Universidad Autónoma de Yucatán

Facultad de Matemáticas

Series de Tiempo

Proyecto Final 1

Profesor: José Luis Batún Cutz

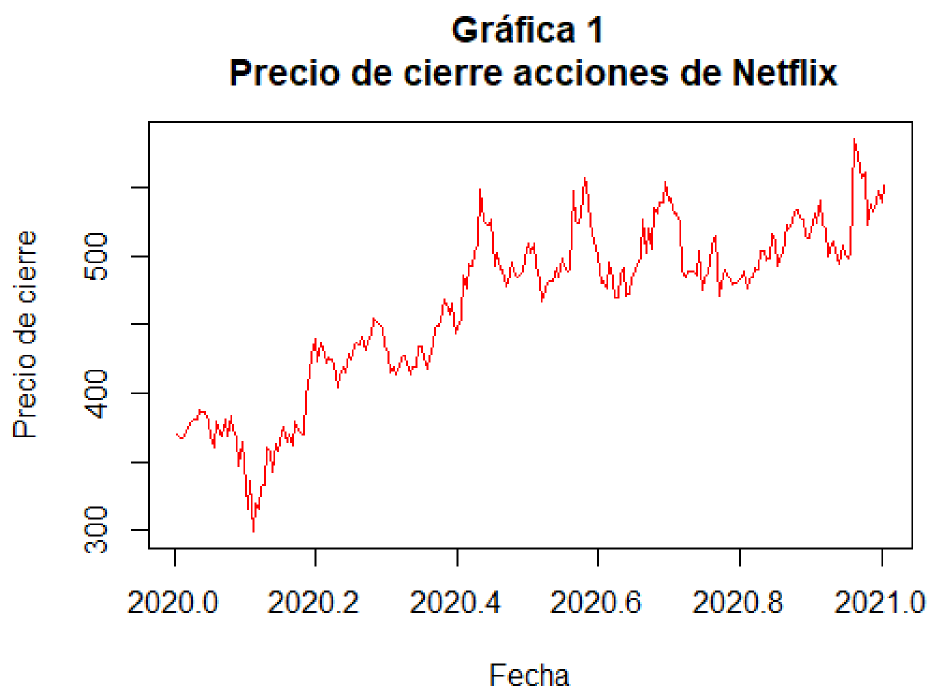
Alumno: Jorge Miguel Guzmán Arjona

Fecha de entrega: 16 de mayo de 2022

Introducción

En este reporte se trabajará con una base de datos que contiene información diaria respecto a las acciones de **NETFLIX** para el período del 5 de febrero de 2020 al 5 de febrero de 2021 el cual contiene 253 observaciones. En particular nuestra variable de interés será los precios diarios de cierre. Nuestro objetivo será plantear un modelo de serie de tiempo para poder realizar pronósticos de dicha variable. La información fue recabada del sitio web Kaggle: [Enlace]

Identificación del modelo

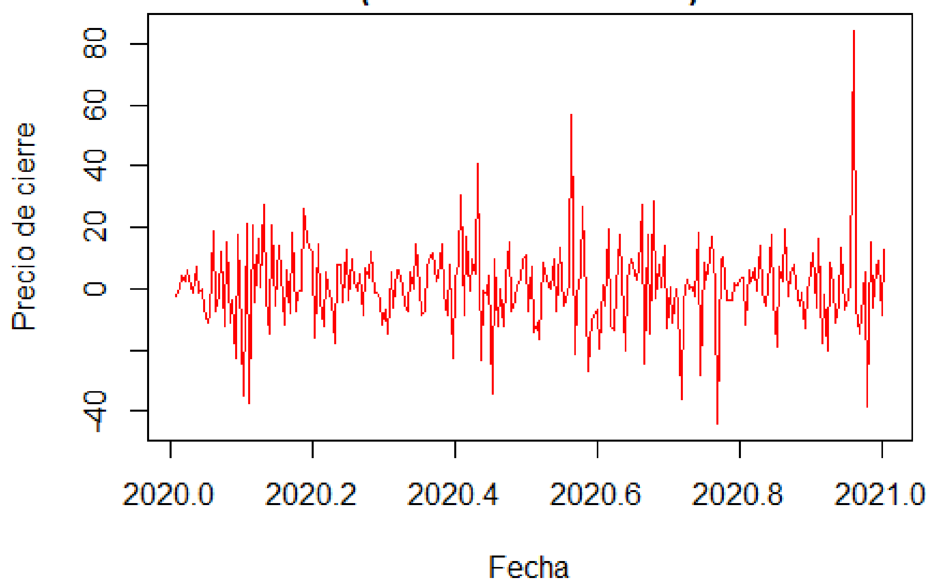


De la **gráfica 1** observamos:

- La serie parece presentar una tendencia alcista, pues el valor medio de los datos en la primera mitad de 2020 pareciera ser 380 pero en la segunda mitad del año los valores parecen tener una media de 500.
- No se observa algún comportamiento que indique que no tenga varianza constante, pues las fluctuaciones a lo largo de la serie parecen mantener su magnitud.
- Tampoco se observa estacionalidad pues no parece haber algún comportamiento repetitivo a lo largo de la serie.

Debido a que no parece tener media constante se hará una primera diferencia, para poder tener una serie estacionaria.

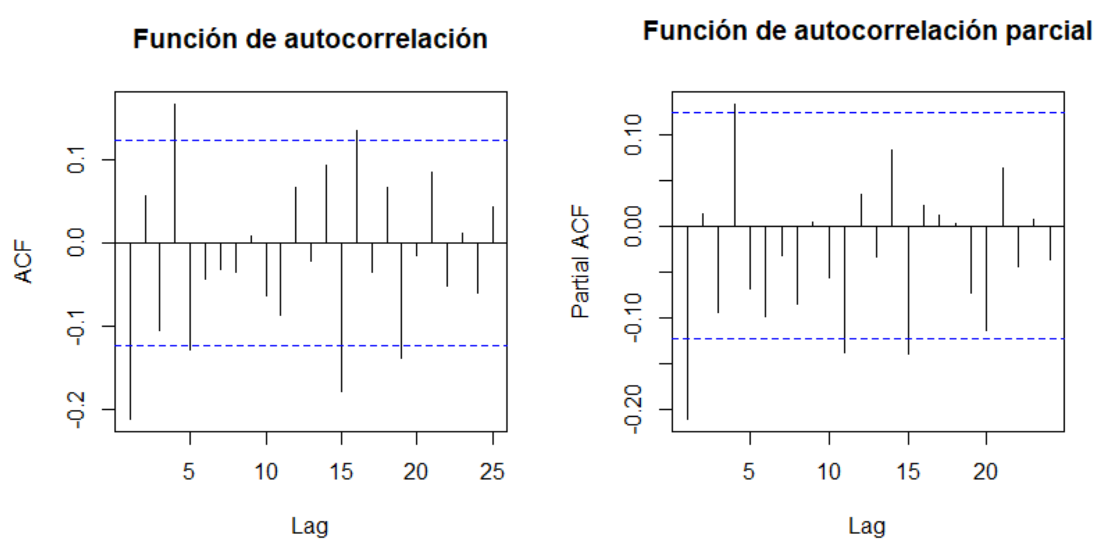
Gráfica 2
Precio de cierre acciones de Netflix
(Primera diferencia)



De la **gráfica 2** observamos que ya se ha ajustado la media. Existen algunos valores que parecen tener comportamiento de outliers pero son pocos para la gran cantidad de datos que tenemos, además siempre habrá la probabilidad de que sucedan. Debido a lo anterior consideraremos que la serie presenta varianza constante y por lo tanto nuestra serie ya es **estacionaria**.

Ahora que nuestra serie es estacionaria, calcularemos el acf y el pacf para poder proponer un modelo.

Gráfica 3



De la **gráfica 3**, al observar el acf y pacf notamos que ninguno tiene un comportamiento de decrecimiento exponencial y ambos caen muy rápidamente dentro de los límites de las bandas de Bartlett por lo que consideraremos que nos encontramos ante un modelo ARMA. Debido a lo anterior estimaremos el eacf a través del esacf.

Gráfica 4.

Extended sample autocorrelation function

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	o	o	x	x	o	o	o	o	o	o	o	o	o
1	o	o	o	o	x	o	o	o	o	o	o	o	o	o
2	x	o	o	o	x	o	o	o	o	o	o	o	o	o
3	x	x	o	o	x	o	o	o	o	o	o	o	o	o
4	x	x	o	x	x	o	o	o	o	o	o	o	o	o
5	x	x	x	x	x	o	o	o	o	o	o	o	o	o
6	x	x	o	o	x	x	o	o	o	o	o	o	o	o
7	x	x	o	o	x	x	x	o	o	o	o	o	o	o

Nuestro objetivo es encontrar un modelo parsimonioso, de la **gráfica 4**, pareciera ser que el modelo que cumple esta característica es el ARMA (p=0, q=5) el cual tiene la siguiente ecuación:

$$\boldsymbol{X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} + \theta_5 \varepsilon_{t-5}}$$

Donde $X_t = Y_t - Y_{t-1}$ y Y_t es nuestra variable de interés, el precio de cierre de la acción de Netflix en el tiempo t, es decir la ecuación para la variable original sería:

$$\boldsymbol{Y_t - Y_{t-1} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} + \theta_5 \varepsilon_{t-5}}$$

Es decir, se propone un modelo ARIMA (p=0, d=1, q=5) para la variable Y_t.

Estimación del modelo

Ya que hemos propuesto un modelo el paso siguiente es estimar sus parámetros, en este caso los coeficientes, esto lo haremos a través de la función "arima" en el software R:

```
z test of coefficients:

     Estimate Std. Error z value Pr(>|z|)
ma1 -0.191750   0.062945 -3.0463 0.002317 **
ma2  0.028818   0.063091  0.4568 0.647841
ma3 -0.087012   0.063740 -1.3651 0.172220
ma4  0.157948   0.063829  2.4746 0.013340 *
ma5 -0.190515   0.069369 -2.7464 0.006026 **
```

De los coeficientes sólo 2 no fueron significativos así que tomaremos como válido el modelo, además los coeficientes que no fueron significativos también son los de menor magnitud así que puede que no tengan gran impacto en el modelo. La ecuación del modelo sería la siguiente:

$$\boldsymbol{Y_t - Y_{t-1} = \varepsilon_t - 0.19\varepsilon_{t-1} + 0.02\varepsilon_{t-2} - 0.08\varepsilon_{t-3} + 0.15\varepsilon_{t-4} - 0.19\varepsilon_{t-5}}$$

Dado que nuestro modelo es un MA (5), por definición es estacionario, verificaremos si es invertible analizando las raíces del polinomio característico:

$$\boldsymbol{X_t = \varepsilon_t - 0.19\varepsilon_{t-1} + 0.02\varepsilon_{t-2} - 0.08\varepsilon_{t-3} + 0.15\varepsilon_{t-4} - 0.19\varepsilon_{t-5}}$$

El cual reescribimos como:

$$\boldsymbol{X_t = \varepsilon_t\left(1 - 0.19B + 0.02B^2 - 0.08B^3 + 0.15B^4 - 0.19B^5\right)}$$

Usando un software externo obtenemos que el polinomio tiene 1 raíz real y 4 complejas:

$$\boldsymbol{r_1 = 1.43}$$

$$\boldsymbol{r_2 = 1.29i + 0.64}$$

$$\boldsymbol{r_3 = -1.29i + 0.64}$$

$$\boldsymbol{r_4 = 0.89i - 0.96}$$

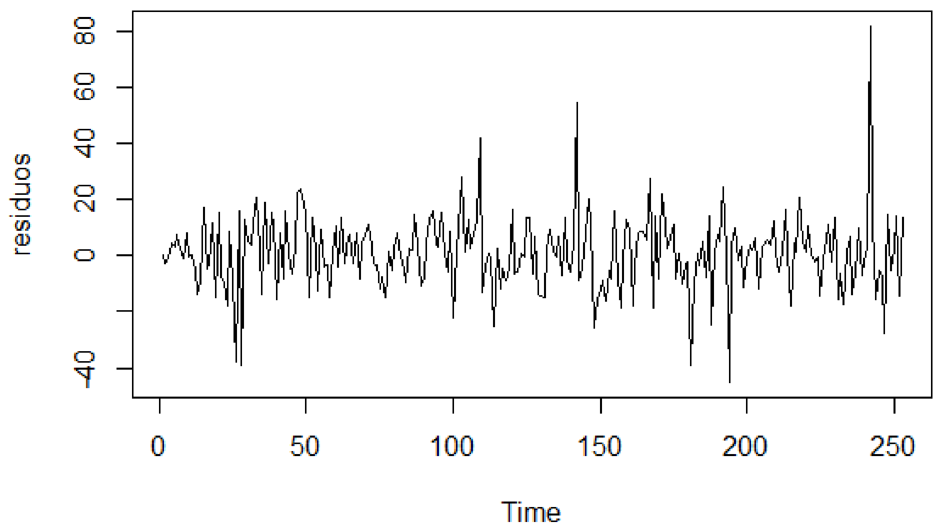
$$\boldsymbol{r_5 = -0.89i - 0.96}$$

Recordemos que para que se cumpla la condición de invertibilidad la norma de las raíces debe ser mayor a 1, lo cual se cumple para las raíces obtenidas, por lo tanto, nuestro modelo es estacionario e invertible.

Validación del modelo

Para darle validez a nuestro modelo necesitamos verificar si nuestros residuales son o no un ruido blanco Gaussiano. Primero graficaremos los residuos del modelo:

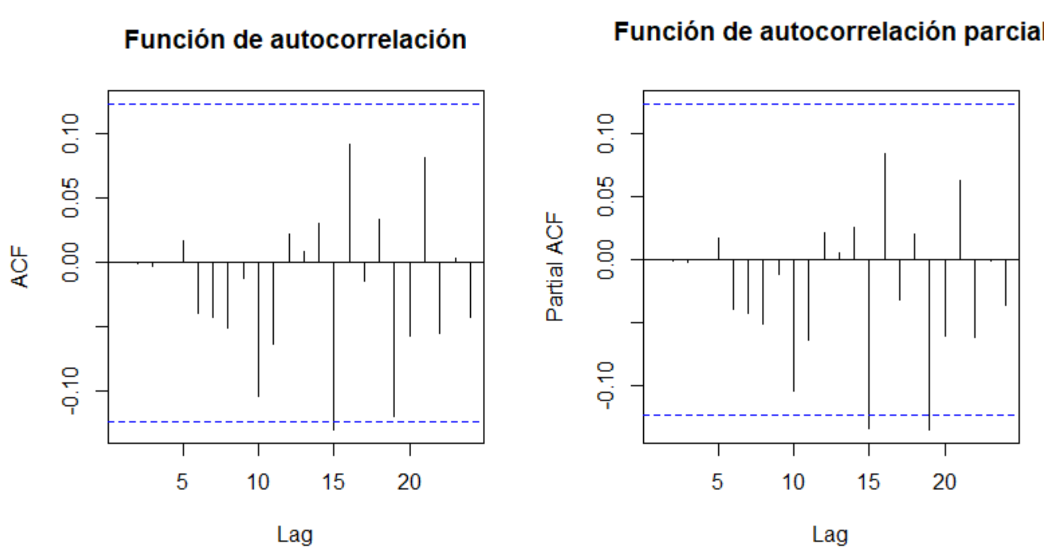
Gráfica 5.
Gráfico de residuos



De la **gráfica 5** observamos que los residuos oscilan alrededor del 0 y parecen tener varianza constante.

También graficamos su acf y pacf para observar su comportamiento:

Gráfica 6.



La **gráfica 6** parece sugerir un comportamiento de ruido blanco, puesto que no parece haber correlación entre los residuos, caen dentro de las bandas de Bartlett desde el lag 0 y no hay un comportamiento o tendencia en ellos. Para comprobar lo anterior realizaremos la prueba Ljung-Box hasta el lag 20:

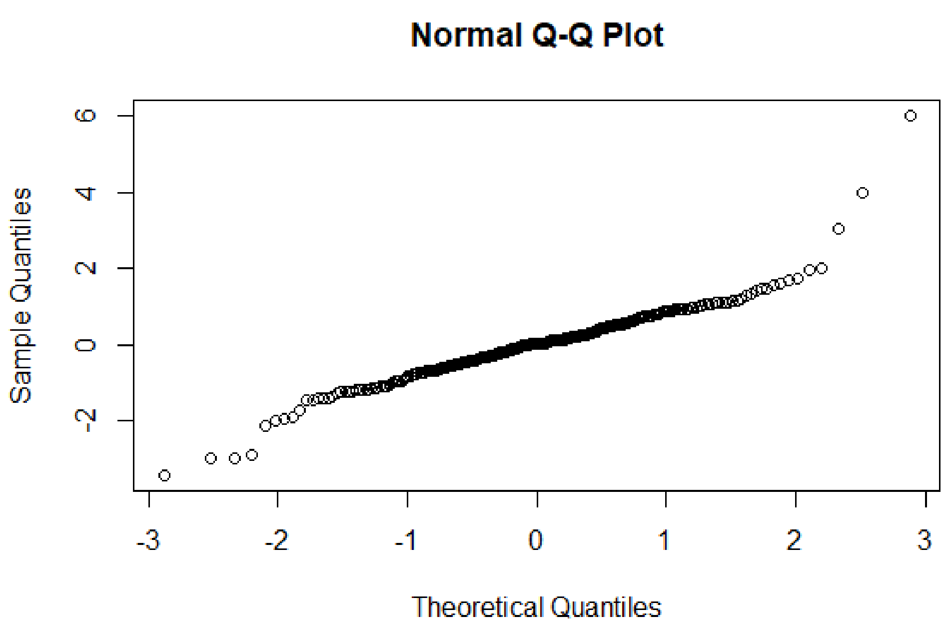
Prueba Ljung-Box

$$H_0: \rho_1 = \rho_2 = \cdots = \rho_{20} = 0$$

Al realizar la prueba obtenemos un p-valor de $0.5852 > 0.05$ por lo que no se rechaza la hipótesis nula, es decir, con un nivel de significancia del 5% no hay evidencia suficiente para rechazar que los residuales se comporten como ruido blanco.

Ahora necesitamos verificar que los residuos tienen una distribución normal para poder demostrar que se tratan de ruido blanco Gaussiano, para ello primero realizaremos un QQ-PLOT como se observa en la **gráfica 7**.

Gráfica 7.



El **gráfico 7** debería mostrarnos una línea recta, observamos que sí parece tener esa tendencia, pero lo verificaremos con una prueba de hipótesis para la normalidad. Los residuos se estandarizan y verificaremos si tienen la distribución de una normal estándar.

Prueba de normalidad Kolmogorov-Smirnov

Sea ε'_t los residuos estandarizados

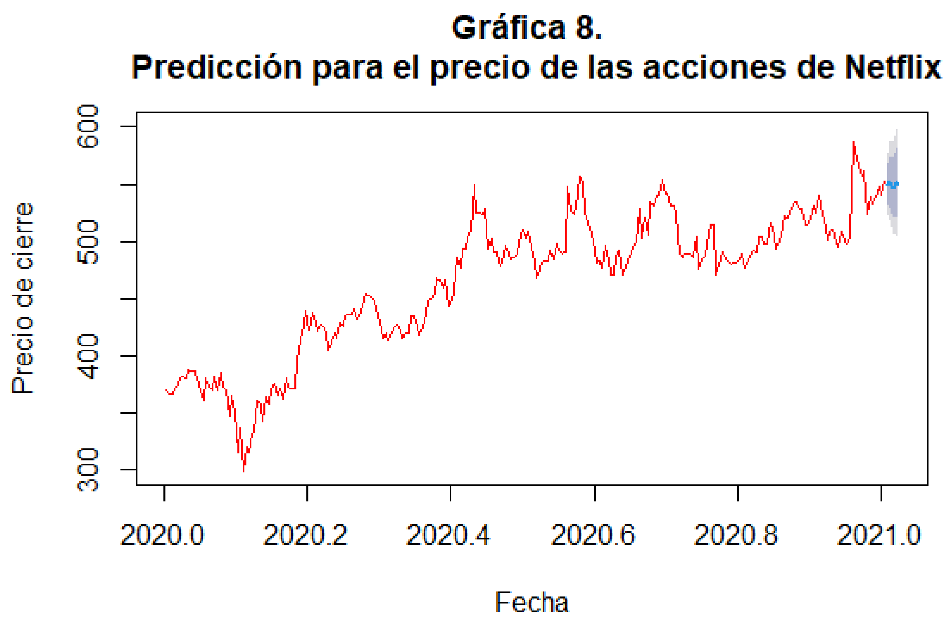
$$H_0: \varepsilon'_t \sim N(0,1)$$

Al realizar la prueba se obtuvo p-valor de $0.2644 > 0.05$ por lo que no se rechaza la hipótesis nula. Es decir, con un nivel de significancia del 5% no existe evidencia suficiente para rechazar que los residuos tengan una distribución normal.

Ya que hemos demostrado que los residuos tienen una distribución normal y son no correlacionados podemos concluir que se tratan de un ruido blanco Gaussiano.

Pronósticos

A continuación, utilizaremos el modelo que hemos estimado para realizar predicciones 5 días hacia adelante, para ello utilizaremos la paquetería "forecast" del software R. Obtenemos lo siguiente:



De la **gráfica 8** podemos observar el rango de valores en el que esperamos que se encuentren los precios de cierre de las acciones de Netflix en los siguientes 5 días, utilizando intervalos de confianza del 95% y 80% como se observa en la siguiente tabla:

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2021.008	549.4140	532.0069	566.8212	522.7921	576.0360
2021.012	552.7111	530.3291	575.0931	518.4808	586.9415
2021.016	546.6255	519.9185	573.3326	505.7806	587.4705
2021.020	551.4322	521.7046	581.1599	505.9677	596.8968
2021.024	548.8670	515.1987	582.5353	497.3757	600.3583

Una estrategia que podríamos realizar sabiendo lo anterior sería la siguiente:

Dado que el último valor conocido fue el del 4 de febrero de 2021 y fue de $552.16, dentro de 5 días, usando un intervalo de confianza del 80% el precio de cierre se espera se encuentre entre $515.19 y $582.53. Estos últimos valores podríamos utilizarlos como nuestras señales de compra y de venta, pues nos dicen cuales son los valores que nosotros estimamos como mínimos y máximos de la serie para el período.

Por ejemplo, si ya contábamos con una acción el 4 de febrero, entonces si el precio de cierre llegase a superar los $582.53 en los siguientes días entonces nuestro modelo nos indica que deberíamos vender nuestra acción.

Es importante notar que las decisiones que tomemos en particular en el problema que analizamos sobre el precio de las acciones de Netflix pueden depender del horizonte de inversión que estemos analizando, pues no es lo mismo observar y analizar la serie de manera diaria a hacerlo de manera semanal, mensual o por el otro lado hacerlo por hora o minuto.

En este trabajo se realizó un análisis de series de tiempo desde la construcción del modelo hasta su aplicación para pronósticos, pudimos ver su utilidad y la importancia de cada uno de sus pasos para su uso correcto.