UNAP

FINESI

Series de Tiempo

VI Semestre

Grupo: B

EJEMPLOS

Construcción de Modelos ARIMA

Mtr. Alcides Ramos Calcina

FINESI

Ejemplo 2

Se tiene la serie de Índices reales de precios de combustibles y de tarifas de servicios públicos - Precios de Combustibles - Gasohol 90 Octanos, del periodo enero 2010 a marzo 2023. Los datos se muestran en la Tabla 6.2.

Tabla 6.2

Precios de Combustibles - Gasohol 90 Octanos (Enero 2010 – Marzo 2023).

Año	Ene	Feb	Mar	Abr	May	Jun	Jul	Ago	Set	Oct	Nov	Dic
2010	95.24	95.2	94.85	96.25	100.92	100.85	101.99	101.54	100.6	101.45	105.6	105.5
2011	110.19	109.86	108.66	107.83	107.77	111.38	112.22	113.46	116.23	115.78	117.31	116.91
2012	113.32	113.62	117.52	117.39	121.33	121.37	113.05	111.99	115.4	115.43	108.59	105.11
2013	105.89	109.59	113.49	108.66	106.18	106.79	109.1	113.31	113.1	107.95	106.19	110.33
2014	112.13	112.09	114.74	114.6	112.86	111.74	111.96	110.03	109.86	109.37	105.34	92.11
2015	80.14	79.05	85.1	84.16	86.33	88.31	89.04	88.78	85.54	83.92	82.97	80.82
2016	78.16	71.98	68.55	67.15	69.86	73.49	72.33	71.13	73.09	75.31	75.74	77.86
2017	81.18	79.56	77.62	77.75	77.02	76.58	75.01	75.91	80.29	79.68	79.62	79.36
2018	80.18	82.09	80.36	80.82	84.88	86.21	84.63	85.28	85.4	87.76	86.68	84.24
2019	82.66	80.34	80.86	84.74	86.67	86.4	87.59	87.81	86.58	86.48	85.98	84.57
2020	84.46	82.98	80.63	79.33	77.29	75	71.84	71.18	73.16	72.48	71.83	72.33
2021	77.36	82.76	88.18	88.66	91.38	93.72	96.66	98.72	99.28	101.43	104.1	103.1
2022	102.44	105.74	116.93	113.68	119.32	130.51	133.63	122	111.44	112.99	114.09	106.03
2023	100.67	102.84	98.56									

Nota: Tomado del *Banco Central de Reserva del Perú – Gerencia Central de Estudios Económicos*.

Ejemplo 2

- **Librería e importación de datos**

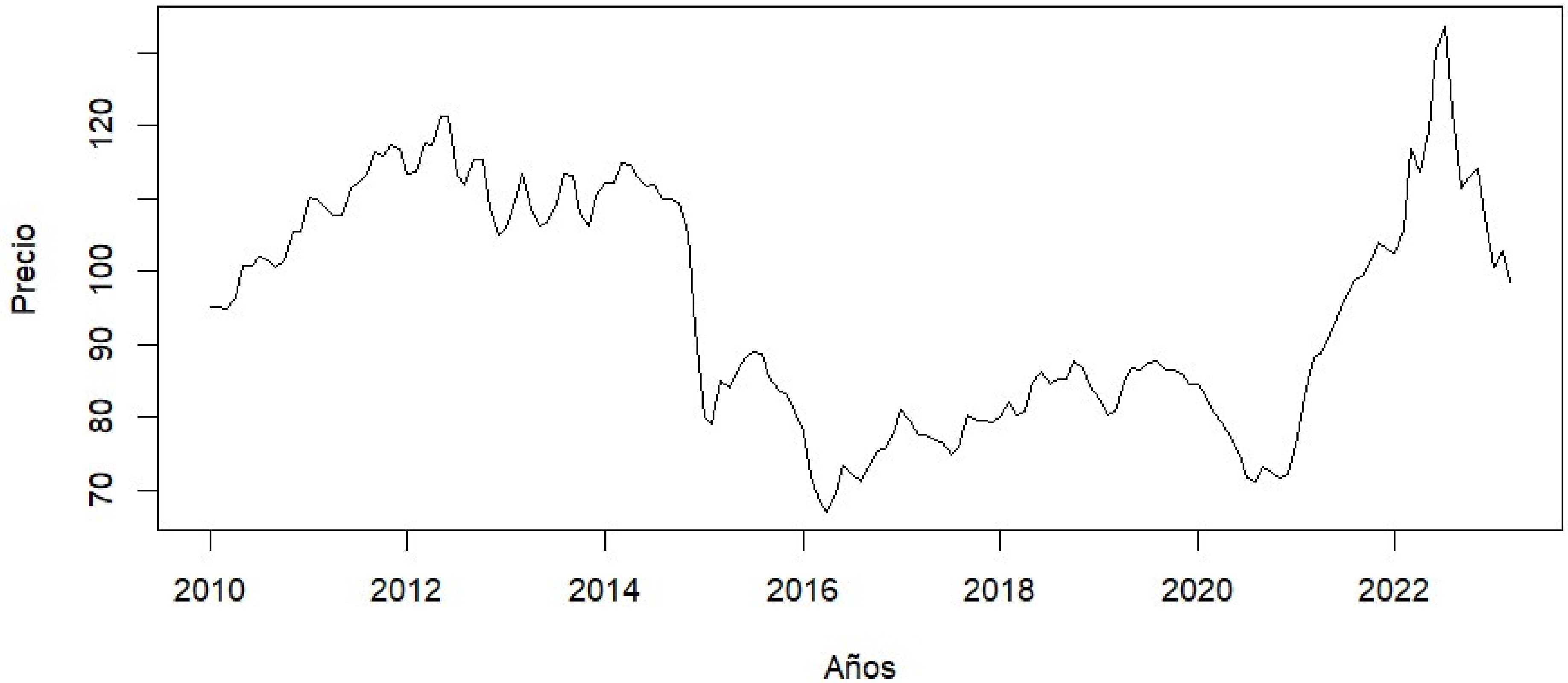
```
# Libreria necesaria
library(forecast) # Modelo ARIMA
library(tseries)  # Para series de tiempo
library(TSA)      # Para series de tiempo
library(urca)     # Para hacer el Test de Raiz Unitaria (detectar si hay o no estacionariedad)
library(ggplot2)  # Para hacer gráficos
library(ggfortify)
library(dplyr)    # Para la manipulación de datos (filtrar, seleccionar, agregar, transformar)

#Importar datos
library(readxl)
Combustible <- read_excel("D:/.../Ejm_6_1-Combustible.xlsx")
View(Combustible)
```

Gráfica de la serie

```
Yt <- ts(Combustible$precio, start = 2010, frequency = 12)
plot(Yt, xlab="Años", ylab="Precio")
```

Ejemplo 2

Precios de Combustibles - Gasohol 90 Octanos (Yt), periodo 2010 – 2023.



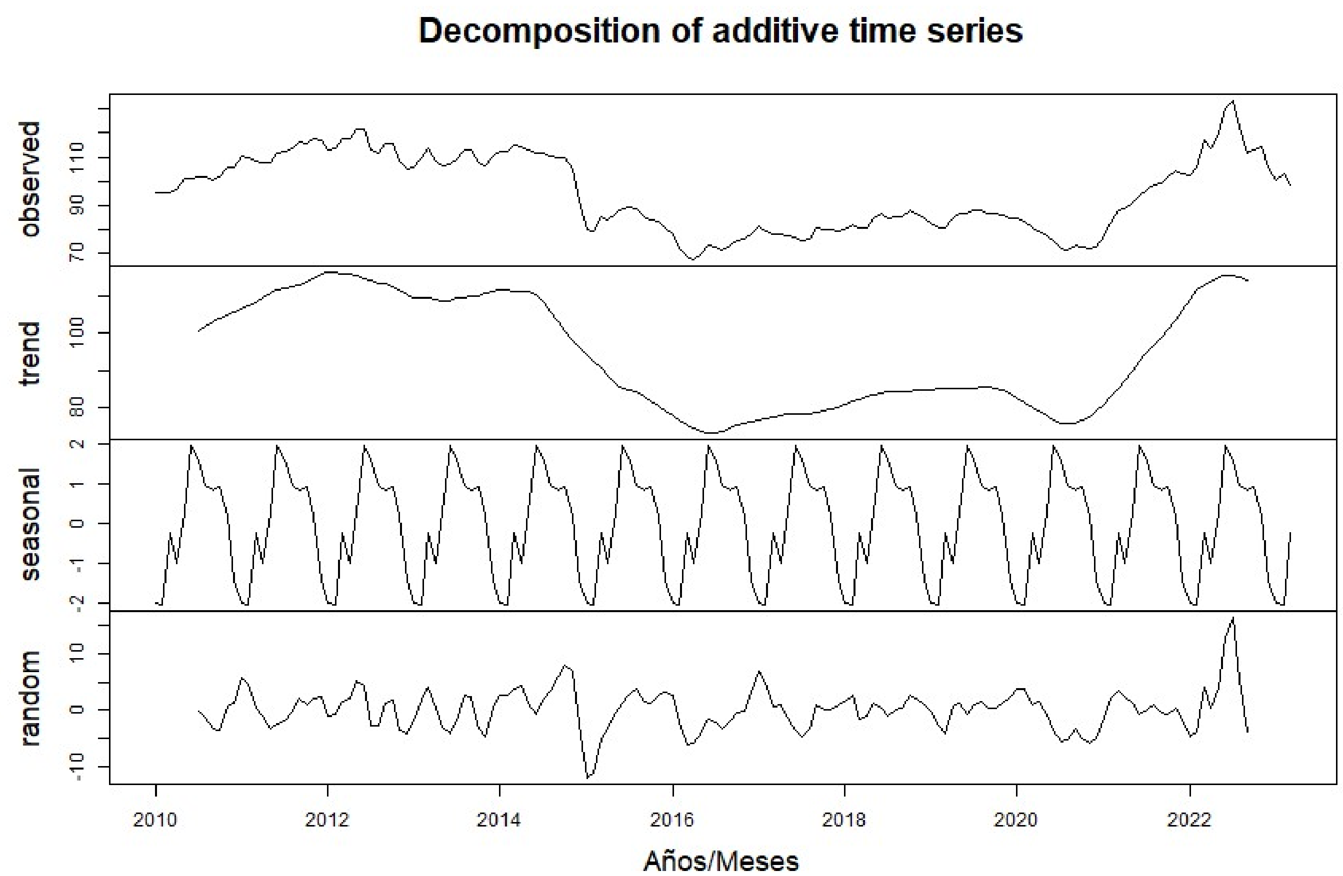
Ahora realizaremos el proceso detallado de la construcción de modelos ARIMA apropiados para la serie Yt basados en la metodología de Box-Jenkins.

Ejemplo 2

1) IDENTIFICACIÓN

Descomposición de la Serie.

```
Yt_desc <- decompose(Yt, type="additive")
plot(Yt_desc, xlab="Años/Meses")
```



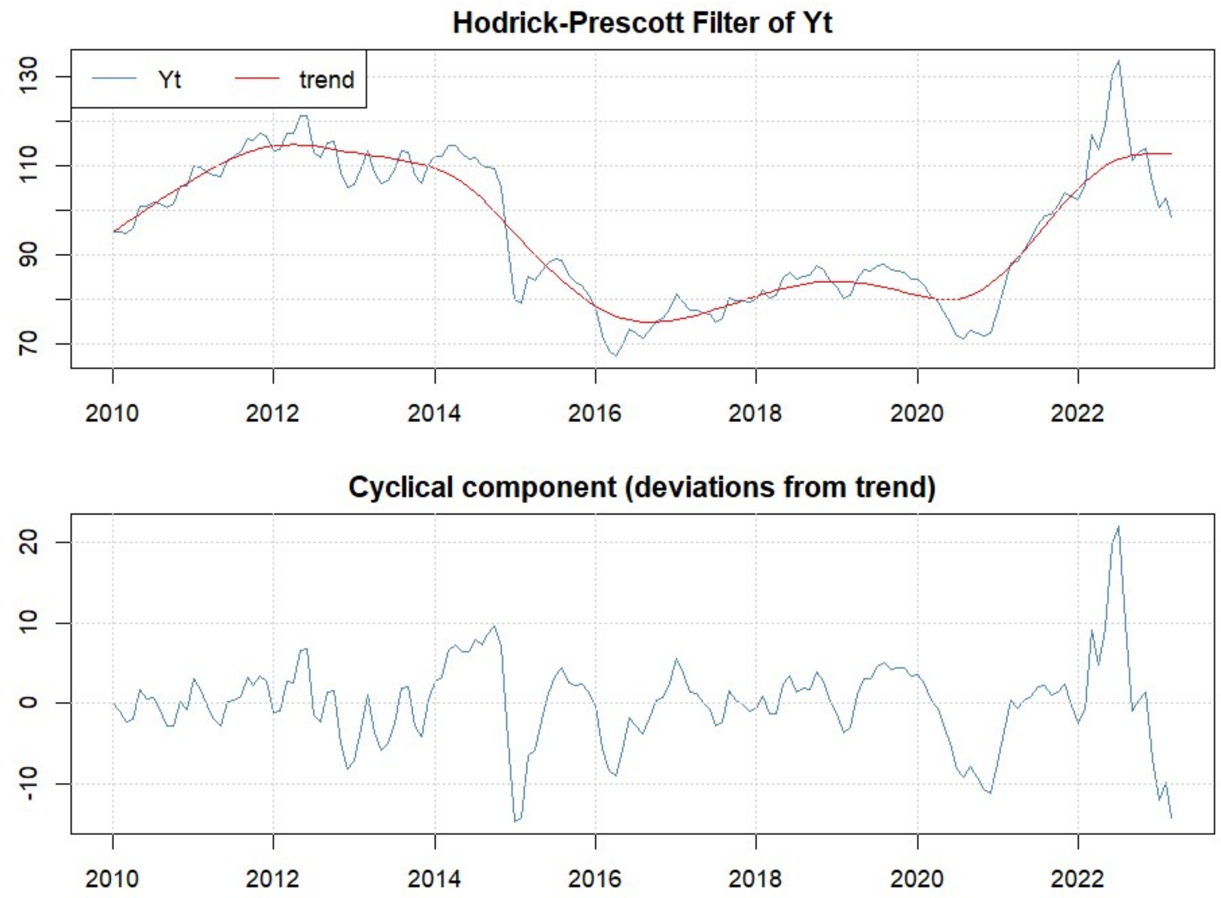
Ahora se identificará de manera individual la estacionalidad y la tendencia para indicar que la serie es no estacionaria en media y varianza.

Ejemplo 2

- Análisis de la tendencia

Aplicando Filtro de Hodrick – Prescott

```
library(mFilter)
lambda_hp <- 1600
Yhp <- hpfilter(Yt, type="lambda", freq=lambda_hp)
plot(Yhp)
```



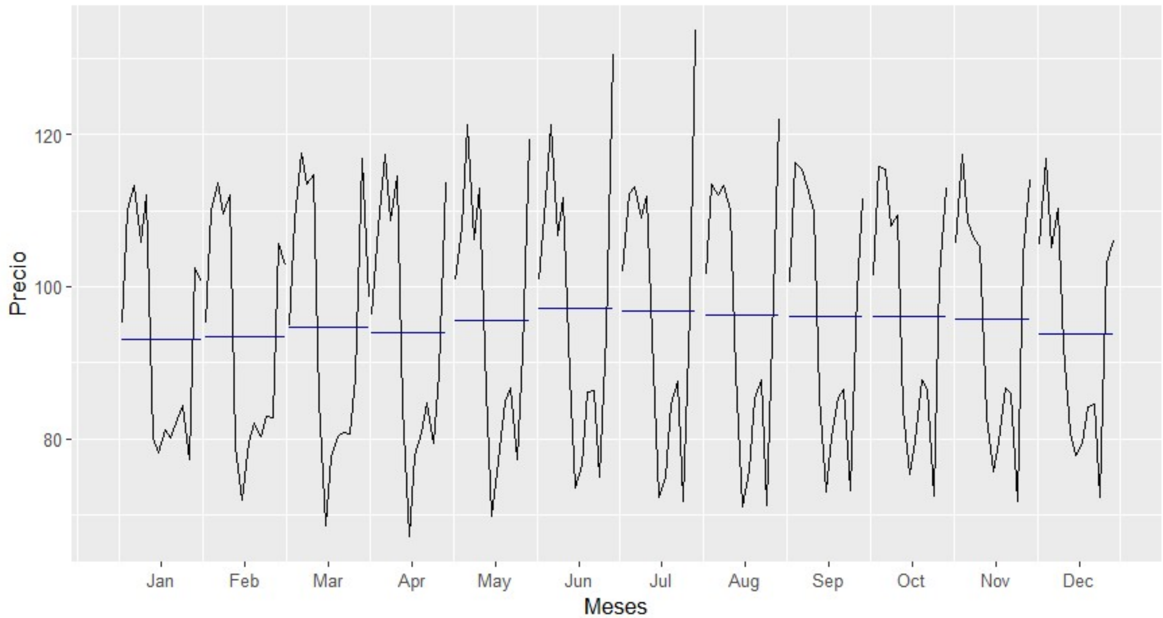
Muestra la evolución temporal de la tendencia, la cual tiene un comportamiento cíclico de conformidad con su evolución marcada por las fluctuaciones del precio de petróleo en el mercado mundial.

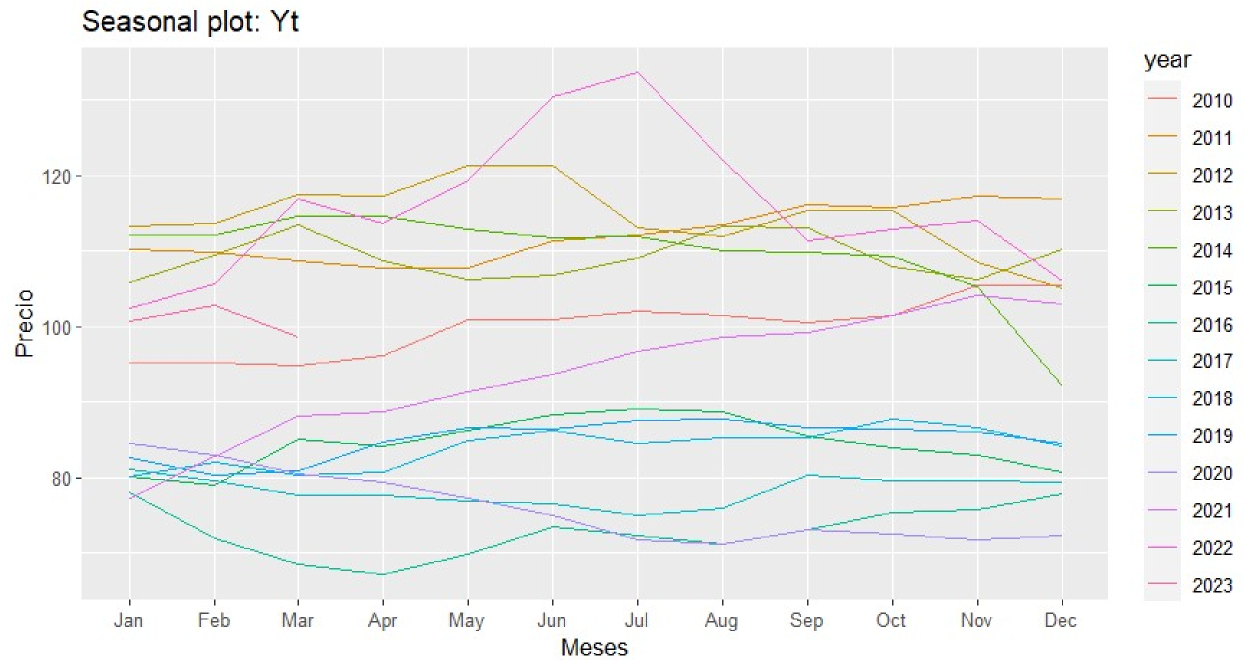
Ejemplo 2

- Análisis de la estacionalidad

✓ Utilizaremos directamente métodos gráficos como los gráficos de líneas separadas y apiladas.
✓ Se hace importante detectar este componente, debido a que, si son series estacionales se usará S-ARIMA (Seasonal).

```
ggsubseriesplot(Yt, xlab = "Meses", ylab="Precio")
ggseasonplot(Yt, xlab = "Meses", ylab="Precio")
```

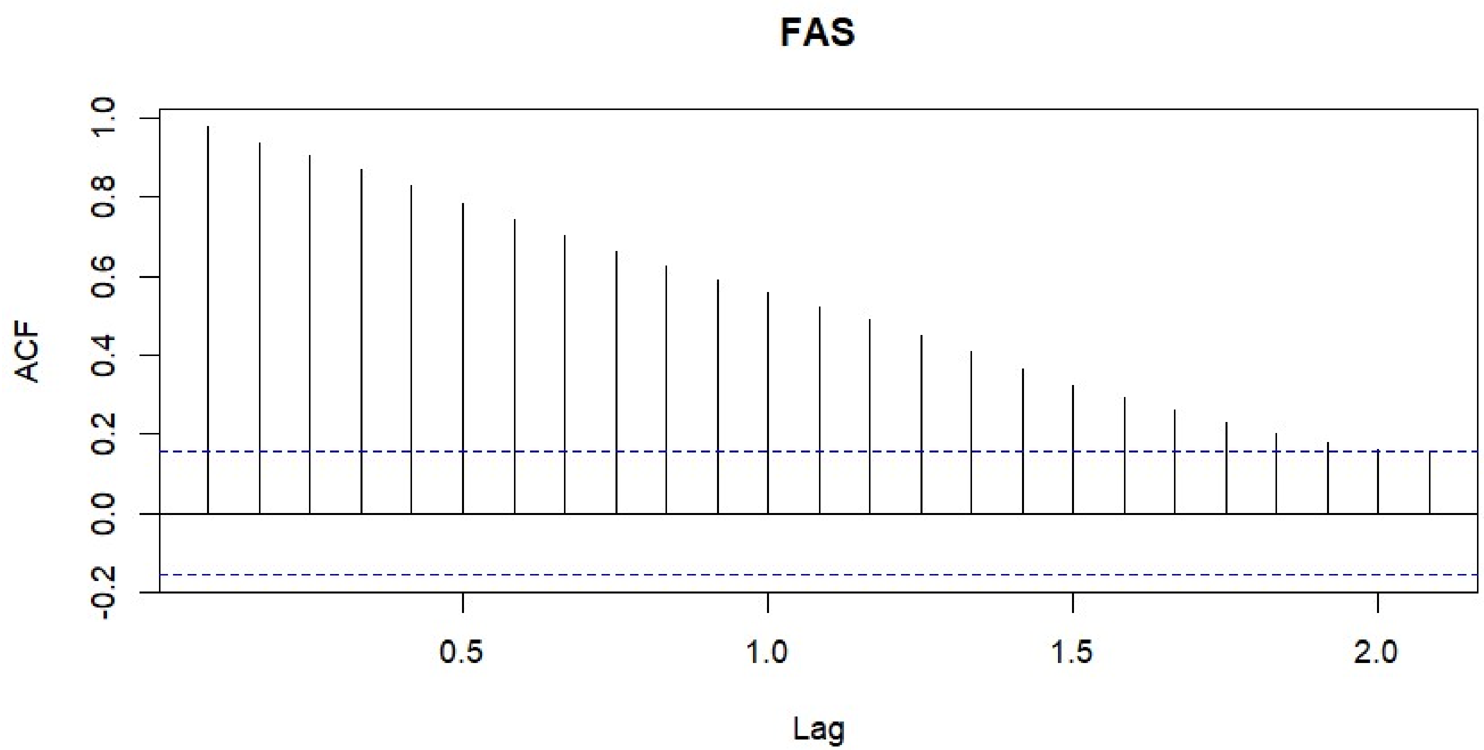




Ejemplo 2

Correlograma de los Precios de Combustibles

```
FAS <- acf(Yt, lag.max = 25, main="FAS", level = 0.95)
```



Ejemplo 2

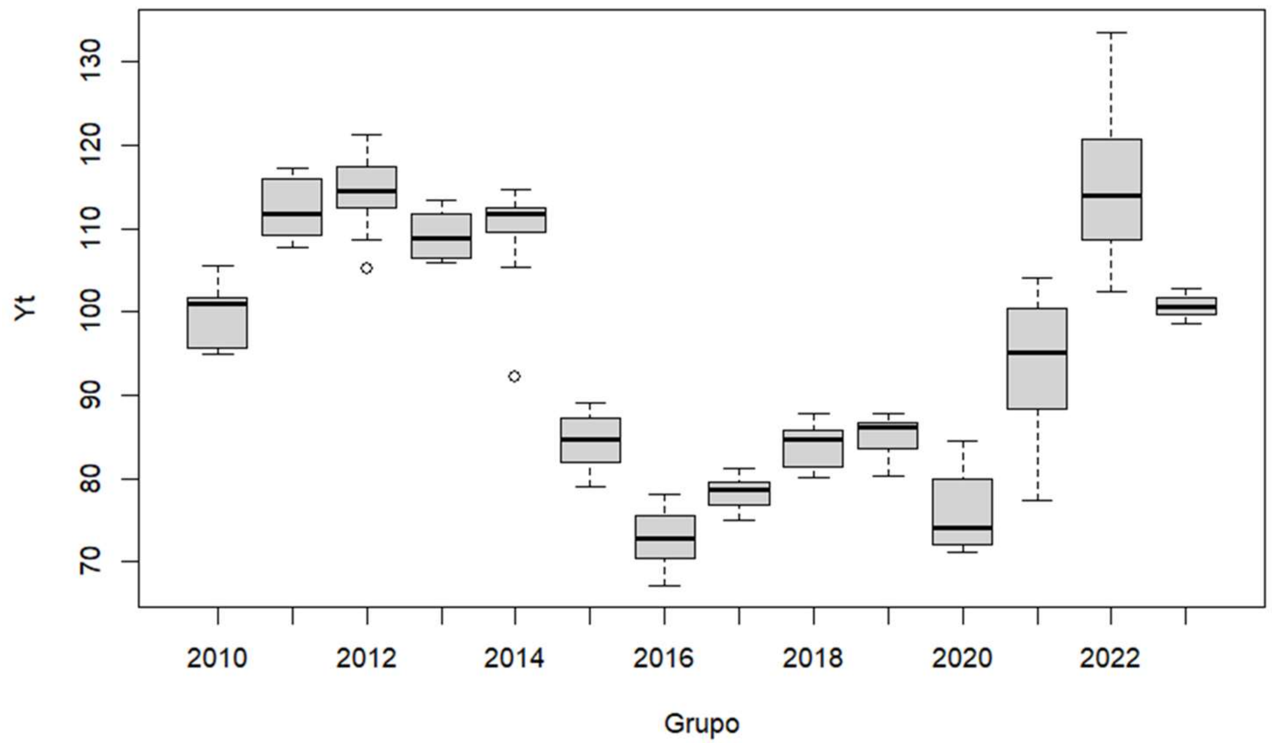
Análisis de Estacionariedad en Variaza

A continuación, realizaremos un análisis gráfico a través de diagrama de cajas por grupos, en este caso, los grupos naturales serían los años.

```
boxplot(Combustible$precio ~ Combustible$año, xlab = "Grupo", ylab="Yt")
```

Podemos ver que los grupos tienen similar variación a lo largo de la evolución de la serie hasta el año 2019, los años 2020, 2021 y 2022 muestran una mayor variación, entonces hay indicios que la serie no sea estacionaria en varianza.

Por tanto, habrá que verificar si es necesario realizar una transformación de la serie a través de la metodología de Box-Cox.



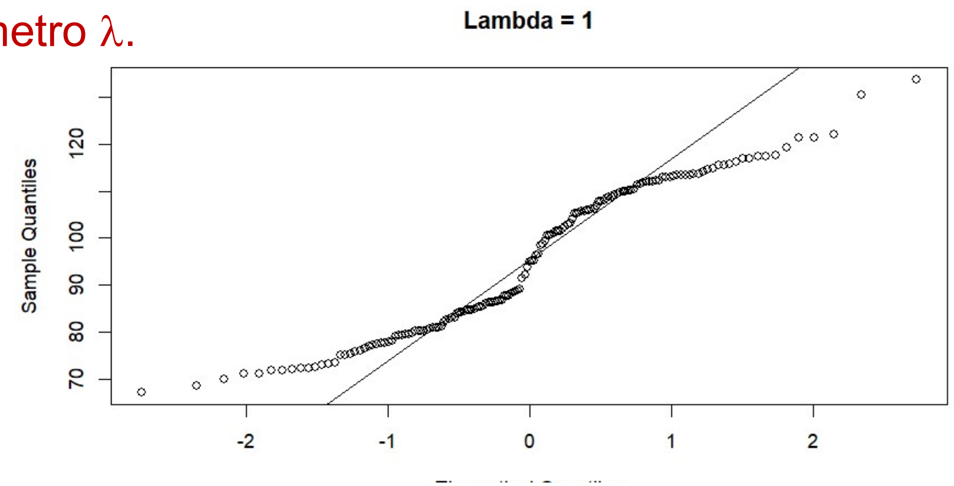
Ejemplo 2

A continuación, mostramos dos formas de **elegir el parámetro de lambda** apropiado para la transformación de la serie.

i) Transformación de la serie para elegir el valor apropiado del parámetro λ.

Para λ = 1, es la serie original Y

```
qqnorm(Yt, main = "Lambda = 1") # Yt: Original
qqline(Yt)
```

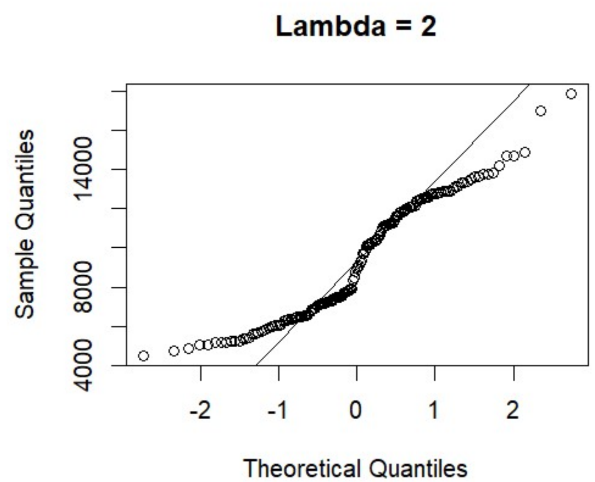


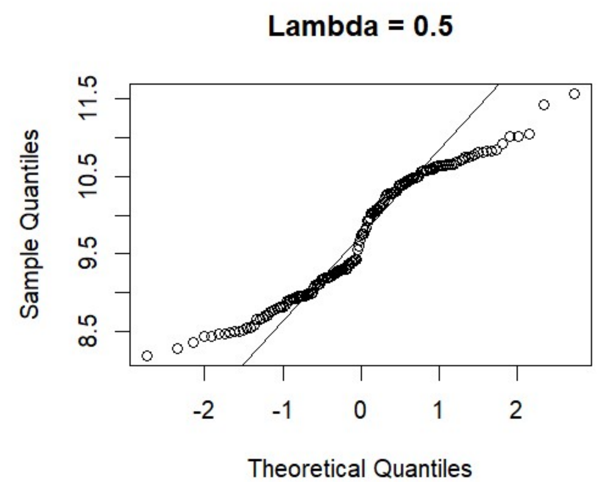
Para λ = 2

```
T1.Yt <- sqrt(Yt)
qqnorm(T1.Yt, main = "Lambda = 2")
qqline(T1.Yt
```

Para λ = 0.5

```
T1.Yt <- sqrt(Yt)
qqnorm(T1.Yt, main = "Lambda = 0.5")
qqline(T1.Yt
```





Ejemplo 2

Para $\lambda = 0$

```
T4.Yt <- log(Yt)
qqnorm(T4.Yt, main = "Lambda = 0")
qqline(T4.Yt)
```

Para $\lambda = -0.5$

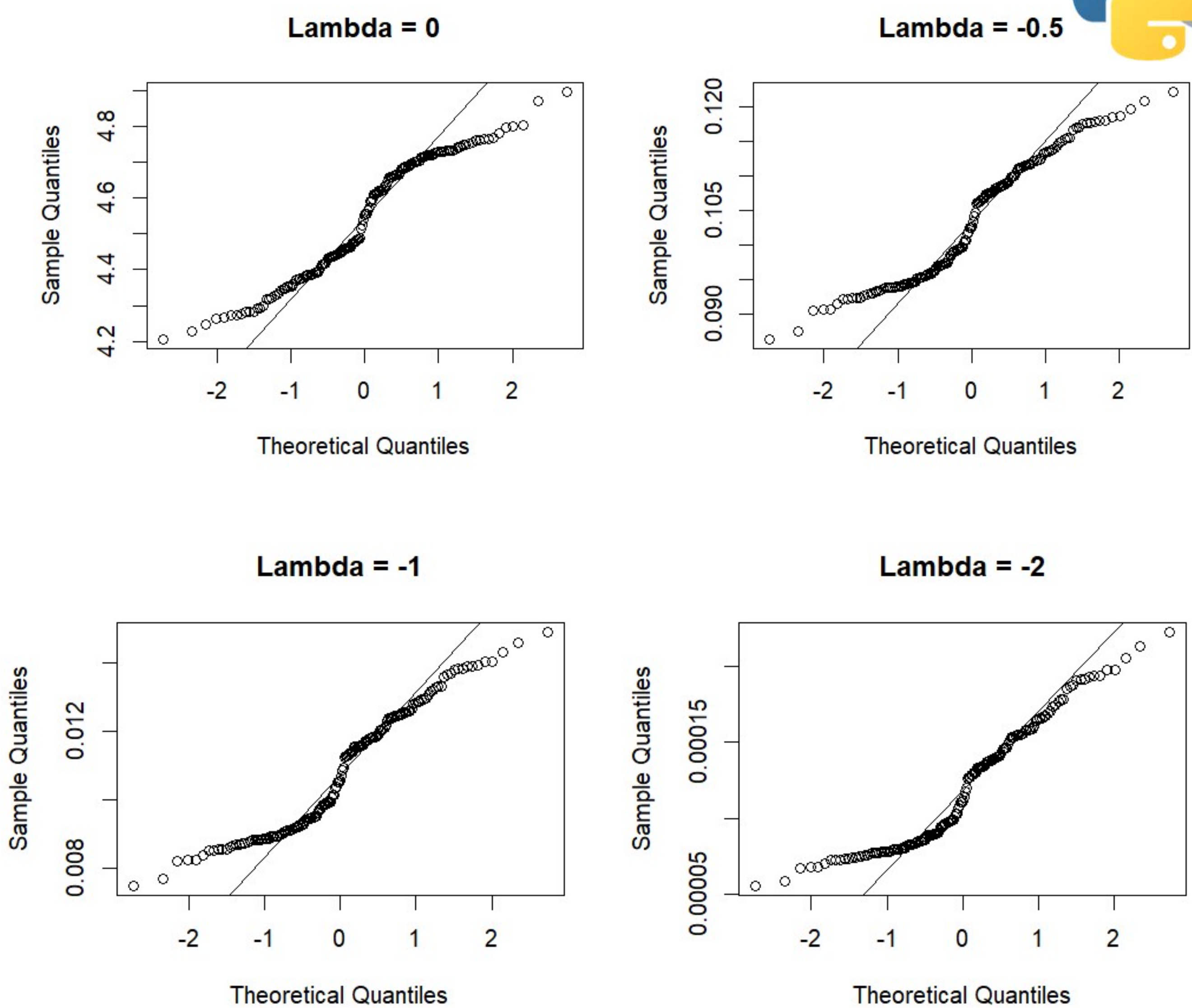
```
T5.Yt <- 1/sqrt(Yt)
qqnorm(T5.Yt, main = "Lambda = -0.5")
qqline(T5.Yt)
```

Para $\lambda = -1$

```
T6.Yt <- 1/Yt
qqnorm(T6.Yt, main = "Lambda = -1")
qqline(T6.Yt)
```

Para $\lambda = -2$

```
T7.Yt <- 1/(Yt^2)
qqnorm(T7.Yt, main = "Lambda = -2")
qqline(T7.Yt)
```



Se observa el ajuste a la distribución normal de la serie original Yt y en las demás figuras se tiene todas las demás transformaciones, las cuales no son sensibles al valor que vaya a tomar lambda (λ)

Ejemplo 2

ii) Aproximación directa del valor de lambda optimo para realizar la transformación.

En R, podemos hacer uso de la función **boxcox** de la librería **MASS** para estimar el parámetro de transformación λ.

```
b=BoxCox.ar(Yt)
```

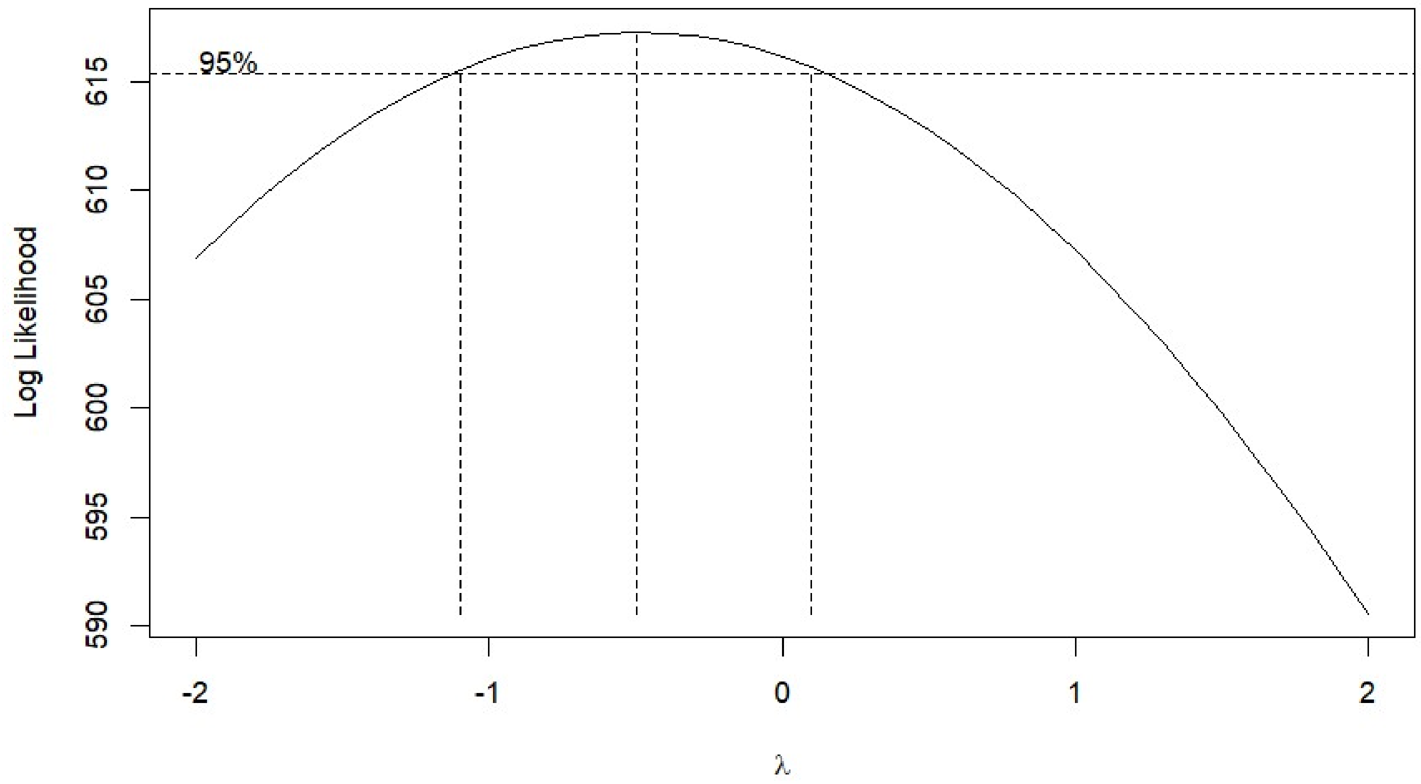
Valor estimado de λ.

```
lambda <- b$mle
round(lambda,1)
```

```
[1] -0.5
```

Por lo tanto, debemos hacer la siguiente transformación para buscar conseguir una aproximación a la normalidad.

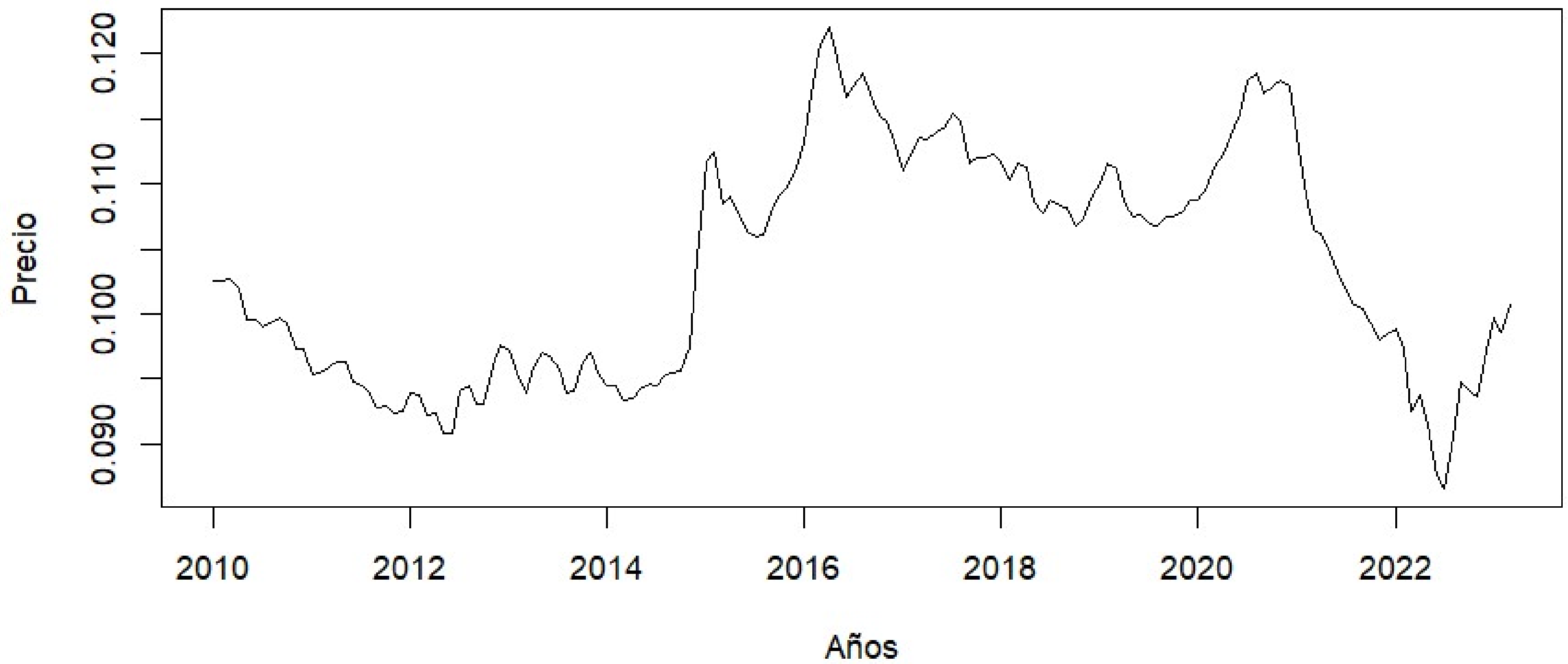
$$Y_t^{(\lambda)} = \frac{Z_t^{\lambda} - 1}{\lambda} = \frac{1}{\sqrt{Y_t}}, \quad \text{para } \lambda = -0.5$$



Ejemplo 2

Transformación de la serie Yt

```
Tr.Yt <- 1/sqrt(Yt)
plot(Tr.Yt, xlab="Años", ylab="Precio")
```



La serie transformada de la figura es muy similar a la serie original, con lo cual ratificamos que la serie no es sensible a las transformaciones, por consiguiente, en este ejemplo tomaremos la decisión de no realizar la transformación de la serie y asumiremos que es estacionaria en varianza.

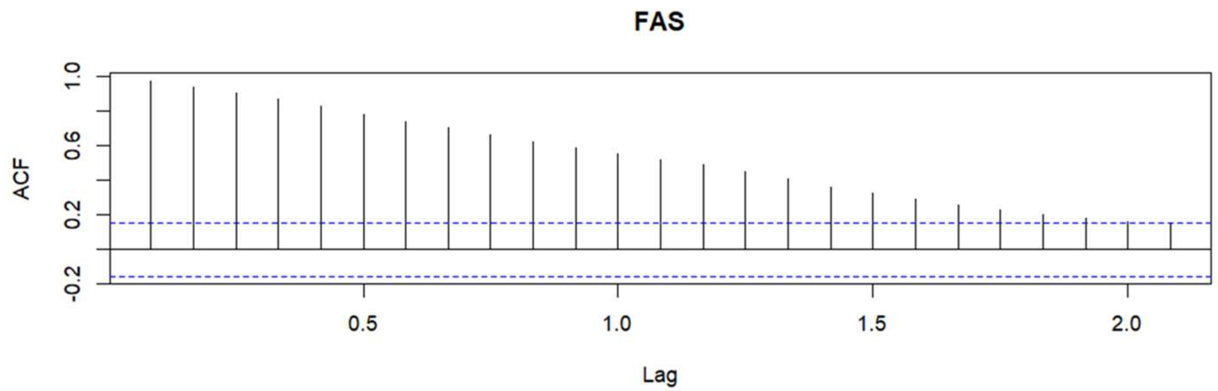
Ejemplo 2

Análisis de Estacionariedad en Media

Correlogramas del FAS y FAP

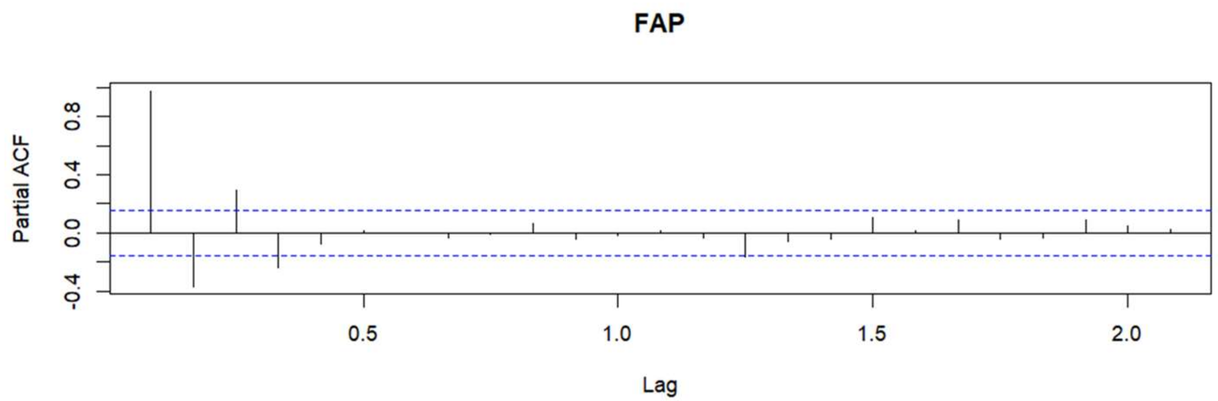
```
FAS <- acf(Yt, lag.max = 25, main="FAS", level = 0.95)
FAP <- pacf(Yt, lag.max = 25, main="FAP", level = 0.95)
```

Decrece lentamente



+

Primer FAP > 0.9



Serie es No estacionaria

Ejemplo 2

Verificación con la prueba de Raíz unitaria de Dickey-Fuller Aumentada.

```
Yt_adf <- ur.df(Yt, type="drift", lags = 1)
summary(Yt_adf)

###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################
Test regression drift
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.233 on 154 degrees of freedom
Multiple R-squared:  0.1422,	Adjusted R-squared:  0.1311
F-statistic: 12.76 on 2 and 154 DF,  p-value: 7.427e-06

Value of test-statistic is: -1.9782 1.9568

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81
```

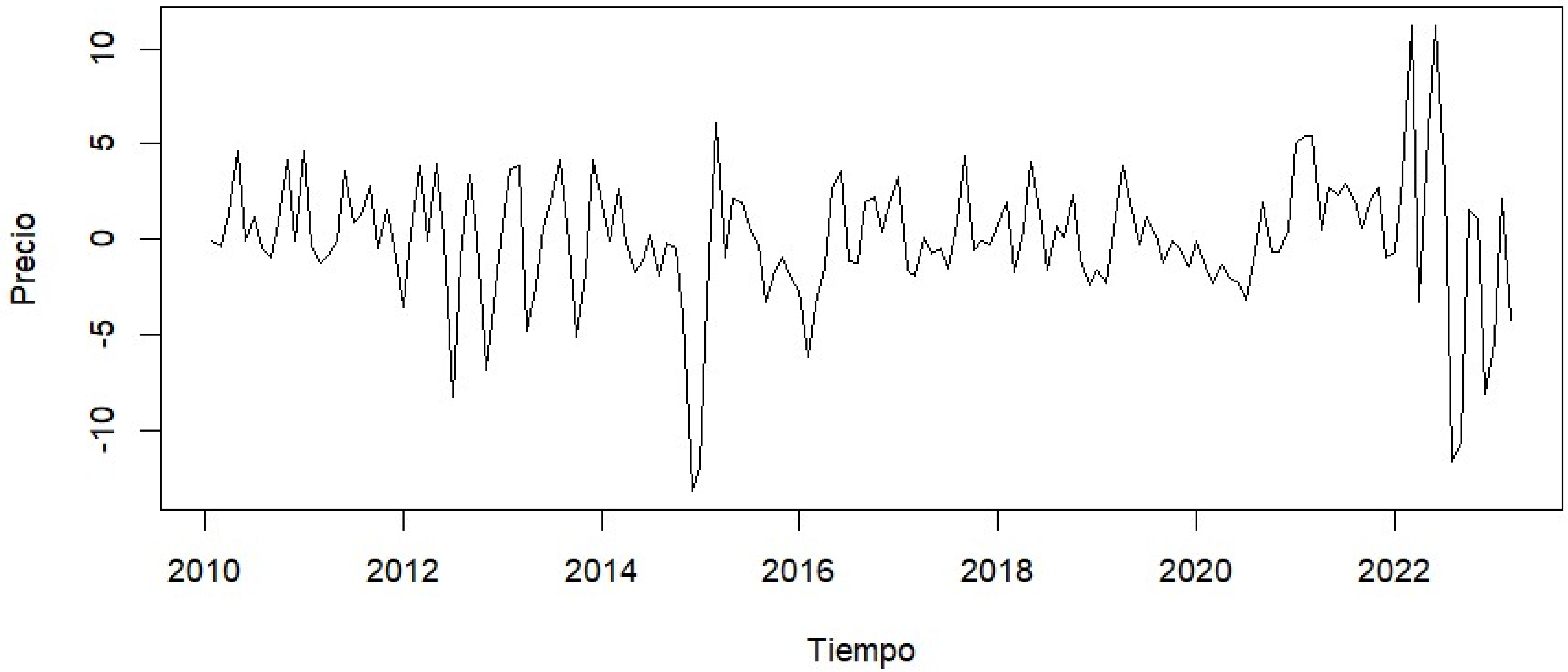
De los resultados, el valor de $\tau_{calculado} = -1.9782 < \tau_{crítico} = -2.88$, por tanto, se acepta la hipótesis nula (Ho) de la existencia de raíz unitaria, es decir, que la serie Yt es NO ESTACIONARIA.

Por consiguiente, se debe diferenciar la serie Yt.

Ejemplo 2

Primera diferencia de la serie de los Precios de Combustibles (**D1.Yt**)

```
D1.Yt <- diff(Yt)
plot(D1.Yt, xlab="Tiempo", ylab="Precio")
```

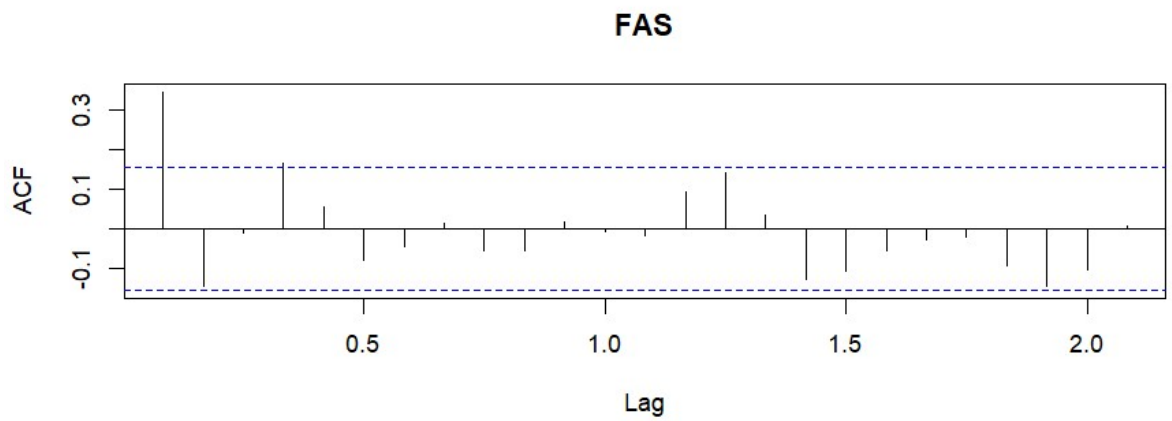


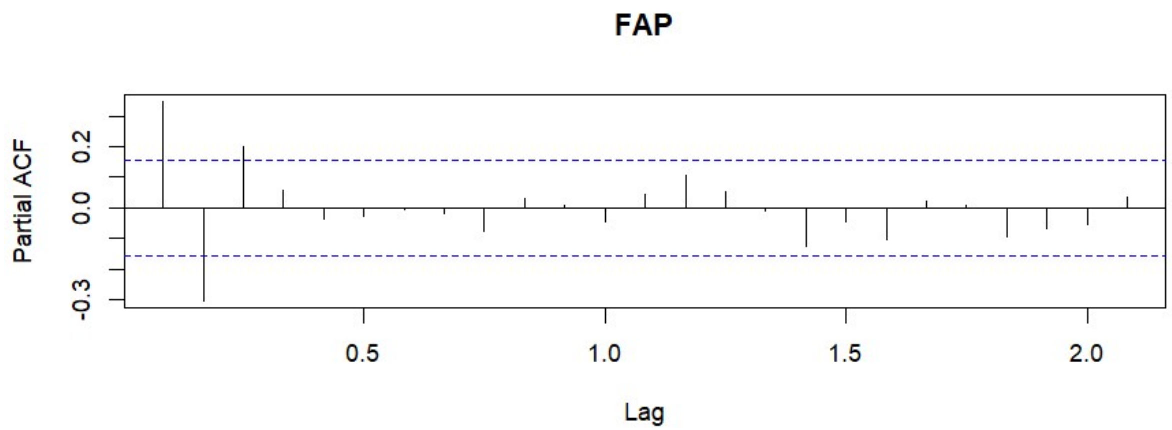
Ejemplo 2

Correlogramas FAS y FAP de la serie diferenciada.

```
FASd <- acf(D1.Yt, lag.max = 25, main="FAS", level = 0.95)
FAPd <- pacf(D1.Yt, lag.max = 25, main="FAP", level = 0.95)
FASd$acf[1]]
```

```
[1] 0.3452626
```





Ejemplo 2

Verificamos con la prueba de la raíz unitaria de Dickey-Fuller Aumentada la serie **D1.Yt**.

```
D1.Yt_adf <- ur.df(D1.Yt, type="drift", lags = 1)
summary(D1.Yt_adf)
```

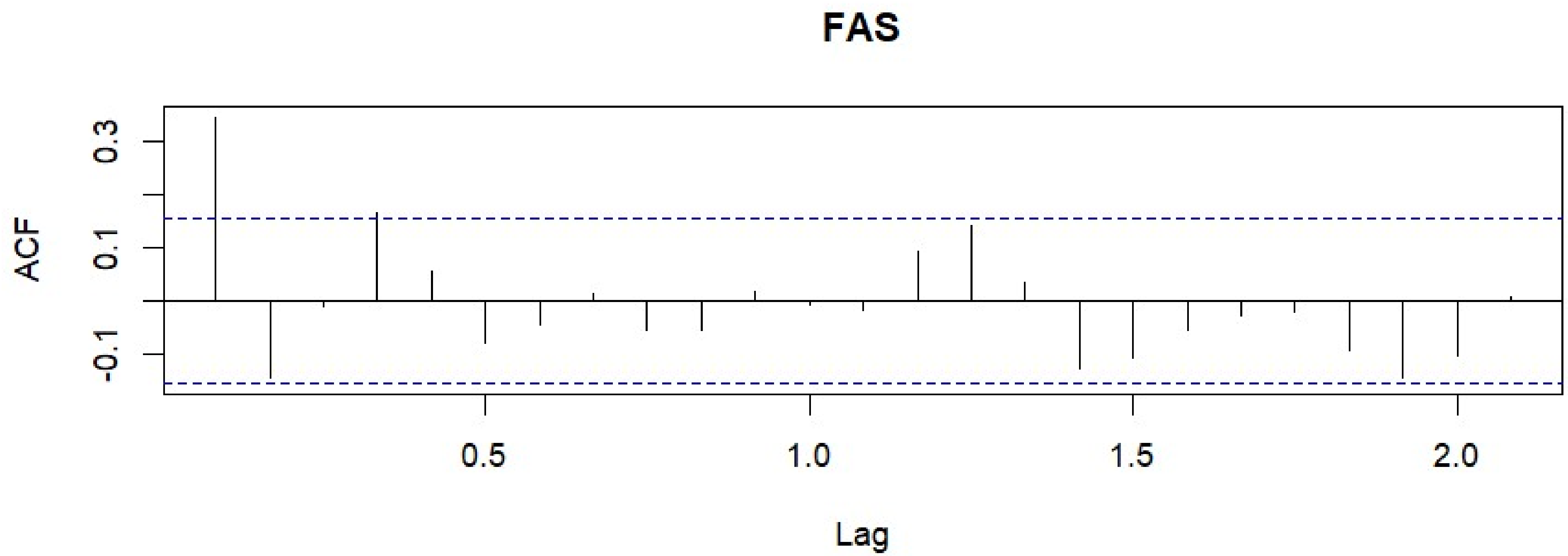
```
###############################################
# Augmented Dickey-Fuller Test Unit Root Test #
###############################################
Test regression drift

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

---
Residual standard error: 3.124 on 153 degrees of freedom
Multiple R-squared:  0.3874,      Adjusted R-squared:  0.3794
F-statistic: 48.39 on 2 and 153 DF,  p-value: < 2.2e-16


Value of test-statistic is: -9.6678 46.738

Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
phi1  6.52  4.63  3.81
```

De los resultados, el valor de $\tau_{calculado}$ = -9.6678 < $\tau_{crítico}$ = -2.88, por tanto, se rechaza la hipótesis nula (Ho), es decir, que la serie D1.Yt ahora si es ESTACIONARIA.

Ejemplo 2

Identificación del orden del modelo

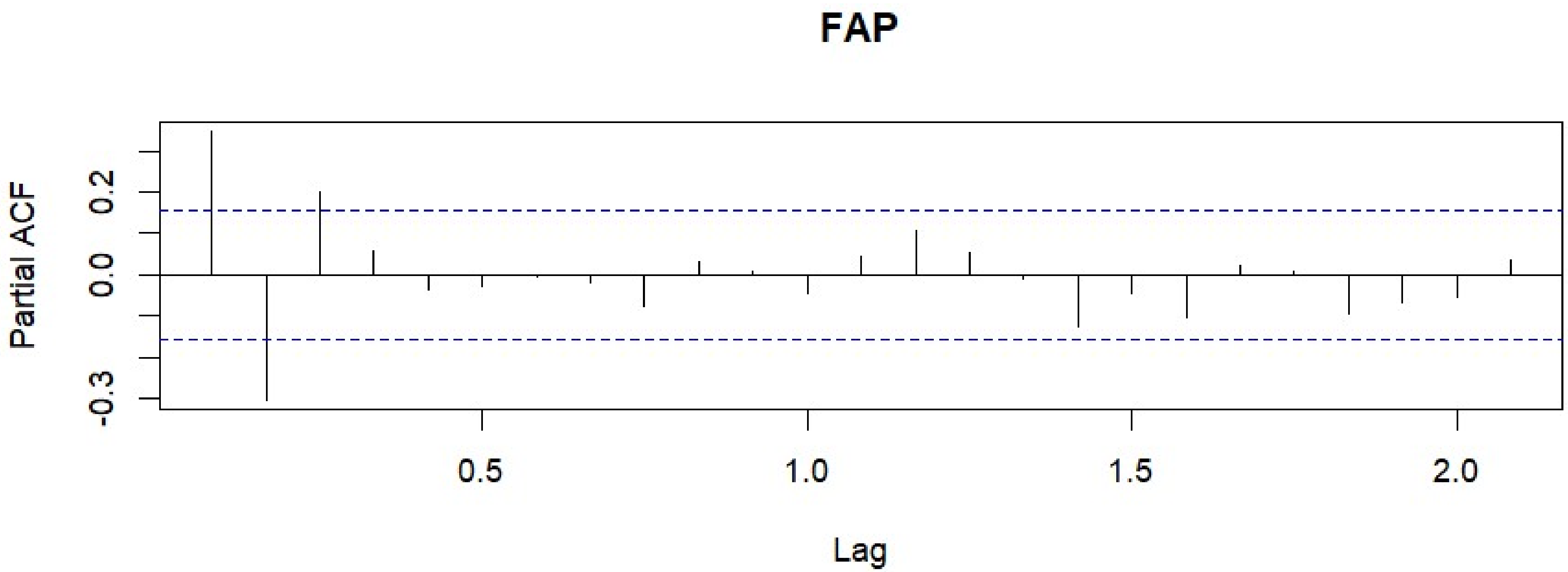
FAS para parte de Media Móvil (MA)



- La FAS estimada muestra una estructura infinita decreciendo en forma de onda sinusoidal, el primer y segundo coeficiente tienen signos opuestos siendo este comportamiento no definido para la parte de media móvil, por tanto, sugiere un modelo **MA(0)** y podríamos proponer un **MA(1)**.

Ejemplo 2

FAP para parte Autorregresiva (AR)



- Los tres primeros coeficientes son significativamente distintos de cero y el resto no. Según esta interpretación, la FAP estaría truncada en el retardo 3 y el modelo adecuado para haber generado la serie sería un **AR(3)**.

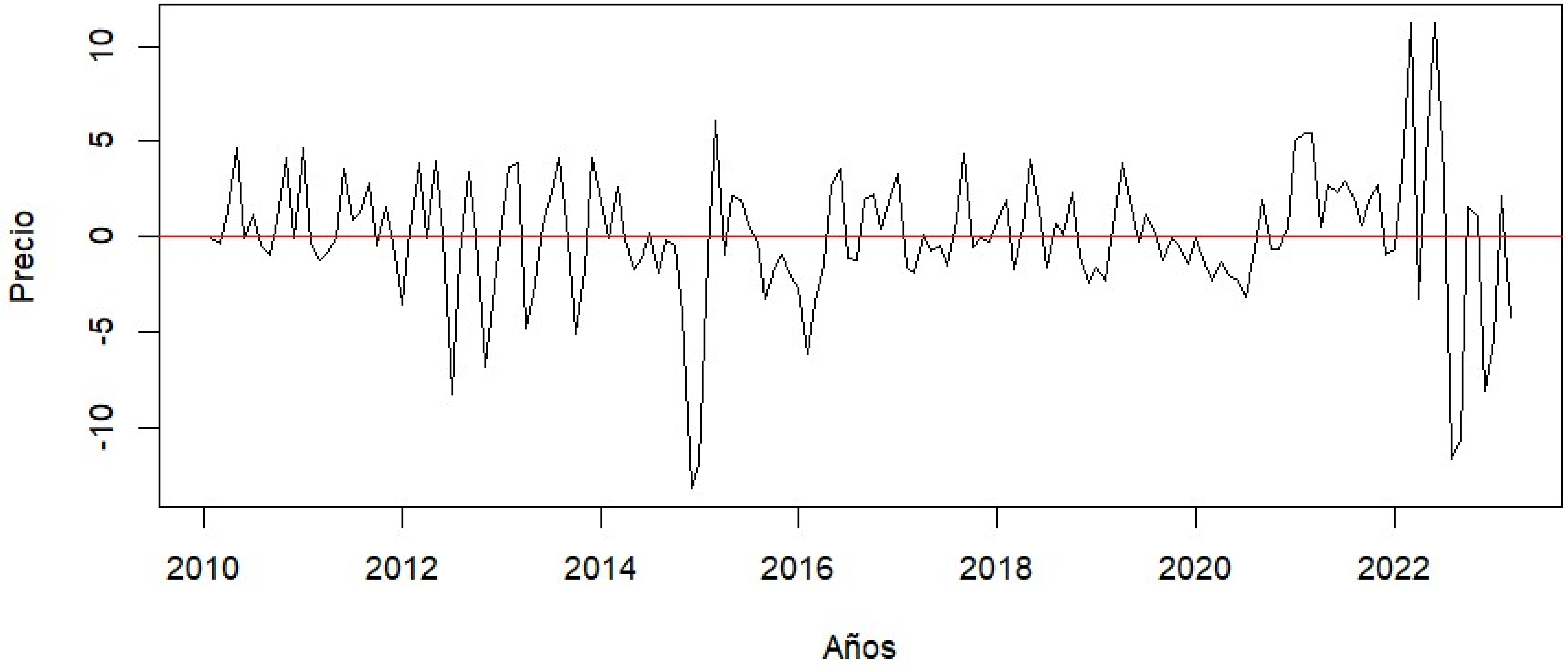
- Modelos propuestos
    - Modelo 1: ARIMA(3, 1, 0)
    - Modelo 2: ARIMA(3, 1, 1)

Ejemplo 2

Inclusión del intercepto

Determinamos la inclusión del término constante, haremos un análisis gráfico para ver entorno a que valor oscilan las variaciones.

```
plot(D1.Yt, xlab="Años", ylab="Precio")
abline(h = mean(D1.Yt), col = "red")
```



Ejemplo 2

Prueba estadística.

```
Z <- mean(D1.Tr.Yt)
Co <- var(D1.Tr.Yt)
Tn <- length(Yt)
Ta <- Tn - 1
Sigma <- Co/Ta
t <- Z/Sigma
tt <- qt(1-0.05/2,Ta-1)
pruebaT <- c(t, tt)
names(pruebaT) <- c("t-calculado","t-critico")
pruebaT
```

```
t-calculado   t-critico
  0.2778186   1.9751892
```

Se acepta la hipótesis nula H_0: $E(Zt) = 0$ para un α del 5% y no se incluye un término constante en el modelo.

Ejemplo 2

Resumiendo, se proponen los dos modelos siguientes para la serie estacionaria Zt = ∆Yt, en caso que fuesen:

- ARMA(3, 0) : $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + a_t$
- ARMA(3, 1): $Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \phi_3 Z_{t-3} + a_t + \theta_1 a_{t-1}$

también:

- ARIMA(3, 1, 0) : $\Delta Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \phi_3 \Delta Y_{t-3} + a_t$
- ARIMA(3, 1, 1): $\Delta Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \phi_3 \Delta Y_{t-3} + a_t - \theta_1 a_{t-1}$

Ejemplo 2

2) ESTIMACIÓN

Estimación de los modelos propuestos:

- Modelo 1: ARIMA(3, 1, 0): $\Delta Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \phi_3 \Delta Y_{t-3} + a_t$

```
mod1 <- Arima(Yt, order = c(3,1,0),include.constant = F)
summmary(mod1)
```

```
Series: Yt
ARIMA(3,1,0)

Coefficients:
         ar1      ar2     ar3
      0.5314  -0.4161  0.2272
s.e.  0.0787   0.0830  0.0790

sigma^2 = 9.15:  log likelihood = -397.81
AIC=803.62   AICc=803.88   BIC=815.87

Training set error measures:
                       ME     RMSE      MAE         MPE    MAPE      MASE         ACF1
Training set -0.005364241 2.986552 2.159041 -0.02703663 2.24548 0.2004655 -0.009397707
```

Parámetros estimados del modelo:

ARIMA(3, 1, 0):

$$\Delta Y_t = 0.5314 \Delta Y_{t-1} - 0.4161 \Delta Y_{t-2} + 0.2272 \Delta Y_{t-3} + a_t$$

Ejemplo 2

- Modelo 2: ARIMA(3, 1, 1): $\Delta Y_t = \phi_1 \Delta Y_{t-1} + \phi_2 \Delta Y_{t-2} + \phi_3 \Delta Y_{t-3} + a_t - \theta_1 a_{t-1}$

```
mod2 <- Arima(Yt, order = c(3,1,1),include.constant = F)
coeftest(mod2)
```

```
Series: Yt
ARIMA(3,1,1)

Coefficients:
         ar1      ar2     ar3      ma1
      0.6918  -0.4902  0.2784  -0.1681
s.e.  0.2714   0.1447  0.1059   0.2774

sigma^2 = 9.188:  log likelihood = -397.63
AIC=805.27   AICc=805.66   BIC=820.58

Training set error measures:
                       ME     RMSE      MAE         MPE     MAPE      MASE         ACF1
Training set -0.009519068 2.983173 2.154947 -0.02743029 2.241676 0.2000854 0.0005440147
```

Parámetros estimados del modelo:

ARIMA(3, 1, 1): $\Delta Y_t = 0.6918 \Delta Y_{t-1} - 0.4902 \Delta Y_{t-2} + 0.2784 \Delta Y_{t-3} + a_t - 0.1681 a_{t-1}$

Ejemplo 2

3) VALIDACIÓN

Significancia de los coeficientes

- Modelo 1: ARIMA(3, 1, 0):

```
coeftest(mod1)

z test of coefficients:

     Estimate Std. Error z value  Pr(>|z|)
ar1  0.531439   0.078670  6.7553 1.425e-11 ***
ar2 -0.416141   0.083049 -5.0108 5.420e-07 ***
ar3  0.227151   0.079050  2.8735  0.004059 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- AR(1) : $\phi_1 = 0.53144$ → p = 0.00000 < 0.01, altamente significativo.
- AR(2) : $\phi_2 = -0.41614$ → p = 0.00000 < 0.01, altamente significativo.
- AR(3) : $\phi_3 = 0.22715$ → p = 0.00372 < 0.01, altamente significativo.

Ejemplo 2

- Modelo 2: ARIMA(3, 1, 1):

```
coeftest(mod2)

z test of coefficients:

    Estimate Std. Error z value  Pr(>|z|)
ar1  0.69184    0.27137  2.5495 0.0107890 *
ar2 -0.49016    0.14472 -3.3869 0.0007069 ***
ar3  0.27844    0.10588  2.6298 0.0085430 **
ma1 -0.16811    0.27735 -0.6061 0.5444225
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- AR(1) : $\phi_1 = 0.69184$ $\rightarrow$ p = 0.00492 < 0.01, altamente significativo.
- AR(2) : $\phi_2 = -0.49016$ $\rightarrow$ p = 0.00027 < 0.01, altamente significativo.
- AR(3) : $\phi_3 = 0.27844$ $\rightarrow$ p = 0.00383 < 0.01, altamente significativo.
- MA(1) : $\theta_1 = -0.16811$ $\rightarrow$ p = 0.44316 > 0.05, no significativo.

Por tanto, se recomienda retirar la parte de media móvil y tomar únicamente la parte autorregresiva.

Ejemplo 2

Examen de la matriz de correlaciones entre los coeficientes

La matriz de correlaciones de los coeficientes para los dos modelos estimado es el siguiente:

- Modelo 1: ARIMA(3, 1, 0)

```
vcov(mod1)
 ar1          ar2          ar3
ar1  0.006188929 -0.002955558  0.002076041
ar2 -0.002955558  0.006897091 -0.002994448
ar3  0.002076041 -0.002994448  0.006248874
```

- Modelo 2: ARIMA(3, 1, 1)

```
vcov(mod2)
 ar1         ar2         ar3         ma1
ar1  0.07364111 -0.03379478  0.02067023 -0.07212521
ar2 -0.03379478  0.02094450 -0.01201836  0.03208593
ar3  0.02067023 -0.01201836  0.01120999 -0.01980303
ma1 -0.07212521  0.03208593 -0.01980303  0.07692453
```

Se observa claramente que ningún coeficiente esta próximo ni cercano a 0.9, por tanto, podemos indicar que no hay problema de multicolinealidad en los modelos propuestos.

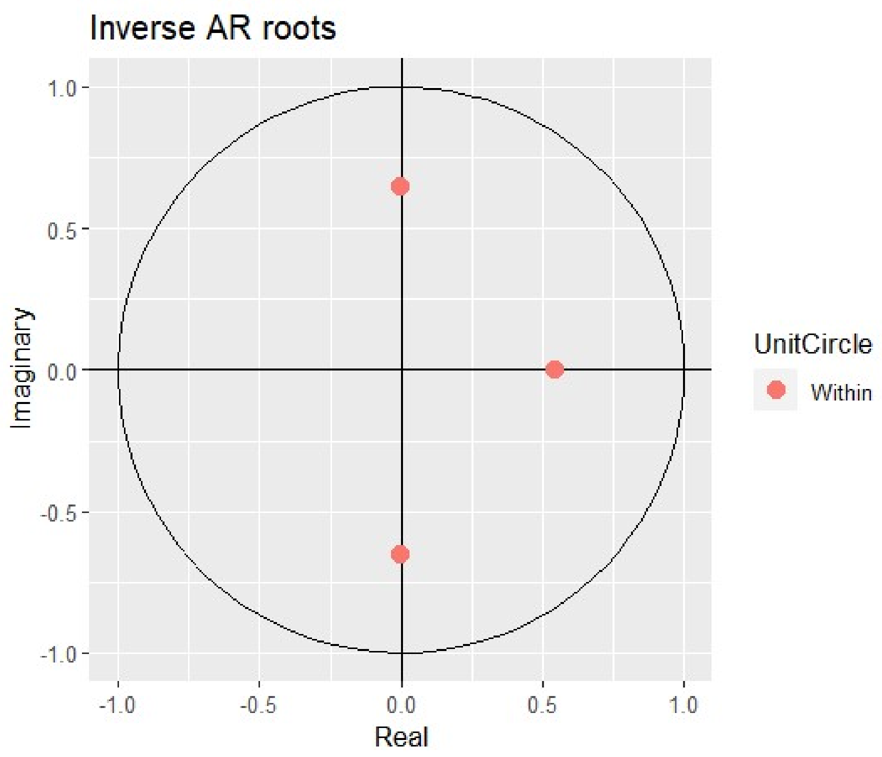
Ejemplo 2

Condición de Convergencia e Invertibilidad

Es importante indicar que, para graficar las raíces inversas es necesario estimar los parámetros del modelo haciendo uso de la función **Arima()**.

- Modelo 1: ARIMA(3, 1, 0)

```
autoplot(mod1)
```

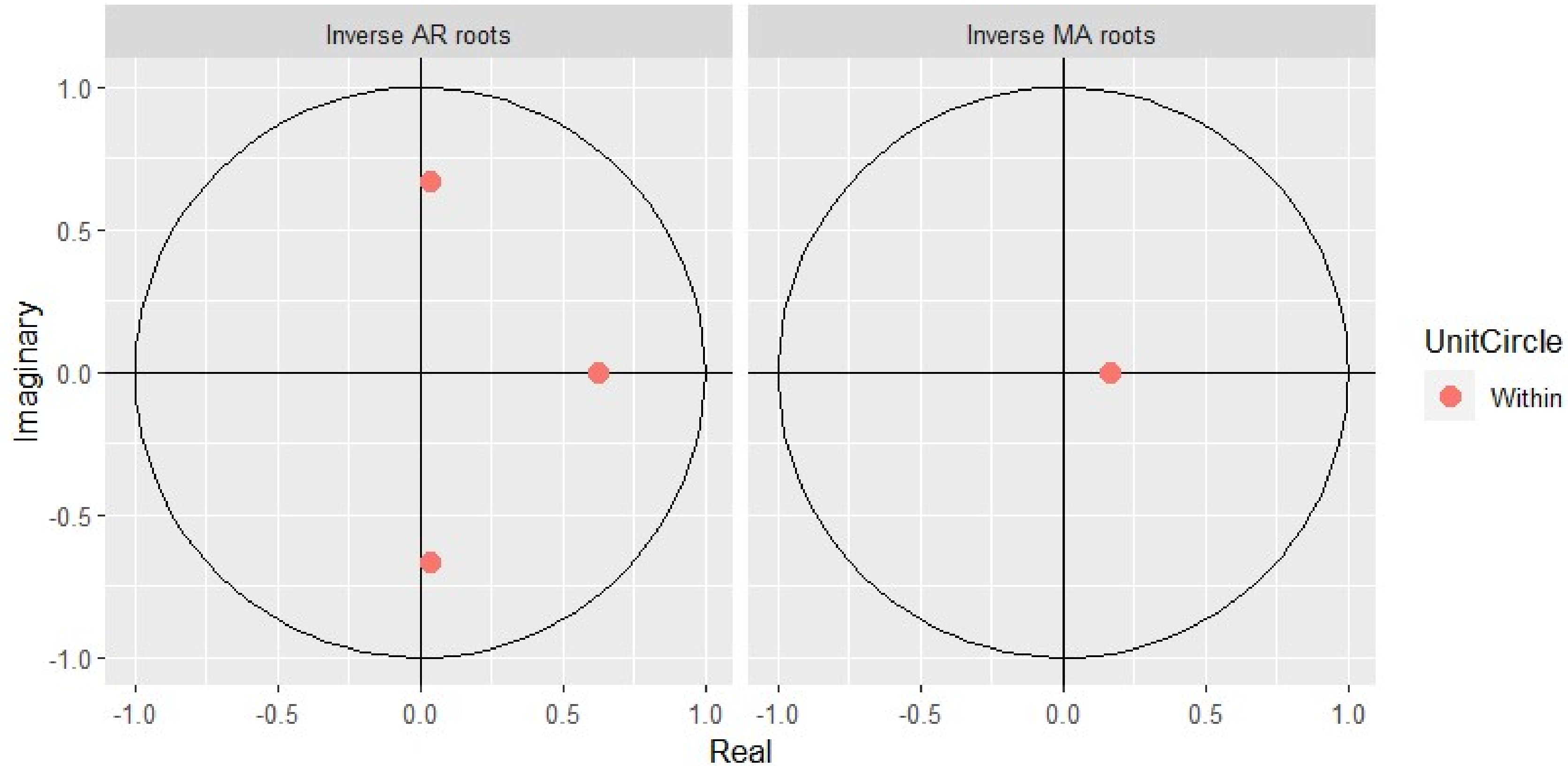


En la figura de raíces inversas de AR, se observa en el gráfico las raíces características se encuentran dentro del círculo, es decir que cumplen con la condición de estacionariedad para la parte autorregresiva.

Ejemplo 2

- Modelo 2: ARIMA(3, 1, 1)

```
autoplot(mod2)
```



Al estar los valores dentro de la circunferencia unitaria es un indicativo de que el modelo se ajusta correctamente. Tanto en su parte AR, como en su parte MA.

Ejemplo 2

Análisis de la estabilidad

- Si observamos nuevamente la gráfica de la serie diferenciada D1.Yt, aproximadamente a inicios del año 2014 puede presentar una cambio estructural y esto también se puede replicar en los modelos estimados, por tanto, ese punto de quiebre correspondería 48/159 = 0.30 de la muestra, entonces vamos a considerar 48 observaciones y aplicamos la prueba de Chow, en R utilizaremos la librería **strucchange** y las funciones **Fstats** y **sctest.**

- **Modelo 1**

```
library(strucchange)
# Generar un modelo respecto a una constante
Chow_mod1 <- Fstats(mod1$fitted ~ 1, from = 0.3)
sctest(Chow_mod1)

           supF test

data:  Chow_mod1
sup.F = 139.36, p-value < 2.2e-16
```

Ejemplo 2

- **Modelo 2**

```
Chow_mod2 <- Fstats(mod2$fitted ~ 1, from = 0.3)
sctest(Chow_mod2)

           supF test

data:  Chow_mod2
sup.F = 138.84, p-value < 2.2e-16
```

De los resultados, en los dos modelos se rechaza la hipótesis nula ($p < \alpha = 0.05$), es decir, no existe estabilidad estructural de coeficientes.

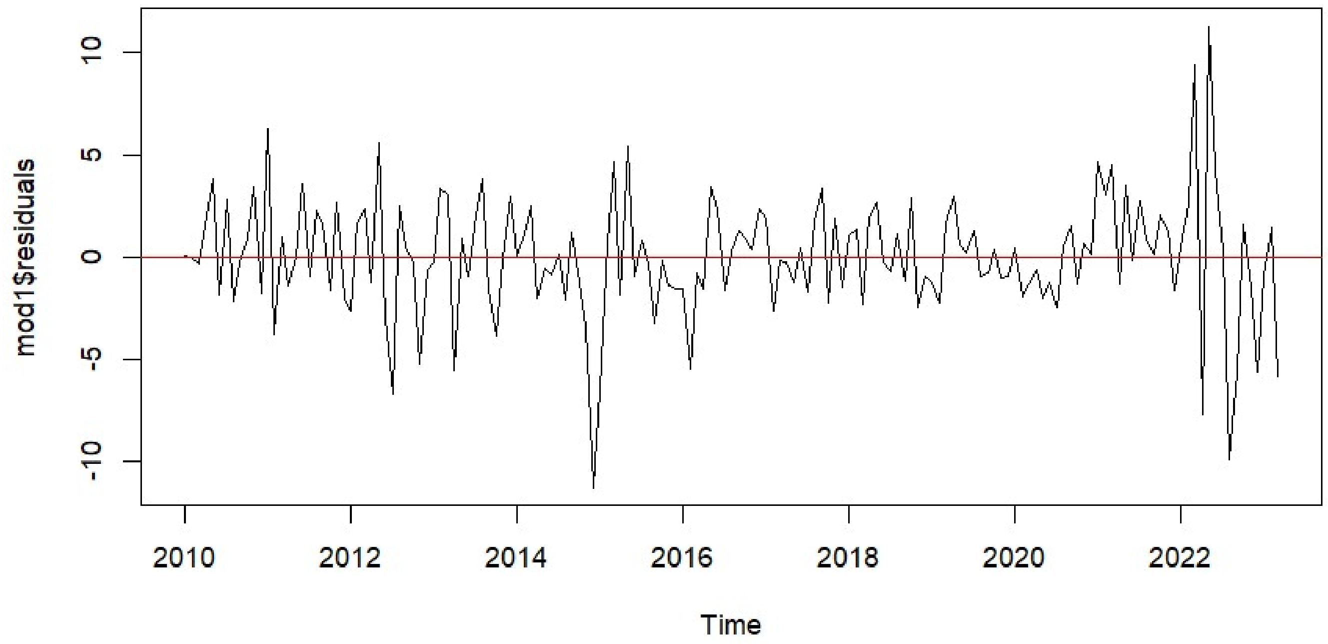
Ejemplo 2

Análisis de los residuos

- Media igual a cero

- **Modelo 1**

```
plot(mod1$residuals)
abline(h = 0, col = "red")
```



Prueba t

```
t.test(mod1$residuals, mu = 0)

            One Sample t-test

data:  mod1$residuals
t = -0.022577, df = 158, p-value = 0.982
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.4746400  0.4639115
sample estimates:
   mean of x
-0.005364241
```

Parece indicar que un buen número de residuales están en torno a la media igual a cero.

Para confirmar lo mencionado se realiza la prueba t:
Como p = 0.982 > α = 0.05, se acepta Ho, es decir la media es igual a cero.

Ejemplo 2

- **Modelo 2**

```
plot(mod2$residuals)
abline(h = 0, col = "red")
```

Prueba t

```
t.test(modelo2$residuals, mu = 0)

            One Sample t-test

data:  mod2$residuals
t = -0.040109, df = 158, p-value = 0.9681
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -0.4782623  0.4592241
sample estimates:
   mean of x
-0.009519068
```

También indica que un buen número de residuales están en torno a la media igual a cero.

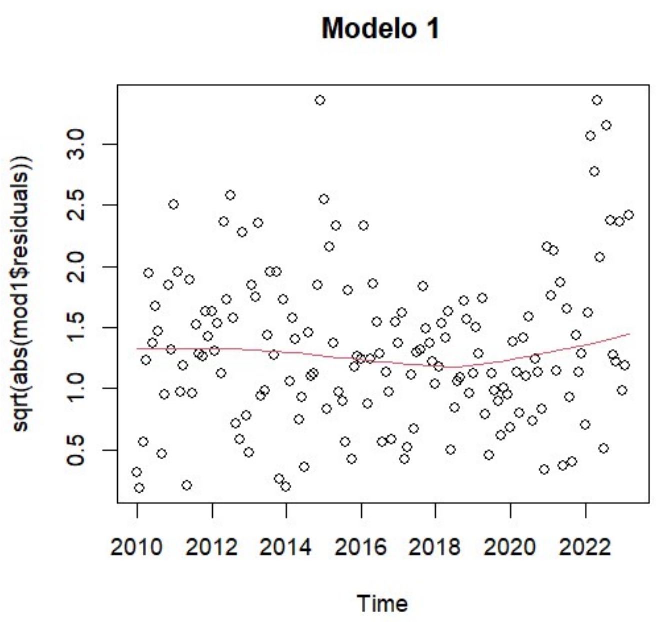
Para confirmar lo mencionado se realiza la prueba t:
Como p = 0.9681 > α = 0.05, se acepta Ho, es decir la media es igual a cero.

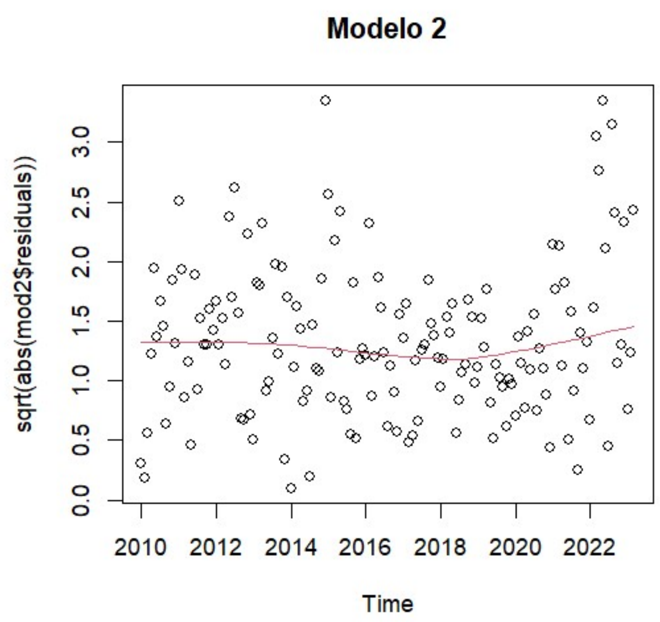
Ejemplo 2

Homocedasticidad de los residuales

Para analizar la homocedasticidad de los residuales realizaremos el diagrama de dispersión de la **raíz cuadrada de los residuales absolutos**.

```
par(mfrow = c(1,2))
scatter.smooth(sqrt(abs(mod1$residuals)), lpars=list(col=2), main = "Modelo 1")
scatter.smooth(sqrt(abs(mod2$residuals)), lpars=list(col=2), main = "Modelo 2")
```





Se observa que los datos presentan una variabilidad considerable, por tanto, será necesario realizar la prueba de Breusch-Pagan para determinar finalmente si las varianzas constantes para ambos modelos

Ejemplo 2

Prueba de Breusch - Pagan

Modelo 1:

```
obs=get(mod1$series)
bptest(resid(mod1)~I(obs-resid(mod1)))

          studentized Breusch-Pagan test

data:  resid(mod1) ~ I(obs - resid(mod1))
BP = 9.204, df = 1, p-value = 0.002415
```

Modelo 2:

```
obs=get(mod2$series)
bptest(resid(mod2)~I(obs-resid(mod2)))

          studentized Breusch-Pagan test

data:  resid(mod2) ~ I(obs - resid(mod2))
BP = 9.0842, df = 1, p-value = 0.002578
```

El valor de probabilidad (p-valor) asociado al estadístico BP asume un valor de 0.002415 y 0.002578 para el modelo 1 y modelo 2 respectivamente y estos son menores $\alpha = 0.05$, entonces se rechaza Ho.

Se concluye que los residuos no tienen varianza constante. Finalmente, podemos indicar que a través de este método los residuos no son ruido blanco.

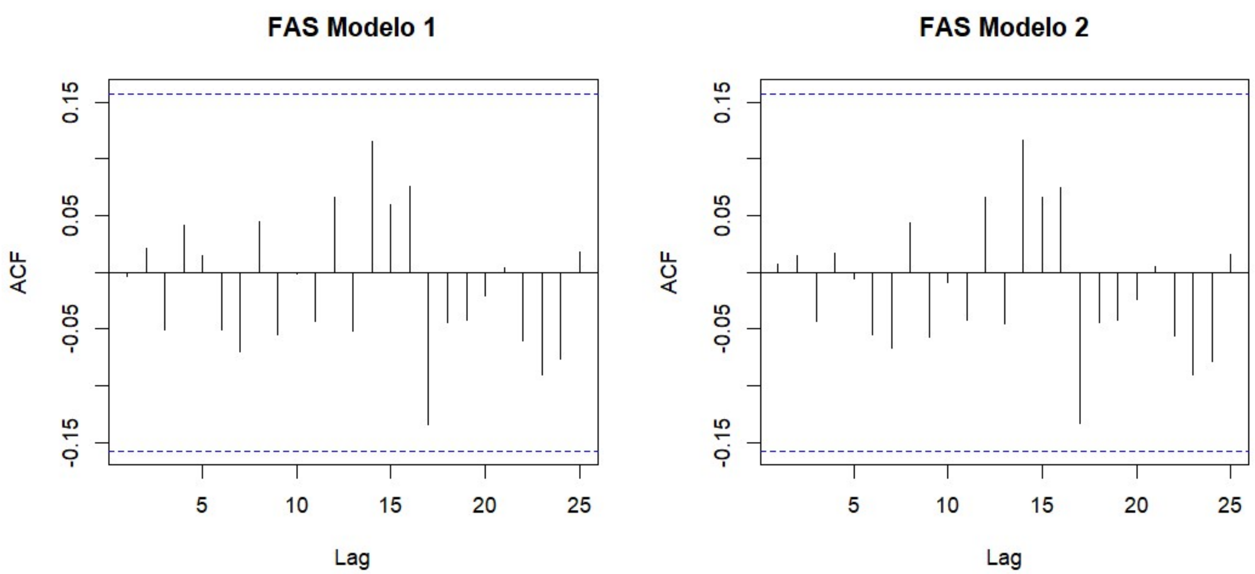
Para solucionar este problema, aparecen modelos nombrados como modelos con heterocedasticidad condicional autoregresiva (**ARCH**), los cuales se analizarán en los siguientes capítulos.

Ejemplo 2

Ausencia de correlación serial

Correlograma de los residuos

```
resid_m1 <- mod1$residuals[4:158]
resid_m2 <- mod2$residuals[4:158]
par(mfrow = c(1,2))
FAS_e.m1 <- acf(resid_m1, lag.max = 25, main="FAS Modelo 1", level = 0.95)
FAS_e.m2 <- acf(resid_m2, lag.max = 25, main="FAS Modelo 2", level = 0.95)
```



La totalidad de los coeficientes de autocorrelación simple estimados se encuentran dentro de las bandas de no significación, sobre todo los de los primeros retardos.

Por tanto, tenemos altos indicios de que los residuos del modelo estimado sean ruido blanco.

Ejemplo 2

Prueba de Ljung - Box

```
Box.test(resid_m1,type = "Ljung-Box")

          Box-Ljung test

data:  resid_m1
X-squared = 0.0018387, df = 1, p-value = 0.9658

Box.test(resid_m2,type = "Ljung-Box")

          Box-Ljung test

data:  resid_m2
X-squared = 0.0085309, df = 1, p-value = 0.9264
```

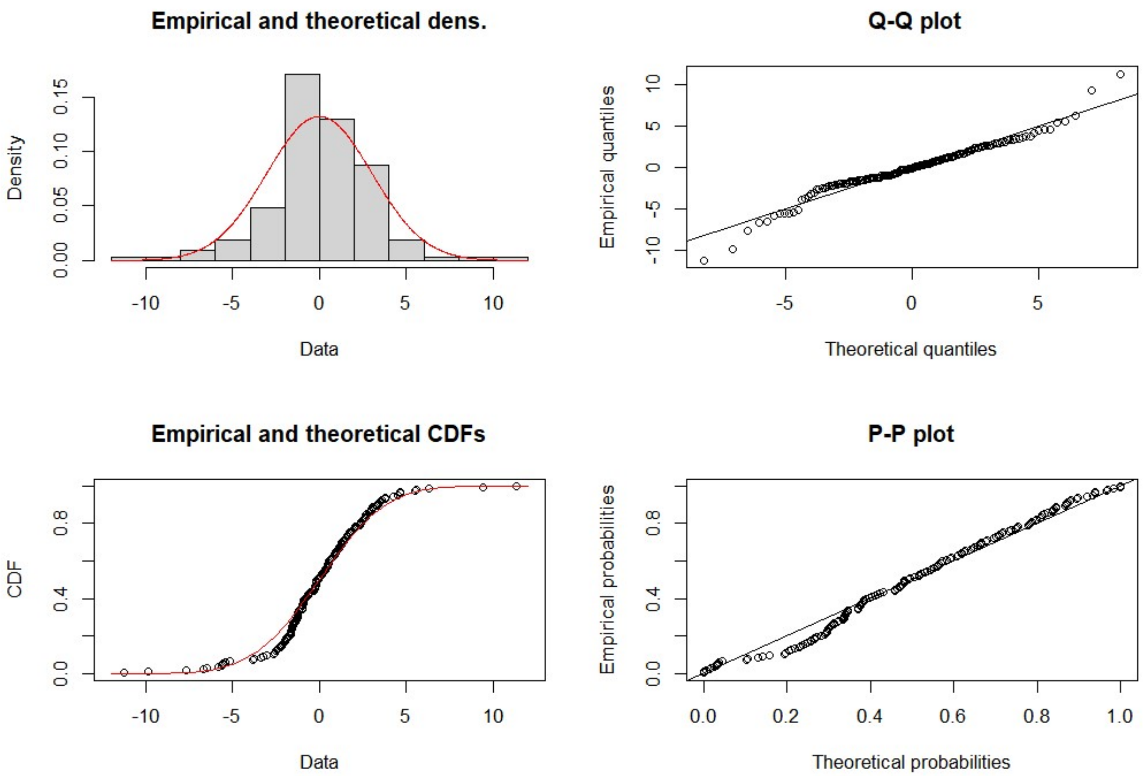
En ambos casos se acepta la hipótesis nula de que los coeficientes de autocorrelación son cero; es decir, los residuos son independientes o están incorrelacionados.

Ejemplo 2

Contraste de normalidad

Método gráfico - Modelo 1

```
library(fitdistrplus)
ajuste_m1<-fitdist(data = resid_m1, distr="norm")
plot(ajuste_m1)
```



Prueba de normalidad de Jarque-bera

```
JB_m1 <- jarque.bera.test(resid_m1)
JB_m1

           Jarque Bera Test

data:  resid_m1
X-squared = 47.274, df = 2, p-value = 5.428e-11
```

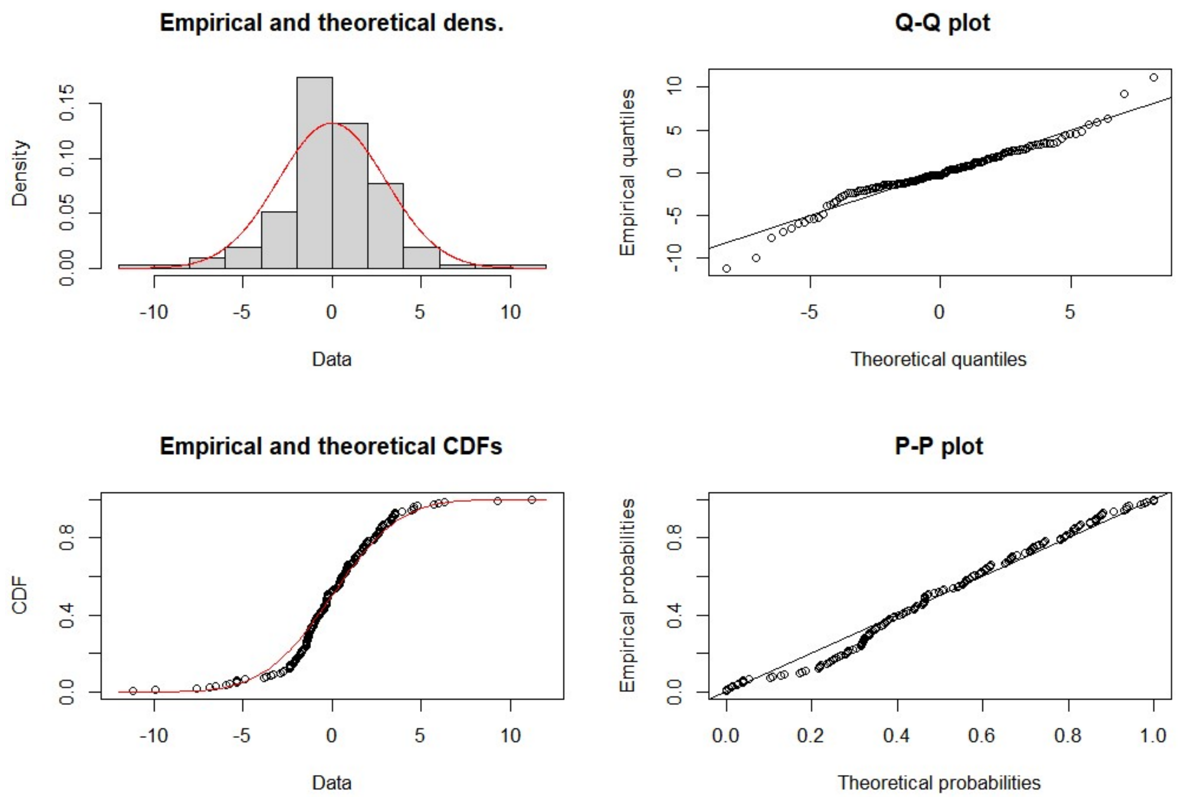
En las figuras se observa que los residuales del modelo 1 presentan indicios de ajuste hacia la distribución normal.

En la prueba JB, como p = 0.0000 < 0.05, se rechaza Ho, es decir, los residuos no se aproximan a una distribución normal.

Ejemplo 2

Método gráfico - Modelo 2

```
ajuste_m2<-fitdist(data = resid_m2, distr="norm")
plot(ajuste_m2)
```



Prueba de normalidad de Jarque-bera

```
JB_m2 <- jarque.bera.test(resid_m2)
JB_m2

        Jarque Bera Test

data:  resid_m2
X-squared = 45.969, df = 2, p-value = 1.042e-10
```

En las figuras se observa que los residuales del modelo 2 también presentan indicios de ajuste hacia la distribución normal.

En la prueba JB, como p = 0.0000 < 0.05, se rechaza Ho, es decir, los residuos no se aproximan a una distribución normal.

También se puede utilizar la prueba de Kolmogorov – Smirnov, pues esta no es tan estricta.

Ejemplo 2

Conclusiones

- En vista a los resultados, aunque la normalidad de los residuales es un supuesto común en muchas técnicas estadísticas, su violación no invalida automáticamente el modelo.

- Muchos modelos pueden ser robustos a desviaciones menores de la normalidad, especialmente si el tamaño de la muestra es lo suficientemente grande.

- Además, es posible aplicar transformaciones a los datos o ajustar modelos alternativos que se adapten mejor a la estructura de los residuales.

- En resumen, un modelo ARIMA con residuales sin correlación y una distribución no normal aún puede ser valioso para el análisis de series de tiempo y la toma de decisiones, siempre y cuando se comprendan sus limitaciones y se realicen verificaciones adecuadas de su rendimiento.

GRACIAS

https://aulavirtual2.unap.edu.pe/