

LABORATORIO 2

Serie de Tiempo

Análisis exploratorio

Son 25 variables las cuales todas son numéricas. Las primeras 2 Año y Mes son categóricas mientras que las otras 23 son variables continuas por lo que se le harán pruebas de distribución.

Anio	Mes	GLP	GasAviacion	GasSuperior
Min. : 1.00	Min. : 1.000	Min. :100562	Min. : 0	Min. :193485
1st Qu.: 5.00	1st Qu.: 3.000	1st Qu.:206347	1st Qu.: 426	1st Qu.:354383
Median :10.00	Median : 6.000	Median :352909	Median : 1759	Median :434578
Mean :10.13	Mean : 6.442	Mean :362948	Mean : 3050	Mean :455769
3rd Qu.:15.00	3rd Qu.: 9.000	3rd Qu.:502568	3rd Qu.: 3291	3rd Qu.:545364
Max. :20.00	Max. :12.000	Max. :935987	Max. :27979	Max. :809640

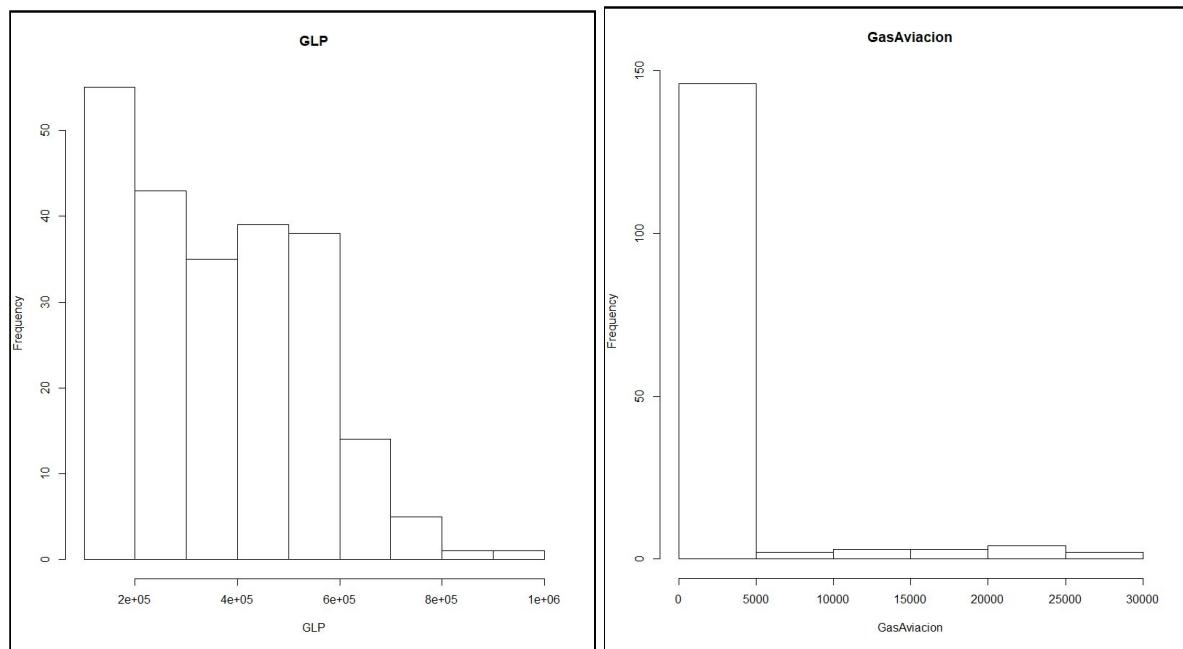
GasRegular	Kerosina	rTurboJet	Diesel	DieselLS
Min. : 81015	Min. : 1000	Min. : 23906	Min. : 229765	Min. : 696963
1st Qu.:189887	1st Qu.: 33762	1st Qu.: 57268	1st Qu.: 620918	1st Qu.: 897927
Median :271554	Median : 51860	Median : 76542	Median : 767285	Median :1056569
Mean :324593	Mean : 54898	Mean : 81523	Mean : 782290	Mean :1067625
3rd Qu.:437175	3rd Qu.: 67657	3rd Qu.:102974	3rd Qu.: 904976	3rd Qu.:1202102
Max. :896841	Max. :184094	Max. :158719	Max. :1595699	Max. :1592580

DieselULS	Bunker	Asfalto	PetCoke	AceitesLub
Min. : 1991	Min. : 7342	Min. : 74.88	Min. : 0	Min. :11558
1st Qu.:10936	1st Qu.: 193047	1st Qu.: 3114.36	1st Qu.:148562	1st Qu.:16325
Median :25967	Median : 307344	Median : 6093.80	Median :161718	Median :17980
Mean :23156	Mean : 320033	Mean : 7469.18	Mean :196651	Mean :18462
3rd Qu.:35768	3rd Qu.: 423443	3rd Qu.:10375.20	3rd Qu.:181821	3rd Qu.:20840
Max. :39867	Max. :1051764	Max. :48364.50	Max. :849463	Max. :27097

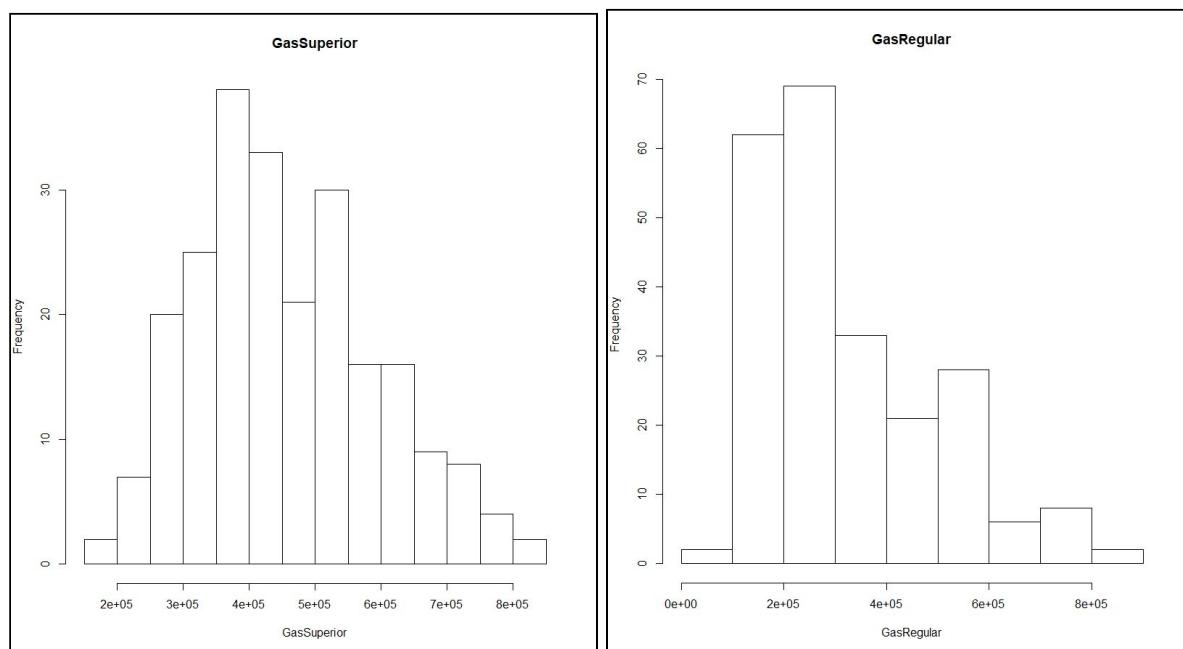
GrasasLub	Solventes	Naftas	Ceras	Butano
Min. :115.8	Min. : 3681	Min. :340.1	Min. : 51.12	Min. : 0.37
1st Qu.:169.9	1st Qu.: 5688	1st Qu.:341.9	1st Qu.:219.72	1st Qu.: 57.47
Median :313.8	Median : 7671	Median :346.7	Median :404.40	Median : 72.29
Mean :351.9	Mean : 7785	Mean :475.6	Mean :352.01	Mean : 68.68
3rd Qu.:483.2	3rd Qu.: 9110	3rd Qu.:480.4	3rd Qu.:471.31	3rd Qu.: 79.64
Max. :996.5	Max. :13552	Max. :869.0	Max. :596.33	Max. :149.09

PetroleoReconst	MTBE	Orimulsion	MezclasOleosas	Total
Min. :356364	Min. : 0	Min. :249983	Min. : 166.7	Min. :1381787
1st Qu.:360619	1st Qu.: 0	1st Qu.:312030	1st Qu.:1647.6	1st Qu.:2036868
Median :367573	Median :11062	Median :315916	Median :2170.2	Median :2436512
Mean :489360	Mean : 8904	Mean :315469	Mean :1972.9	Mean :2514304
3rd Qu.:718114	3rd Qu.:13299	3rd Qu.:325083	3rd Qu.:2607.1	3rd Qu.:2912599
Max. :730957	Max. :19431	Max. :344685	Max. :3134.8	Max. :4322010

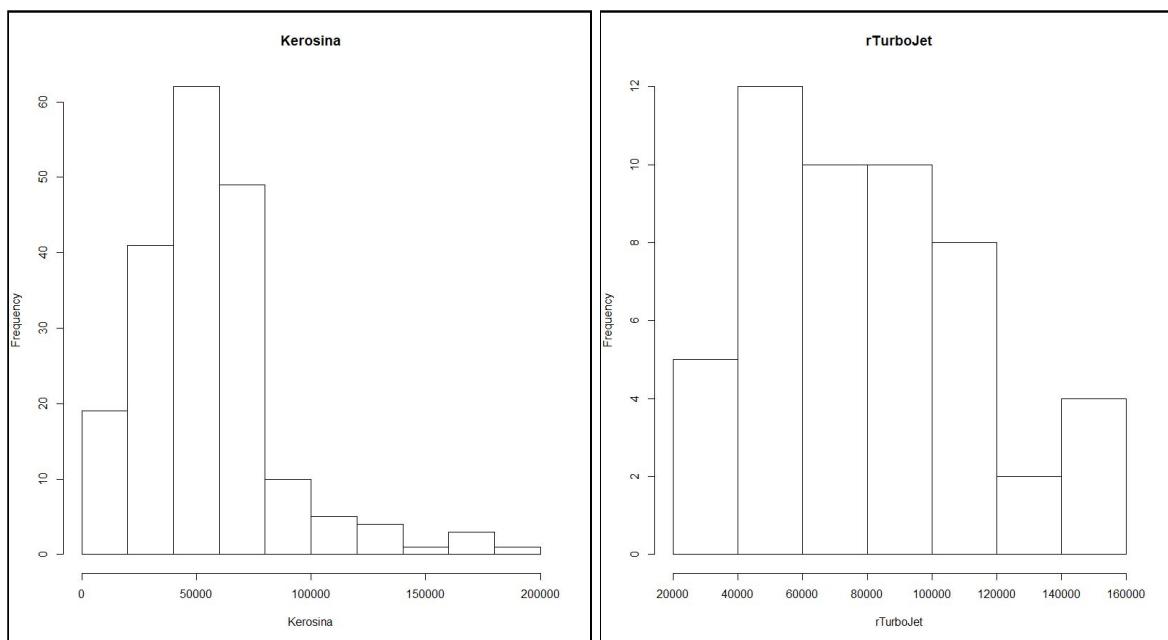
GLP tiene su pico en los 200,000 y está asimétrico hacia a la izquierda.
Gas Aviación está distribuido casi por completo de 0 a 5000.



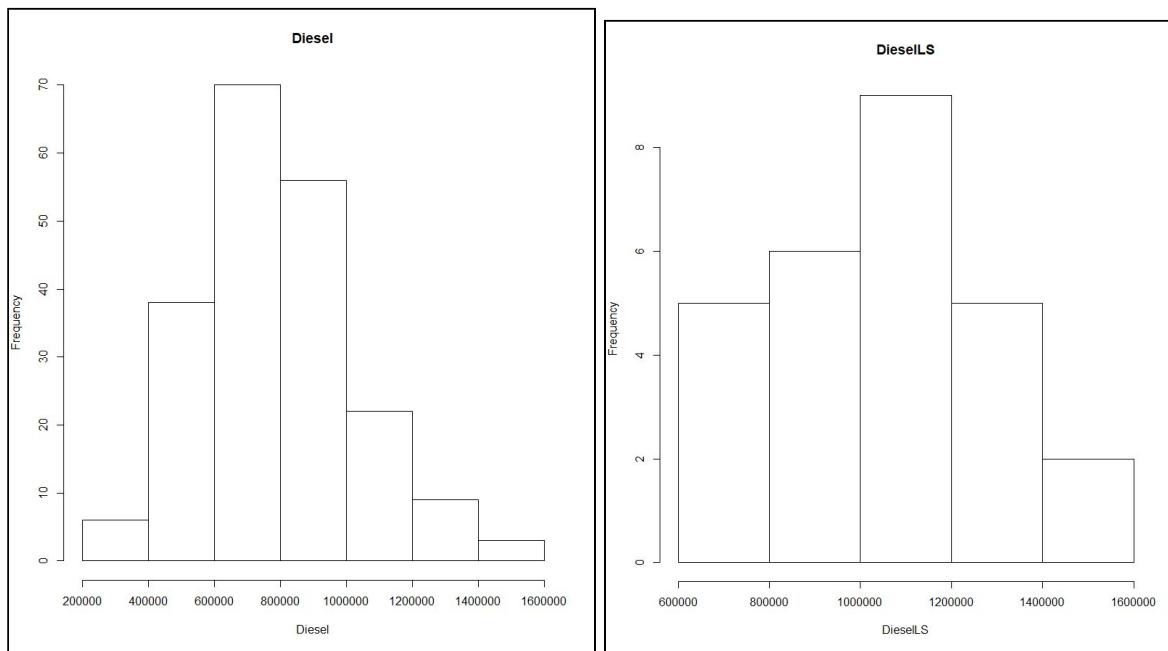
Gas superior está cerca de estar distribuido normalmente.
Gas regular está asimétrico hacia a la izquierda.



Kerosina tiene su pico en los 50,000 y está asimétrico hacia la izquierda.
Turbo Jet está distribuido casi por completo.

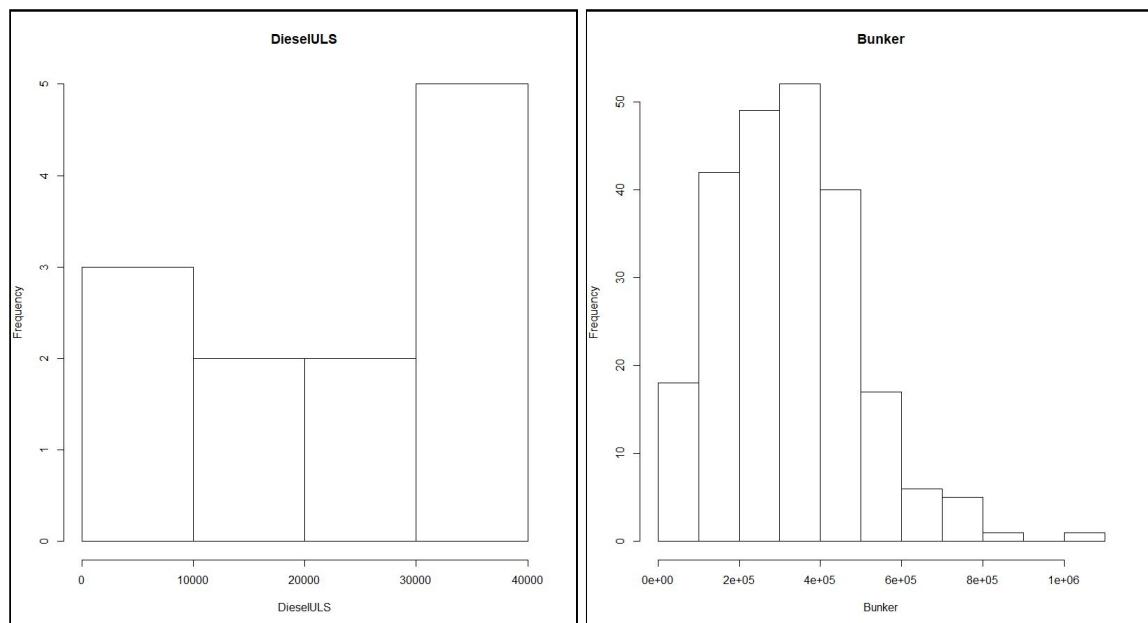


Diesel está distribuido normalmente.
Diesel LS está distribuido casi normalmente.



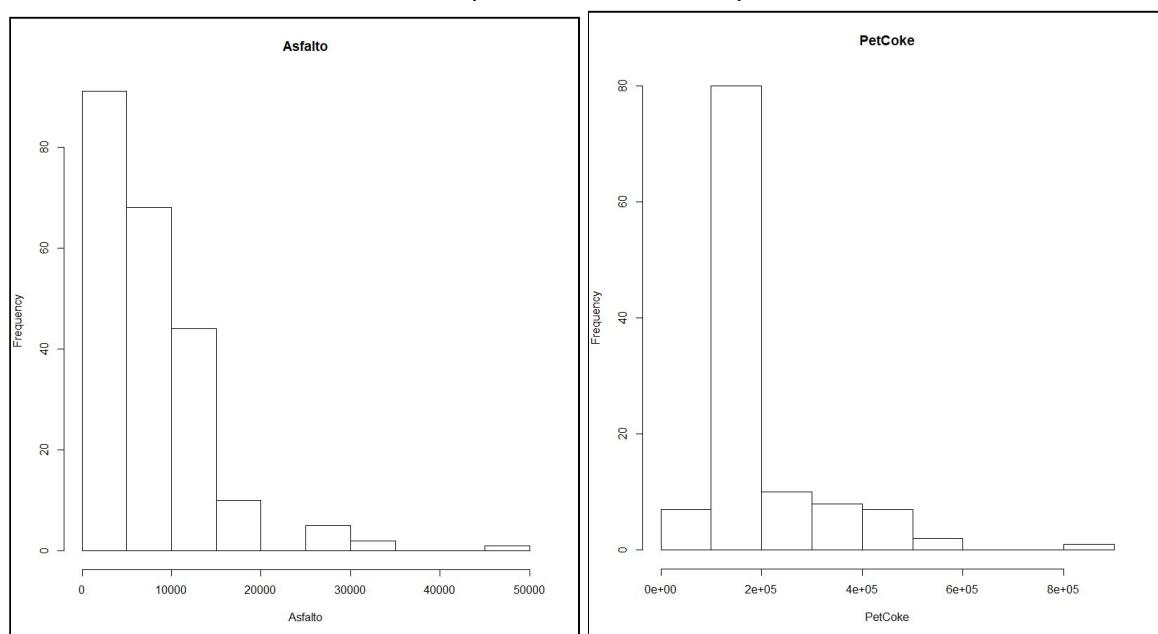
Diesel ULS no está distribuido normalmente.

Bunker tiene su pico en los 300,000 y está asimétrico hacia la izquierda con datos atípicos.



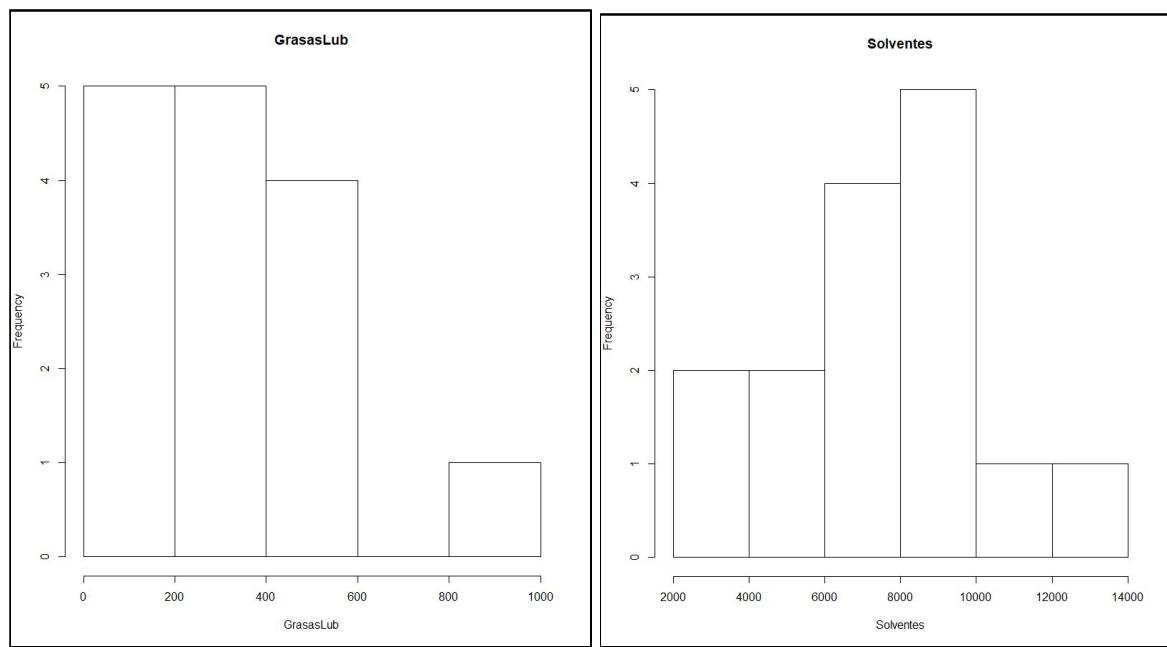
Asfalto está asimétrico hacia la izquierda con datos atípicos.

PetCoke está asimétrico hacia la izquierda con datos atípicos.



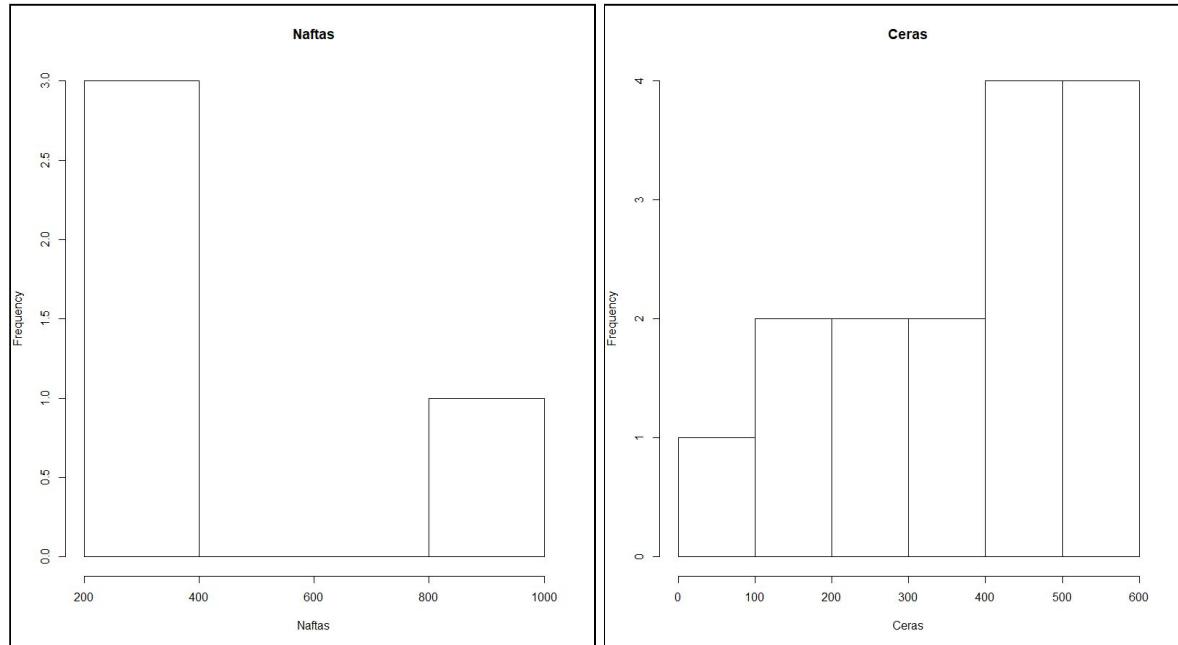
GrasasLub está asimétrico hacia la izquierda con datos atípicos.

Solventes está más o menos simétrico.

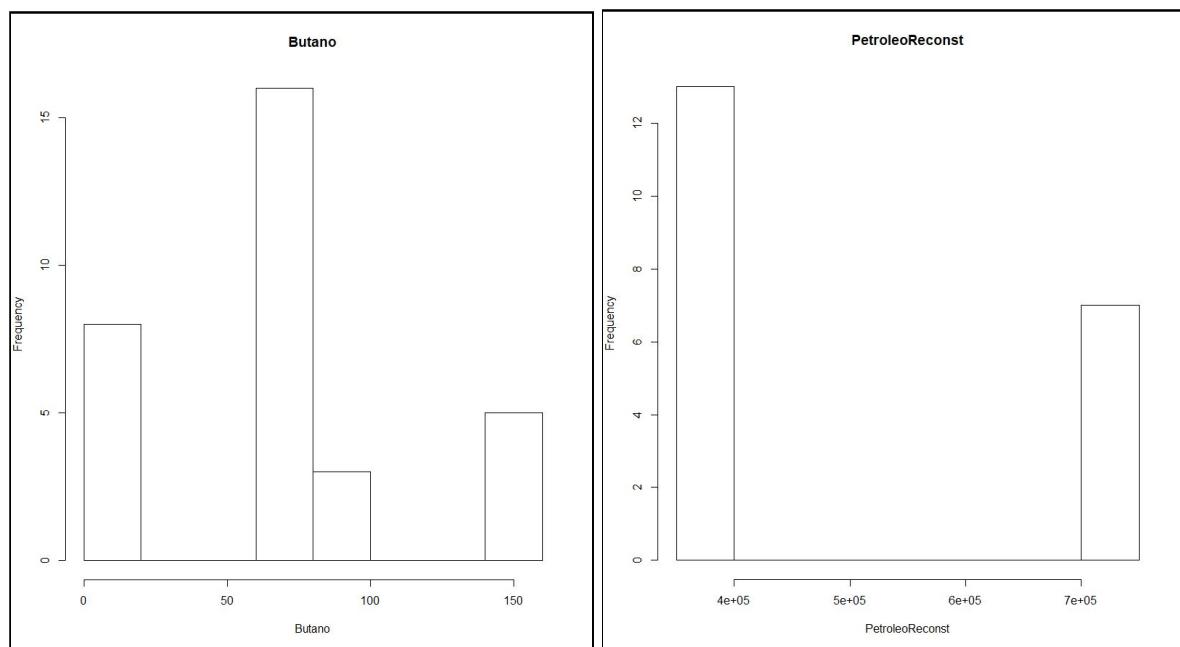


Naftas tiene tendencias asimétricas.

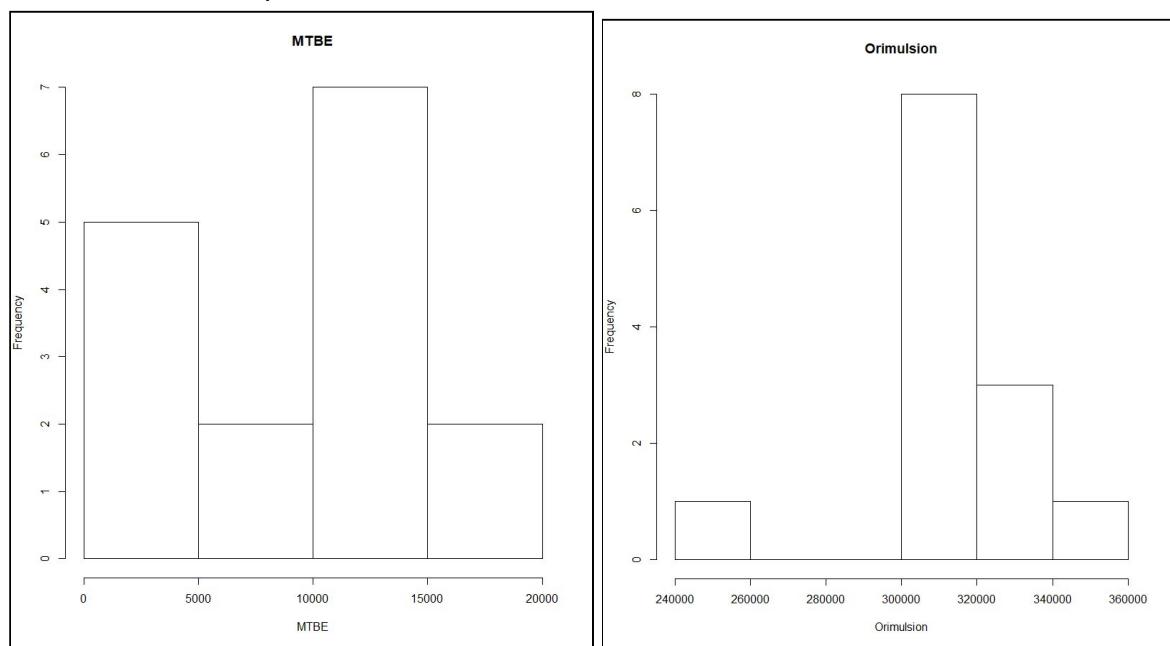
Ceras está asimétrico hacia la derecha con datos atípicos.



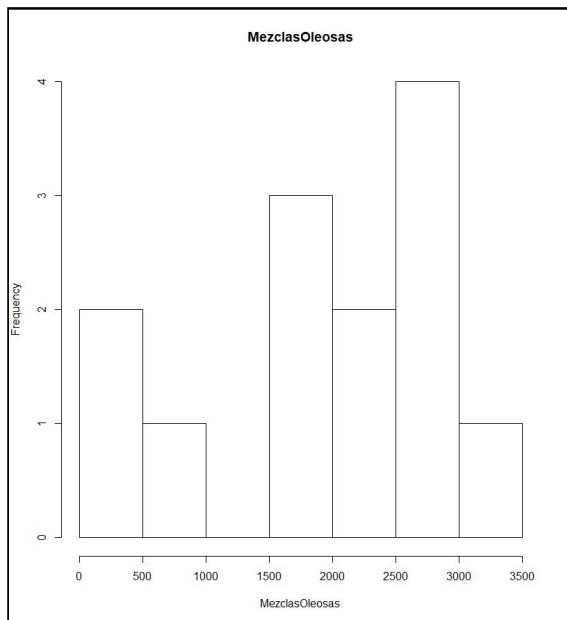
Butano tiene datos atípicos con pico en 50-100.
Petrolero Reconst se distribuye en tener o no tener



MTBE está asimétrico.
Orimulsión tiene su pico alrededor de 300000



Mezclas Oleosas tiene un par de datos atípicos.



Meses con más importaciones

Mes	Total
3	55905684
1	53673128
5	51293182
12	50003176
4	49328007
2	47540808
6	46480093
11	46407319
10	46278079
7	45728612
8	45495570
9	42670456

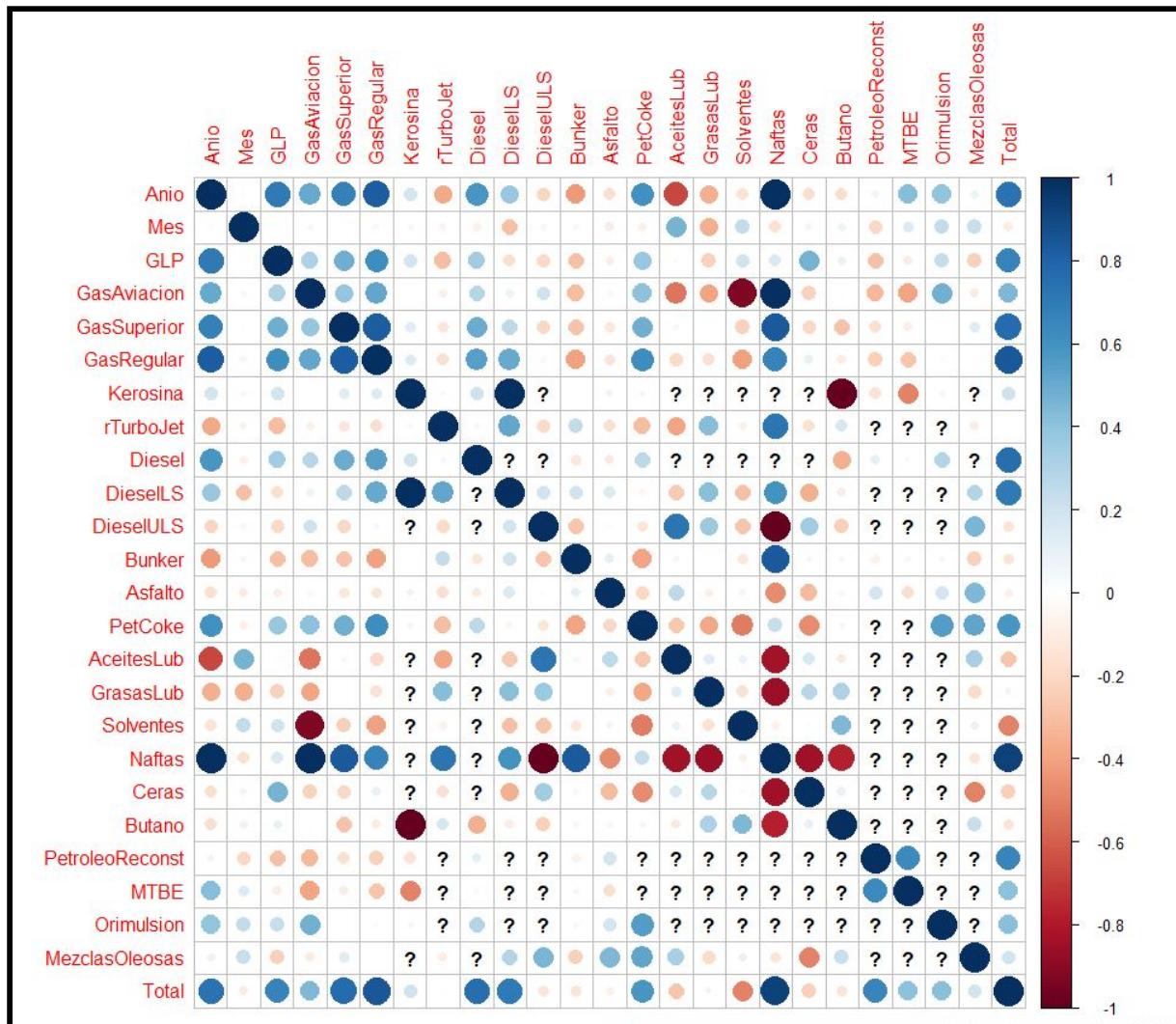
Como se puede apreciar en la tabla Marzo, Enero y Mayo son los meses con más importaciones siendo todos del primer semestre del año. Y los últimos 3 son Julio, Agosto y Septiembre siendo todos del segundo semestre del año un dato a tomar en cuenta.

Análisis de últimos años

Año	Total
2019	42024267
2016	38234309
2017	37456551
2018	37343170
2015	36837524
2014	31510637
2009	30292766
2007	29937287
2013	28426231
2005	27602641
2012	27473824
2011	27363726
2006	27007372
2010	26671154
2008	26016023
2004	25439778
2003	23725964
2002	23612569
2001	22382822
2020	11445499

Como se puede apreciar en la tabla a medida que los años pasan se está importando más combustible. Siendo el año con más importaciones el año pasado, este año se encuentra de último debido a que solo están registrados de momento los primeros 3 meses del año y aún así ya va por la mitad de todo lo que se importó en 2001.

Cruces de variables



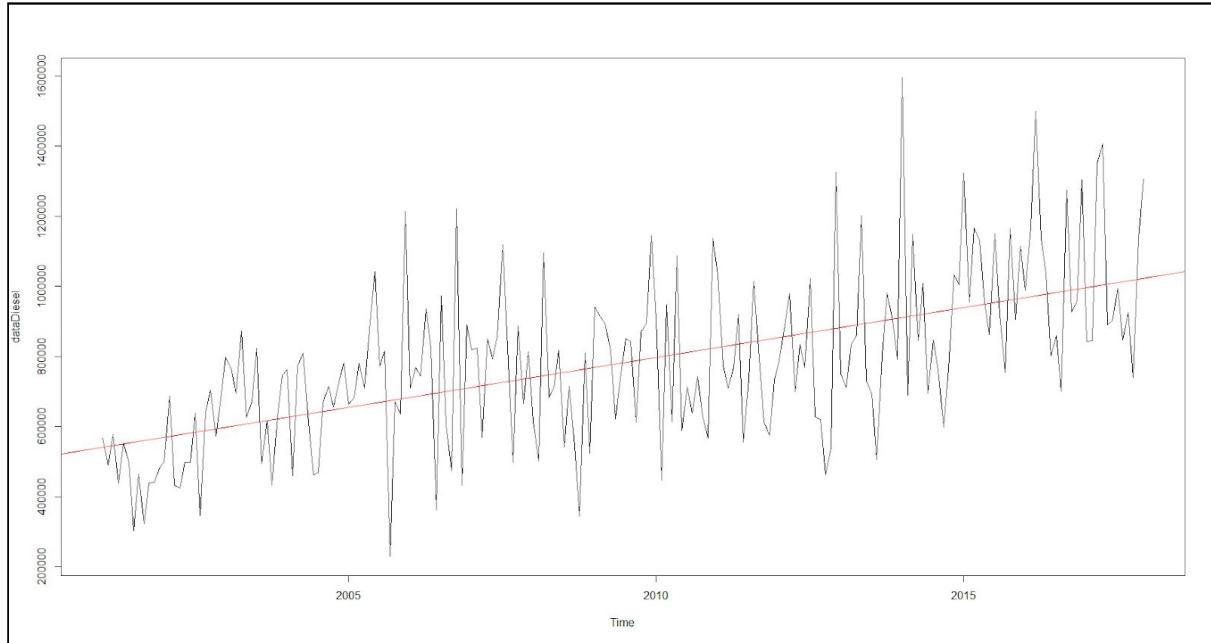
Existen variables que se relacionan entre sí como la Kerosina con el DieselLS, y a su vez que no se relacionan para nada como la Kerosina con el Butano, dándonos a entender que cuando se usa un material el otro no es necesario o no funcionan entre sí. Hay varios ejemplos así como las ceras y las naftas o los solventes con el gas Aviación.

Serie de tiempo de DIESEL

Inicio: 2001, 01

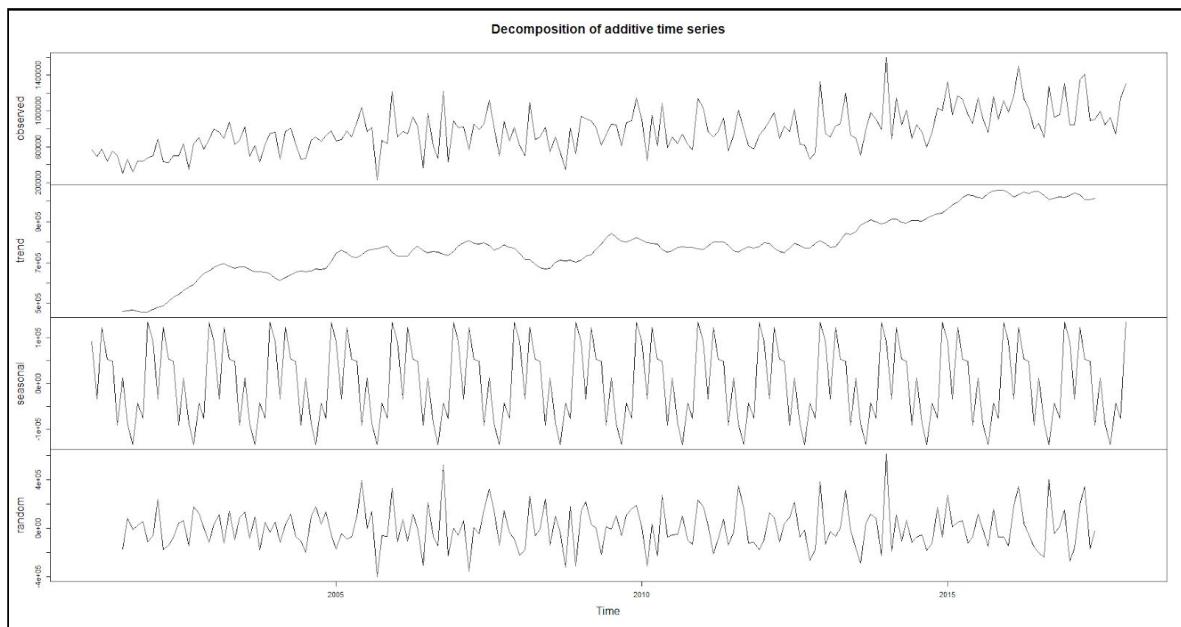
Fin: 2017, 12

Frecuencia: 12

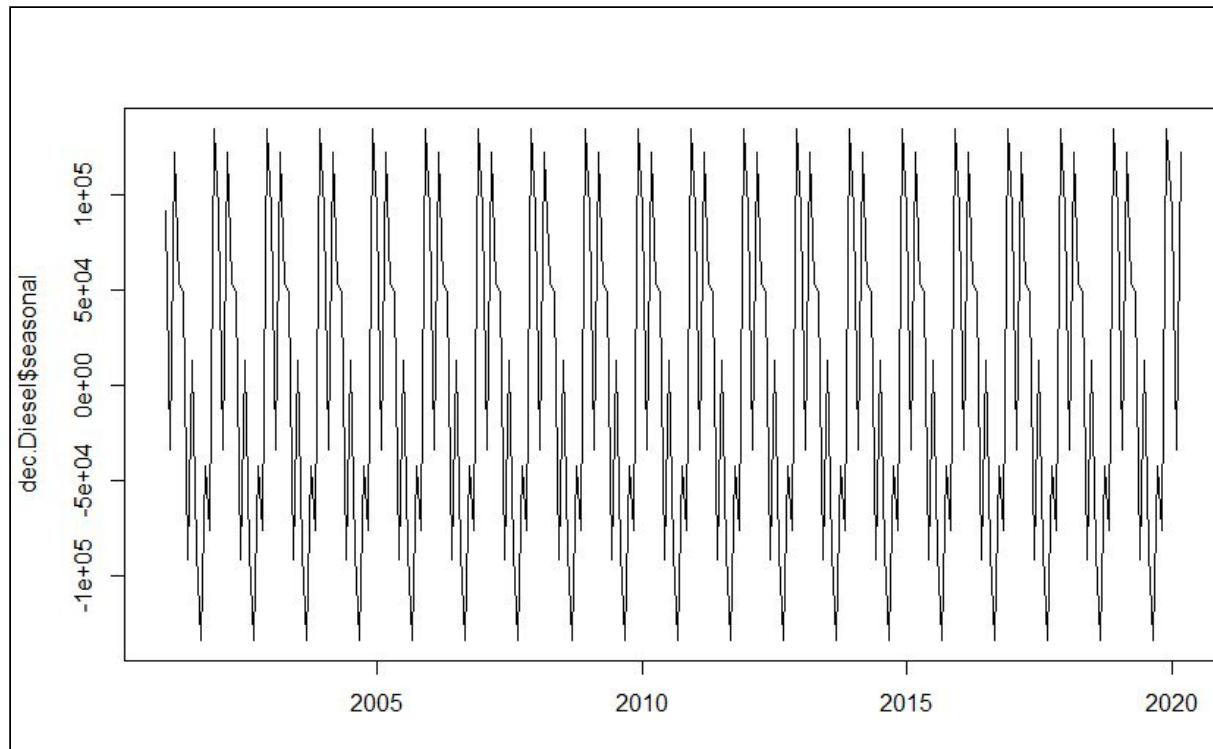


Se aprecia que a medida que pasan los años se importan más galones de Diesel como se había demostrado en el análisis exploratorio. La línea roja nos ayuda a ver como la media va en aumento

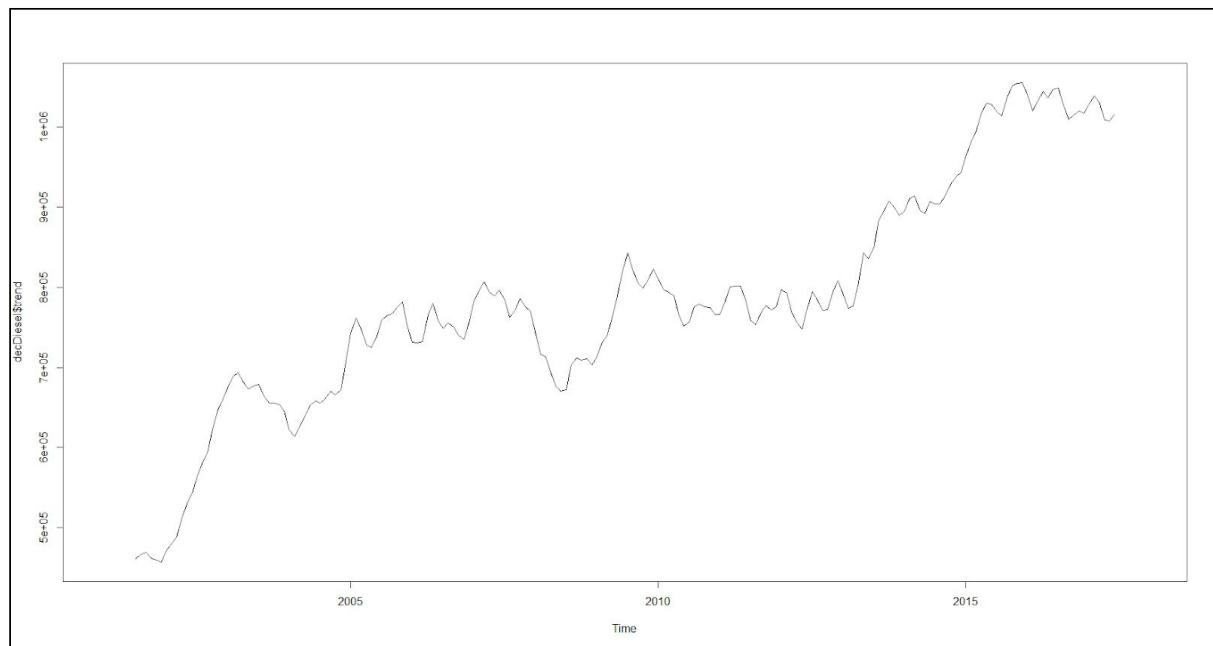
Descomposición en componentes de DIESEL



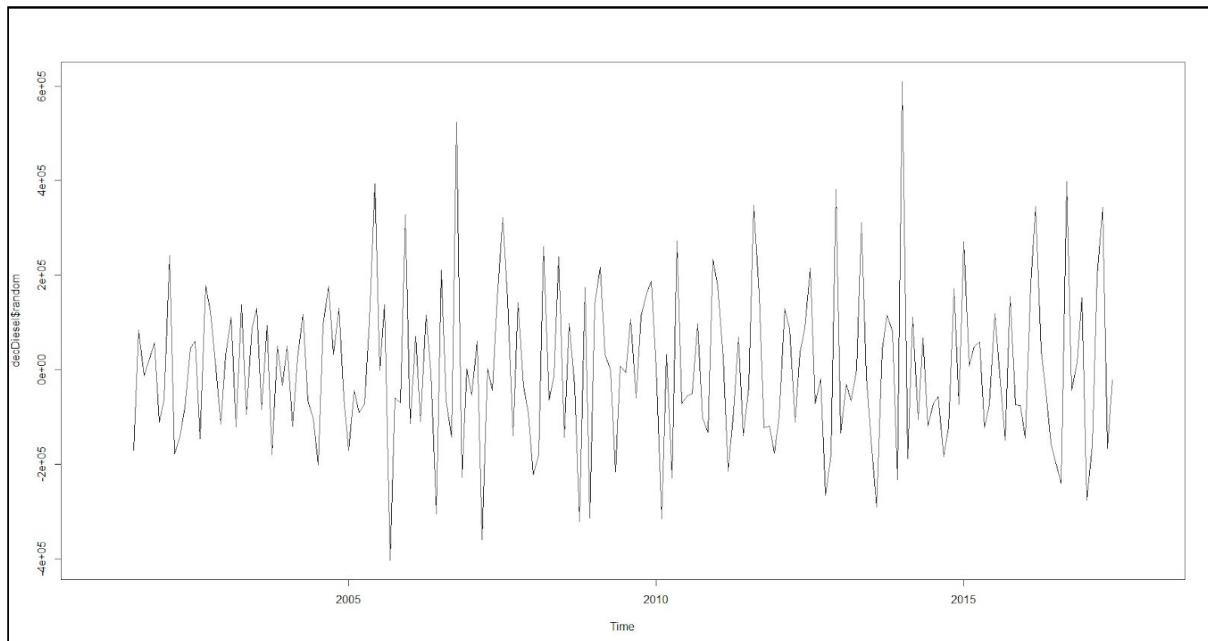
Seasonal



Trend



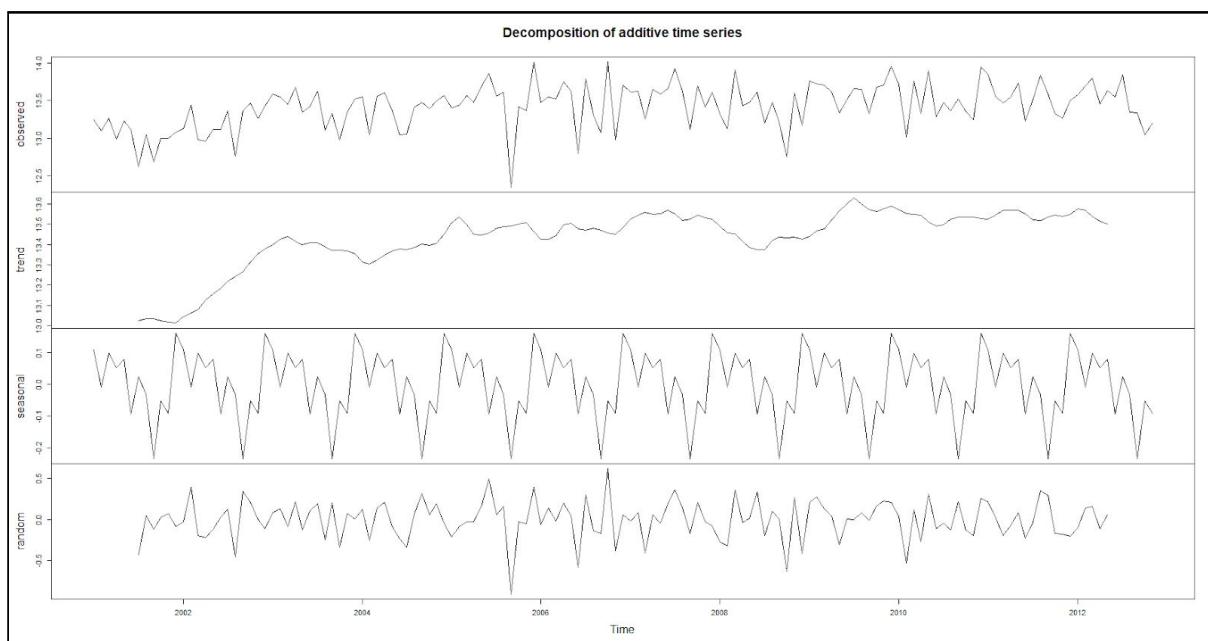
Random



Con estos diagramas podemos observar que está la tendencia a aumentar por lo que no es estacionaria en media y tampoco en varianza.

Si es necesario realizar una transformación logarítmica para intentar conseguir una varianza constante.

Descomposición en componentes de DIESEL (TRANSFORMADA)



Se logró hacer constante la serie en varianza., pero hay que verificar si es estacionaria en media. Si tiene raíces unitarias podemos decir que no es estacionaria en media y hay que aplicar procesos de diferenciación.

Dickey-Fuller Test ($p > 0.05$)

```
Title:  
Augmented Dickey-Fuller Test  
  
Test Results:  
PARAMETER:  
Lag Order: 1  
STATISTIC:  
Dickey-Fuller: -1.1276  
P VALUE:  
0.2546
```

Raíces Unitarias ($p > 0.05$)

```
Title:  
Augmented Dickey-Fuller Test  
  
Test Results:  
PARAMETER:  
Lag Order: 1  
STATISTIC:  
DF: -1.1276  
P VALUE:  
t: 0.2352  
n: 0.4491
```

Basados en las pruebas Dickey-Fuller Aumentada y UnitRoot, se obtuvo un valor $p > 0.05$, por lo cual la serie no es estacionaria en media. Luego, se realizó de nuevo la prueba Dickey-Fuller pero con una diferenciación, y se obtuvo un valor de $p < 0.05$. De esta manera se determinó que la serie no tiene raíces unitarias, siendo estacionaria en media.

Diferenciación, es necesario

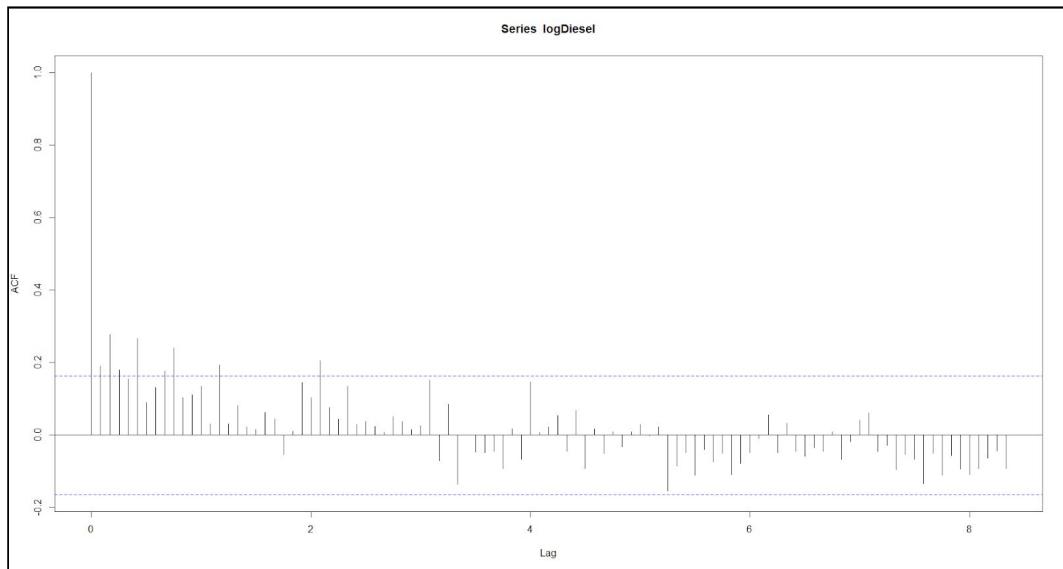
```
Title:  
Augmented Dickey-Fuller Test  
  
Test Results:  
PARAMETER:  
Lag Order: 1  
STATISTIC:  
DF: -14.209  
P VALUE:  
t: < 2.2e-16  
n: 0.00755
```

Basados en la prueba anterior, se elige $d=1$. Para elegir los valores de p y q , se realizaron gráficos de autocorrelación y autocorrelación parcial. Se determinaron los valores $q=3$ y $p=2$.

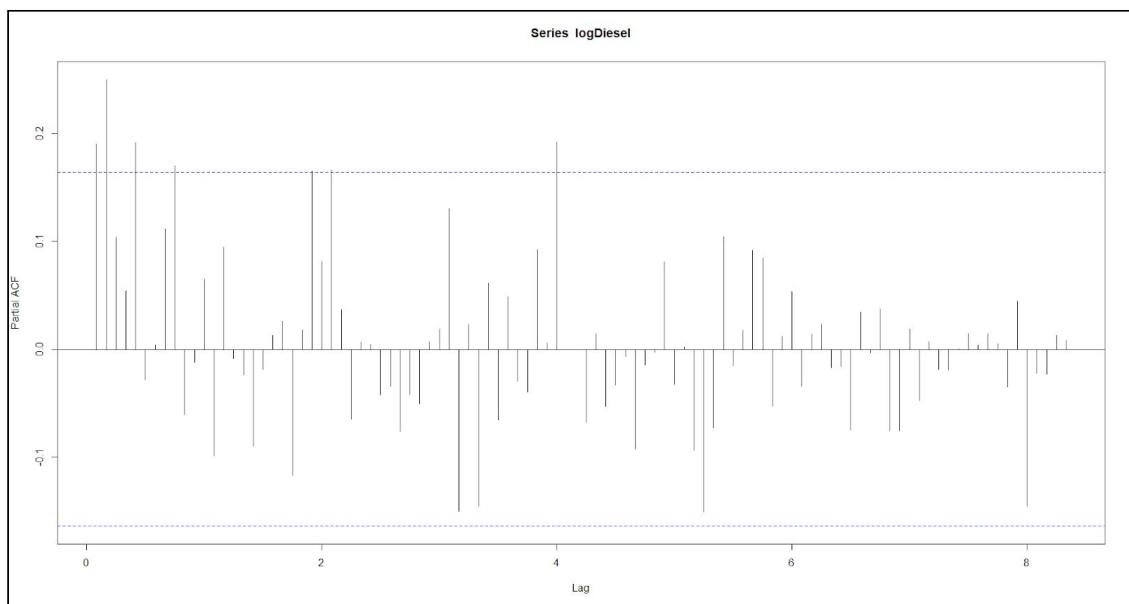
Gráficos de función

Intentar identificar parámetros p y q

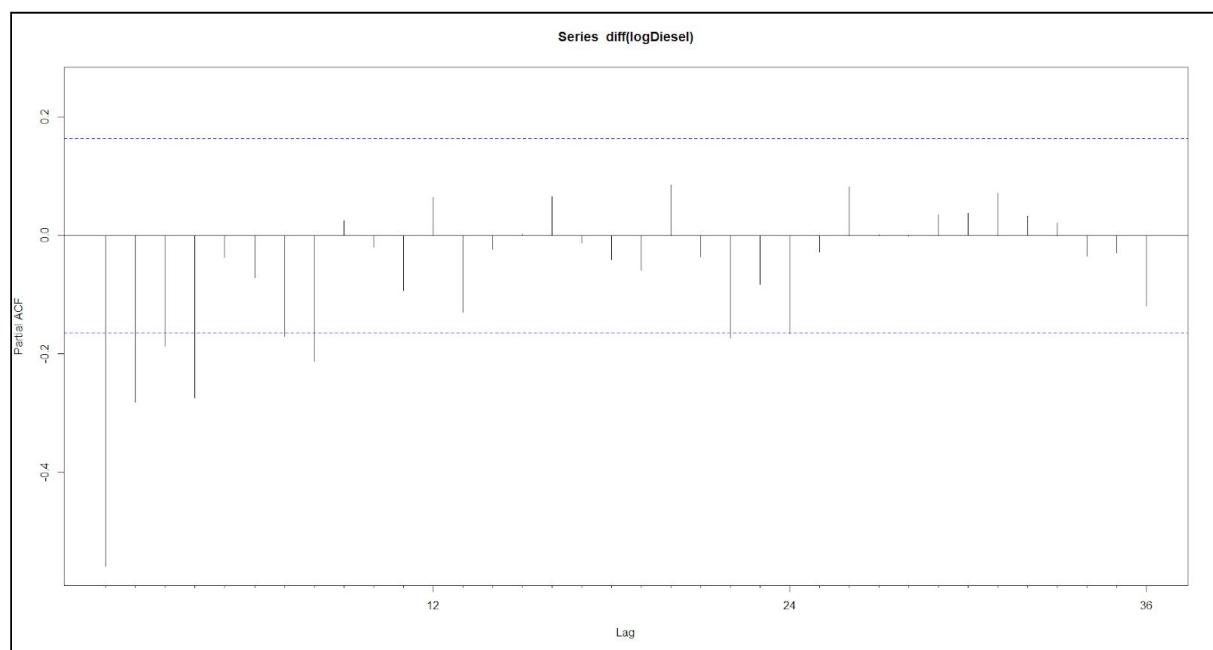
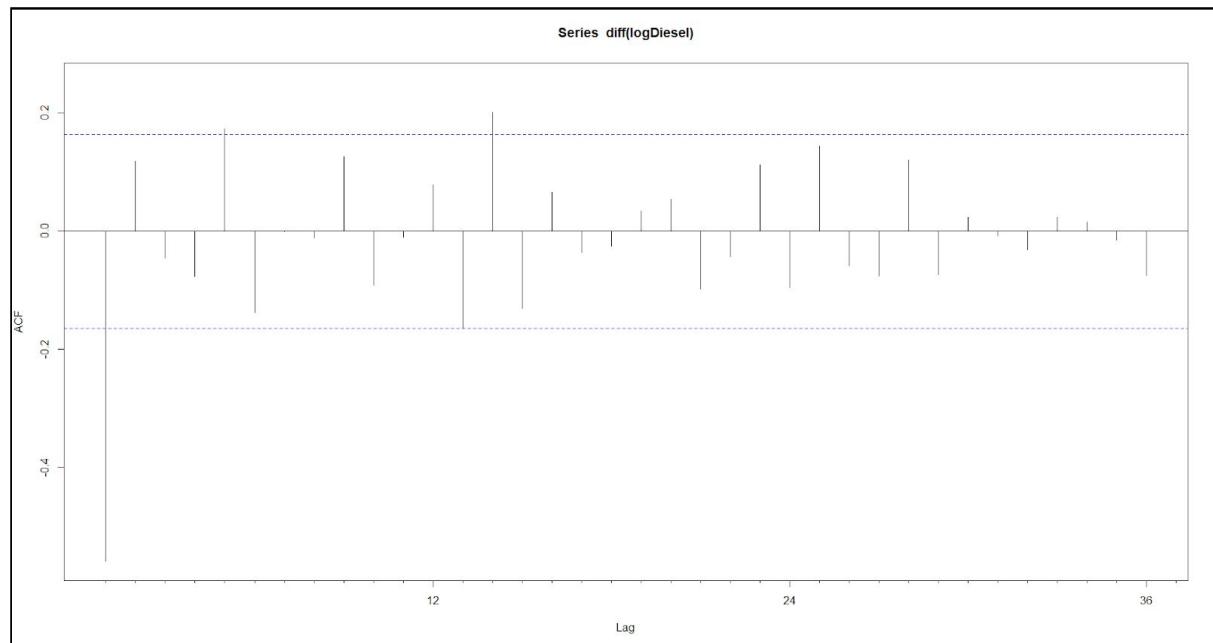
$q = 3$



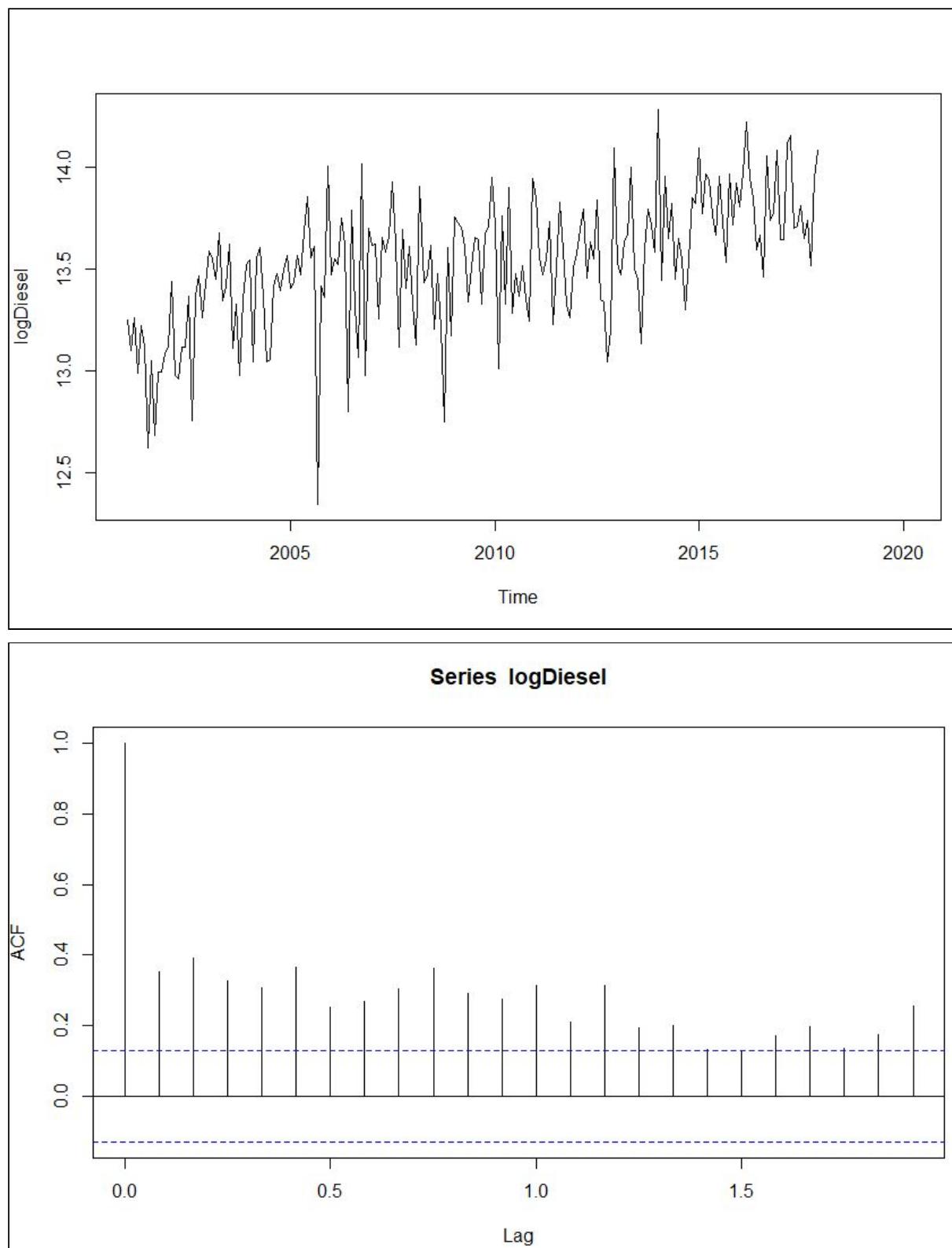
$p = 2$

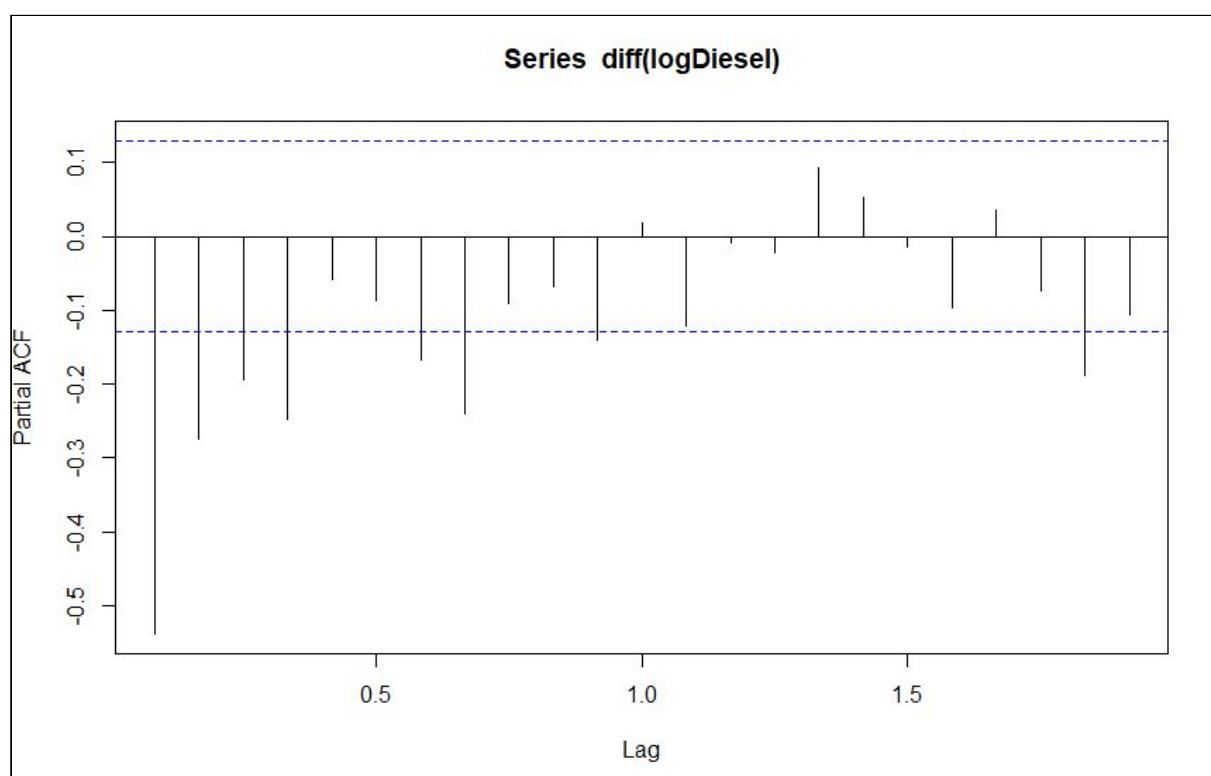
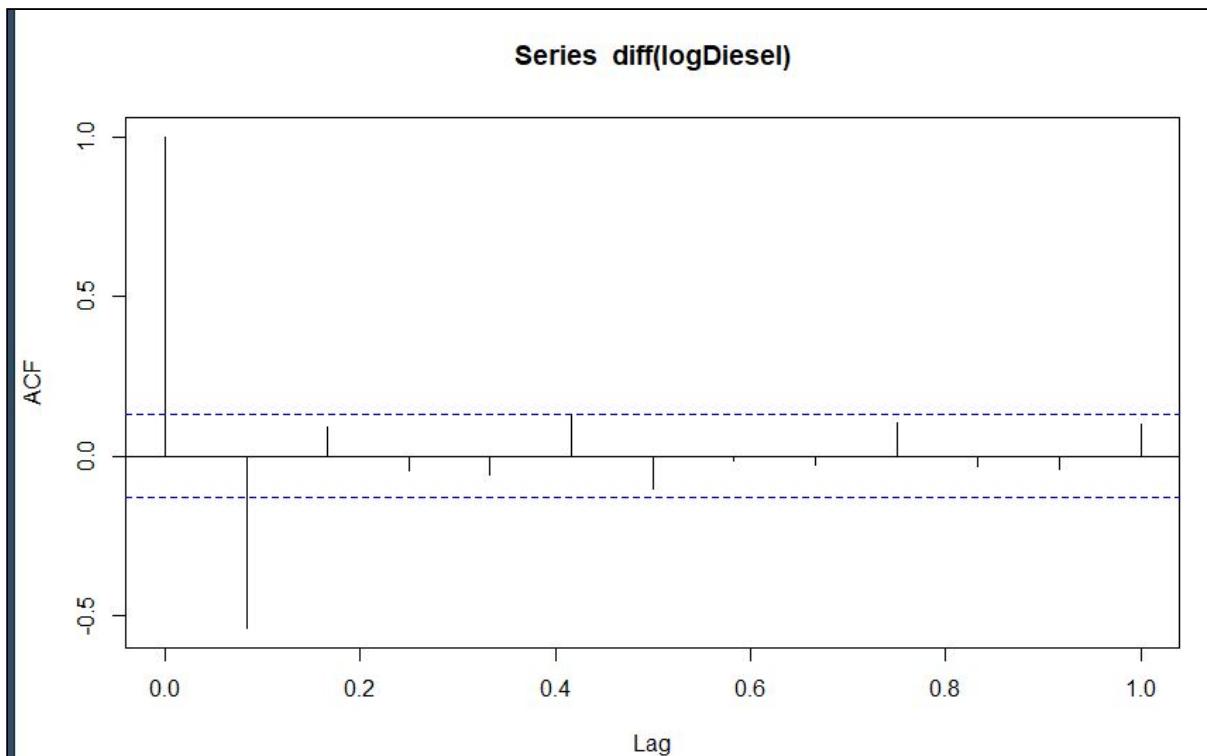


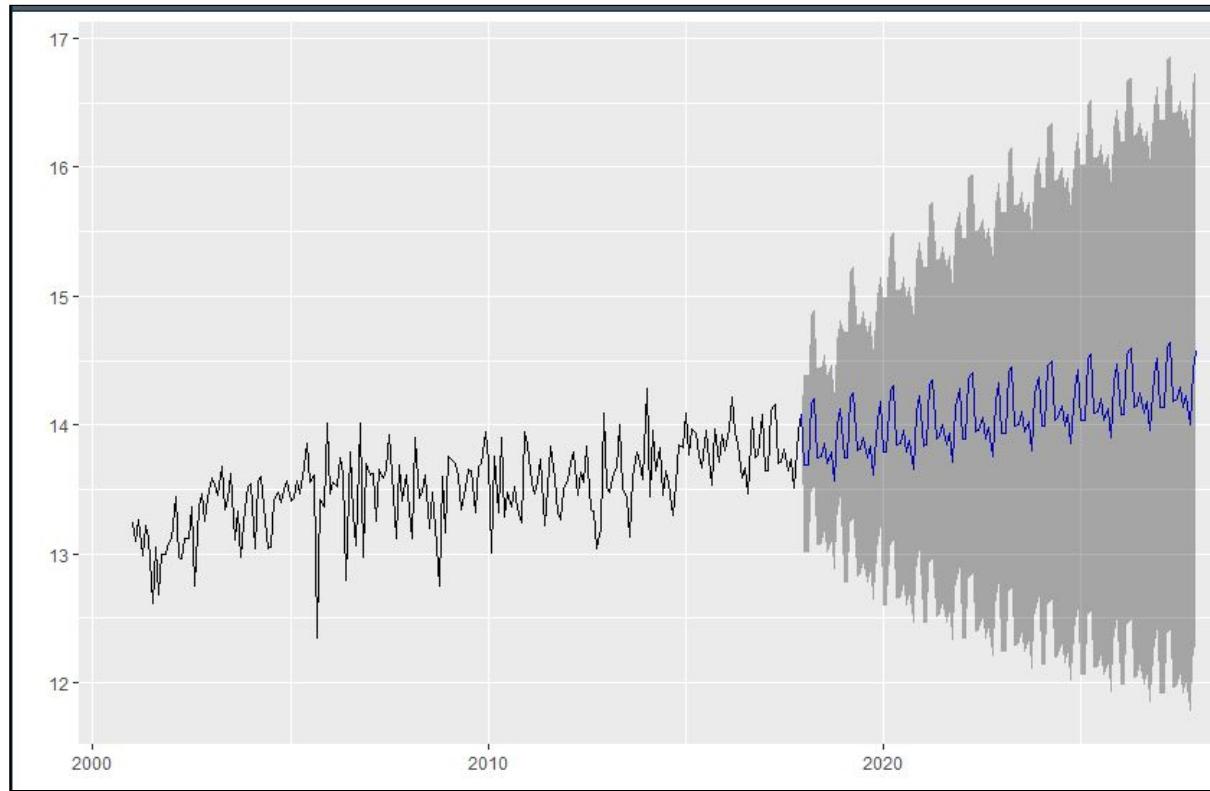
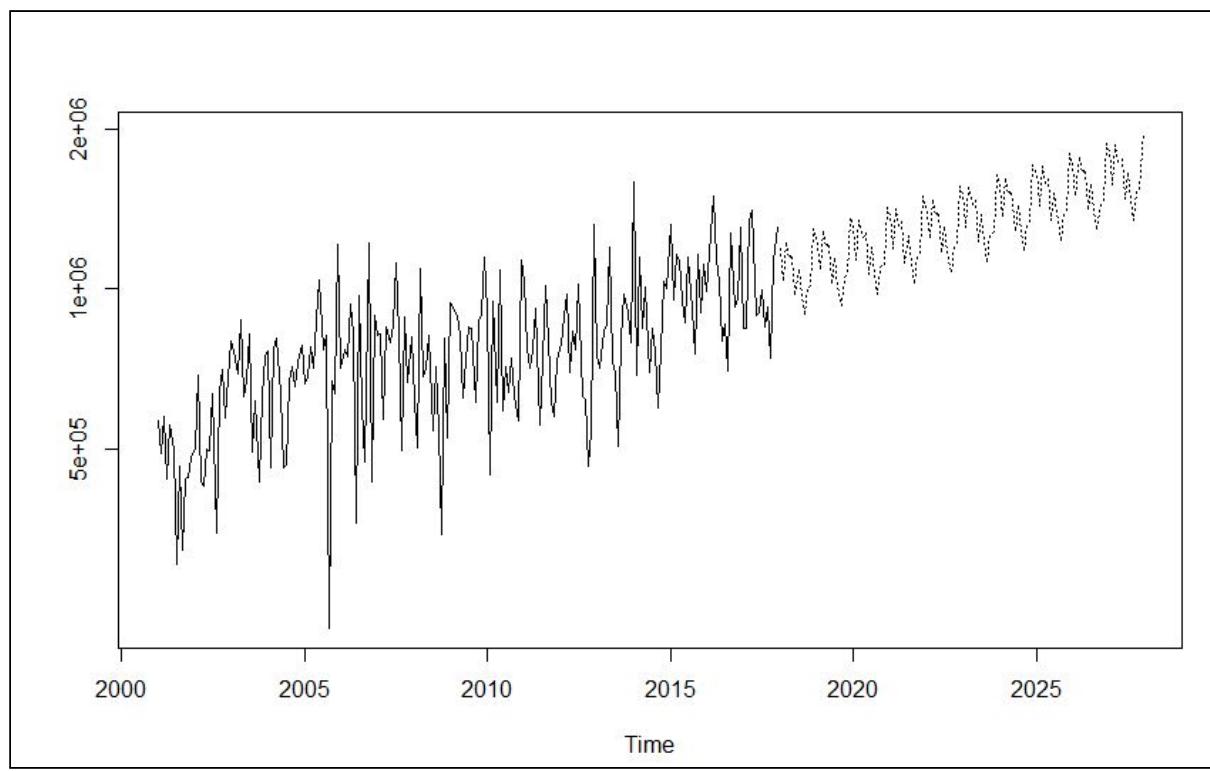
Se usará la serie estacionalizada.



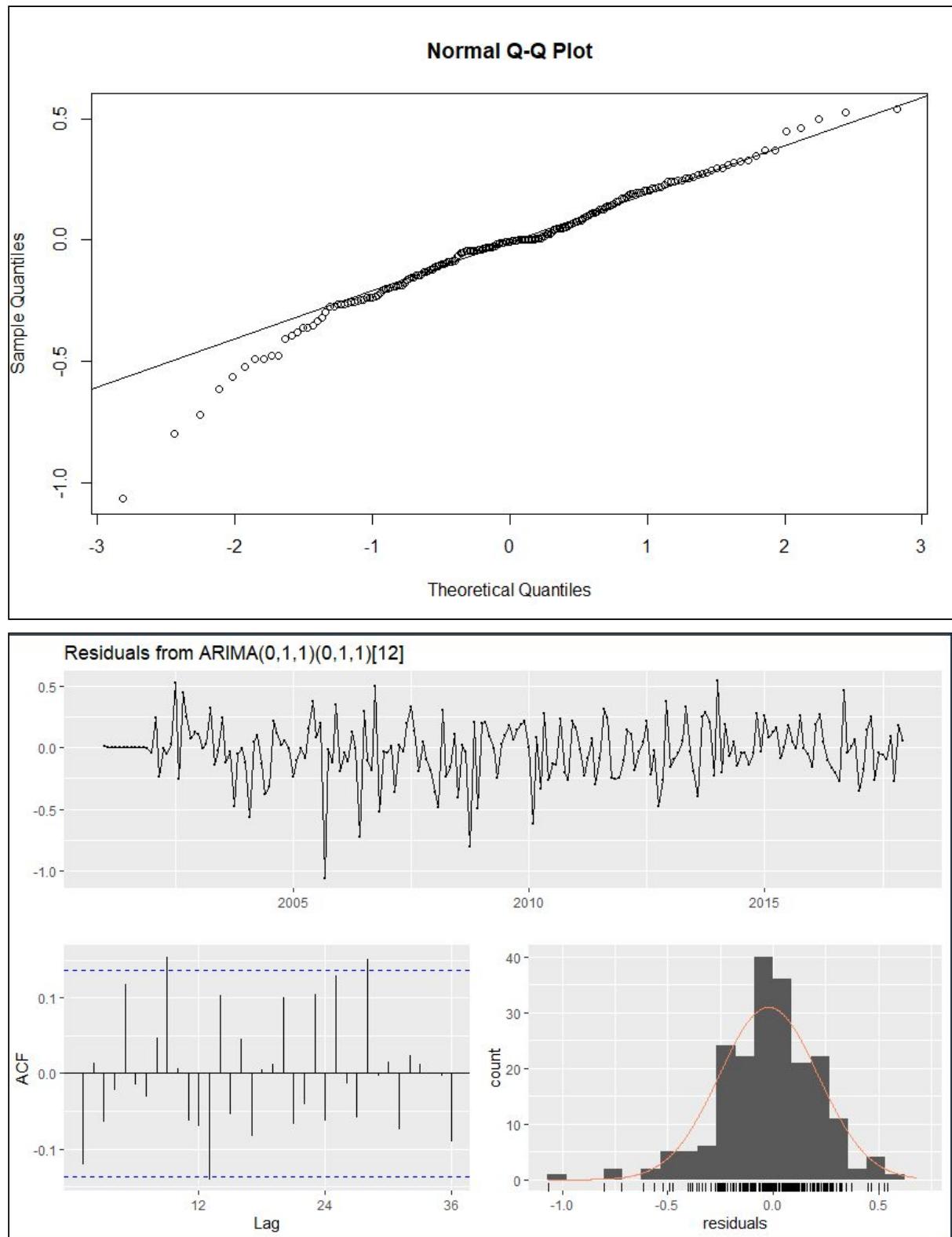
Arima





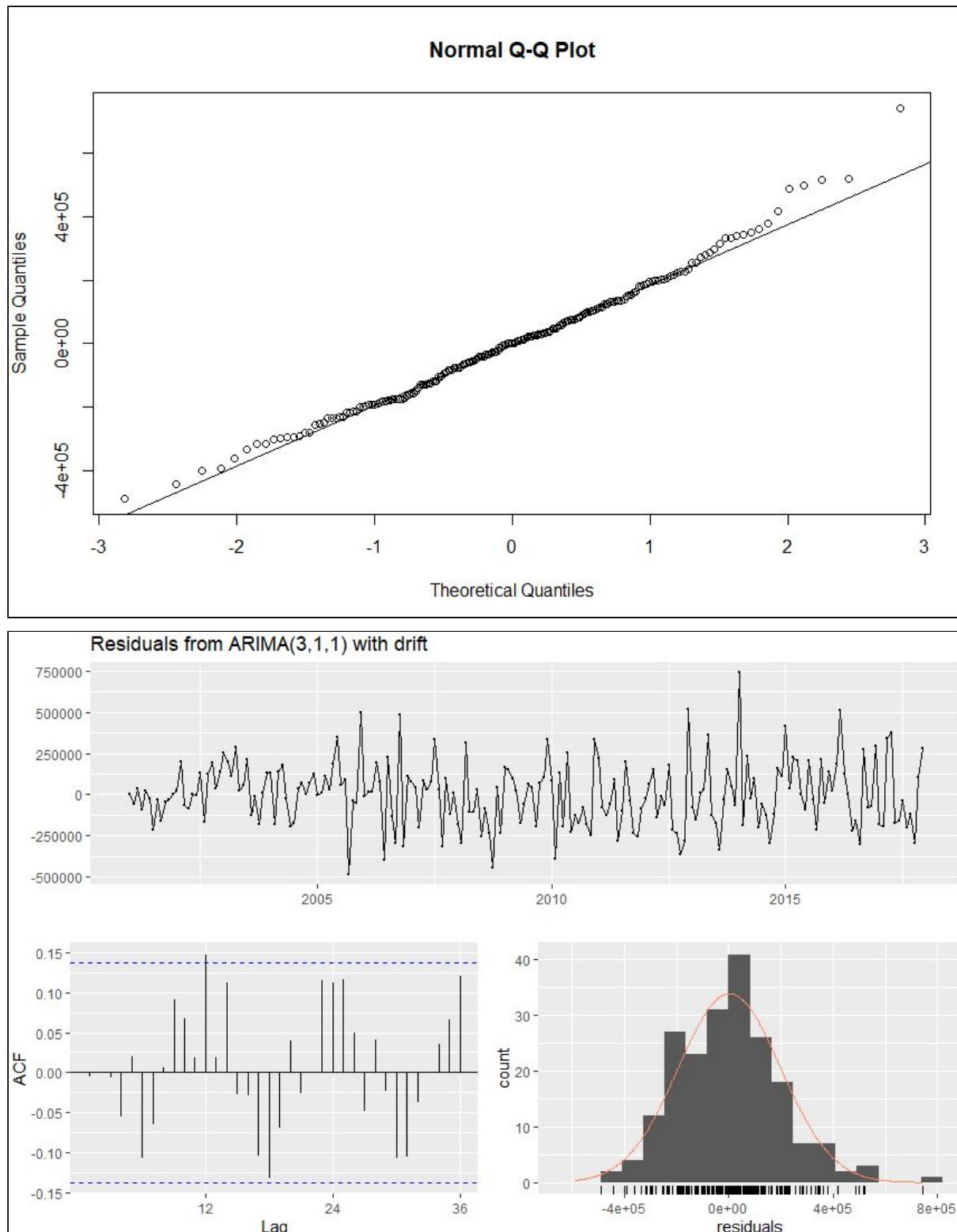


FIT NORMAL

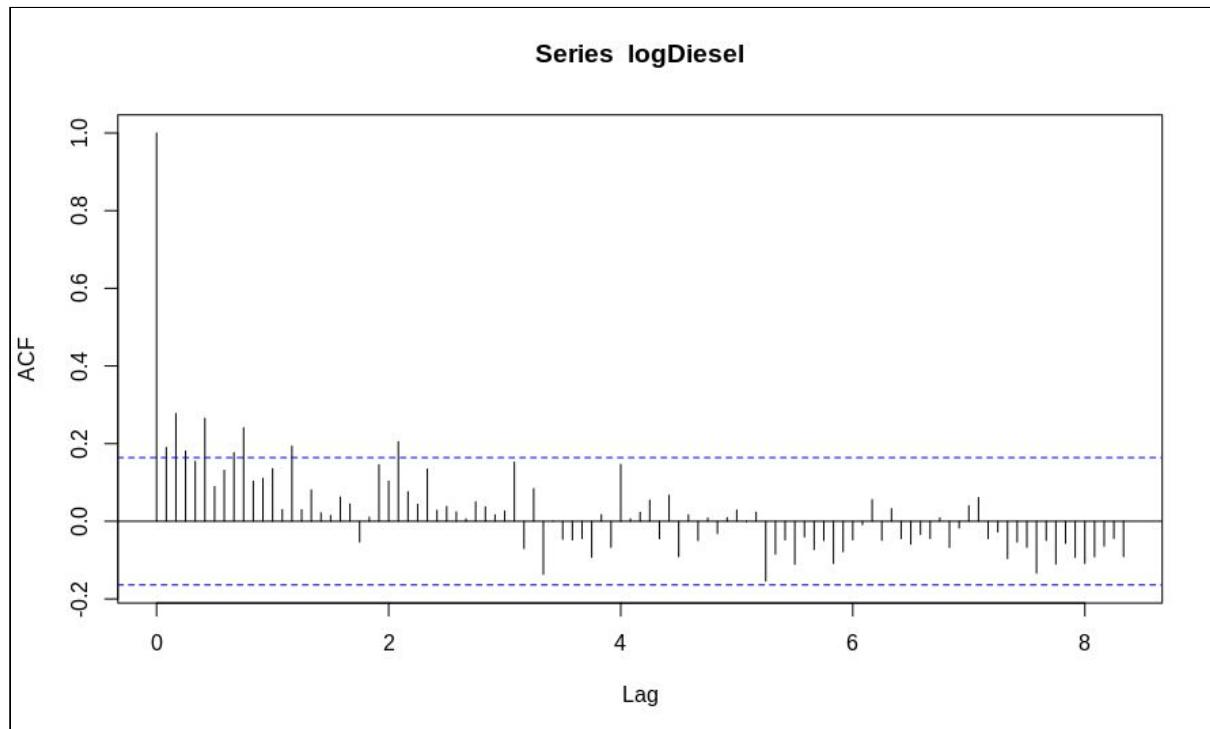
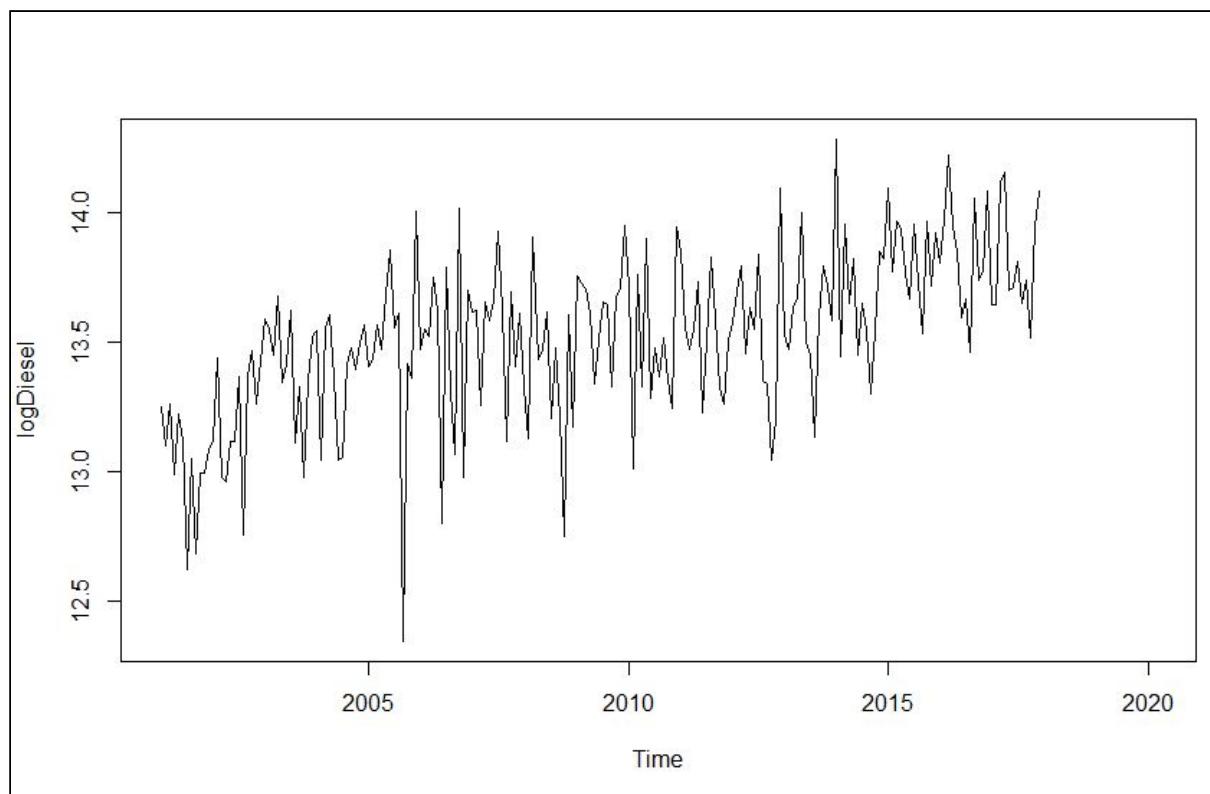


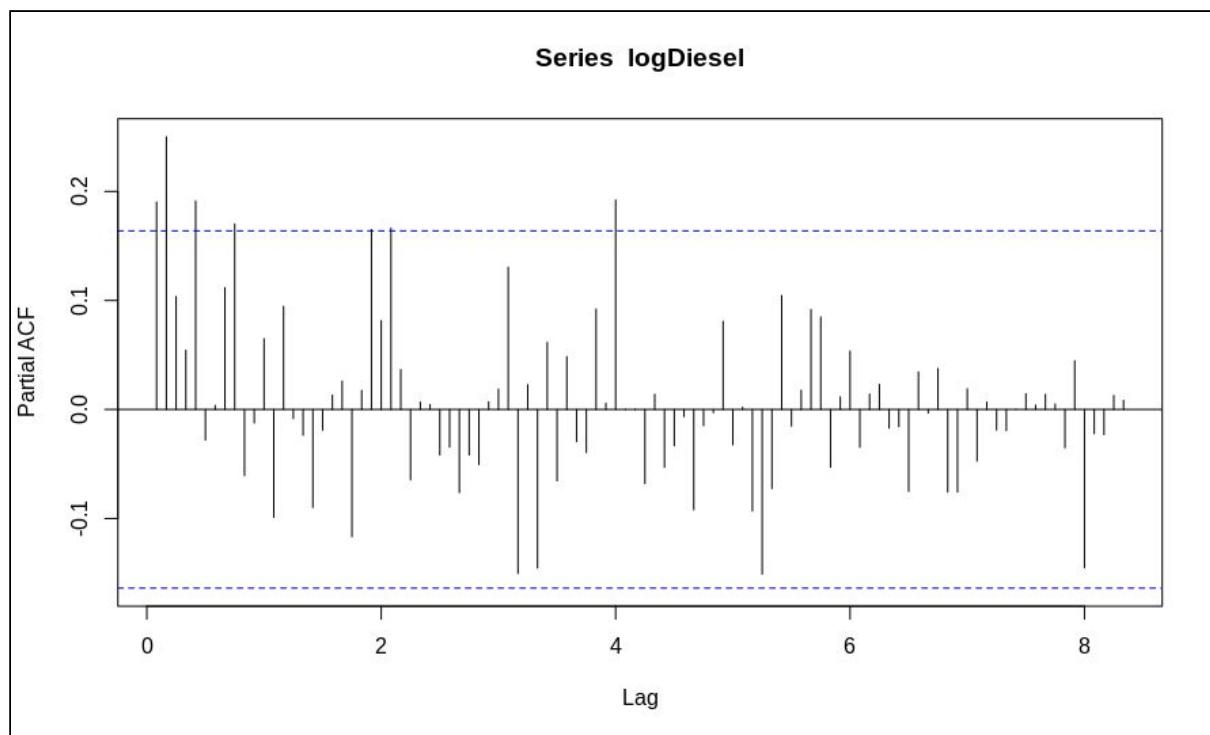
La distribución de los residuales es normal y no hay correlaciones significativas.

FIT AUTO



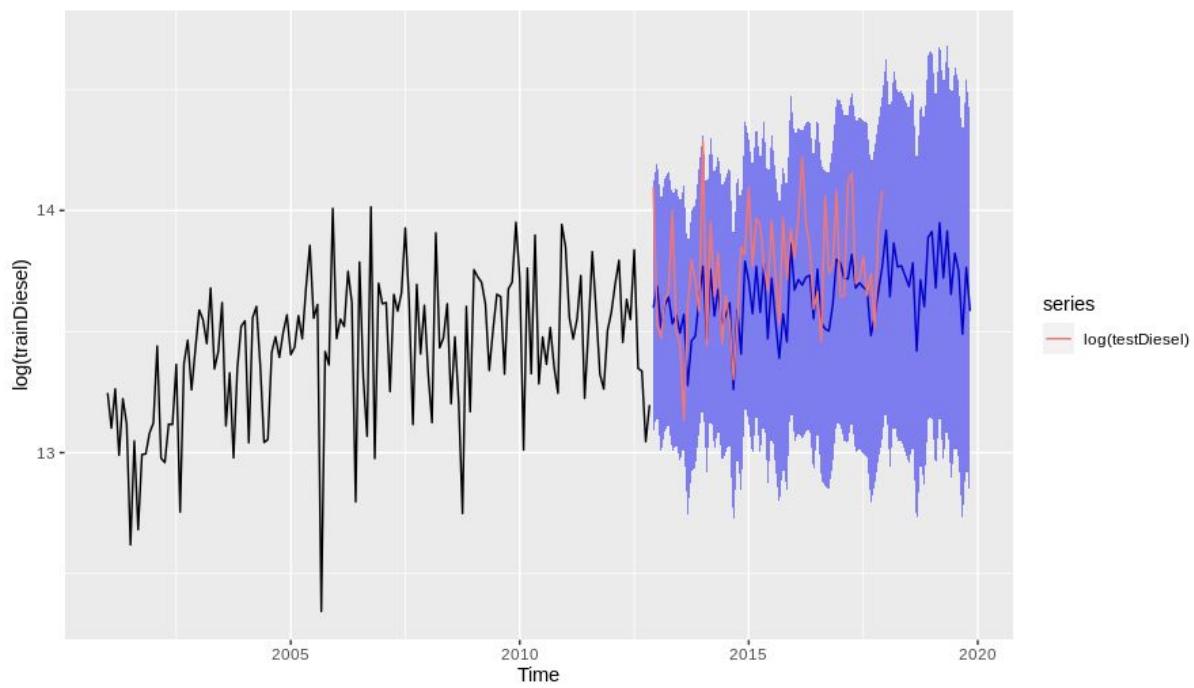
Parece un buen modelo para predecir



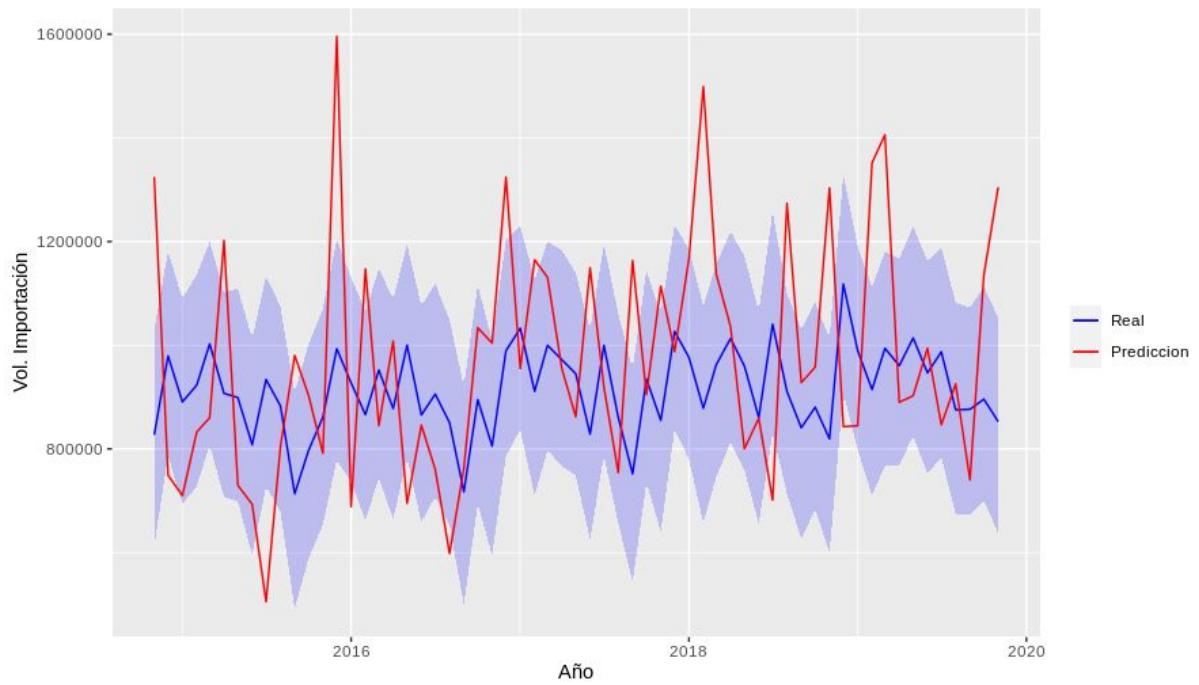


- i. Se realizaron los modelos siguientes:
 - 1. arima(2, 1, 3)x(1, 1, 0)
 - 2. arima(2, 1, 3)x(0, 1, 1)
- ii. El mejor modelo fue el arima(2, 1, 3)x(0, 1, 1)

Forecasts from ARIMA(2,1,3)(0,1,1)[12]



Prophet

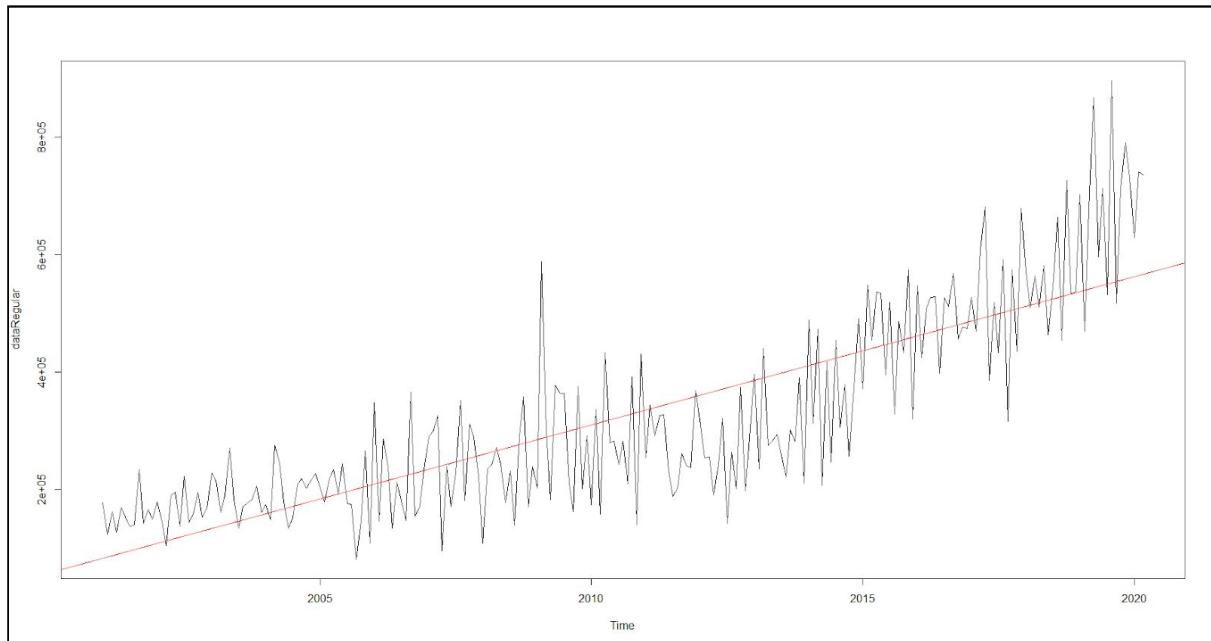


Serie de tiempo de REGULAR

Inicio: 2001, 01

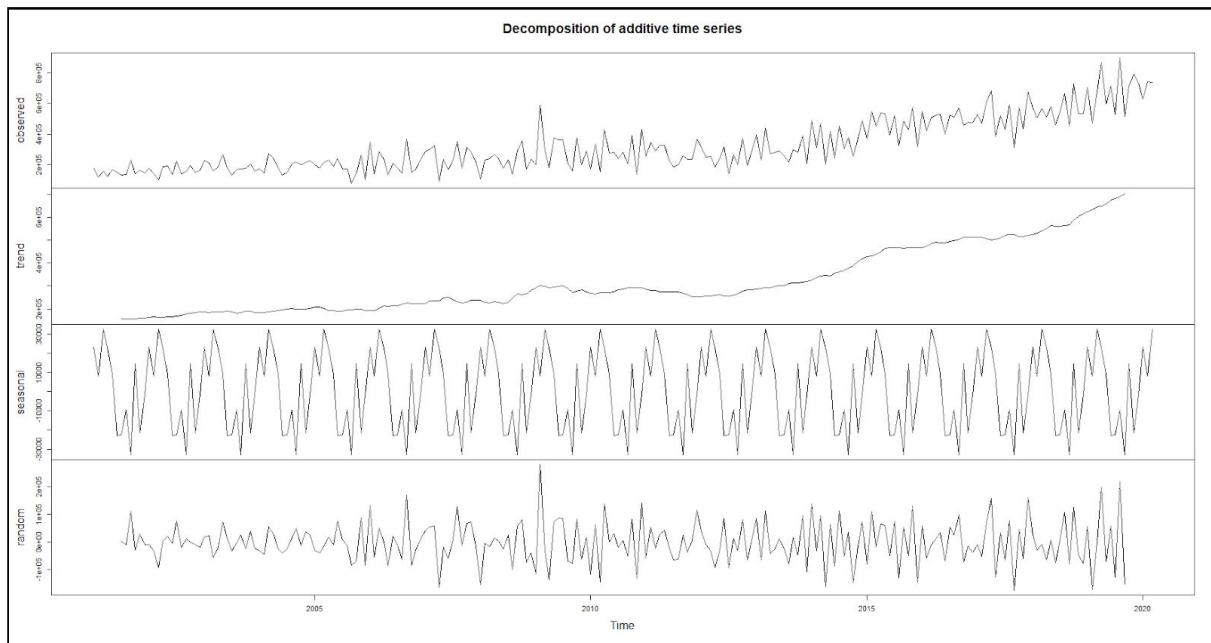
Fin: 2020, 3

Frecuencia: 12

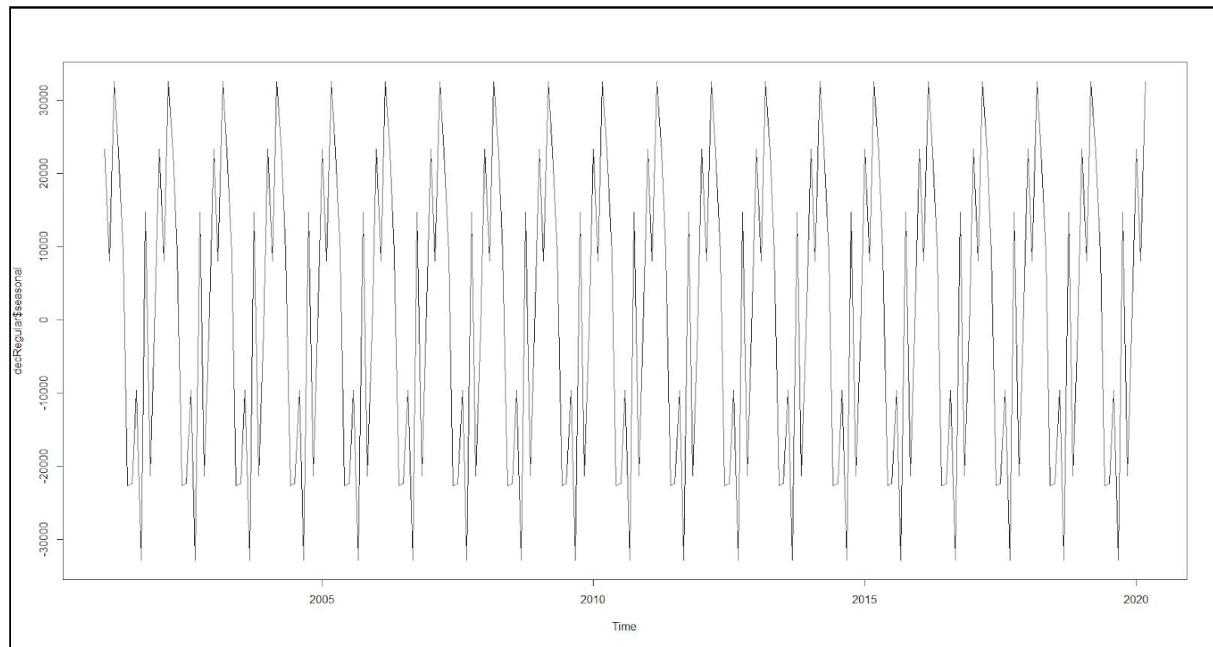


Se aprecia que a medida que pasan los años se importan más galones de Regular como se había demostrado en el análisis exploratorio. Se ve un aumento de media considerable de varianza que no se aprecia tanto a simple vista.

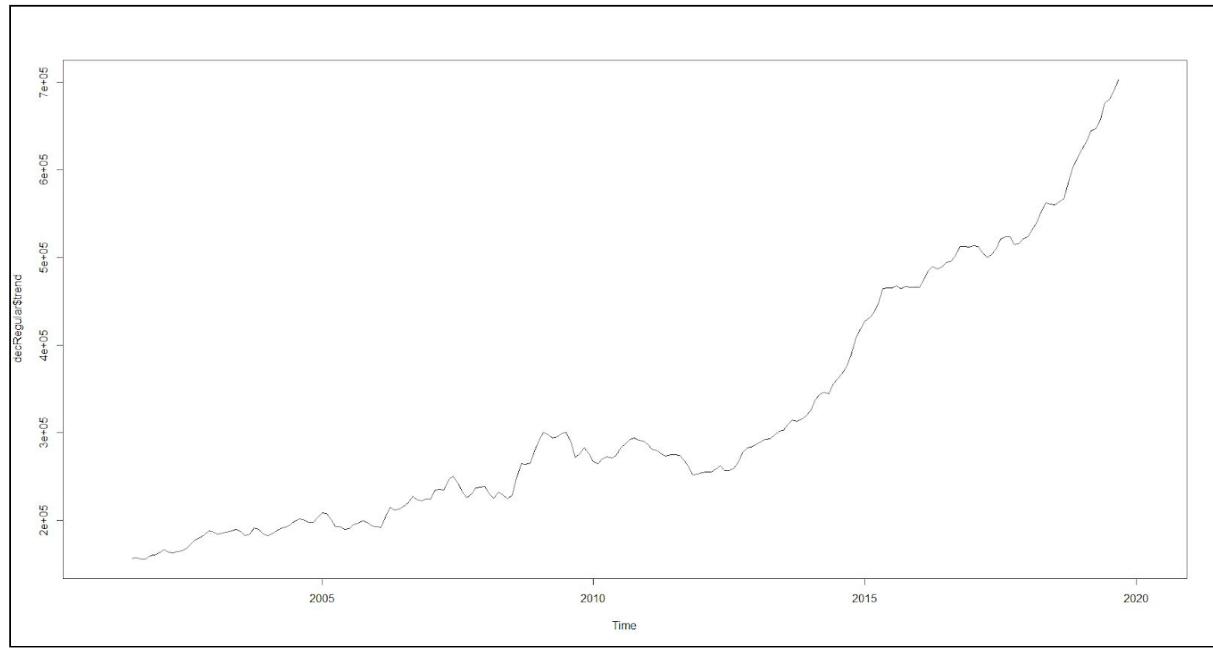
Descomposición en componentes de REGULAR



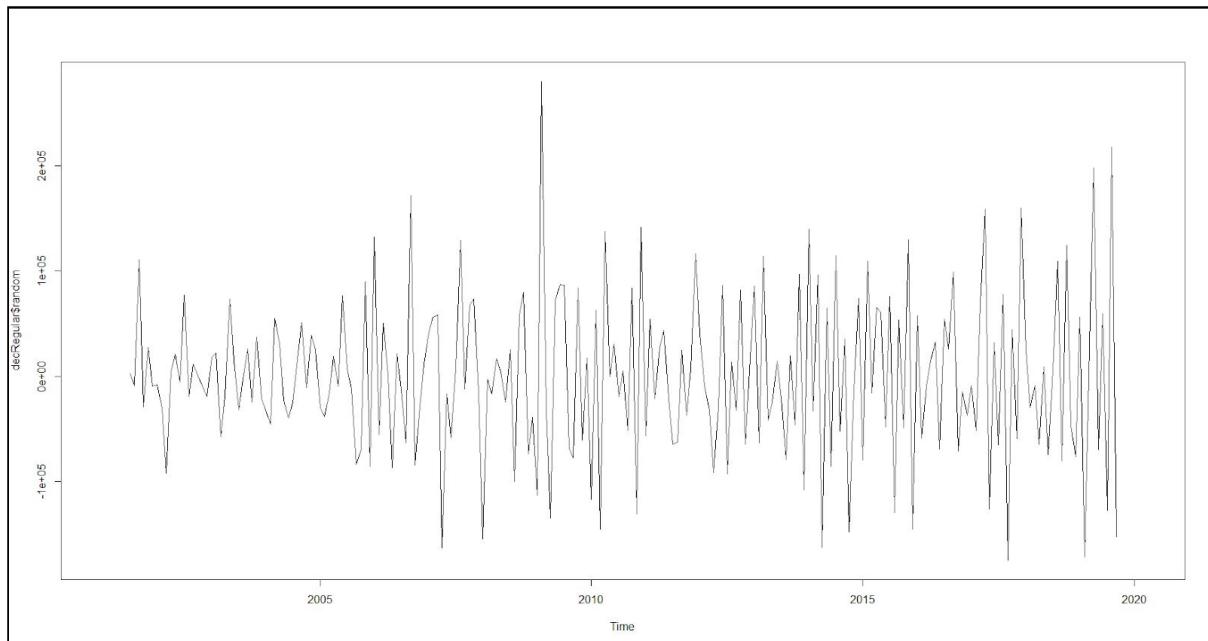
Seasonal



Trend



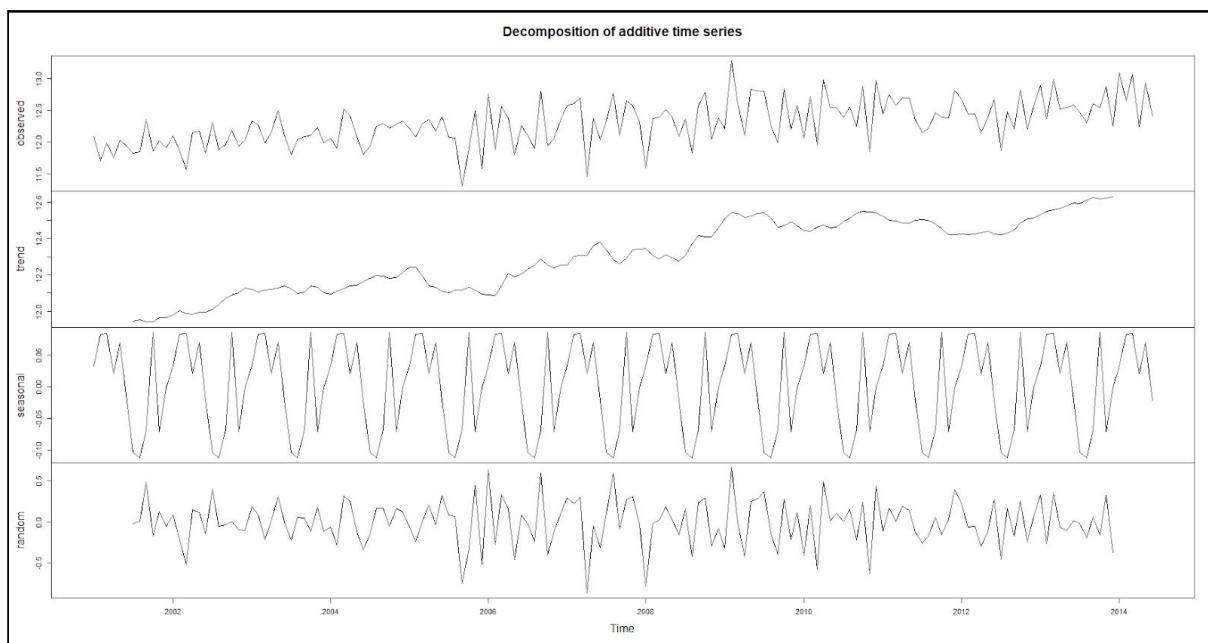
Random



Con estos diagramas podemos observar que está la tendencia a aumentar por lo que no es estacionaria en media y tampoco en varianza.

Si es necesario realizar una transformación logarítmica para intentar conseguir una varianza constante.

Descomposición en componentes de REGULAR (TRANSFORMADA)



Se logró hacer constante la serie en varianza., pero hay que verificar si es estacionaria en media. Si tiene raíces unitarias podemos decir que no es estacionaria en media y hay que aplicar procesos de diferenciación.

Dickey-Fuller Test ($p > 0.05$)

```
Title: Augmented Dickey-Fuller Test
Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
Dickey-Fuller: -1.0843
P VALUE:
0.2689
```

Raíces Unitarias ($p > 0.05$)

```
Title: Augmented Dickey-Fuller Test
Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
DF: -1.0843
P VALUE:
t: 0.2512
n: 0.4567
```

Diferenciación, es necesario

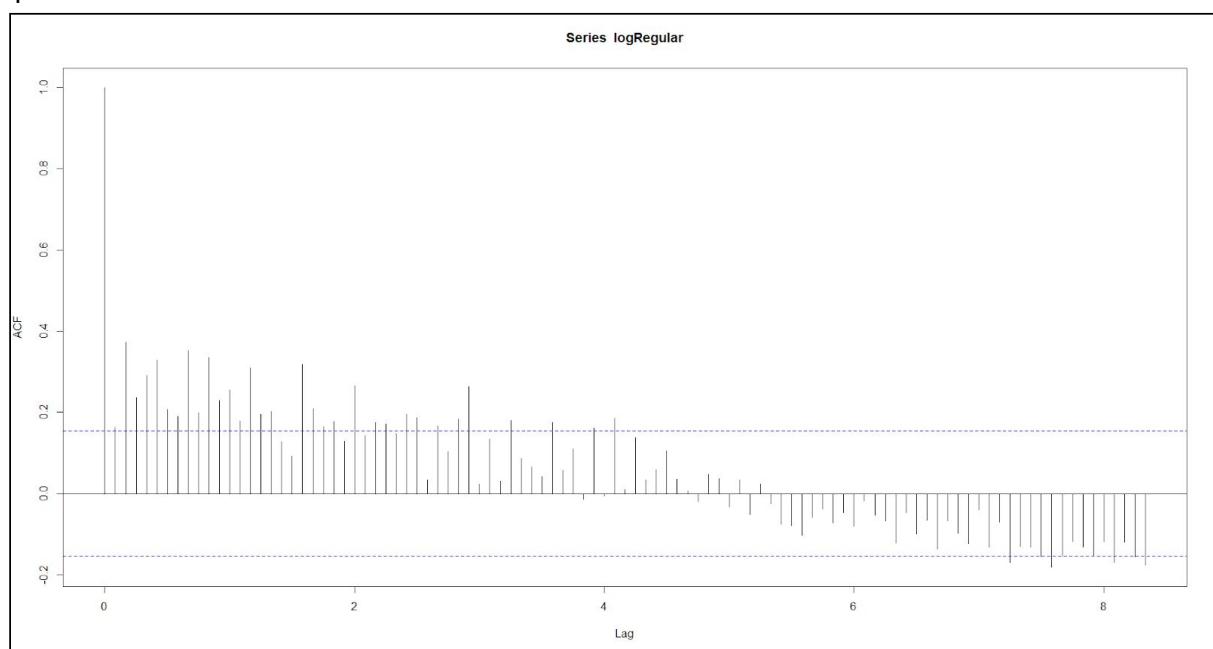
```
Title: Augmented Dickey-Fuller Test
Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
DF: -15.1922
P VALUE:
t: < 2.2e-16
n: 0.005762
```

Basados en la prueba anterior, se elige $d=1$. Para elegir los valores de p y q , se realizaron gráficos de autocorrelación y autocorrelación parcial. Se determinaron los valores $q=3$ y $p=2$.

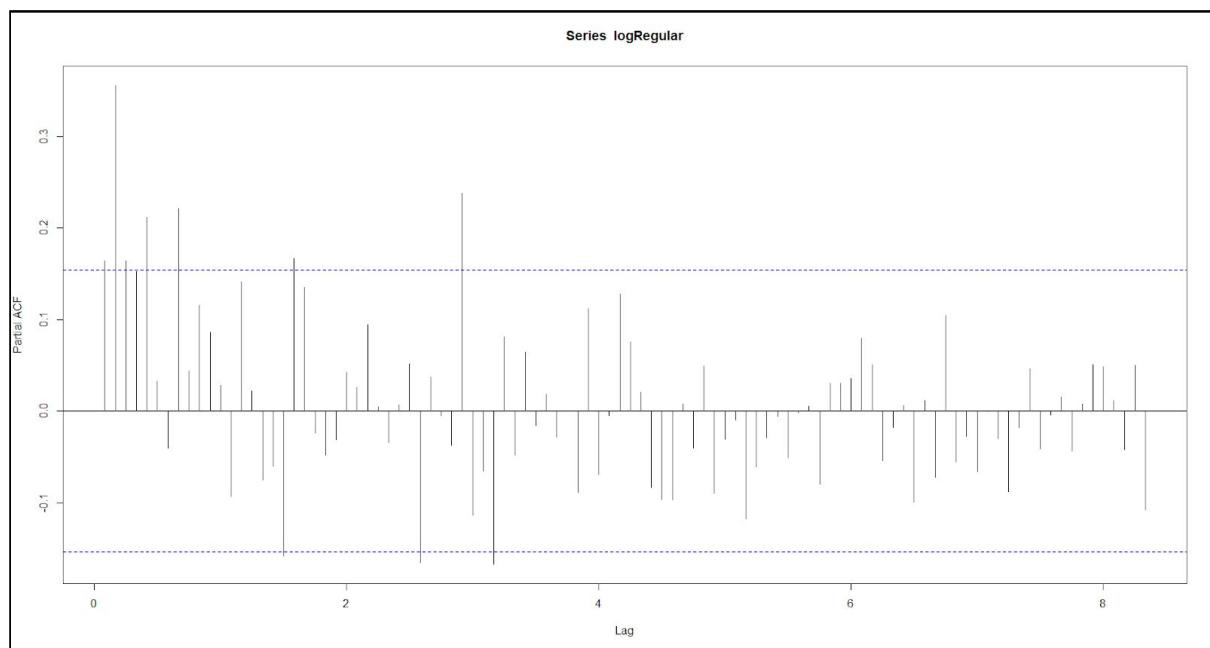
Gráficos de función

Intentar identificar parámetros p y q

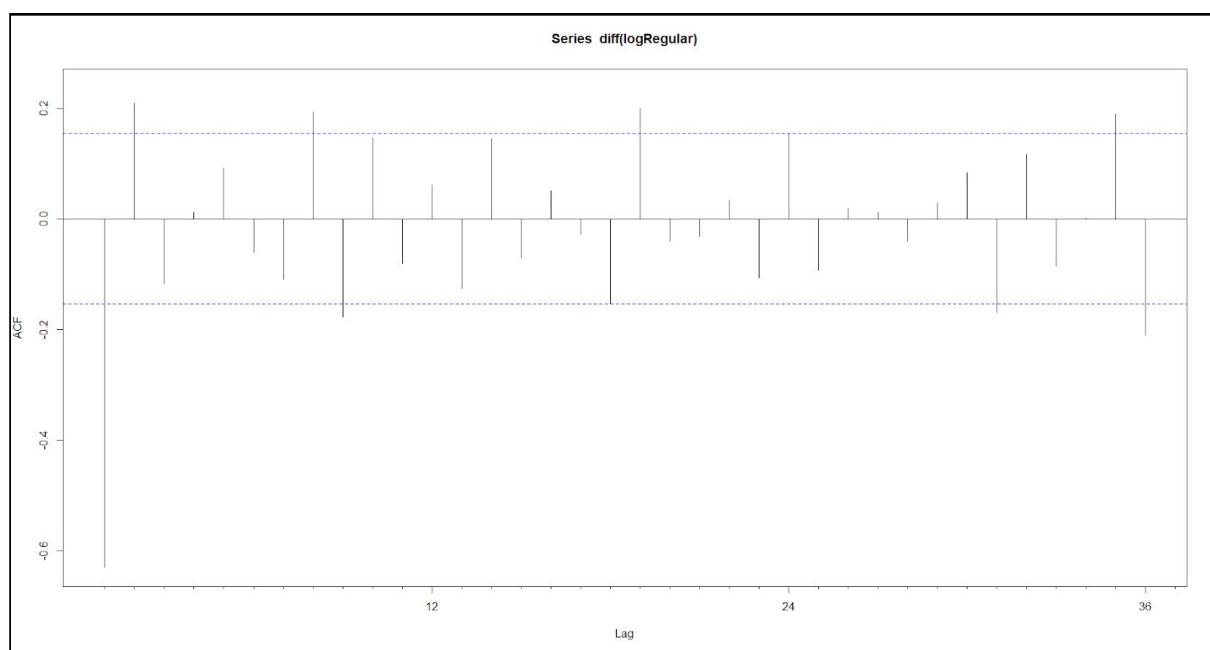
$q = 3$

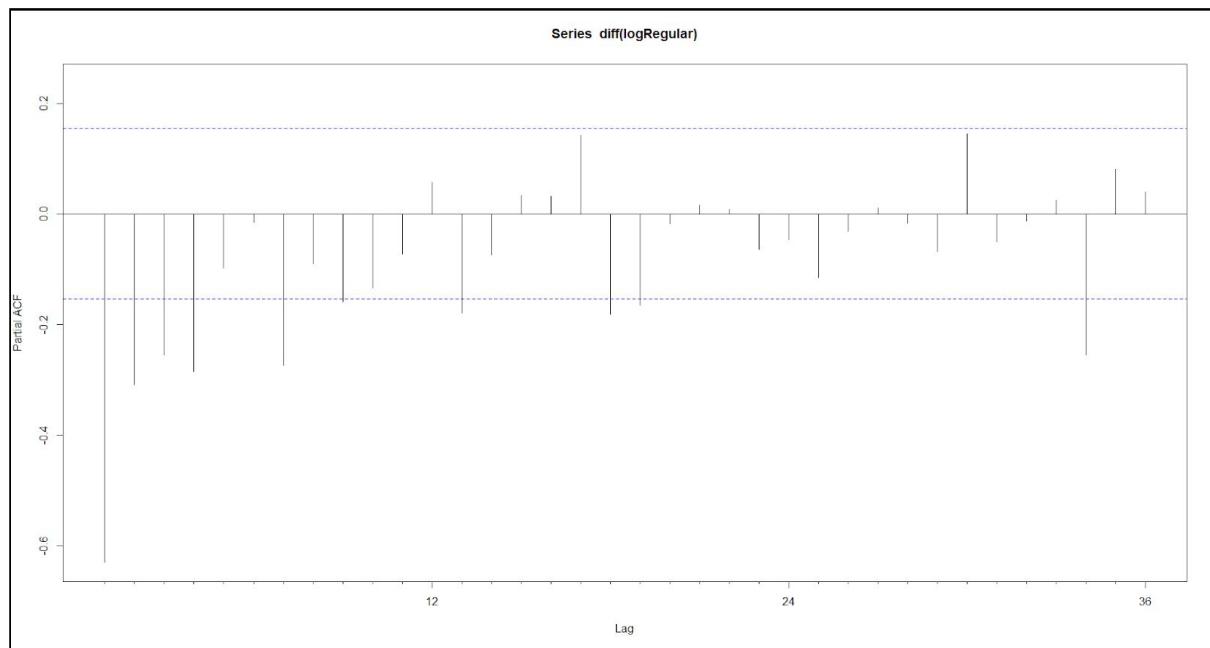


$p = 2$

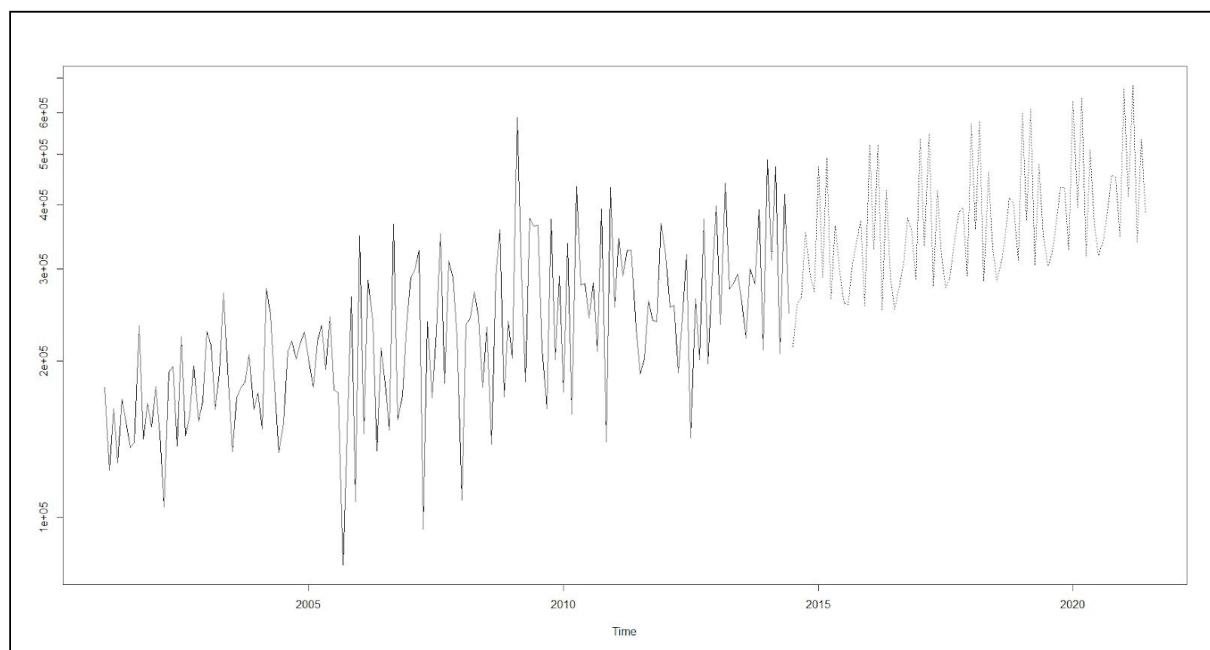


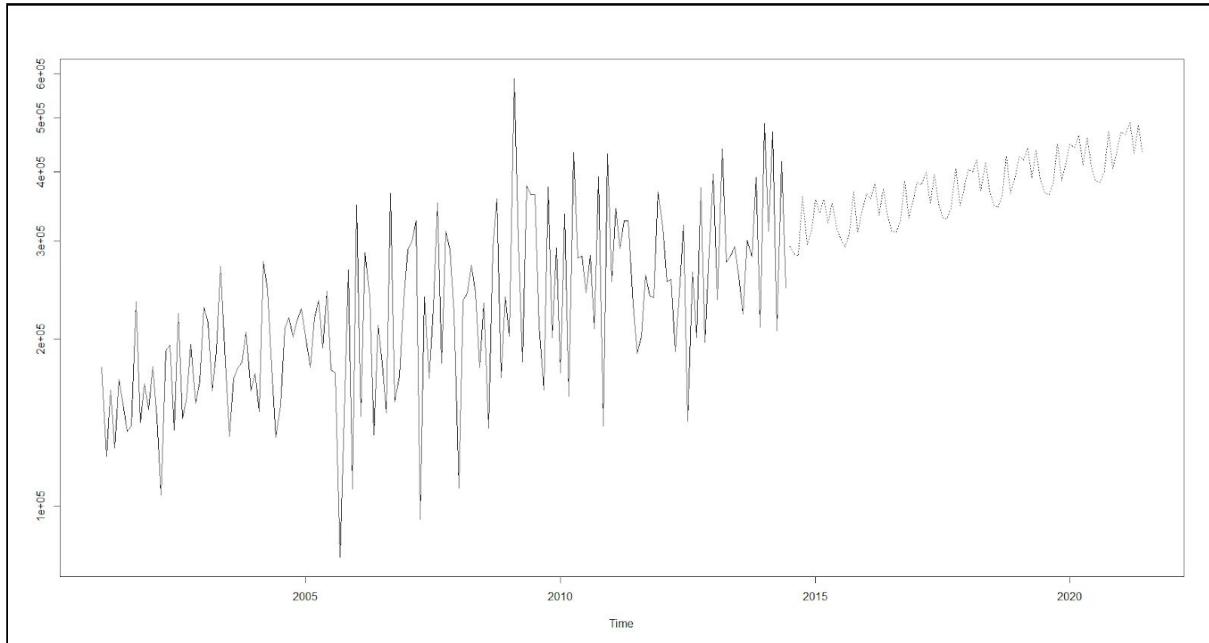
Se usará la serie estacionalizada.





Arima



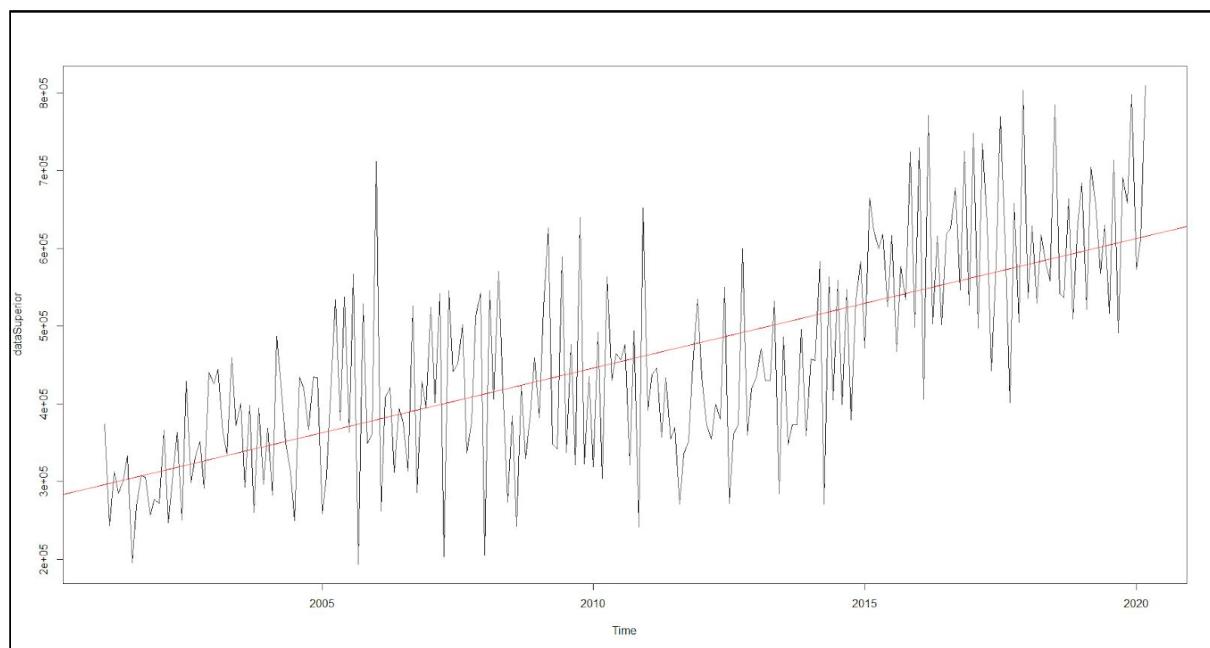


Serie de tiempo de SUPER

Inicio: 2001, 01

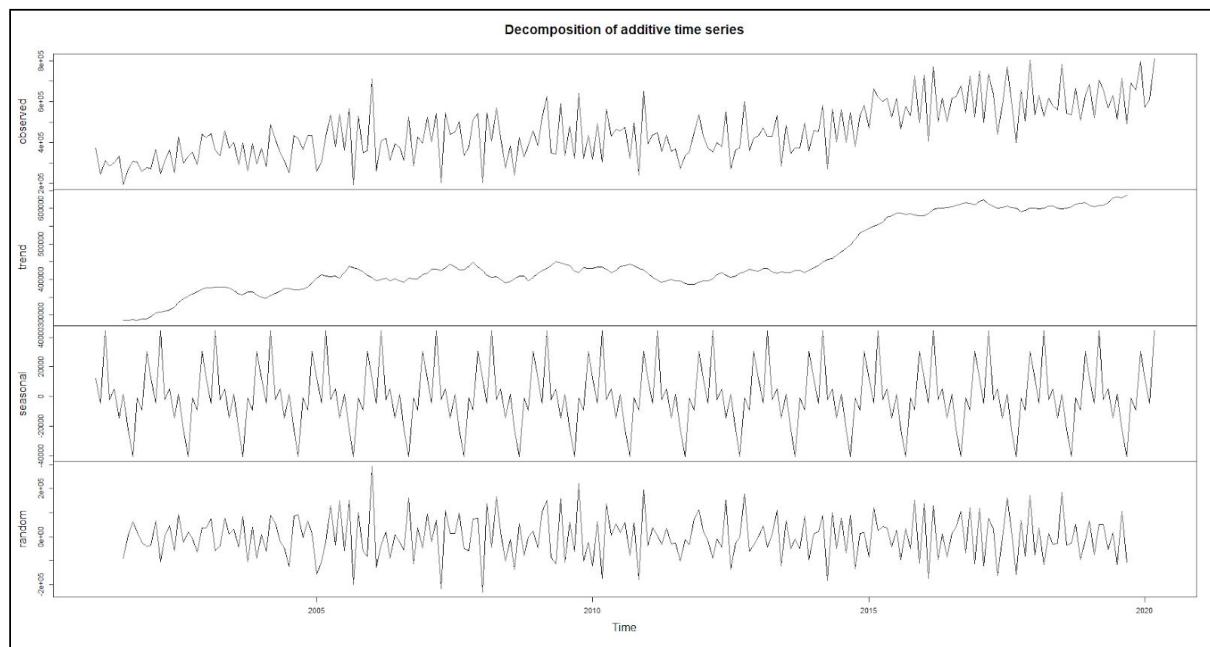
Fin: 2020, 3

Frecuencia: 12

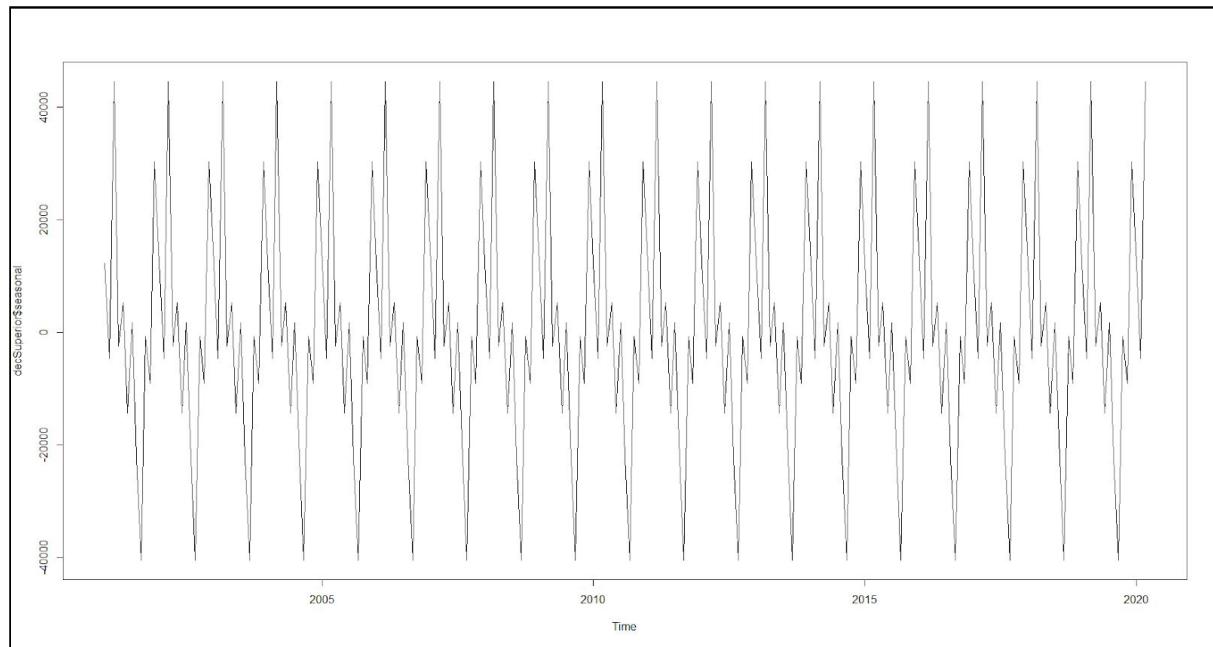


Al igual que los otros 2 se aprecia que a medida que pasan los años se importan más galones o sea un aumento en la media.

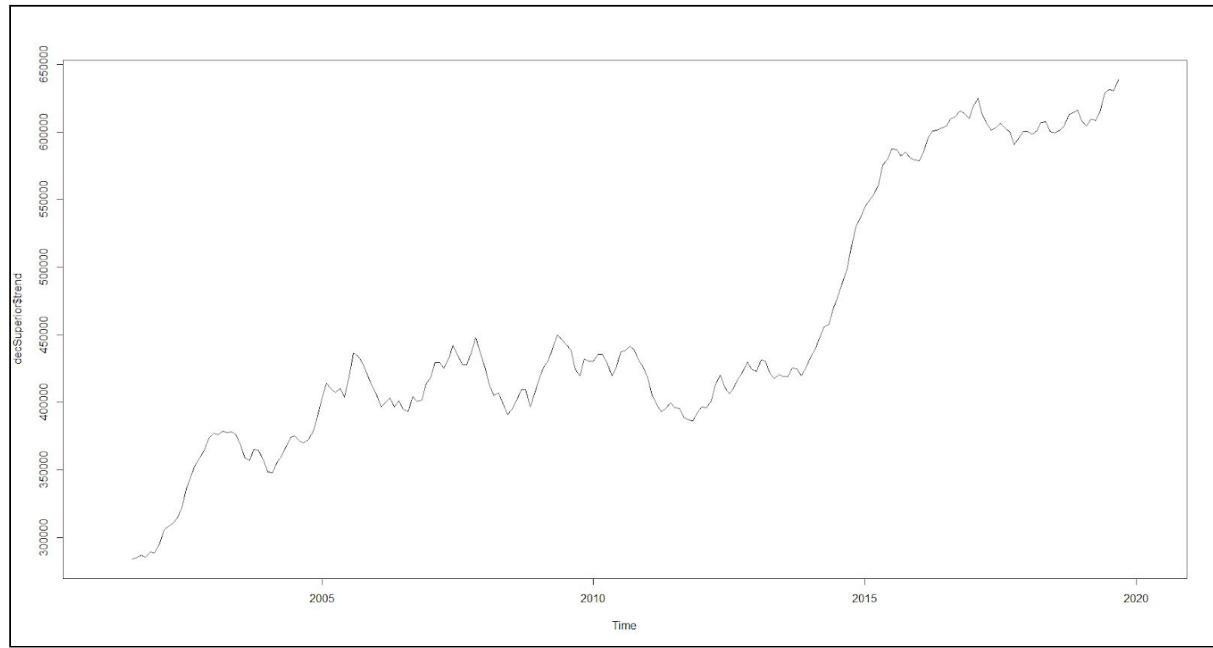
Descomposición en componentes de SUPER



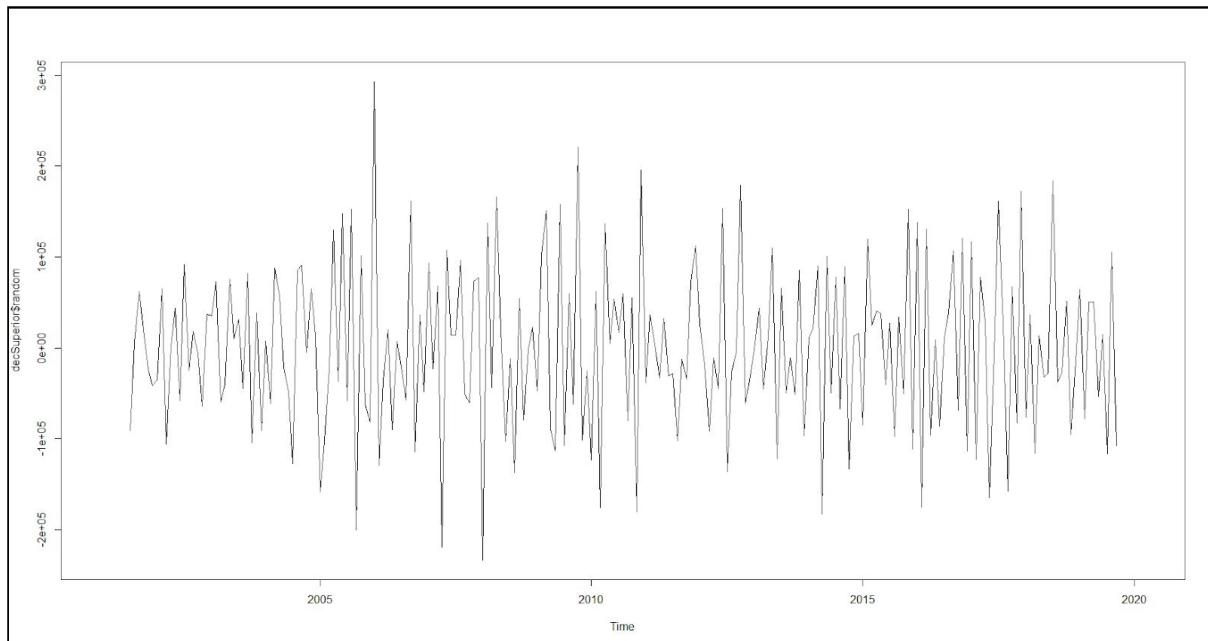
Seasonal



Trend



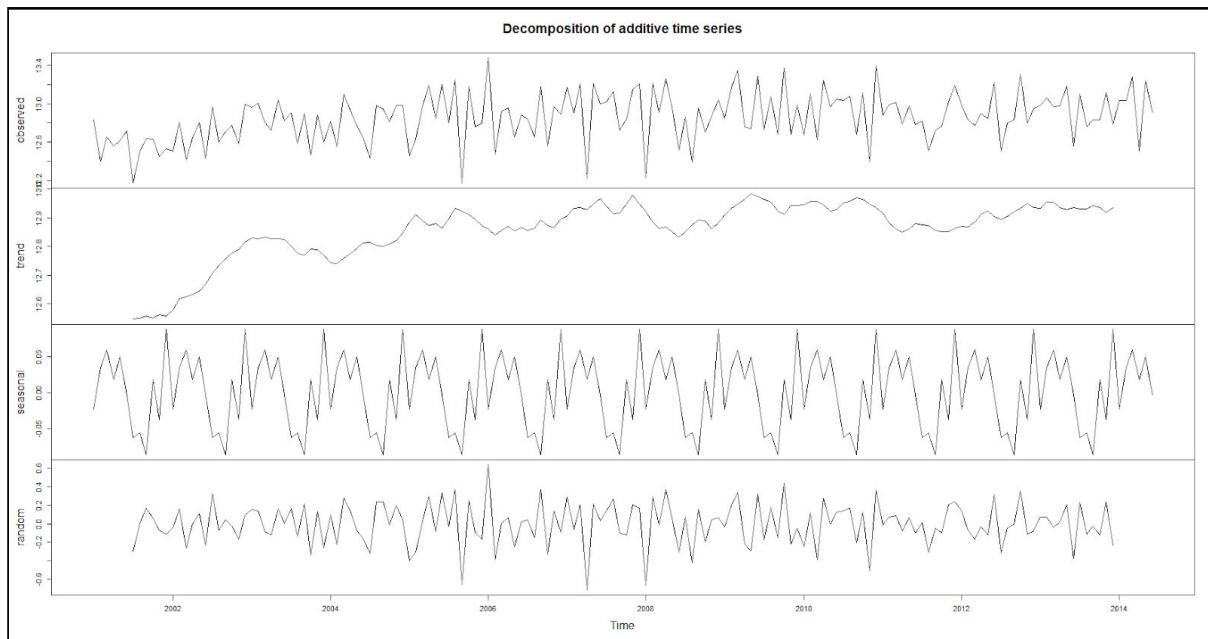
Random



La media y la varianza no se mantiene constante como en ninguno de los demás casos.

Si es necesario realizar una transformación logarítmica para intentar conseguir una varianza constante.

Descomposición en componentes de SUPER (TRANSFORMADA)



Se logró hacer constante la serie en varianza., pero hay que verificar si es estacionaria en media. Si tiene raíces unitarias podemos decir que no es estacionaria en media y hay que aplicar procesos de diferenciación.

Dickey-Fuller Test ($p > 0.05$)

```
Title: Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
Dickey-Fuller: -0.7879
P VALUE:
0.3633
```

Raíces Unitarias ($p > 0.05$)

```
Title: Augmented Dickey-Fuller Test

Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
DF: -0.7879
P VALUE:
t: 0.3734
n: 0.5102
```

Diferenciación, es necesario

```
Title: Augmented Dickey-Fuller Test

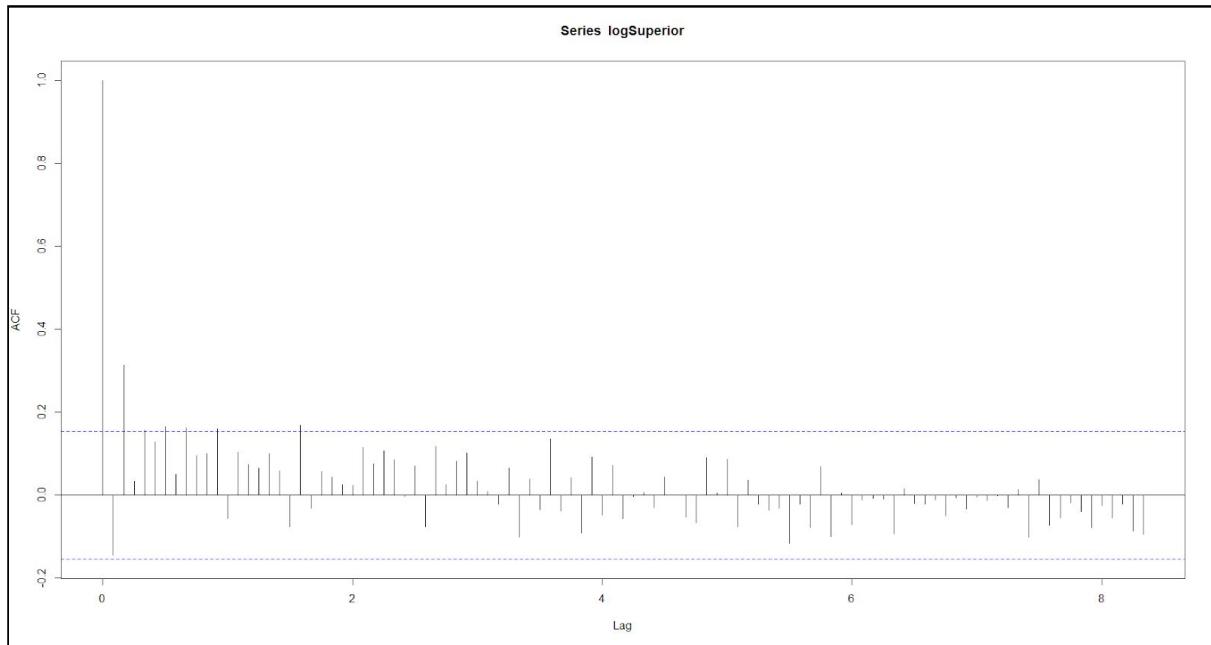
Test Results:
PARAMETER:
Lag Order: 1
STATISTIC:
DF: -16.4935
P VALUE:
t: < 2.2e-16
n: 0.003944
```

Basados en la prueba anterior, se elige $d=1$. Para elegir los valores de p y q , se realizaron gráficos de autocorrelación y autocorrelación parcial. Se determinaron los valores $q=3$ y $p=2$.

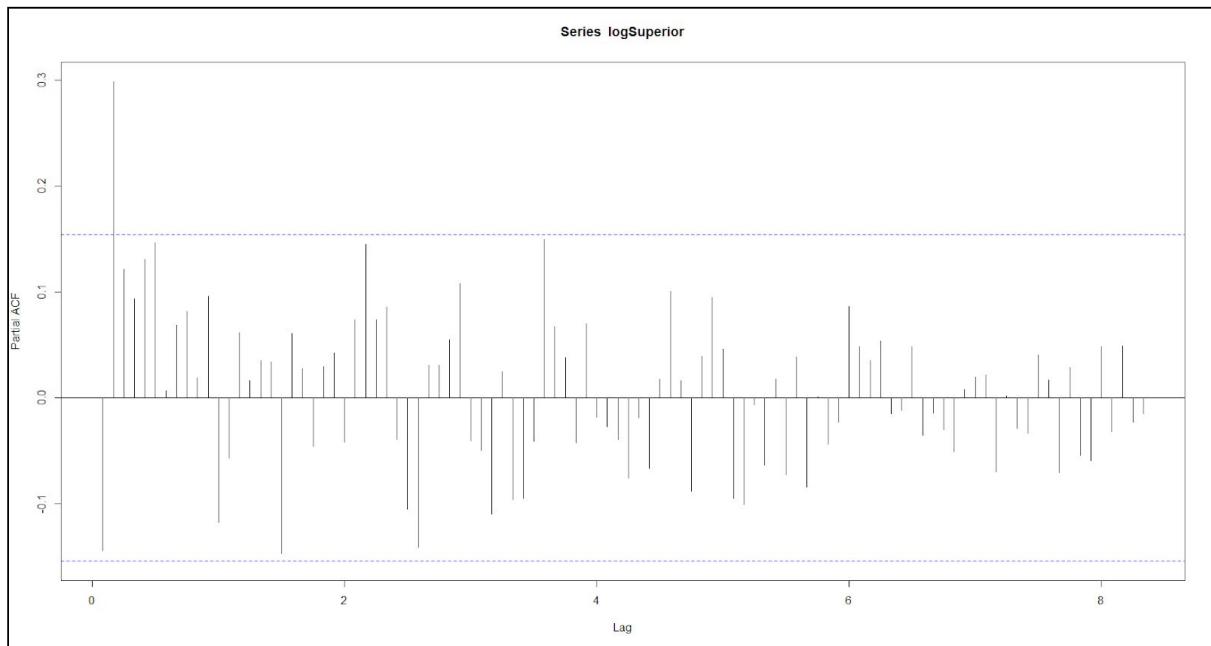
Gráficos de función

Intentar identificar parámetros p y q

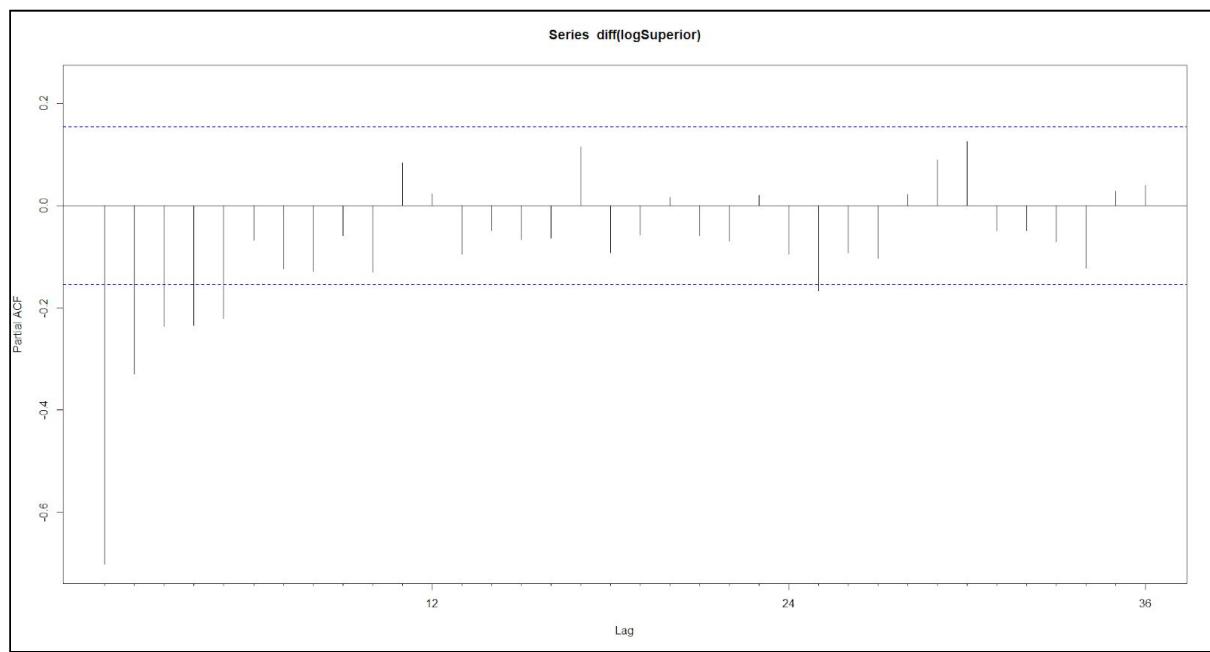
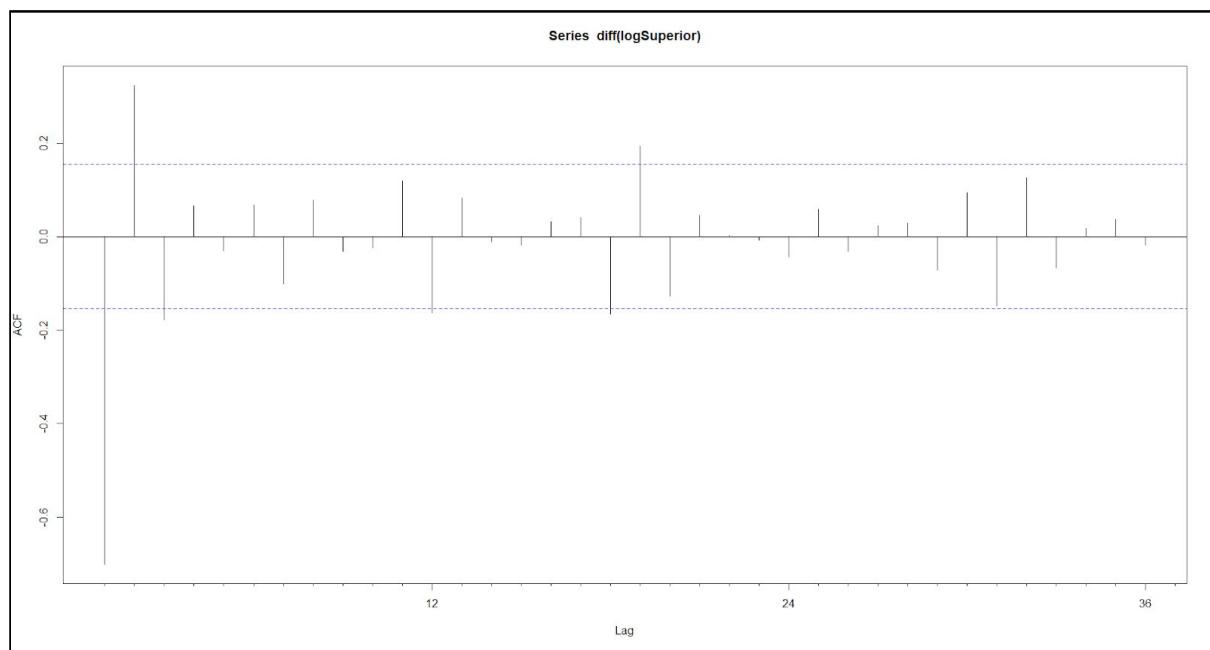
$q = 3$



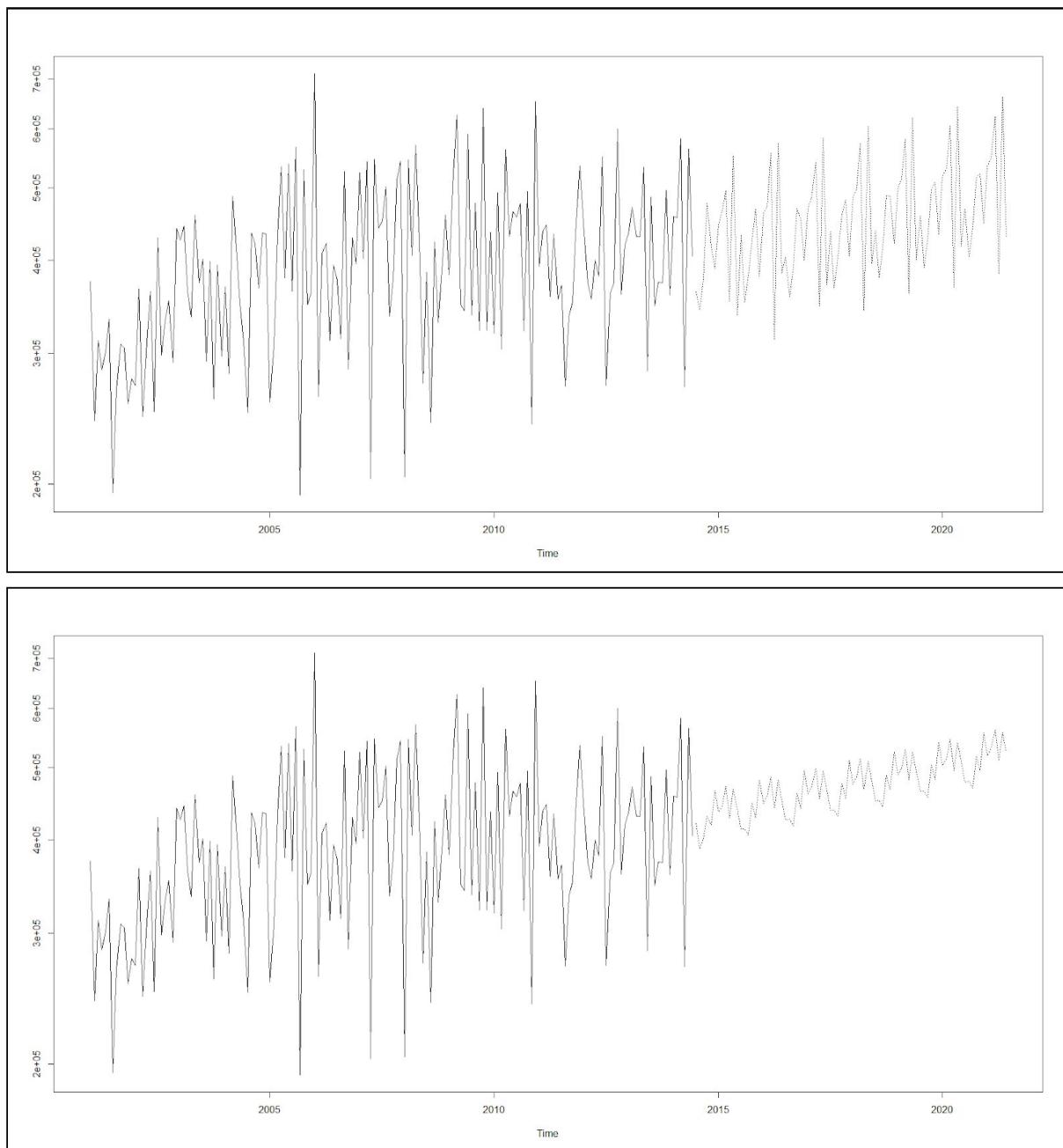
$p = 2$



Se usará la serie estacionalizada.



Arima

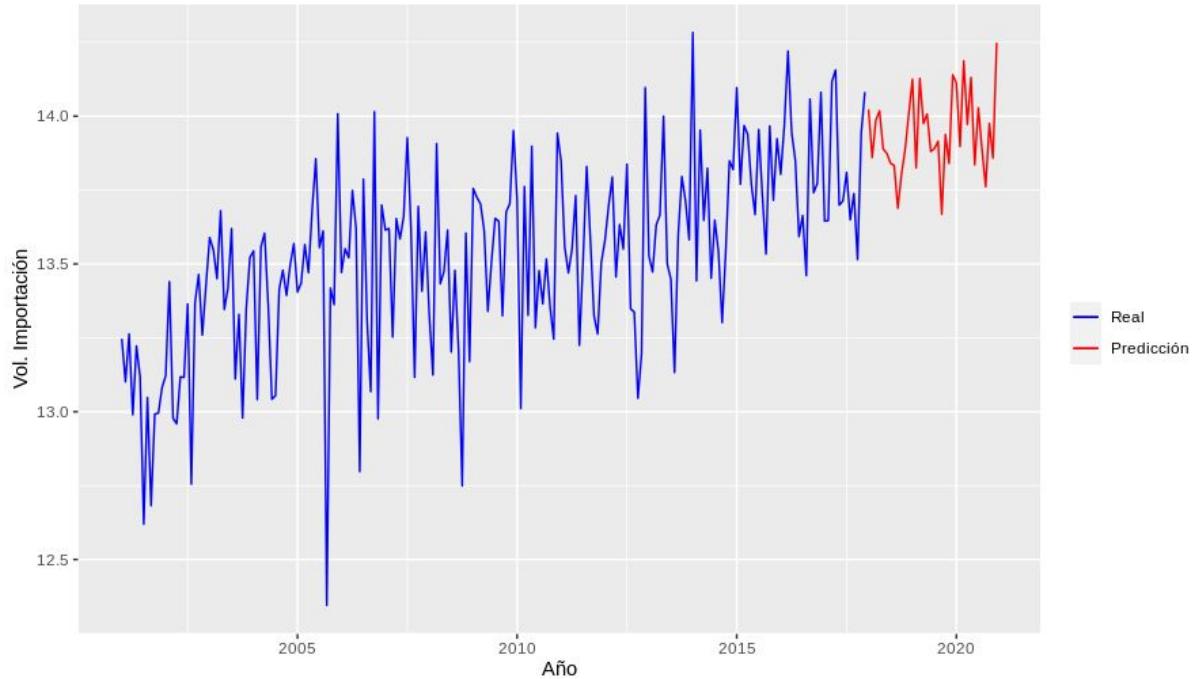


Predicción 2018-2020

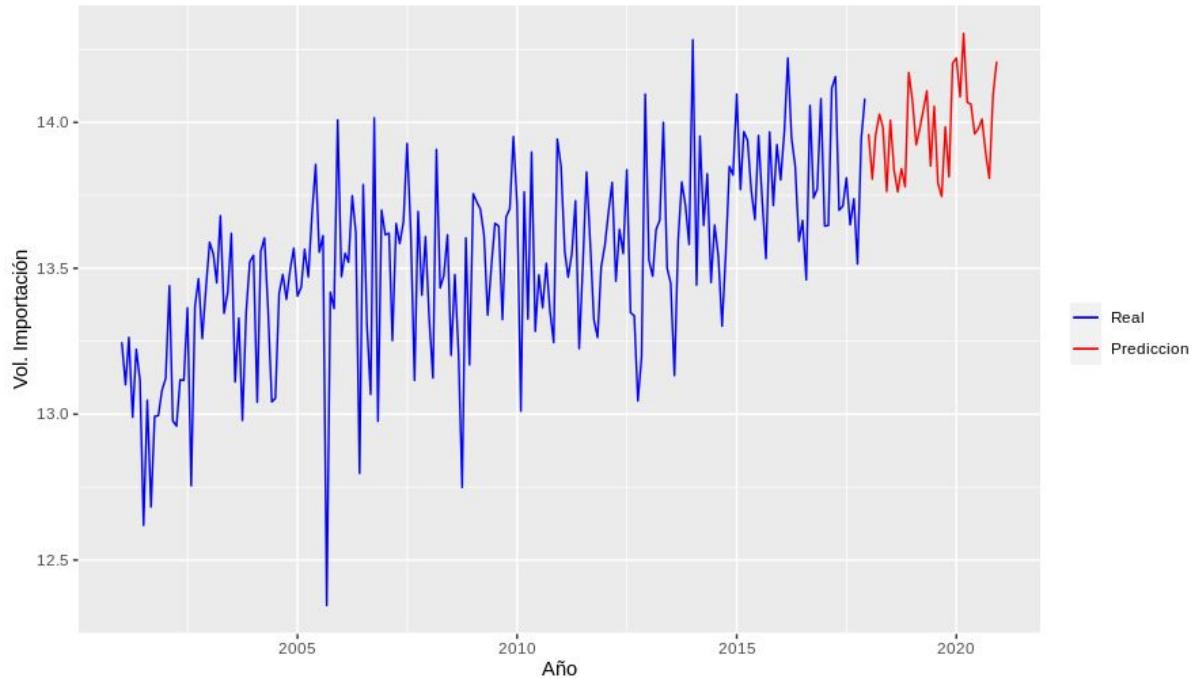
Haga una predicción de los valores de la serie para los años 2018, 2019, y 2020. Explique qué tan buena es su predicción.

DIESEL

Predicción 2018 - 2020 para Diesel (ARIMA)

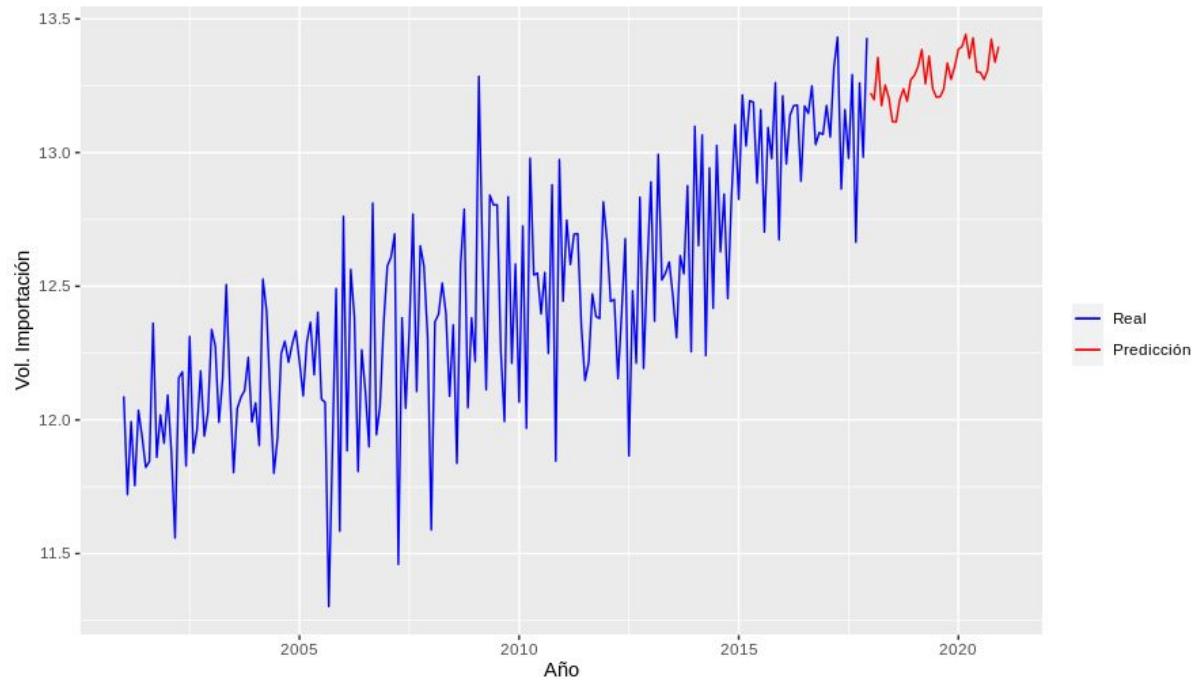


Predicción 2018 - 2020 para Diesel (PROPHET)

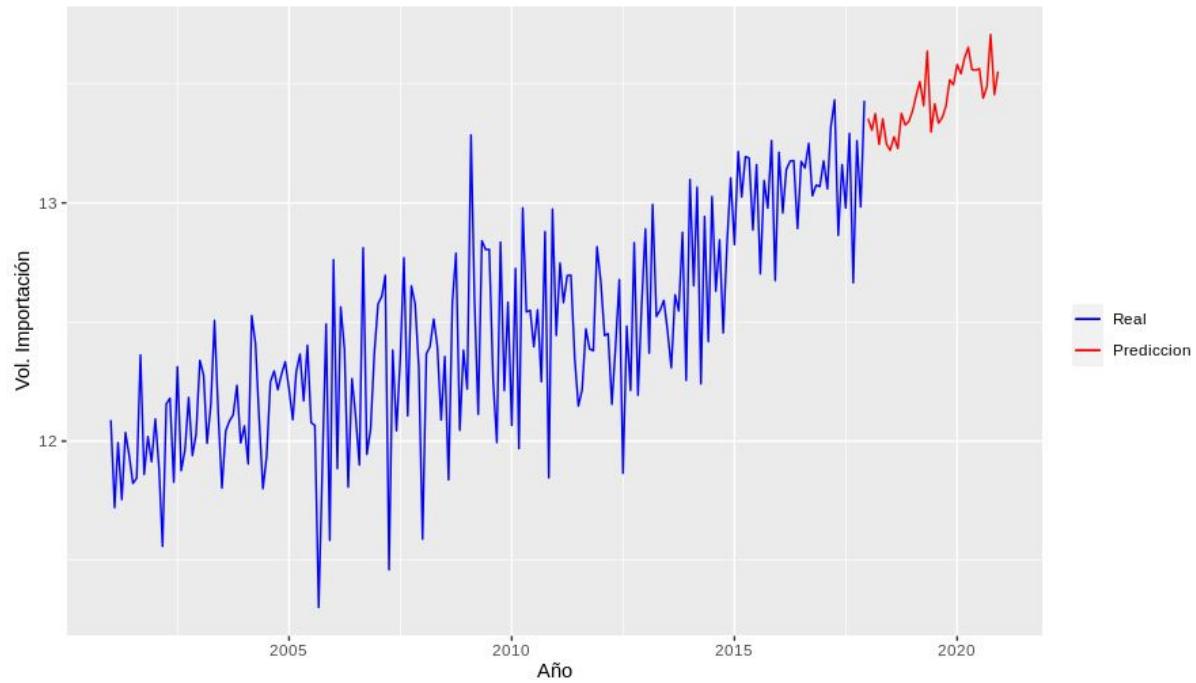


REGULAR

Predicción 2018 - 2020 para Regular (ARIMA)

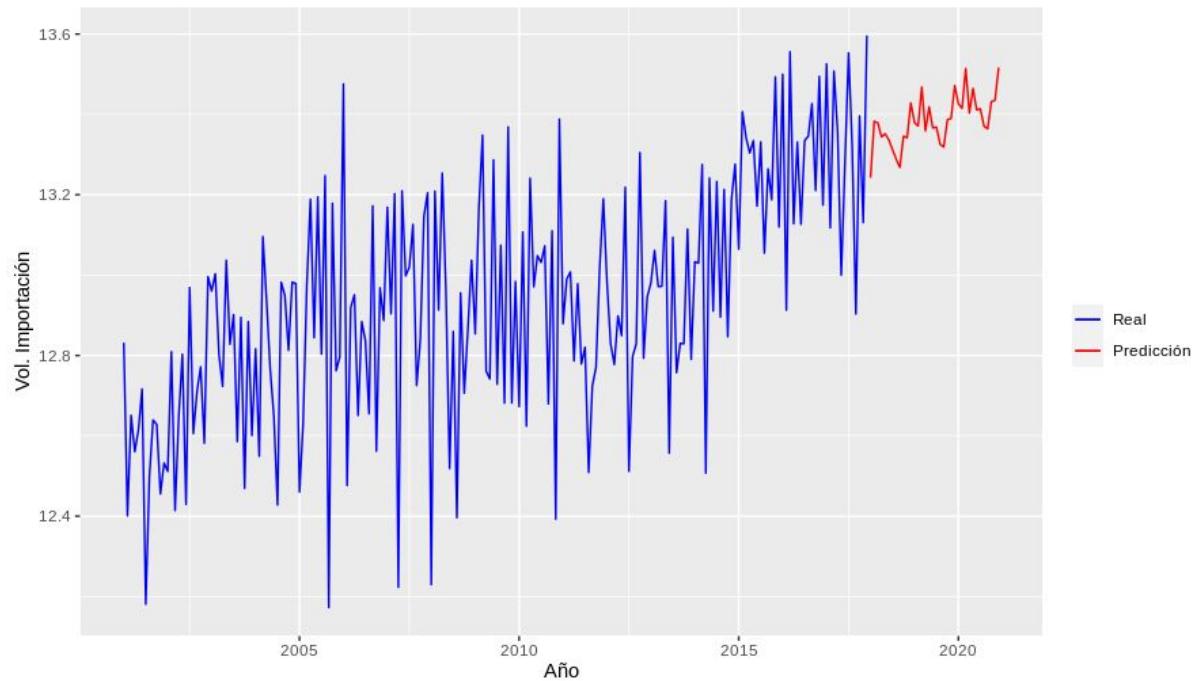


Predicción 2018 - 2020 para Regular (PROPHET)

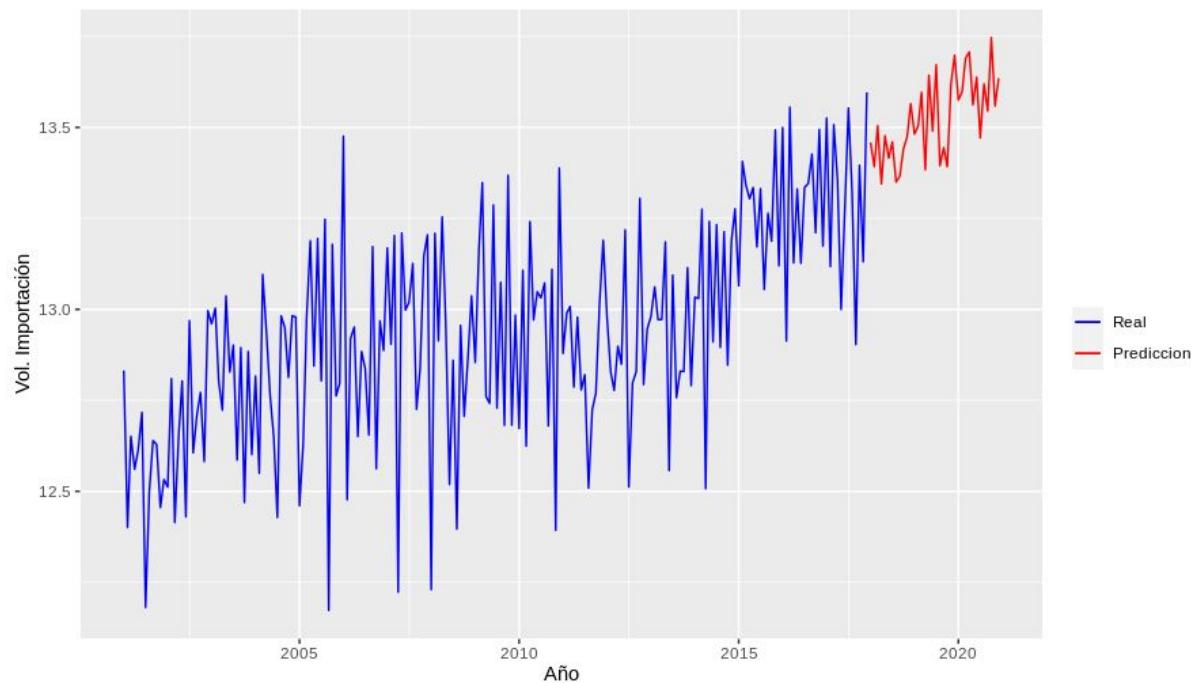


SUPERIOR

Predicción 2018 - 2020 para Superior (ARIMA)



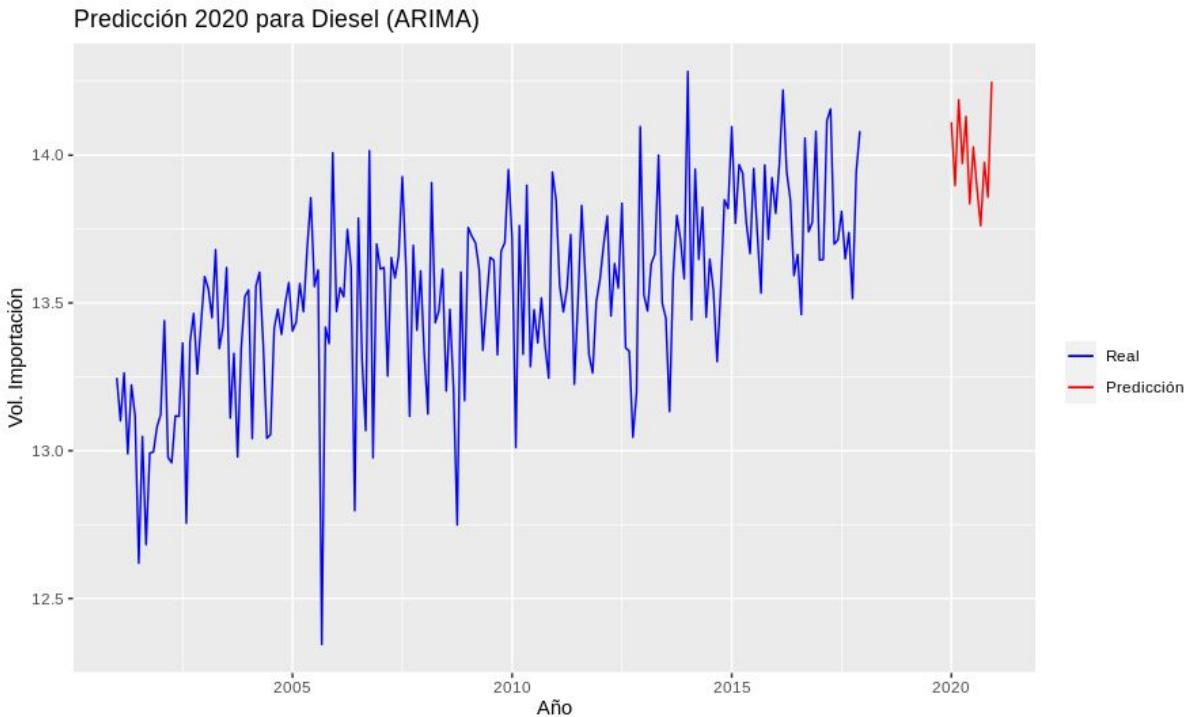
Predicción 2018 - 2020 para Superior (PROPHET)



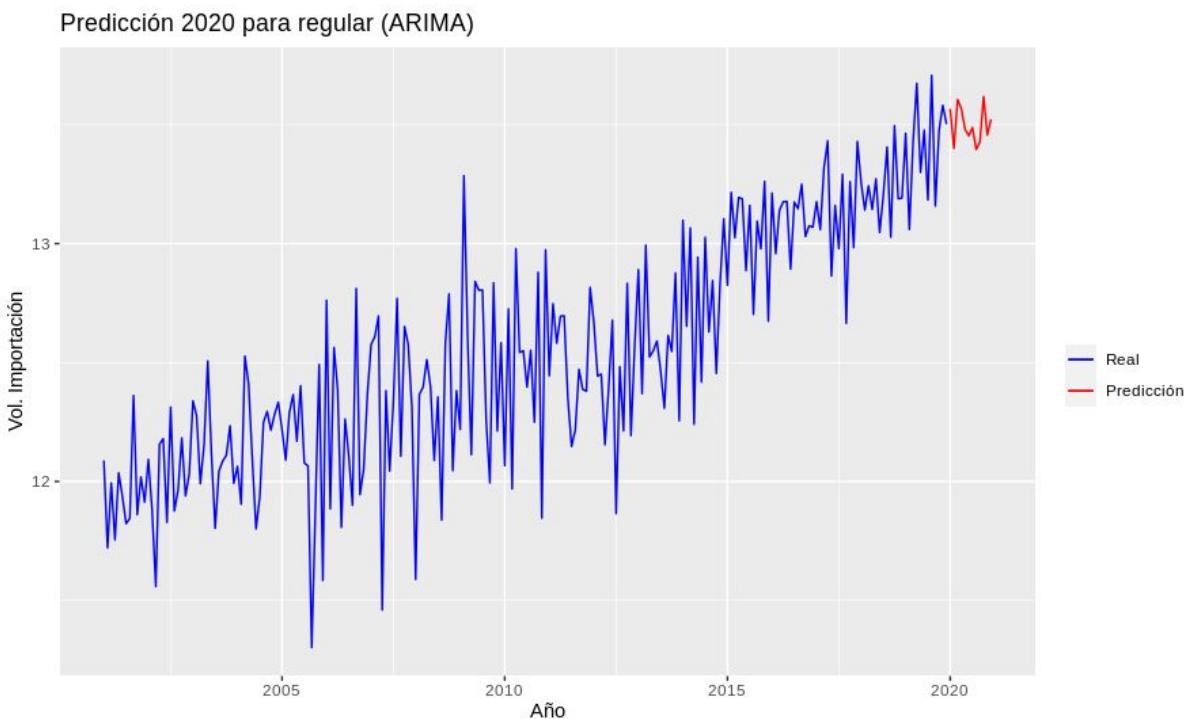
Predicción 2020

Haga una predicción del año 2020 usando los años anteriores y determine qué tan apegada es a la realidad teniendo en cuenta los meses pasados del año actual.

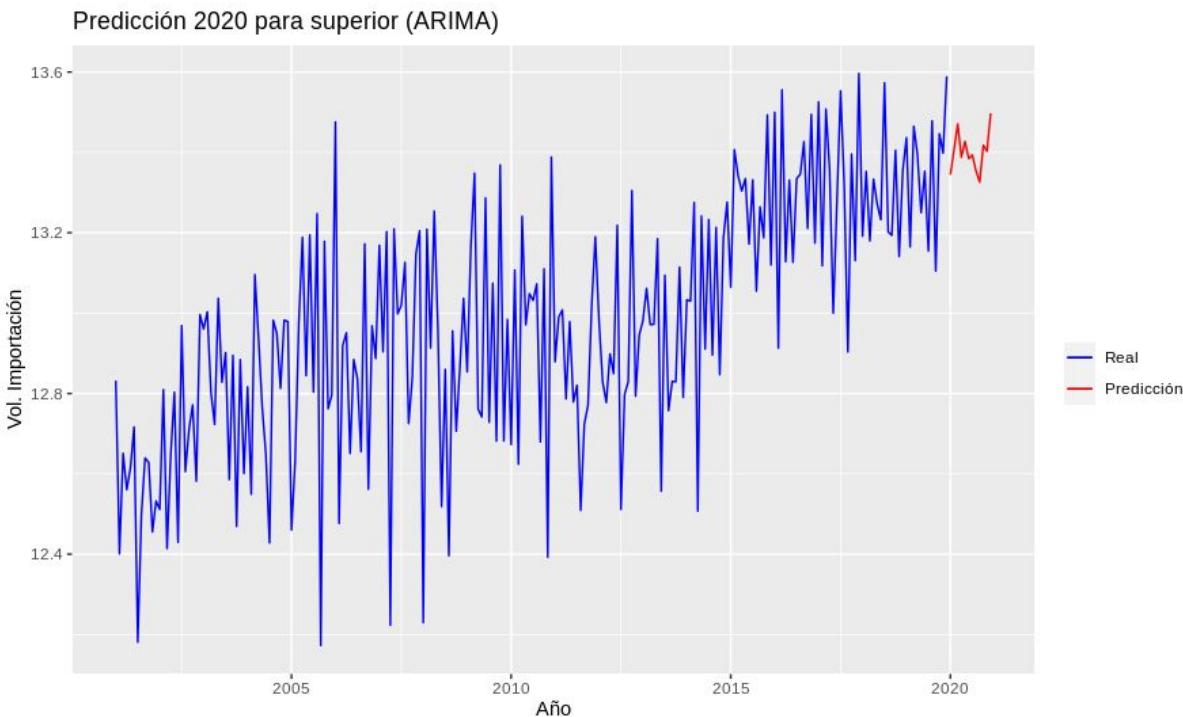
DIESEL



REGULAR



SUPERIOR



Se explica que tan bueno es el modelo prediciendo los volúmenes de importación de combustible para el conjunto de prueba.

El modelo es bueno prediciendo los volúmenes de importación de combustible. Al comparar con los datos reales, la predicción realizada parece producir datos similares, indicando un crecimiento en los volúmenes de importación así como una gráfica con un comportamiento acertado.

Se comparan los modelos generados con ambos algoritmos

El modelo generado con ARIMA produjo resultados aceptables. Al comparar con la gráfica generada con Prophet, ambos modelos parecen indicar un comportamiento similar con respecto al volumen de importación de los tres combustibles.

Es posible observar que en una de las gráficas, la predicción de Prophet es más apegada a una mejor predicción. La predicción en el tipo de gasolina regular, es más acertado en la gráfica que se realizó con Prophet a la realizada con ARIMA.

En conclusión, Prophet produce una predicción más detallada y similar en comportamiento a los datos originales, mientras la producción de ARIMA parece tener un patrón repetido, pero indicando una predicción de comportamiento similar al otro algoritmo.