

Predicción Cuota de mercado Colgate - Crest

Sergio Casares

20/11/2020

Contents

1. Introducción al Trabajo	2
2. Importación de librerías	2
2.1. Importación y Transformación del dataset	2
2.2. Formato xts y zoo	2
3.Exploratory Data Analysis	3
4. Creación Test y Training	5
4.1. Autocorrelación y Autocorrelación parcial	5
5. Auto-Arima	6
5.1.Crest	6
5.2. Colgate	8
6. Cálculo de valores atípicos	9
7. Modelo de Intervención	10
7.1. Crest	10
7.2. Colgate	10
8. Función de transferencia	11
9. Resumen utilizando la librería Casual Impact	12
10. Bibliografía	13

1. Introducción al Trabajo

La finalidad del trabajo es realizar un análisis el cual debe incluir para las dos series la estimación de un modelo ARIMA, la detección de atípicos/outlier y un modelo de interacción. Por último se debe realizar un modelo de función de transferencia entre las dos cuotas.

Introducción histórica:

Procter y Gamble introdujeron Crest en el mercado estadounidense en 1956. Antes de ese momento, Colgate disfrutaba de un liderazgo de mercado con una participación de mercado cercana al 50%. Durante los siguientes 4 años, Colgate siguió siendo un competidor dominante y Crest solo logró una participación de mercado relativamente modesta pero estable del 15%. Sin embargo, el 1 de agosto de 1960, el Consejo de Terapéutica Dental de la American Dental Association (ADA) aprobó a Crest como una “ayuda importante en cualquier programa de higiene dental”.

Mientras tanto, Procter y Gamble revitalizaron su campaña de marketing para aprovechar el respaldo de la ADA, lo que dio como resultado un salto casi inmediato en la cuota de mercado de Crest en detrimento de la de Colgate.

2. Importación de librerías

Importamos las librerías necesarias para el trabajo.

2.1. Importación y Transformación del dataset

Creamos la secuencia del periodo de trabajo y la incorporamos a nuestro dataset eliminando las columnas de año y semana.

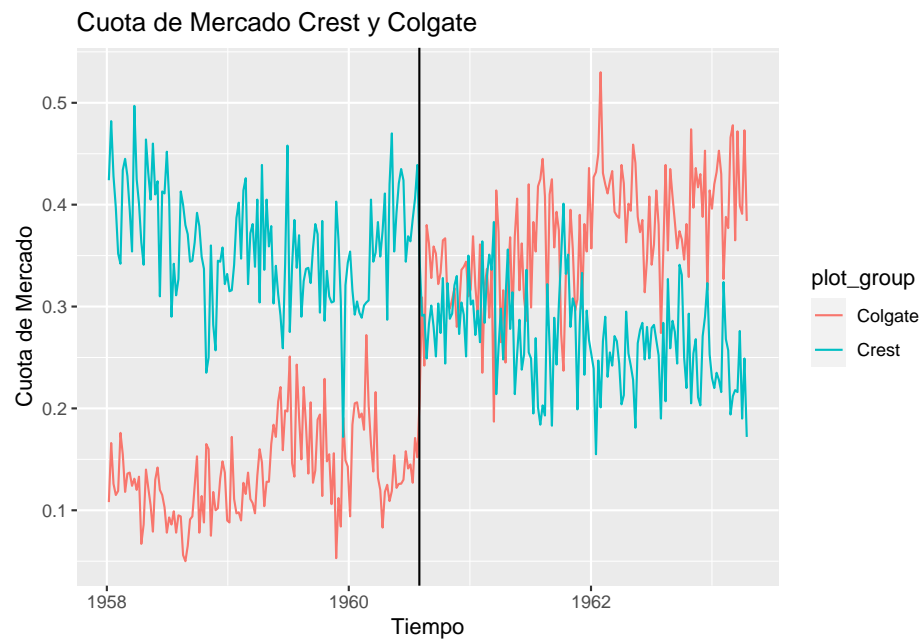
2.2. Formato xts y zoo

Transformación de datos para poder manejar mejor los datos.

3.Exploratory Data Analysis

Realizamos un análisis exploratorio con el fin de analizar la estructura de los datos.

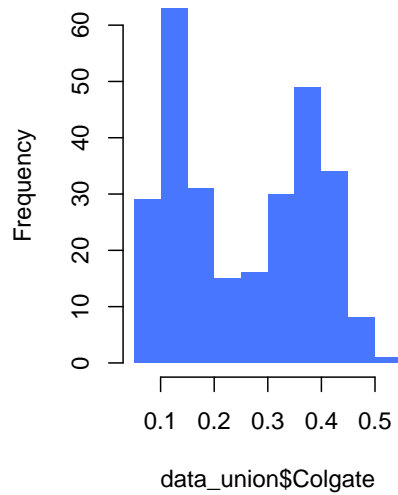
Realizamos una visualización de las series temporales de Colgate y de Crest, se observa una diferenciación hasta mediados de 1960, pero que tras un momento corto de tiempo, se produce un salto de Colgate (y una caída de Crest) que terminan convirtiéndose en escalones porque la tendencia a la inversa se prolonga en el tiempo.



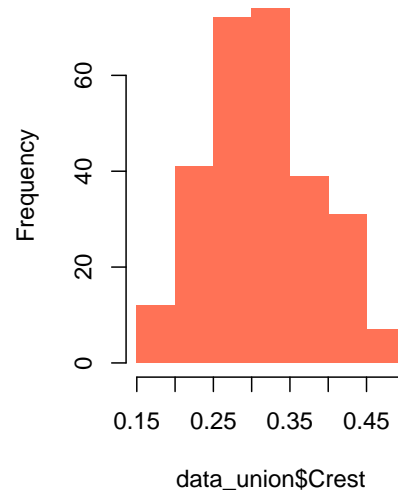
Por otra parte, podemos observar los histogramas de frecuencias de la cuota de mercado de ambas empresas. De **colgate** podemos observar cómo ha tenido dos espacios bien diferenciados, por una parte, presentaba na cuota de mercado inferior al 20% durante un periodo largo de tiempo y también ha tenido otra distribución de la cuota de mercado pero con valores superiores, de entre el 30% y 50%.

Crest, en cambio, presenta solo una distribución normal con valores entorno al 20% y al 40% debido quizás porque su caída no fue tan brusca como la subida de Colgate, por eso puede haber cierta solapación entre ambos momentos (antes y después del break point).

Histogram of data_union\$Colgat



Histogram of data_union\$Crest



```
## [1] -0.7659797
```

```
##      Index      Colgate      Crest
## Min.   :1958-01-07 Min.   :0.050 Min.   :0.1550
## 1st Qu.:1959-05-03 1st Qu.:0.131 1st Qu.:0.2607
## Median :1960-08-26 Median :0.248 Median :0.3055
## Mean   :1960-08-26 Mean   :0.255 Mean   :0.3127
## 3rd Qu.:1961-12-20 3rd Qu.:0.373 3rd Qu.:0.3563
## Max.   :1963-04-16 Max.   :0.530 Max.   :0.4970
```

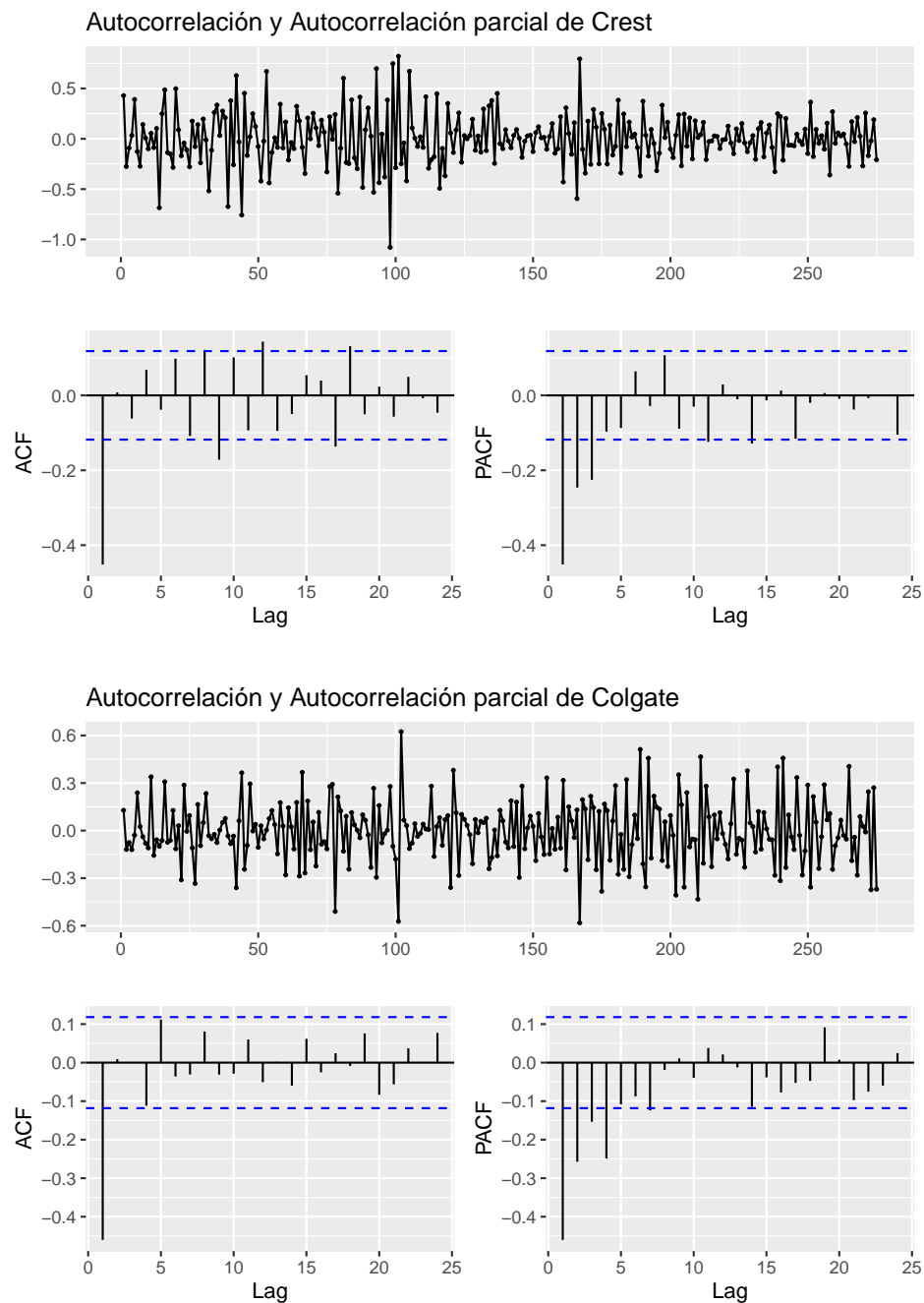
La correlación entre ambas empresas es del -0.77% y además se presenta un resumen de los valores de cada una de las series temporales

4. Creación Test y Training

Para nuestro estudio, dividimos la muestra en Training y Test, apartando del modelo las últimas 16 semanas.

4.1. Autocorrelación y Autocorrelación parcial

Se transforman la varianza y la media para convertir a la serie en estacionaria para garantizar la buena estimación de la predicción. La varianza se realiza mediante la toma de logaritmos y la media mediante la diferenciación (establecer tantos períodos de retardos como sean necesarios). Se ve una correlación parcial a 3-4 semanas en ambos casos (quizá correlación a un mes)



5. Auto-Arima

Es un modelo estadístico que utiliza variaciones y regresiones de datos estadísticos con el fin de encontrar patrones para una predicción hacia el futuro. Se trata de un modelo dinámico de series temporales, es decir, las estimaciones futuras vienen explicadas por los datos del pasado y no por variables independientes.

Para nuestro trabajo, utilizaremos la función `auto.arima` para que nos calcule el modelo tipo de modelo.

El modelo resultante es un $ARIMA(0,1,1)$ sin componente estacional, solo con componente regular.

- AR: Orden del autorregresivo regular $p = 0$
- Número de diferencias regulares $d = 1$
- MA: Orden de la media móvil regular $q = 1$

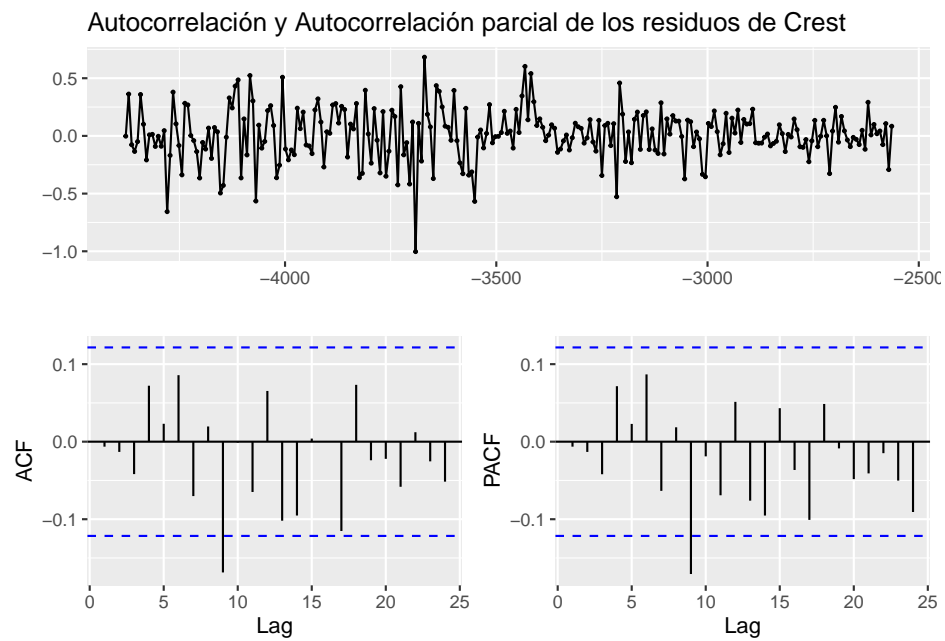
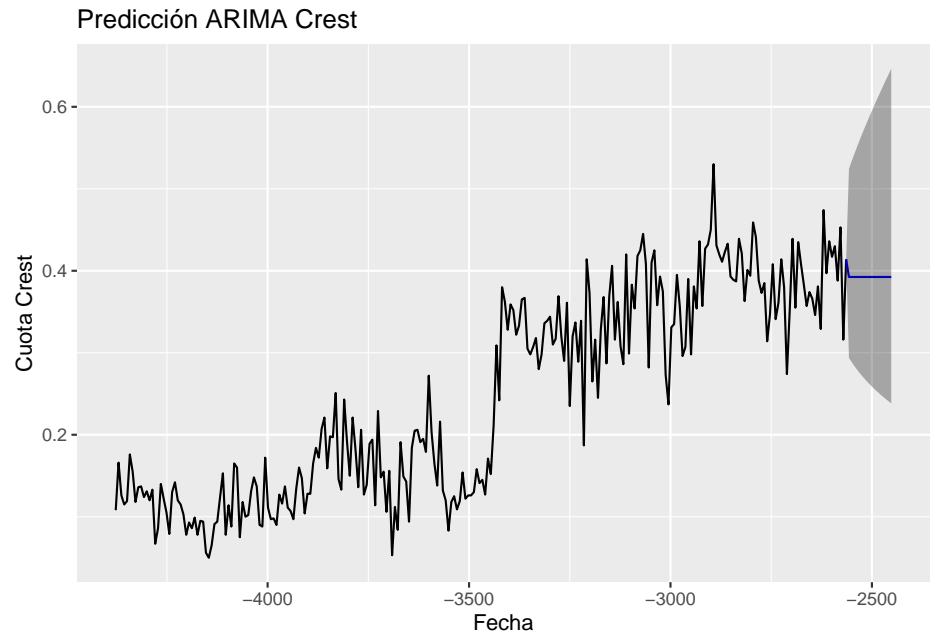
También se deben comprobar los residuos, para ver si no introducen información el modelo (como están entre las líneas azules esta se cumple) y se cumple la hipótesis del ruido blanco (media 0, varianza constante y covarianza nula).

Posteriormente, se realiza el test de Box-Ljung, el cual es un tipo de prueba estadística de si un grupo cualquiera de autocorrelaciones de una serie de tiempo son diferentes de cero. En lugar de probar la aleatoriedad en cada retardo distinto, esta prueba la aleatoriedad “en general” basado en un número de retardos. En este caso se reafirma la aleatoriedad general con un p-value del 0,77. Se rechaza la hipótesis nula; se acepta H_1 (no hay correlación), lo que quiere decir que los residuos son independientes uno del otro.

Se predicen las 16 semanas restantes y posteriormente se grafica, la predicción es plana. Debido a, probablemente, las continuas variaciones de la compañía.

5.1.Crest

```
## Series: training_crest
## ARIMA(0,1,1)
## Box Cox transformation: lambda= 0
##
## Coefficients:
##          ma1
##        -0.6377
## s.e.      0.0471
##
## sigma^2 estimated as 0.05107:  log likelihood=17.94
## AIC=-31.89  AICc=-31.84  BIC=-24.77
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 0.005791997 0.04579936 0.03497255 -1.357479 17.19282 0.1426577
##              ACF1
## Training set -0.04351554
```

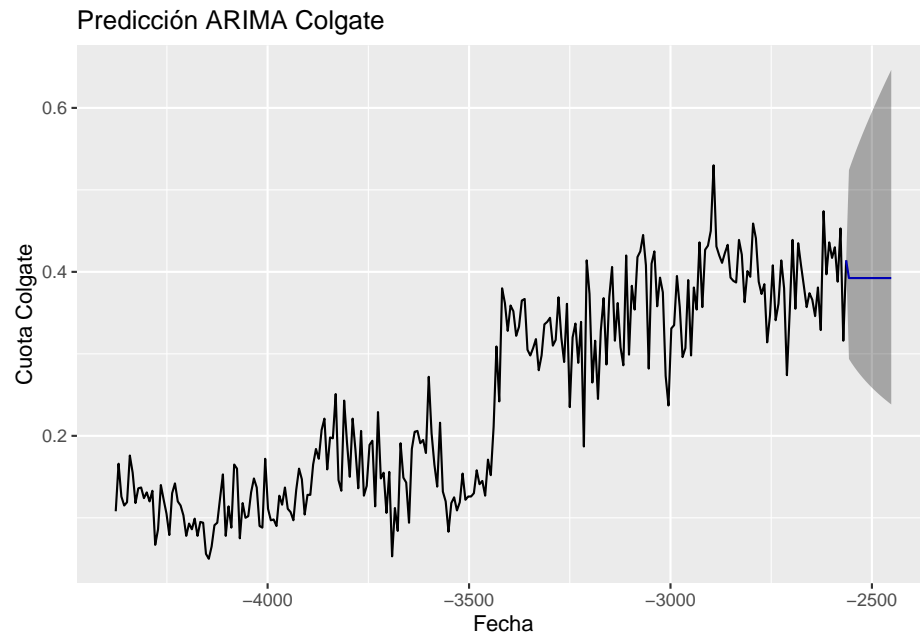


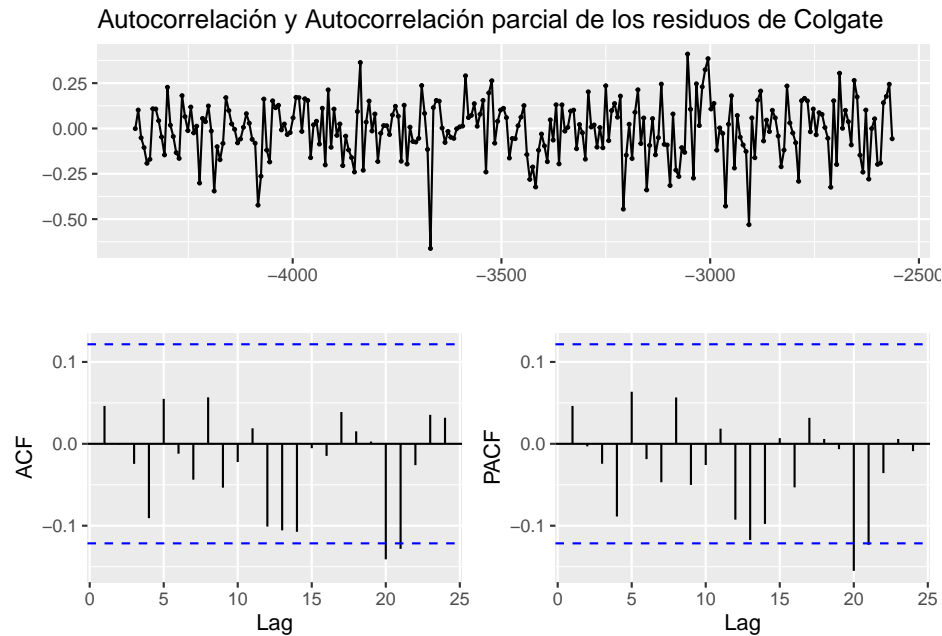
```
##
## Box-Ljung test
##
## data:  arima_crest$residuals
## X-squared = 0.51958, df = 2, p-value = 0.7712
```

5.2. Colgate

La interpretación de Colgate es la misma que la de Crest para el modelo Arima.

```
## Series: training_colgate
## ARIMA(0,1,1)
## Box Cox transformation: lambda= 0
##
## Coefficients:
##          ma1
##        -0.7688
## s.e.      0.0480
##
## sigma^2 estimated as 0.02572:  log likelihood=106.6
## AIC=-209.2   AICc=-209.15   BIC=-202.08
##
## Training set error measures:
##              ME          RMSE          MAE          MPE          MAPE          MASE
## Training set 0.0001499824 0.04794299 0.03787449 -2.139878 12.73175 0.119219
##              ACF1
## Training set 0.06055707
```





```
##
## Box-Ljung test
##
## data: arima_colgate$residuals
## X-squared = 0.72137, df = 2, p-value = 0.6972
```

6. Cálculo de valores atípicos

En series temporales los valores atípicos o outliers son:

Aditivos (AO): afecta a la serie temporal Innovativos (IO): afecta al error

AO_Crest = día 136

IO_Crest = día 99

IO_Colgate = día 102

Se detectan atípicos aditivos (que afectan a la serie temporal) e innovativos (que afectan al error de la serie).

En el caso de **Crest** se detecta un aditivo la semana posterior a la decisión de la ADA, y un innovativo a finales del año 1959 (semana 99). Quizá provocada por las expropiaciones realizadas por el Gobierno cubano con la entrada de Fidel Castro en el poder.

En **Colgate** ocurre lo mismo en el lado innovativo, porque aditivo no existe. El atípico se presenta en la semana 102 y puede estar provocada por lo mismo (Cuba). Sin embargo, se presentan muchas dudas pues ante esto se espera un efecto “más largo” pues apenas dura 1 semana. Luego ambas cuotas se recuperan

```
##           [,1]
## ind      136.000000
## lambda2   3.733484

## [1] "No AO detected"
```

```
##           [,1]
## ind      99.000000
## lambda1 -4.753225
```

```
##           [,1]
## ind     102.000000
## lambda1 -4.275533
```

7. Modelo de Intervención

Realizaremos un modelo de Intervención para poder comprobar el efecto que tuvo la ADA en la cuota de mercado de ambas empresas.

Siendo obvio por lo visto antes que la ADA influyó se busca ver como influyó si mediante un impulso (transitorio) o un escalón (permanente). Además se incluyen los anómalos aditivos (en xreg) e innovativos (en io) para ver su significación.

En ambos casos, afecta la intervención de la ADA y el atípico innovativo. Se descarta la idea del impulso y se reafirma la idea del escalón

7.1. Crest

```
##
## Call:
## arimax(x = training_crest_log, order = c(0, 1, 1), xreg = data.frame(Agosto_60_sem2 = 1 *
##   (seq(training_crest_log) == 136)), method = "ML", io = c(99), xtransf = data.frame(Agosto_60_step = 1 *
##   (seq(training_crest_log) > 135), Agosto_60_pulse = 1 * (seq(training_crest_log) ==
##   135)), transfer = list(c(0, 0), c(0, 0)))
##
## Coefficients:
##          ma1 Agosto_60_sem2      IO.99 Agosto_60_step-MA0 Agosto_60_pulse-MA0
##        -0.7358        -0.0339  -1.0255          0.7827          0.3672
## s.e.    0.0463          0.2087   0.2089          0.1628          0.2088
##
## sigma^2 estimated as 0.04344:  log likelihood = 38.27,  aic = -66.53
```

7.2. Colgate

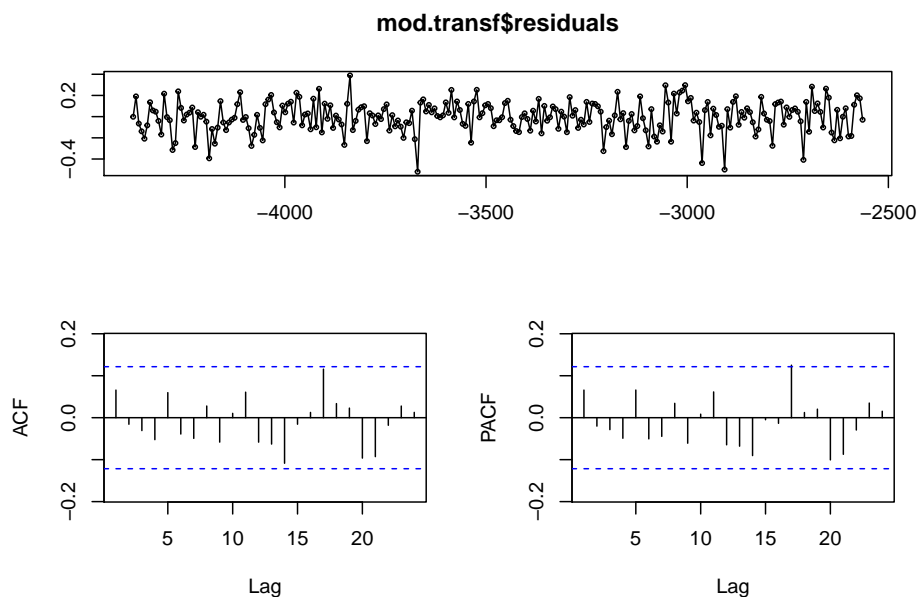
```
##
## Call:
## arimax(x = training_colgate_log, order = c(0, 1, 1), method = "ML", io = c(102),
##   xtransf = data.frame(Agosto_60_step = 1 * (seq(training_colgate_log) > 135),
##     Agosto_60_pulse = 1 * (seq(training_colgate_log) == 135)), transfer = list(c(0,
##     0), c(0, 0)))
##
## Coefficients:
##          ma1      IO.102 Agosto_60_step-MA0 Agosto_60_pulse-MA0
##        -0.8105  -0.6619        -0.3330        -0.1364
## s.e.    0.0445   0.1512          0.0939          0.1515
##
## sigma^2 estimated as 0.02286:  log likelihood = 121.27,  aic = -234.55
```

8. Función de transferencia

Se estudia el impacto de la decisión de la ADA en Colgate, el orden es el ARIMA de Colgate y el transfer es 0,0 pues el impacto es de impulso y se mantiene en el tiempo. ζ

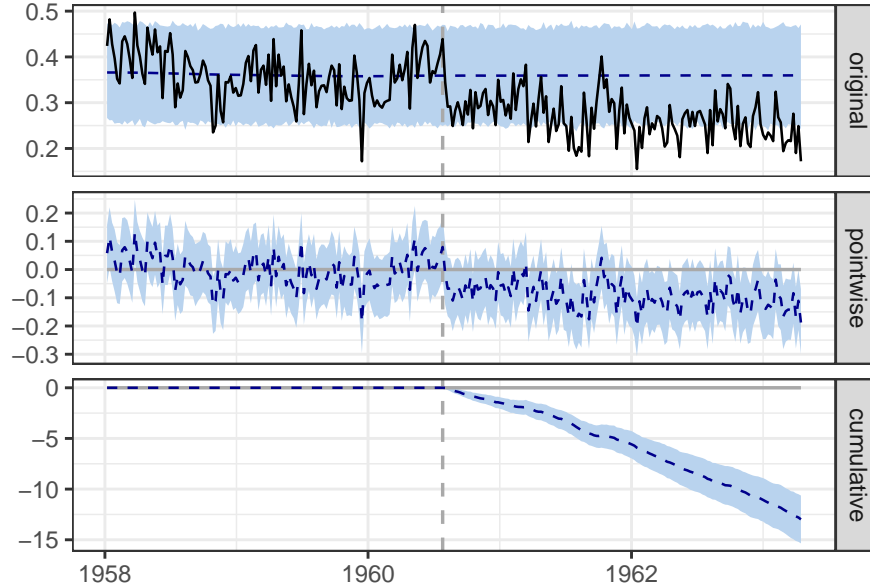
Posteriormente, se comprueba la irrelevancia de los residuos a no sobrepasar las bandas azules ninguno de los valores.

```
##
## Call:
## arimax(x = training_colgate_log, order = c(0, 1, 1), include.mean = TRUE, method = "ML",
##       xtransf = training_crest_log, transfer = list(c(0, 0)))
##
## Coefficients:
##          ma1  Cuota Crest-MA0
##       -0.8134        -0.2746
## s.e.   0.0432         0.0389
##
## sigma^2 estimated as 0.02154:  log likelihood = 128.94,  aic = -253.88
```



9. Resumen utilizando la librería Casual Impact

Con esta librería, podemos obtener un informe sobre el impacto de la actuación de ADA en las cuotas de mercado.



During the post-intervention period, the response variable had an average value of approx. 0.27. By contrast, in the absence of an intervention, we would have expected an average response of 0.36. The 95% interval of this counterfactual prediction is $[0.34, 0.38]$. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is -0.091 with a 95% interval of $[-0.11, -0.075]$. For a discussion of the significance of this effect, see below.

Summing up the individual data points during the post-intervention period (which can only sometimes be meaningfully interpreted), the response variable had an overall value of 38.04. By contrast, had the intervention not taken place, we would have expected a sum of 51.03. The 95% interval of this prediction is $[48.64, 53.45]$.

The above results are given in terms of absolute numbers. In relative terms, the response variable showed a decrease of -25%. The 95% interval of this percentage is $[-30\%, -21\%]$.

This means that the negative effect observed during the intervention period is statistically significant. If the experimenter had expected a positive effect, it is recommended to double-check whether anomalies in the control variables may have caused an overly optimistic expectation of what should have happened in the response variable in the absence of the intervention.

The probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability $p = 0.001$). This means the causal effect can be considered statistically significant.

10. Bibliografía

Historia de las expropiaciones de Cuba:

- <https://www.radiotelevisionmarti.com/a/la-habana-tendria-que-pagar-unos-7-mil-millones-por-expropiaciones-a-estados-unidos-/82885.html>

Modelos de Intervención:

- <https://core.ac.uk/download/pdf/250157957.pdf>

Series Temporales:

- https://sebastiansauer.github.io/figure_sizing_knitr/

Modelo ARIMA:

- <http://www.estadistica.net/ECONOMETRIA/SERIES-TEMPORALES/modelo-arima.pdf>

Prueba de Ljung-Box:

- https://es.wikipedia.org/wiki/Prueba_de_Ljung-Box

Estructuración del R markdown:

- https://sebastiansauer.github.io/figure_sizing_knitr/