Análisis, modelizado y pronóstico de la serie de índices de precios al consumidor en Medellín 2010-2015

Tatiana Espinal Rodríguez
Estudiante de Ingeniería Administrativa
Autora al que se le dirige la correspondencia
E-mail: tatiana.espinal@upb.edu.co

Paula Andrea Giraldo Botero
Estudiante de Ingeniería Administrativa.

Natalia Peña Córdoba
Estudiante de Ingeniería Administrativa.

Laura Lotero
Ingeniera Industrial, M.Sc., Ph.D. en Ingeniería. Docente de la Facultad de Ingeniería Industrial de la UPB sede Medellín. Grupo de investigación de Sistema Aplicados a la Industria, Gisai.

Universidad Pontificia Bolivariana, sede Medellín.

Resumen

En este artículo se presenta un análisis de la serie del Índice de Precios al Consumidor (IPC) de la ciudad de Medellín, que estima la evolución del costo promedio de la canasta de bienes y servicios representativa del consumo final de los hogares. Se elige una muestra compuesta por 72 datos mensuales, desde enero de 2010 hasta diciembre de 2015, para desarrollar un pronóstico del trimestre posterior a través de un modelo Arima y comparar con un método de descomposición usando el filtro Holt-Winters. Partiendo de lo anterior, se elige el modelo Arima (0, 2 ,1) como el más acertado para ajustar este tipo de series, dado que presenta bajos valores de error y coeficientes significativos, con un nivel de confianza del 95%.

Palabras clave

IPC, Arima, pronóstico, serie de tiempo, Holt-Winters.

Abstract

This paper presents an analysis of the Consumer Price Index (CPI) series of Medellin, which is an index that estimates the evolution of the average cost of the basket of products and services representative of final consumption of households. The sample consisted of 72 monthly data from January 2010 to December 2015, and then time series models such as ARIMA were compared to decomposition methods such as Holt Winters filter in order to forecast the subsequent 2016 quarter. Taking into account the above, the ARIMA(0,2,1) model is chosen as the most appropriate to model this kind of series, because it has low error values and significant coefficients at a confidence level of 95%, allowing to develop a forecast for the next three months that look like the real values of 2016.

Keywords

CPI, ARIMA, forecasting, time series, Holt-Winters.

Introducción

El Índice del Precio al Consumidor (IPC) se establece como un instrumento estadístico que a partir de la información recolectada por el Departamento Administrativo Nacional de Estadística – DANE, valora la evolución del costo promedio de una canasta de bienes y servicios representativa del consumo final de los hogares, expresado en relación con un periodo base. (Banco de la República).

La importancia del IPC radica en su uso para el desarrollo de análisis de situaciones de carácter económico, empleadas para la toma de decisiones por parte del Gobierno y otros agentes privados, como ajustes salariales y poder adquisitivo de la moneda (DANE, 2009).

El presente artículo pretende analizar las variaciones que se han generado en el IPC de la ciudad de Medellín en los últimos seis años, y tiene como objeto modelar su comportamiento para llevar a cabo una predicción de los valores futuros a corto plazo, mediante una transformación de la serie no estacionaria con el programa estadístico “R” (R Core Team, 2012).

Antecedentes

Investigaciones que pretenden analizar la serie de IPC en diversos contextos han sido desarrollas por autores como Santana (2006), quien ha empleado las redes neuronales, que constan de diversas capas por nivel, para hacer un procesamiento no lineal de los patrones recibidos en la predicción de los valores futuros de la serie de inflación colombiana, contrastó los resultados obtenidos con los de otras metodologías tradicionales como Sarima de Box-Jenkins y el suavizado exponencial, concluyendo que este primero arroja resultados más precisos.

Asimismo, Huwiler & Kaufmann (2013) desarrollaron un estudio del IPC desagregado de la economía suiza para llevar a cabo pronósticos a corto plazo empleando modelos Arima, concluyendo que la estimación mediante estos modelos para el ítem de gasto del IPC permite obtener resultados más precisos que la aplicación directa del método Arima en el IPC total.

Otro de los estudios en el que se evidencia el uso de modelos estadísticos para el tratamiento de series de tiempo financieras es el de Gallego (2010), en el cual se introducen diversas técnicas de análisis con el fin de detectar el comportamiento no lineal de la serie de la tasa representativa del mercado colombiano (TRM) y ajustarla a un modelo para su pronóstico. Se contrastó con modelos ARMA y Arima (1,1,3), sin embargo, se concluyó que ninguno de los modelos se adaptó satisfactoriamente, dado a que el error RMSE resultó mayor que la variación promedio de la variable.

Adicionalmente, en México, López (2004) evalúa la contribución de tres modelos ARCH para el análisis de series de tiempo en los mercados accionarios, específicamente del Índice de Precios y Cotizaciones del mercado, también llamado IPC. En el estudio se ajusta un modelo autorregresivo para series de tiempo heterocedásticas y se concluye que aquel que mejor se adapta es el Igarch, es decir, un modelo Garch integrado, que resulta equivalente a un modelo de promedios móviles ponderados exponencialmente para explicar la volatilidad del Índice en la bolsa de valores mexicana que tiene un comportamiento similar al Índice de Precios al Consumidor.

Análisis de series de tiempo

Una serie de tiempo es una sucesión de variables aleatorias ordenadas y equidistantes de acuerdo a una unidad de tiempo, Yt, t = 1…T, esta se puede descomponer en tres componentes, tendencia Tt, estacionalidad St y error (Tróchez & Valencia, 2014):

$$Y_t = T_t + S_t + \varepsilon \qquad (1)$$

Las componentes T_t y S_t son funciones determinísticas que dependen del tiempo. Su evolución es perfectamente predecible. (Giraldo, 2006).

Modelo Arima

Una serie de tiempo X_t, que presente las características AR y MA de manera conjunta, seguirá un proceso ARMA (p, q), con p términos autorregresivos y q términos de media móvil, estos modelos permiten aproximar la estructura de covarianza de un proceso estacionario hasta el nivel que se fije previamente. Los modelos Arima son una extensión de los ARMA que se utiliza para modelar algunos procesos no estacionarios.

De manera que, se dice que un proceso X_t tiene estructura Arima (p,d,q) si existen polinomios $\Phi(x)$ y $\Theta(x)$ de grado p y q, respectivamente, verificando que (Carvajal, 2014):

$$\Phi(B)\ (1-B)^d\ (X_t-\mu)=\theta(B)\varepsilon_t \qquad (2)$$

El patrón que debe seguir la Función de Auto-correlación (FAC) y la Función de Auto-correlación Parcial (FACP) para la identificación del orden de un modelo puro AR (p) es el siguiente: la FACP presenta los p primeros valores distintos de cero y el resto de valores son cero o muy próximos a cero con un comportamiento sinusoidal, y la FAC presenta un decrecimiento exponencial y un comportamiento sinusoidal. (Guerrero, 2003).

Asimismo, el patrón que deben seguir la FAC y la FACP para la identificación del orden de un modelo puro MA (q) es el siguiente: la FAC presenta los q primeros valores distintos de cero y el resto de valores son cero o muy próximos a cero con un comportamiento sinusoidal, y la FACP presenta un decrecimiento exponencial y un comportamiento sinusoidal. (Guerrero, 2003).

Criterio AIC y BIC

Los índices AIC y BIC (Criterio de información de Akaike y criterio de información bayesiano, respectivamente) son de uso frecuente para la selección de modelos. (López A. M., 2011).

El criterio de información de Akaike AIC, se define como:

$$AIC = e^{2k/T}\frac{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}{T} = \frac{(T-k)\hat{\sigma}^2}{T}e^{2k/T} \qquad (5)$$

Donde $\sum_{t=1}^{T}\hat{\varepsilon}_t^2 = (T-k)\sigma^{2k/T}$. El AIC es un estimador de σ^2 pero penalizado por el número de grados de libertad, es decir, aumenta cuando k aumenta. El AIC también se define como el logaritmo de la ecuación (2).

Asimismo, el criterio de información de Schwarz BIC, se define como:

$$BIC = T^{k/T}\frac{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}{T} \qquad (6)$$

El BIC también se define como el logaritmo de la ecuación (5).

La regla para utilizar AIC y BIC indica para escoger entre varios modelos; se elige aquél que presente menor AIC o menor BIC. (Giraldo, 2006).

Significancia de los coeficientes

Para evaluar la significancia de los coeficientes, después de haber ajustado un modelo, es necesario aplicar la siguiente fórmula:

$$Significancia\ coef = \frac{Valor\ Estimado}{Error\ Estándar} \quad (7)$$

Si el valor absoluto del valor obtenido es superior a 1,96 entonces el coeficiente es estadísticamente diferente de cero con un 95% de confianza. De igual manera, si es superior a 1,645 es diferente de cero con un 90% de confianza.

Indicadores

Estos sirven para comparar la efectividad de diferentes modelos utilizados. Siempre se busca el valor menor en los indicadores tales como MAPE, MAD y MSD, ya que representan un mejor ajuste del modelo. (Reyes, 2007).

MAPE

El Porcentaje Promedio Absoluto de Error (MAPE por sus siglas en inglés) mide la exactitud de los valores estimados de la serie de tiempo. La exactitud se expresa como un porcentaje con y_t igual al valor observado, $\hat{y}_t$ es el valor estimado y n el número de observaciones.

$$MAPE = \frac{\sum\left|\frac{(y_t-\hat{y}_t)}{y_t}\right|}{n} * 100\ (y_t \neq 0) \quad (8)$$

MAD

La desviación media absoluta (MAD, por sus siglas en inglés), mide la exactitud de los valores estimados de la serie de tiempo. Expresa la exactitud en las mismas unidades de los datos.

$$MAD = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|}{n} \tag{9}$$

MSD

La desviación cuadrática media (MSD, por sus siglas en inglés), es más sensible a errores anormales de pronóstico que el MAD.

$$MSD = \frac{\sum_{t=1}^{n} |y_t - \hat{y}_t|^2}{n} \tag{10}$$

Serie de IPC Medellín

La serie se realizó a través del programa estadístico "R" (R Core Team, 2012) empleando las cifras provenientes del DANE.

La Figura 1 presenta los datos del Índice de Precios al Consumidor (IPC) de la ciudad de Medellín para un periodo de seis años, comprendido entre el mes de enero de 2010 y el mes de diciembre de 2015. Para este último, según el periódico El Colombiano (2015), se observa un incremento significativo, hecho que no se evidenciaba desde el 2008. Para el 2015, las otras ciudades de Colombia presentaron niveles bajos en comparación con la ciudad de Medellín, debido a que la variable que lideraba correspondía a la de alimentos.

La serie del IPC en la ciudad de Medellín presenta los porcentajes de forma mensual, está conformada por 72 observaciones y se caracteriza por:

- Poseer un componente tendencial de comportamiento positivo respecto a la media a largo plazo.

- No poseer factores estacionales, dado que no presenta un comportamiento periódico.
- Poseer un comportamiento aleatorio, es decir, que existe cierta irregularidad en los datos.
- Ser homocedástica, ya que su varianza es aproximadamente constante a lo largo de la serie.

Debido a esto, la serie mensual del IPC en la ciudad de Medellín en el periodo 2010-2015 es no estacionaria, por su componente tendencial creciente a largo plazo.

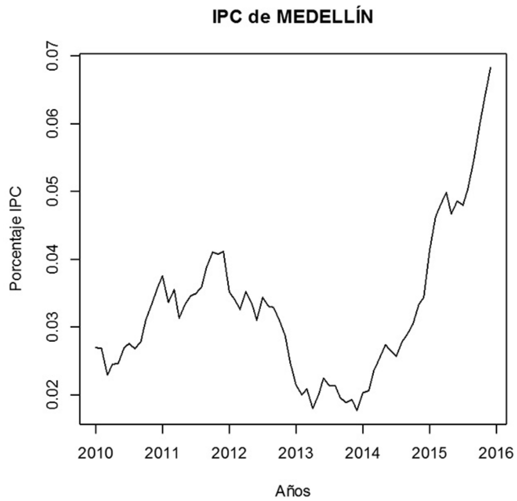

Figura 1. Serie IPC en la ciudad de Medellín para el periodo 2010-2015.

Fuente: Elaboración propia con cifras del DANE.

Modelo Arima para la serie IPC de Medellín

Con el fin de desarrollar un modelo Arima para la serie del IPC de la ciudad de Medellín, inicialmente se procedió a la transformación de la misma, mediante la aplicación de una doble diferencia que permitió eliminar la tendencia. En la Figura 2 se muestra el gráfico de la serie transformada (Estacionaria):

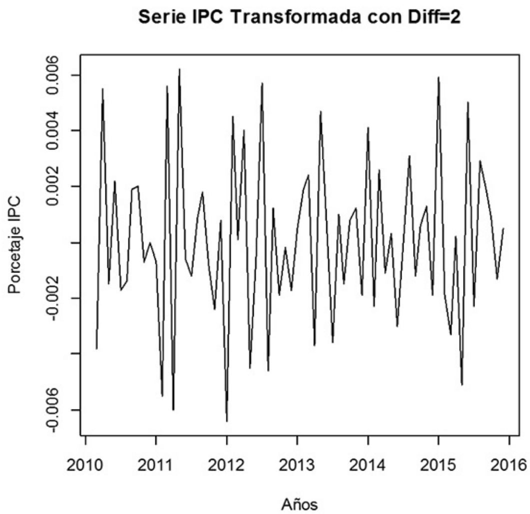

Figura 2. Serie IPC en la ciudad de Medellín transformada con dos diferencias para el periodo 2010-2015.

Fuente: Elaboración propia en R.

Posteriormente, se desarrolló el correlograma de la FAC y FACP para la serie transformada, presentados en la Figura 3.

A partir del análisis de la Figura 3, se propone el siguiente modelo que representa el IPC de la ciudad de Medellín para el periodo 2010-2015: Arima (0, 2, 1) ya que la FAC de la serie presenta un único valor significativo y, se observa una tendencia cuadrática en la serie original.

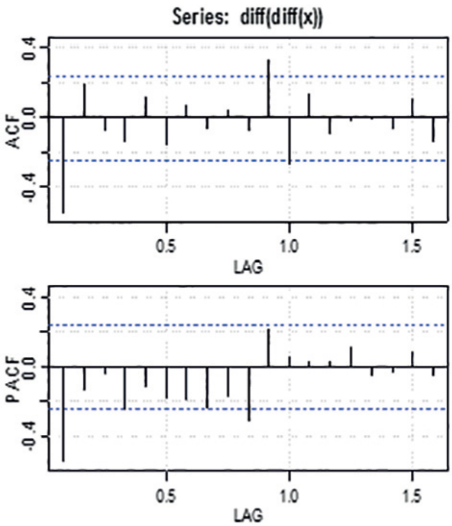

Figura 3. FAC y FACP de la serie transformada con dos diferencias.

Fuente: Elaboración propia en R.

Modelo Arima (0, 2, 1)

Diagnóstico:

a. La gráfica de residuos estandarizados no presenta tendencia ni heterocedasticidad.

b. La FAC de los residuales tiene un único dato significativo.

c. El gráfico QQ de los residuos estandarizados es aproximadamente normal.

d. Tan solo dos valores “p” del estadístico “Ljung-Box” se encuentran por fuera de las bandas de significancia.

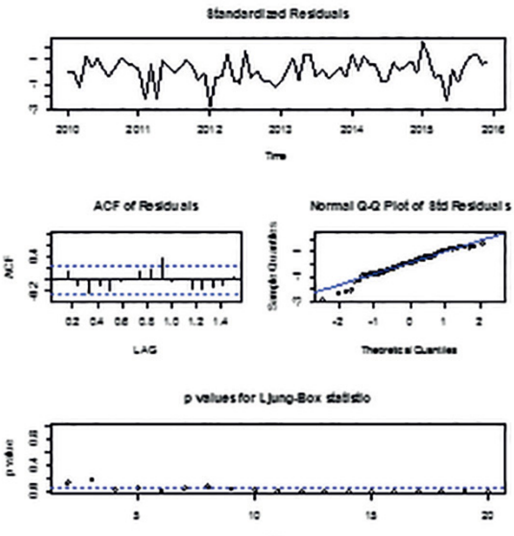

Figura 4. Diagnóstico del modelo Arima (0, 2, 1)

Fuente: Elaboración propia en R.

Por tanto, se verifican las medidas de error:

AIC: -11.07127
BIC: -12.03965

Análisis de los coeficientes:

Dado que la significancia del coeficiente MA es de 10,83 y, por ende, mayor a 1,96 es posible asumir que es diferente de cero con un 95% de confianza.

Tabla 1. Coeficientes del Arima (0, 2, 1)

	MA(1)
Valor estimado	-0.8296
Error estándar	0.0766

Métodos de descomposición y suavizado

Se procede a desarrollar una descomposición por medio del método Holt-Winters y un suavizado a través del método de Promedios Móviles, con el fin de observar mejor los patrones de la serie.

Pese a que la serie IPC de la ciudad de Medellín no presenta un componente marcado de estacionalidad se procedió a aplicar un suavizado con promedios móviles a la serie original con un periodo igual a 12 meses, el cual se ajustaba de mejor manera, para observar con exactitud la tendencia (Ver figura 5).

De igual forma, se desarrolló una descomposición Holt-Winters de la serie original respecto a su nivel, tendencia y posible estacionalidad, presentada en la Figura 6. El filtro de Holt-Winters es un método de suavizado exponencial que incluye componentes estacionales y de tendencia, y es uno de los más usados para el pronóstico de series (Goodwin, 2010).

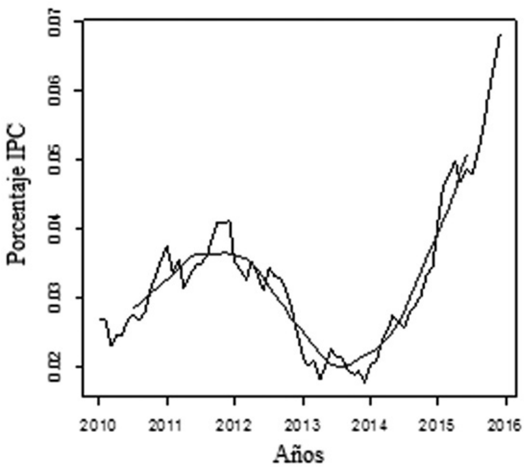

Figura 5 Suavizado serie IPC con Promedios Móviles.

Fuente: Elaboración propia en R.

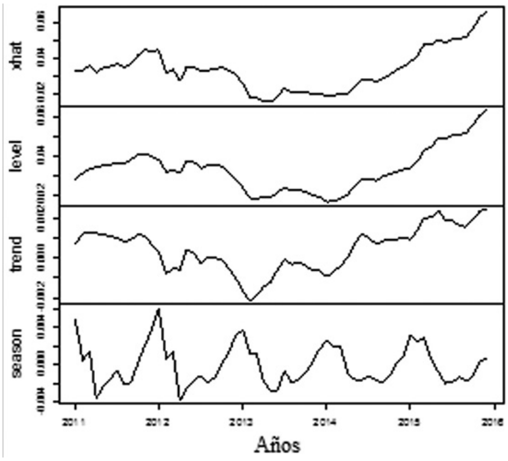

Figura 6. Descomposición Holt-Winters de la serie.

Fuente: Elaboración propia en R.

Pronósticos

A continuación, en la Tabla 3 se presentan los pronósticos del IPC de la ciudad de Medellín para el primer trimestre del año 2016, obtenidos a través del modelo Arima (0, 2, 1), y la descomposición Holt-Winters y se comparan con los valores reales reportados por el Banco de la República.

Tabla 3. Pronósticos con los diferentes métodos para el primer trimestre de 2016 vs valores reales.

	Enero	Febrero	Marzo
Holt-Winters	0.07378	0.07617	0.07818
ARIMA(0,2,1)	0.07118	0.07415	0.07713
Valor Real	0,07130	0.07290	0,07600

Al contrastar los valores pronosticados con los reales se obtienen los indicadores de ajuste del pronóstico para cada método, reportados en la Tabla 4, en la que se evidencia que el modelo Arima (0,2,1) proporciona un pronóstico más acertado que el filtro Holt-Winters, ya que sus indicadores de error son menores, dada la cercanía con los valores reales.

Tabla 4. Indicadores para evaluar pronósticos

	MAPE	MAD	MSD
Holt-Winters	3,61%	0,00264	7,190E-06
ARIMA(0,2,1)	1,13%	0,00084	9,553E-07

Discusión

Entre diversos modelos de series de tiempo Arima analizados, el que mejor se ajustó a la serie IPC de la ciudad de Medellín para el periodo 2010-2015 fue el modelo Arima (0,2,1), siendo sus coeficientes estimados significativos a un nivel de confianza del 95%.

Con base en este se comparó con uno de los métodos de descomposición y suavizado más utilizado para el pronóstico de series de tiempo (filtro Holt-Winters) y se determinó que el método que presenta mejores pronósticos para la serie del IPC de Medellín es el modelo Arima (0,2,1) dado que los valores de sus errores son menores a los que presenta el método de Holt-Winters.

Conclusiones

La aplicación de modelos Arima en el tratamiento de series de tiempo resulta ser bastante útil para la modelación de las mismas, debido a que transforman los componentes no estacionarios de la serie para desarrollar un análisis más efectivo. Partiendo de lo anterior, en este trabajo se ajustó la serie IPC de la ciudad de Medellín para el periodo 2010-2015 a través de un modelo Arima (0,2,1), usando los métodos de selección del orden del modelo Arima (p,d,q) y validando los supuestos por medio del diagnóstico de los residuos.

Los pronósticos se realizaron empleando un método de suavizado muy usado filtro Holt-Winters y comparando con un modelo de serie de tiempo Arima, para estimar los valores del IPC del trimestre posterior a los datos de la muestra. Al comparar los valores estimados con los reales se evidenciaron resultados similares, sin embargo, se observa que usando el modelo Arima para el mes de enero el pronóstico se ubica un 0.17% por debajo del valor real, para febrero y marzo se encuentra un 1.69% y 1.46%, respectivamente, por encima del valor real, mientras que con el método de suavizado la diferencia entre el valor pronosticado y el real es mayor.

Los hallazgos obtenidos a partir del modelo Arima (0,2,1) fueron satisfactorios, sin embargo, es posible que se logre un modelo más preciso si se lleva a cabo un análisis del IPC desagregado, ya que los modelos de series de tiempo solo se tienen en cuenta a sí mismas (la variable o los errores) y, de esta manera, se podrían tener en cuenta otras variables para desarrollar el estudio, que permitirían que los pronósticos a largo plazo no se desvíen de manera significativa de los valores reales.

Referencias

1. Argáez, J., Batún, J., Guerrero, E., Kantúm, D., Medina, S., & Pantí, H. (2014). Un paseo por el modelo Garch y sus variantes. Abstraction & Application 10, 35-50.

2. Arias, F. (6 de diciembre de 2015). IPC mayor a 6,0% vuelve a Medellín luego de 7 años. El Colombiano. Recuperado de: http://www.elcolombiano.com/negocios/inflacion-medellin-noviembre-2015-NX3231492

3. Banco de la República. Índices de precios. Recuperado de: http://www.banrep.gov.co/series-estadisticas/see_precios.htm

4. Carvajal, A. (2014). Series Temporales: Modelos Heterocedásticos condicionales. Una aplicación usando R. Granada: Tesis de Máster Universidad de Granada. Granada.

5. DANE. Índice de Precios al Consumidor. Recuperado de: http://www.dane.gov.co/files/investigaciones/boletines/ipc/ipc_autocapacita.pdf

6. DANE. (2009). Metodología Índice de Precios al Consumidor. Recuperado de: www.dane.gov.co/files/investigaciones/fichas/IPC.pdf

7. Gallego, J. P. (2010). Aplicación de la teoría de caos para el análisis y pronóstico de series de tiempo financieras en Colombia. Medellín: Universidad Nacional de Colombia. Tesis de Maestría.

8. Giraldo, N. (2006). Notas de Clase - Series de Tiempo con R. Medellín: Universidad Nacional de Colombia. Recuperado de: http://www.unalmed.edu.co/~ndgirald/Archivos%20Lectura/Archivos%20curso%20Series%20EIO/Notas%20de%20Clase.%20Series%20de%20Tiempo%20con%20R.pdf

9. Goodwin, P. (2010). The holt-winters approach to exponential smoothing: 50 years old and going strong. Foresight, 19, 30-33.

10. Guerrero, V.M. (2003). Análisis estadístico de series de tiempo económicas. International Thomson Editores. Ciudad de México.

11. Huwiler, M., & Kaufmann, D. (2013). Combining disaggregate forecasts for inflation: The SNB´s ARIMA model. Swiss National Bank Economic Studies.

12. ICESI, U. (2012). Índice de Precios al Consumidor (IPC). Recuperado de: http://www.icesi.edu.co/cienfi/es/glosario.php

13. López, A. M. (2011). Estudio del AIC y BIC en la selección de modelos de vida con datos censurados. Guanajuato: Tesis de Maestría Centro de Investigación en Matemáticas, Cimat.

14. López, F. (2004). Modelado de la Volatilidad y Pronóstico del Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores. Contaduría y Administración. 213, 43-72.

15. R Core Team. (2012). R: A language and environment for stadistical computing.

16. Reyes, P. (2007). Metodología de Análisis de Series de Tiempo. Recuperado de: www.icicm.com/files/SeriesDeTiempoComp.doc

17. Santana, J. C. (2006). Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana. Revista Colombiana de Estadística, 29, 77-92.

18. Tróchez, J., & Valencia, M. (2014). Análisis de series temporales en el sector lácteo de Antioquia para detectar efectos de la apertura comercial. Revista Investigaciones Aplicadas, 8, 140-151.