

## ANEXO 11

# Código R-Studio “modelo estadístico para predecir el índice COLCAP”

María Camila Cardona<sup>a,c</sup>, María José Martínez<sup>a,c</sup>, Paula Andrea Bravo<sup>a,c</sup>, Carlos Andrés Hernández<sup>a,c</sup>

Isabel Cristina García<sup>b,c</sup>

<sup>a</sup>Estudiante de Ingeniería Industrial

<sup>b</sup>Profesora, Directora del Proyecto de Grado, Departamento de Ciencias Naturales y Matemáticas

<sup>c</sup>Pontificia Universidad Javeriana, Cali, Colombia

Una serie temporal o serie de tiempo es una secuencia ordenada de observaciones, cada una de las cuales está asociada a un momento de tiempo. Ejemplos de series temporales las podemos encontrar en el campo de la economía cuando buscamos datos para estudiar el comportamiento de una variable económica y su relación con otras a lo largo del tiempo, estos datos se presentan frecuentemente en forma de series temporales [1].

Los modelos ARIMA, son modelos paramétricos que tratan de obtener la representación de la serie en términos de la interrelación temporal de sus elementos. Este tipo de modelos que caracterizan las series como sumas o diferencias, ponderadas o no, de variables aleatorias o de las series resultantes, fue propuesto por Yule y Slutsky en la década de los 20 [1].

En el siguiente documento se desarrollará el modelo empleado en el software R-Studio, por medio del cual se planteó, análisis y desarrollo el objetivo esperado, de establecer una predicción del índice COLCAP a través de la base de datos establecida.

- **Supuestos desarrollados para el modelo arima:**

```
mr<-modelo$residuals # estos son los residuos del modelo
```

```
t.test(mr, alternative='two.sided',
```

```
conf.level=0.95, mu=0)
```

```
##
```

```
## One Sample t-test
```

```
##
```

```
## data: mr
```

```
## t = 0.40508, df = 730, p-value = 0.6855
```

```
## alternative hypothesis: true mean is not equal to 0
```

```
## 95 percent confidence interval:
```

```
## -1.534945 2.333044
```

```
## sample estimates:
```

```
## mean of x
```

```
## 0.3990496
```

Se concluye el supuesto de no rechazar la hipótesis nula, es decir, que la media de los residuos es cero.

- **Supuesto de independencia de los residuos:**

Se verifica el supuesto que asume que los residuos que no están autocorrelacionados son independientes. Para validar la independencia de los residuos se usó la prueba Ljung-Box:

```
independencia <- Box.test(mr, type="Ljung-Box") # Test de Ljung-Box
```

```
independencia$p.value
```

```
## [1] 0.9902458
```

Al desarrollar la prueba anterior se concluye que son independientes, por consecuencia no existe dependencia entre los residuos del modelo presentado.

- **Supuesto de distribución:**

Establecer qué tipo de distribución se ajusta frente a los datos que se tienen:

```
qqnorm(mr, col = "blue")
```

```
qqline(mr, col = "red")
```

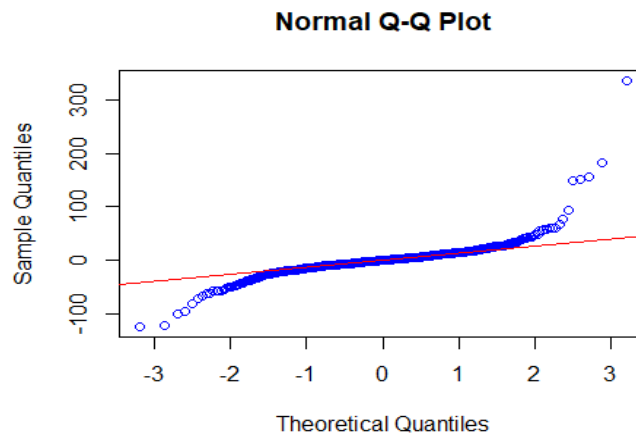


Fig. 208. Prueba para validar el supuesto de distribución normal

- Prueba Shapiro-Wilk, con el fin de contrastar la normalidad del conjunto de datos.

```
shapiro.test(mr)
```

```
##
```

```
## Shapiro-Wilk normality test
```

```
##
```

```
## data: mr
```

```
## W = 0.72477, p-value < 2.2e-16
```

```
library(car)
```

```
## Warning: package 'car' was built under R version 4.1.1
```

```
## Loading required package: carData
```

```
qqPlot(mr)
```

```
## [1] 375 376
```

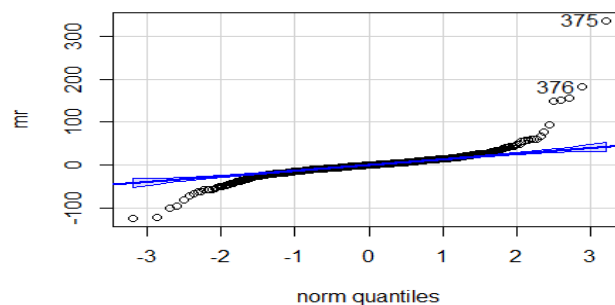


Fig. 209. Prueba de Shapiro-Wilk (normalidad)

Se obtiene que la hipótesis de normalidad se rechaza.

- **Prueba de t-distribution o t-student:**

```
library(fitdistrplus)
```

```
library(moments)
```

```
library(MASS)
```

```
library(fGarch)
```

```
summary(mr)
```

```
##   Min.   1st Qu.   Median   Mean   3rd Qu.   Max.
```

```
## -124.2791 -8.9308   0.1179  0.3990   8.9929 336.0226
```

```
skewness(mr)
```

```
## [1] 3.457803
```

```
## attr(,"method")
```

```
## [1] "moment"
```

```
kurtosis(mr)
```

```
## [1] 41.93158
```

```
## attr(,"method")
```

```
## [1] "excess"
```

```
ec<-ecdf(mr)
```

```
plot(ec)
```

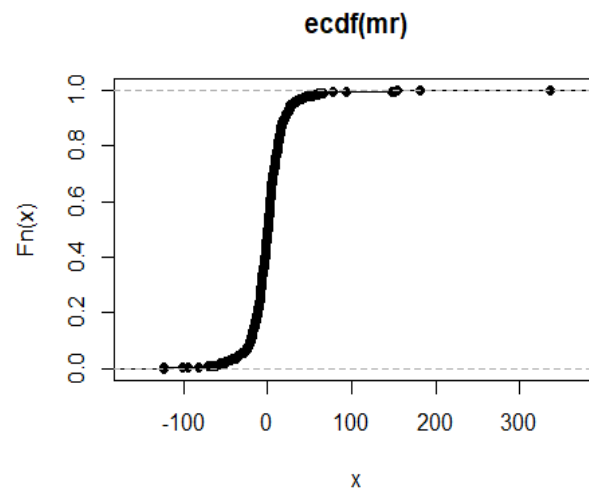


Fig. 210. Grafica de distribución de ajuste (t-distribution)

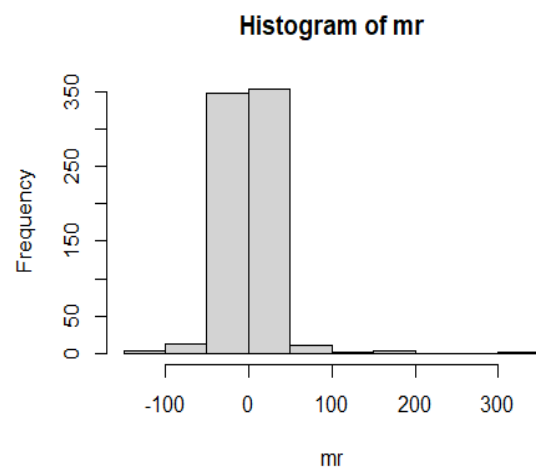


Fig. 211. Histograma de distribución de ajuste (t-distribution)

```
mr<-as.numeric(mr)
```

```
#descdist(mr)
```

```
stdFit(mr)
```

```
## $par
```

```
##      mean      sd      nu
```

```
## -0.1702331 38.3591677 2.1941099
```

```
##
```

```
## $objective
```

```
## [1] 3176.056
```

```
##
```

```
## $convergence
```

```
## [1] 0
```

```
##
```

```
## $iterations
```

```
## [1] 27
```

```
##
```

```
## $evaluations
```

```
## function gradient
```

```
##      34      108
```

```
##
```

```
## $message
```

```
## [1] "relative convergence (4)"
```

```
#ajust<-fitdistr(mr, "t", start = list(m=mean(mr),s=sd(mr), df=2.2), lower=c(-1, 0.001,1))
```

```
#summary(ajust)
```

```
$(ajust$estimate)
```

```
length(mr)
```

```
## [1] 731
```

- Finalmente se vuelve a generar una función de distribución empírica que sirve para ver si se ajusta la t-student

```
x <- seq( -100, 100, by = 0.01)
```

```
ed<-ecdf(rt(x,df=2.2))
```

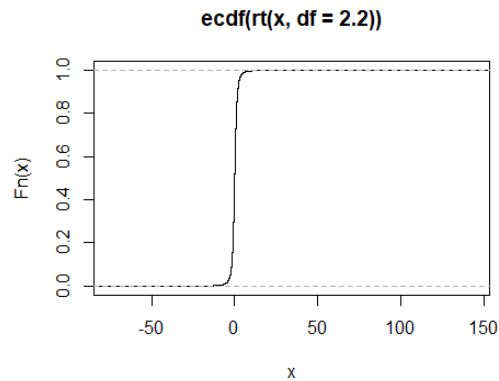


Fig. 212. Gráfico de distribución empírica (t-student)

### PASOS EN R DESARROLLADOS PARA PREDECIR EL ÍNDICE COLCAP:

Se debe desarrollar una serie de pasos para finalmente poner en marcha el modelo planteado y para el modelo ejecutado se requirieren algunas librerías.

- **Librerías para que las funciones se ejecutarán correctamente a través del software:**

*library(forecast)*

*library(tseries)*

*library(timsac)*

*library(ggplot2)*

*library(changepoint)*

*library(readxl)*

*library(readxl)*

- **Se hace uso de la base de datos obtenida, inicia el día 84, 2019:**

*Datos<- read\_excel("C:/Users/portatil/Desktop/2021-2/PD2/Bolsa/Datos históricos COLCAP.xlsx")*

*View(Datos)*

*Indice.ts<-ts(Datos\$COLCAP, start = c(2019,84), frequency = 365)*

*(Indice.ts)*

*## Time Series:*

*## Start = c(2019, 84)*

*## End = c(2021, 85)*

## Frequency = 365

- Finalmente, a la información de entrada agregamos el gráfico:

plot(Indice.ts)

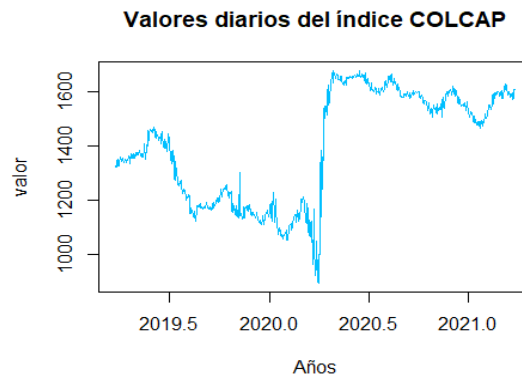


Fig. 213. Gráfico de serie de tiempo del índice COLCAP en R-Studio

- Se establecen los componentes de la serie temporal, verificando que sea la serie sea estacionaria (Dicker-Fuller 'adf.test'):

```
adf.test(Indice.ts,, alternative = c("stationary", "explosive"))
```

```
##
```

```
## Augmented Dickey-Fuller Test
```

```
##
```

```
## data: Indice.ts
```

```
## Dickey-Fuller = -2.4885, Lag order = 9, p-value = 0.3715
```

```
## alternative hypothesis: stationary
```

- Para conseguir estacionariedad:

```
ndiffs(Indice.ts)
```

```
## [1] 1
```

nos dice que 1 así que diferenciamos una vez y la llamamos dif.Indice.ts

```
dif.Indice.ts<-diff(Indice.ts)
```

```
#la graficamos
```

```
plot(dif.Indice.ts, main=" ", ylab="valor", col="deepskyblue", xlab="Años")
```

*title(main="DIF Precios del índice COLCAP")*

- Se presenta un gráfico de diferencia:

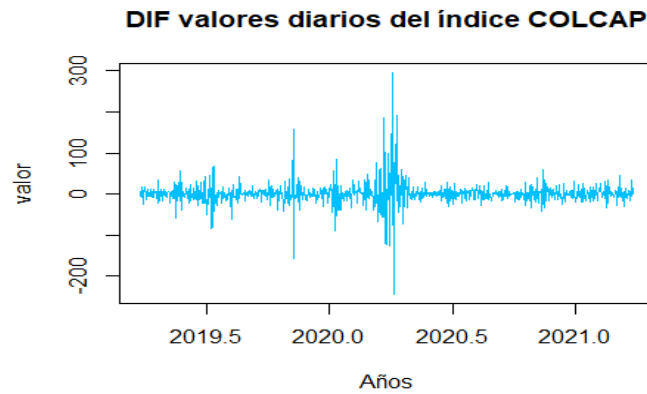


Fig. 214. Gráfico de diferencia en diferencia del valor del índice COLCAP en R-Studio

- Se verifica nuevamente con la función anterior y proceder a un pronóstico acertado:

```
adf<-adf.test(dif.Indice.ts)
```

```
## Warning in adf.test(dif.Indice.ts): p-value smaller than printed p-value
```

```
adf
```

```
##
```

```
## Augmented Dickey-Fuller Test
```

```
##
```

```
## data: dif.Indice.ts
```

```
## Dickey-Fuller = -6.6882, Lag order = 9, p-value = 0.01
```

```
## alternative hypothesis: stationary
```

- Se realiza un gráfico de rezagos en la serie:

```
lag.plot(Indice.ts, 9, do.lines = FALSE)
```

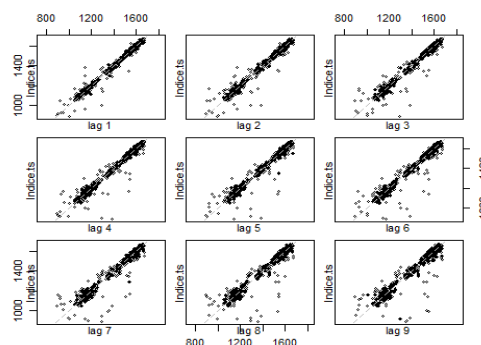


Fig. 215. Gráfico de rezagos en la serie de valores diarios del índice COLCAP



- Se procede a obtener las autocorrelaciones de la serie obtenida en donde a través de R se calcula por medio de:

```
library(ggplot2)
```

```
ACF<-acf(dif.Indice.ts)
```

- Función de autocorrelación simple de una serie temporal (`acf()`), proporciona la autocorrelación en todos los retrasos posibles.

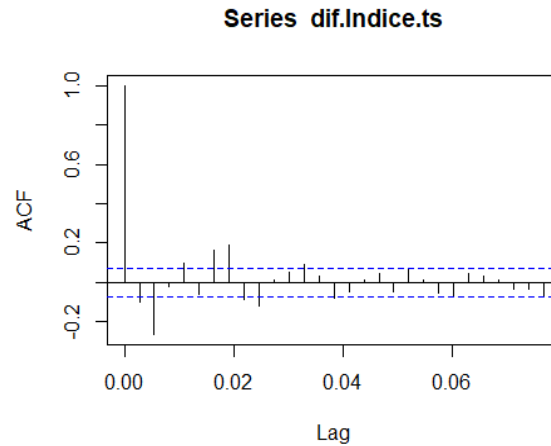


Fig. 216. Función de autocorrelación simple de una serie temporal

- Así mismo se aplica `pacf()` la función de autocorrelación parcial:

```
PACF<-pacf(dif.Indice.ts)
```

- Grafica de autocorrelación generada:

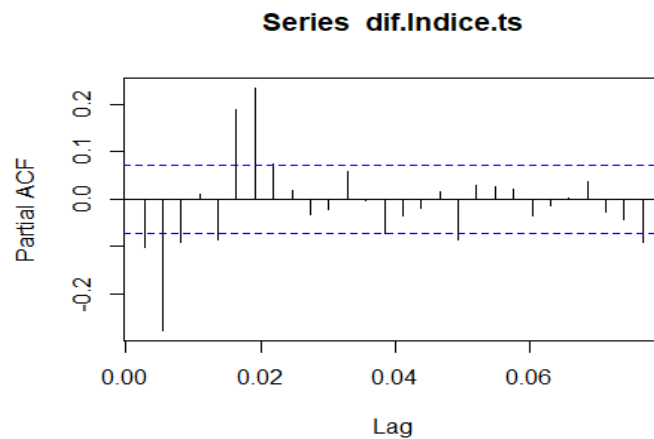


Fig. 217. Función de autocorrelación parcial

```
autoplot(stl(dif.Indice.ts, s.window = "periodic"), ts.colour = "blue")
```

- **Obtener los componentes de la serie temporal:**

```
plot(stl(dif.Indice.ts, s.window = "periodic"))
```

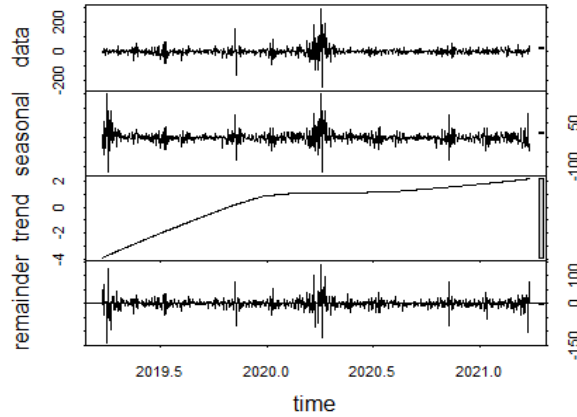


Fig. 218. Componentes de la serie temporal del índice COLCAP

- **Modelo Arima en R:**

```
modelo<-auto.arima(dif.Indice.ts)
```

```
modelo
```

```
## Series: dif.Indice.ts
```

```
## ARIMA(2,0,4) with zero mean
```

```
##
```

```
## Coefficients:
```

```
##      ar1      ar2      ma1      ma2      ma3      ma4
```

```
##      0.2857 0.2404 -0.4829 -0.4885 0.1161 0.3220
```

```
## s.e. 0.1295 0.1100 0.1242 0.1265 0.0457 0.0497
```

```
##
```

```
## sigma^2 estimated as 714.5: log likelihood=-3436.49
```

```
## AIC=6886.97 AICc=6887.13 BIC=6919.13
```

En este caso el modelo ajustado que R-Studio retorna es un ARIMA (2,1,4)

- **Función para detectar un punto de cambio en la media que se desarrolló:**

```
mval<-cpt.mean(dif.Indice.ts,method = "AMOC")
```

```
cpts(mval)
```

## [1] 372

```
plot(mval, type = "l", cpt.col = "blue", xlab = "Value", cpt.width = 4, main = "default penalty")
```

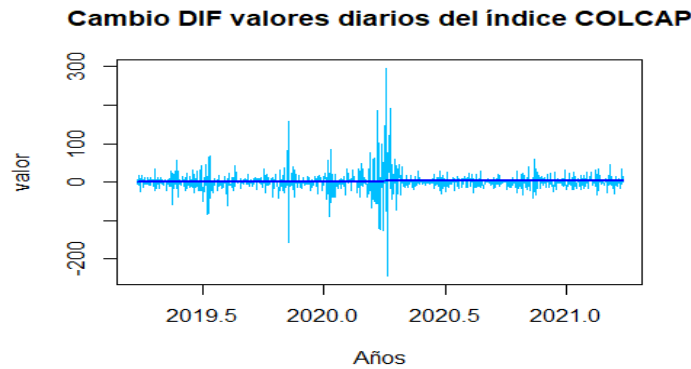


Fig. 219. Gráfica de puntos de cambio en R-Studio

- **Predecir 12 días utilizando un modelo ARIMA:**

```
pred<-forecast(dif.Indice.ts,h=12)
```

```
pred
```

```
plot(pred, main=" ", ylab="valor", col="deepskyblue", xlab="Años")
```

```
title(main="Predicción DIF Precios del índice COLCAP")
```

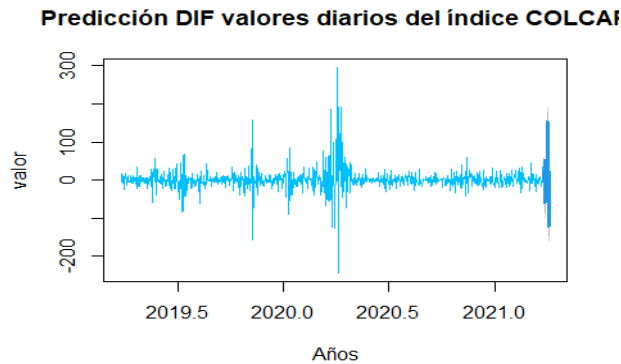


Fig. 220. Predicción del modelo para el índice COLCAP

Para validar el pronóstico entregado por R-Studio, y la veracidad de los valores de diferencia del modelo, se busca analizar los puntos críticos de cambios: `library("tsoutliers")` [5].

- **Puntos críticos de cambio:**

```
dat.ts<- ts(dif.Indice.ts,frequency=1)
```

```
data.ts.outliers <- tso(dat.ts)
```

*data.ts.outliers*

*## Series: dat.ts*

*## Regression with ARIMA(2,0,2) errors*

*##*

*## Coefficients:*

*## ar1 ar2 ma1 ma2 TC227 AO229 TC360 AO363*

*## 0.9310 -0.6006 -1.2909 0.6401 72.1572 -190.9670 -51.9647 141.8149*

*## s.e. 0.0841 0.0446 0.0973 0.0704 12.6956 26.0103 11.8287 23.5220*

*## TC367 AO369 AO375 AO382*

*## -68.3681 128.8834 303.0247 238.7274*

*## s.e. 13.3930 27.6457 19.2944 17.7135*

*##*

*## sigma^2 estimated as 451.8: log likelihood=-3265.93*

*## AIC=6557.86 AICc=6558.37 BIC=6617.59*

*##*

*## Outliers:*

*## type ind time coefhat tstat*

*## 1 TC 227 227 72.16 5.684*

*## 2 AO 229 229 -190.97 -7.342*

*## 3 TC 360 360 -51.96 -4.393*

*## 4 AO 363 363 141.81 6.029*

*## 5 TC 367 367 -68.37 -5.105*

*## 6 AO 369 369 128.88 4.662*

*## 7 AO 375 375 303.02 15.705*

*## 8 AO 382 382 238.73 13.477*

*plot(data.ts.outliers)*

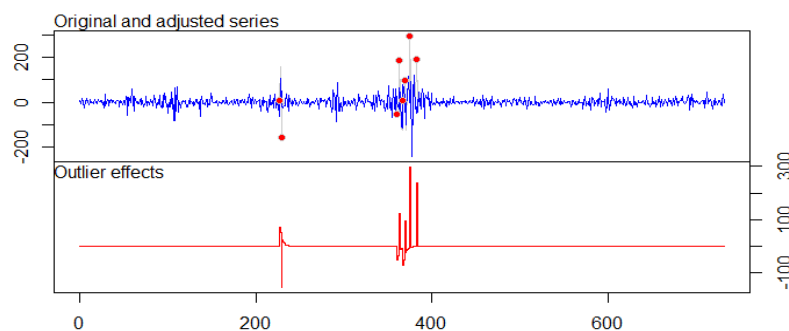


Fig. 221. Validación de los puntos de cambio

Finalmente se validó también a través de la predicción, desde el 26 de marzo del 2021 al 6 de abril del 2021, con el fin de poder realizar una comparación entre el modelo generado a través del programa R-Studio y la base de datos real presentada en el Banco de la república:

TABLA XXXII.  
 INTERVALOS DE CONFIANZA DEL ÍNDICE COLCAP EN R-STUDIO

#	Fecha	Pronóstico índice COLCAP generado por R-Studio					Índice COLCAP real	
		Forecast	Lo 80	Hi 80	Lo 95	Hi 95	Valor	Diferencia
0	25/03/2021						1,320	
1	26/03/2021	1	- 25,489,538	27,365,473	- 3,947,941,437	4,135,535	1,314	- 6
2	27/03/2021	19	- 7,410,854	45,444,157	- 2,140,073,039	5,943,403	1,309	- 6
3	28/03/2021	55	28,555,035	81,410,046	1,456,515,816	9,539,992	1,303	- 6
4	29/03/2021	-61	- 87,213,057	- 34,358,045	- 10,120,293,351	- 2,036,817	1,321	18
5	30/03/2021	-21	- 47,271,591	5,583,421	- 6,126,146,727	1,957,330	1,327	6
6	31/03/2021	-1	- 27,679,330	25,175,682	- 4,166,920,660	3,916,556	1,317	- 11
7	1/04/2021	64	37,442,144	90,297,157	2,345,226,748	10,428,703	1,318	1
8	2/04/2021	3	- 23,444,404	29,410,609	- 3,743,428,071	4,340,049	1,316	- 2
9	3/04/2021	154	127,767,740	180,622,753	11,377,786,366	19,461,263	1,314	- 2
10	4/04/2021	40	13,958,962	66,813,975	-0.03091469	8,080,385	1,312	- 2
11	5/04/2021	-123	- 149,329,702	- 96,474,689	- 16,331,957,913	- 8,248,481	1,324	12
12	6/04/2021	24	- 2,667,634	50,187,379	- 1,665,751,129	6,417,726	1,324	0

Se observa como los datos pronosticados tienen fluctuaciones, sin embargo, estos se encuentran dentro de los intervalos de confianza que ha calculado el programa a partir del modelo.

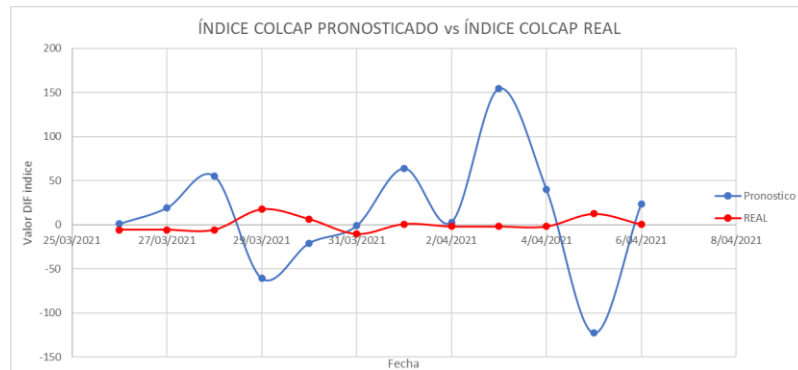


Fig. 222. Comparación entre el pronóstico y los datos reales

## REFERENCIAS

- [1]P. González, Análisis de series temporales: Modelos ARIMA. 2009.
- [2] S. Chatterjee, "Time Series Analysis Using ARIMA Model In R", DataScience+, 2021. [En línea]. Disponible: <https://datascienceplus.com>. [Accedido: 21- Oct- 2021].
- [3]E. Santana, E. Santana and V. perfil, "Supuestos en Regresión Lineal", Apuntes-r.blogspot.com, 2021. [En línea]. Disponible: <http://apuntes-r.blogspot.com/2015/04/supuestos-en-regresion-lineal.html>. [Accedido: 21- Oct- 2021].
- [4] Sergas.es, 2021. [En línea]. Disponible: [https://www.sergas.es/Saude-publica/Documents/1899/Ayuda\\_Epidat\\_4\\_Distribuciones\\_de\\_probabilidad\\_Octubre2014.pdf](https://www.sergas.es/Saude-publica/Documents/1899/Ayuda_Epidat_4_Distribuciones_de_probabilidad_Octubre2014.pdf). [Accedido: 21- Oct- 2021].
- [5] J. López-de-Lacalle, "Detection of Outliers in Time Series", Cran.r-project.org, 2019. [En línea]. Disponible: <https://cran.r-project.org/web/packages/tsoutliers/tsoutliers.pdf>. [Accedido: 08- Oct- 2021].