

2021-2

TECNICAS DE PRONOSTICOS



1 TABLA DE CONTENIDO

2	Resumen.....	2
3	Introduccion	2
4	Materiales y métodos	3
5	Descripción de la serie de tiempo	4
6	Modelo exponencial con indicadoras	7
7	Modelo estructural básico (BSM).....	9
8	Análisis de resultados.....	10
9	Conclusiones.....	13
10	Referencias.....	14

Trabajo 1 técnicas de pronóstico, uso de modelos de tendencia exponencial lineal con variables indicadoras para la estacionalidad y modelo estructural básico para el análisis de la serie de tiempo del precio de la hamburguesa.

Tomás S. Gómez, Andrés H. Cerquera

2 RESUMEN

El presente trabajo incluye la descripción y análisis exploratorio de la serie de tiempo para el precio promedio de venta de la hamburguesa por libra en Estados Unidos desde enero del 2001 hasta abril del 2008. Se hizo uso de 2 modelos para el análisis de la serie: un modelo aditivo con tendencia exponencial lineal y variables indicadoras para la estacionalidad, y un modelo estructural básico (BSM). El modelo BSM mostró mejor ajuste según los indicadores de bondad de ajuste; sin embargo, el modelo aditivo de componentes evidenció un mejor resultado en los criterios de pronóstico empleados. Además, se identificó que la serie no tiene una tendencia monótona, y que los parámetros asociados a la tendencia en ambos modelos resultaron no significativos.

3 INTRODUCCION

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.

4 MATERIALES Y MÉTODOS

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 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.

$$Y_t = \mu_t + \beta_t + S_{t-s} + \varepsilon_t$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \alpha \varepsilon_t$$

$$\beta_t = \beta_{t-1} + \delta \varepsilon_t$$

$$S_t = S_{t-s} + \gamma \varepsilon_t$$

Ecuación 2

Donde la varianza 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.

5 DESCRIPCIÓN DE LA SERIE DE TIEMPO

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.

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	

La figura 1 muestra la mediana y cuartiles para cada uno de los meses del año en los diferentes años, a través de un gráfico de cajas y bigotes. En dicho gráfico se observa la variación del precio promedio de la hamburguesa, el eje x indica el número del mes, por lo que 1 corresponde a enero, hasta doce que sería diciembre.

Se deduce del Box-plots que la mediana del precio promedio de la hamburguesa disminuye drásticamente en diciembre y noviembre y aumenta considerablemente desde julio hasta octubre. Además, tan solo hay cuatro meses con valores atípicos que corresponden a los meses de febrero, mayo, julio y octubre.

Por otro lado, los meses de marzo y abril presentan mucha fluctuación del precio a lo largo de los años, en comparación con el mes de noviembre el cual es el mes con menor varianza.

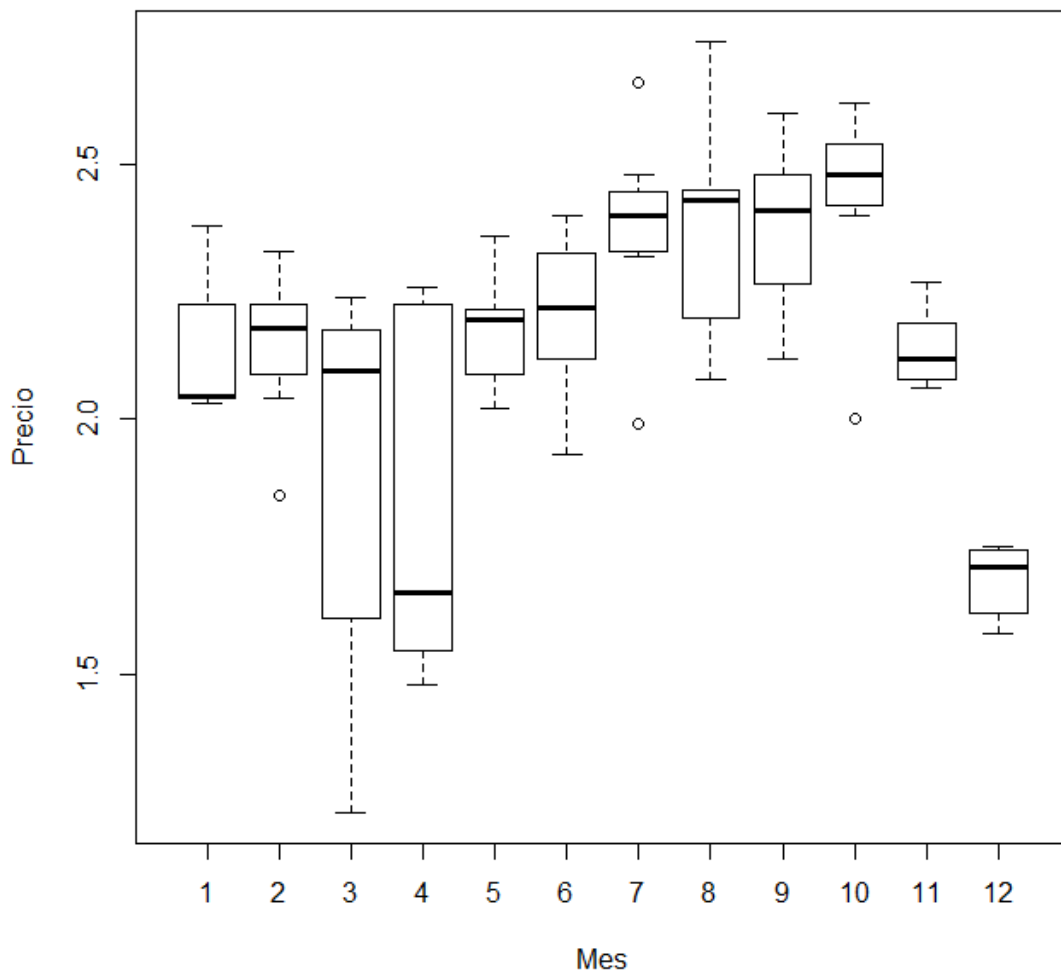


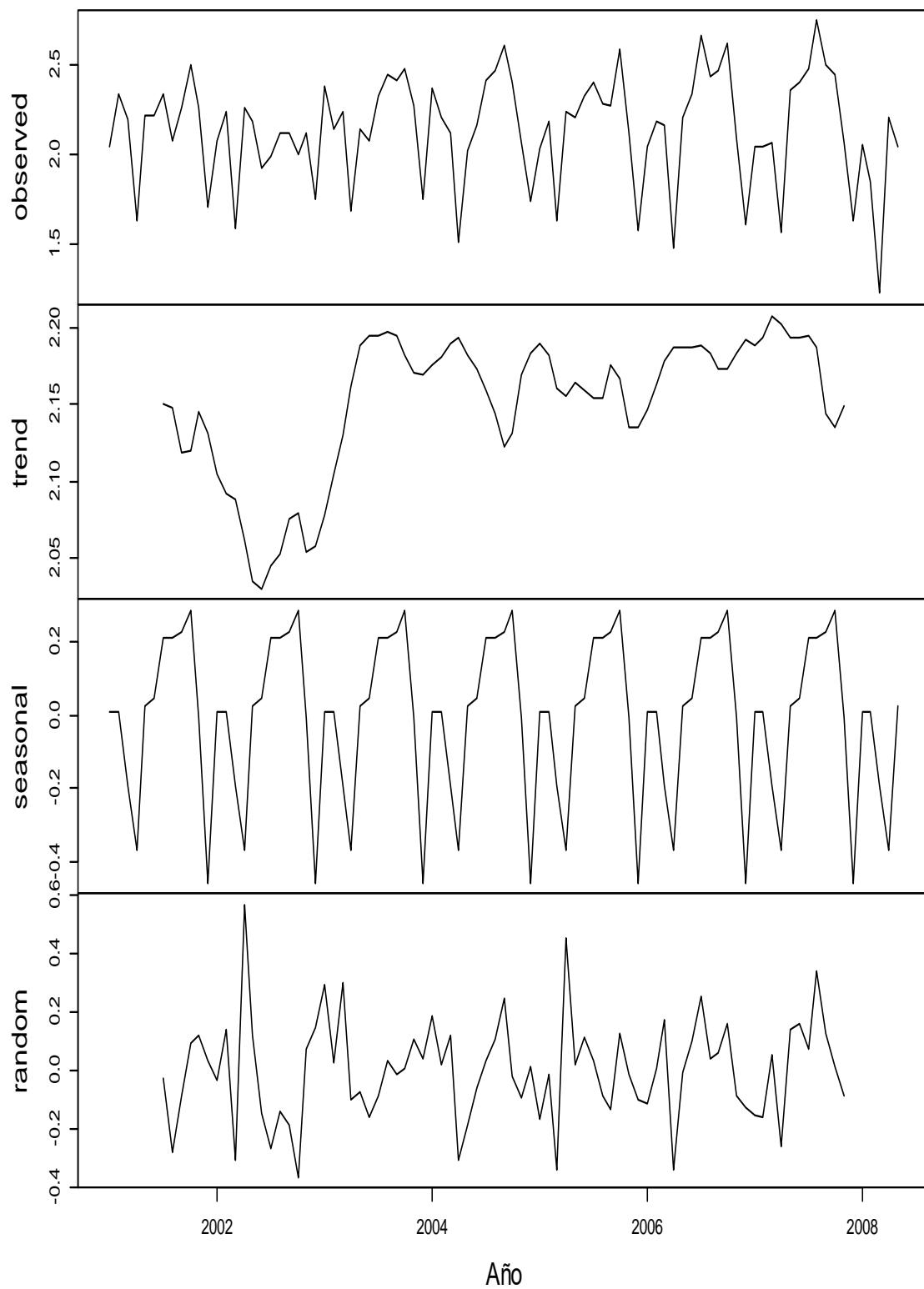
Figura 1. Distribución mes a mes de la media del precio de las hamburguesas en el periodo de observación del conjunto de datos.

En la gráfica 2 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 $s=12$

En la sección de los datos observados se evidencia que la varianza del precio de las hamburguesas tiene una tendencia creciente, es decir, parece ser no constante; sin embargo, parece tener un aumento particularmente alto desde el 2006 con comportamiento del precio saliéndose de las bandas entre 1.5 a 2.5 en comparación de los otros años.

Por otro lado, la componente estacional muestra una periodicidad clara, con una caída drástica del precio de la hamburguesa cada fin de año, seguido por un aumento de los precios hasta el noviembre siguiente.

Decomposition of additive time series



Grafica 2. Serie de tiempo observada y descomposición de la serie de tiempo por tendencia, estacionalidad y error.

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.

6 MODELO EXPONENCIAL CON INDICADORAS

Los parámetros ajustados del modelo exponencial con variables indicadoras, de acuerdo con la Ecuación 1, se presentan en la Tabla 2.

Para verificar la utilidad del modelo se realizaron pruebas de hipótesis de significancia de los parámetros del modelo. Para identificar los parámetros significativos se realizan pruebas t independientes sobre cada parámetro, el valor p para cada parámetro se presenta en la Tabla 2.

En esta prueba, la hipótesis nula es que el parámetro β_i con $i=0,...,13$ no es significativo. La hipótesis alternativa indica que el parámetro β_i con $i=0,...,13$ es significativo; y el criterio de rechazo es que el valor p sea menor a 5%.

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

Como conclusión preliminar de las pruebas de hipótesis de los parámetros, de acuerdo con la información de la Tabla 2, las variables indicadoras muestran significancia, excepto por el mes de abril (β_5), lo cual indica que hay estacionalidad en el modelo. Por otro lado, β_1 que es el parámetro relacionado con

la tendencia exponencial lineal no fue significativo, lo cual, puede representar que la serie no tenga una tendencia monótona.

Por lo tanto, se realiza una prueba Mann Kendall para validar esta hipótesis. Donde, la hipótesis nula es que la serie no tiene una tendencia monótona y la hipótesis alternativa es que tiene una tendencia monótona. Con un criterio de rechazo de valor P menor a 5%.

El valor p obtenido para esta prueba fue de 59.1%, por consiguiente, no se puede rechazar la hipótesis nula, entonces la serie no presenta una tendencia lineal.

Una vez estudiados los parámetros del modelo y sus implicaciones, se procedió a comparar los datos observados contra el modelo exponencial lineal con indicadores propuesto, la figura 3 ilustra el resultado mencionado previamente.

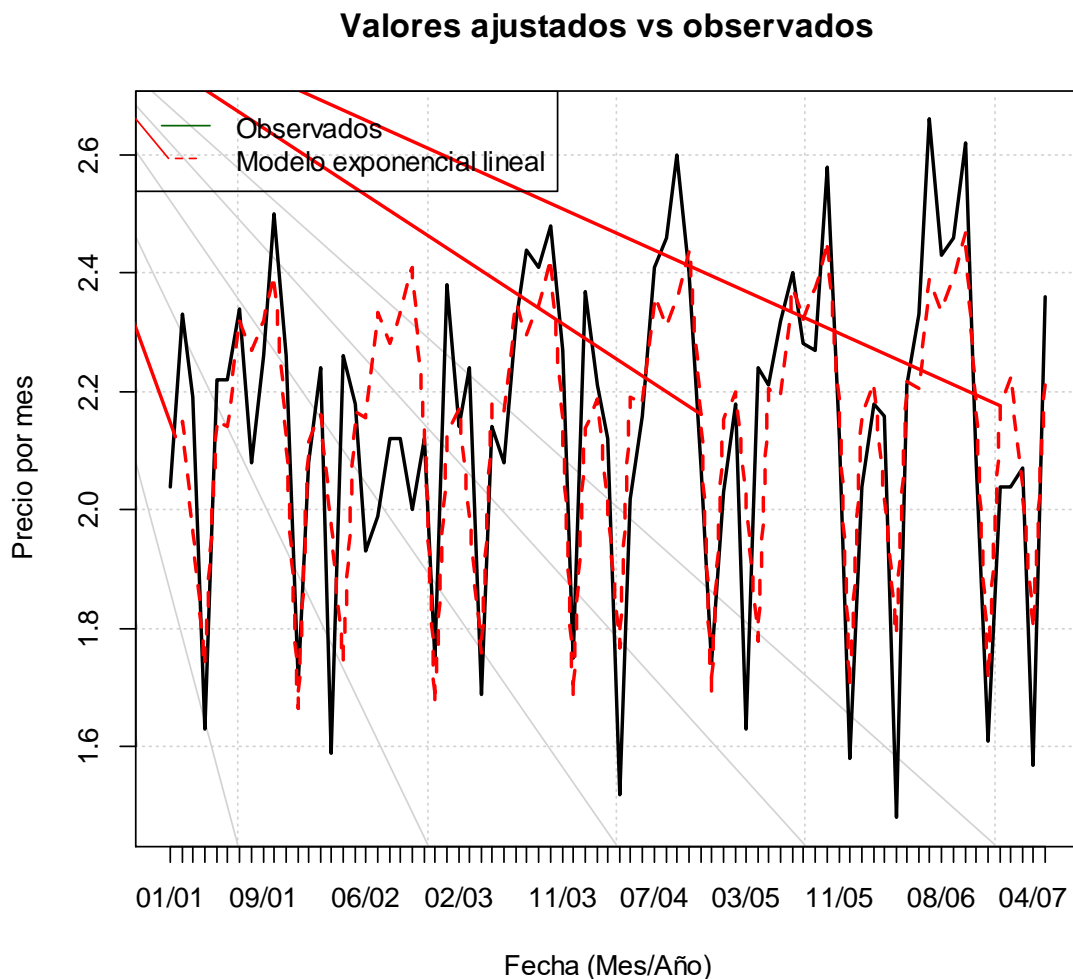


Figura 3. Valores observados contra ajuste del modelo de componentes con tendencia exponencial lineal y estacionalidad con variables indicadoras.

7 MODELO ESTRUCTURAL BÁSICO (BSM)

Para el modelo estructural estudiar las componentes del modelo estructural gaussiano o básico, se extraen las varianzas de los errores para determinar cuales son los componentes que aportan dentro del modelo. La tabla 3 muestra las varianzas de los parámetros de este, correspondientes a: μ , β , S_t , ε , de acuerdo con la Ecuación 2

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

Nivel	Pendiente	Estacionalidad	Error
4.1E-05	0.0E+00	1.4E-02	2.9E-02

Si bien para el estudio de estas componentes no se realizan pruebas de hipótesis, la magnitud de las varianzas de los parámetros determina el aporte de estas componentes al modelo. Como se puede observar el nivel y la pendiente, es decir, los parámetros asociados a la tendencia son los que menos aportan al modelo; considerando esto y el resultado de la prueba de Mann Kendall realizada previamente, podemos interpretar que la serie no tiene una tendencia clara. Por otro lado, si cuenta con una estacionalidad y componente aleatorio con mayor influencia.

Una vez estudiados los parámetros del modelo y sus implicaciones, se procedió a comparar los datos observados contra el modelo BSM propuesto, la figura 4 ilustra el resultado mencionado previamente.

Valores ajustados BSM vs observados

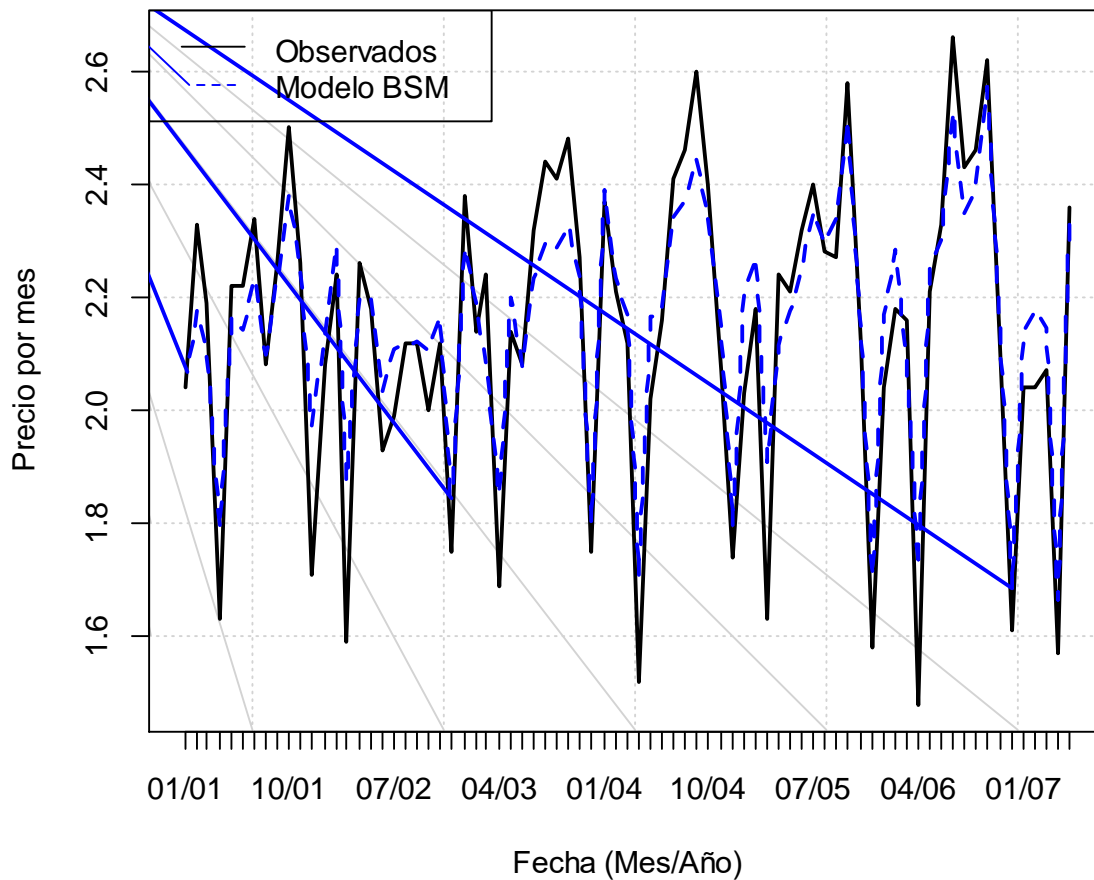


Figura 3. Valores observados contra ajuste del modelo de estructural básico (BSM).

8 ANÁLISIS DE RESULTADOS

En la figura 4 se muestra un contraste entre ambos modelos ajustados y la serie de tiempo observada, se puede observar que en el periodo correspondiente al año 2002, hubo una discrepancia considerable entre el modelo exponencial lineal y la serie de tiempo observada, esto debido a que en este año hubo un cambio en el patrón estacional global esperado. Por otro lado, el modelo BSM al ser un modelo local, dependiente del tiempo, pudo predecir este cambio en la serie y no hubo una diferencia considerable con el modelo. A partir de un análisis visual preliminar, se espera que el modelo BSM tenga un mejor ajuste a la serie de tiempo.

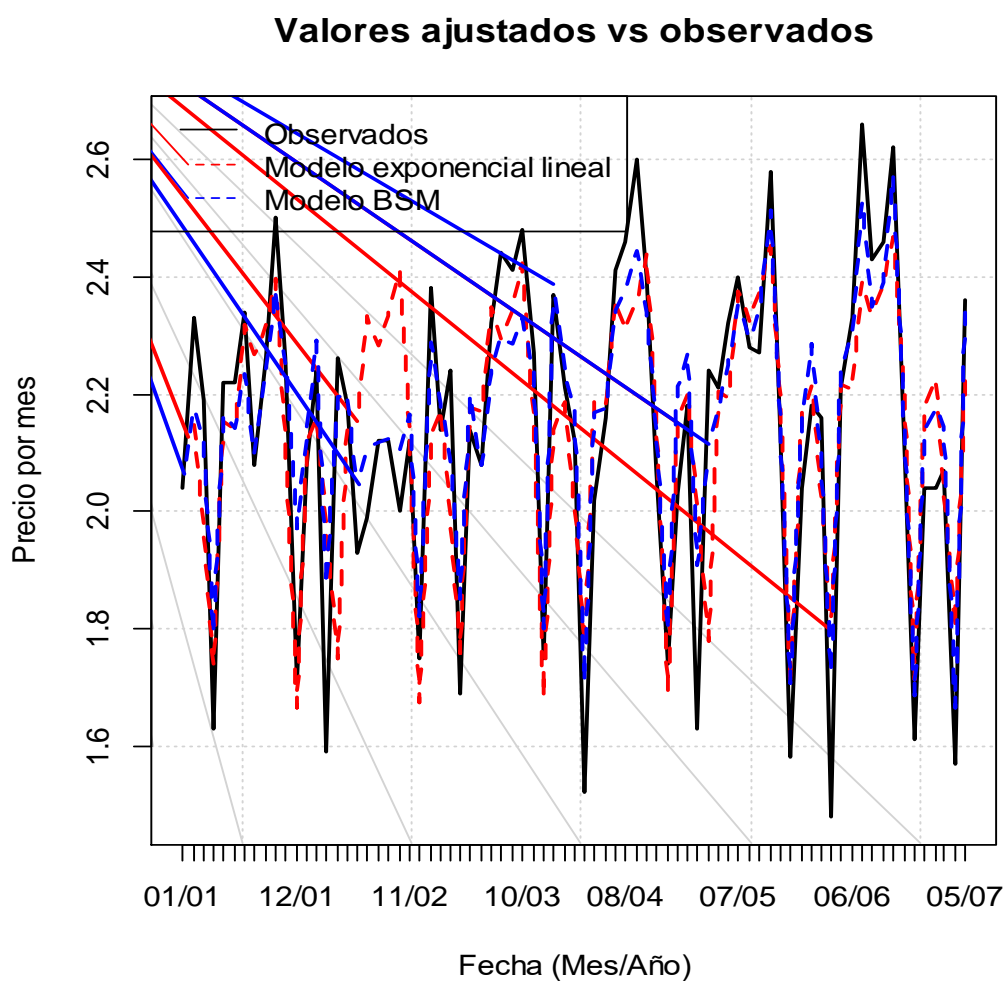


Figura 4. Modelos ajustados y serie de tiempo observada

Se analizó la comparación de los modelos a partir de criterios de ajuste del modelo que se presentan en la tabla 4, acá se puede evidenciar, que el modelo BSM presenta mejor ajuste a partir de los criterios RMSE (menor), R^2 ajustado (mayor), AIC (menor) y BIC (menor).

Tabla 4. Criterios de comparación de ajuste para los modelos estudiados

Modelo	RMSE	R^2 ajust.	Log AIC	Log BIC
Exponencial lineal con variables indicadoras	0.189	0.538	-3.174	-2.779
Estructural básico (BSM)	0.109	0.846	-4.389	-4.297

Una vez comprobados los ajustes del modelo, 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 cual modelo tiene mayor precisión de pronóstico. La figura 5 muestra la predicción de los modelos y la serie de tiempo observada, se puede observar una discrepancia considerable de ambos modelos con la serie de tiempo desde enero del 2008 hasta el final de la serie.

Predicción datos validación cruzada por modelo

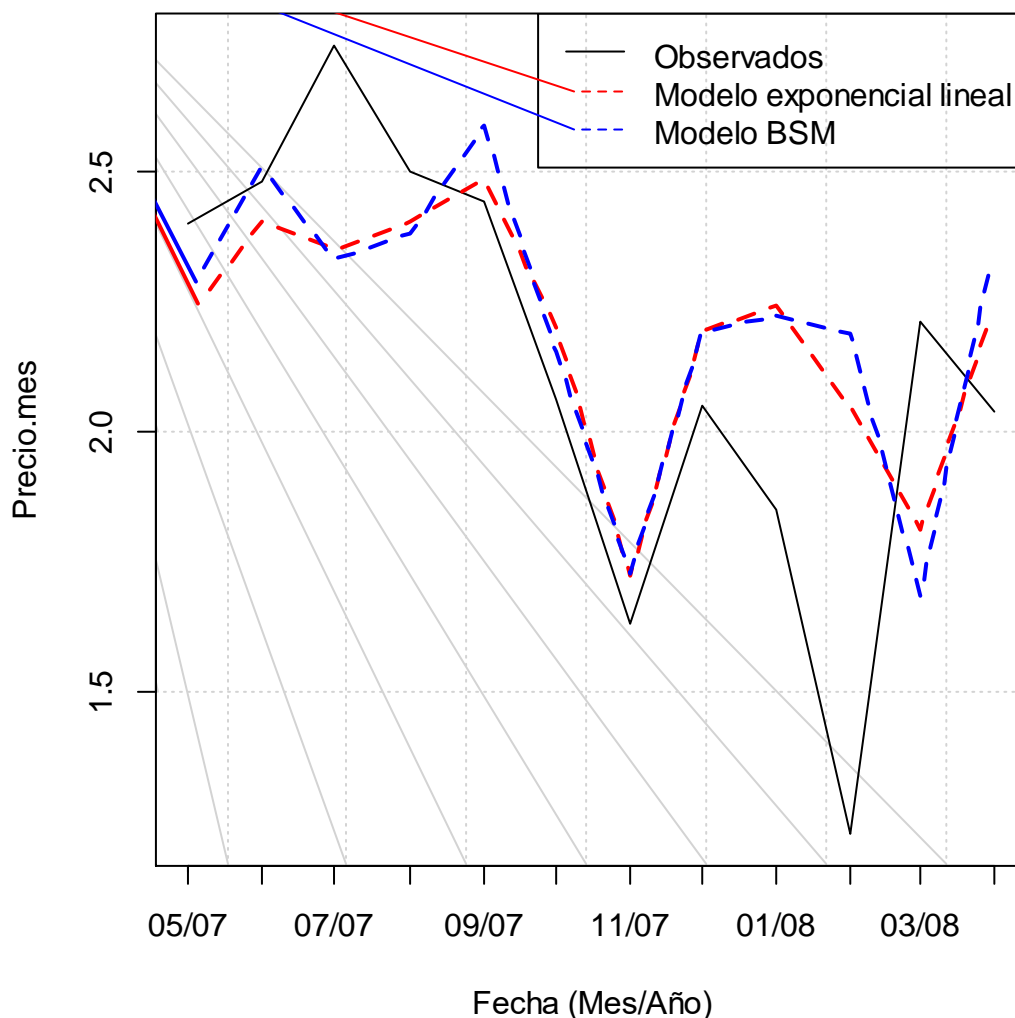


Figura 5. Pronósticos por modelo y serie de tiempo observada mayo 2007 a abril 2008

La tabla 5 muestra los criterios de pronóstico para ambos modelos. En todos los criterios el modelo exponencial lineal con variables indicadoras mostró mayor precisión en la predicción; sin embargo, la distancia entre estos dos modelos no es muy amplia al menos para ME, RMSE y MAE. Una posible razón por la cual el modelo estructural básico sea menos preciso para el pronóstico, pero tenga tan buen ajuste es que esté sobre ajustado.

El MAPE, RMSE y U de Theil deben ser lo más pequeños posible, por lo tanto, el modelo exponencial lineal con indicadoras presenta menor error, y mayor precisión en el pronóstico (U de Theil)

Tabla 5. Criterios de pronóstico para modelos estudiados

Modelo	RMSE	MAPE	U de Theil
Exponencial lineal con variables indicadoras	0.325	13.785	0.648
Estructural básico (BSM)	0.376	15.631	0.774

9 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 dos modelos bajo la estrategia de validación cruzada, donde se dejaron 77 datos para el ajuste y 12 datos para verificar los pronósticos.

Se comparó el desempeño de ajuste y pronóstico para un modelo global de componentes exponencial lineal y estacionalidad con variables indicadoras, y un modelo BSM.

Los resultados de este estudio arrojaron que el modelo de componentes muestra mayor precisión en los pronósticos a pesar de tener menor ajuste que el modelo BSM, esto puede ser consecuencia de la flexibilidad del modelo BSM, ya que este se basa en marchas aleatorias, por lo que se puede ajustar fácilmente a cambios temporales en la serie que no representen un cambio a largo plazo.

La serie de tiempo estudiada tiene un componente estacional muy marcado, mientras que no muestra una tendencia clara, lo cual fue probado por ambos modelos y por la prueba Mann Kenndall.

10 REFERENCIAS

- [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.