

Tarea 3 –Modelos de Componentes Estructurales
versus Modelos ARIMA-SARIMA

Técnicas de pronóstico

Universidad Nacional De Colombia

Andrés Hernando Cerquera Mejía

C.C 1075235138

Tomás Simón Gómez Méndez

C.C 1214720739

Medellín -Antioquia

2022

Contenido

Resumen	3
Introducción.....	3
Desarrollo.....	5
Pruebas de raíz unitaria ordinaria y estacional	7
Estimación de modelo SARIMA.....	10
Modelo exponencial con indicadoras	12
Modelo BSM	12
Conclusiones	14
Bibliografía.....	14

Resumen

En el presente trabajo se estudiaron las raíces unitarias de una serie de tiempo del precio de la carne de hamburguesa promedio mensual entre el 2001 y 2008 para ajustar un modelo SARIMA, que se comparara con las predicciones de los modelos con tendencia lineal exponencial + indicadoras y modelo estructural básico (BSM) + AR (1). Se determinó que la serie presentaba raíces unitarias estacionales, por lo que tiene un comportamiento estocástico en el componente estacional. Por otro lado, se encontró que los residuos de dicho modelo se comportaban como ruido blanco, acorde a la definición de los modelos SARIMA. Se identificó un modelo SARIMA (0,0,0) (2,1,1) para representar la serie de tiempo el cual presentaba menor error de pronósticos comparado con los modelos previamente estudiados.

Introducción

Las series de tiempo permiten desarrollar modelos estadísticos que ayuden a reproducir y/o pronosticar el comportamiento de una variable aleatoria medida en puntos sucesivos del tiempo.

En este trabajo se estudia una serie temporal que mide el precio promedio mensual de venta de la carne de hamburguesa por libra en Estados Unidos, entre el año 2001 al 2008. La información de la serie fue extraída de la base de datos de precios del escáner de venta al por menor de carne del Centro de Información del Mercado del Ganado. Esta serie ha sido estudiada en el libro *“R in a Nutshell: A desktop quick reference”* donde se analiza su autocorrelación con la serie del precio de la carne de pavo por libra en el mismo periodo y país, ya que la fuente de información de ambas series es la misma (Adler, 2010; Rdrr.io, 2021).

Para analizar esta serie se implementa la estrategia de validación cruzada, dejando un total de 77 datos para el ajuste y los 12 datos finales de la serie para la verificación del pronóstico. Se analiza la capacidad de reproducir el pasado (ajustar) y pronosticar dos modelos, un modelo exponencial-lineal con variables indicadoras para la tendencia (modelo paramétrico) y un modelo espacio de estados BSM (modelo no paramétrico). Finalmente se compara el desempeño de ambos modelos tanto para ajuste como para pronóstico utilizando gráficos y criterios de error.

A continuación se realiza una breve descripción de los modelos estudiados, basados en los contenidos de las notas de clase (Giraldo Gomez, 2006).

El modelo de tendencia exponencial lineal con estacionalidad descrita con variables indicadoras se define en la Ecuación 1:

$$Y_t = e^{(\beta_0 + \beta_1 t) + \sum_{j=2}^{13} \beta_j I_j(t)} + \varepsilon_t$$

Ecuación 1

Donde, los β 's son parametros a estimar, y β_0 y β_1 se asocian a la tendencia, mientras β_2 hasta β_{13} se asocian a la estacionalidad, considerando una serie estacional con período $s=12$ y por tanto 11 variables indicadoras. En este tipo de modelos la componente de tendencia y estacionalidad se estiman de manera conjunta, es decir, la componente estructural de la serie de tiempo sería la combinación lineal de tendencia y estacionalidad conjunta.

Cabe indicar que las variables indicadoras solo se estiman para el periodo de la serie menos una posición, pues esto evita colinealidad de las indicadoras, y una correlación con el intercepto β_0 y el coeficiente lineal β_1 .

Por otro lado, el modelo estructural básico (BSM, por sus siglas en inglés), incorpora una hipótesis de variación estocástica, acá se consideran los parámetros β , μ , α δ y γ .

α δ y γ controlan la varianza del error aleatorio ε_t a partir de un sistema de ecuaciones recursivo que afectan el modelo regido por Y_t definidas de acuerdo con la ecuación 2.

$$\begin{aligned} Y_t &= \mu_t + S_t + \varepsilon_t \\ \mu_t &= \mu_{t-1} + \beta_{t-1} + \alpha_t \\ \beta_t &= \beta_{t-1} + \delta_t \\ S_t &= S_{t-1} \dots + S_{t-s+1} \gamma_t \end{aligned}$$

Ecuación 2

Donde la varianza del coeficiente del error α representa el nivel inicial de la serie, la varianza del error δ representa la pendiente de la tendencia y la varianza del error γ representa la estacionalidad. En este tipo de modelos la magnitud de dichas varianzas se relaciona con su efecto en el modelo, a mayor magnitud, mayor efecto tiene la componente asociada al error en el modelo.

Los residuos estructurales $\hat{\varepsilon}_t$, que son el resultado de restarle a la serie de tiempo Y_t la componente de tendencia y estacionalidad, como se muestra en la ecuación 3.

$$\hat{\varepsilon}_t = Y_t - \hat{T}_t - \hat{S}_t$$

Ecuación 3

Para calcular los pronósticos se debe analizar la serie de los residuos estructurales y detectar si presentan autocorrelación o por el contrario son solo ruido blanco, de ser ruido blanco, los pronósticos del modelo estructural original no presentarán alteraciones dado que se estima el pronóstico de ruido blanco como cero, de lo contrario, se debe estimar un modelo para los residuos que logre capturar la correlación existente, esto mediante modelos autoregresivos de media móvil, ARMA (p,q), o autoregresivos de media móvil estacional SARMA(p,q)(p_s,q_s).

Finalmente, en este trabajo se hará un análisis de la serie de tiempo mediante modelos Box-Jenkins o ARIMA-SARIMA, estos modelos se presentan como una alternativa al modelo de componentes. Los modelos SARIMA-ARIMA son modelos ARMA pero que están integrados en un orden 1 o superior ya sea en la componente de tendencia o en la estacional, estas componentes estacionales se pueden eliminar mediante diferenciación. Dicha integración se evidencia a través de las raíces unitarias, ya sean ordinarias o estacionales, cuando se identifica un modelo con raíces unitarias ordinarias se generaría un modelo ARIMA (p,s,q), cuando se encuentran raíces unitarias estacionales se generaría un modelo SARIMA (p,s,q) (p,s,q) y si existen ambas raíces unitarias se generaría un modelo SARIMA (p,s,q) (p,s,q), donde s es un número entero superior a 1.

La presencia de raíces unitarias se hace posible a través de diferentes pruebas de hipótesis, Dickey-Fuller aumentada, Phillips-Perron, Canova-Hansen, HEGY y OCSB (Giraldo Gomez, 2006).

Desarrollo

En la tabla 1 se presenta la información del precio de la hamburguesa desde enero del año 2001 hasta el año 2008, para el año 2008 faltan los meses de mayo a diciembre, los datos aportados tienen una frecuencia mes a mes cada años, es decir una frecuencia de 12 meses. Por otro lado, los meses de marzo y abril presentan mucha fluctuación del precio a lo largo de los años, en comparacion con el mes de noviembre el cual es el mes con menor varianza.

Tabla 1. Fechas y precio de hamburguesas.

Mes/Año	2001	2002	2003	2004	2005	2006	2007	2008
Enero	2.04	2.08	2.38	2.37	2.03	2.04	2.04	2.05
Febrero	2.33	2.24	2.14	2.21	2.18	2.18	2.04	1.85
Marzo	2.19	1.59	2.24	2.12	1.63	2.16	2.07	1.23
Abril	1.63	2.26	1.69	1.52	2.24	1.48	1.57	2.21
Mayo	2.22	2.18	2.14	2.02	2.21	2.21	2.36	
Junio	2.22	1.93	2.08	2.16	2.32	2.33	2.4	
Julio	2.34	1.99	2.32	2.41	2.4	2.66	2.48	
Agosto	2.08	2.12	2.44	2.46	2.28	2.43	2.74	

Septiembre	2.26	2.12	2.41	2.6	2.27	2.46	2.5
Octubre	2.5	2	2.48	2.4	2.58	2.62	2.44
Noviembre	2.26	2.12	2.27	2.07	2.12	2.09	2.06
Diciembre	1.71	1.75	1.75	1.74	1.58	1.61	1.63

Para el ajuste y selección de los modelos de pronóstico se dividió la serie en dos conjuntos de datos, con el objetivo de utilizar un conjunto para el entrenamiento (ajuste) y otro para verificar el pronóstico, tomando de los 89 datos disponibles 77 datos para el entrenamiento, y los últimos 12 para la prueba del pronóstico.

En la figura 1 se observa la descomposición de la serie de tiempo usando el método Loes, en esta se pueden observar componentes de tendencia (*trend*), estacional (*seasonal*), residuales (*remainder*) y datos observados (*observed*), donde se identifica la periodicidad en la componente estacional, la cual parece tener un período $s=12$. Se observa un componente de tendencia lo cual implica una posible raíz unitaria ordinaria y existe una periodicidad lo cual da indicios de la posible presencia de raíces unitarias estacionales.

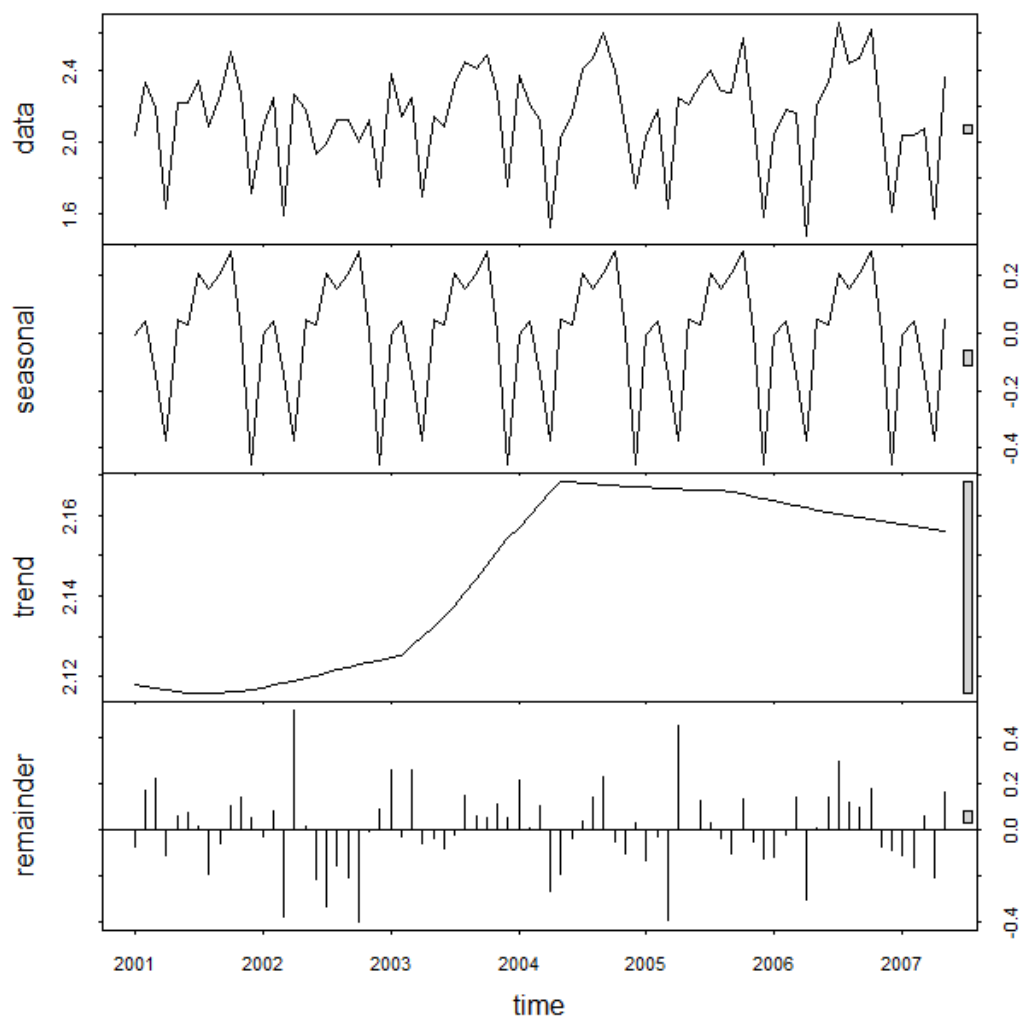


Figura 1. Serie de tiempo observada y descomposición de la serie de tiempo por tendencia, estacionalidad y error.

Pruebas de raíz unitaria ordinaria y estacional

La serie de tiempo previamente descrita se le aplicaron las pruebas Dickey-Fuller Aumentada y Phillips-Perron para determinar la presencia de raíces unitarias ordinarias y las pruebas Canova-Hansen y Hegy, para verificar la existencia de raíces unitarias estacionales, finalmente se aplicó la prueba OCSB para determinar la presencia conjunta de raíces unitarias estacionales y ordinarias

Para la prueba Dickey-Fuller aumentada (Elder & Kennedy, 2001) y Phillips-Perron (Phillips & Perron, 1988) se escogió la opción 3, es decir con media y tendencia, en estas pruebas la hipótesis nula es la existencia de raíces unitarias, esta se rechaza cada vez que el valor p sea menor a 0.05; la tabla 2, muestra los resultados de la prueba ADF y Phillips-Perron para una serie de tiempo con tendencia y media. Con base en los resultados de las pruebas no existen raíces unitarias ordinarias, por lo que la tendencia de la serie de tiempo no es de naturaleza estocástica sino determinística y la serie es estacionaria en covarianza.

Tabla 2. Resultados de la prueba ADF y Phillips-Perron

Lag	Valor - P	Conclusión
Prueba Dickey-Fuller aumentada		
0	0.010	Se rechaza H_0
1	0.010	Se rechaza H_0
2	0.010	Se rechaza H_0
3	0.015	Se rechaza H_0
Prueba Phillips-Perron		
5	0.01	Se rechaza H_0

Para la aplicación de la prueba Canova Hansen se requiere la eliminación del componente de tendencia de la serie, esto se logra a partir de la substracción del componente estacional de la serie, previamente descompuesto a través del método Loes. La figura 2, muestra la serie de tiempo sin la tendencia, gráficamente se puede ver que el componente estacional de la serie presenta una variabilidad considerable conforme pasa el tiempo, por lo que se espera que las pruebas muestren la presencia de raíces unitarias estacionales.

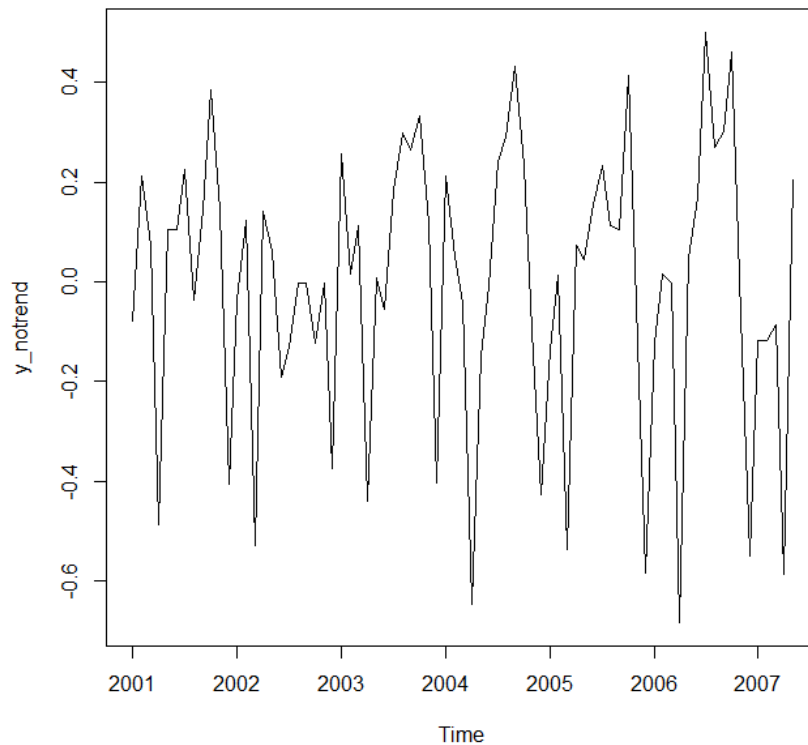


Figura 2. Componente estacional + residuos de la serie de tiempo estudiada.

La prueba de hipótesis de Canova Hansen tiene por H_0 que el componente estacional de la serie de tiempo es determinístico, y H_a que es estocástica o que existe presencia de raíces unitarias estacionales. H_0 se rechaza si el valor p es mayor a 0.05 (Canova & Hansen, 1995).

La tabla 3 muestra los resultados de la prueba Canova-Hansen, para las raíces en el semiplano positivo del círculo unitario, se puede visualizar, que el valor p para todas las componentes es mayor a 0.05, por lo tanto, no se identifica un comportamiento estocástico en el componente estacional de la serie.

Tabla 3. Resultados de la prueba Canova-Hansen y

Frecuencia	Valor – P	Conclusión
$\pi/6$	0.104	Se rechaza H_0
$2\pi/6$	0.685	Se rechaza H_0
$3\pi/6$	0.553	Se rechaza H_0
$4\pi/6$	0.504	Se rechaza H_0
$5\pi/6$	0.422	Se rechaza H_0
π	0.736	Se rechaza H_0

Conjunta	0.418	Se rechaza H_0
----------	-------	------------------

La prueba HEGY, Permite decidir cuáles de las frecuencias de una serie corresponden a raíces estacionales significativas. H_0 en esta prueba corresponde a que la serie está integrada estacionalmente, es decir aceptar que la frecuencia corresponde a una raíz unitaria estacional. Con un valor P menor a 0.05 se rechaza H_0 (Hylleberg et al., 1990). La tabla 5 muestra los resultados de la prueba Hegy, se puede ver que existe una discrepancia entre los resultados de la prueba Canova-Hansen y Hegy, sin embargo, se tomarán los resultados de la prueba Hegy, ya que la presencia de raíces unitarias estacionales es coherente con el comportamiento de la serie sin tendencia.

Tabla 4. Resultados de la prueba Hegy

Frecuencia	Valor P	Conclusión
0	0.112	No se rechaza H_0
$\pi/2$	0.0739	No se rechaza H_0
$3\pi/2$	0.7469	No se rechaza H_0
$5\pi/6$	0.1516	No se rechaza H_0
$7\pi/8$	0.2913	No se rechaza H_0
$9\pi/10$	0.0982	No se rechaza H_0
$11\pi/12$	0.0191	Se rechaza H_0

Finalmente, se aplicó la prueba OCSB, la cual tiene como H_0 la presencia de raíces unitarias estacionales y ordinarias, esta se rechaza cuando el estadístico de prueba es menor que el valor crítico ($t_{\hat{\beta}_2} < t_\alpha$). H_a es la presencia de únicamente raíces unitarias estacionales (Osborn et al, 1988).

La tabla 6 muestra los resultados de aplicar la prueba OCSB a la serie de tiempo, como se puede observar, al rechazarse la hipótesis nula, la serie de tiempo solo presenta raíces unitarias estacionales, siendo coherente con los resultados de la prueba ADT y Phillips-Perron.

Tabla 4. Resultados de la prueba Hegy

$t_{\hat{\beta}_2}$	t_α	Conclusión
-3.201	-1.803	Se rechaza H_0

Se compilaron las conclusiones de las pruebas aplicadas a la serie de tiempo en la tabla #, esta muestra que en conclusión la serie no presenta raíces unitarias ordinarias, sin embargo, si se detectan raíces unitarias estacionales, mostrando coherencia entre casi todas las pruebas excepto la Canova-Hansen.

Tabla 2. Resultados de las pruebas aplicadas

Prueba	Conclusión
ADT	No hay raíces unitarias ordinarias
Phillips-Perron	
Canova-Hansen	No hay raíces unitarias estacionales
Hegy	Hay raíces unitarias estacionales
OCSB	Hay raíces unitarias estacionales, más no ordinarias

Estimación de modelo SARIMA

Una vez identificadas las raíces unitarias de la serie, se procedió a estimar un modelo SARIMA. Utilizando la función `auto.arima()` se encontró un modelo SARIMA (0,0,0) (2,1,1), donde se puede corroborar las raíces unitarias estacionales detectadas en las pruebas anteriores, y al calcularse el modelo donde se evidencia el termino D con orden mayor a 0.

Ya estimado el modelo SARIMA se procedió a verificar que los residuos de dicho modelo fueran ruido blanco, La autocorrelación del modelo se rechazó a través de la prueba Ljung-Box, la cual obtuvo un valor P de 0.442 aceptando la hipótesis nula de que los residuales se distribuyen como ruido blanco (H_0 : se acepta que los residuales se aproximan a ruido blanco con media cero y varianza constante, H_1 : No se cumple H_0).

Adicionalmente, la función de autocorrelación de los residuos del modelo SARIMA y las bandas de Bartlett se calcularon, la figura # muestra la función de autocorrelación y las bandas de Bartlett, donde se corrobora el resultado obtenido por la prueba Ljung-Box

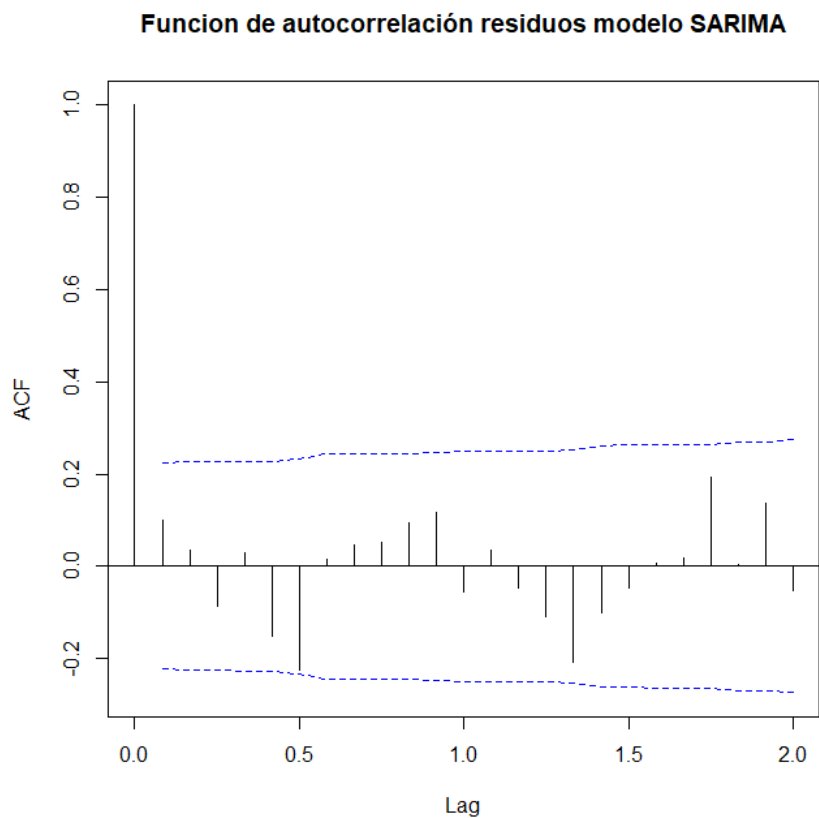
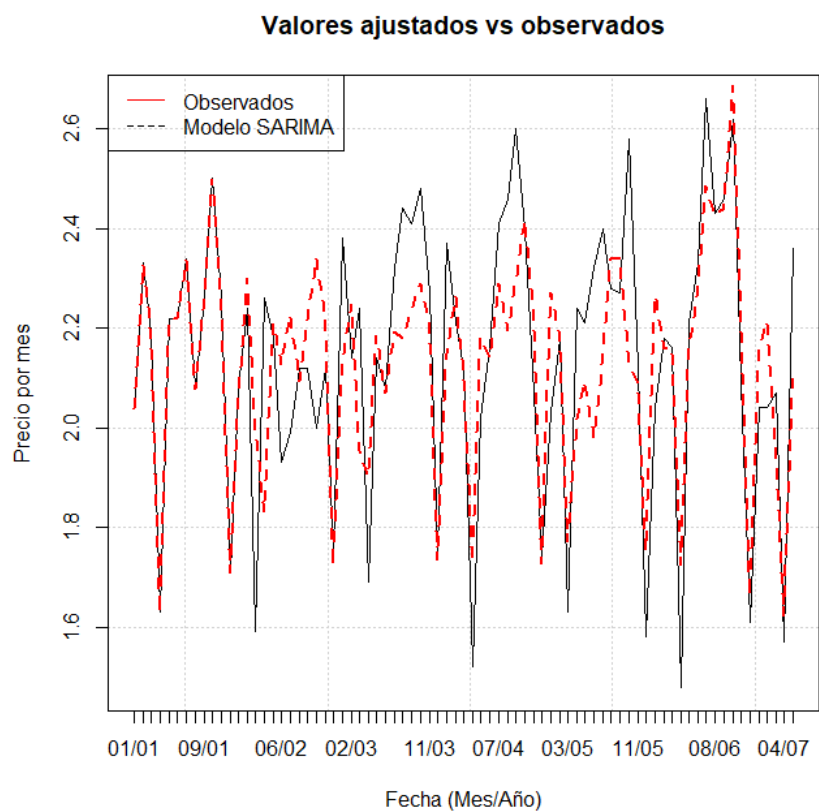


Figura #: Función de Autocorrelación para los residuos del modelo SARIMA



Modelo exponencial con indicadoras

En la entrega anterior se hizo el ajuste del modelo exponencial con indicadoras, descrito en la ecuación 1, cuyos parámetros se resumen en la tabla 2. El análisis de los residuos estructurales de este modelo se presenta en esta sección.

Tabla 2: Parámetros del modelo exponencial lineal con indicadoras

Tipo de parámetro	Parámetro	Valor ajustado	Valor t	Valor P	Significancia
Tendencia	β_0	0.50	10.13	6.30E-15	Significativo
	β_1	0.00	1.08	0.28218	No Significativo
Estacionalidad	β_2	0.23	4.20	8.30E-05	Significativo
	β_3	0.26	4.62	1.88E-05	Significativo
	β_4	0.17	2.92	0.00472	Significativo
	β_5	0.04	0.76	0.44563	No Significativo
	β_6	0.26	4.63	1.80E-05	Significativo
	β_7	0.25	4.39	4.24E-05	Significativo
	β_8	0.33	5.93	1.33E-07	Significativo
	β_9	0.31	5.48	7.51E-07	Significativo
	β_{10}	0.33	5.91	1.44E-07	Significativo
	β_{11}	0.36	6.54	1.17E-08	Significativo
	β_{12}	0.24	4.18	8.81E-05	Significativo

La incorrelación aparente del modelo se aceptó a través de la prueba Ljung-Box, la cual obtuvo un valor P de 0.37 aceptando la hipótesis nula de que los residuales se distribuyen como ruido blanco. (Ho: se acepta que los residuales se aproximan a ruido blanco con media cero y varianza constante, H1: No se cumple Ho)

Con base en los resultados de incorrelación para los residuos del modelo descrito, no fue necesario el uso de un modelo ARMA-SARMA para modelar los residuos, dado que se comportan como ruido blanco.

Modelo BSM

Anteriormente se había extraído las varianzas de los errores del modelo básico estructural para determinar cuáles eran los componentes más significativas. La tabla 3 muestra las varianzas de los parámetros de acuerdo con la Ecuación 2

Tabla 3. Varianzas de los parámetros del modelo estructural básico (BSM)

Nivel	Pendiente	Estacionalidad	Error
0.000051	0.000	0.014	0.029

La autocorrelación del modelo se aceptó a través de la prueba Ljung-Box, la cual obtuvo un valor P de 0.0043 rechazando la hipótesis nula de que los residuales se distribuyen como ruido blanco (H_0 : se acepta que los residuales se aproximan a ruido blanco con media cero y varianza constante, H_1 : No se cumple H_0). Por lo tanto, los residuales se representaron a través de un modelo AR (1,0,0), mostrando este el mejor ajuste en comparación con otros modelos SARMA generados por las funciones auto.arima y armasubset.

Se realizó una predicción usando los modelos anteriormente presentados y los datos de validación cruzada, esto con el fin de verificar las métricas de error en la predicción para determinar cuál modelo tiene mayor precisión de pronóstico. La figura # muestra la predicción de los modelos y la serie de tiempo, se puede observar una discrepancia considerable de los modelos con la serie de tiempo desde enero del 2008 hasta el final de la serie.

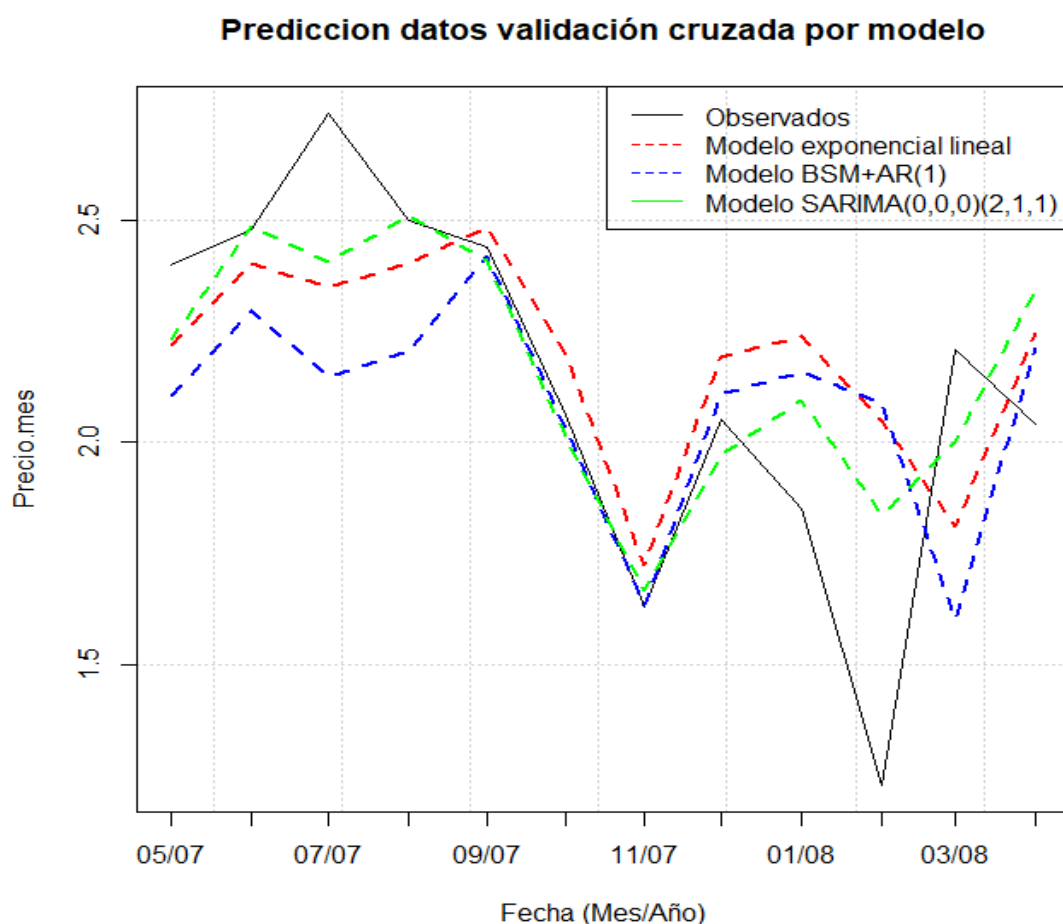


Figura 10. Pronósticos por modelo y serie de tiempo observada mayo 2007 a abril 2008

La tabla 5 muestra los criterios de pronóstico para los modelos. Se puede observar resaltado en verde los resultados con menor error, o mayor precisión de pronóstico, con base en estos resultados, se puede afirmar que el modelo SARIMA estimado tiene mayor precisión y cumple con los supuestos de que sus residuos son ruido blanco, por ende, este es el más apropiado para modelar serie.

Tabla 5. Criterios de pronóstico para modelos estudiados

Modelo	RMSE	MAPE	U de Theil
Exponencial lineal con indicadoras	0.429	20.054	0.754
Modelo BSM	0.412	20.039	0.714
SARIMA (0,0,0) (2,1,1)	0.363	16.38	0.690

Conclusiones

Para la serie de tiempo del precio promedio de venta de carne de hamburguesa por libra en Estados Unidos desde enero del 2001 hasta abril del 2008, se analizaron las raíces unitarias ordinarias y estacionales que podría presentar la serie. Se detectó la presencia de raíces unitarias estacionales, lo cual implica que la serie presenta un comportamiento estocástico en el componente estacional. Si bien la prueba Hegy y Canova-Hansen mostraron resultados opuestos, se adoptó el resultado de la prueba Hegy, al ser coherente con el comportamiento del componente estacional de la serie en el tiempo.

Se estimó un modelo SARIMA (0,0,0) (2,1,1), donde se corrobora que es integrado estacionalmente, adicionalmente se encontró que los residuos de este modelo se comportaban como ruido blanco.

Las métricas de precisión del modelo SARIMA estimado muestran una mejora significativa comparado al modelo BSM+AR (1) y al exponencial lineal con indicadoras.

Bibliografía

- Adler, J. (2010). *R in a Nutshell: A Desktop Quick Reference* (M. Mukherji (ed.); 1st ed.). O'Reilly Media, Inc.
- Canova, F., & Hansen, B. E. (1995). Are Seasonal Patterns Constant over Time? A Test for Seasonal Stability. *Journal of Business & Economic Statistics*, 13(3), 237–252. <https://doi.org/10.2307/1392184>

- Elder, J., & Kennedy, P. E. (2001). Testing for unit roots: what should student Taught? *Journal of Economic Education*, 32.
- Giraldo Gomez, N. (2006). *Técnicas de Pronósticos Aplicaciones con R*. Universidad Nacional de Colombia Sede Medellín Facultad de Ciencias Escuela de Estadística.
- Hylleberg, S., Engle, R., Granger, C., & Yoo, B. S. (1990). Seasonal integration and cointegration. *Journal of Econometrics*, 44(1–2), 215–238.
- Osborn et al, D. R. (1988). Seasonality and the Order of Integration for Consumption. *Oxford Bulletin of Economics and Statistics*, 50(4), 361–377.
- Phillips, P. C. B., & Perron, P. (1988). Testing for a Unit Root in Time Series Regression Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–381.
- Rdrr.io. (2021, November 19). *ham.price.ts: Ham Price Time Series*.