

## Tarea 2 –Modelos ARMA-SARMA para residuos estructurales

Técnicas de pronóstico

Universidad Nacional De Colombia

Andrés Hernando Cerquera Mejía

C.C 1075235138

Tomás Simón Gómez Méndez

C.C 1214720739

Medellin -Antioquia

2022

## Resumen

En el presente trabajo se estudiaron los residuos estructurales de una serie de tiempo del precio de la carne de hamburguesa promedio mensual entre el 2001 y 2008 para un modelo con tendencia lineal exponencial + indicadoras y un modelo estructural básico (BSM). se determinó que los residuos del modelo lineal exponencial + indicadoras no se comportaban como ruido blanco, mientras que en el modelo BSM sí. Se identificó un modelo ARMA-SARMA para representar el error del modelo BSM. Se encontró que el modelo con menor AIC fue el modelo ARMA (1,0,0) [12]. Adicionalmente, al comprar el pronóstico de los modelos, se encontró que el modelo BSM + AR (1), presentaba menor error de pronósticos.

## 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 [1], [2].

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 de 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 [3].

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  $\beta_s$  son parametros a estimar, y  $\beta_0$  y  $\beta_1$  se asocian a la tendencia, mientras  $\beta_2$  hasta  $\beta_{13}$  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  $\beta_0$  y el coeficiente lineal  $\beta_1$ .

Por otro lado, el modelo estructural básico (BSM, por sus siglas en inglés), incorpora una hipótesis de variación estocástica, acá se consideran los parámetros  $\beta$ ,  $\mu$ ,  $\alpha$   $\delta$  y  $\gamma$ .

$\alpha$   $\delta$  y  $\gamma$  controlan la varianza del error aleatorio  $\varepsilon_t$  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  $\alpha$  representa el nivel inicial de la serie, la varianza del error  $\delta$  representa la pendiente de la tendencia y la varianza del error  $\gamma$  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.

En este trabajo se analizan 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.

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

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, estos comportamientos se observan a través de funciones de autocorrelación

(FAC) y se prueba mediante las pruebas de incorrelación Durbin-Watson o Ljung-Box.

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<sub>s</sub>,q<sub>s</sub>).

## 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 en sus componentes de tendencia (*trend*), estacional (*seasonal*), residuales (*random*) y

datos observados (*observed*), donde se identifica la periodicidad en la componente estacional, la cual parece tener un período  $s=12$

### Decomposition of additive time series

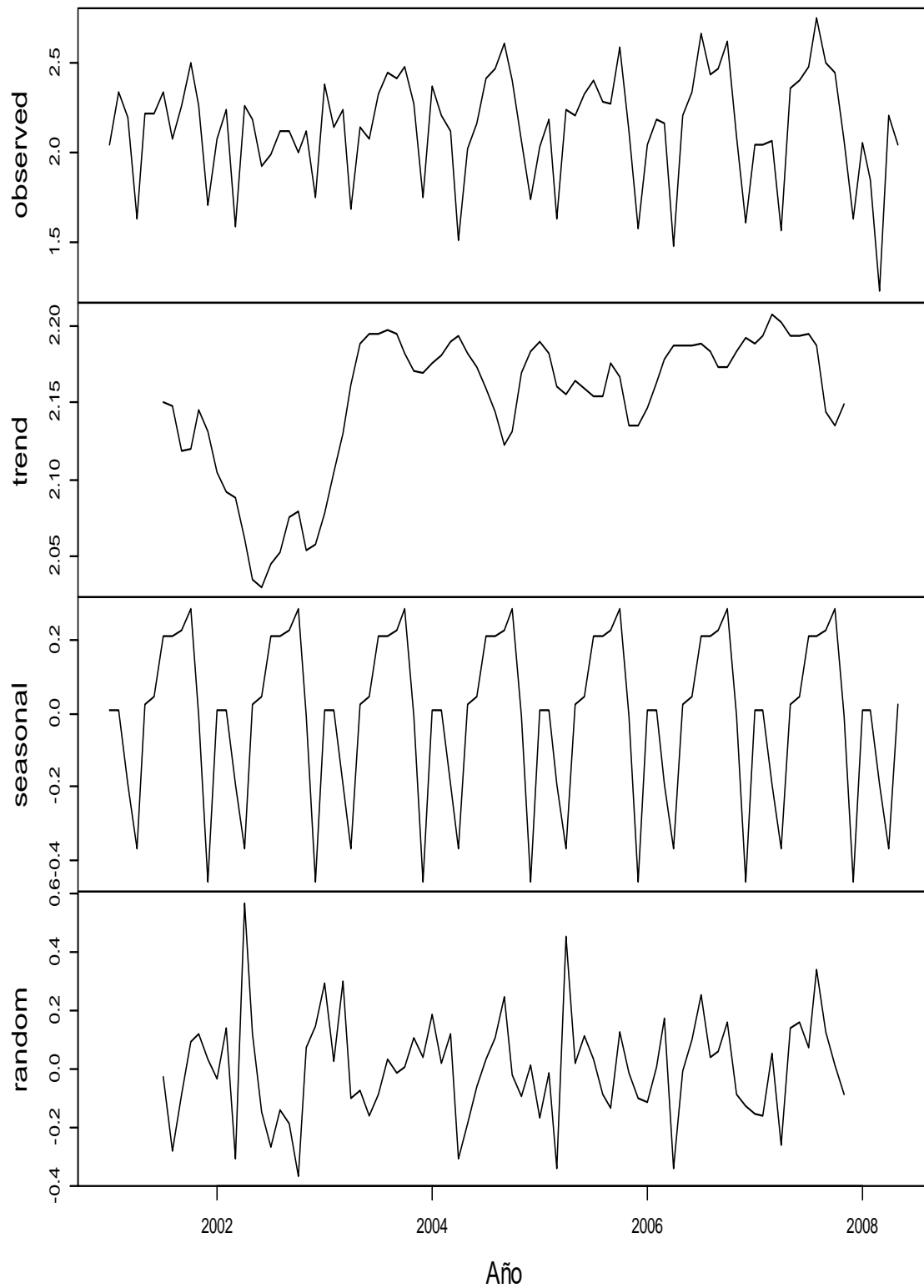


Figura 1. Serie de tiempo observada y descomposición de la serie de tiempo por tendencia, estacionalidad y error.

### 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	$\beta_0$	0.50	10.13	6.30E-15	Significativo
	$\beta_1$	0.00	1.08	0.28218	No Significativo
Estacionalidad	$\beta_2$	0.23	4.20	8.30E-05	Significativo
	$\beta_3$	0.26	4.62	1.88E-05	Significativo
	$\beta_4$	0.17	2.92	0.00472	Significativo
	$\beta_5$	0.04	0.76	0.44563	No Significativo
	$\beta_6$	0.26	4.63	1.80E-05	Significativo
	$\beta_7$	0.25	4.39	4.24E-05	Significativo
	$\beta_8$	0.33	5.93	1.33E-07	Significativo
	$\beta_9$	0.31	5.48	7.51E-07	Significativo
	$\beta_{10}$	0.33	5.91	1.44E-07	Significativo
	$\beta_{11}$	0.36	6.54	1.17E-08	Significativo
	$\beta_{12}$	0.24	4.18	8.81E-05	Significativo

La figura 2 muestra la serie de los residuales para el modelo exponencial lineal con indicadoras. Se observa que es una serie estacionaria en covarianza, ya que no muestra tendencias monótonas, fluctúa alrededor de una media y no muestra patrones de heterocedasticidad.

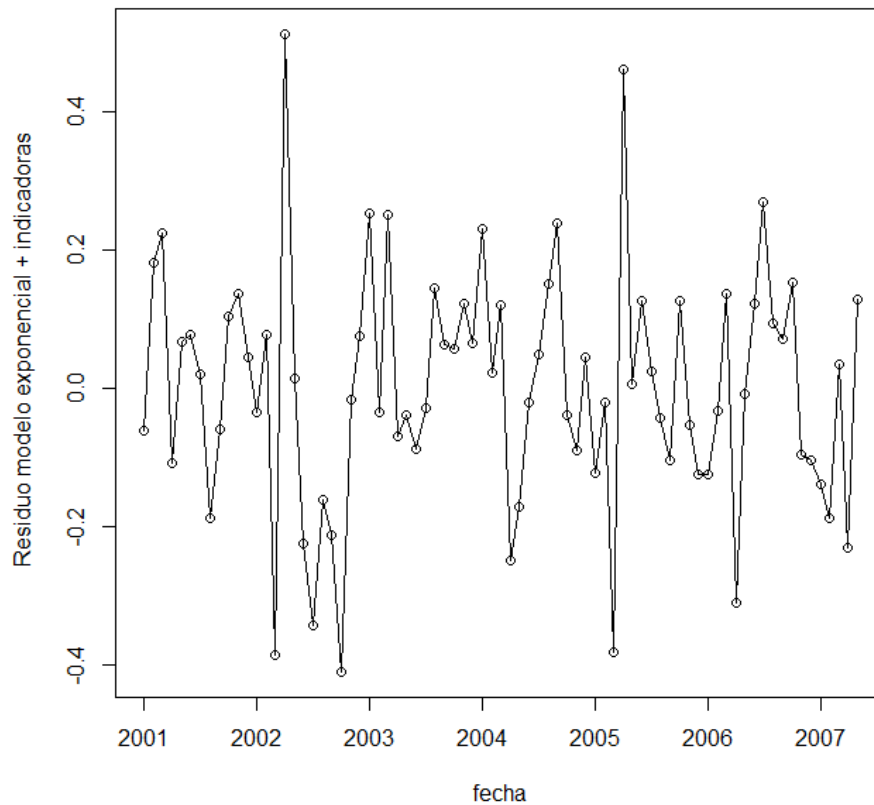


Figura 2. Serie de tiempo de los residuales para el modelo exponencial con indicadoras.

La figura 3 muestra la autocorrelación de los residuales con las bandas de Bartlett para el modelo utilizando 24 rezagos, se selecciona este número de rezagos utilizando el 25% de los datos y aproximando al múltiplo finito más cercano del periodo estacional de la serie.

Se observa que los residuos para este modelo no presentan autocorrelación por encima de las bandas de Bartlett, lo cual, es un indicio de que se distribuyen como ruido blanco. Ahora, 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 con media cero y varianza constante, H1: No se cumple H0). Por otro lado, la prueba Durbin Watson no es apropiada para evaluar la incorrelación de los residuales en este caso, dado que esta se aplica a modelos lineales.

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

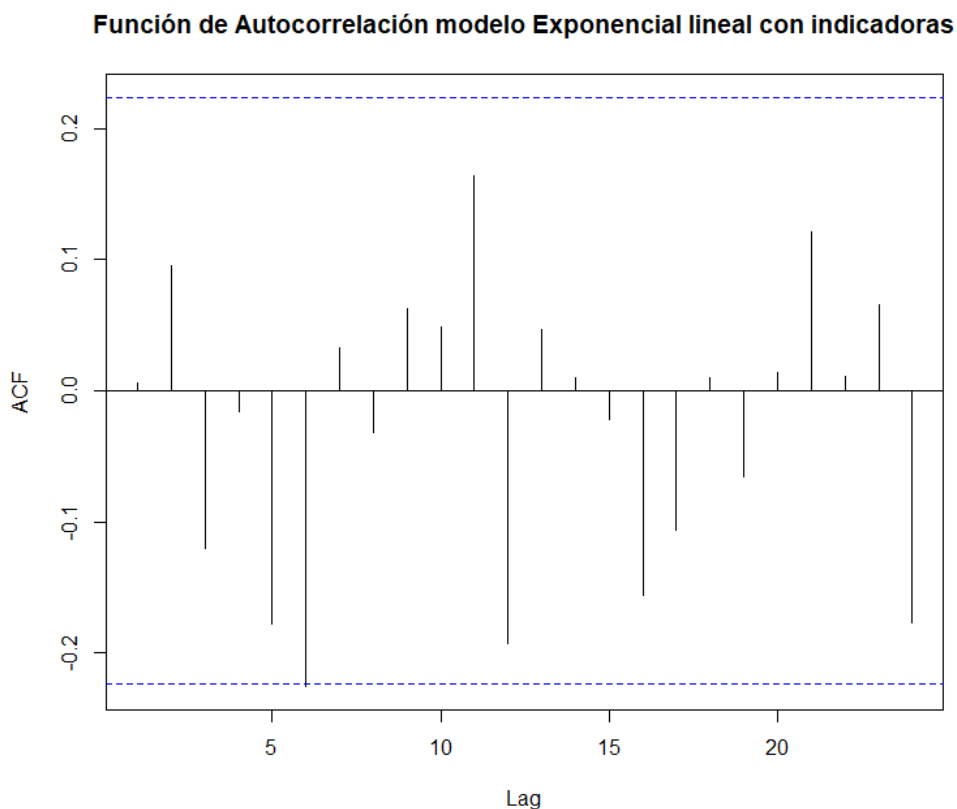


Figura 3. Función de autocorrelación modelo exponencial lineal con indicadoras y bandas de Bartlett.

## Modelo BSM

Anteriormente se había extraído las varianzas de los errores del modelo básico estructural para determinar cuáles eran los componentes que 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 figura 4 muestra la serie de los residuales para el modelo BSM. Se observa que es una serie es estacionaria en covarianza, ya que no muestra tendencias monótonas, fluctúa alrededor de una media y no muestra patrones de heterocedasticidad.



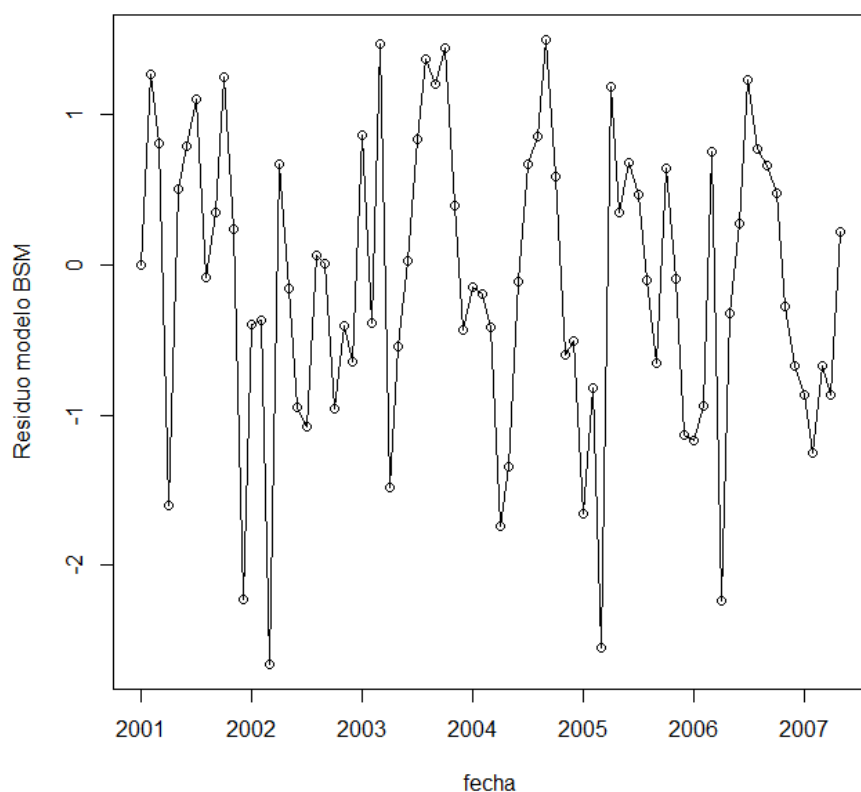


Figura 4. Serie de tiempo de los residuales para el modelo BSM

La figura 5 muestra la autocorrelación de los residuales con las bandas de Bartlett para el modelo utilizando 24 rezagos, se selecciona este número de rezagos utilizando el 25% de los datos y aproximando al múltiplo finito más cercano del periodo estacional de la serie.

Se observa que los residuos para este modelo presentan autocorrelación por encima de las bandas de Bartlett, en los rezagos 6,7,12 (dentro de la muestra graficada), lo cual, muestra que, aunque es una serie estacionaria, no es ruido blanco.

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 pueden ser modelados con ARMA o SARMA para capturar la estructura de autocorrelación.

Por otro lado, la prueba Durbin Watson no es apropiada para evaluar la incorrelación de los residuales en este caso, dado que esta se aplica a modelos lineales.

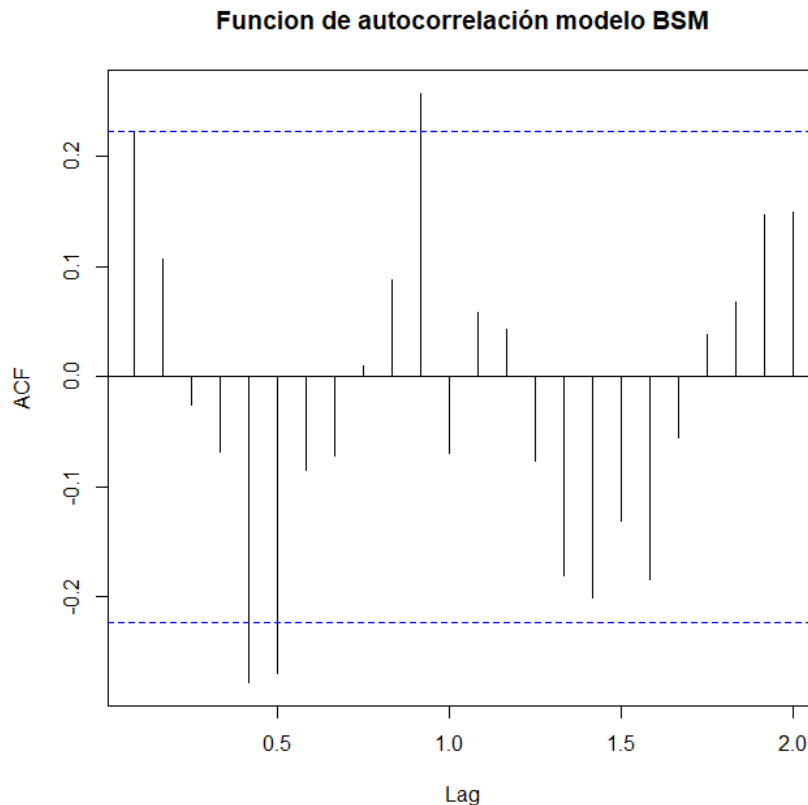


Figura 5. Función de autocorrelación modelo BSM y bandas de Bartlett.

Para identificar el modelo ARMA-SARMA apropiado para modelar el comportamiento de los residuos se utilizaron las funciones `auto.arima()` y `armasubset()`. Con la función `auto.arima()` se obtuvo un modelo AR(1), con las características presentadas en la tabla 4, obtenidas del ajuste con la función ARIMA.

La medida de bondad de ajuste AIC se calculó para el modelo AR (1,0,0) y AR (5,0,0), se muestra en la tabla 4, con base en el resultado presentado, el modelo con mayor ajuste a los residuos es el AR (1,0,0).

Tabla 4. Parámetros para los modelos ARMA/SARMA para los residuos

Modelo	AIC	Parámetro	Valor ajustado	Desviación estándar	Valor P	Conclusión
AR (1,0,0)	218.07	$\Phi_1$	0.22	0.11	0.04	Significativo
		Intercepto del modelo	-0.12	0.14	0.37	No significativo
AR (5,0,0)	219.62	$\Phi_1$	0.190	0.10	0.08	No significativo

		$\Phi_2$	0.059	0.11	0.59	No significativo
		$\Phi_3$	-0.026	0.11	0.81	No significativo
		$\Phi_4$	-0.006	0.11	0.95	No significativo
		$\Phi_5$	-0.270	0.11	0.01	Significativo
		Intercepto del modelo	-0.120	0.10	0.23	No significativo

Usando la función `armasubsets()` se encontraron 2 modelos con el menor BIC un modelo ARMA(5,0,0) y ARMA(6,0,0), como se muestra en la figura 6, se seleccionó el modelo con menor rezagos. Los parámetros del modelo están presentados en la tabla 4.

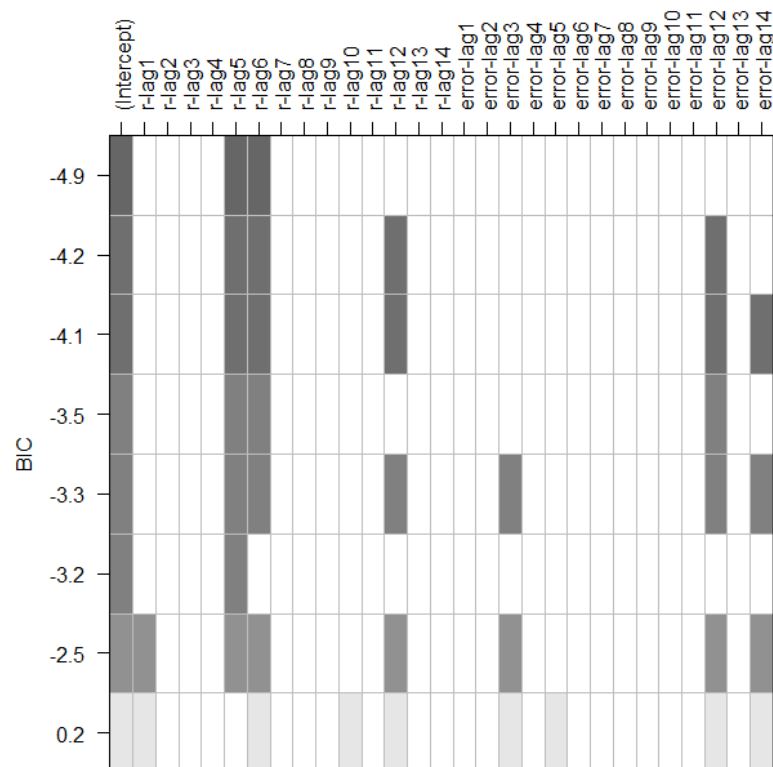


Figura 6. Posibles modelos para estimar el error basados en rezagos y BIC usando la función `armasubsets()`.

La figura 7 muestra la serie de los residuales para el modelo AR (1), donde se observa que es una serie es estacionaria en covarianza

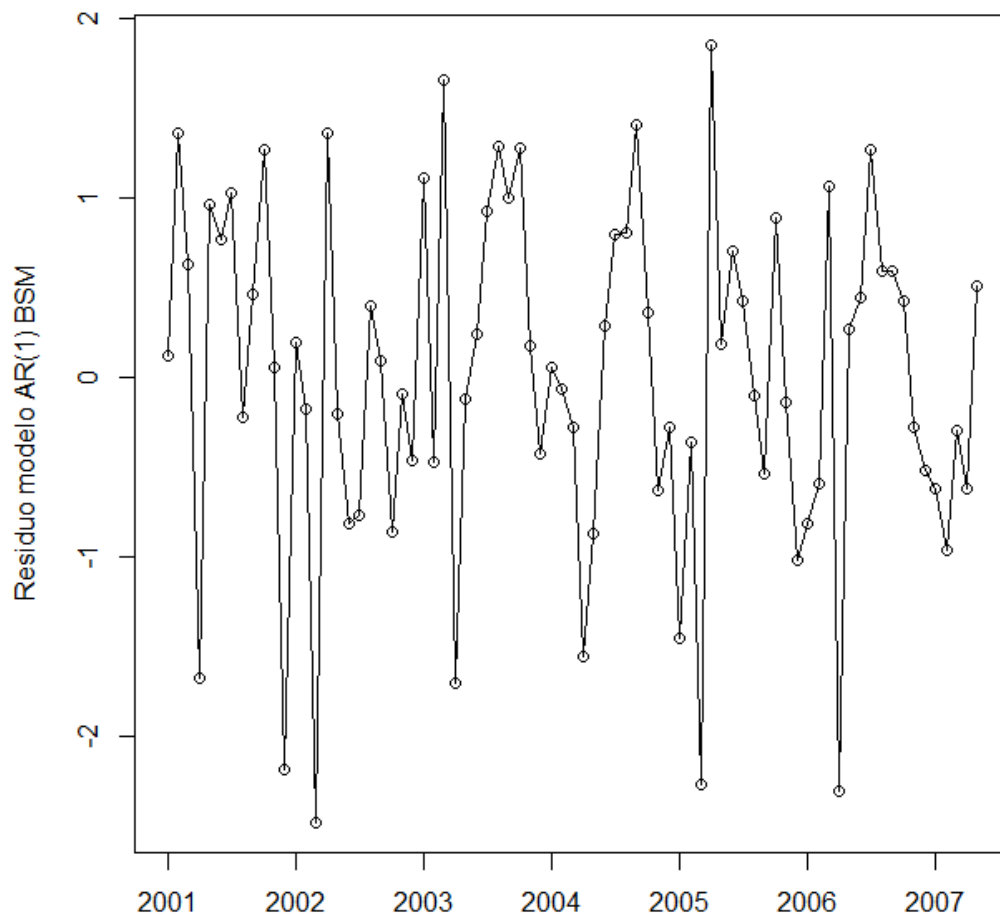


Figura 7. Serie de tiempo de los residuales para el modelo AR (1)

La figura 8 muestra la autocorrelación de los residuales con las bandas de Bartlett para el modelo AR (1) utilizando 24 rezagos, se selecciona este número de rezagos utilizando el 25% de los datos y aproximando al múltiplo finito más cercano del periodo estacional de la serie.

Se observa que los residuos para este modelo presentan autocorrelación por encima de las bandas de Bartlett, en los rezagos 5,11 (dentro de la muestra graficada), lo cual podría evidenciar que los residuos no se comportan como ruido blanco. Sin embargo, la incorrelación aparente del modelo AR (1) se probó a través de la prueba Ljung-Box, la cual obtuvo un valor P de 0.21 aceptando la hipótesis nula de que los residuales se distribuyen como ruido blanco.

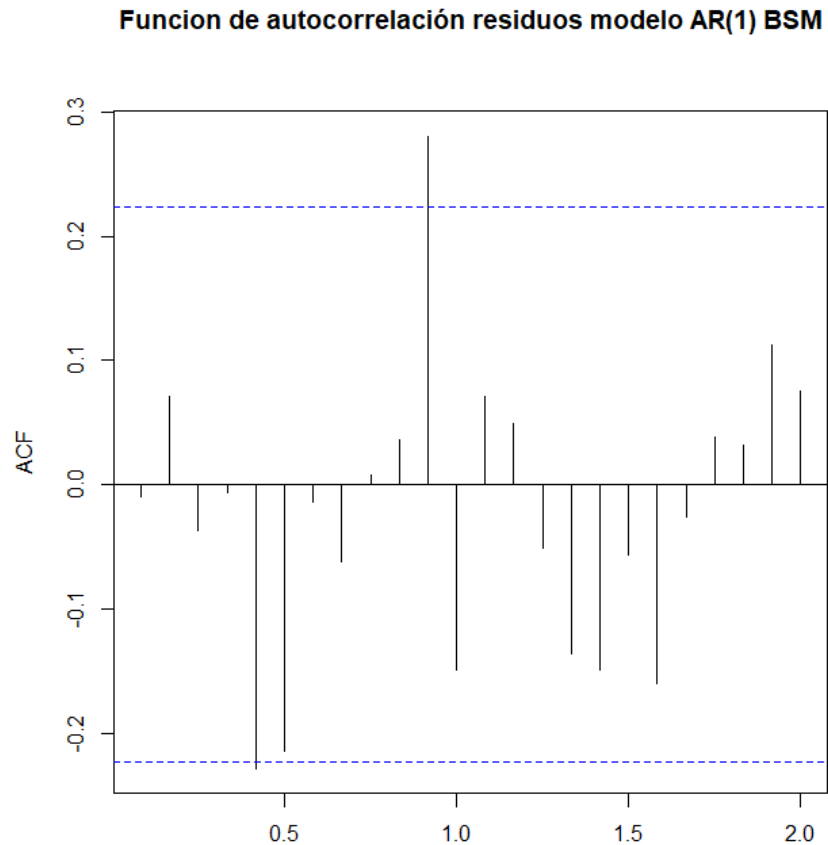


Figura 8. Función de autocorrelación modelo AR (1) y bandas de Bartlett.

Se realizó una predicción usando los modelos 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 9 muestra la predicción de los modelos y la serie de tiempo observada, se puede observar una discrepancia considerable de los modelos con la serie de tiempo desde enero del 2008 hasta el final de la serie.

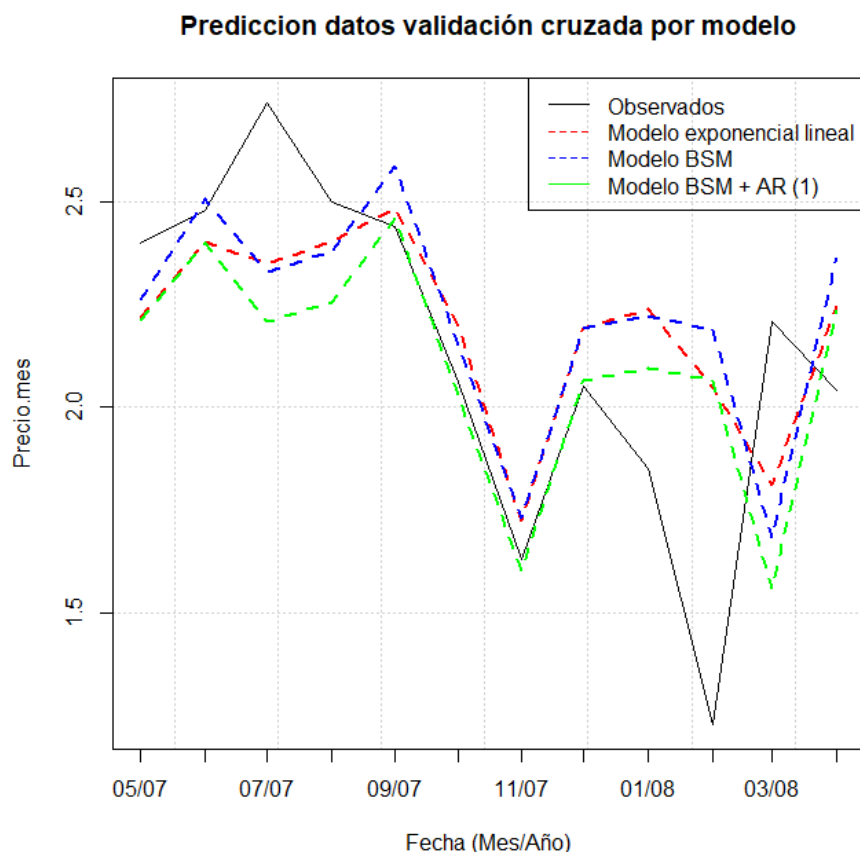


Figura 9. Pronósticos por modelo y serie de tiempo observada mayo 2007 a abril 2008

La tabla 5 muestra los criterios de pronóstico para los modelos. Se puede observar resultado en verde los resultados con menor error, o mayor precisión de pronóstico, con base en estos resultados, se puede afirmar que el modelo BSM + AR (1) tiene mayor precisión y cumple con los supuestos de que los errores del modelo sean ruido blanco, por ende, este es el más apropiado para modelar serie.

Tabla 5. Criterios de pronóstico para modelos estudiados

<b>Modelo</b>	<b>RMSE</b>	<b>MAPE</b>	<b>U de Theil</b>
Exponencial lineal con indicadoras	0.429	20.054	0.754
Modelo BSM	0.427	19.756	0.741
Modelo BSM + AR (1)	0.412	20.039	0.714

## Conclusiones

Para la serie de tiempo del precio promedio de venta de la hamburguesa por libra en Estados Unidos desde enero del 2001 hasta abril del 2008, se analizaron los residuos estructurales de los modelos BSM y exponencial lineal + indicadoras.

Se encontró que para el modelo exponencial lineal + indicadoras, los residuos se comportaban como ruido blanco, por lo tanto, no se requería modelarlos. Por otro lado, los residuos estructurales del modelo BSM, no se comportaban como ruido blanco, al presentar autocorrelación, lo cual implicó que se ajustara un modelo ARMA que capturara el comportamiento de autocorrelación y que sus residuos se comportaran como ruido blanco.

El modelo ajustado con menor AIC fue un modelo ARMA (1,0,0), se comparó el desempeño de ajuste y pronóstico para los modelos mencionados y los resultados de este estudio mostraron que el modelo BSM + AR (1) muestra mayor precisión en los pronósticos.

A pesar de que las métricas de precisión no mejoran significativamente al agregar el modelo ARMA al BSM, la inclusión de este implica que el modelo ahora cumple con el supuesto de que el error se comporta como ruido blanco, por lo cual se debe dar prelación a la selección de dicho modelo.

## Bibliografía

- [1] J. Adler, R in a nutshell: A desktop quick reference. "O'Reilly Media, Inc.," 2010.
- [2] Rdrr.io, "ham.price.ts: Ham Price Time Series," Nov. 19, 2021.
- [3] N. Giraldo Gomez, Técnicas de Pronósticos Aplicaciones con R. Medellin: Universidad Nacional de Colombia Sede Medellín Facultad de Ciencias Escuela de Estadística., 2006