



# **MAESTRÍA EN EXPLOTACIÓN DE DATOS Y GESTIÓN DEL CONOCIMIENTO**

## **ANÁLISIS DE SERIES TEMPORALES**

### **TRABAJO PRÁCTICO N°1**

**ALUMNOS:**     **DEL VILLAR, JAVIER**  
**OTRINO, FACUNDO DAMIÁN**  
**PISTOYA, HAYDEÉ SOLEDAD**  
**ROJAS, MARIANO ARTURO**  
**SORZA, EDWIN ANDRÉS**  
**VAILLARD, LEANDRO CARLOS**

**FECHA:**        **31 DE JULIO DE 2021**

## Índice

Consigna.....	3
Origen de los Datos.....	3
Código Empleado para el Desarrollo del Presente Trabajo .....	3
Introducción (Punto 1) .....	4
Marco Teórico (Punto 1) .....	6
Modelo Ruido Blanco (White Noise).....	7
Modelo Autorregresivo (AR) .....	7
Modelo Medias Móviles (MA) .....	8
Modelo ARMA.....	9
Modelo ARIMA.....	9
Modelo SARIMA .....	10
Serie Faena.....	11
Análisis Series Originales (Punto 2).....	11
<b>Análisis FAS, FAC y FACP (Punto 3)</b> .....	16
<b>Modelo SARIMA (Punto 4)</b> .....	17
<b>Métricas de Performance (Punto 5)</b> .....	19
<b>Comparación con otros modelos estimados (Punto 6)</b> .....	20
<b>Análisis de Diagnóstico de las Series Temporales (Punto 7)</b> .....	22
<b>Pronósticos con el modelo Seleccionado (Punto 8)</b> .....	23
Serie Consumo .....	25
Análisis Series Originales (Punto 2).....	25
Serie Consumo .....	25
<b>Análisis FAS, FAC y FACP (Punto 3)</b> .....	28
<b>Modelo SARIMA (Punto 4)</b> .....	30
Métricas de Performance (Punto 5) .....	31
<b>Comparación con otros modelos estimados (Punto 6)</b> .....	33
<b>Análisis de Diagnóstico de las Series Temporales (Punto 7)</b> .....	35
<b>Pronósticos con el modelo Seleccionado (Punto 8)</b> .....	35
Bibliografía .....	37

## Consigna

A continuación, se presenta la consigna del presente trabajo práctico:

1. Buscar dos series de tiempo y exponer la problemática de interés analítico (es importante detallar cuáles fueron los motivos de elección y situar al lector en el contexto adecuado)
2. Graficar las series originales. Analizar si es necesario diferenciarlas para convertirlas en estacionarias. Exponer los conceptos relacionados con la estacionariedad de la serie. Utilizar como criterio un test de raíces unitarias que consideren pertinente.
3. Graficar la FAS, FAC y FACP de las mismas en un solo gráfico y analizarlas. ¿Qué puede inferir de los gráficos? Justificar con teoría.
4. Efectuar una estimación de la serie temporal con distintos órdenes de parámetros de modelos (S)ARIMA y encontrar aquel que se ajuste mejor a los datos (pueden emplear Criterios de Información y analizar significatividad individual y global de los parámetros del modelo).
5. Evaluar alguna métrica de performance teniendo en cuenta el entrenamiento del modelo entre Training Set y Testing Set.
6. Detallar una comparación con otros modelos estimados.
7. Efectuar un Análisis de Diagnóstico de las series temporales (Análisis sobre los Residuos del modelo, análisis gráfico de FAC, test de Ljung-Box, etc.).
8. Pronosticar con el modelo seleccionado para cada serie para una ventana temporal razonable, en función de la periodicidad y el comportamiento de las mismas.

## Origen de los Datos

El conjunto de datos a ser analizado proviene de la página web de datos del Ministerio de Agricultura, Ganadería y Pesca. Se puede acceder a ellos por medio del siguiente vínculo:

<https://datos.agroindustria.gob.ar/dataset/indicadores-mensuales-sector-bovino/archivo/7afe10d1-e9bc-4383-9e3c-c8066bc21f65>

## Código Empleado para el Desarrollo del Presente Trabajo

Para el desarrollo del presente trabajo, se ha denominado:

- “Serie Faena” al análisis correspondiente a la faena de ganado vacuno durante el período bajo análisis.
- “Serie Consumo” al análisis correspondiente al consumo per cápita de carne vacuna durante el período bajo análisis.

El código empleado para el desarrollo del presente trabajo se encuentra almacenado en los siguientes repositorios de Kaggle:

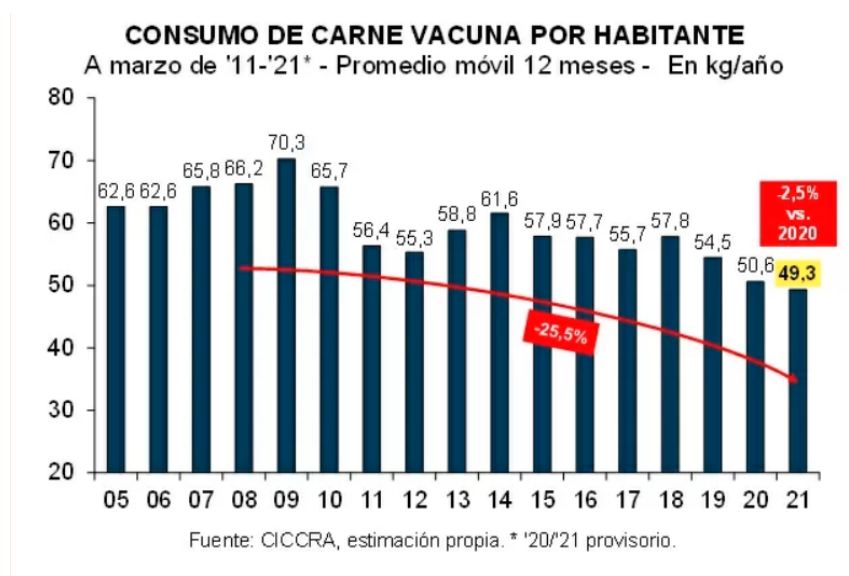
- Serie Faena: <https://www.kaggle.com/ultraxav/tp-series-temporales-carne>
- Serie Consumo: <https://www.kaggle.com/ultraxav/tp-series-temporales-carne-variable-2>

## Introducción (Punto 1)

Históricamente, Argentina ha sido reconocido como el mayor consumidor de carne vacuna (bobina) del mundo. Sin embargo, con las repetidas crisis económicas vividas por el país en los últimos años, al igual que los cambios en los hábitos alimenticios, ha llevado a poner en duda dicha reputación. Según Télam, la agencia oficial de noticias del país, el año 2020 fue el de menor consumo de carne vacuna en los últimos 100 años, mientras que la faena había tenido un crecimiento interanual del 0,6% hasta 14 millones de cabezas, el mayor volumen en más de una década (Télam Digital, 2021). En dicho artículo se explica que parte del motivo por el cual se redujo el consumo de carne vacuna fue el incremento del precio (78% de incremento anual) y de la baja en el poder adquisitivo de las familias.

La baja en el consumo interno de carne vacuna y el incremento de la faena puede que no sean tan malas noticias para la Nación. Por un lado, puede significar que la dieta de los argentinos se está volviendo más variada, dado que se ha incrementado el consumo de otros tipos de carne como la porcina y aviar (dando lugar al crecimiento de nuevas industrias). Por otro lado, el incremento en la faena y la reducción del consumo interno significan que hubo mayores saldos exportables de carne, lo que genera un flujo de divisas que ayudan a engrosar las arcas del Estado.

El primer trimestre del año 2021 fue el que mostró los peores resultados en relación con el consumo de carne vacuna en el país (Boyadjian, 2021). Esto se puede explicar por la caída en los ingresos de las familias y el aumento en el nivel de desocupación resultantes de la pandemia por el COVID-19 donde las políticas económicas de los últimos tiempos no pudieron mitigar. El gráfico a continuación, extraído del artículo de (Boyadjian, 2021) cuya fuente es CICCRA, muestra la evolución del consumo de carne vacuna per cápita en el período 2005 a 2021.



El presente trabajo pretende analizar desde una perspectiva de series de tiempo la evolución en el consumo de carne vacuna y el nivel de faena con la finalidad de predecir cómo ambas variables se comportarán en el futuro cercano. Esto se debe a que cualquier pronóstico de índole económica en Argentina sólo puede realizarse a corto plazo por los vaivenes que sufre dicha economía y la falta de políticas de estado de largo plazo que permitan sostener el consumo interno de carnes acompañado de mayores volúmenes de producción. Si bien no es objeto de este trabajo, no hay que dejar de mencionar

que la matriz productiva Agrícola/ganadera de la República Argentina, se encuentran estrechamente relacionada, donde el creciente incremento de los precios de los granos y mayores superficies destinadas a producción de los mismo con ciclos más cortos de liquidación, hacen que la producción de carnes se vea desalentada.

## Marco Teórico (Punto 1)

Según (Enders, 2014) un modelo de series de tiempo es una metodología desarrollada para descomponer una serie en componentes de tendencia, estacionalidad, cíclicos y un componente irregular. El componente de tendencia es representado por el comportamiento a largo plazo de la serie y, el componente cíclico, por los movimientos periódicos regulares. El componente irregular es estocástico y ha de ser estimado y pronosticado. Para poder realizar dichos pronósticos y estimaciones se ha de encontrar una ecuación de movimiento que conduce el proceso estocástico y utilizarla para predecir valores subsecuentes. Cabe aclarar que, en el caso de las series temporales univariantes, donde las estimaciones futuras obtenidas se basan en las hipótesis de que las condiciones futuras serán análogas a las pasadas y son especialmente útiles para la previsión a corto plazo (Peña, 2014).

Continuando con lo propuesto por (Enders, 2014), una ecuación diferencial expresa el valor de una variable como función de sus valores anteriores, el tiempo y otras variables. Las ecuaciones diferenciales son las que permitirán encontrar los coeficientes de las series de tiempo. Dependiendo de la complejidad del problema, para poder determinar los coeficientes de la ecuación de movimiento se deberán utilizar ecuaciones diferenciales de distintos grados. En términos generales la ecuación diferencial de orden-n se puede expresar de la siguiente manera:

$$y_t = a_0 + \sum_{i=1}^n a_i y_{t-i} + x_t$$

Siendo n el orden de la ecuación diferencial,  $y_t$  el valor a predecir,  $y_{t-i}$  los valores anteriores,  $a_0$  el coeficiente independiente,  $a_i$  el coeficiente correspondiente al período i y  $x_t$  es un componente aleatorio que recoge todos los demás efectos que actúan sobre la serie que tiene una estructura estable a lo largo del tiempo (media cero, varianza constante y distribución normal (Enders, 2014; Peña 2010)).

Para poder encontrar la ecuación de movimiento relacionada a una serie temporal se pueden aplicar distintos métodos como son la iteración sin una condición inicial, la reconciliación de dos métodos iterativos o las secuencias no-convergentes, pero todos estos métodos se vuelven sumamente complejos rápidamente. Por lo tanto, (Enders, 2014) propone la siguiente metodología de solución:

1. Formar la ecuación homogénea y encontrar las n soluciones homogéneas.
2. Encontrar una solución particular.
3. Obtener la solución general como la suma de las soluciones particulares y una combinación lineal de todas las soluciones homogéneas.
4. Por último, eliminar las constantes arbitrarias al imponer las condiciones iniciales en la solución general.

Según (Peña, 2010), los fenómenos dinámicos que se observan mediante series temporales pueden ser de dos clases: los estacionarios, que toman valores constantes en el tiempo sin mostrar tendencias crecientes/decrecientes a lo largo del tiempo; y, los no estacionarios, que pueden mostrar tendencia, estacionalidad y otros efectos evolutivos en el tiempo. La estacionalidad de una serie puede variar en el tiempo, siendo estacionaria en un período corto, pero no así si se amplía el margen temporal.

(Peña, 2010) indica que un proceso estocástico (serie temporal) es estacionario en sentido estricto si las distribuciones marginales de todas las variables son idénticos y las distribuciones finito-dimensionales de

cualquier conjunto de variables sólo dependen de los retardos entre ellos. Las condiciones de estacionariedad en sentido débil son:

1.  $\mu_t = \mu = cte$
2.  $\sigma_t^2 = \sigma^2 = cte$
3.  $\gamma(t, t - k) = E[(z_t - \mu)(z_{t-k} - \mu)] = \gamma_k \quad k = 0, \pm 1, \pm 2$

Las dos primeras condiciones indican que la media y la varianza son constantes. La tercera, que la covarianza entre dos variables depende solo de su separación.

Para estudiar los procesos estacionarios se pueden aplicar distintos modelos, entre ellos se encuentran los procesos de ruido blanco, autorregresivos (AR), medias móviles (MA) y combinaciones de ellos como ARMA, ARIMA, SARIMA, etc.

### Modelo Ruido Blanco (White Noise)

(Del Rosso, 2021) menciona que un proceso estocástico  $\{\varepsilon_t\}$  se comporta como ruido blanco (white noise) si  $\varepsilon_t$  son independientes e idénticamente distribuidos (IID) con valor esperado cero y varianza finita. Esto es:

- $E(\varepsilon_t) = 0$
- $Var(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$
- $cov(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s$

En el proceso de ruido blanco, la primera condición establece que la esperanza es siempre constante e igual a cero; la segunda condición que la varianza es contante; y, la tercera que las variables del proceso no se encuentran correlacionadas para todos los retardos (Peña, 2010).

### Modelo Autorregresivo (AR)

En el caso general, una serie temporal  $z_t$  estacionaria sigue un proceso autorregresivo de orden  $p$  (AR( $p$ )) si:

$$\bar{z}_t = \phi_1 \bar{z}_{t-1} + \dots + \phi_p \bar{z}_{t-p} + a_t$$

Donde  $\bar{z}_t = z_t - \mu$ , siendo  $\mu$  la media del proceso estacionario  $z_t$  y  $a_t$  un proceso de ruido blanco. Los coeficientes de autocorrelación de un AR( $p$ ) verifican la siguiente ecuación en diferencias:

$$p_k = \phi_1 p_{k-1} + \dots + \phi_p p_{k-p}, \quad k > 0$$

La función de autocorrelación simple de un proceso  $AR(p)$  es una mezcla de exponenciales, debidas a los términos con raíces reales, y sinusoidales, debidas a las raíces complejas conjugadas. Su estructura puede ser muy compleja. (Peña, 2010).

Como determinar el orden de un proceso autorregresivo a partir de la función de autocorrelación simple (fas) es difícil, se ha desarrollado la función de autocorrelación parcial (fap) que es la representación de los coeficientes de autocorrelación parcial en función del retardo:

$$\bar{z}_t = a_{k1} \bar{z}_{t-1} + \dots + a_{kk} \bar{z}_{t-k} + \eta_{kt}$$

Un proceso  $AR(p)$  tendrá los  $p$  primeros coeficientes de autocorrelación parcial distintos de cero, y, por tanto, en la fap el número de coeficientes distintos de cero indica el orden del proceso  $AR$ . El coeficiente de correlación parcial del orden  $p$  coincidirá siempre con el parámetro  $\phi_p$  (Peña, 2010).

Un proceso  $AR(0)$  indica que los resultados anteriores no afectan su comportamiento futuro; un proceso Markoviano ( $AR(1)$ ) sólo se ve afectado por el período inmediato anterior; un proceso de Yule ( $AR(p)$ ) se ve afectado por los dos períodos inmediatos anteriores y así sucesivamente hasta llegar al caso general  $AR(p)$ .

En el caso de los modelos  $AR$ , estos son estacionarios bajo las siguientes condiciones:

$$|Raíces EC| < 1 \quad |Raíces PC| > 1$$

Si el modelo  $AR$  es estacionario y se cumple que  $AR(p) \leftrightarrow MA(\infty)$ , entonces será invertible (Del Rosso, 2021).

### Modelo Medias Móviles (MA)

Según (Uriel, 1985), un proceso de medias móviles de orden  $q$  ( $MA(q)$ ) está dado por:

$$Y_t = \varepsilon_t - \theta_t \varepsilon_{t-1} - \dots - \theta_t \varepsilon_{t-q}$$

Esta expresión hace referencia a que la variable  $Y_t$  se obtiene como un promedio de variables de ruido blanco, siendo las  $\theta_t$  los coeficientes de ponderación.

Al utilizar un proceso de ruido blanco para construir una serie temporal más interesante se aplica la siguiente ecuación:

$$\varepsilon_t = \sum_{i=0}^q \theta_i \varepsilon_{t-i}$$

Para cada período  $t$ ,  $x_t$  se construye tomando los valores de  $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$  y multiplicando cada uno por el valor asociado de  $\theta_i$  (Enders, 2014). La media de un modelo  $MA$  es cero, cualesquiera que sean los valores de  $\theta$ .

(Del Rosso, 2021) menciona que los modelos  $MA$  son siempre estacionarios con independencia de los valores de los coeficientes, ya que son combinaciones lineales de errores ruido blanco. Dado que estos modelos son estacionarios, pueden ser invertibles en un  $AR(\infty)$  bajo ciertas condiciones. La invertibilidad se puede evaluar a partir de las raíces de la ecuación característica (EC) o a partir de las raíces del polinomio característico (PC):

$$|Raíces EC| < 1 \quad |Raíces PC| > 1$$

A diferencia de los modelos  $AR$ , los modelos de  $MA$  tienen memoria finita.



## Modelo ARMA

En (Enders, 2014), se explica que el modelo de autorregresivo medias móviles (*ARMA*) surge de la combinación de un proceso de medias móviles con una ecuación de diferencias lineal:

$$y_t = \theta_0 + \sum_{i=1}^p a_i y_{t-i} + \sum_{i=0}^q \theta_i \varepsilon_{t-i}$$

La parte autorregresiva del modelo es la ecuación de diferencias dada por la parte homogénea de:  $y_t = \theta_0 + \sum_{i=1}^p a_i y_{t-i} + x_t$ , y la parte de medias móviles es la secuencia  $x_t$ . Si la parte homogénea de la ecuación de diferencias contiene  $p$  retrasos y el modelo para  $x_t$  contiene  $q$  retrasos, el modelo se llama *ARMA*( $p, q$ ). Si  $q=0$ , el proceso se denomina un proceso puramente autorregresivo denotado por *AR*( $p$ ) y si  $p=0$ , el proceso es de medias móviles puras, denotado por *MA*( $q$ ).

(Del Rosso, 2021) dice que para que un modelo *ARMA*( $p, q$ ) sea estacionario se requiere que las raíces de la ecuación polinomial caigan fuera del círculo unitario. Por tanto, como un proceso *MA* es siempre estacionario, para evaluar la estacionariedad de estos modelos solo basta con probar la estacionariedad de su parte autorregresiva:

- Si se cumplen las condiciones de estacionariedad, el modelo *ARMA*( $p, q$ ) se puede expresar como un *MA*( $\infty$ ).
- Si se cumplen las condiciones de invertibilidad, el modelo *ARMA*( $p, q$ ) se puede expresar mediante un *AR*( $\infty$ ).

Continuando con los mencionado por (Del Rosso, 2021), si se cumplen las condiciones de Estacionalidad, entonces es posible transformar a un proceso finito en un *MA*( $\infty$ ). Asimismo, si se cumplen las condiciones de invertibilidad, entonces es posible transformar a un proceso finito en un *AR*( $\infty$ ).

## Modelo ARIMA

Según (Athanasopoulos & Hyndman, 2018) si se combina la diferenciación con una autorregresión y un modelo de medias móviles, se obtiene un modelo *ARIMA* no-estacional. Este modelo se puede describir de la siguiente manera:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

Donde  $y'_t$  es la serie diferenciada (puede haber sido diferenciada más de una vez). Los predictores en el lado derecho incluyen valores rezagados de  $y_t$  y errores rezagados. Al modelo que así se desarrolla se lo denomina *ARIMA*( $p, d, q$ ) siendo  $p$  el orden de la parte autoregresiva,  $d$  el grado de la primera diferencial y  $q$  el orden de la parte de medias móviles.

Continuando con lo que indican (Athanasopoulos & Hyndman, 2018), muchos de los modelos anteriores se corresponden a casos especiales del modelo *ARIMA*, a saber:

- |                                  |  |
|----------------------------------|--|
| • Ruido Blanco                   | <i>ARIMA</i> (0,0,0)                   |
| • Random Walk                    | <i>ARIMA</i> (0,1,0) sin una constante |
| • Random Walk con una derivación | <i>ARIMA</i> (0,1,0) con una constante |
| • Autorregresión                 | <i>ARIMA</i> (1,0,0)                   |

- Medias móviles

$ARIMA(0,0,1)$

### Modelo SARIMA

Según (Athanasopoulos & Hyndman, 2018) un modelo  $ARIMA$  estacional ( $SARIMA$ ) está formado por medio de la inclusión de términos estacionales en el modelo  $ARIMA$ . Este modelo se escribe de la siguiente manera:  $ARIMA(p, d, q)(P, D, Q)_m$  donde el primer paréntesis denota las componentes de la parte no-estacional, el segundo paréntesis la parte estacional y la  $m$  es el número de observaciones por período estacional (en el que se repite el patrón).

La parte estacional del modelo consiste en términos similares a la parte no-estacional, pero implica retrocesos en el período estacional. Los términos estacionales adicionales se multiplican por los términos no-estacionales (Athanasopoulos & Hyndman, 2018). Un modelo  $ARIMA(1,1,1)(1,1,1)_{12}$  puede escribirse de la siguiente forma:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})y_t = (1 + \theta_1 B)(1 + \Theta_1 B^{12})\varepsilon_t$$

La parte estacional de un modelo  $AR$  o  $MA$  se puede observar en los retrasos de las funciones de autocorrelación y de autocorrelación parcial.

## Serie Faena

### Análisis Series Originales (Punto 2)

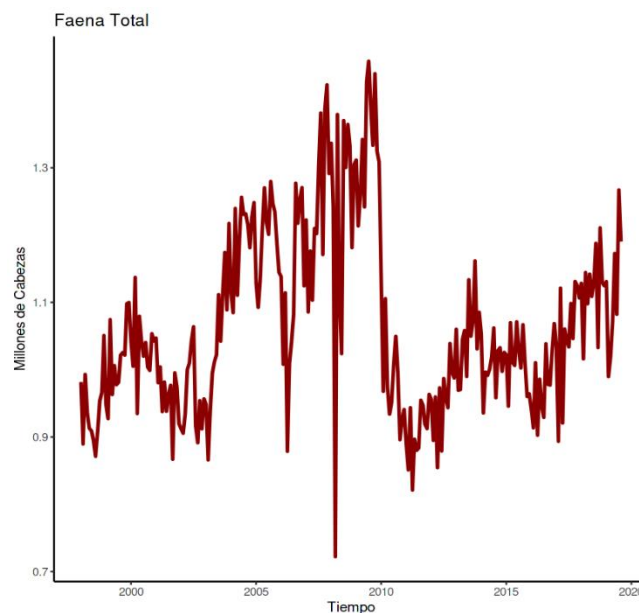
En primera instancia se procederá a realizar un análisis visual de los datos. Luego se aplicará el test de raíces unitarias para evaluar si las conclusiones referentes a estacionariedad de la serie coinciden con aquellas generadas por medio de las visualizaciones.

Habiendo realizado el diagnóstico, y en el caso que la serie no respete las condiciones necesarias para que pueda aplicarse un análisis de series de tiempo, se procederá a realizar una serie de transformaciones a los datos en el caso de considerarse adecuado.

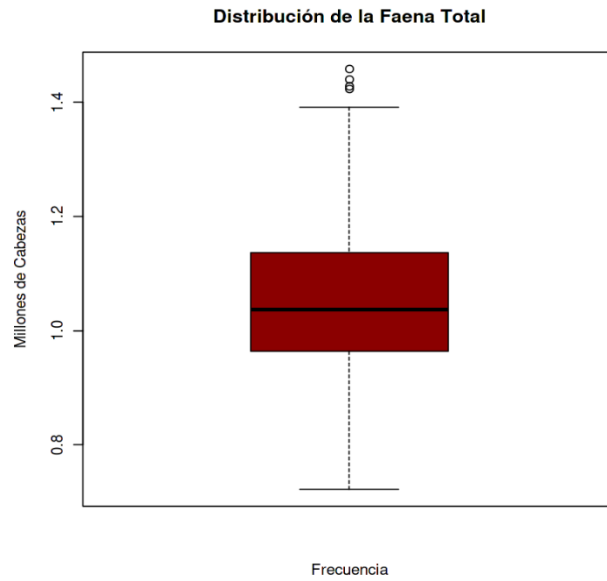
Uno de los supuestos importantes a validar en un análisis de series temporales es el supuesto de estacionariedad. El mismo puede ser en sentido estricto/determinístico o débil.

Siendo que no conocemos el proceso generador de la serie bajo análisis. Es que se probará la existencia de estacionariedad de manera débil, esto es, se analizarán sus momentos de primer orden y segundo orden.

Una serie es estacionaria en sentido débil cuando la serie presenta media y varianza constante a lo largo de la serie.



Del gráfico anterior se puede observar que en la presente serie habría tendencia, por otro lado, la media no sería constante, ya que esta resulta diferente según el período de tiempo considerado, por ende, sería una serie no estacionaria.



Por medio del Boxplot se detecta la presencia de valores extremos los cuales podrían afectar el modelo.

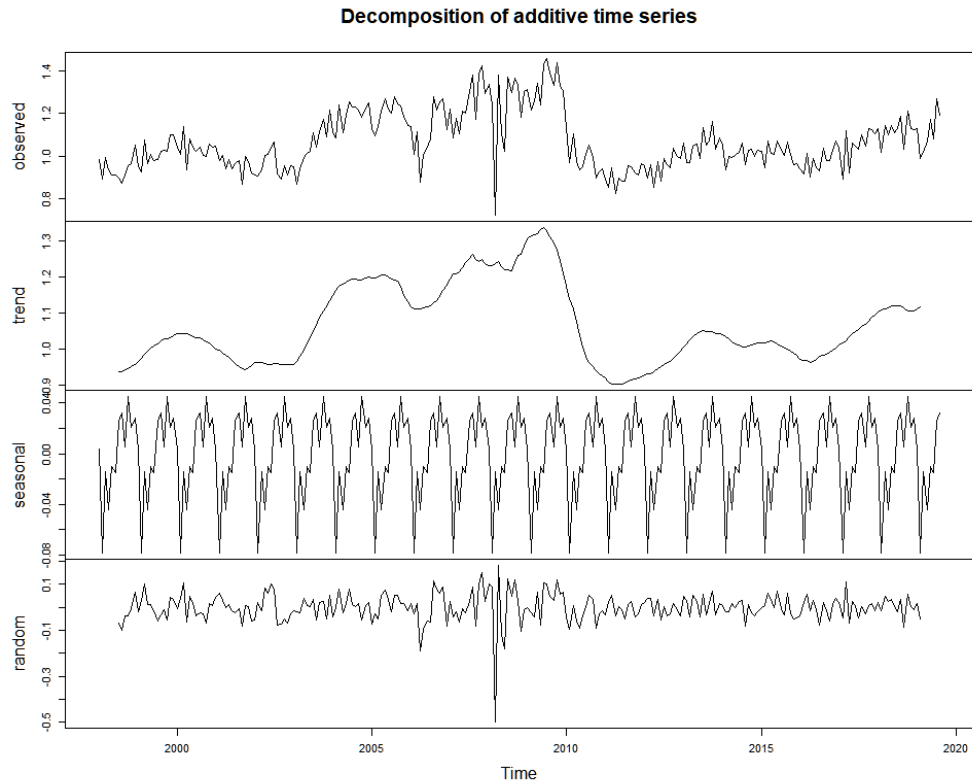
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.722	0.964	1.037	1.065	1.136	1.458

Una serie puede ser descompuesta en los siguientes factores, según Athanasopoulos & Hyndman, se detallan los siguientes:

$$Y_t = C_t + T_t + S_t + \epsilon_t$$

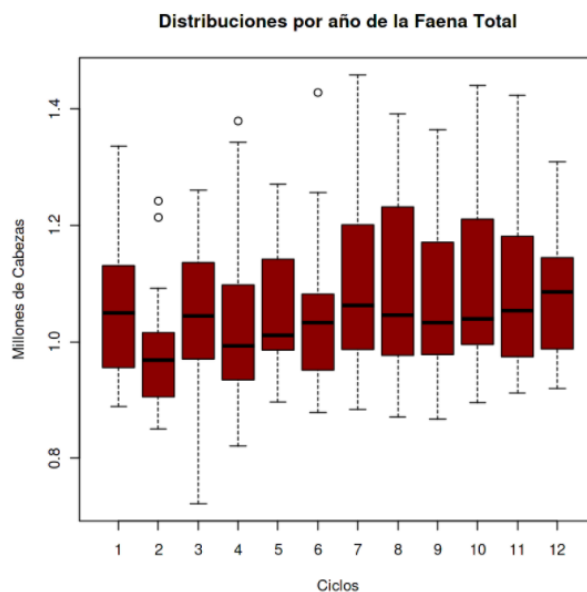
Una serie temporal puede estar compuesta por los siguientes factores, se detallan los siguientes:

- **Tendencia ( $T_t$ ):** una tendencia existe cuando existe un incremento o decremento a lo largo del tiempo en los datos.
- **Estacionalidad ( $S_t$ ):** un patrón estacional ocurre cuando una serie de tiempo se encuentra afectada por factores estacionales como ser un cierto período del año o día de la semana. Para que exista un proceso estacional se requiere que este sea fijo cada cierto tiempo y que se conozca su frecuencia de aparición.
- **Ciclo ( $C_t$ ):** Un ciclo ocurre cuando los datos muestran altas y bajas que no responden a una frecuencia determinada.
- Cabe señalar la diferencia que existe entre la presencia de un componente estacional de un componente cíclico. En el caso que las fluctuaciones no sean fijas entonces esto se refiere a un ciclo. La estacionalidad requiere que la frecuencia sea constante respecto al tiempo.
- **Componente Aleatorio ( $\epsilon_t$ ):** Definido por Peña como una secuencia de variables incorrelacionadas de media cero y varianza constante.



Al descomponer la serie, visualmente se observa que la serie tiene presencia de tendencia, estacionalidad y que hay presencia de valores extremos durante el 2008.

En cuanto a la estacionalidad respecta, por medio de este *boxplot* que grafica los valores por cada mes dentro de la serie, se puede observar que la mediana presenta variaciones según el mes que se considere, esto aplica para el caso de los valores del mes de febrero y diciembre.



Dado que hasta el momento se ha realizado un análisis visual de los datos con el objeto de detectar si esta cumple con el supuesto de estacionariedad, se procederá a aplicar el test de Dicky Fuller para validar o refutar lo observado.

Dicho test supone cuenta con las siguientes hipótesis:

- $H_0: \varphi_1 - 1 = 0$  es decir, el proceso es no estacionario, la serie tiene una raíz unitaria.
- $H_1: \varphi_1 - 1 < 0$ , es decir, hay evidencia suficiente para suponer que el proceso es estacionario, la serie no tiene una raíz unitaria.

El p-value del conjunto analizado alcanza el valor del 0.4123, en función de ello, no se puede rechazar la hipótesis nula, en otras palabras, no existe evidencia suficiente para suponer que esta serie sea estacionaria. Esto confirma lo que se había supuesto en base a la mera observación del gráfico.

Habiendo verificado tanto visualmente como por medio del test la no estacionariedad de la serie, se proceden a realizar los siguientes ajustes a la serie, algunos de los cuales se encuentran descritos por (Athanasopoulos & Hyndman, 2018):

- **Transformaciones:** por medio de logaritmos que pueden ayudar a estabilizar la varianza de una serie de tiempo.
- **Diferenciaciones:** las cuales se utilizan para intentar eliminar la tendencia y la estacionalidad en la media de la serie.

En el caso de la diferenciación simple esta ayuda a eliminar o reducir el impacto de la tendencia de una serie, y puede ser escrita de la siguiente manera

$$y'_t = y_t - y_{t-1}$$

En el caso de la diferenciación estacional, busca eliminar o mitigar el impacto que puede tener la presencia de estacionalidad en la serie, en este sentido se refiere a la diferencia entre la observación y la observación previa referida a la misma estación, esto es, por ejemplo, en el caso que la estacionalidad tuviera una duración de 12 meses, este ajuste debiera restar al valor de enero 2020, el valor de enero 2019.

En términos formales:

$$y_t = y_{t-m} + \varepsilon_t$$

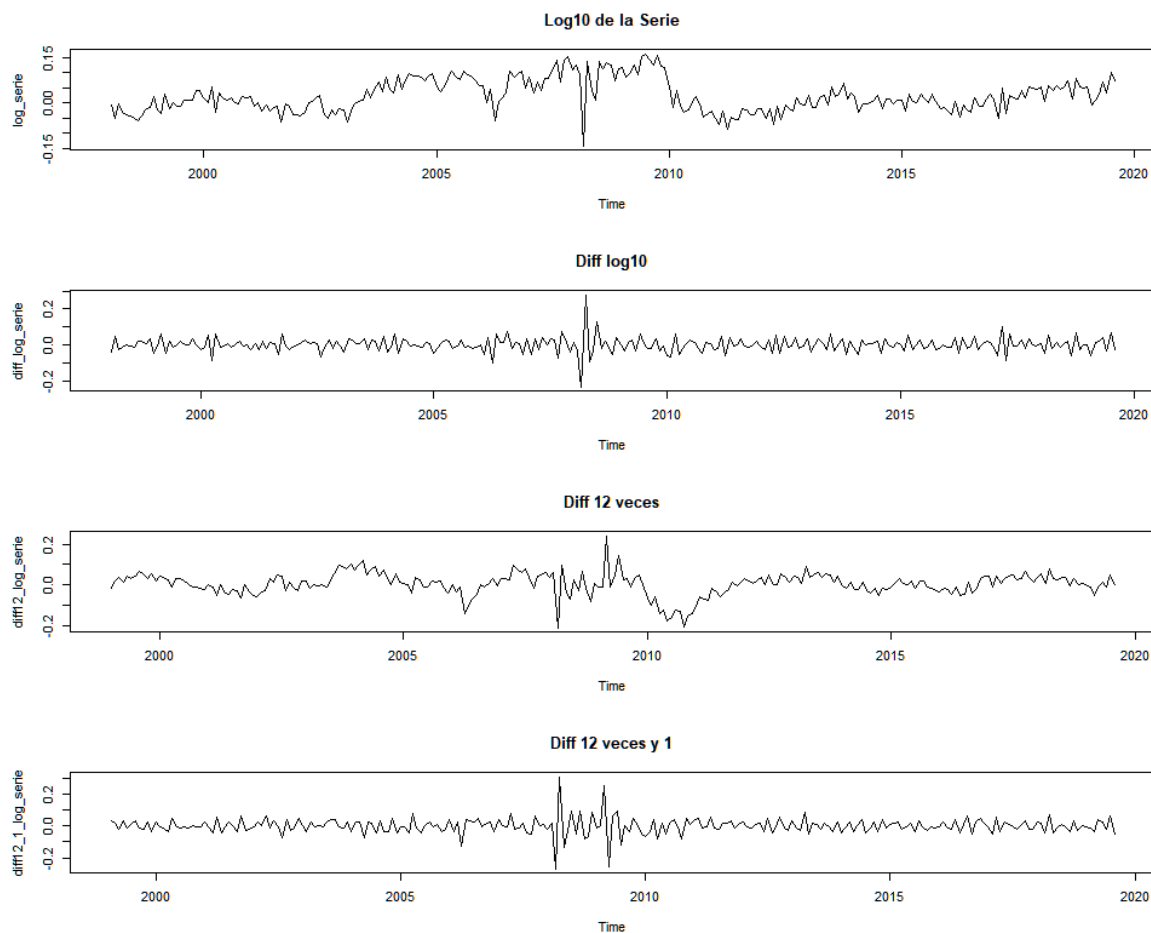
Transformación de Box & Cox, esta última será aplicada en el caso que se detecte la necesidad de estabilizar la varianza de la serie.

Por otro lado, a lo largo del presente trabajo, se realizarán ajustes que se consideren estrictamente necesarios con el fin de alcanzar un modelo parsimonioso a fin de evitar un sobreajuste del modelo alcanzado.

Con base en lo expuesto anteriormente se aplicaron los siguientes ajustes iniciales a la serie:

- Log 10 de la serie original (log\_serie)
- Primera diferenciación de la serie anterior (diff\_log\_serie)
- Diferencia estacional de la serie original (diff12\_log\_serie)
- Primera diferenciación de la serie diferenciada estacionalmente (diff12\_1\_log\_serie)

En los siguientes gráficos pueden observarse los resultados alcanzados por medio de las transformaciones mencionadas.



En función de las gráficas previas podría suponerse que se alcanza la estacionariedad en los casos (b), (c) y (d). A fin de corroborar lo observado, se procede a aplicar el test de Dicky-Fuller con el objeto de evaluar la presencia de estacionariedad de las series transformadas. De los resultados alcanzados, para los casos (b), (c) y (d) se rechaza  $H_0$ , es decir habría presencia de estacionariedad en tales series transformadas.

	(a)	(b)	(c)	(d)
	log_serie	diff_log_serie	diff12_log_serie	diff12_1_log_serie
P value <sup>1</sup>	0.4126	0.01	0.01	0.01
Resultado	No Rechazo $H_0$	Rechazo $H_0$	Rechazo $H_0$	Rechazo $H_0$

Como se observa en los gráficos, la serie original es no-estacionaria y por dicho motivo se han debido aplicar distintas transformaciones para alcanzar la estacionariedad. Si bien se ha experimentado con diferentes transformaciones, y siendo que los resultados alcanzados por medio de las otras no han

<sup>1</sup> Nivel de Significación considerado 5%

generado resultados superiores, se ha optado por la serie más simple en lo que a transformaciones respecta, respetando el principio de parsimonia que afecta a estos modelos.

### Análisis FAS, FAC y FACP (Punto 3)

Se detallan a continuación los gráficos realizados en el código de FAC y FACP y su correspondiente análisis.

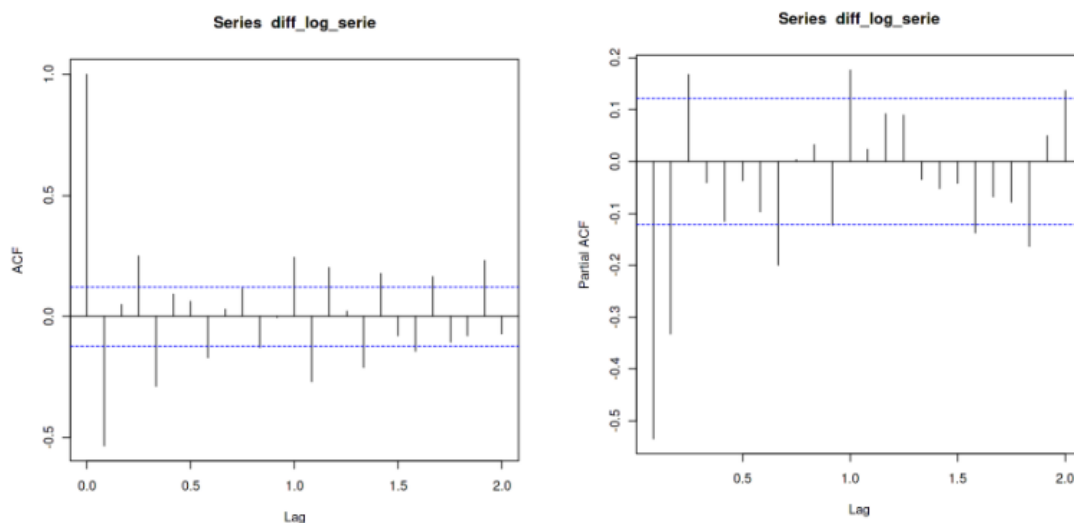
- (a) data: **log\_serie**
- (b) data: **diff\_log\_serie**
- (c) data: **diff12\_log\_serie**
- (d) data: **diff12\_1\_log\_serie**

(a) Dado que la serie original y esta serie son no estacionarias no se pueden obtener mayores conclusiones de los gráficos por ende no se adjuntan en el presente trabajo.

(b) En cuanto al gráfico correspondiente a la primera diferenciación de la serie logarítmica se puede apreciar por medio de la función de autocorrelación que se estaría en presencia de un modelo ARIMA ya que en su gráfico de autocorrelación presenta una fuerte caída en sus primeros 'lags' y el resto de los 'lags' presentan un comportamiento irregular y poco significativo.

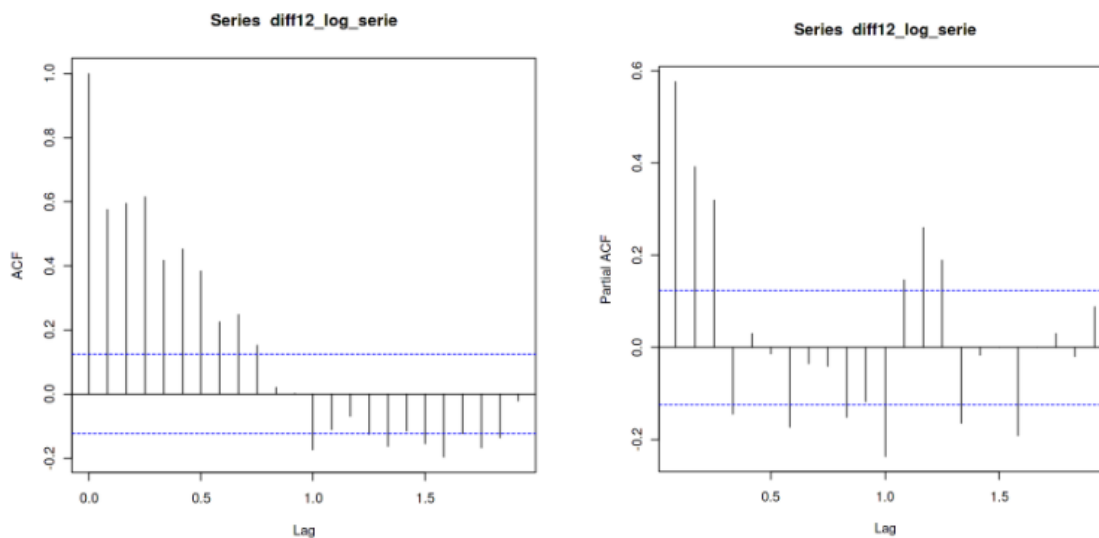
Con respecto a la función de autocorrelación parcial se puede observar que la misma tiene 'lags' significativos en el nivel 1 y 2 y que cada 12 meses existen picos significativos, lo cual podría estar indicando presencia de estacionalidad al considerar este último 'lag' mencionado, ya que los coeficientes crecen en ese momento y luego decrecen rápido con periodicidades iguales.

Esto nos permite suponer que el modelo que mejor se puede ajustar a los datos y tener un mayor poder de predicción es aquel que se compone de la parte Autorregresiva de orden 2 AR(2), siendo necesaria una diferenciación de primer orden que ya se encuentra incluida en la serie.

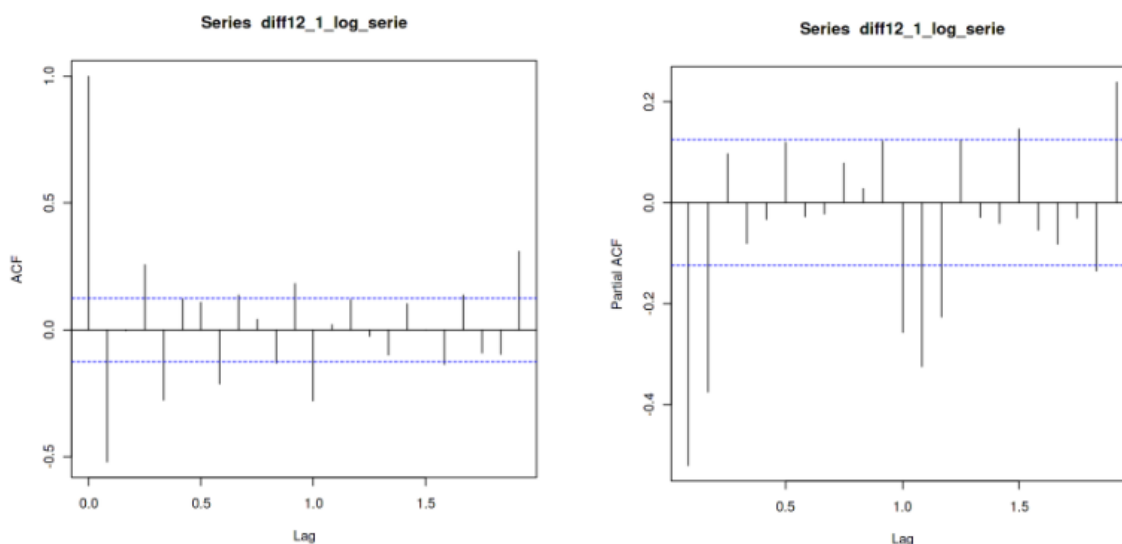


(c) Se podría inferir que estamos en presencia de un modelo AR ya que en la función de autocorrelación se observa que los coeficientes poseen un decrecimiento atenuado y ondas sinusoidales presentando en la función de autocorrelación parcial coeficientes representativos sólo en sus primeros tres valores.





- (d) La serie diferenciada por estacionalidad y diferenciada una vez, parece replicar el mismo patrón que el indicado en el punto anterior, resultando más marcada la presencia de estacionalidad en los valores cercanos a 12 y a 24 en la función de autocorrelación parcial.



Los gráficos ACF y PAC muestran una aproximación del modelo a elegir, sin embargo, para elegir el mejor candidato resulta necesario realizar diversas pruebas que validen con cuál de ellos se obtienen mejores resultados para el caso bajo análisis.

### Modelo SARIMA (Punto 4)

Al considerar los resultados obtenidos en los puntos anteriores, se llegó a la conclusión que los datos exhiben una estacionalidad cada 12 meses (1 período anual). A continuación, se detallan los distintos modelos SARIMA que han sido desarrollados con la finalidad de encontrar el mejor modelo que se ajuste a los datos de la serie:

$$(1) \text{SARIMA}(1,0,0)(0,0,0)_{12} \rightarrow \text{AR}(1)$$

$$(2) \text{SARIMA}(2,0,0)(0,0,0)_{12} \rightarrow \text{AR}(2)$$

(3)  $SARIMA(2,0,1)(0,0,0)_{12} \rightarrow ARMA(2,1)$

(4)  $SARIMA(3,0,0)(0,0,0)_{12} \rightarrow AR(3)$

(5)  $SARIMA(2,0,0)(1,0,1)_{12}$

(6)  $SARIMA(2,1,0)(1,0,0)_{12}$

(7)  $SARIMA(2,1,1)(1,0,0)_{12}$

(8)  $SARIMA(2,1,1)(0,0,1)_{12}$

(9)  $SARIMA(2,1,2)(1,0,0)_{12}$

(10)  $SARIMA(1,1,0)(1,0,0)_{12}$

(11)  $SARIMA(2,1,0)(0,0,1)_{12}$

(12)  $SARIMA(2,1,0)(0,1,1)_{12}$

Los resultados obtenidos para cada uno de los modelos mencionados anteriormente se detallan a continuación (los números que anteceden a cada uno de ellos corresponden al número de encabezado en cada una de las columnas:

	(1)	(2)	(3)	(4)	(5)	(6)
$a_0$	0.0013	0.0013	0.0013	-0.0033	-0.0043	
p-value	0.4239	0.1836	0.2066	0.8778	0.7646	
$a_1$	-0.6690	-0.9990	-1.0983	-0.0082	0.2871	-0.9002
p-value	$2.00 \times 10^{-16}$	$2.20 \times 10^{-16}$	$4.80 \times 10^{-12}$	0.9323	0.0027	$2.20 \times 10^{-12}$
$a_2$		-0.4888	-0.5570	0.4752	0.4792	-0.4736
p-value		$2.55 \times 10^{-7}$	$6.37 \times 10^{-6}$	$1.26 \times 10^{-8}$	$1.39 \times 10^{-6}$	$8.89 \times 10^{-7}$
$a_3$				0.4505		
p-value				$5.70 \times 10^{-6}$		
$\beta_1$			0.1229			
p-value			0.4724			
$\beta_2$						
p-value						
$sar_1$					0.1852	0.4207
p-value					0.5011	$3.93 \times 10^{-4}$
$sma_1$					0.3141	
p-value					0.2397	
$\sigma^2$	0.0005879	0.0004442	0.0004414	0.0004500	0.0004588	0.0004011
Log-likelihood	190.65	202.00	202.25	203.44	201.64	206.69
AIC	-375.3	-396.01	-394.51	-396.89	-391.27	-405.37

	(7)	(8)	(9)	(10)	(11)	(12)
$a_0$						
p-value						
$a_1$	-1.0182	-1.0414	-0.8995	-0.6072	-0.9155	-0.9035
p-value	$3.74 \times 10^{-9}$	$4.37 \times 10^{-9}$	$6.96 \times 10^{-9}$	$5.63 \times 10^{-12}$	$2.2 \times 10^{-16}$	$2.2 \times 10^{-16}$
$a_2$	-0.5477	-0.5320	-0.5567		-0.4506	-0.5864
p-value	$9.29 \times 10^{-6}$	$3.88 \times 10^{-5}$	$6.00 \times 10^{-5}$		$5.49 \times 10^{-6}$	$7.13 \times 10^{-10}$
$a_3$						
p-value						
$\beta_1$	0.1515	0.1578	0.0391			
p-value	0.4364	0.4273	0.8723			
$\beta_2$			0.1793			
p-value			0.3508			
$sar_1$	0.4230		0.4100	0.4143		
p-value	0.0004		$9.55 \times 10^{-4}$	$2.31 \times 10^{-4}$		
$sma_1$		0.3447			0.3420	-1.000

	(7)	(8)	(9)	(10)	(11)	(12)
p-value		0.00198			0.0018	0.0310
$\sigma^2$	0.0003838	0.0003984	0.003806	0.0005009	0.0004015	0.0003027
Log-likelihood	206.98	205.86	207.36	196.23	205.55	174.71
AIC	-403.96	-401.72	-402.71	-386.46	-403.11	-341.43

Las tablas anteriores muestran los coeficientes de los modelos al igual que algunas métricas que hacen posible la selección del modelo más apropiado.

El valor de *log-likelihood* suele ser mejor para los casos en que los modelos cuenten con más grados de libertad. Cabe recordar que por cada diferenciación se le está restando un grado de libertad al modelo propuesto. Sin embargo, esta no es una buena medida de selección de modelos ya que, considerando el principio de parsimonia, los modelos más complejos suelen tener mayores *log-likelihood*. Por este motivo, se debe emplear otro criterio de selección de modelos.

En cuanto a los criterios de información, este criterio de selección de Akaike (AIC) considera que el modelo con el menor valor de AIC será el mejor de todos ellos.

Basándose en lo antepuesto, se considera que el modelo *SARIMA(2,1,0)(1,0,0)<sub>12</sub>*, resaltado en verde, es el que arroja el menor valor de AIC y, además, es el modelo cuyos coeficientes resultan todos significativos (con *p-value* < 0.05). Este resultado fue confirmado con la función *auto.arima* con el conjunto de prueba diferenciado y se obtuvo el modelo mencionado anteriormente.

El código empleado para la elaboración de cada uno de los modelos se encuentra en el link indicado en la sección “Código Empleado para el Desarrollo del Presente Trabajo” de la presente presentación.

### Métricas de Performance (Punto 5)

Para la evaluación del performance de los distintos modelos, se seleccionó la métrica Mean Absolute Error (MAE – por sus siglas en inglés). La utilización de esta métrica conlleva a las siguientes ventajas:

- La medida del error está en la misma unidad de la serie temporal. En este caso como solo se está analizando una sola serie temporal, no habrá problemas de comparación de resultados.
- Es una medida de performance muy utilizada en series temporales ya que es fácil de computar y de interpretar.

Adicionalmente se explorarán otras métricas de performance y analizar cómo se comparan entre ellas.

Respecto a los datos utilizados, la serie fue dividida en dos conjuntos de datos, uno de entrenamiento (*train*) y otro de testeo (*test*), en la siguiente configuración:

	Train	Test	Total
Fecha inicio	01/2011	01/2018	01/2011
Fecha Fin	12/2017	12/2018	12/2018
Cant. Observaciones	84	12	96

### Consideraciones en la selección de los datos:

- Si bien se cuenta con información histórica desde el año 1998, se excluyeron desde esa fecha hasta diciembre de 2010 debido a que hubo grandes cambios en el comportamiento de la faena.

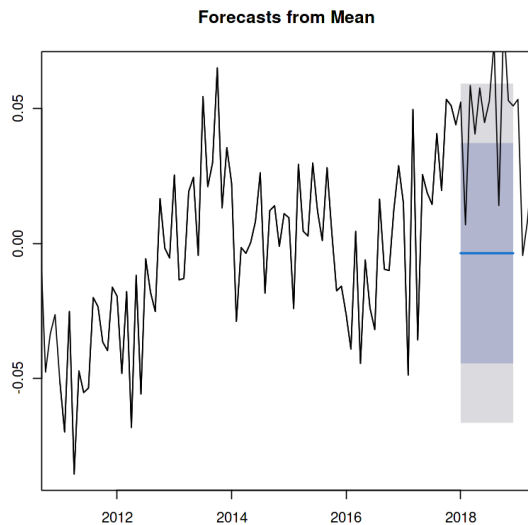
El periodo excluido incluye los eventos de profundo impacto, entre los que se pueden mencionar la crisis económica del 2001, el conflicto que tuvieron los productores y el gobierno en el primer semestre del año 2008 a partir de la resolución 125, que si bien no afectaba a la producción ganadera, se puede apreciar la complejidad y relacionada matriz productiva agrícola/ ganadera, que combinada a los efectos climáticos que causaron grandes sequías y combinado al escaso nivel de inversiones originado en la falta de políticas de estado que den previsibilidad al sector entendemos que el período seleccionado es el más representativo.

- Debido a que el dataset presenta estacionalidad se decidió utilizar un año completo para realizar las predicciones (dataset de testeo).
- Cabe señalar que la serie utilizada en el modelo (S)ARIMA ha sido la serie original aplicado el logaritmo, restringida al período indicado.

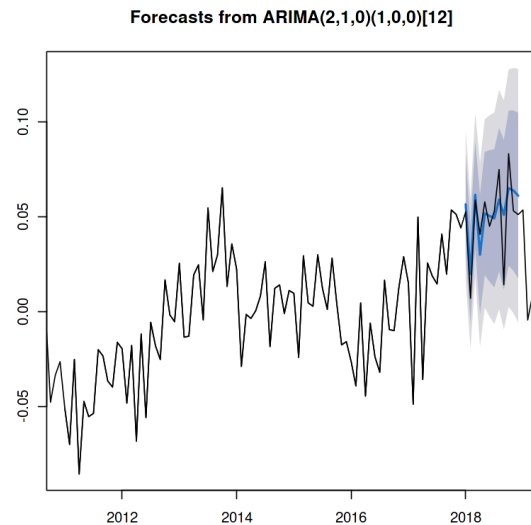
### Comparación con otros modelos estimados (Punto 6)

Se han elaborado seis modelos predictivos a partir de los datos de entrenamiento y testeo:

#### Promedio Simple:

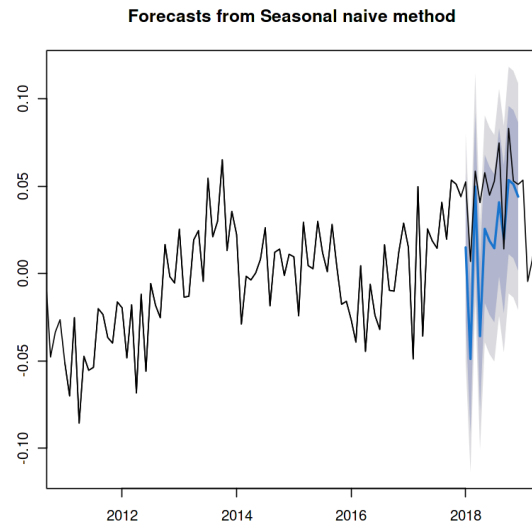
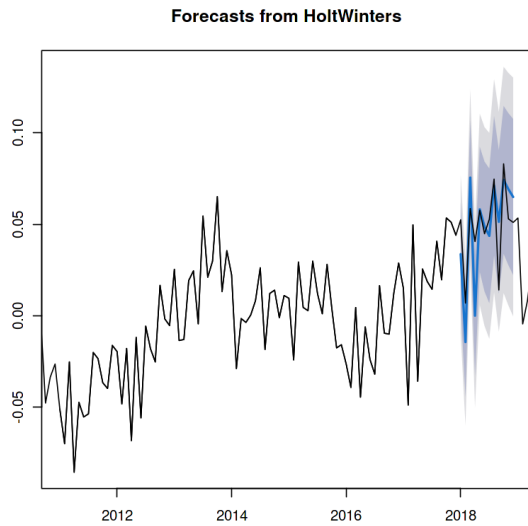


#### Auto Arima:

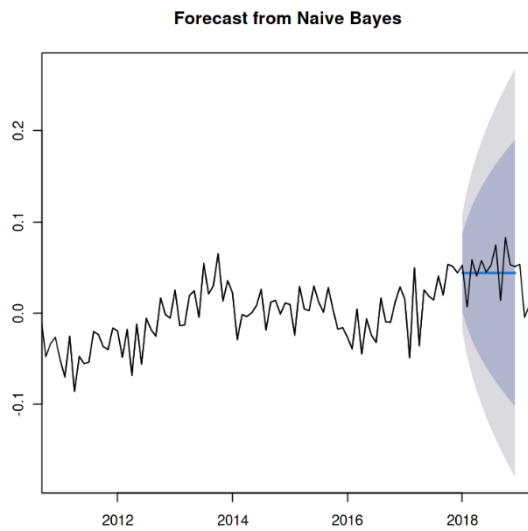


#### Holt Winters:

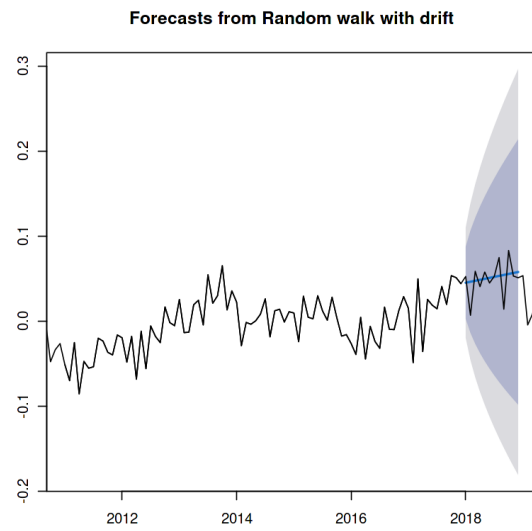
#### Naïve Estacional:



### Naïve Bayes:



### Drift:



### Métricas de Evaluación:

Modelo	ME	RMSE	MAE	MPE	MAPE	MASE
Auto Arima	-0.0023	0.0145	0.0114	-35.41	49.89	0.4096
Holt Winters	0.0011	0.0199	0.0159	9.65	69.83	0.5709
Naïve Bayes	0.0051	0.0213	0.0169	-46.41	80.03	0.6055
Drift	-0.0023	0.0200	0.0150	-66.66	84.99	0.5394
Naïve Estacional	0.0285	0.0362	0.0294	110.67	117.26	1.0564
Promedio Simple	0.0528	0.0567	0.0528	112.17	112.17	1.8953

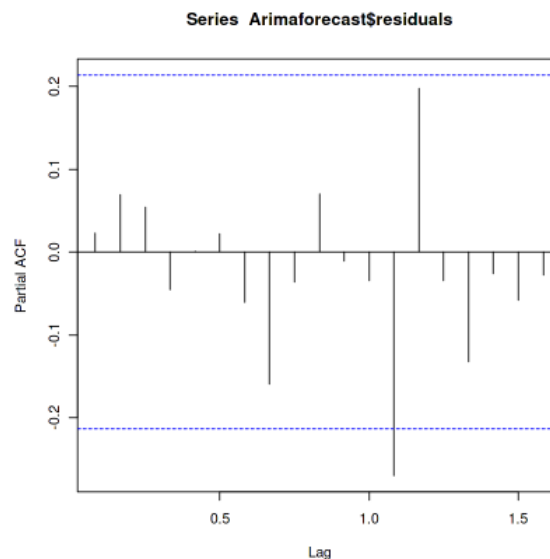
Partiendo del análisis de las distintas métricas se puede concluir que el modelo de Auto Arima es el modelo que mejor captura la tendencia de la serie bajo análisis.

Las métricas relacionadas a medidas absolutas como, el RMSE, MAE, MAPE, MASE indican que el modelo Auto Arima es el que posee una menor magnitud en su error.

Cabe destacar que el modelo de Holt Winters al contar con una magnitud levemente mayor que el modelo Auto Arima, lo convierte en una opción viable a considerar ya que una de sus características es la de poseer un menor sesgo en sus predicciones.

### Análisis de Diagnóstico de las Series Temporales (Punto 7)

Se seleccionó  $SARIMA(2,1,0)(1,0,0)_{12}$  como el modelo con las mejores métricas de validación, se procede a realizar la validación de los residuos. Como un primer paso se realiza la prueba de autocorrelación en los residuos mediante el gráfico ACF de los residuales y el test de Ljung-Box.



#### Ljung-Box test

```
data: Residuals from ARIMA(2,1,0)(1,0,0)[12]
Q* = 14.286, df = 14, p-value = 0.4286

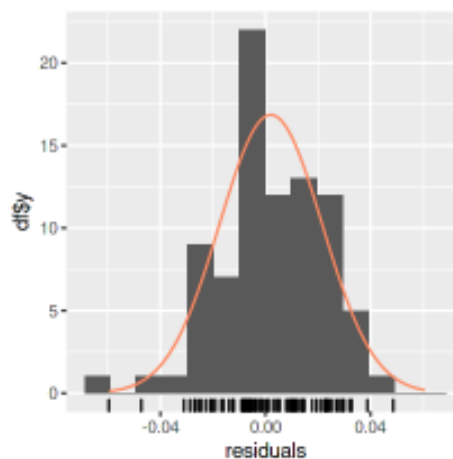
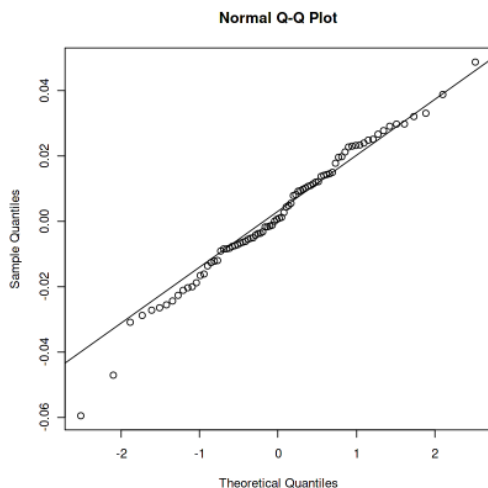
Model df: 3. Total lags used: 17

>>> Note: residuos is from the workspace, not in a data frame (table)
```

De la gráfica del ACF se puede observar que los residuales no superan los intervalos de confianza exceptuando un caso, indicando falta de correlación de estos, para completar el análisis se realiza el test de Ljung Box donde se observa que no se rechaza la Hipótesis nula indicando independencia de estos mismos.

Como un segundo paso se procede a realizar el análisis de normalidad, para esto se realiza el grafico de qq-plot e histograma y se confirma con los test de Jarque- Bera, Anderson Darling y Shapiro Wilk.

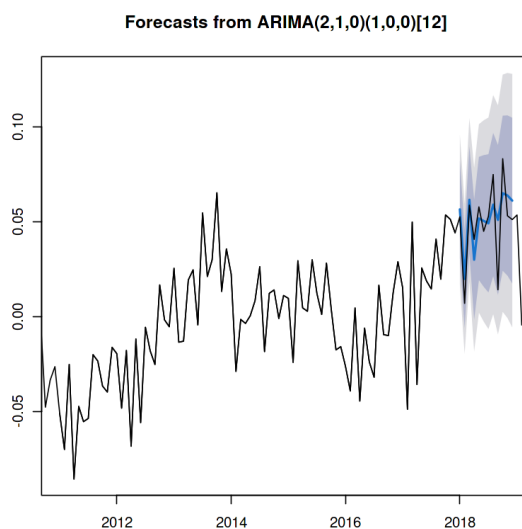
Prueba	Jarque - Bera	Anderson Darling	Shapiro Wilk
P-value	0,43	0,53	0,56



En las diferentes pruebas realizadas se puede identificar que los residuos tienen una distribución que se asemeja a la normal indicando que podemos obtener buenas predicciones del modelo.

### Pronósticos con el modelo Seleccionado (Punto 8)

- (1) El modelo seleccionado para realizar las predicciones es el  $SARIMA(2,1,0)(1,0,0)_{12}$ , en función a lo descrito en los puntos anteriores.



Las predicciones generadas con este modelo son las siguientes:

Date	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Jan-2018	0.0564	0.0307	0.0820	0.0171	0.0956
Feb-2018	0.0198	-0.0060	0.0456	-0.0197	0.0592
Mar-2018	0.0614	0.0333	0.0895	0.0184	0.1044
Apr-2018	0.0299	-0.0017	0.0615	-0.0184	0.0782
May-2018	0.0516	0.0192	0.0841	0.0020	0.1012
Jun-2018	0.0503	0.0157	0.0850	-0.0027	0.1033
Jul-2018	0.0491	0.0126	0.0855	-0.0066	0.1048
Aug-2018	0.0589	0.0212	0.0966	0.0012	0.1166
Sep-2018	0.0509	0.0114	0.0903	-0.0094	0.1112
Oct-2018	0.0649	0.0241	0.1058	0.0024	0.1274
Nov-2018	0.0637	0.0215	0.1059	-0.0009	0.1283
Dec-2018	0.0610	0.0174	0.1046	-0.0057	0.1277



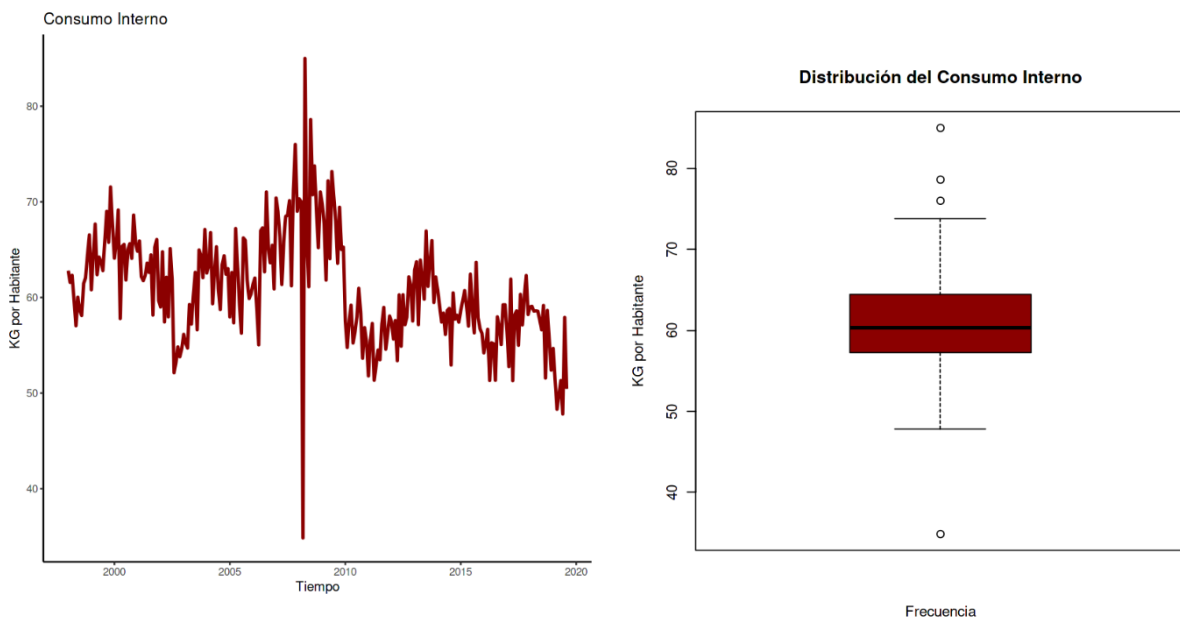
## Serie Consumo

### Análisis Series Originales (Punto 2)

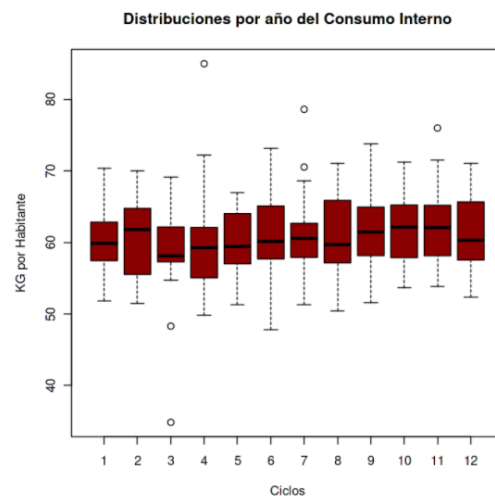
#### Serie Consumo

En la siguiente sección se analizará la segunda serie, la cual se refiere a Consumo interno en kg. per cápita. Se seguirá el mismo ordenamiento que aquel utilizado para el análisis de la primera serie.

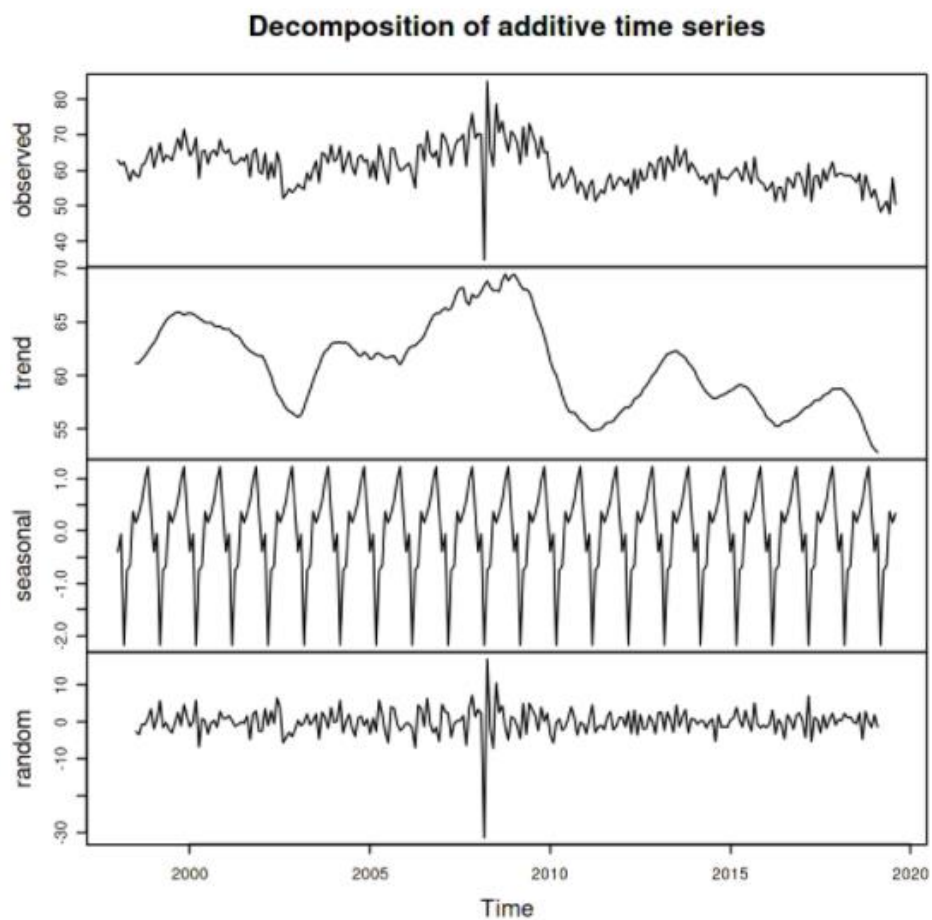
Por medio del siguiente gráfico, se observa el comportamiento de la serie. Según lo observado se detecta que, la presencia de cierta tendencia, un promedio que no mantiene constante a lo largo del tiempo, aspectos que indican la presencia de no estacionariedad. Asimismo, se detecta la presencia de valores extremos durante el 2008, lo cual también es capturado por medio del boxplot.



En cuanto a la estacionalidad, se detecta que habría un leve cambio en el consumo en el mes de febrero vs el resto de los meses.



Finalmente, se adjunta la descomposición de la serie en sus 3 componentes. Donde se observa la presencia de tendencia y estacionalidad.

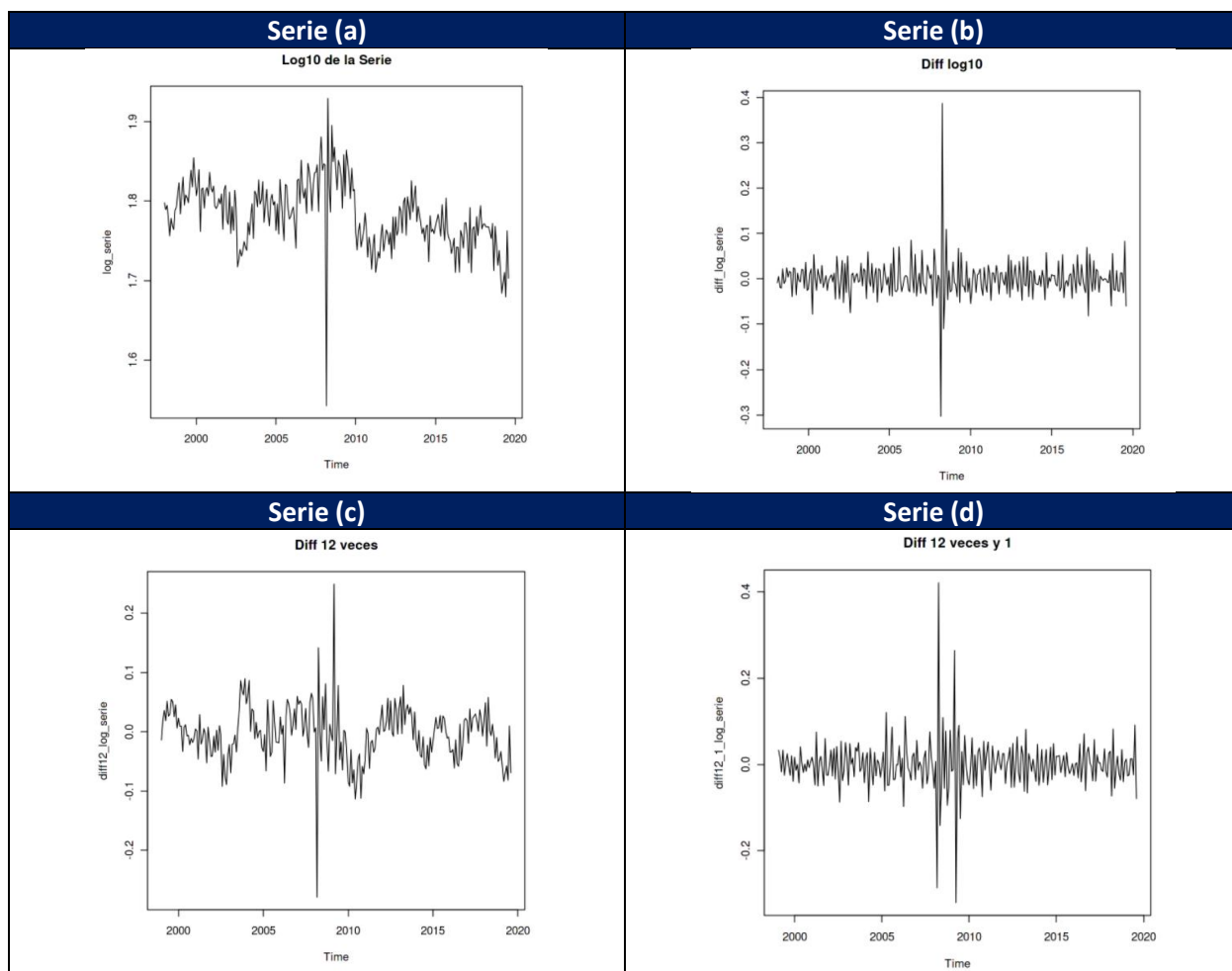


Siendo que la serie según la inspección visual realizada no sería estacionaria, se procedió a aplicar el test de Dicky Fuller, sus resultados indican que con un p-value de 0,249 no hay evidencia suficiente para rechazar la hipótesis nula, por lo cual estaríamos en presencia de una serie no estacionaria.

Habiendo analizado la no estacionariedad de la serie a analizar se procedieron a realizar los siguientes ajustes a la serie:

- a) Log 10 de la serie original (log\_serie)
- b) Primera diferenciación de la serie anterior (diff\_log\_serie)
- c) Diferencia estacional de la serie original (diff12\_log\_serie)
- d) Primera diferenciación de la serie diferenciada estacionalmente (diff12\_1\_log\_serie)

En la siguiente gráfica se visualizan los efectos que tuvieron las transformaciones indicadas en la serie bajo análisis.



Con excepción del caso (a), el resto de las series parecerían haber alcanzado la estacionariedad. Esto a su vez es corroborado por medio del test de Dicky-Fuller cuyos resultados coinciden con la inspección visual. Se rechaza la hipótesis nula de no estacionariedad para los casos (b), (c), (d).

	(a)	(b)	(c)	(d)
	log_serie	diff_log_serie	diff12_log_serie	diff12_1_log_serie
P value <sup>2</sup>	0.2185	0.01	0.01	0.01
Resultado	No Rechazo H0	Rechazo H0	Rechazo H0	Rechazo H0

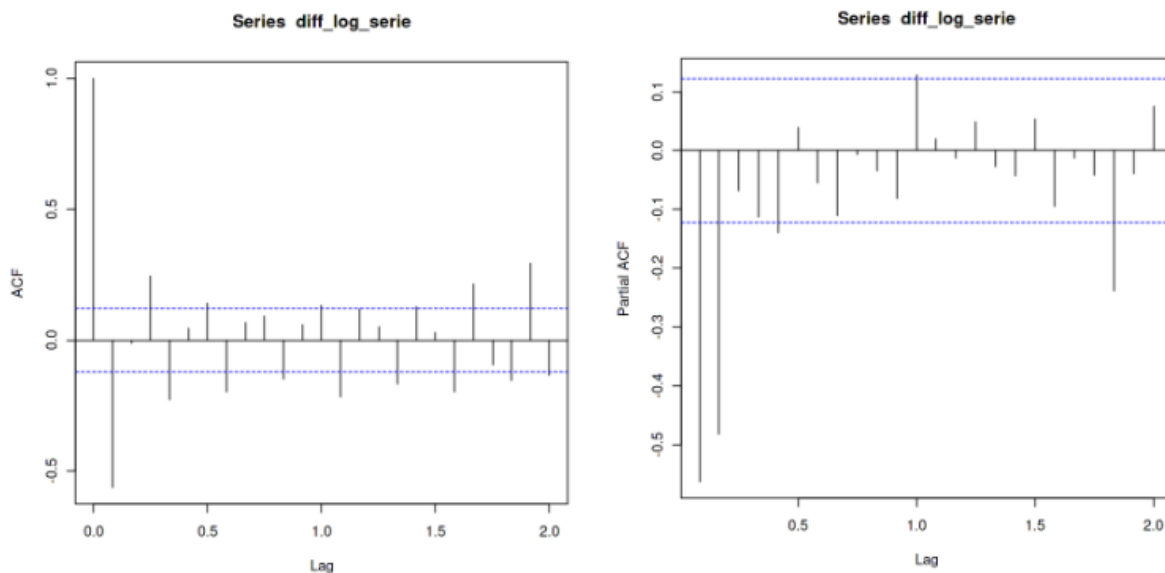
Como se observa en los gráficos, la serie original es no-estacionaria y por dicho motivo se han debido aplicar distintas transformaciones para alcanzar la estacionariedad. Si bien se ha experimentado con diferentes transformaciones, y siendo que los resultados alcanzados por medio de las otras no han generado resultados superiores, se ha optado por la serie más simple en lo que a transformaciones respecta, respetando el principio de parsimonia que afecta a estos modelos.

<sup>2</sup> Nivel de Significación considerado 5%

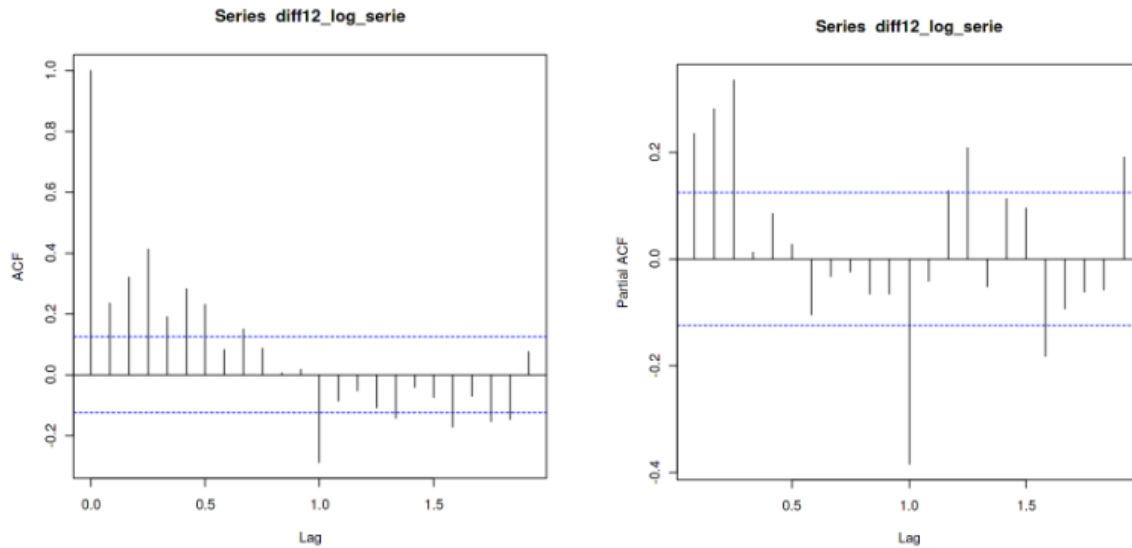
### Análisis FAS, FAC y FACP (Punto 3)

- (a) data: **log\_serie**
- (b) data: **diff\_log\_serie**
- (c) data: **diff12\_log\_serie**
- (d) data: **diff12\_1\_log\_serie**

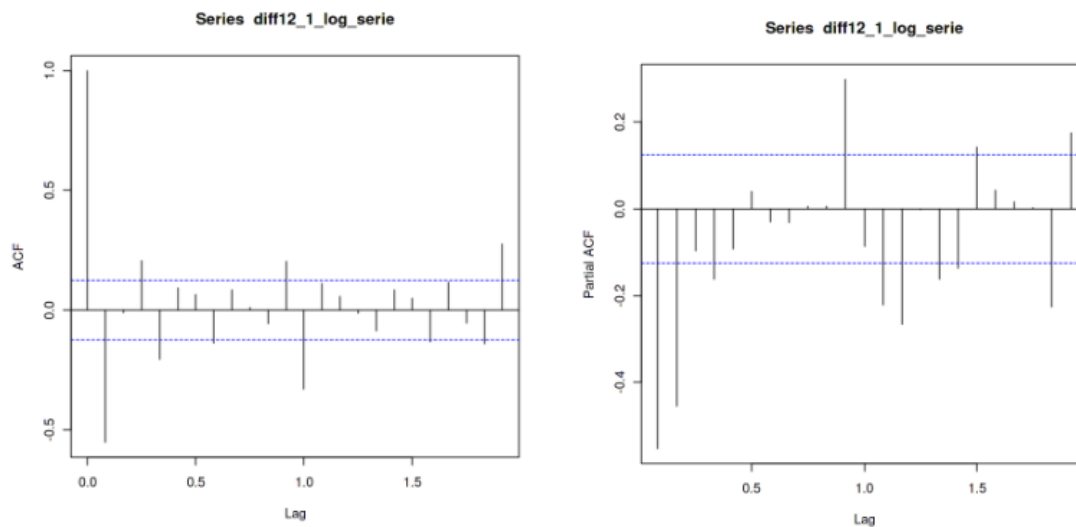
- (a) Dado que la serie original y esta serie son no estacionarias no se pueden obtener mayores conclusiones de los gráficos.
- (b) En el siguiente gráfico se observa caída marcada en el gráfico de autocorrelación, lo cual podría estar indicando un modelo Autorregresivo, el cual sería de orden 2 al observar su gráfico de autocorrelación parcial. Por otro lado, siendo que el gráfico de Autocorrelación parcial, también se observa caídas marcadas, lo cual sería indicativo de un modelo MA, siendo su orden de nivel 1, según se obtiene del gráfico de autocorrelación.



- (c) Se observa en el gráfico de autocorrelación que los coeficientes no se anulan marcadamente, sino que presentan un decrecimiento rápido de forma sinusoidal, por su parte el gráfico de autocorrelación parcial muestra 3 picos significativos al inicio y otro en el *lag* 12. Lo cual podría inferirse que se podría aplicar un modelo AR de orden 3 y evaluar la estacionalidad con un SARIMA siendo que llama la atención el *lag* de orden 12.



- (d) Se observan caídas marcadas en ambos gráficos, tanto en el de autocorrelación como en el de autocorrelación parcial, lo cual podría estar indicando la presencia de un modelo ARIMA, de orden  $p=1$  y  $q=2$ , orden 1 para el modelo autorregresivo ya que presenta 2 valores marcados en el gráfico de autocorrelación parcial y orden 1 en la parte MA ya que presenta un *lag* marcado en el gráfico de autocorrelación. De todas maneras, cabe señalar que resulta llamativo la presencia de un valor elevado en lag 11 lo cual generaría la necesidad de realizar diferentes pruebas para tratar de capturar la presencia de dicho valor dentro del modelo.



Los gráficos ACF y PAC muestran una aproximación del modelo a elegir, sin embargo, para elegir el mejor candidato resulta necesario realizar diversas pruebas que validen con cuál de ellos se obtienen mejores resultados para el caso bajo análisis.

## Modelo SARIMA (Punto 4)

Dados los resultados expuestos en las secciones anteriores, para esta serie se llegó a la conclusión de que los datos no presentan un comportamiento de estacionalidad. Sin embargo, se evaluarán algunos modelos *SARIMA* de estacionalidad de 12 períodos para poder rechazar la hipótesis de no presencia de estacionalidad. Para el desarrollo de los modelos para encontrar el óptimo se emplearán los datos de entrenamiento de la serie original aplicado el logaritmo y diferenciada una vez. El motivo de realizar la diferenciación se debe a que se optó por ingresar al modelo una serie estacionaria. Los modelos generados se detallan a continuación:

- |   |   |
|---|---|
| (1) $SARIMA(1,1,0)(0,0,0)_{12} \rightarrow AR(1)$     | (7) $SARIMA(0,1,2)(0,0,0)_{12} \rightarrow MA(2)$     |
| (2) $SARIMA(2,1,0)(0,0,0)_{12} \rightarrow AR(2)$     | (8) $SARIMA(1,1,1)(0,0,0)_{12} \rightarrow ARMA(1,1)$ |
| (3) $SARIMA(2,1,1)(0,0,0)_{12} \rightarrow ARMA(2,1)$ | (9) $SARIMA(1,1,2)(0,0,0)_{12} \rightarrow ARMA(1,2)$ |
| (4) $SARIMA(3,1,0)(0,0,0)_{12} \rightarrow AR(3)$     | (10) $SARIMA(0,1,1)(1,0,0)_{12}$                      |
| (5) $SARIMA(2,1,0)(1,0,1)_{12}$                       | (11) $SARIMA(0,1,1)(0,0,1)_{12}$                      |
| (6) $SARIMA(0,1,1)(0,0,0)_{12} \rightarrow MA(1)$     | (12) $SARIMA(0,1,3)(0,0,0)_{12} \rightarrow MA(3)$    |

Los resultados obtenidos para cada uno de los modelos mencionados anteriormente se detallan a continuación (los números que anteceden a cada uno de ellos corresponden al número de encabezado en cada una de las columnas:

	(1)	(2)	(3)	(4)	(5)	(6)
$a_0$	0.0006	0.0005	0.0005	0.0005	0.0006	0.0005
p-value	0.706	0.6595	0.5142	0.5146	0.5290	0.4879
$a_1$	-0.6057	-0.9839	-0.9978	-0.9733	-0.9728	
p-value	$4.08 \times 10^{-12}$	$2.20 \times 10^{-16}$	$2.54 \times 10^{-11}$	$2.20 \times 10^{-16}$	$2.20 \times 10^{-16}$	
$a_2$		-0.6109	-0.6194	-0.5930	-0.6024	
p-value		$1.71 \times 10^{-12}$	$2.43 \times 10^{-9}$	$2.05 \times 10^{-5}$	$2.65 \times 10^{-11}$	
$a_3$				0.0185		
p-value				0.8679		
$\beta_1$			0.0218			-0.7254
p-value			0.8810			$2.00 \times 10^{-16}$
$\beta_2$						
p-value						
$sar_1$					0.2492	
p-value					0.6691	
$sma_1$						
p-value						
$\sigma^2$	0.0005573	0.003487	0.003486	0.0003485	0.0003434	0.0004750
Log-likelihood	192.93	211.93	211.94	211.94	212.47	199.42
AIC	-379.87	-415.85	-413.87	-413.88	-412.95	-392.84

	(7)	(8)	(9)	(10)	(11)	(12)
$a_0$	0.0005	0.0005	0.0005			0.0005
p-value	0.6734	0.4816	0.6676			0.6581
$a_1$		-0.3961	-0.0226			
p-value		0.001002	0.9054			
$a_2$						
p-value						
$a_3$						
p-value						
$\beta_1$	-1.0428	-0.5465	-1.0276	-0.7048	-0.6968	-1.0607
p-value	$2.2 \times 10^{-16}$	$5.78 \times 10^{-12}$	$6.00 \times 10^{-11}$	$2.00 \times 10^{-16}$	$2.00 \times 10^{-16}$	$4.21 \times 10^{-16}$
$\beta_2$			0.5784			0.6209
p-value			$1.03 \times 10^{-5}$			0.001626
$\beta_3$						-0.0328
p-value						0.8506
$sar_1$				0.1983		
p-value				0.1015		
$sma_1$					0.2166	
p-value					0.06844	
$\sigma^2$	0.0003767	0.0004280	0.0003766	0.0004580	0.0004556	0.0003765
Log-likelihood	208.71	203.66	208.72	200.72	200.91	208.73
AIC	-409.43	-399.31	-407.44	-393.45	-393.82	-407.46

Las tablas anteriores muestran los coeficientes de los modelos al igual que algunas métricas que hacen posible la selección del modelo más apropiado.

El valor de *log-likelihood* suele ser mejor para los casos en que los modelos cuenten con más grados de libertad. Cabe recordar que por cada diferenciación se le está restando un grado de libertad al modelo propuesto. Sin embargo, esta no es una buena medida de selección de modelos ya que, considerando el principio de parsimonia, los modelos más complejos suelen tener mayores *log-likelihood*. Por este motivo, se debe emplear otro criterio de selección de modelos.

El criterio de selección de Akaike (AIC) considera que el modelo con el menor valor de AIC será el mejor de todos ellos.

Basándose en lo antepuesto, se considera que el modelo  $SARIMA(2,1,0)(0,0,0)_{12}$ , equivalente a un modelo  $ARIMA(2,1,0)$ , es el que arroja el menor valor de AIC y, además, es el modelo cuyos coeficientes resultan casi todos significativos (con *p-value* < 0.05). El único valor que no resulta significativo es el *intercept* el cual no tiene relevancia para la selección del modelo.

El código empleado para la elaboración de cada uno de los modelos se encuentra en el link indicado en la sección “Código Empleado para el Desarrollo del Presente Trabajo” de la presente presentación.

### Métricas de Performance (Punto 5)

Para la evaluación del performance de los distintos modelos, se seleccionó la métrica Mean Absolute Error (MAE – por sus siglas en inglés). La utilización de esta métrica conlleva a las siguientes ventajas:

- La medida del error está en la misma unidad de la serie temporal. En este caso como solo se está analizando una sola serie temporal, no habrá problemas de comparación de resultados.
- Es una medida de performance muy utilizada en series temporales ya que es fácil de computar y de interpretar.

Adicionalmente se explorarán otras métricas de performance y analizar cómo se comparan entre ellas.

Respecto a los datos utilizados, la serie fue dividida en dos conjuntos de datos, uno de entrenamiento (*train*) y otro de testeo (*test*), en la siguiente configuración:

	Train	Test	Total
Fecha inicio	01/2011	01/2018	01/2011
Fecha Fin	12/2017	12/2018	12/2018
Cant. Observaciones	84	12	96

#### Consideraciones en la selección de los datos:

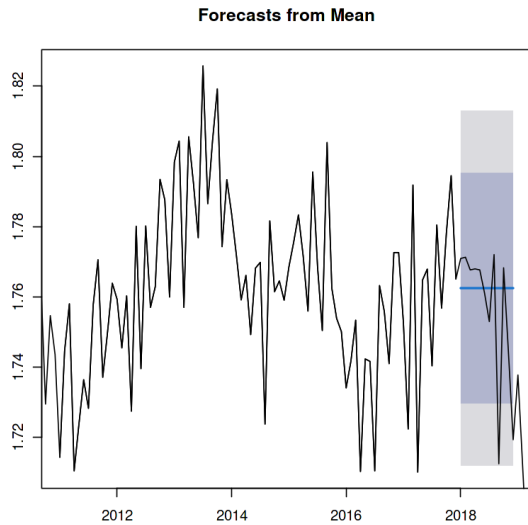
- Si bien se cuenta con información histórica desde el año 1998, se excluyeron desde esa fecha hasta diciembre de 2010 debido a que hubo grandes cambios en el comportamiento del consumo. El periodo excluido incluye los mismos eventos mencionados en la serie Faena los que afectaron la oferta de carne al mercado interno que sumado a un fuerte deterioro del poder adquisitivo y aceleración de los niveles de inflación desde es el periodo seleccionado hacen que el periodo seleccionado sea el más representativo.
- Cabe señalar que para el modelo (S)ARIMA seleccionado se ha utilizado la serie original aplicado el logaritmo, restringida al período indicado.
- Por otro lado, cabe indicar que, al entrenar el modelo con un AutoARIMA el modelo alcanzado fue  $SARIMA(2,0,0)(0,0,0)_{12}$ , el cual no era adecuado ya que presentaba correlación en los residuos. Ante tal situación, se procedió a elegir el mejor modelo del análisis realizado en el punto 4,  $SARIMA(2,1,0)(0,0,0)_{12}$ .



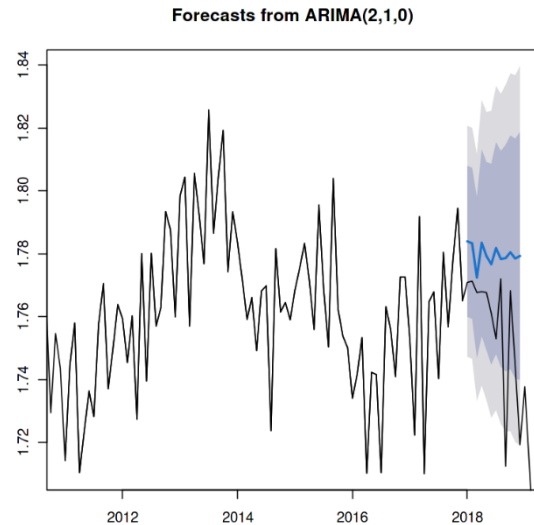
## Comparación con otros modelos estimados (Punto 6)

Se han elaborado seis modelos predictivos a partir de los datos de entrenamiento y testeo:

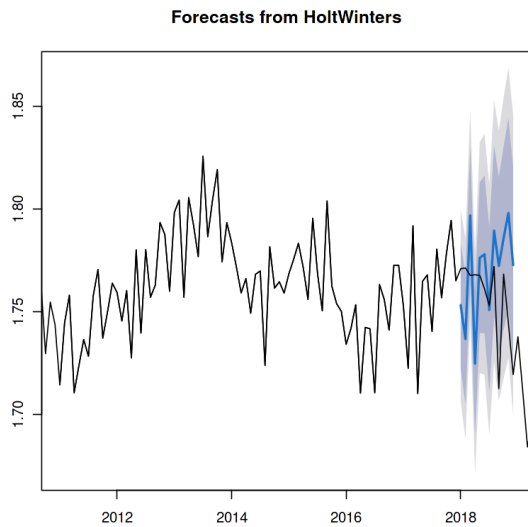
### Promedio Simple:



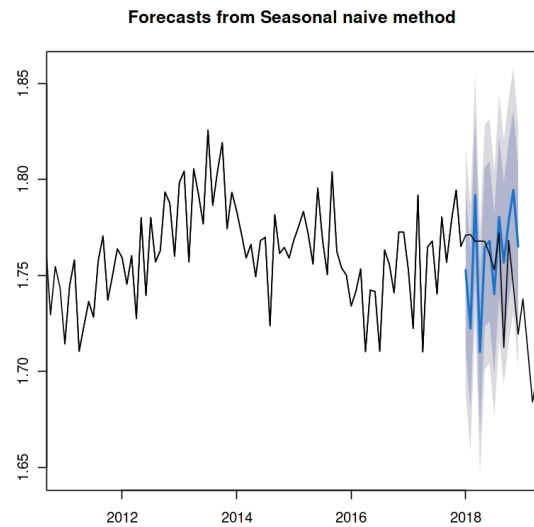
### ARIMA:



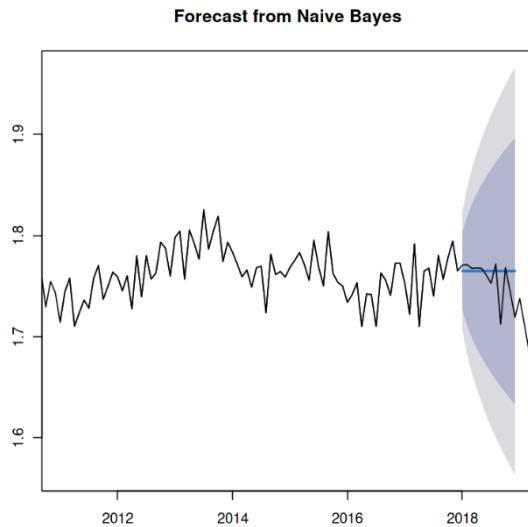
### Holt Winters:



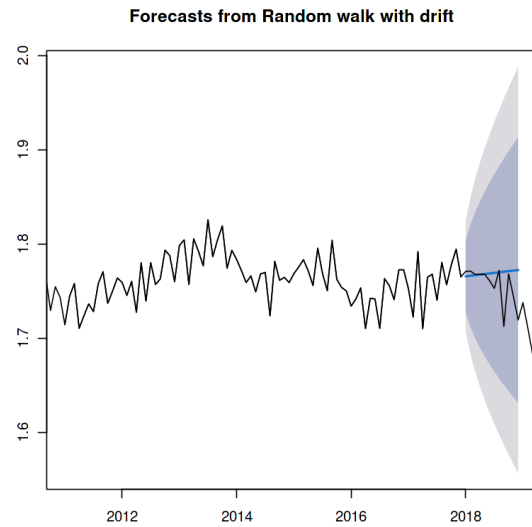
### Naïve Estacional:



## Naïve Bayes:



## Drift:



Modelo	ME	RMSE	MAE	MPE	MAPE	MASE
ARIMA(2,1,0)	-0,0061	0,0207	0,0142	-0,3624	0,8185	0,5423
Holt Winters	-0,0041	0,0337	0,0275	-0,2459	1,5706	1,0488
Naïve Bayes	-0,0087	0,0216	0,0138	-0,5097	0,7953	0,5259
Drift	-0,0132	0,0348	0,0295	-0,7687	1,6899	1,1270
Naïve Estacional	-0,0127	0,0247	0,0150	-0,7368	0,8642	0,5706
Promedio Simple	1,7552	1,7553	1,7552	99,9328	99,9328	61,3707

En base a la inspección de las predicciones de los gráficos, ningún modelo pudo predecir con certeza las caídas pronunciadas (cambio de tendencia) al final del período de testeo, sin embargo, los modelos que mejor se ajustan ARIMA (2,1,0) y Naive Bayes siendo que arrojan las mejores métricas de ser comparados con los otros modelos. En particular en el caso del score en MASE, el cual según lo indicado por (Athanasopoulos & Hyndman, 2018), un error escalado es menor que uno si surge de un pronóstico mejor que el pronóstico ingenuo promedio calculado sobre los datos de entrenamiento.

Considerando que el modelo de Naive Bayes es un modelo más simple y alcanza similares e incluso resultados levemente superiores al modelo más complejo, ARIMA, es que se procede a elegir como mejor modelo al primero.

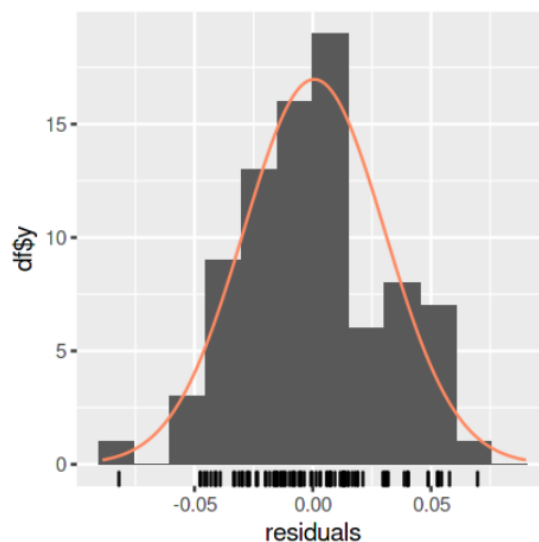
Esto nos hace concluir que para capturar el cambio de tendencia se debería abordar líneas de investigación con un modelo par que nos permita incorporar otras variables que estén relacionadas con el cambio de tendencia como ser el poder adquisitivo, inflación, entre otras.

### Análisis de Diagnóstico de las Series Temporales (Punto 7)

Se procederá al análisis de Naive Bayes dado que presenta un comportamiento similar al modelo ARIMA(2,1,0) y como este es un modelo más simple, es preferible sobre uno con mayor complejidad.

Para un modelo como el de Naive Bayes por el tipo de modelo que es requiere analizar la normalidad de los residuos para evaluar la calidad de sus predicciones.

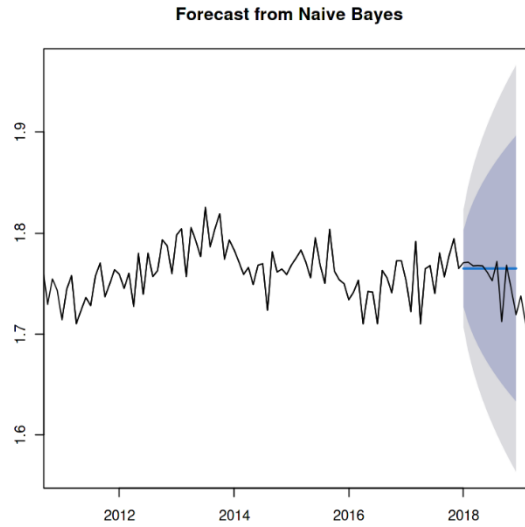
Prueba	Jarque - Bera	Anderson Darling	Shapiro Wilk
P-value	0.8526126	0.5681762	0.597984



En base a las pruebas de normalidad sobre los residuos efectuadas en el apartado anterior, se puede observar normalidad en los residuos analizados, dándonos un punto de partida aceptable para este modelo.

### Pronósticos con el modelo Seleccionado (Punto 8)

Siendo que ambos modelos muestran similar performance se adjuntan las alcanzadas por el modelo Naive Bayes, donde se puede apreciar que las predicciones efectuadas no logran pronosticar el cambio de tendencia presente en la serie, lo que hace que sea necesaria la incorporación de otro conjunto de variables que ayuden al desarrollo de un nuevo modelo que contenga información adicional para la construcción del mismo:



Las predicciones generadas con este modelo son las siguientes:

Date	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Feb-2018	1.7651	1.7270	1.8031	1.7069	1.8232
Mar-2018	1.7651	1.7113	1.8188	1.6828	1.8473
Apr-2018	1.7651	1.6992	1.8309	1.6643	1.8658
May-2018	1.7651	1.6890	1.8411	1.6488	1.8814
Jun-2018	1.7651	1.6800	1.8501	1.6350	1.8951
Jul-2018	1.7651	1.6719	1.8582	1.6226	1.9075
Aug-2018	1.7651	1.6645	1.8657	1.6112	1.9189
Sep-2018	1.7651	1.6575	1.8726	1.6006	1.9296
Oct-2018	1.7651	1.6510	1.8791	1.5906	1.9395
Nov-2018	1.7651	1.6448	1.8853	1.5812	1.9490
Dec-2018	1.7651	1.6390	1.8912	1.5722	1.9579

## Bibliografía

1. 4.1 Seasonal ARIMA models / STAT 510. (s. f.). PennState: Statistics Online Courses. Recuperado 23 de julio de 2021, de <https://online.stat.psu.edu/stat510/lesson/4/4.1>
2. Athanasopoulos, G., & Hyndman, R. J. (2018). *Forecasting: Principles and Practice* (2nd Edition). Monash University.
3. Boyadjian, C. (2021, abril 7). ¿Adiós al asado? El consumo de carne vacuna es el más bajo en 18 años. *El Cronista*. <https://www.cronista.com/economia-politica/sigue-cayendo-el-consumo-de-carne-vacuna-y-se-ubica-en-49-3-kg-anuales-por-persona/>
4. Del Rosso, R. (2021, junio). *Análisis de Series Temporales* [Diapositivas de Clase]. Maestría en Ciencia de Datos, Universidad Austral.
5. En 2020 el consumo de carne vacuna fue el más bajo en cien años. (2021, enero 7). *Télam Digital*. <https://www.telam.com.ar/notas/202101/540926-en-2020-el-consumo-de-carne-vacuna-fue-el-mas-bajo-en-cien-anos.html>
6. Enders, W. (2014). *Applied Econometric Time Series* (4th Edition). Wiley Series.
7. Peña, D. (2010). *Análisis de Series Temporales* (1ra Edición). Alianza Editorial.
8. *Time series—Arima sarima model mathematical formula*. (s. f.). Cross Validated. Recuperado 24 de julio de 2021, de <https://stats.stackexchange.com/questions/90476/arima-sarima-model-mathematical-formula>
9. Uriel, E. (1985). *Análisis de Series Temporales: Modelos ARIMA* (1ra Edición). Paraninfo.