Series de Tiempo

Gómez Rodríguez Angel Giovanni

28 de noviembre de 2021

Introducción.

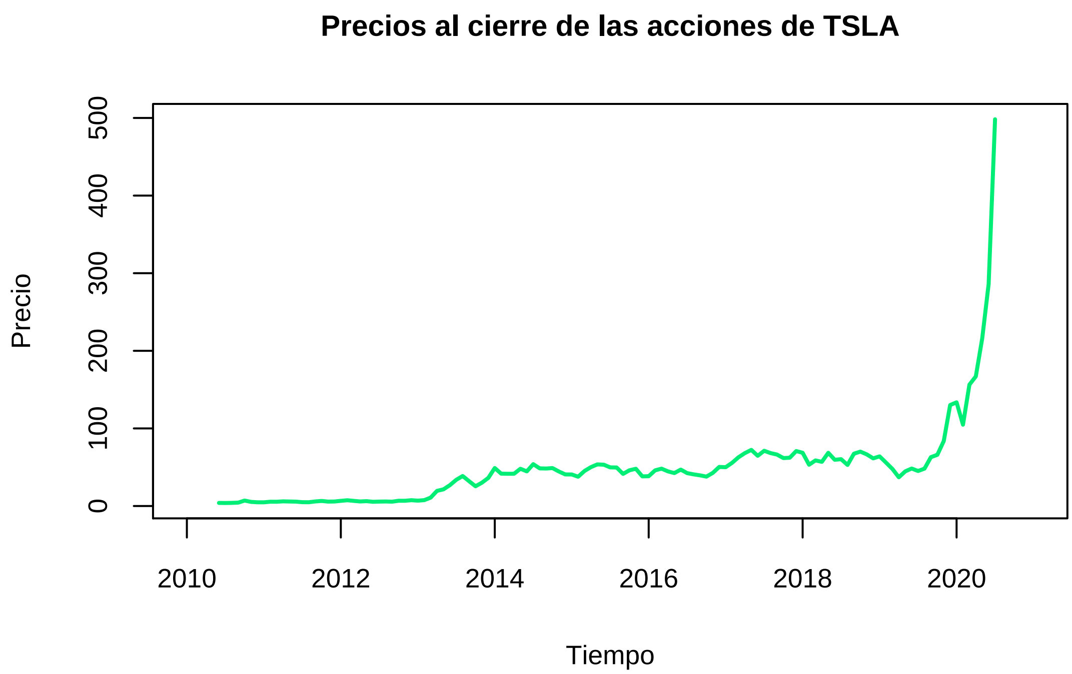
La elección de Tesla como empresa para el proyecto es considerar una entidad que acelere la transición del mundo hacia la energía sostenible.

En la actualidad, Tesla fabrica no solo vehículos totalmente eléctricos, sino también productos de almacenamiento y generación de energía limpia infinitamente escalables. Al igual que Tesla, pienso que el mundo será mejor en cuanto se deje de depender de los combustibles fósiles y se avance hacia un futuro de cero emisiones.

1. Analizar los datos de manera descriptiva.

Grafique los datos, describa lo que observe (varianza constante o no constante, descomposición clásica, tendencia, ciclos estacionales, periodicidad de los ciclos).

Para la serie de tiempo se tienen datos mensuales, de donde:



Descripción de la serie

Varianza

A simple vista se observa una varianza no constante, ya que, si se la trazan bandas a la gráfica, estas cambian con el paso del tiempo. Pero para verificarlo, se debe realizar alguna prueba. De esta manera se tienen las hipótesis:

$$H_0\text{: Varianza constante (homocedasticidad) vs } H_a\text{: Varianza no constante (heterocedasticidad)}$$

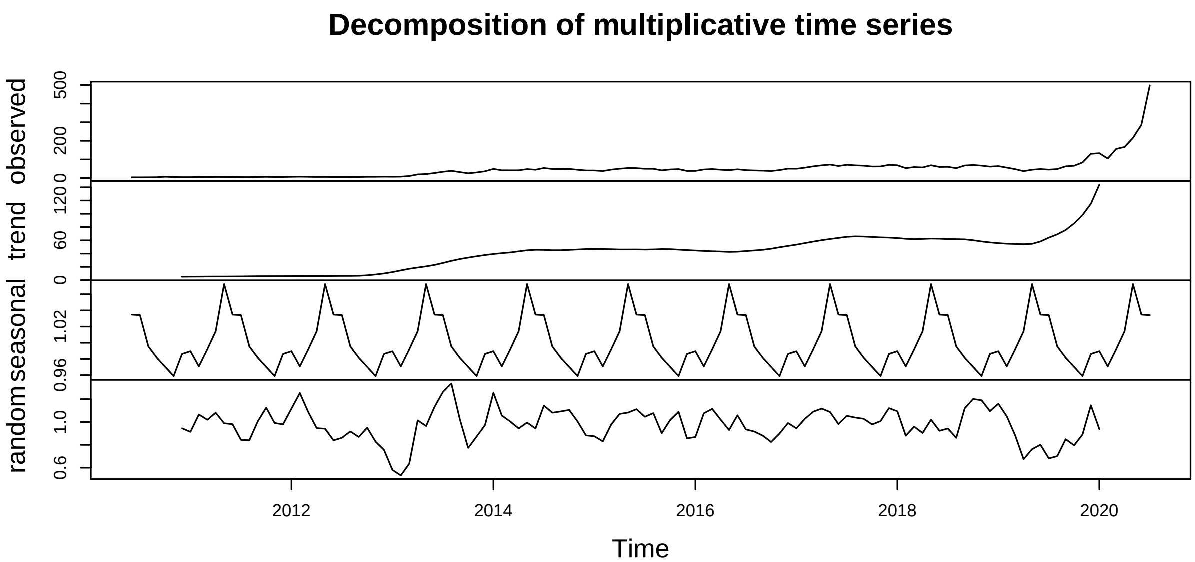
Aplicando la *Prueba de Breusch-Pagan*, se obtiene:

```
##
##  studentized Breusch-Pagan test
##
## data:  Y1 ~ X1
## BP = 6, df = 1, p-value = 0.01
```

Y como el $p-value = 0.01 < 0.05$, se rechaza H_0 y se concluye que **la serie no tiene varianza constante.**

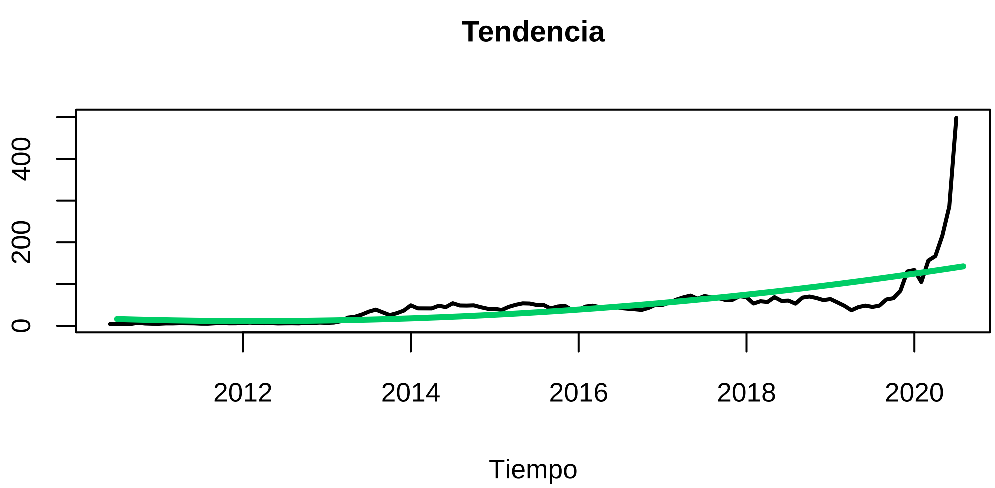
Descomposición clásica

Por la observación anterior, se tiene que la varianza no es constante. Además, se observa que hay cambios en la dispersión de los datos con respecto del tiempo, porque si se crearan bandas éstas cambiarían a lo largo del tiempo. Con estas 2 últimas observaciones se puede concluir de la gráfica que la serie es de un modelo multiplicativo. De donde:



Tendencia

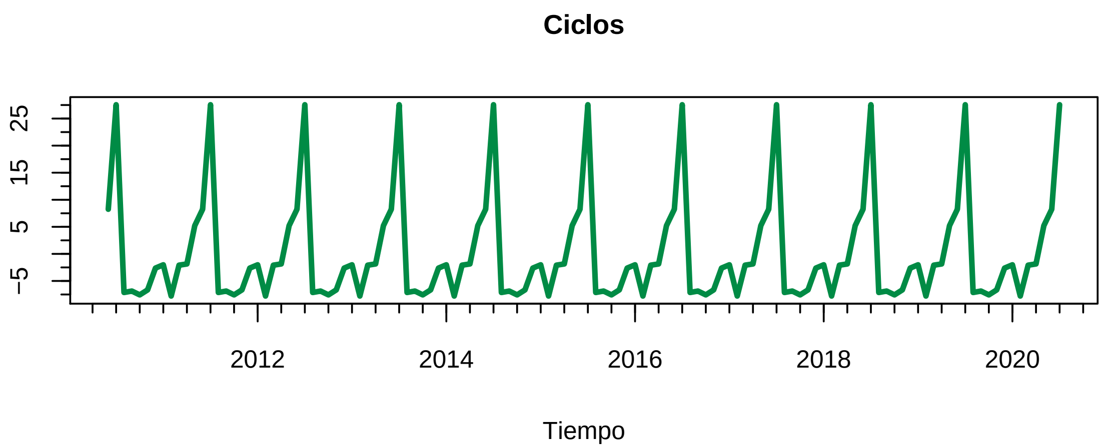
Al analizar la tendencia, se le puede ajustar una recta mediante un *modelo de regresión lineal.*



Se puede apreciar que a pesar de tener pequeñas fluctuaciones, **la serie tiene una tendencia que va a la alza.**

Ciclos

Los ciclos se ven de la siguiente manera:



El comportamiento de cada ciclo indica que el precio de las acciones:

- Tiene su máximo valor en junio y luego disminuye, posteriormente cambia de manera constante
- Incrementa su valor en el mes de septiembre
- Comienza a disminuir en diciembre, antes de que termine el año; alcanza su valor mínimo
- Empieza a subir abruptamente en enero
- Se mantiene constante durante marzo
- Sube bastante a partir del mes de abril hasta llegar a junio

Periodicidad

La periodicidad de los ciclos es anual. Cada periodo comienza a partir del mes de junio, esto se debe a que las acciones de Tesla, dan inicio el 29 de junio de 2010. Además, para contar con los 10 años completos, se concluye con el análisis de datos en julio de 2020, pues así se pueden considerar los datos hasta inicios de julio de 2020, en una fecha a cercana a cuando Tesla comenzó con el precio de sus acciones.

Observaciones adicionales

Los datos históricos corresponden a una serie de tiempo que muestra el comportamiento de los precios al cierre del mercado; el precio se maneja en dólares. Se contemplan precios mensuales.

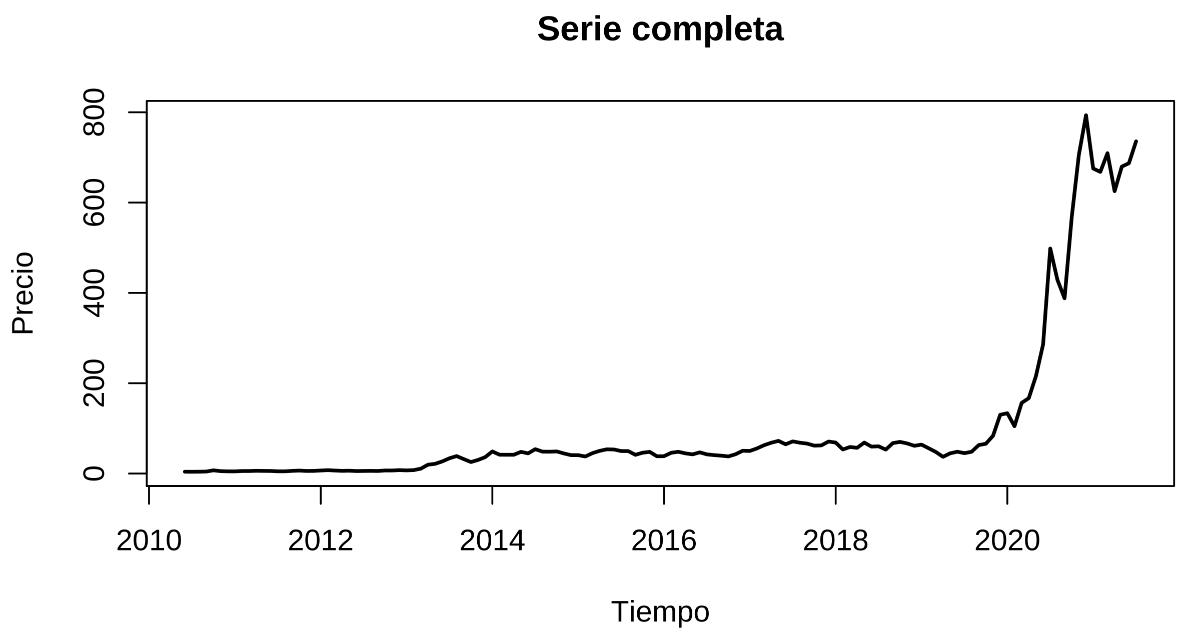
El análisis de datos comienza a finales de junio de 2010 y termina en julio de 2020 con el objetivo de que se contemplen años completos y se tomen en cuenta los datos de finales de junio, ya que las acciones de Tesla empezaron el 29 de junio de 2010.

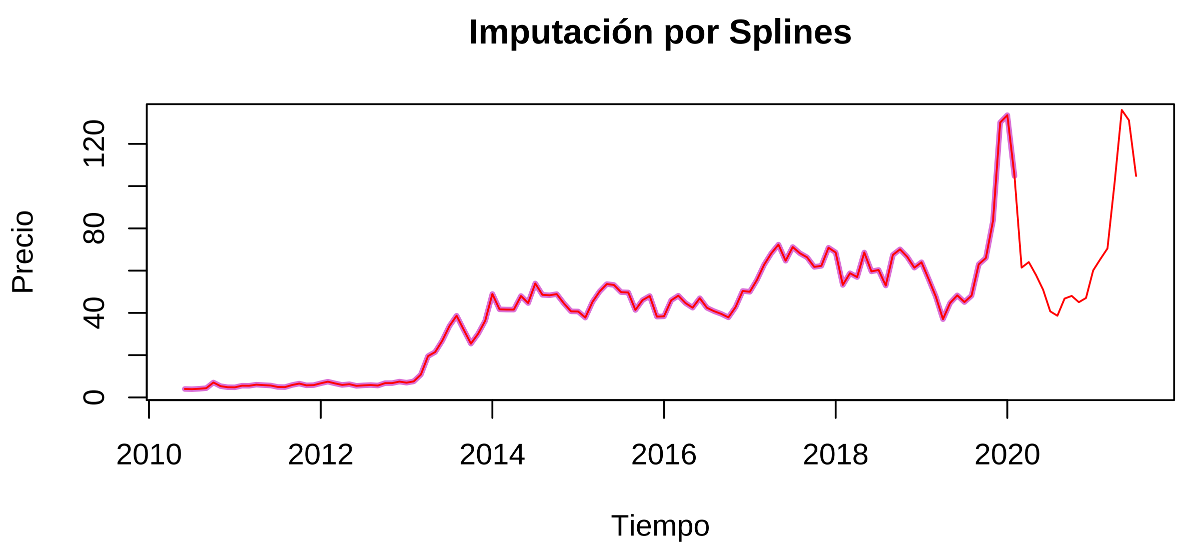
Finalmente, tras comparar la descomposición clásica con el modelo de regresión lineal, se puede observar que los ciclos y la tendencia son bastante similares en ambos modelos.

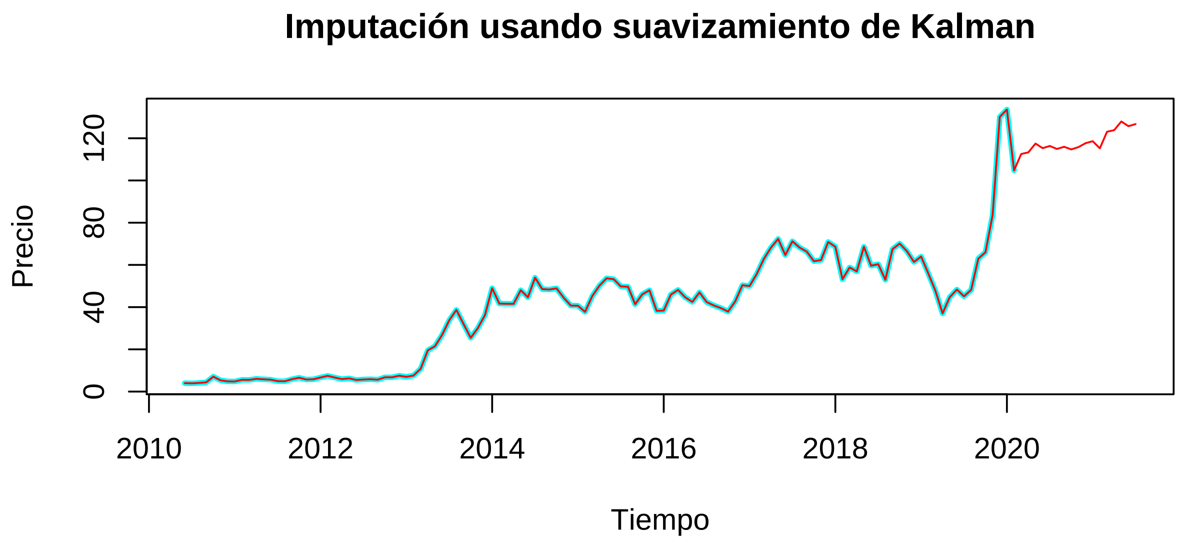
2. Métodos de imputación.

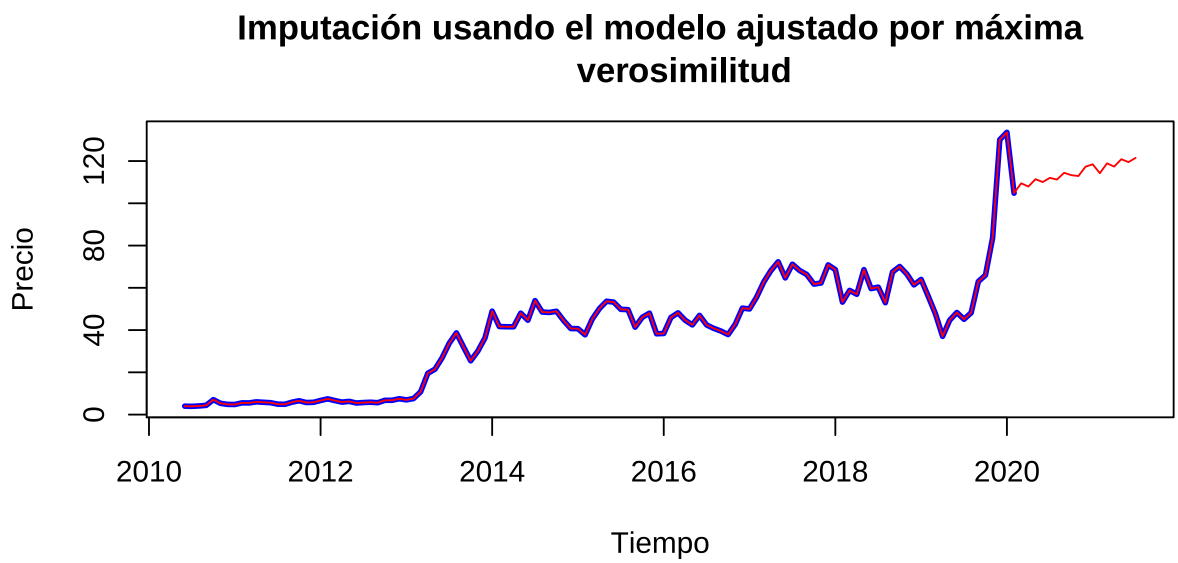
Si la base presenta datos faltantes NA. Use algún método de imputación de la paquetería imputeTS.

En este caso se contemplan los datos históricos de 11 años, es decir hasta julio del año 2021; posteriormente se eliminan los datos a partir de marzo de 2020, es decir, a partir de que comienza la emergencia sanitaria a nivel mundial. El propósito de realizar esto es ver si la imputación es acertada.









Comparativa:

Periodo	Valor real	Splines	Kalman	Máxima verosimilitud
Marzo 2020	156.38	61.42	112.53	109.47
Agosto 2020	429.01	38.65	114.93	111.24
Enero 2021	675.50	60.11	118.61	118.47
Junio 2021	687.20	131.17	125.72	119.51

Conclusiones

- El “mejor” método para imputar estos datos es con el modelo ajustado por máxima verosimilitud. El método más deficiente es el de splines.
- Realmente ninguno de los métodos se acerca a los datos reales, esto debido a que Tesla ha tenido un crecimiento enorme en meses recientes.

3. Usar un método de descomposición clásico.

Pueden ser diferencias, filtros, modelos de regresión, suavizamientos exponenciales para calcular sus componentes.

Regresando a la serie original que contempla datos de 2010 a 2020.

Para realizar la descomposición clásica de esta serie de tiempo, hay que estabilizar la varianza, para esto, se eligirá a la transformación logaritmica:

Y verificamos si pasa la Prueba De esta manera se tienen las hipótesis:

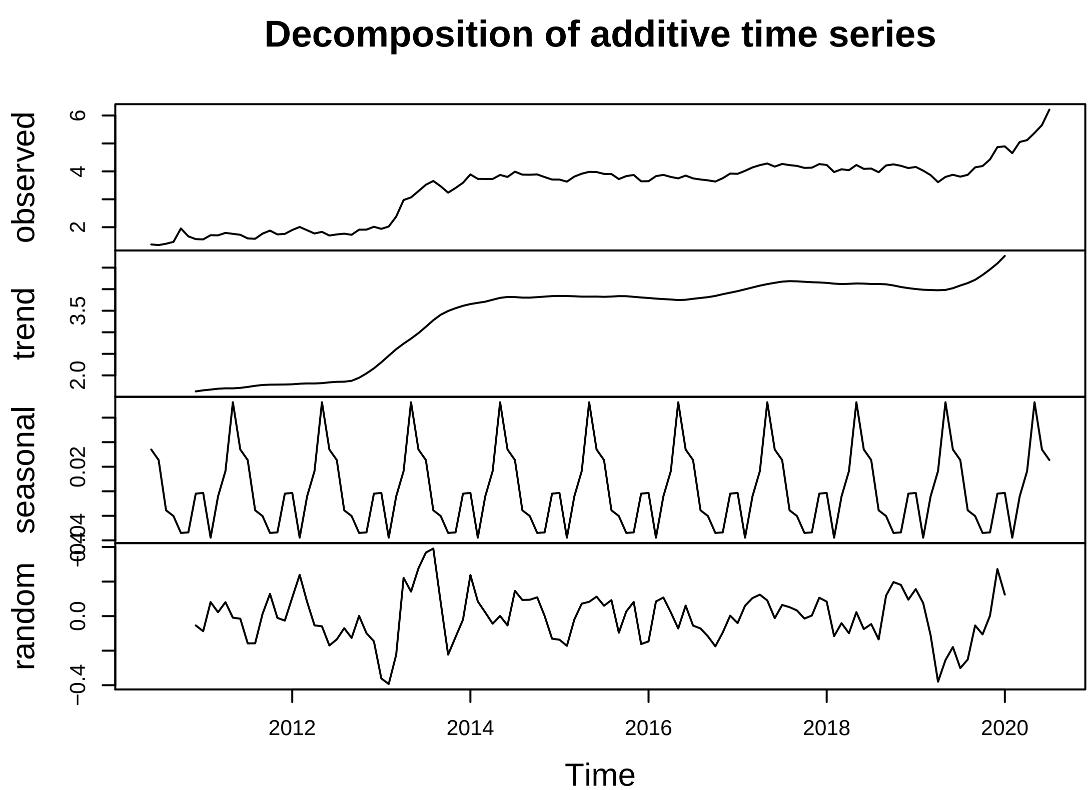
H_0: Varianza constante (homocedasticidad) vs H_a: Varianza no constante (heterocedasticidad)

Aplicando la *Prueba de Breusch-Pagan*, se obtiene:

```
##
##  studentized Breusch-Pagan test
##
## data:  Y1 ~ X1
## BP = 4, df = 1, p-value = 0.06
```

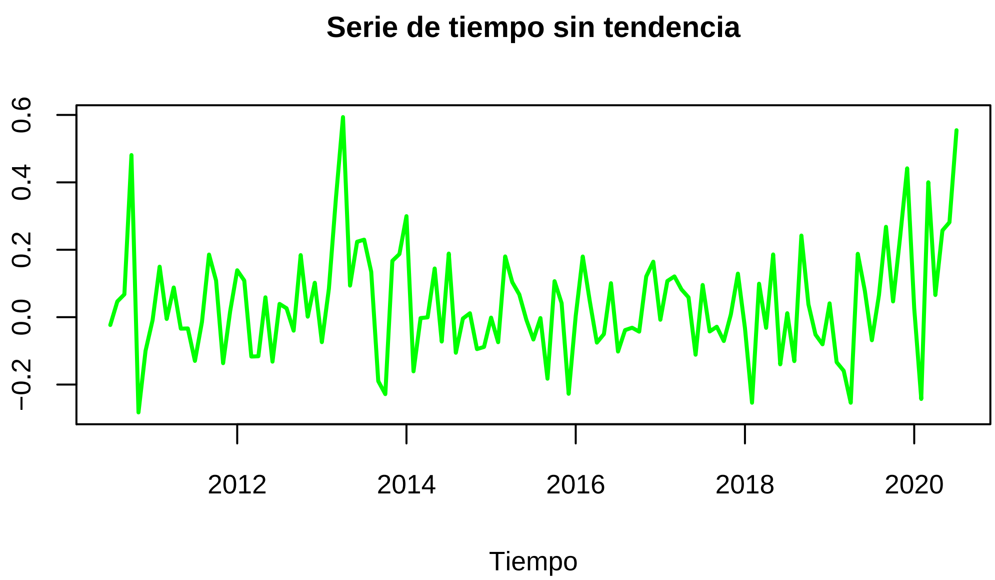
Y como el $p - value = 0.06 > 0.05$, se acepta H_0 y se concluye que **la serie ya tiene varianza constante.**

Descomposición por defecto de R.

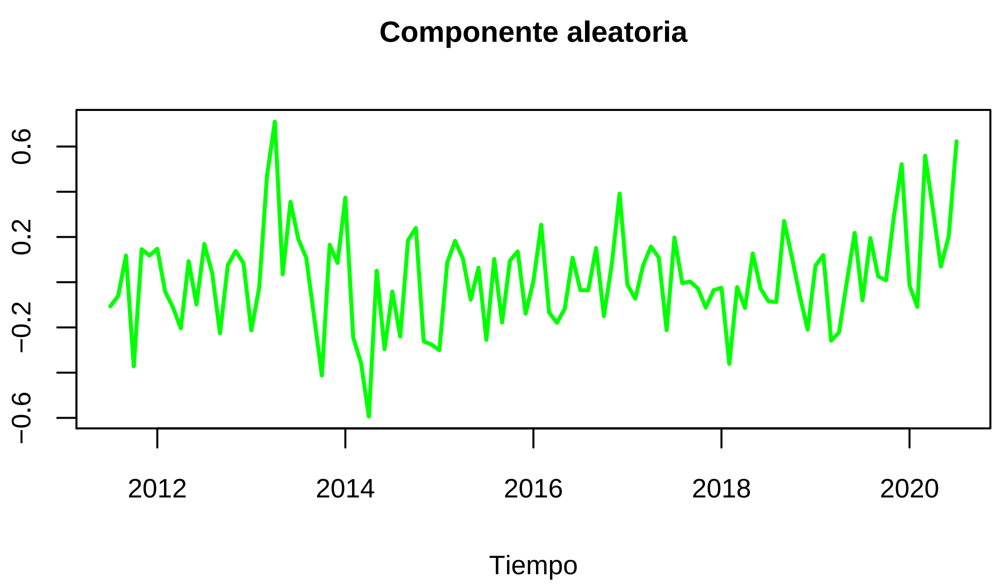


Diferencias.

Empleando el operador *diff* dado por $Y_t = \nabla X'_t - X'_{t-1}$ se elimina la tendencia:



Eliminando los ciclos estacionarios con $W_t = \nabla_{12} X'_t = X'_t - X'_{t-12}$ obtenemos la componente aleatoria:



4. Realizar un ajuste de series de tiempo adecuado.

El modelo puede ser ARIMA, SARIMA o modelos GARCH.

Para empezar a modelar hay que verificar que la serie sea homocedástica. Recordemos que los datos de la serie original no tienen varianza constante.

```
##
##  studentized Breusch-Pagan test
##
## data:  Y4 ~ X4
## BP = 6, df = 1, p-value = 0.01
```

Así que se tienen que transformar los datos para que tengan una varianza constante.

Para ello se realizará la transformación por medio de dos métodos:

- En el primero, se le aplicará un logaritmo a la serie original
- En segundo, se realizará la transformación Box-Cox por método Guerrero

Logaritmo

```
Teslalog <- log(TeslaStockData)
```

Para saber si se cumple la varianza constante, se realiza el *bptest*, de donde:

```
##
##  studentized Breusch-Pagan test
##
## data:  Y5 ~ X5
## BP = 0.03, df = 1, p-value = 0.9
```

Así, se obtiene un $p - value = 0.9 > 0.05$, con lo cual, ya no se rechaza H_0 y así, se confirma que los datos ya tienen varianza constante.

BoxCox Guerrero

En este caso primero se genera la lambda y posteriormente se generan los datos:

```
TeslaGuerrero = BoxCox(TeslaStockData, lambda1)
```

Y finalmente para saber si existe varianza constante, se realiza la *Prueba de Breusch-Pagan*

```
##
##  studentized Breusch-Pagan test
##
## data:  Y6 ~ X6
## BP = 3, df = 1, p-value = 0.1
```

De donde, resulta un $p - value = 0.1$, con lo cual, no se rechaza H_0 y se confirma que los datos tienen varianza constante.

Realización auto.arima

Logaritmo

```
(fit1 <- auto.arima(Teslalog, allowdrift = F))
```

```
## Series: Teslalog
## ARIMA(1,1,1)
##
## Coefficients:
##          ar1     ma1
##        0.934  -0.815
## s.e.   0.082   0.115
##
## sigma^2 = 0.0259:  log likelihood = 50.3
## AIC=-94.6   AICc=-94.4   BIC=-86.2
```

Por lo que para estos datos tenemos un ARIMA(1,1,1)

BoxCox Guerrero

```
(fit2 <- auto.arima(TeslaGuerrero, allowdrift = F))
```

```
## Series: TeslaGuerrero
## ARIMA(1,1,0)
##
## Coefficients:
##          ar1
##        0.161
## s.e.   0.092
##
## sigma^2 = 0.0121:  log likelihood = 95.9
## AIC=-188   AICc=-188   BIC=-182
```

Obtenemos un ARIMA(1,1,0)

Significancia de los parámetros.

Logaritmo

```
##      2.5 % 97.5 %
## ar1  0.772   1.10
## ma1 -1.041  -0.59
```

Todos los parámetros son significantes, ya que, ninguno contiene al 0 en su intervalo de confianza.

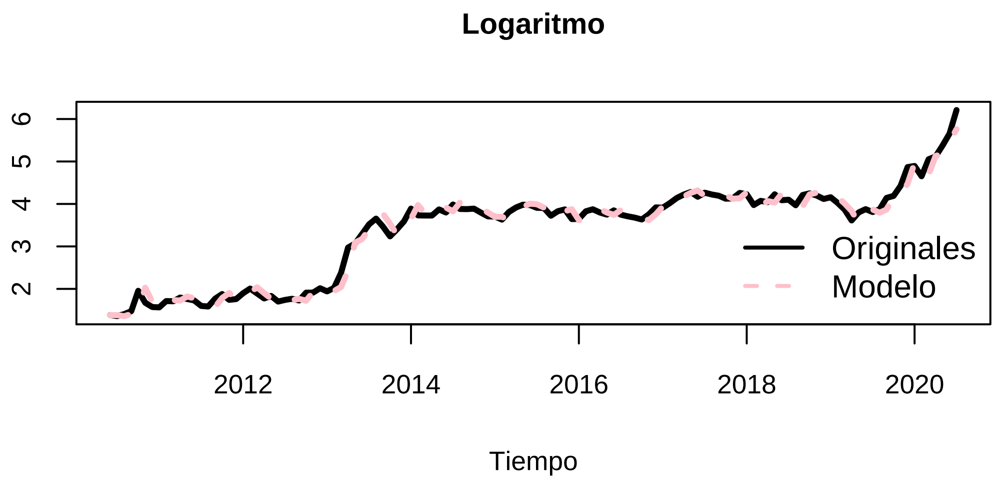
BoxCox Guerrero

```
##      2.5 % 97.5 %
## ar1 -0.019  0.341
```

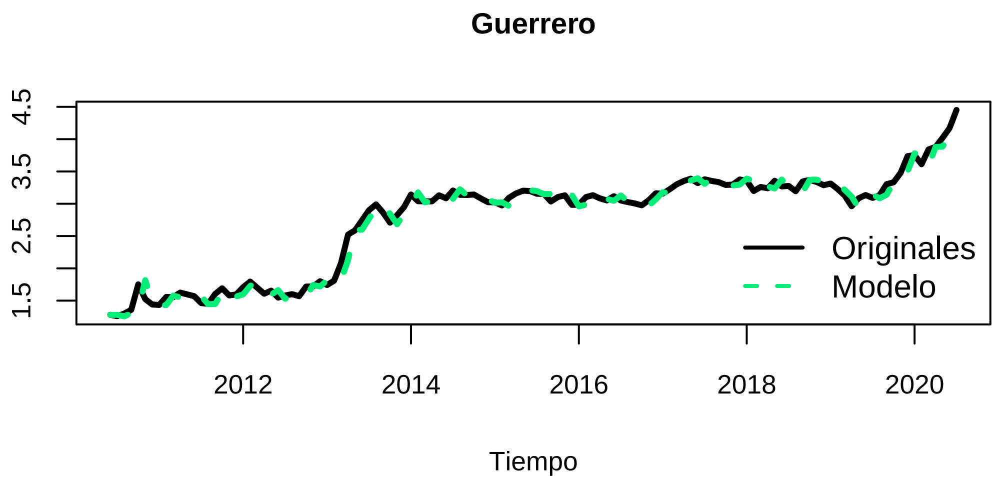
En este caso el parámetro no es significante, ya que, $ar1$ contiene al 0 en su intervalo de confianza.

Ajuste de los datos.

Logaritmo

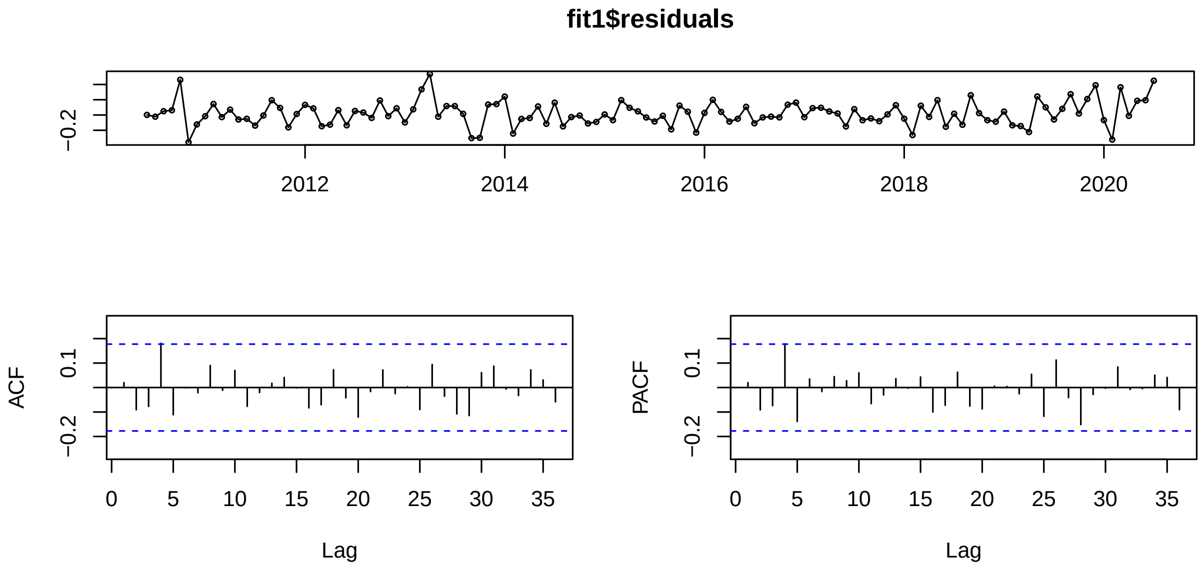


BoxCox Guerrero



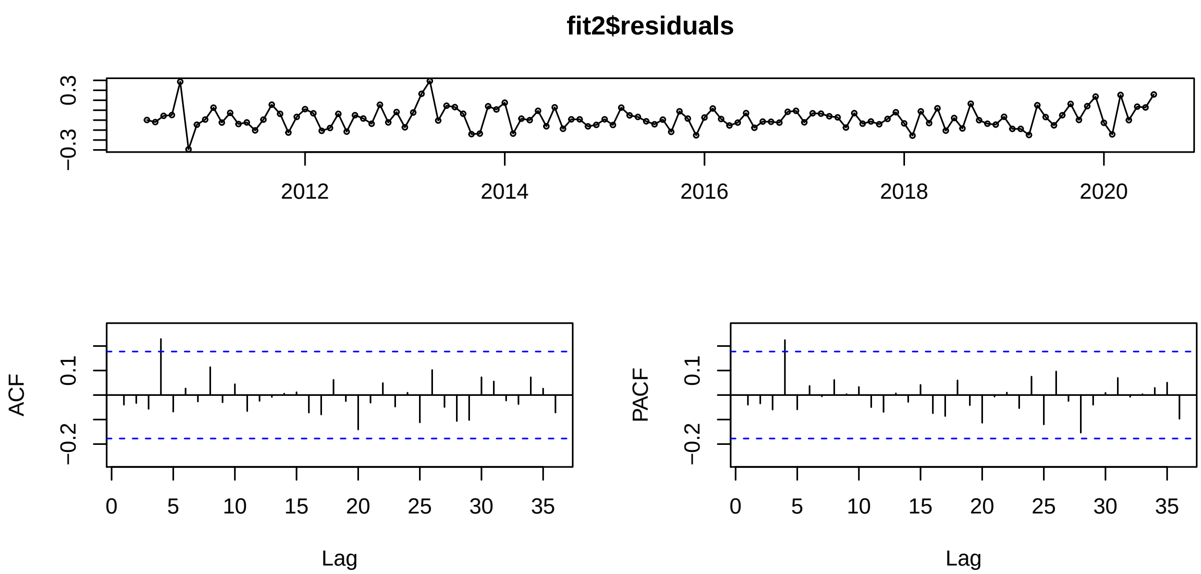
5. Revisar supuestos del modelo.

Correlograma para logaritmo



Parece que en el lag=4 está al limite de la banda de confianza, pero no se sale, por lo que aparenta no tener errores significativos.

Correlograma para Método Guerrero



Vemos que solo en el lag 4 podría dar problemas, pero no es tan significativo su valor sobrante después de la banda de confianza.

Residuos con media cero para logaritmo

```
##
##  One Sample t-test
##
## data:  fit1$residuals
## t = 1, df = 121, p-value = 0.2
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.00776  0.04893
## sample estimates:
## mean of x
##    0.0206
```

Con un $p - value = 0.2$, y una media estimada con un valor de 0.0206, se puede concluir que los residuos tienen media cero.

Residuos con media cero por Método Guerrero

```
##
##  One Sample t-test
##
## data:  fit2$residuals
## t = 2, df = 121, p-value = 0.02
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.00299 0.04142
## sample estimates:
## mean of x
##    0.0222
```

Con un $p-value = 0.02, y una media estimada con un valor de 0.0222, se puede decir que los residuos no tienen media cero.

Varianza constante para residuos de logaritmo

```
##
##  studentized Breusch-Pagan test
##
## data:  fit1$residuals ~ t
## BP = 0.3, df = 1, p-value = 0.6
```

Se obtiene un $p - value = 0.6$, con lo cual no se rechaza H_0 y se confirma que la varianza es constante.

Varianza constante para residuos por Método Guerrero

```
##
##  studentized Breusch-Pagan test
##
## data:  fit2$residuals ~ t
## BP = 0.8, df = 1, p-value = 0.4
```

Se obtiene un $p - value = 0.4$, con lo cual no se rechaza H_0 y se confirma que la varianza también es constante.

Residuos con distribución normal para logaritmo

```
##
##  Anderson-Darling normality test
##
## data:  fit1$residuals
## A = 0.7, p-value = 0.07
```

```
##
##  Shapiro-Wilk normality test
##
## data:  fit1$residuals
## W = 1, p-value = 0.04
```

```
##
##  Jarque Bera Test
##
## data:  fit1$residuals
## X-squared = 9, df = 2, p-value = 0.01
```

Para este modelo solo pasa la *Prueba de Anderson-Darling*, por lo que, no hay suficientes pruebas para decir que los datos se distribuyen normal.

Residuos con distribución normal para Método Guerrero

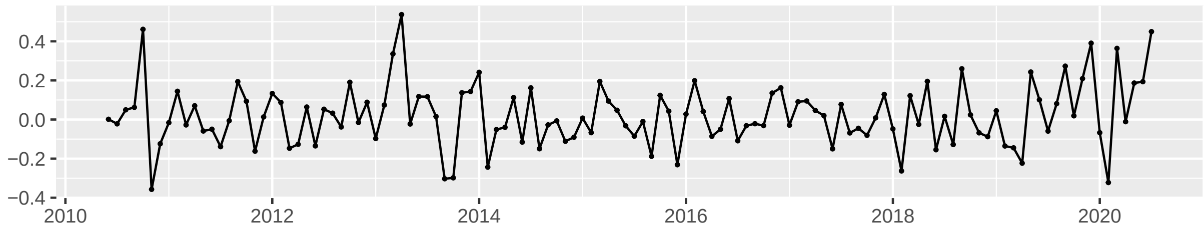
```
##
##  Anderson-Darling normality test
##
## data:  fit2$residuals
## A = 0.8, p-value = 0.04
```

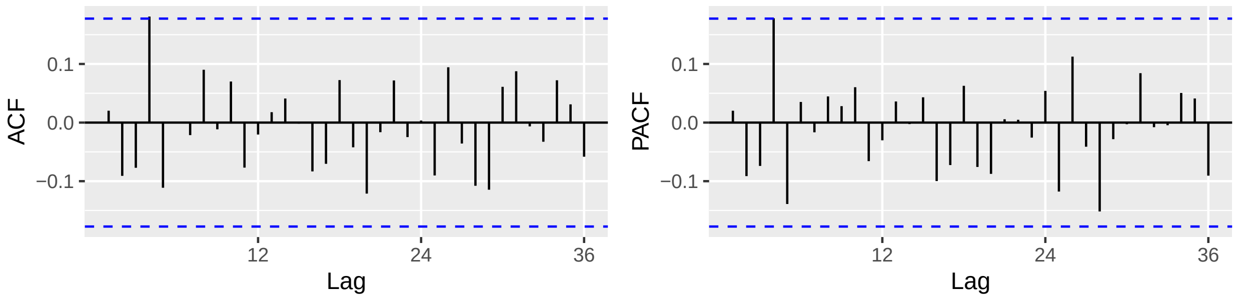
```
##
##  Shapiro-Wilk normality test
##
## data:  fit2$residuals
## W = 1, p-value = 0.005
```

```
##
##  Jarque Bera Test
##
## data:  fit2$residuals
## X-squared = 19, df = 2, p-value = 9e-05
```

Para este modelo no pasa los Test, por lo que, no hay pruebas para decir que los datos se distribuyen normal.

Datos independientes para Logaritmo





Parece que tenemos problemas con el lag = 4.

Primero realizamos la prueba general:

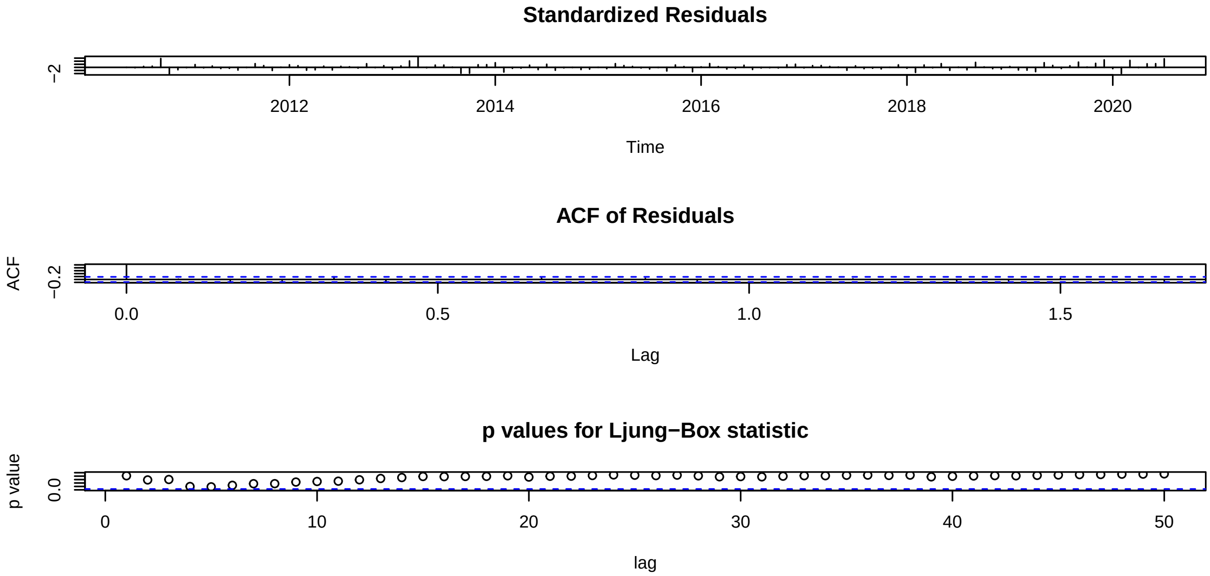
```
##
##  Box-Pierce test
##
## data:  fit1$residuals
## X-squared = 0.05, df = 1, p-value = 0.8
```

Al tener un $p-value = 0.8$, no rechazamos H_0 y podemos decir que los datos son independientes.

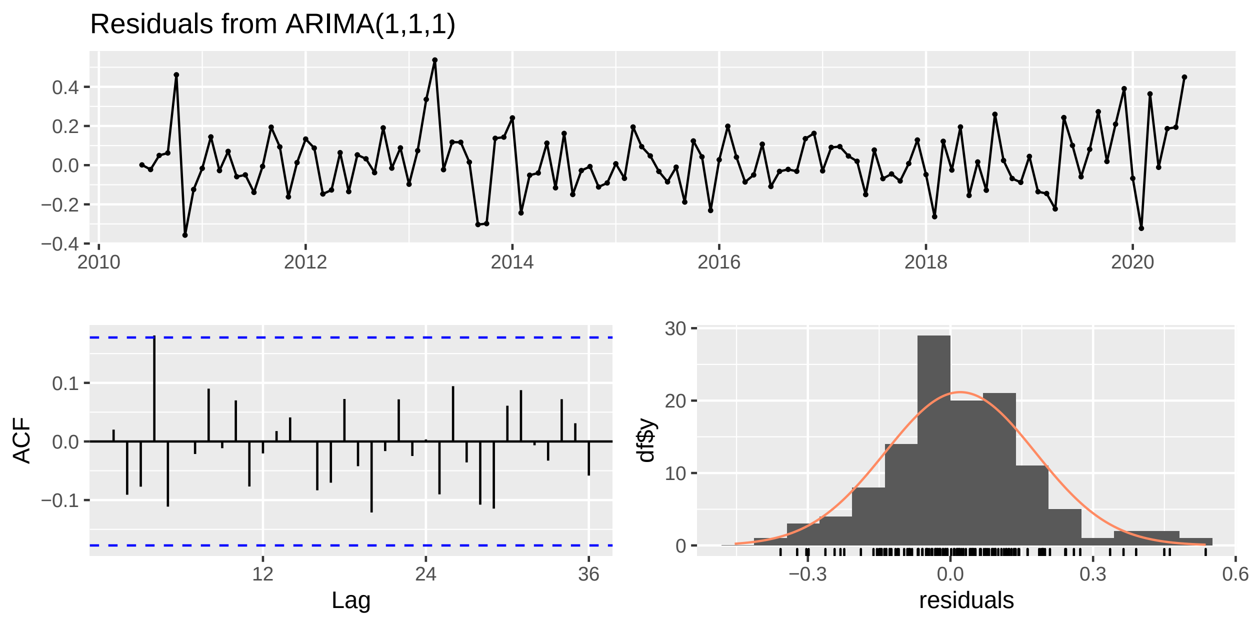
Ahora, para el lag = 4:

```
##
##  Box-Pierce test
##
## data:  fit1$residuals
## X-squared = 6, df = 4, p-value = 0.2
```

Tiene un $p-value = 0.2$, por lo que no hay problema y así se puede decir que los datos son independientes.



Vemos que por el lag 4 y 5 tenemos un pequeño problema.

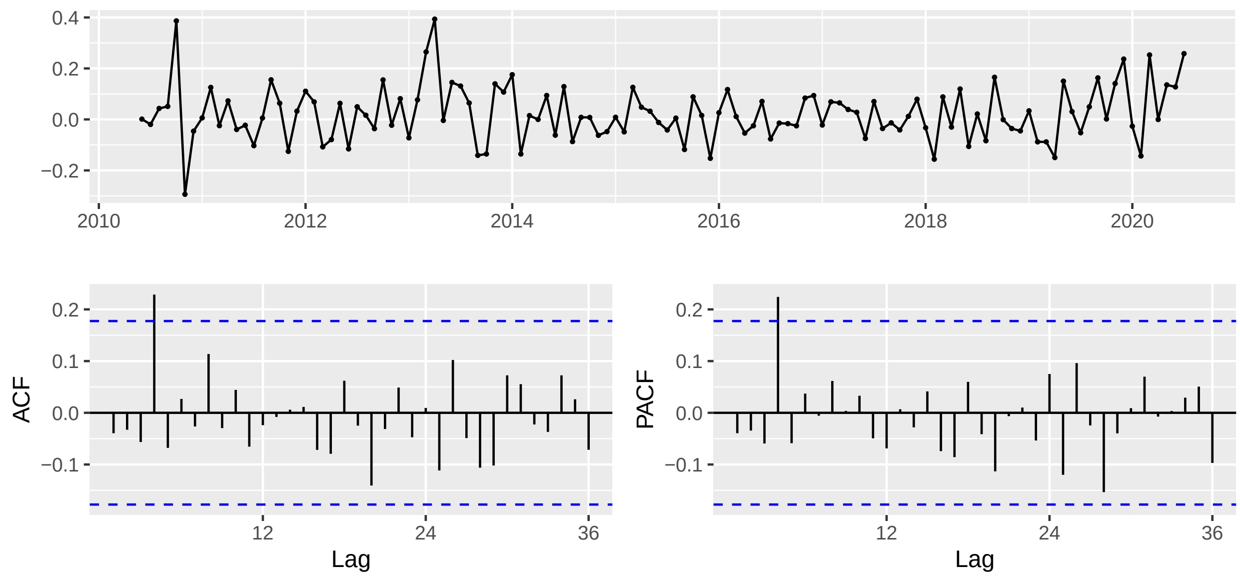



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,1)
## Q* = 16, df = 22, p-value = 0.8
##
## Model df: 2.   Total lags used: 24
```

Con un $p - value = 0.8$, si pasa la prueba Ljung Box.

Por lo que si son datos independientes.

Datos independientes por Método Guerrero



Parece que nuevamente hay problemas con el lag=4

Así que se realiza la prueba general:

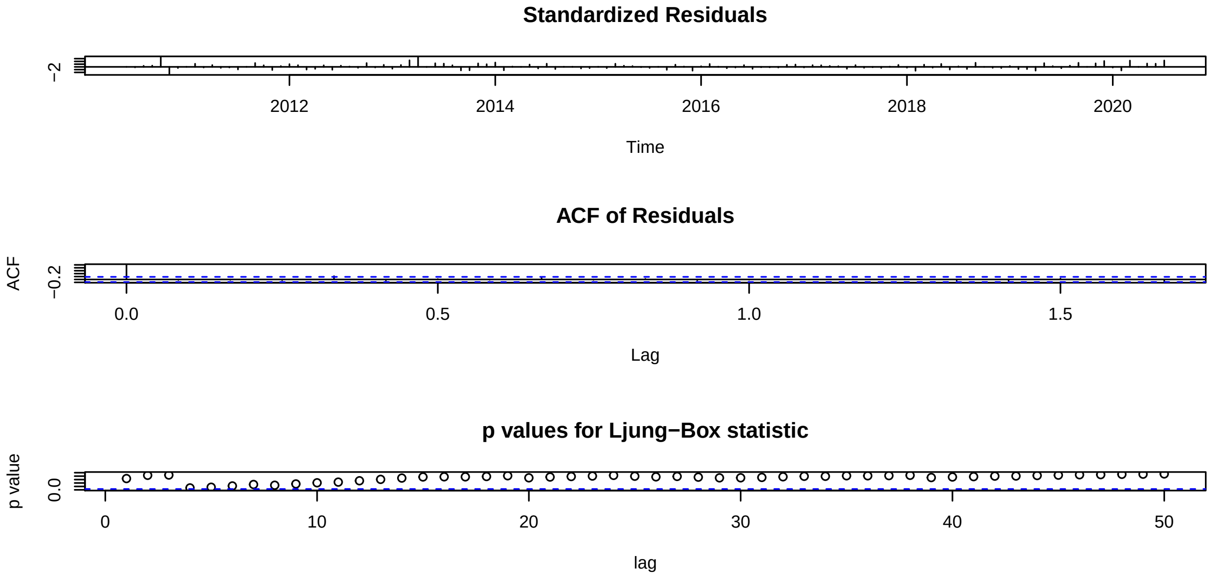
```
##
##  Box-Pierce test
##
## data:  fit2$residuals
## X-squared = 0.2, df = 1, p-value = 0.7
```

Al tener un $p-value = 0.7$, no rechazamos H_0 y se puede concluir que los datos son independientes.

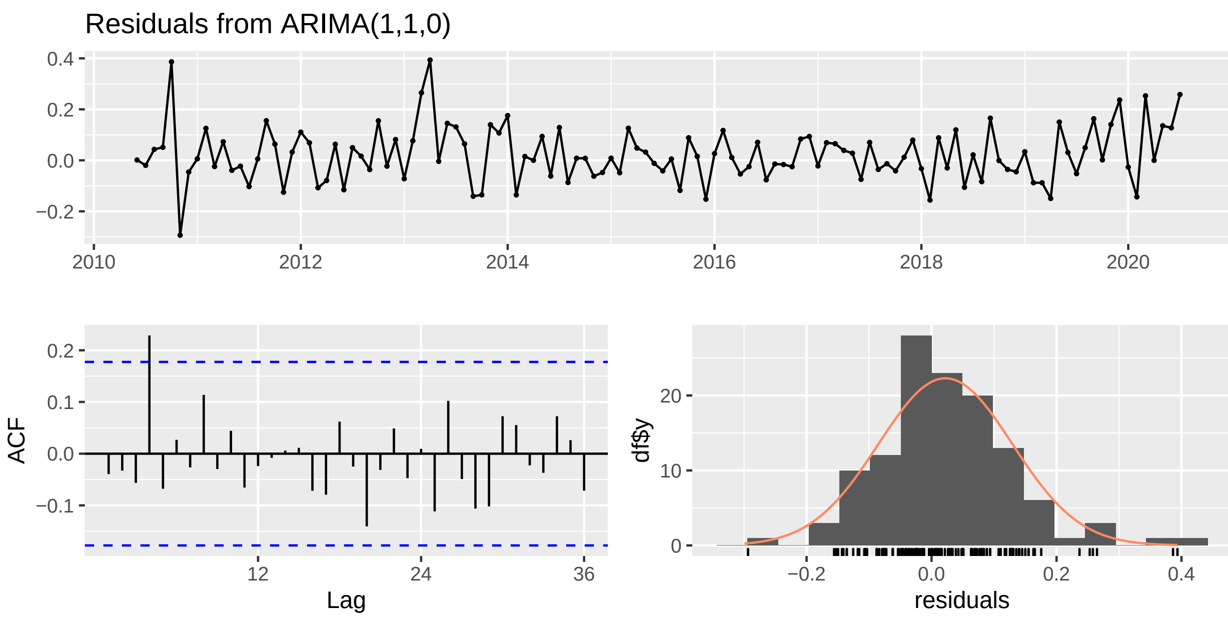
Ahora, para el lag $= 4$:

```
##
##  Box-Pierce test
##
## data:  fit2$residuals
## X-squared = 7, df = 4, p-value = 0.1
```

Tiene un $p-value = 0.1$, por lo que no hay problema y así se puede decir que los datos son independientes.



Se aprecia que por el lag 4 y 5 hay un pequeño problema.




```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,1,0)
## Q* = 17, df = 23, p-value = 0.8
##
## Model df: 1.   Total lags used: 24
```

Con un $p - value = 0.8$, si pasa la prueba Ljung Box.

Por lo que si son datos independientes.

Comparativa.

```
##   DATOS     MODELO       PARÁMETROS MEDIA CERO VARIANZA CONSTANTE NORMALIDAD
##   LOGARITMO ARIMA(1,1,1) 2          Sí         Sí                 No
##   GUERRERO  ARIMA(1,1,0) 1          No         Sí                 No
##   INDEPENDENCIA
##   Sí
##   Sí
```

En este caso se tiene una disyuntiva para elegir el mejor modelo:

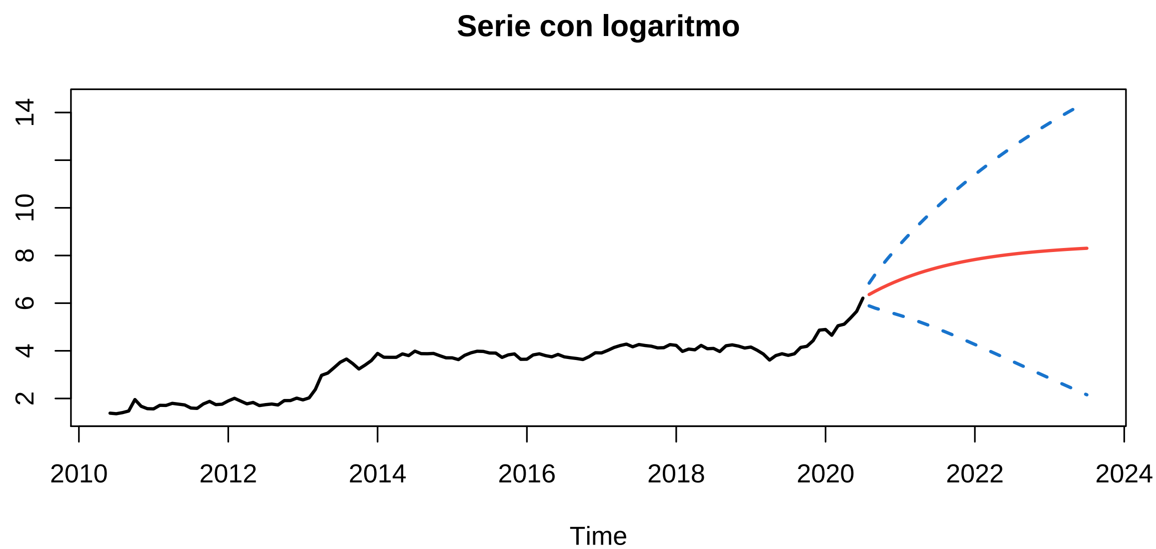
- El modelo ARIMA(1,1,1) tiene dos parámetros, pero tiene media cero.
- Por otro lado, el modelo ARIMA(1,1,0) solo tiene un parámetro, pero no tiene media cero.

Pero al tomar como criterio la cantidad de parámetros, el mejor modelo es el ARIMA(1,1,0).

5. Realizar pronósticos.

Para Logaritmo tenemos que la predicción a dos años es:

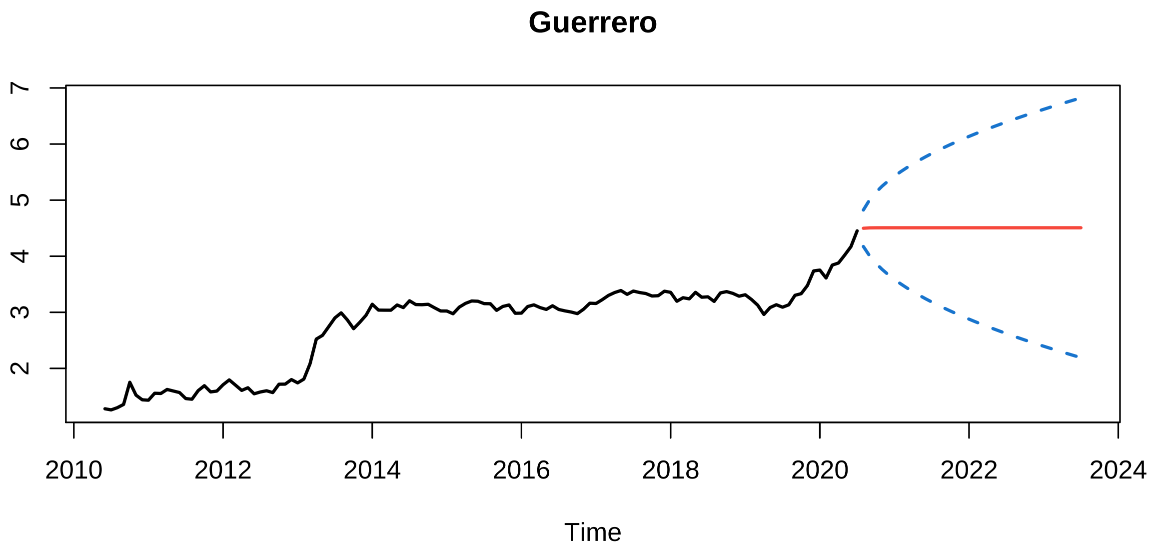
Para los datos transformados:



Para los datos originales:

Para Guerrero tenemos que la predicción a dos años es:

Para los datos transformados:



Para los datos originales:

Conclusiones.

En conclusión, las predicciones no son las mejores, pero no son tan alejadas considerando que son años de mucha incertidumbre.

Todavia se puede mejorar el modelo, quizá a un modelo GARCH para series financieras.