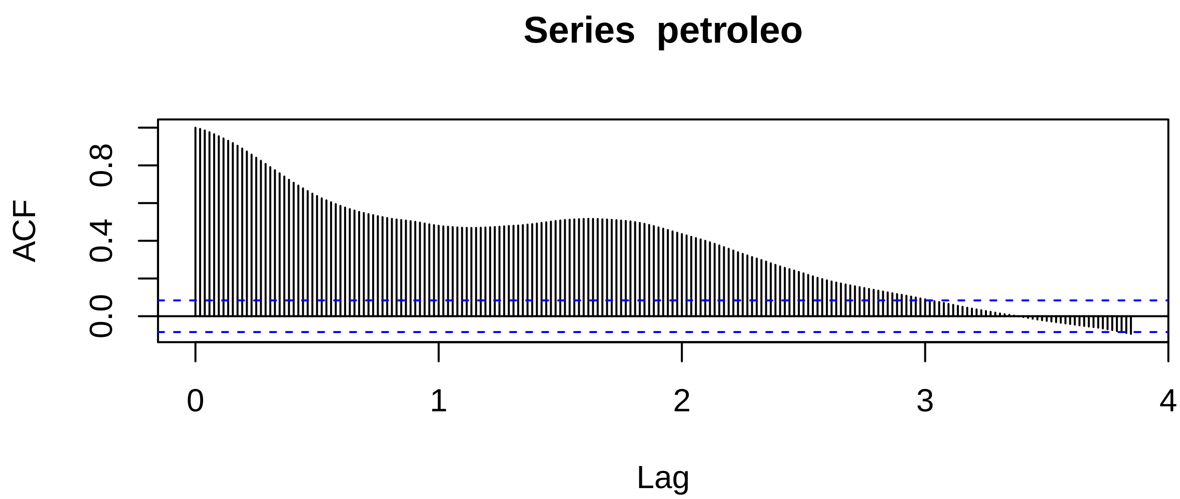
```
## Time Series:
## Start = c(2000, 1)
## End = c(2000, 6)
## Frequency = 52
## [1] 26.20 26.07 26.34 24.95 26.27 29.37
```

```r
plot.ts(petroleo)
```



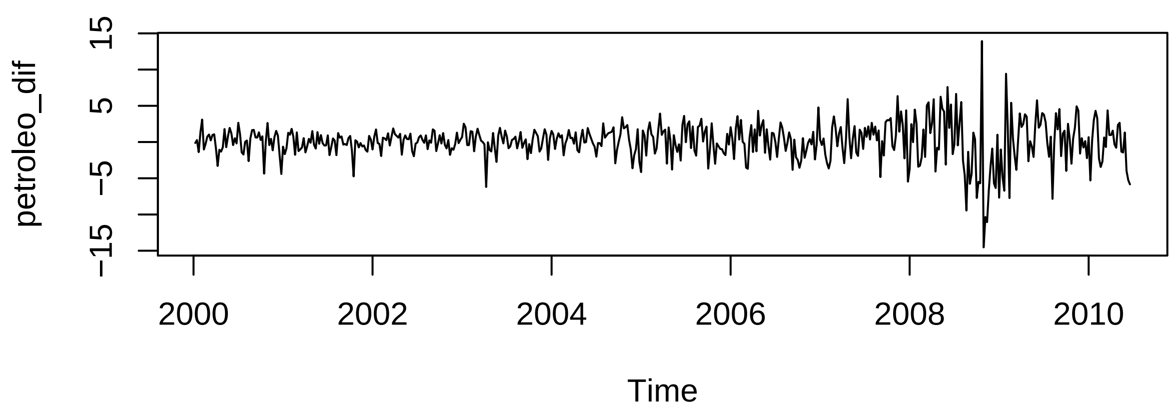
De esto se comienza a observar que la varianza sí parece ser constante pero la media no lo es, pues esta cambia en el tiempo. Esto se puede observar también el gráfico del ACF:

```r
acf(petroleo,lag.max = 200)
```



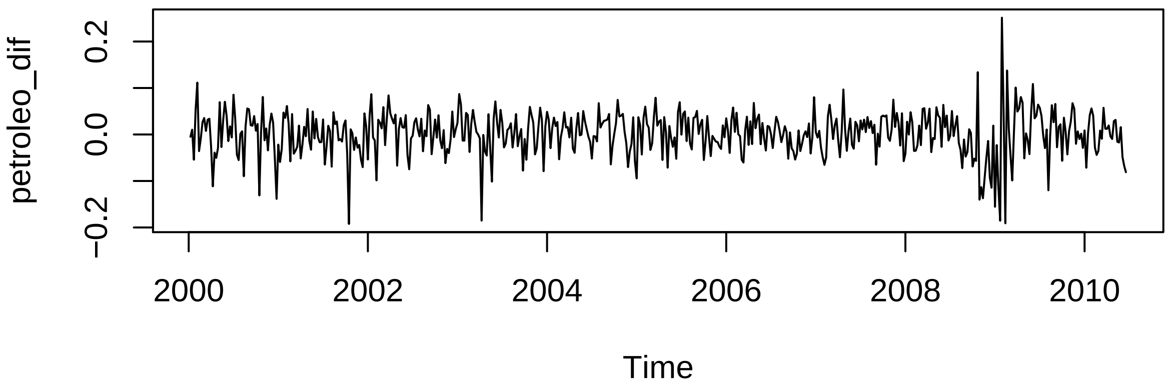
Se observa que el decaimiento es lento y por lo tanto se dice que la serie no es estacionaria en media, por lo que para que lo sea es necesario diferenciarla. Se procede a diferenciar la serie:

```
petroleo_dif <- diff(petroleo)
plot.ts(petroleo_dif)
```



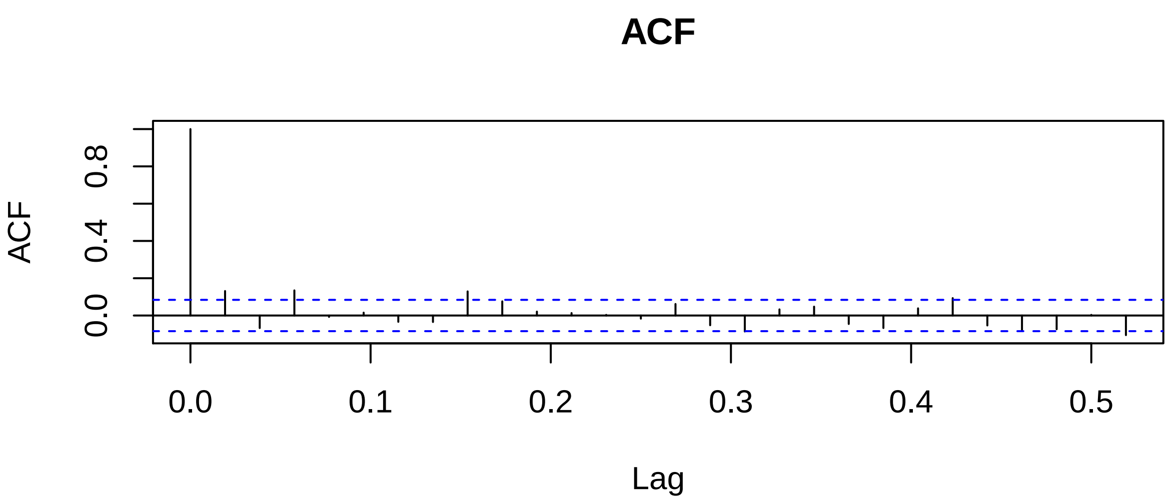
Se ve que la serie ya es estacionaria en media pero se puede notar el cambio que se genera en la varianza. Para esto, antes de diferenciar la serie se aplica el logaritmo natural a esta y luego se procede a diferenciarla.

```
petroleo_dif <- log(petroleo)
petroleo_dif <- diff(petroleo_dif)
plot.ts(petroleo_dif)
```

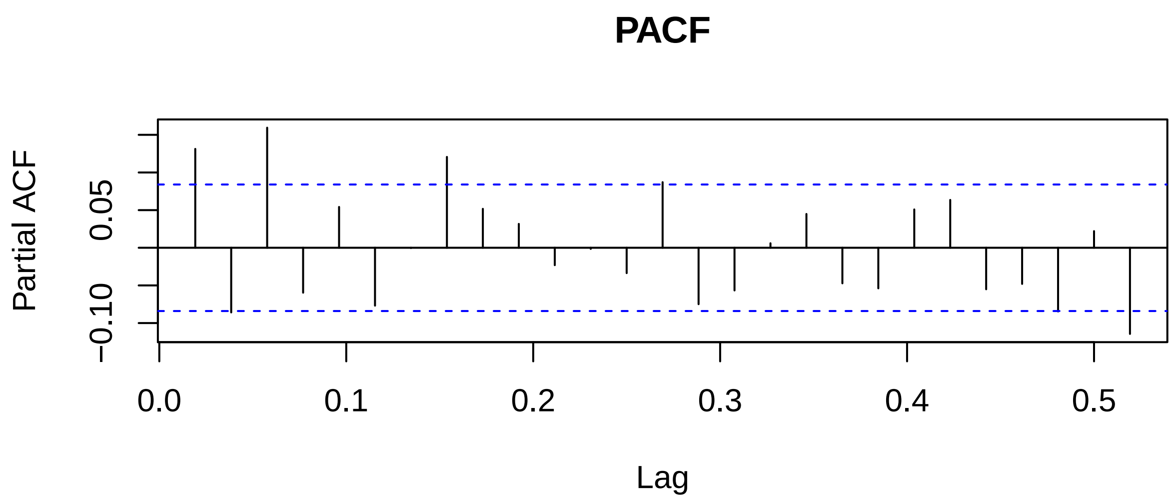


Ahora la serie parece si ser estacionaria en media y varianza, y los siguientes son los gráficos del ACF y PACF para ver si se pueden observar características para definir el modelo que mejor se ajusta a la serie:

```r
acf(petroleo_dif,main='ACF')
```




```r
pacf(petroleo_dif,main='PACF')
```



De los gráficos no se infiere de manera directa cual sería el modelo a ajustar para esta serie, por lo que se ajustarán varios modelos y se elegirá el que tenga menor coeficiente de información, usando el AIC y el BIC. Si se puede inferir que el orden del modelo ARIMA(p,d,q) que puede tener un buen ajuste, no es alto. Por lo tanto, se ajustan algunos modelos ARIMA(p,d,q) desde ARIMA(0,0,0) hasta ARIMA(4,2,3) pasando por todas las combinaciones de p, d y q , luego se comparan bajo los criterios de AIC y BIC:

```
z <- oil
  modelo <- function(i,k,j){
    m=arima(z,order=c(i,k,j))
    svec1 <- data.frame(m$aic,BIC(m),paste(as.character(c(i,k,j))
                                              ,collapse = '-'))
    return(svec1)
  }
  #ajuste de todos los posibles modelos hasta (4,2,4)
  n=4;i=0
  m=4;j=0
  l=2;k=0

  s <- modelo(0,0,0)
  while (i<=n){
    while (k<=l){
      while (j<=m){
        a=0
        try(a <- modelo(i,k,j), silent=TRUE)
        if ('try-error' %in% class(a)) next
        s <- rbind(s,a)
        j=j+1
      }
      j=0
      k=k+1
    }
    k=0
    i=i+1
  }
  colnames(s) <- c("AIC","BIC","P-D-Q")
```

```r
s <- na.omit(s)
s <- s[-1,]
#se muestra la tabla con los aic, bic de los modelos, con indice
s2 <- data.frame(s[1,],"|",s[2,]);i=3
while(i<nrow(s)){
  s2 <- rbind(s2,data.frame(s[i,],"|",s[i+1,]))
  i=i+2
}
colnames(s2) <- c("AIC","BIC","P-D-Q",".","AIC.","BIC.","P-D-Q.")
rownames(s2) <- NULL
kable(s2, format = "pandoc",digits = 4,
      booktabs = TRUE,caption="Modelos Ajustados")
```

Table 1: Modelos Ajustados

AIC	BIC	P-D-Q	.	AIC.	BIC.	P-D-Q.
5096.926	5105.527	0-0-0	|	4416.821	4429.723	0-0-1
3893.618	3910.821	0-0-2	|	3555.442	3576.946	0-0-3
3323.890	3349.694	0-0-4	|	2581.713	2586.012	0-1-0
2567.297	2575.895	0-1-1	|	2568.911	2581.808	0-1-2
2565.999	2583.195	0-1-3	|	2567.197	2588.691	0-1-4
2845.759	2850.056	0-2-0	|	2576.469	2585.063	0-2-1
2571.205	2584.096	0-2-2	|	2572.794	2589.982	0-2-3
2569.804	2591.290	0-2-4	|	2593.390	2606.292	1-0-0
2578.504	2595.707	1-0-1	|	2580.023	2601.526	1-0-2
2576.767	2602.571	1-0-3	|	2577.784	2607.889	1-0-4
2566.347	2574.945	1-1-0	|	2561.818	2574.715	1-1-1
2561.353	2578.549	1-1-2	|	2561.506	2583.001	1-1-3
2563.390	2589.183	1-1-4	|	2736.568	2745.163	1-2-0
2570.213	2583.104	1-2-1	|	2574.471	2591.659	1-2-2
2576.465	2597.951	1-2-3	|	2564.778	2590.560	1-2-4
2577.394	2594.597	2-0-0	|	2569.763	2591.267	2-0-1
2569.316	2595.121	2-0-2	|	2569.296	2599.402	2-0-3
2573.863	2608.269	2-0-4	|	2567.843	2580.740	2-1-0
2561.701	2578.897	2-1-1	|	2561.448	2582.943	2-1-2
2563.186	2588.979	2-1-3	|	2565.186	2595.278	2-1-4
2667.104	2679.995	2-2-0	|	2575.520	2592.709	2-2-1
2576.463	2597.948	2-2-2	|	2568.039	2593.821	2-2-3
2567.513	2597.593	2-2-4	|	2578.744	2600.248	3-0-0
2581.164	2606.968	3-0-1	|	2569.336	2599.442	3-0-2
2574.758	2609.164	3-0-3	|	2574.426	2613.133	3-0-4
2564.502	2581.698	3-1-0	|	2561.906	2583.400	3-1-1
2563.181	2588.975	3-1-2	|	2554.099	2584.191	3-1-3
2555.735	2590.127	3-1-4	|	2647.048	2664.236	3-2-0
2568.209	2589.694	3-2-1	|	2565.181	2590.964	3-2-2
2570.633	2600.713	3-2-3	|	2557.355	2591.732	3-2-4
2574.852	2600.657	4-0-0	|	2569.707	2599.812	4-0-1
2571.026	2605.433	4-0-2	|	2565.472	2604.179	4-0-3
2575.954	2618.961	4-0-4	|	2566.292	2587.787	4-1-0
2563.709	2589.503	4-1-1	|	2565.168	2595.261	4-1-2
2564.194	2598.586	4-1-3	|	2557.961	2596.651	4-1-4
2626.086	2647.572	4-2-0	|	2569.970	2595.753	4-2-1
2572.209	2602.288	4-2-2	|	2572.017	2606.393	4-2-3

AIC	BIC	P-D-Q	.	AIC.	BIC.	P-D-Q.

```r
#seleccion de los mejores cinco modelos
s1 <- s[1,]
for (i in 1:5){
  pos <- which(s[,1]==min(s[,1]), arr.ind=TRUE)
  s1[(nrow(s1)+1),] <- s[pos,]
  s[pos,] <- NA
  s <- na.omit(s)
}
s1 <- s1[-1,]
s1mod <- c("Modelo 1","Modelo 2","Modelo 3","Modelo 4","Modelo 5")
s1 <- cbind(s1mod,s1)
colnames(s1) <- c("Modelos",colnames(s1[,-1]))
rownames(s1) <- NULL

kable(s1, format = "pandoc",digits = 4,
      booktabs = TRUE,caption="Los 5 Mejores Modelos")
```

Table 2: Los 5 Mejores Modelos

Modelos	AIC	BIC	P-D-Q
Modelo 1	2554.099	2584.191	3-1-3
Modelo 2	2555.735	2590.127	3-1-4
Modelo 3	2557.355	2591.732	3-2-4
Modelo 4	2557.961	2596.651	4-1-4
Modelo 5	2561.353	2578.549	1-1-2

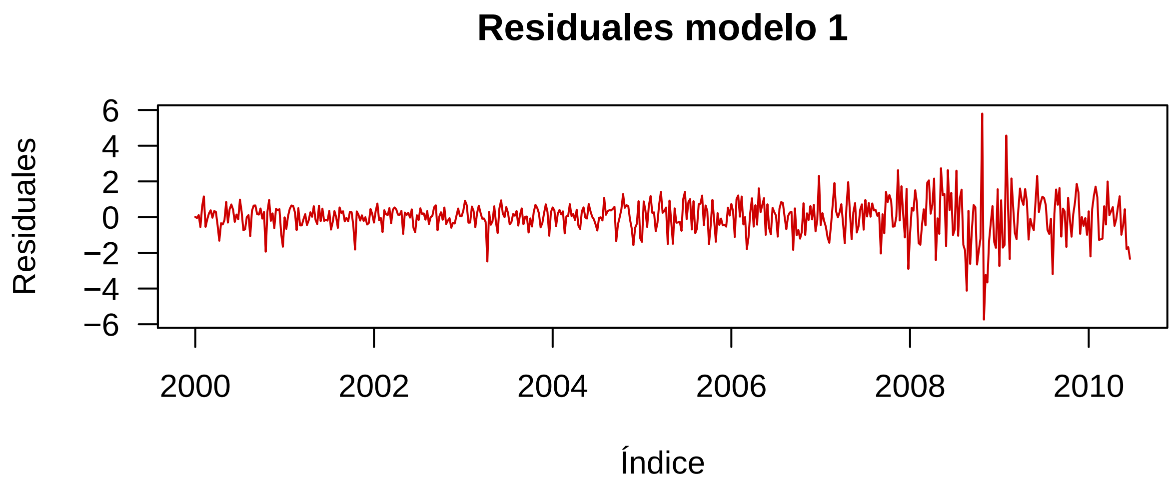
Se observa que el mejor modelo según los criterios de información es el ARIMA(3,1,3) seguido por el ARIMA(4,1,3). Entre estos dos se elige como mejor modelo el ARIMA(3,1,3) ya que tiene menor cantidad de parámetros y menor valor en los criterios de información calculados.

Luego de esto, se hacen los diagnósticos para este modelo, los cuales se hacen para los residuales del modelo ajustado.

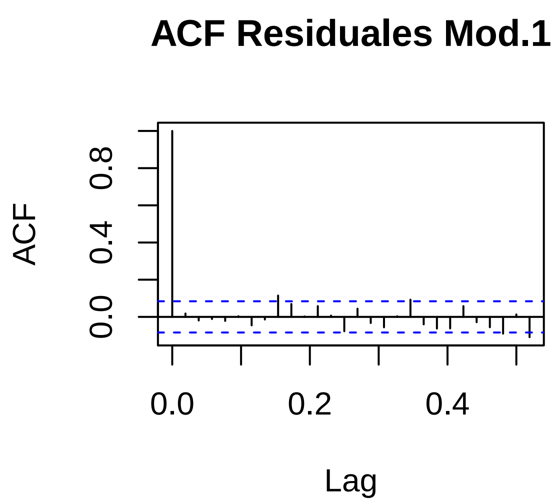
Primero, el gráfico de los residuales estandarizados:

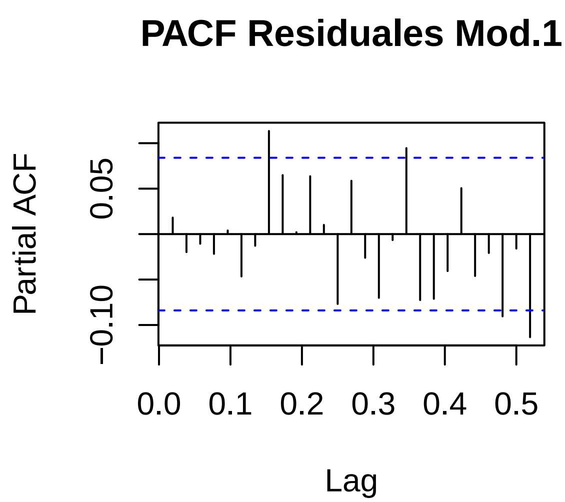
```r
m1=arima(z,order=c(3,1,3))

par(mfrow=c(1,1))
plot(residuals(m1)/sqrt(as.numeric(m1[2]))
    ,main="Residuales modelo 1",
    ylab="Residuales",
    xlab="Índice",col="Red3",las=1,lwd=1)
```




```r
par(mfrow=c(1,2))
acf(residuals(m1)/sqrt(as.numeric(m1[2] ))
    ,main="ACF Residuales Mod.1")
pacf(residuals(m1)/sqrt(as.numeric(m1[2] ))
    ,main="PACF Residuales Mod.1")
```





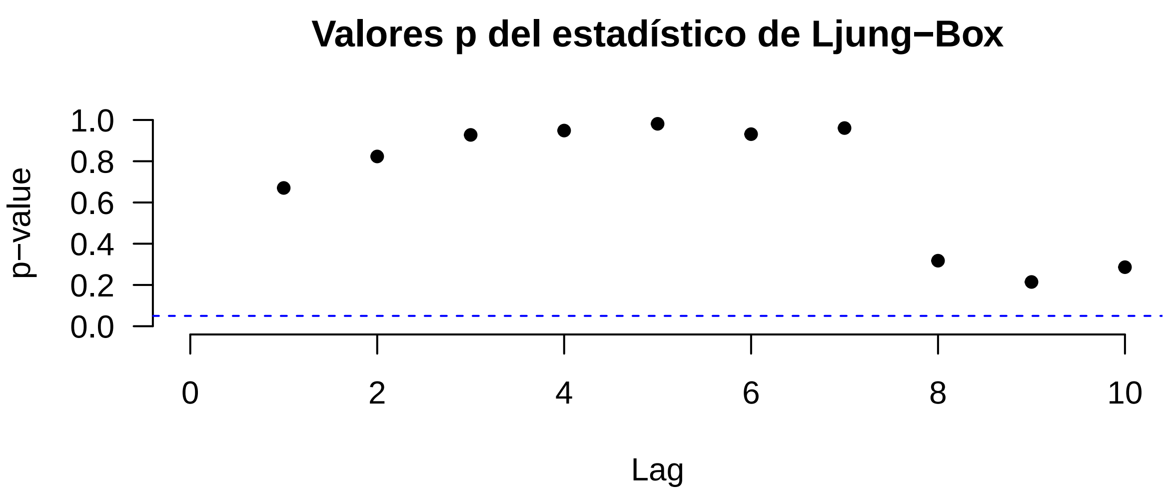
No se observan tendencias del gráfico de los residuales y además el ACF y el PACF sugieren que estos son aleatorios. Luego, se hace la prueba de Ljung-Box para saber si estos son independientes, es decir, para probar H0: los residuales se distribuyen de forma independiente vs. H1: los residuales no se distribuyen de forma independiente.

```r
adec1 <- LjungBox(residuals(m1)/sqrt(as.numeric(m1[2] )))
rownames(adec1) <- NULL
kable(adec1, format = "pandoc",digits = 4,
      booktabs = TRUE,caption="Modelo Adecuado")
```

Table 3: Modelo Adecuado

lags	statistic	df	p-value
5	0.7279	5	0.9814
10	11.9824	10	0.2862
15	18.9343	15	0.2167
20	28.7257	20	0.0933
25	39.7377	25	0.0310
30	49.4885	30	0.0140

```
par(mfrow=c(1,1))
Box.Ljung.Test(residuals(m1)/sqrt(as.numeric(m1[2] )), main = 'Valores p del estadístico de Ljung-Box')
```

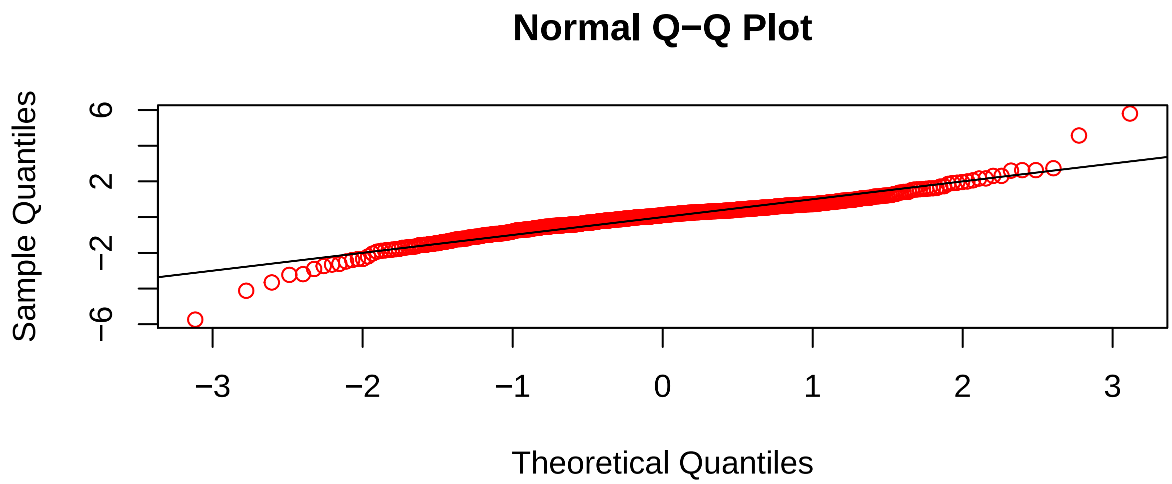


Se observa que hasta 20 rezagos el valor p es mayor a 0.05 y por tanto no se rechaza la hipótesis nula de que los residuales son independientes. Esto mismo se observa en el gráfico que tiene hasta 10 rezagos. Y por último se hace el QQplot para ver la normalidad de los mismos.

```
shapiro.test(residuals(m1)/sqrt(as.numeric(m1[2])))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(m1)/sqrt(as.numeric(m1[2]))
## W = 0.93407, p-value = 8.857e-15
```

```
qqnorm(residuals(m1)/sqrt(as.numeric(m1[2] )),col='red')
abline(0,1)
```



El qqplot puede sugerir que se cumple normalidad, aunque se observa que las colas de la distribución son pesadas, pero con la prueba de Shapiro-Wilk se concluye que se rechaza la hipótesis nula de normalidad y por lo tanto se dice que los residuales no provienen de una distribución normal. Esto puede ser influido principalmente por los datos extremos de los residuales los cuales hacen que las colas de su distribución sean pesadas, pero esta característica se presenta en varios modelos ajustados. De esta manera, se dice que el ARIMA(3,1,3) es el mejor modelo ajustado según los criterios del AIC y BIC.

Punto 7

Ajuste un ARIMA(p,d,q) para los datos de la temperatura global gtemp realizando todos los diagnósticos necesarios. Luego realice un pronóstico para los próximos 10 años. Comente.

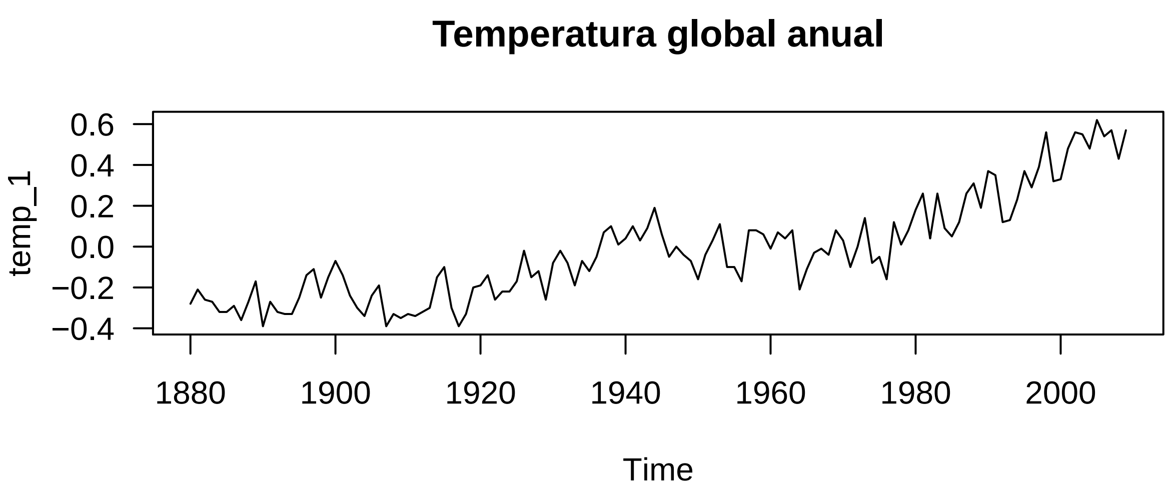
Primero, se procede a cargar los datos de gtemp de la librería astsa, los cuales son observaciones anuales de la temperatura media global, desde 1880 hasta 2009, tomada en grados centígrados.

Los datos son los siguientes:

```
temp_1 <- gtemp
head(temp_1)
```

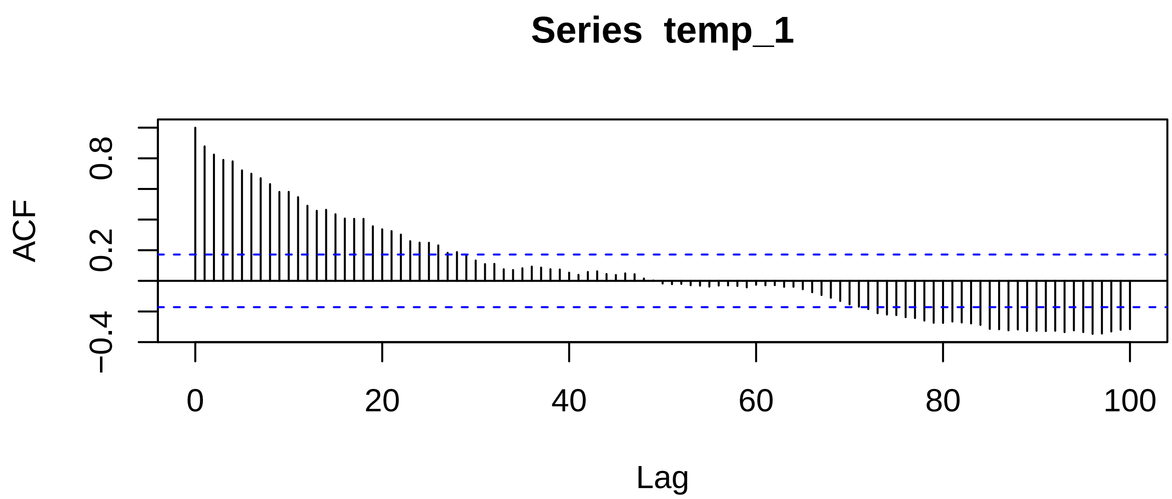
```
## Time Series:
## Start = 1880
## End = 1885
## Frequency = 1
## [1] -0.28 -0.21 -0.26 -0.27 -0.32 -0.32
```

```
plot.ts(temp_1,las=1,main='Temperatura global anual')
```



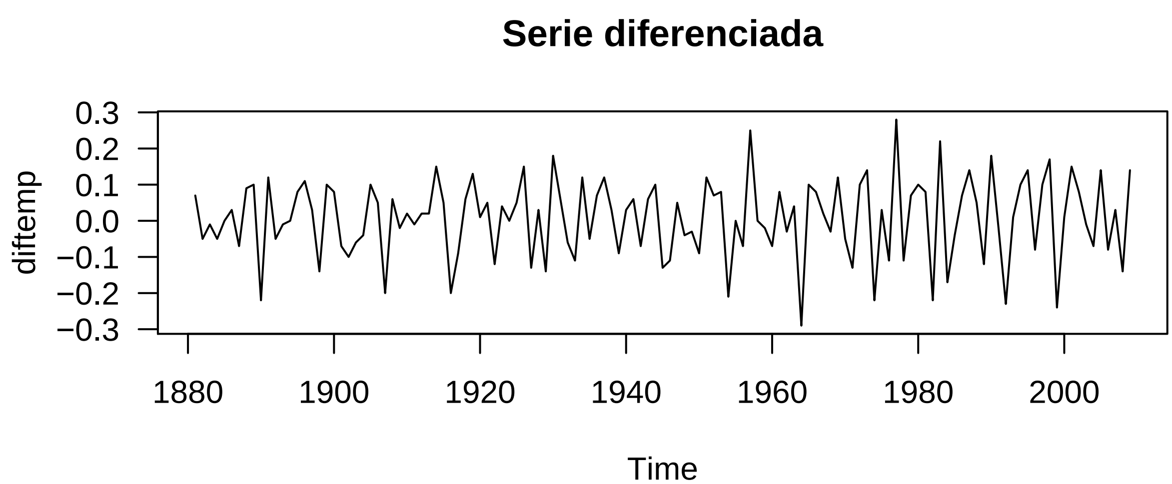
Se observa que la varianza puede ser constante pero la media notablemente no es constante en el tiempo y esto se verifica con el ACF:

```
acf(temp_1,lag.max = 100)
```

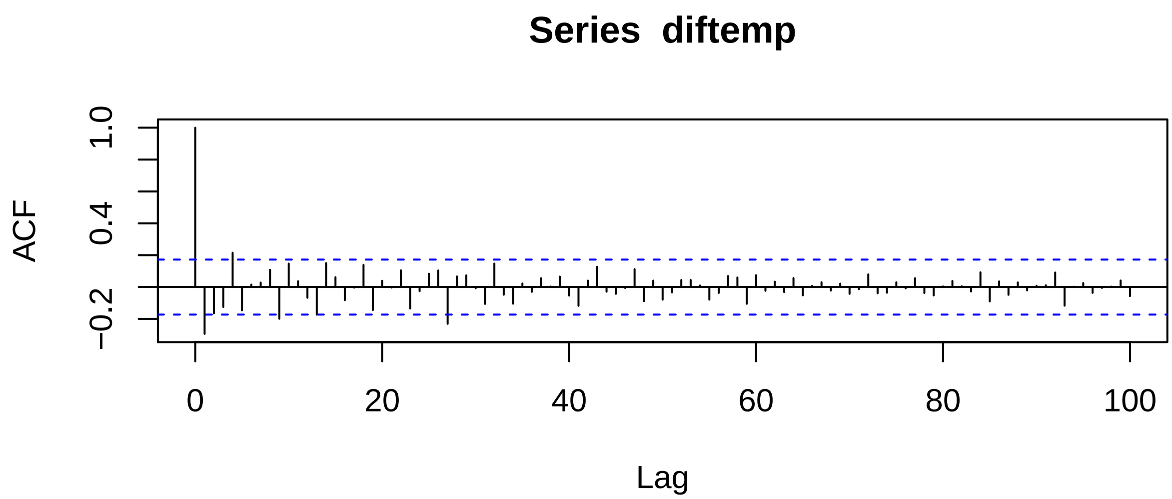


Se concluye que la serie no es estacionaria en media y por tanto la mejor opción es diferenciarla:

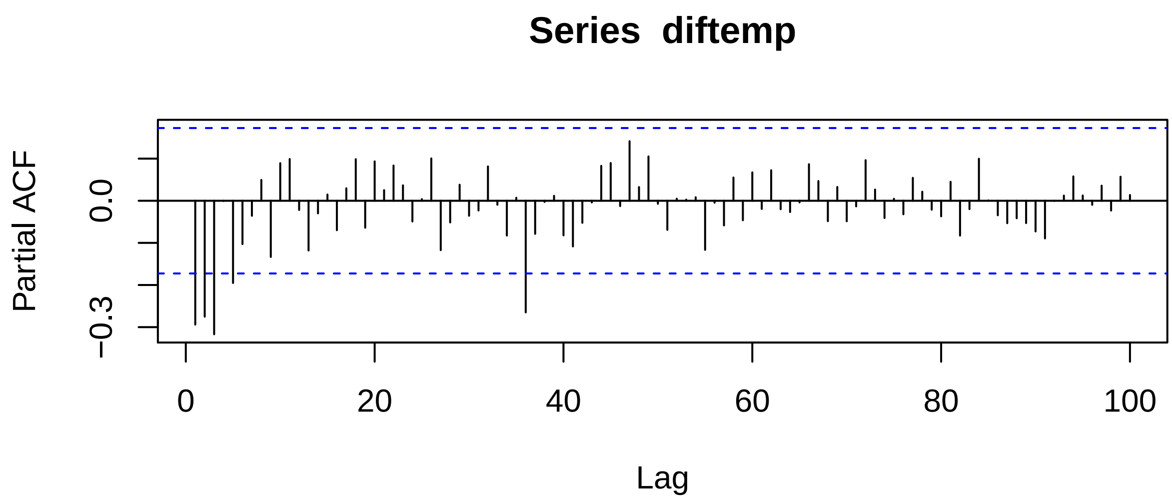
```
diftemp <- diff(temp_1)
plot.ts(diftemp,las=1,main='Serie diferenciada')
```




```
acf(diftemp,lag.max = 100)
```




```
pacf(diftemp,lag.max = 100)
```



Se observa así, que la serie se vuelve estacionaria cuando esta se diferencia, pero no se infiere de manera directa un modelo ARIMA desde los gráficos ACF y PACF, pero se sospecha que el orden de este modelo no es alto. Como siguiente paso se ajustan todos las combinaciones de modelos desde el ARIMA(0,0,0) hasta el ARIMA(4,2,3) de manera similar al punto anterior, y se obtiene:

```
z <- gtemp
  modelo <- function(i,k,j){
    m=arima(z,order=c(i,k,j))
    svec1 <- data.frame(m$aic,BIC(m),paste(as.character(c(i,k,j))
                                              ,collapse = '-'))
    return(svec1)
  }
  #ajuste de todos los posibles modelos hasta (4,2,4)
  n=4;i=0
```

```r
m=4;j=0
l=2;k=0

s <- modelo(0,0,0)
while (i<=n){
  while (k<=l){
    while (j<=m){
      a=0
      try(a <- modelo(i,k,j), silent=TRUE)
      if ('try-error' %in% class(a)) next
      s <- rbind(s,a)
      j=j+1
    }
    j=0
    k=k+1
  }
  k=0
  i=i+1
}
colnames(s) <- c("AIC","BIC","P-D-Q")
s <- na.omit(s)
s <- s[-1,]
#se muestra la tabla con los aic, bic de los modelos, con indice
s2 <- data.frame(s[1,],"|",s[2,]);i=3
while(i<nrow(s)){
  s2 <- rbind(s2,data.frame(s[i,],"|",s[i+1,]))
  i=i+2
}
colnames(s2) <- c("AIC","BIC","P-D-Q",".","AIC.","BIC.","P-D-Q.")
rownames(s2) <- NULL
kable(s2, format = "pandoc",digits = 4,
      booktabs = TRUE,caption="Modelos Ajustados")
```

Table 4: Modelos Ajustados

AIC	BIC	P-D-Q	.	AIC.	BIC.	P-D-Q.
11.3523	17.0874	0-0-0	|	-78.2661	-69.6635	0-0-1
-133.7879	-122.3177	0-0-2	|	-140.7901	-126.4525	0-0-3
-161.3933	-144.1881	0-0-4	|	-202.6198	-199.7599	0-1-0
-226.3860	-220.6663	0-1-1	|	-228.4156	-219.8362	0-1-2
-226.4234	-214.9841	0-1-3	|	-230.0916	-215.7926	0-1-4
-79.4941	-76.6421	0-2-0	|	-193.6422	-187.9381	0-2-1
-218.4133	-209.8572	0-2-2	|	-222.0425	-210.6344	0-2-3
-220.1279	-205.8677	0-2-4	|	-202.9290	-194.3264	1-0-0
-212.0653	-206.3456	1-1-0	|	-227.6518	-219.0724	1-1-1
-226.4179	-214.9787	1-1-2	|	-229.1147	-214.8156	1-1-3
-228.3384	-211.1795	1-1-4	|	-122.8357	-117.1317	1-2-0
-202.7846	-194.2285	1-2-1	|	-221.3286	-209.9205	1-2-2
-220.0752	-205.8150	1-2-3	|	-223.0767	-205.9646	1-2-4
-209.9385	-198.4683	2-0-0	|	-205.3035	-190.9658	2-0-1
-221.8152	-204.6100	2-0-2	|	-219.5121	-210.9327	2-1-0
-227.5186	-216.0793	2-1-1	|	-226.7058	-212.4067	2-1-2
-231.9874	-214.8286	2-1-3	|	-227.7494	-207.7307	2-1-4

AIC	BIC	P-D-Q	.	AIC.	BIC.	P-D-Q.
-140.8224	-132.2664	2-2-0	|	-210.2413	-198.8332	2-2-1
-220.7992	-206.5390	2-2-2	|	-219.0900	-201.9778	2-2-3
-222.6262	-202.6620	2-2-4	|	-216.0319	-201.6942	3-0-0
-222.9568	-205.7516	3-0-1	|	-221.4917	-201.4189	3-0-2
-229.7594	-218.3201	3-1-0	|	-228.3175	-214.0184	3-1-1
-229.5338	-212.3749	3-1-2	|	-227.5721	-207.5534	3-1-3
-225.8622	-202.9837	3-1-4	|	-176.5644	-165.1563	3-2-0
-221.0264	-206.7663	3-2-1	|	-219.3283	-202.2161	3-2-2
-222.3051	-202.3409	3-2-3	|	-220.4835	-197.6673	3-2-4
-225.4958	-208.2906	4-0-0	|	-224.2544	-204.1817	4-0-1
-227.8408	-213.5417	4-1-0	|	-226.4421	-209.2833	4-1-1
-227.5814	-207.5627	4-1-2	|	-227.3668	-204.4883	4-1-3
-179.7482	-165.4880	4-2-0	|	-219.0452	-201.9330	4-2-1
-219.2832	-199.3190	4-2-2	|	-220.5718	-197.7555	4-2-3

```r
#seleccion de los mejores cinco modelos
s1 <- s[1,]
for (i in 1:5){
  pos <- which(s[,1]==min(s[,1]), arr.ind=TRUE)
  s1[(nrow(s1)+1),] <- s[pos,]
  s[pos,] <- NA
  s <- na.omit(s)
}
s1 <- s1[-1,]
s1mod <- c("Modelo 1","Modelo 2","Modelo 3","Modelo 4","Modelo 5")
s1 <- cbind(s1mod,s1)
colnames(s1) <- c("Modelos",colnames(s1[,-1]))
rownames(s1) <- NULL

kable(s1, format = "pandoc",digits = 4,
      booktabs = TRUE,caption="Los 5 Mejores Modelos")
```

Table 5: Los 5 Mejores Modelos

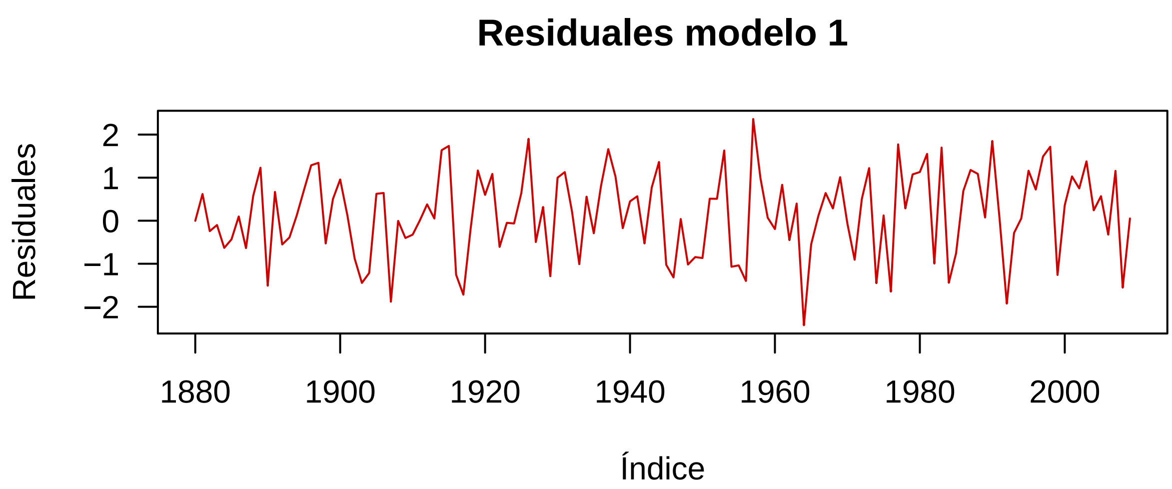
Modelos	AIC	BIC	P-D-Q
Modelo 1	-231.9874	-214.8286	2-1-3
Modelo 2	-230.0916	-215.7926	0-1-4
Modelo 3	-229.7594	-218.3201	3-1-0
Modelo 4	-229.5338	-212.3749	3-1-2
Modelo 5	-229.1147	-214.8156	1-1-3

Se dice que el mejor modelo ARIMA según los criterios de información es el ARIMA(2,1,3), al cual se le realizarán las pruebas de diagnóstico:

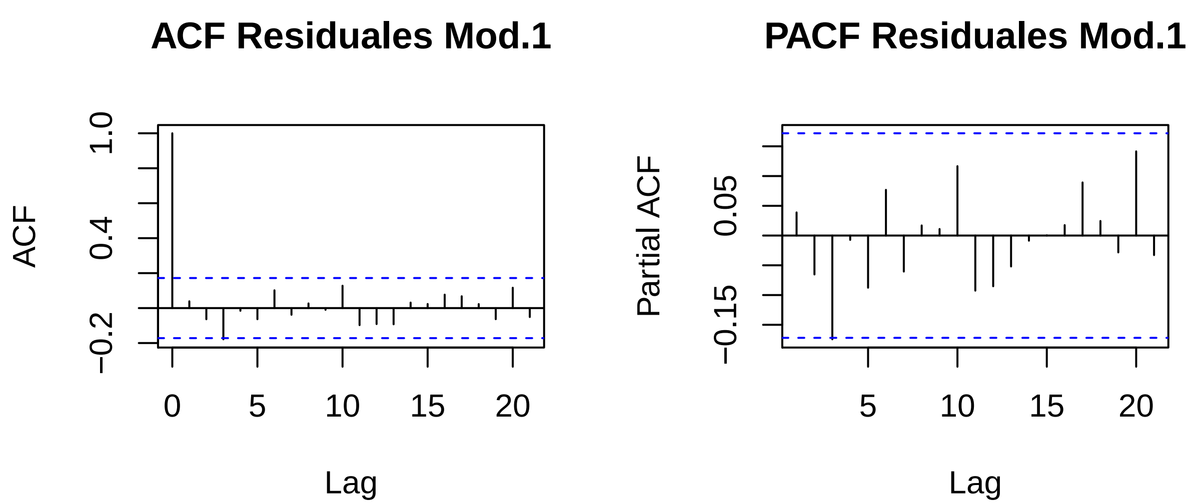
```r
m1=arima(z,order=c(2,1,3))

par(mfrow=c(1,1))
plot(residuals(m1)/sqrt(as.numeric(m1[2]))
    ,main="Residuales modelo 1",
```

```
    ylab="Residuales",
    xlab="Índice",col="Red3",las=1,lwd=1)
```




```
par(mfrow=c(1,2))
acf(residuals(m1)/sqrt(as.numeric(m1[2] ))
    ,main="ACF Residuales Mod.1")
pacf(residuals(m1)/sqrt(as.numeric(m1[2] ))
    ,main="PACF Residuales Mod.1")
```

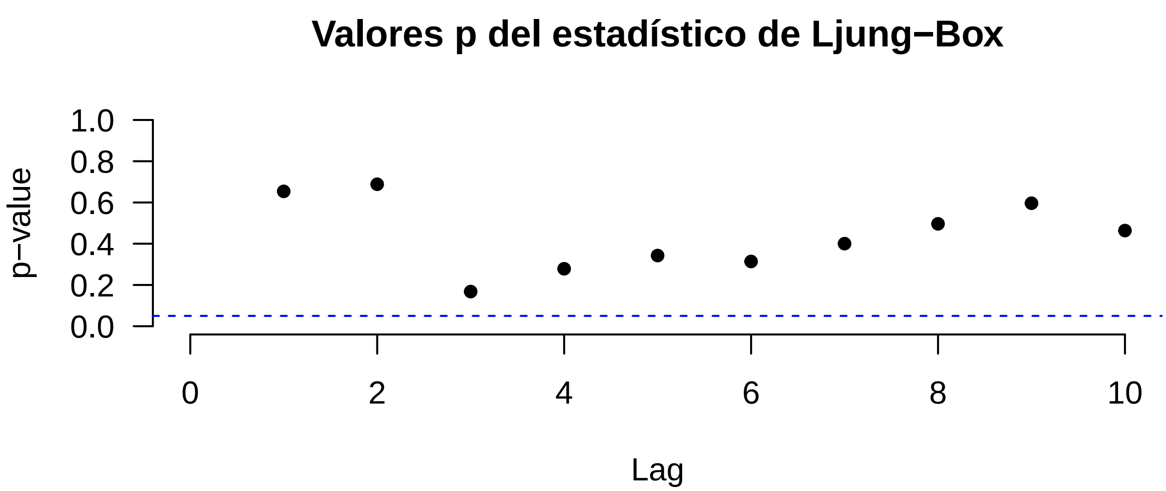



```
adec1 <- LjungBox(residuals(m1)/sqrt(as.numeric(m1[2] )))
rownames(adec1) <- NULL
kable(adec1, format = "pandoc",digits = 4,
      booktabs = TRUE,caption="Modelo Adecuado")
```

Table 6: Modelo Adecuado

lags	statistic	df	p-value
5	5.6424	5	0.3426
10	9.7381	10	0.4638
15	13.8116	15	0.5399
20	18.2146	20	0.5733
25	18.9272	25	0.8006
30	24.5578	30	0.7462

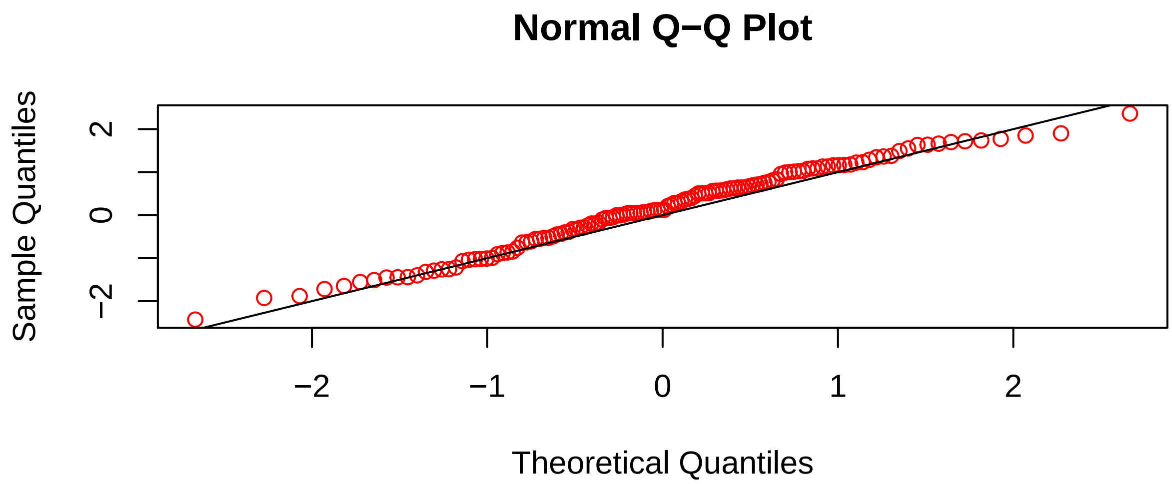
```r
par(mfrow=c(1,1))
Box.Ljung.Test(residuals(m1)/sqrt(as.numeric(m1[2] )), main = 'Valores p del estadístico de Ljung-Box')
```




```r
shapiro.test(residuals(m1)/sqrt(as.numeric(m1[2])))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  residuals(m1)/sqrt(as.numeric(m1[2]))
## W = 0.9862, p-value = 0.2146
```

```r
qqnorm(residuals(m1)/sqrt(as.numeric(m1[2] )),col='red')
abline(0,1)
```



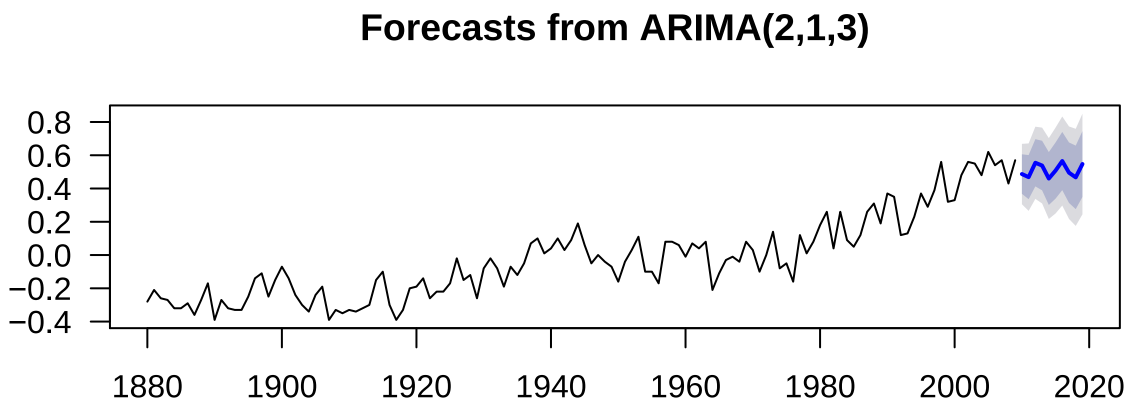
Del gráfico de los residuales estandarizados se observa la aleatoriedad de los mismos, no tienen una tendencia en particular como para decir que son descritos por un modelo y esto se confirma con la anulación de los rezagos en los gráficos del ACF y PACF de estos. Para la prueba de Ljung-Box, no se rechaza la hipótesis nula de independencia de los residuales para ningún rezago, es decir, estos son independientes. Por último, de la prueba de normalidad de Shapiro-Wilk y del qqplot se concluye que los residuales provienen de una población con distribución normal, es decir, que son normales o cumplen el supuesto de normalidad, pues no se rechaza la hipótesis nula de normalidad.

Luego de esto, se procede a realizar el pronóstico para los siguientes 10 años, es decir, desde 2010 hasta 2019. Esto se hace usando forecast basándose en el modelo ARIMA ajustado y se obtiene:

```
pronostico <- forecast(m1,h=10)
pronostico$mean
```

```
## Time Series:
## Start = 2010
## End = 2019
## Frequency = 1
##  [1] 0.4869025 0.4684238 0.5547372 0.5380692 0.4599999 0.5078698 0.5652524
##  [8] 0.4950756 0.4670146 0.5472765
```

```
plot(pronostico,las=1)
```



Así, se obtiene el pronóstico para la temperatura media anual desde 2010 hasta 2019 con el gráfico del mismo.