

Modelización de serie de tiempo de precios mayoristas de manzana de Uruguay.

Emanuelle Marsella, Maximiliano Saldaña

Junio 2022

Resumen

Se modeliza la serie de precios mensuales mayoristas de manzana mediante la metodología SARIMA propuesta por Box y Jenkins. Se llegó a un conjunto de modelos, en los cuales se incluyen elementos autorregresivos (AR) y/o de medias móviles (MA), tanto regulares como estacionales. El modelo principal fue especificado usando herramientas de identificación tales como las funciones de autocorrelación y pruebas de significación, teniendo un orden 2 en su componente AR regular y estacional. Los otros modelos identificados por métodos automáticos (en general usando el criterio de Akaike corregido -AICc-) tienen especificaciones similares. Se realizan predicciones a 12 pasos en una muestra de prueba con los distintos modelos, siendo los de mejor desempeño el modelo identificado manualmente y el modelo identificado mediante AICc, ambos sin intervenciones. Los modelos en los cuales se incluyeron intervenciones tienen un desempeño predictivo marcadamente peor.

1. Análisis descriptivo

La serie a ser estudiada es la de precios promedio mensuales del kilo de manzana en la Unidad Agroalimentaria Metropolitana (ex Mercado Modelo). Los precios de los distintos rubros transados en este mercado mayorista de frutas y hortalizas son relevados por el Observatorio Granjero dos veces a la semana, los lunes y los jueves, mediante encuestas a los distintos vendedores informantes. Se relevan precios por distintas variedades, calidades y calibres. Empleando los distintos precios obtenidos los técnicos del Observatorio llegan a un precio de referencia por consenso.

Se cuentan con los datos desde enero de 2013 a mayo de 2022 y se considerará el promedio mensual de los precios, por lo que se cuenta con 113 observaciones. En lugar de emplear los datos bisemanales o el promedio semanal se opta por la frecuencia mensual debido a la dificultad de emplear las herramientas de modelado SARIMA para tales tipos de series, en particular para el tratamiento de la estacionalidad.

Resulta de particular interés al modelizar los precios de este alimento de importancia la predicción, debido a la incidencia que pueden tener aumentos o bajas en los precios en la producción del bien y en el consumo de los uruguayos.

En la Figure 1 se presenta el gráfico de la serie a ser trabajada. La impresión inicial que da es que la serie presenta cierto patrón estacional anual, donde los precios comienzan altos para luego descender hasta el segundo trimestre de los años y luego tienden a elevarse hasta el final de año. Esto se puede observar mejor en el gráfico de los precios coloreados por año y el gráfico de la evolución de los precios año a año por mes de la Figure 2. El año 2020 presenta los precios más altos de la serie y no se observa la caída inicial de precios de los años anteriores sino un aumento sostenido dentro del año. Esto se puede deber al impacto económico que causó la pandemia de Coronavirus, que llegó a nuestro país en marzo dicho año. Ya para 2021 y lo que va de 2022 parece haber una vuelta a patrones previos.

Para ahondar en el análisis descriptivo se realizó la descomposición de la serie en tendencia/ciclo, estacionalidad y componente irregular. En la Figure 8 del anexo se presentan las series de los componentes resultado de una descomposición mediante *STL* (Seasonal Trend Descomposition using LOESS) por separado. Esta metodología consiste en suavizaciones recursivas de la serie basadas en regresiones y un sistema de pesos (Cleveland et al. 1990).

Se puede apreciar una marcada estacionalidad anual y en los últimos 6 años un ciclo corto que se repite cada dos años. La fuerza de la estacionalidad, definida como (Hyndman and Athanasopoulos 2018):

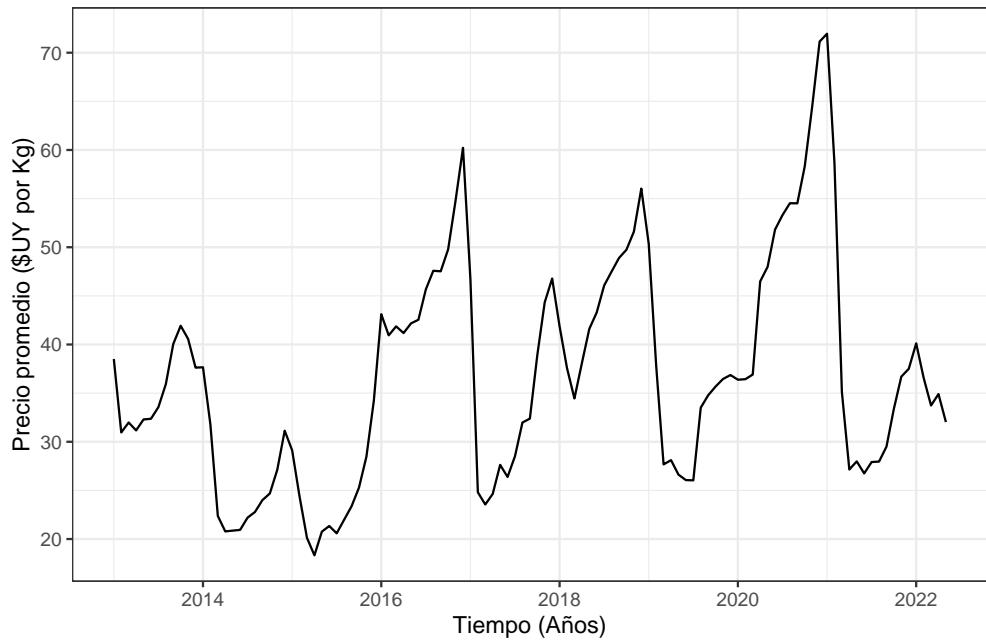


Figura 1: Serie de precios mensuales del Kg de manzana en pesos uruguayos.

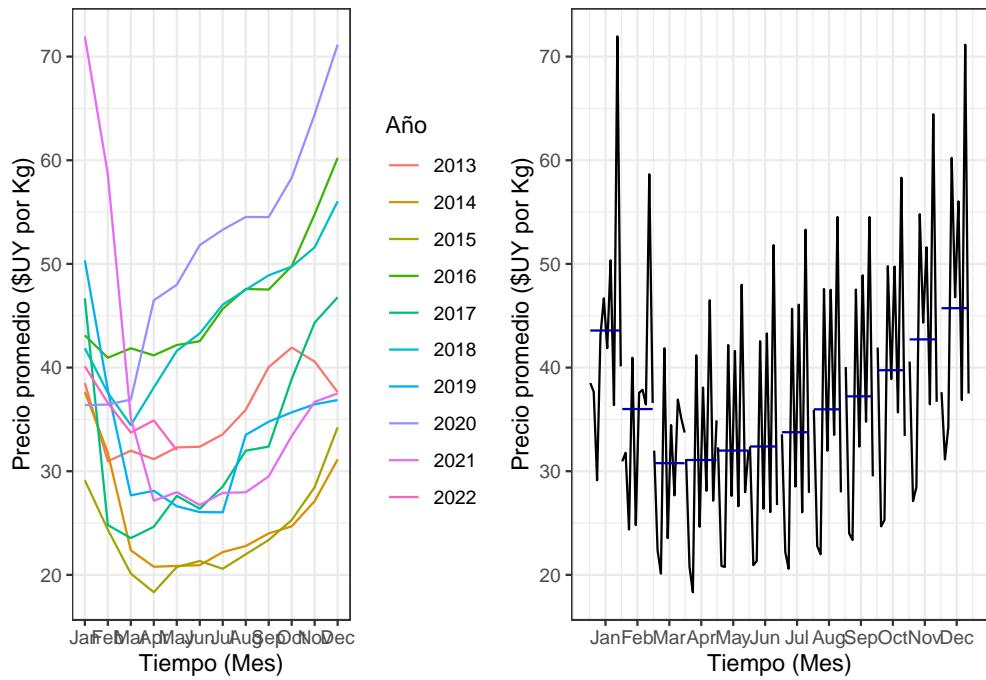


Figura 2: Serie de precios mensuales del Kg de manzana en pesos Uruguayos.

$$F_s = \max \left(0, 1 - \frac{Var(R_t)}{Var(R_t + S_t)} \right)$$

toma el valor 0.51 (entre más cercano a 1 más fuerte el componente). Esto refuerza la necesidad de considerar la estacionalidad a la hora de especificar un modelo.

2. Metodología y resultados

2.1. Muestra de entrenamiento y de prueba

Resulta de interés que el modelo ajustado a la serie sea de utilidad para la predicción. Para poder evaluar la calidad de las predicciones, una manera que busca replicar el proceso de obtención de nuevos datos es dividir la serie en una muestra de entrenamiento y una de prueba. La primera se emplea para ajustar el modelo, a partir del cual se realizarán las predicciones. Se dejan las últimas 12 observaciones para la muestra de prueba, que son los precios que van desde junio de 2021 a mayo de 2022. Debe tenerse en cuenta que el periodo del final de la muestra de entrenamiento y también la muestra de prueba están enmarcados en el contexto de gran incertidumbre que presenta la pandemia, por lo que deberá tenerse especial cuidado con el tratamiento de atípicos y las conclusiones que se tomen sobre las predicciones.

2.2. Identificación

La primera fase para el modelado ARIMA de una serie de tiempo en el marco de la metodología de Box y Jenkins es la identificación del modelo, que consiste en determinar en un principio en detectar la estructura de autocorrelación, la cantidad de parámetros con la que contará la especificación, si la serie necesita diferenciación y si resultará necesaria alguna otra transformación.

2.2.1. Transformación logarítmica

La transformación logarítmica de una serie de tiempo puede tener como resultado una reducción del error de predicción en el caso de que estabilice la varianza (Lütkepohl and Xu 2009). Esto se cumple en particular cuando la varianza aumenta con la media de la serie, lo cual no es el caso de los precios de manzana, que si bien presentan una varianza que aumenta en el tiempo no parece haber una tendencia creciente clara. Por lo tanto, esta transformación no resultaría aconsejable de aplicar. Para confirmar esto, se considera la transformación de Box-Cox, donde siendo y la variable transformada y x la variable a transformar:

$$y_t = \begin{cases} \frac{x_t - 1}{\lambda} & \text{si } \lambda \neq 0 \\ \ln x_t & \text{si } \lambda = 0 \end{cases}$$

Donde el parámetro λ se estima por máxima verosimilitud. En el caso de la serie planteada, dicho parámetro toma el valor -0.59, por lo que la transformación logarítmica no resulta adecuada.

2.2.2. Autocorrelación y autocorrelación parcial

Como primer elemento a considerar en la identificación de un modelo ARIMA resultan útiles las funciones de autocorrelación y autocorrelación parcial. Al graficarlas en ellas se puede visualizar la estructura de dependencia temporal de la serie trabajada e indicios de la estacionariedad o ausencia de ella. La función de autocorrelación (ACF) se define como:

$$\rho(t, t+j) = \frac{Cov(Y_t, Y_{t+j})}{\sigma_t \sigma_{t+j}}$$

Mientras que la función de autocorrelación parcial (PACF) es la autocorrelación entre Y_t y Y_{t+j} una vez se quita el efecto de las correlaciones intermedias que hay entre ambas variables (Rodríguez 2022).

Estas funciones se deben estimar a partir de los datos, obteniéndose la autocorrelación muestral. En la Figure 9 del Anexo se presentan los primeros 72 valores de ambas funciones para la serie original de precios de manzana, junto con el intervalo de confianza de la prueba de hipótesis $H_0: \rho_j = 0$ vs $H_1: \rho_j \neq 0$. En la ACF se puede distinguir que las primeras 5 autocorrelaciones resultan significativas y presentan un decaimiento exponencial y que también son significativas las que están en torno al rezago 24 y entre los

rezagos 36 y 48. Lo primero puede ser indicio de una estructura autorregresiva subyacente y lo segundo un indicio de estacionalidad anual o bianual.

Por otro lado, en la PACF se aprecia que los valores para los primeros dos lags son significativamente distintos a 0. Esto puede ser un indicio de que hay una estructura autorregresiva, posiblemente de segundo orden, en los datos.

2.2.3. Dominio de las frecuencias

Desde la perspectiva del dominio de las frecuencias de la serie se considera esta última en su expresión trigonométrica, mediante una suma ponderada de funciones periódicas coseno y seno. El espectro poblacional puede resultar de utilidad para observar la estructura de variabilidad de la serie, dado que el área por debajo del mismo es la variabilidad asociada a las frecuencias consideradas.

En la Figure 3 se presenta la estimación no paramétrica del espectro poblacional de la serie de precios de manzana (izquierda) y las diferencias de los precios de manzana (derecha). En estas estimaciones se hace uso del periodograma muestral, que es la estimación del espectro poblacional a partir de las autocovarianzas muestrales y luego se realiza un promedio ponderado de sus valores mediante un *kernel* a fines de suavizar el resultado, que en general resulta difícil de interpretar inicialmente. En este caso se pondera con el *kernel* de Daniell modificado ponderando de a 3 valores del periodograma muestral 2 veces sucesivas.

Se puede apreciar para ambos casos como las frecuencias menores, aquellas asociadas a períodos más largos (teniendo en cuenta que $p = 2\pi/w$, siendo p el período y w la frecuencia) son aquellas que acumulan mayor variabilidad. Esto puede considerarse como otro indicio de una dependencia temporal estacional entre las observaciones de la muestra. También es posible apreciar como la primera diferencia, que puede ser vista como un filtro que se aplica a la serie, afecta la forma del espectro. Si bien las menores frecuencias siguen siendo las más acentuadas, mediante la aplicación de este filtro el espectro se ve acentuado en torno a las frecuencias 1, 2 y 3.

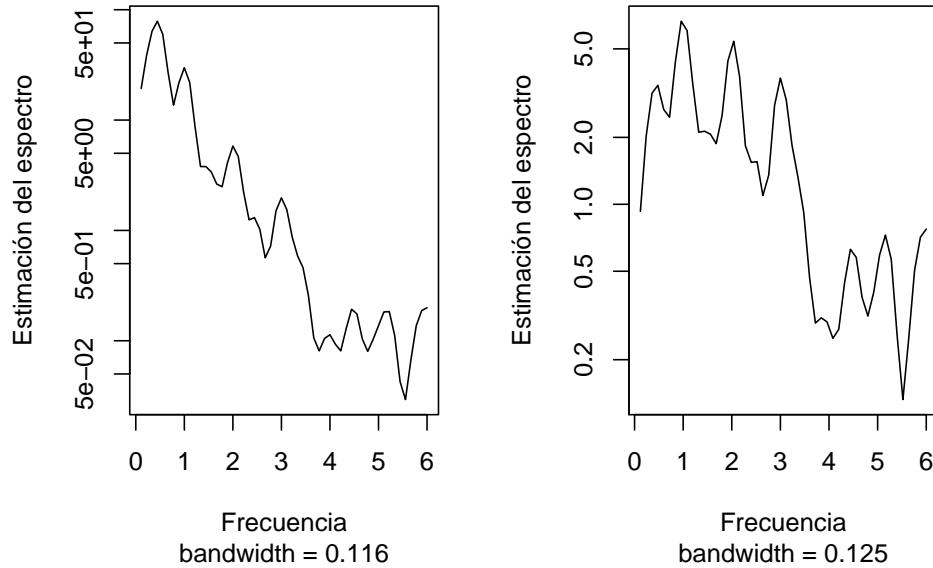


Figura 3: Periodograma muestral, estimación no paramétrica. 3 spans aplicados 2 veces.

2.2.4. Tests de raíces unitarias

Dado que las funciones de autocorrelación dieron indicios de que el proceso no es estacionario, resulta de interés poner a prueba si el proceso es $I(1)$, es decir, si cuenta con una raíz unitaria. En dicho caso el proceso sería no estacionario la cual implicaría que no puede ser modelado en el marco de los ARMA.

Hay múltiples pruebas de hipótesis que han sido desarrolladas con el propósito de identificar raíces unitarias. Dos de ellos son el de Dickey-Fuller aumentado y el de Phillips-Perron. En el caso del primero se especifica el proceso proceso estocástico subyacente como:

$$Y_t = \rho Y_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$$

y se contrasta $H_0: \rho = 1$ vs $H_1: \rho < 1$

- . Empleamos el estadístico $(\hat{\rho} - 1)/\sigma_{\hat{\rho}}$, que si ρ es menor a 1 en valor absoluto se distribuye normal asintóticamente y si es igual a 1 debe usarse una distribución empírica tabulada por Fuller. El test además toma en cuenta la posible autocorrelación de los errores incluyendo rezagos de la variable en la regresión auxiliar para mayor robustez.

Por otro lado la propuesta de prueba de raíz unitaria de Phillip-Perron se basa en la de Dickey y Fuller, pero además los compatibilizan con la presencia de heteroscedasticidad y/o autocorrelación de los errores.

Para la serie planteada, al observar los p-valores, ambos tests llevan a no rechazar la hipótesis de raíz unitaria al 95 % de confianza. Esto lleva a concluir que una primera diferencia resulta necesaria para llevar la serie a la estacionariedad.

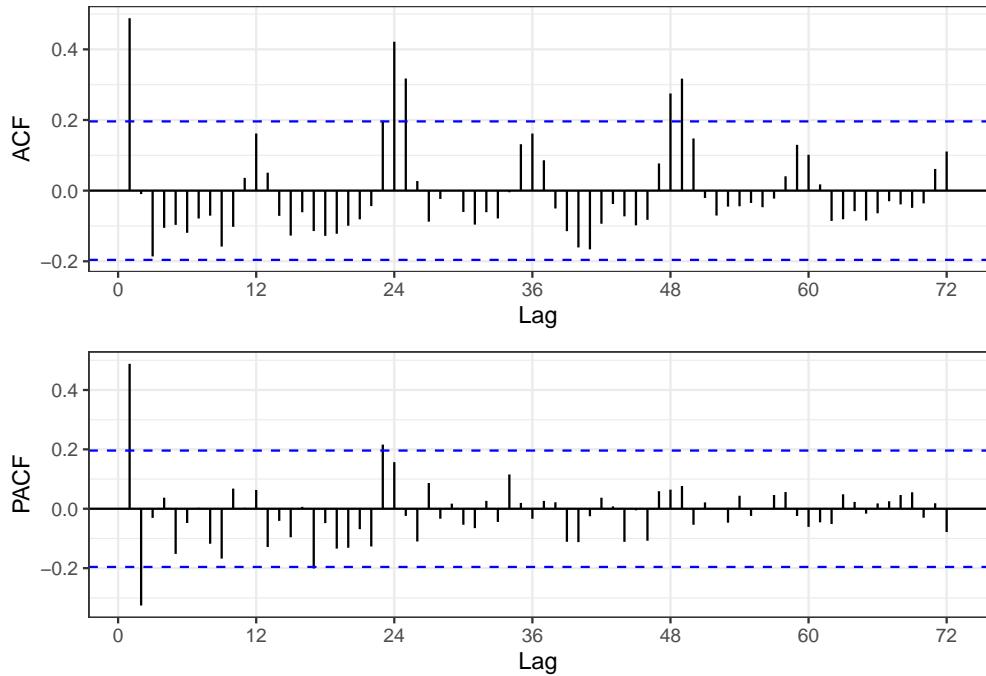


Figura 4: Autocorrelograma y autocorrelograma parcial de la serie de primeras diferencias de precios de manzana.

Observando los nuevos autocorrelogramas (Figure 4) se puede notar como de entre los primeros valores de la ACF solo el primero resulta significativamente distinto a 0 (lo que puede indicar un componente de medias móviles de primer orden) y lo mismo se da para los valores en torno al lag 24 y 48. Esto último vuelve a indicar la presencia de una estructura de dependencia estacional (indicio de un componente estacional autorregresivo de orden 2). Por otro lado, en la PACF, se aprecia que solo los dos primeros valores son significativos, lo cual puede indicar cierta estructura autorregresiva, posiblemente de orden 2. A pesar de esto, hay que tener en cuenta que en las aplicaciones prácticas al tratar de identificar el orden de un proceso ARIMA mediante la FAC y PACF la distinción entre especificaciones puede volverse difusa y no resulta tan claro la cantidad de parámetros a elegir.

Teniendo en cuenta este acercamiento metodológico, inicialmente se plantea un modelo $ARIMA(2, 1, 1)(2, 0, 0)$ (en adelante "modelo manual").

2.2.5. Selección mediante criterios de información

Una forma alternativa de elegir la especificación del modelo, esto es, la cantidad de parámetros y de diferencias, es mediante los criterios de información. Con estos criterios se busca elegir entre una selección de modelos aquel con una mayor verosimilitud, penalizando por la cantidad de parámetros que tenga.

Uno de los más empleados, y en particular el que se considera en la función `auto.arima()` del software R para la estimación, es el criterio de información de Akaike corregido (AICc). Su fórmula para un determinado modelo es:

$$AICc = T \log \hat{\sigma}_{MV}^2 + T \frac{1 + k/T}{1 - (k + 2)/T}$$

Siendo T el número de observaciones, k el de parámetros y $\hat{\sigma}_{MV}^2$ el estimador máximo verosímil de la varianza de los errores. El modelo seleccionado es aquel con el menor valor del AICc, lo cual a nivel algorítmico se hace empleando la selección *stepwise* a partir de un conjunto inicial de modelos (Hyndman and Athanasopoulos 2018).

El modelo seleccionado mediante esta metodología (de ahora en adelante “Modelo AICc”) resulta en un *ARIMA*(0, 1, 2)(0, 0, 2). Se puede observar que no se considera un componente autorregresivo, pero si un componente estacional de medias móviles y al igual que el anterior modelo, una diferencia (el algoritmo de `auto.arima()` realiza múltiples tests de raíz unitaria). Además, se considera un componente estacional de medias móviles de orden 2, en lugar de autorregresivo como era en el caso del modelo manual.

2.3. Estimación

Luego de haber identificado la especificación del modelo, el paso que sigue es estimar sus parámetros, dado que no es posible conocer sus verdaderos valores al ser una construcción teórica. Para esto se recurre a la estimación por máxima verosimilitud, donde las estimaciones obtenidas son las que maximizan la probabilidad de que se haya observado la muestra con la que se cuenta. Es necesario asumir que los errores son gaussianos, un supuesto fuerte pero las estimaciones resultantes que emanen a partir de hacerlo serán razonables aunque no se cumpla (Hamilton 1994).

En particular, el método empleado es la estimación máximo verosímil condicional, donde se supone que la primera observación de la serie es determinística y se maximiza la verosimilitud condicionada a dicha observación. Esto simplifica las expresiones de las funciones, y si el tamaño de muestra es razonablemente grande, la primera observación no tendrá gran efecto sobre la verosimilitud estimada.

2.3.1. Pruebas de significación de los parámetros

Para completar la especificación se realizan pruebas de hipótesis sobre la significación de los parámetros estimados. La hipótesis para cada uno de ellos, siendo λ un parámetro cualquiera del modelo:

$$\begin{aligned} H_0) \lambda &= 0 \\ H_1) \lambda &\neq 0 \end{aligned}$$

Donde el estadístico empleado es:

$$z = \hat{\lambda}/\hat{\sigma}_\lambda$$

Para el que se cumplirá la Normalidad asintótica debido a que la estimación se realizó por máxima verosimilitud (Hamilton 1994, 143).

Realizando esta prueba para los coeficientes del *modelo manual* todos resultan significativos al 5 % de confianza con la excepción del parámetro de medias móviles (MA) y el primer parámetro autorregresivo estacional (sAR). Se opta entonces por eliminar el componente MA de la especificación y restringir el primer parámetro sAR a 0.

2.3.2. La especificación del *modelo manual* final

La forma del modelo identificado de manera manual es:

$$\Phi_2(L^{12})\phi_2(L)\Delta^1 Y_t = \varepsilon_t$$

Donde:

$$\Phi_2(L^{12}) = 1 - \Phi_2 L^{24}$$

$$\phi_2(L) = 1 - \phi_1 L - \phi_2 L^2$$

$$\Delta^1 = 1 - L$$

Y se supone en un principio que los residuos son de media 0, homoscedásticos, incorrelacionados y:

$$\epsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma^2)$$

Los parámetros resultantes de la estimación para el *modelo manual*, son:

- Parte autorregresiva: $\hat{\phi}_1 = 0,557$ y $\hat{\phi}_1 = -0,277$
- Parte medias móviles: $\hat{\theta}_1 = -0,4157$
- Parte autorregresiva estacional: $\hat{\Phi}_2 = 0,415$
- Varianza de las estimaciones: $\hat{\sigma}^2 = 15,82$

2.4. Diagnóstico

Para evaluar si el modelo es adecuado se ponen a prueba los supuestos realizados sobre los residuos, sobre los cuales se sostiene.

2.4.1. Media 0 de los residuos

Para poner a prueba este supuesto se realiza el contraste no paramétrico de Wilcoxon, donde se prueba $H_0: \mu_\epsilon = 0$ contra $H_1: \mu_\epsilon \neq 0$. En el caso de la muestra no se rechaza la hipótesis nula al 99 % de confianza, por lo que la evidencia está a favor de que su media incondicional es 0.

2.4.2. Incorrelación de los errores

Un supuesto cuyo cumplimiento es clave es la incorrelación de los errores, debido a que toda estructura de dependencia que no se esté captando en el modelo irá a parar a los residuos, el proxy con el que contamos para conocer el comportamiento de los errores. Si hay correlaciones entre los errores, la especificación del modelo todavía no capta el comportamiento de los datos.

Para evaluar el cumplimiento de este supuesto un contraste usual es el de autocorrelación conjunta de Ljung-Box, donde la hipótesis nula es que los residuos son incorrelacionados contra la alternativa de que no lo son. El estadístico para realizar el contraste con los primeros h rezagos es:

$$Q_{L-B}(h) = (T(T+2)) \frac{\sum_1^h (\hat{\rho}_j)^2}{T-j}$$

El cual se distribuye asintóticamente χ^2_{h-m} bajo la hipótesis nula, donde m es el número de parámetros del modelo.

Para el *modelo manual*, $m = 3$. En la ?? se presentan los p-valores de la prueba conjunta, aumentando sucesivamente el número de rezagos considerados. Se puede observar que considerando hasta los primeros 36 rezagos, no se rechaza la hipótesis nula de incorrelación de los residuos, por lo que el modelo resulta apropiado en este aspecto.

2.4.3. Homoscedasticidad

Otro supuesto que se realizó fue sobre la varianza de los errores, la cual se desea constante. Para ponerlo a prueba se realiza el test desarrollado en McLeod and Li (1983). Es básicamente una prueba de Box-Ljung sobre los residuos al cuadrado del modelo. La hipótesis nula de esta prueba es la homoscedasticidad entre los k rezagos considerados.

En la Figure 5 se presentan los valores de los p-valores de la prueba para los primeros 19 rezagos. Se puede apreciar que considerando en conjunto hasta el décimo rezago hay presencia de heteroscedasticidad. Esto es una indicación que los modelos del tipo SARIMA pueden no ser los más adecuados y se tendría que recurrir a los del tipo ARCH/GARCH donde se busca modelizar la varianza de los errores.

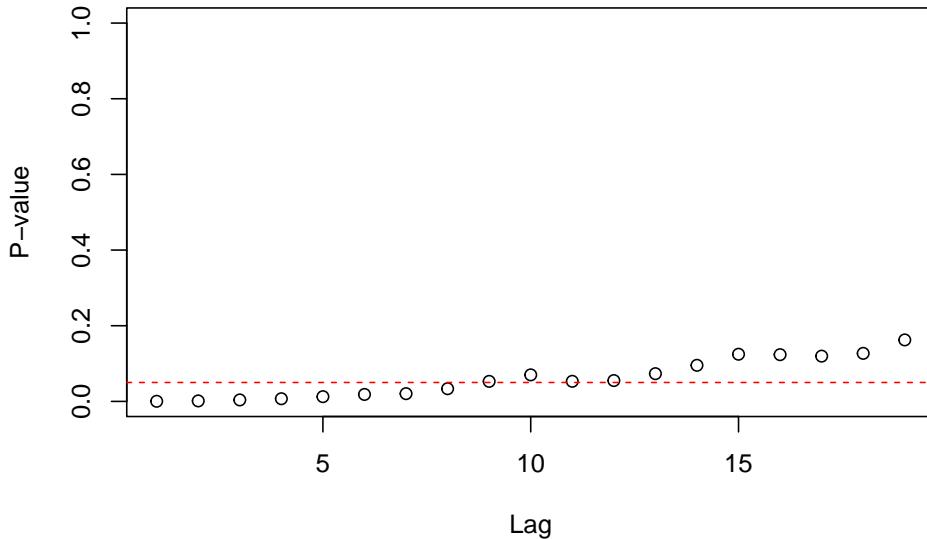


Figura 5: P-valores del test de McLeod-Li para los primeros 19 rezagos

2.4.4. Normalidad

Finalmente, se evalúa si el modelo cumple el supuesto de que sus residuos tienen una distribución gaussiana. Este supuesto entra afecta a la hora de calcular intervalos de predicción, que expresan la incertidumbre que se tiene respecto a la predicción puntual realizada. En caso de que se cumpla la normalidad los intervalos de confianza al $(1 - \alpha)100\%$ para la predicción a s pasos podrán ser calculados con la fórmula:

$$\hat{Y}_{t+s}|Y_t \pm z_{\alpha/2} \sqrt{E[Y_{t+s} - E(Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \dots)]^2} = \hat{Y}_{t+s}|Y_t \pm z_{\alpha/2} \sqrt{\sigma^2(1 + \sum_{j=1}^{s-1} \Psi_j^2)}$$

Donde \hat{Y}_{t+s} es la predicción puntual, $z_{\alpha/2}$ es el percentil $\alpha/2$ de una distribución normal, σ^2 es la varianza de los errores y Ψ_j es el j -ésimo parámetro de la representación $MA(\infty)$ del modelo (que siempre es posible hallarla cuando el modelo es estacionario por el teorema de Wold (Rodríguez 2022)).

Hay múltiples formas de contrastar la normalidad, dos de ellas aplicadas en el presente trabajo son los contrastes de Jarque-Bera y de Shapiro-Wilk. Ambas contrastan las hipótesis $H_0:\varepsilon_t \sim N(0, \sigma^2)$ versus $H_1:\varepsilon_t \not\sim N(0, \sigma^2)$, pero difieren en la forma que lo hacen. En el caso de Jarque-Bera, se calculan los coeficientes de simetría y curtosis de los residuos para contrastarlos con los de una distribución normal teórica. Por otro lado, el test de Shapiro-Wilk fue desarrollado en Shapiro and Wilk (1965) y sigue el enfoque del análisis de varianza y hace uso de un estadístico cuyos valores críticos son calculados mediante simulación Monte-Carlo.

Estas pruebas son particularmente sensibles a la presencia de valores atípicos, que tienen una tendencia a alargar las colas de la distribución de los residuos y haciendo que estas consecuentemente se alejen de la normalidad. Como se mencionó anteriormente.

En ambos contrastes se rechaza la hipótesis nula de normalidad al 95 % de confianza, por lo que se recurrió a incorporar el efecto de los atípicos en el *modelo manual*.

2.4.5. Intervención

Dado que el *modelo manual* no cumple con el supuesto de normalidad, y atribuimos esto a la presencia de algunos outliers, optamos por realizar una intervención a dichos outliers para intentar corregir el

no cumplimiento de este supuesto. Para ello, consideramos como outliers a los residuos e_t que no se encuentran en el intervalo $(Q_1 - 3IQR, Q_3 + 3IQR)$ Hyndman (n.d.). Siguiendo esta metodología, los outliers son los valores correspondientes a enero y febrero de 2017, así como a marzo de 2021. Optamos por considerar al residuo de enero de 2017 como un outlier del tipo de cambio transitorio (TC), y a los otros dos como outliers aditivos (AO). Los primeros representan un impacto cuyo efecto decrece progresivamente a lo largo de un período determinado, mientras que los segundos consideran un impacto en esa observación puntual donde se observa. Incluimos estas intervenciones como regresores del modelo, y luego de testear la significación de sus parámetros comprobamos que los mismos son significativos al 5 %, por lo cual conservamos estas intervenciones en el modelo.

Nuevamente se pusieron a prueba los supuestos para el modelo intervenido, y si bien en este caso se cumple el supuesto de normalidad, no se cumple el supuesto de no autocorrelación de los residuos, que sí se cumplía para el modelo sin intervenir. Por lo tanto, optamos por deshechar la intervención y mantener el modelo manual sin intervenir, dado que si bien para este no se cumple la normalidad, eso se puede subsanar utilizando bootstrap para la estimación de los intervalos de predicción, el cual es una técnica que consiste en estimar la varianza de los datos a partir de un remuestreo de los residuos observados.

2.4.6. Modelos alternativos

Una vez se llegó al final de la etapa de diagnóstico del modelo planteado de forma manual, optamos por evaluar otros modelos alternativos seleccionados a partir de métodos automáticos, para tener varias opciones a la hora de evaluar el desempeño de las predicciones y compararlos con el modelo manual.

2.4.6.1. Modelo seleccionado por AICc Se seleccionó un modelo utilizando el Criterio de Información de Akaike corregido ($AICc$), que tiene la forma $ARIMA(0, 1, 2)(0, 0, 2)$. Al realizar los contrastes de significación de los parámetros observamos que los parámetros de medias móviles son significativos, así como el parámetro de medias móviles estacionales de segundo orden, pero no el de primer orden. En cuanto a los supuestos sobre los residuos, al igual que sucedía para el *modelo manual*, no se rechaza la hipótesis de no autocorrelación ni tampoco la de media nula, pero la homocedasticidad no se cumple para todos los rezagos, y se rechaza la hipótesis de normalidad.

Luego, se realizó una intervención sobre este modelo para corregir el efecto de algunos outliers que se detectaron, y se llegó a un modelo que cumple todos los supuestos sobre los residuos (excepto el de homocedasticidad para algunos rezagos).

2.4.6.2. Modelo seleccionado por tso Se seleccionó otro modelo a partir de las intervenciones aplicadas con el procedimiento automático de detección de outliers de la función *tso* presente en el paquete *tsoutliers* (*modelo tso*). El resultado del mismo es un $ARIMA(0, 1, 1)(2, 0, 0)$, que tiene 7 intervenciones (5 TC y 2 AO), esto último se debe a que al ejecutar el algoritmo se restringió el tipo de outliers posibles a estos dos, luego de haber obtenido malos resultados incluyendo también outliers del tipo cambio de nivel (LS). Todos los parámetros son significativos, excepto el correspondiente a la estacionalidad autorregresiva de orden 1. Por otro lado, se cumplen todos los supuestos que se suelen pedir sobre los residuos, incluyendo el de la homocedasticidad para todos los rezagos (que en modelos anteriores no se cumplía).

2.4.6.3. Modelo seleccionado por tsoutliers Otra opción considerada es la de un modelo intervenido con por métodos de la función *tsoutliers* del paquete *forecast*, que reemplaza los outliers a través de interpolación lineal (*modelo clean*). En este caso, se selecciona un modelo $ARIMA(1, 1, 0)(2, 0, 0)$, cuyos parámetros son significativos excepto el autorregresivo estacional de orden 1, pero que no cumple los supuestos de normalidad y no autocorrelación de los residuos, por lo cual optamos por no utilizarlo.

2.4.6.4. Modelo seleccionado por ets Finalmente se selecciona un modelo de una clase distinta a los *ARIMA*, de la familia de los de suavizado exponencial, que hacen uso de promedios ponderados de las observaciones (*modelo ets*), y tienen como base modelos estructurales o espacio-estado. Este tipo de modelos permiten una estructura de varianzas que cambia en el tiempo, por lo cual no se requiere el supuesto de homocedasticidad de los residuos. Por su parte, el supuesto de no autocorrelación de los mismos sí se cumple.

3. Predicción

Una vez se cuenta con un modelo (o una selección de modelos) que se suponen correctos, tanto en especificación como en estimación, se puede proceder a la predicción de los valores futuros de la serie condicionando a los valores con los cuales se cuenta. Para los datos empleados, los precios mayoristas de manzana, resulta de particular interés la predicción a futuro, dado que lo que ocurría con el precio incidirá tanto en la producción como en el consumo del bien y puede ser un insumo para la toma de decisiones a nivel de política económica.

3.1. Predicción puntual

Se busca tener una predicción puntual de los precios y uno o varios intervalos de confianza (para considerar distintos niveles de confianza). Para esto se va a partir de los modelos seleccionados y se va a buscar realizar una predicción óptima, entendida como aquella que minimiza una función de pérdida. Esta función expresa el costo en el cual se incurre ante el error en la predicción y la idea es minimizar su valor esperado.

La función de costos que se empleará será la cuadrática, es decir $C(e) = ae^2$, donde a es una constante positiva y e el error. Este tipo de función penaliza en mayor medida los errores de gran magnitud más que los pequeños. Minimizar el valor esperado de esta función implica minimizar el error cuadrático medio a un horizonte predictivo de s pasos, condicional a las t observaciones con las que se cuenta:

$$E(Y_{t+1} - Y_{t+1|t}^*)^2$$

Como resultado de la minimización de la expresión anterior, se llega a que la predicción óptima es la esperanza condicional, es decir: $Y_{t+1|t}^* = E(Y_{t+1}|Y_t)$. En el caso de un proceso SARIMA estacionario, se lo puede expresar como un $MA(\infty)$ y la expresión del pronóstico óptimo resultante es:

$$\hat{E}[Y_{t+s}|\varepsilon_t, \varepsilon_{t-1}, \dots] = \psi_s \varepsilon_t + \psi_{s+1} \varepsilon_{t-1} + \psi_{s+2} \varepsilon_{t-2} + \dots$$

3.2. Intervalos de predicción

Adicionalmente resulta de importancia cuantificar la incertidumbre que tenemos sobre la predicción puntual mediante intervalos de predicción, los cuales se pueden interpretar como la dispersión de posibles resultados relativa a dichos valores puntuales calculados, para un nivel de confianza dado (Ericsson 2001). En el caso que se cumpla el supuesto de normalidad, se puede emplear la expresión indicada en la sección subsubsection 2.4.4 y en caso contrario se pueden obtener mediante métodos de bootstrap (como se discutió en la sección subsubsection 2.4.5).

3.3. Medidas de error

Para cuantificar el error que se comete al comparar predicción contra valor verdadero, se recurre a cuatro medidas de error, cada una con sus particularidades que se desarrollan a continuación.

La raíz del error cuadrático medio (RMSE) y el error absoluto medio (MAE) son dos medidas ampliamente usadas, debido a su relevancia en el modelado estadístico (Hyndman and Koehler 2006). Por ejemplo, se minimiza el RMSE para obtener las predicciones como se desarrolló en la subsección anterior. Ambas medidas son dependientes de la escala, por lo que no pueden usarse para comparar el error de modelos que no están en la misma escala y además son sensibles a los outliers. Sea $e_s = Y_{t+s} - Y_{t+s|t}^*$:

- Raíz del error cuadrático medio (RMSE) = $\sqrt{\sum_{i=1}^s e_i^2}$
- Error absoluto medio (MAE) = $\frac{1}{s} \sum_{i=1}^s |e_i|$

Por otro lado, otro tipo de medidas son las basadas en errores porcentuales. Dichos errores están dados por: $p_s = 100e_s/Y_s$. Este tipo de medidas son independientes de la escala, lo que las hace idóneas para comparar el desempeño predictivo de modelos con distintas escalas (como es el caso de nuestra selección). Hay que tener en cuenta que si la serie presenta valores cero o infinitos esta medida no está definida, cuando tiene valores cercanos a cero presenta una distribución sesgada y en el caso de la media de los valores absolutos de los porcentajes de error (MAPE); penaliza más severamente los errores positivos que los negativos. Se emplea esta última medida ya que no se cuenta con valores negativos, cero o cercanos a cero en la muestra:

- Media de los valores absolutos de los porcentajes de error (MAPE) = $\frac{1}{s} \sum_{i=1}^s |p_i|$

Para enfrentarse a las problemáticas de las medidas de error antes presentadas, en Hyndman and Koehler (2006) se propone una medida donde se usan los errores escalados con el MAE dentro de la muestra: $q_s = e_t / (\frac{1}{n-1} \sum_{j=1}^n |Y_j - Y_{j-1}|)$ y la fórmula de la medida es:

- Error escalado absoluto medio (MASE) = $\sum_{i=1}^s |q_i|$

3.4. Predicción dentro de la muestra de entrenamiento

Como una forma de medir el ajuste del modelo a la serie trabajada en general se calculan las predicciones a un paso del modelo dentro de la muestra de entrenamiento, comparándose con el valor verdadero de la serie. Hay que tener en cuenta que el modelo con el que se hacen estas predicciones es ajustado con todos los datos y no con los que se tienen al momento de la predicción, por lo que no resulta una buena referencia para evaluar el modelo en cuanto a sus predicciones fuera de la muestra.

En la Table 1 se presentan los valores de las medidas planteadas evaluadas dentro de la muestra para los cinco modelos seleccionados. Si bien las diferencias no son substanciales, se puede observar que el modelo con ajuste de outliers automático *modelo tso* presenta el mejor desempeño en todas las medidas (aunque su escala es distinta porque no toma en cuenta diferencias como los otros SARIMA, las predicciones son consideradas a niveles, por lo que se pueden emplear todas las medidas). El que le sigue es el *modelo AICc intervenido*. El modelo identificado de manera manual y el obtenido mediante el AICc tienen un desempeño muy similar.

Modelo	RMSE	MAE	MAPE	MASE
modelo_manual	3.9	2.7	8.0	0.20
modelo_aicc	4.0	2.8	8.2	0.20
modelo_aicc_int	3.1	2.4	7.0	0.17
modelo_tso	2.3	1.8	5.0	0.13
modelo_ets	4.1	2.6	7.5	0.19

Tabla 1: Errores de los distintos modelos en la muestra de entrenamiento.

3.5. Predicción fuera de la muestra de entrenamiento

Una forma de evaluar las predicciones que busca simular el proceso de predecir al futuro pero teniéndose la posibilidad de evaluar la calidad de las predicciones es mediante la predicción fuera de la muestra de entrenamiento. En este caso se predice a 12 pasos (a un año dado que se trabaja con datos mensuales) y se compara con los 12 valores que se apartaron para conformar la muestra de prueba. Luego, se evalúan las medidas de error anteriormente discutidas.

Tomando en cuenta esta forma de medir los errores, ahora el modelo con mejor desempeño en todas las medidas es el seleccionado mediante el AICc sin intervenciones por atípicos (Ver cuadro 2). Por otra parte, cabe destacar como los modelos intervenidos, que antes eran los de mejor desempeño dentro de la muestra, ahora presentan el peor desempeño, siendo el modelo seleccionado mediante el AICc e intervenido el que peor predice a 12 pasos. La conclusión que se puede hacer de esto es que no necesariamente el mejor ajuste dentro de la muestra tendrá como resultado una mejor predicción fuera de ella y que incluso se puede estar incurriendo en un sobreajuste. También debe ser tenido en cuenta que en ambos modelos intervenidos se identificaron outliers del tipo TC hacia el final de la muestra empleada para la estimación, lo cual tiene un efecto que perdura en el horizonte predictivo considerado.

Modelo	RMSE	MAE	MAPE	MASE
modelo_manual	4.5	3.9	12.4	0.28
modelo_aicc	3.2	2.6	8.6	0.19
modelo_aicc_int	13.1	12.7	39.4	0.92
modelo_tso	6.3	5.9	18.0	0.42
modelo_ets	4.5	3.6	10.9	0.26

Tabla 2: Errores de los distintos modelos en la muestra de testeо.

Para visualizar las diferentes predicciones en el cuadro 6 se presentan sus valores puntuales junto al verdadero valor que tomaron los precios. Se puede observar como las predicciones captan el comportamiento general de la serie.

miento de los precios (por ejemplo, el pico de comienzos de 2022), con la excepción del *modelo AICc intervenido* y el *modelo manual*, a pesar de que este último es el segundo mejor en términos de medidas de error. Salta a la vista que si bien se consideró la predicción a 12 pasos, una predicción con un horizonte más corto puede arrojar resultados distintos en cuanto al modelo con mejor desempeño.

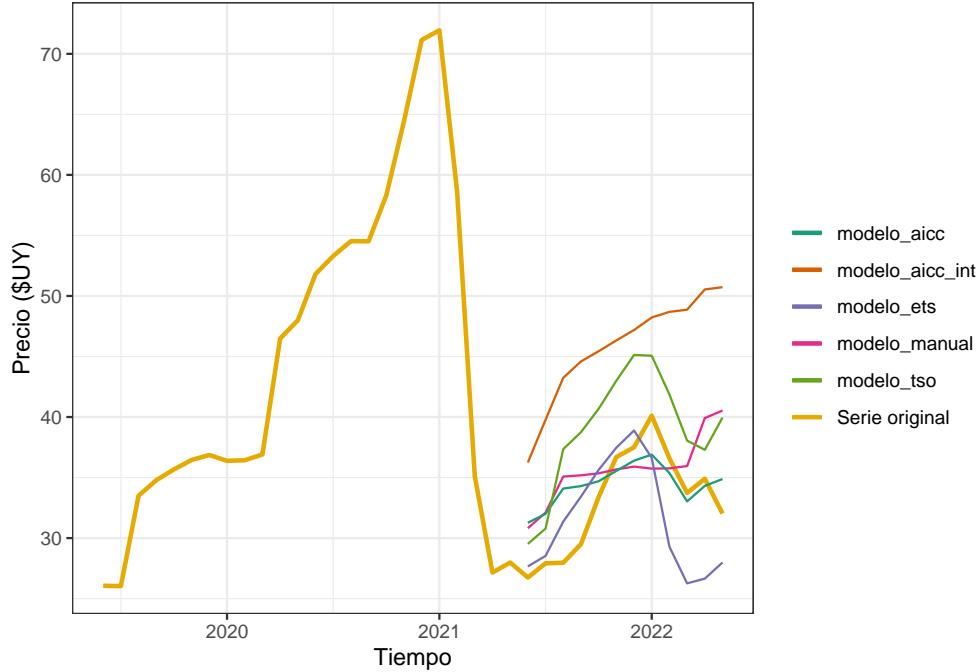


Figura 6: Predicciones a 12 pasos del precio de manzana, muestra de prueba.

En la Figure 7 se presentan las predicciones a 12 pasos junto a sus respectivos intervalos de predicción al 80 % de confianza del modelo de mejor desempeño predictivo; el seleccionado mediante AICc (*modelo AICc*), y el identificado manualmente (*modelo manual*), que fue el segundo con mejor desempeño predictivo.

Se puede observar que a simple vista *modelo AICc* captura mejor el comportamiento de la serie real y de hecho su error es menor. No obstante, sus intervalos de predicción al 80 % son más amplios que los de *modelo manual*, que puede hacer que inicialmente nos inclinemos por este último modelo, aunque el primer valor de la muestra de prueba cae fuera de su intervalo (por \$0.01). Ambos son obtenidos mediante bootstrap, por lo que se puede afirmar que hay una mayor incertidumbre asociada al primero de estos modelos.

Como es de esperarse, al alejarse la predicción del último valor de la serie que se empleó para estimar los intervalos crecen en amplitud. Hacia el final del horizonte predictivo, en el paso 12, el de *modelo aicc* va de 1.05 a 63.6; amplitud que resulta muy amplia pero es entendible al haber tantos factores fuera de control en un lapso de tiempo tan largo.

4. Conclusiones

Como resultado del proceso de modelado de la serie de precios de manzana siguiendo la metodología de Box y Jenkins se llegó a un modelo autorregresivo con una diferencia, con un componente estacional también autorregresivo. Este cumplió los supuestos que se hicieron en su especificación con la excepción del de normalidad y homoscedasticidad. El primero afecta los intervalos de predicción y pudo ser subsanado mediante *bootstrap*, mientras que el segundo puede ser una señal de que debe ser considerado un modelo que modele la heteroscedasticidad, como por ejemplo los del tipo ARCH/GARCH. Como alternativa a este modelo trabajado de manera “manual” se planteó un conjunto de modelos seleccionados mediante metodologías de carácter automatizado, como lo son la selección mediante criterios de información.

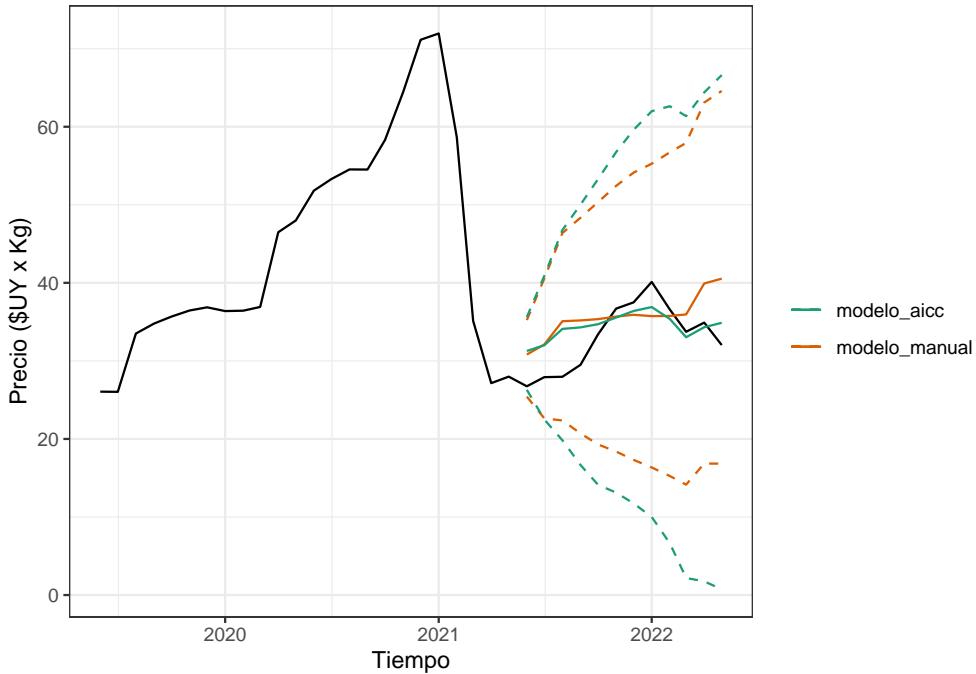


Figura 7: Predicciones puntuales del precio de manzana e intervalos de predicción al 80 %.

La serie cuenta con valores atípicos los cuales se intentaron subsanar incluyendo intervenciones, pero en el caso de los modelos que las incluían se observó mayor error a la hora de la predicción fuera de la muestra. En particular, la intervención de valores atípicos al final de la serie aumentaba el error en el que se incurría, por más que en estos casos se cumplían los supuestos subyacentes a los modelos y el ajuste dentro de la muestra era mejor que para los no intervenidos. Esto lo atribuimos al hecho de que incluir una cantidad alta de intervenciones implica un sobreajuste del modelo, en el caso del *modelo tso* un 6 % de las observaciones fueron intervenidas. La modelización con mejor desempeño predictivo fue la elegida mediante el criterio de Akaike corregido y sin intervenciones, siendo la segunda mejor la especificada manualmente.

El hecho de que la predicción se haga con el trasfondo de la pandemia de Coronavirus aumenta la incertidumbre de una forma que es difícil cuantificar, porque todavía no se tiene la distancia temporal suficiente para concluir si causó un quiebre en la estructura de los precios, en particular de la manzana que fueron los tratados en el presente trabajo. De hecho, las predicciones fuera de la muestra las realizamos para el periodo entre junio de 2021 y mayo de 2022, por lo cual los últimos 3 meses se ubican dos años después del inicio de la pandemia, lo cual puede explicar el pobre desempeño del modelo para dichos meses teniendo en cuenta el orden 2 de los componentes estacionales especificados.

5. Bibliografía

Anexo

5.1. Descomposición STL

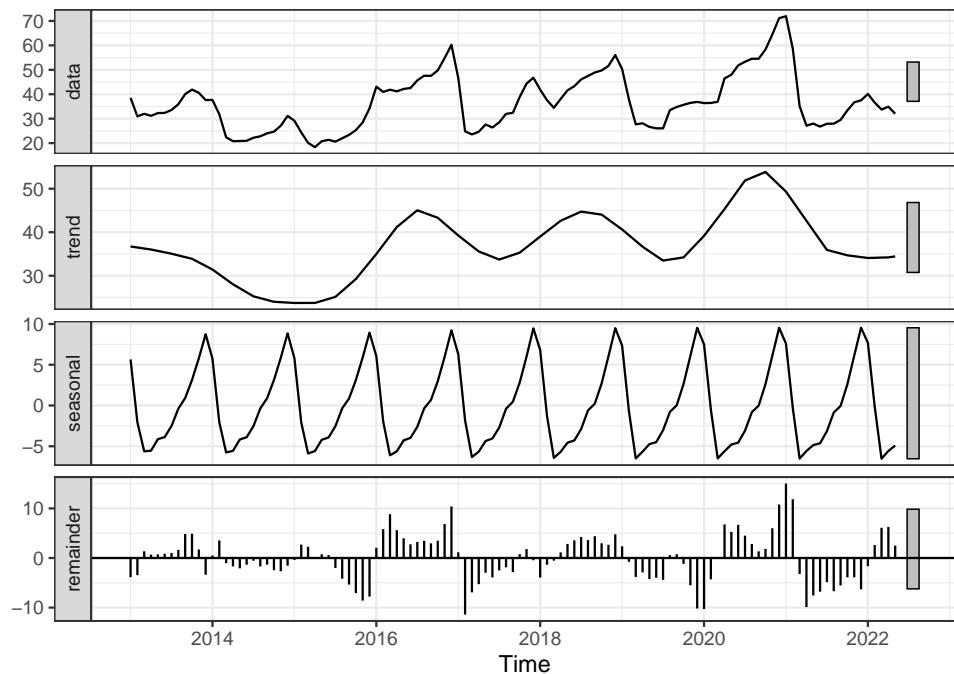


Figura 8: Descomposición STL del precio de manzana.

5.2. ACF a niveles

- Cleveland, Robert, William Cleveland, Jean McRae, and Irma Terpenning. 1990. "STL: A Seasonal-Trend Decomposition Procedure Bases on Loess." *Journal of Official Statistics* 6 (1): 3–73.
- Ericsson, Neil. 2001. "Forecast Uncertainty in Economic Modeling." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.266494>.
- Hamilton, James. 1994. *Time Series Analysis*. Princeton University Press.
- Hyndman, Rob J. n.d. "Detecting Time Series Outliers." Accessed June 29, 2021. <https://robjhyndman.com/hyndtsight/tsoutliers/>.
- Hyndman, Rob J., and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*. 2nd ed. Australia: OTexts.
- Hyndman, Rob J., and Anne B. Koehler. 2006. "Another Look at Measures of Forecast Accuracy." *International Journal of Forecasting* 22 (4): 679–88. <https://doi.org/10.1016/j.ijforecast.2006.03.001>.
- Lütkepohl, Helmut, and Fang Xu. 2009. "The Role of the Log Transformation in Forecasting Economic Variables." Working paper No. 2591. Munich, Alemania: CESifo.
- McLeod, A., and W. Li. 1983. "Diagnostic Checking Arma Time Series Models Using Squared-Residual Autocorrelations." *Journal of Time Series Analysis* 4 (4): 269–73. <https://doi.org/10.1111/j.1467-9892.1983.tb00373.x>.
- Rodríguez, Silvia. 2022. *Series Cronológicas: Notas de Curso*.
- Shapiro, S., and M. Wilk. 1965. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika* 52 (3/4): 591. <https://doi.org/10.2307/2333709>.

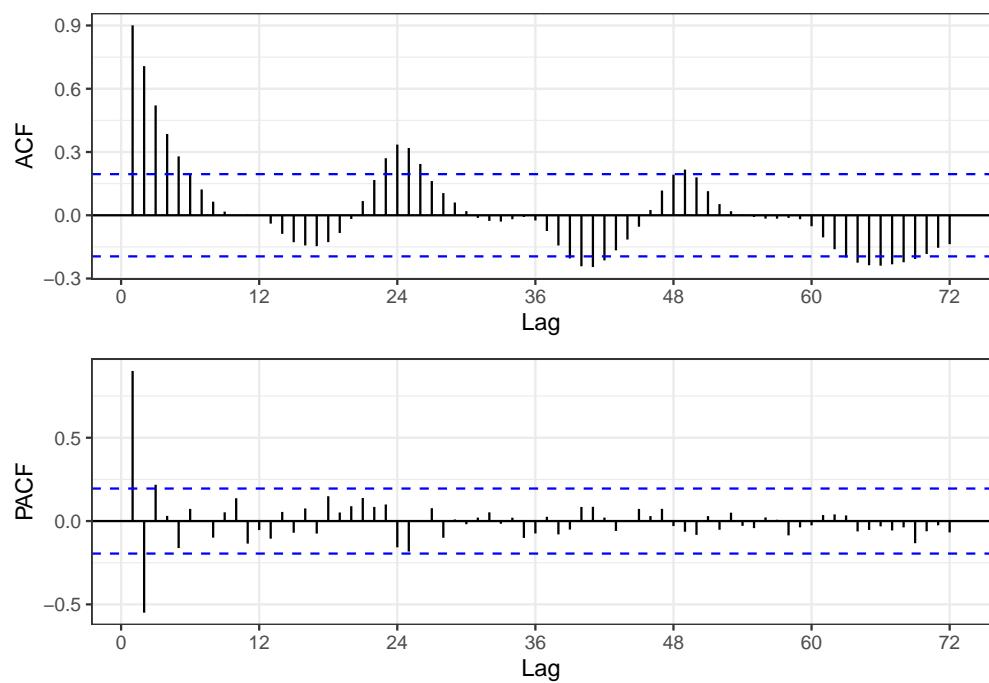


Figura 9: Funciones de autocorrelación y autocorrelación parcial muestrales de la serie de precios de manzana.