

NOTA SOBRE LA PREDICCIÓN DEL ÍNDICE DE PRECIOS AL CONSUMIDOR USANDO REDES NEURONALES ARTIFICIALES*

*Juan David Velásquez H.***

*Carlos Jaime Franco C.****

* Artículo de investigación científica y tecnológica, producto de la investigación realizada por los grupos de Sistemas e Informática y de Computación Aplicada en el modelado y la predicción de variables económicas y financieras. Patrocinado por la Facultad de Minas, Universidad Nacional de Colombia, Medellín, Colombia. El artículo se recibió el 10-08-2009 y se aprobó el 01-11-2010.

** Doctor en Ingeniería, Área de Sistemas Energéticos, Universidad Nacional de Colombia, Medellín, Colombia, 2009; Magister en Ingeniería de Sistemas, Universidad Nacional de Colombia, Medellín, 1997. Profesor asociado, Escuela de Sistemas, Facultad de Minas, Universidad Nacional de Colombia. Director del grupo de investigación Computación Aplicada. Miembro del grupo de investigación Sistemas e Informática, Facultad de Minas, Universidad Nacional de Colombia. Correo electrónico: jdvelasq@unal.edu.co.

*** Doctor en Ingeniería, Área de Sistemas Energéticos, Universidad Nacional de Colombia, Medellín, Colombia, 2002; Magister en Aprovechamiento de Recursos Hidráulicos, Universidad Nacional de Colombia, Medellín, 1996. Profesor asociado, Escuela de Sistemas, Facultad de Minas, Universidad Nacional de Colombia. Miembro de los grupos de investigación Sistemas e Informática y de Computación Aplicada, Facultad de Minas, Universidad Nacional de Colombia. Correo electrónico: cjfranco@unal.edu.co.

Nota sobre la predicción del Índice de Precios al Consumidor usando redes neuronales artificiales

RESUMEN

En este artículo se pronostica la variación porcentual del Índice de Precios al Consumidor en Colombia usando una red neuronal artificial. El modelo obtenido, una red neuronal tipo perceptrón multicapa, es capaz de capturar el ciclo estacional presente en los datos usando sólo los datos pasados, como entradas a la red neuronal, y no es necesario incluir variables ficticias que representen dicha componente estructural. La precisión de esta aproximación es mejor que la obtenida por otros modelos presentados en la literatura, incluidos aquellos con suavizaciones exponenciales, aproximaciones a modelos SARIMA y otras configuraciones de redes neuronales.

Palabras clave:

Predicción, modelos no lineales, macroeconomía, SARIMA, suavizado exponencial.

Note on Predicting Consumer Price Indexes Using Artificial Neuronal Networks

ABSTRACT

This article forecasts the percentage variation in the Colombian Consumer Price Index, using an artificial neuronal network. The model obtained, a multiple-layer perceptron neuronal network, is capable of capturing the seasonal cycle present in the data, by solely using past data as entries into the neuronal network; it is not necessary to include fictitious variables to represent said structural component. This approach is more precise than other models presented in the literature on the topic, including those with exponential smoothing, SARIMA model approximations, and other neuronal network configurations.

Key words:

Prediction, non-linear models, macroeconomics, SARIMA, exponential smoothing.

Nota sobre a previsão do Índice de Preços ao Consumidor usando redes neurais artificiais

RESUMO

Neste artigo se prognostica a variação porcentual do Índice de Preços ao Consumidor na Colômbia usando uma rede neuronal artificial. O modelo obtido, uma rede neuronal tipo perceptron multicamada, é capaz de capturar o ciclo sazonal presente nos dados usando somente os dados passados, como entradas a rede neuronal, e não é necessário incluir variáveis fictícias que representem tal componente estrutural. A precisão desta aproximação é melhor que a obtida por outros modelos apresentados na literatura, incluídos aqueles com suavizações exponenciais, aproximadas a modelos SARIMA e outras configurações de redes neuronais.

Palavras chave:

Previsão, modelos não lineares, macro-economia, SARIMA, suavizado exponencial.

Introducción

Las últimas décadas se han caracterizado por un interés creciente en la evaluación de modelos no lineales para la predicción de series económicas (Van Djck, 1999; Granger y Teräsvirta, 1993); consecuentemente, no es una casualidad que exista una gran cantidad de publicaciones sobre este tema, que abordan desde tópicos relacionados con el problema de la estimación de sus parámetros hasta la comparación entre diferentes tipos de modelos (Heravi, Osborn y Birchenhall, 2004; Swanson y White, 1997a y 1997b), así como el desarrollo de nuevos modelos (Friedman, 1991; Tseng, Tzeng, Yu y Yuan, 2001; Ghiasi, Saidane y Zimbra, 2005). Zhang, Patuwo y Hu (1998) presentan una revisión detallada sobre la predicción de series de tiempo usando redes neuronales artificiales.

En el mundo, los modelos de redes neuronales artificiales se han venido aplicando para predecir variables económicas y financieras (Swanson y White, 1997a y 1997b; Stock y Watson, 1998; Chen, Racine y Swanson, 2001), de las cuales la inflación es uno de los principales temas de interés. En la literatura se ha reportado el uso de diferentes tipos de redes neuronales que incluyen: perceptrones multicapa (Aiken, 1999; Moshiri, Cameron y Scuse, 1999; McNelis, 2002; Nakamura, 2005; Santana, 2006; Gungor y Berk, 2006; Binner et al., 2005), redes recurrentes (Moshiri, Camero y Scuse, 1999; Binner, Elger, Nilsson y Tepper, 2004 y 2006; Binner, Jones, Kendall, Tepper y Tino, 2006), redes de funciones de base radial (Moshiri, Cameron y Scuse, 1999), redes neuronales de regresión generalizada (Düzgün, 2010) y modelos hí-

bridos (Moshiri y Cameron, 2000; McAdam y McNelis, 2005).

Los estudios que utilizan redes neuronales artificiales han sido realizados para diferentes economías, como: Estados Unidos (Aiken, 1999; McNelis, 2002; Nakamura, 2005; McAdam y McNelis, 2005; Binner, Elger, Nilsson y Tepper, 2006; Binner et al., 2006; Nakamura, 2005); Colombia (Santana, 2006), Japón (McAdam y McNelis, 2005), Turquía (Gungor y Berk, 2006; Düzgün, 2010), Unión Económica Europea (Binner et al., 2005; McNelis, 2002), Reino Unido (Binner et al., 2004) y Canadá (Moshiri, Cameron y Scuse, 1999).

Un factor común en toda la literatura analizada es que se favorece el uso de los modelos de redes neuronales sobre los métodos más tradicionales de predicción, tal como la Curva de Phillips (McNelis, 2002; McAdam y McNelis, 2005), los modelos ARIMA (Rodríguez y Siado, 2003; Düzgün, 2010), los modelos de Markov (Binner, Elger, Nilsson y Tepper, 2006) o las técnicas no paramétricas (Rodríguez y Siado, 2003).

No obstante, la problemática de la predicción de la inflación va más allá del pronóstico con redes neuronales artificiales, tal como puede evidenciarse a partir del volumen de publicaciones que se han realizado en la última década. Por ejemplo, una búsqueda en Scopus usando las cadenas TITLE(*consumer price index forecast*) OR TITLE(*consumer price index predict*) OR TITLE(*inflation forecast*) OR TITLE(*inflation predict*) permitió recuperar un total de 105 artículos de los cuales 53 fueron publicados entre el 2005 y el 2010.

Nótese que la búsqueda fue restringida al título de los artículos, y no se incluyeron ni el resumen ni las palabras clave, lo que podría aumentar la cantidad de artículos reportados. En este sentido, es necesario llevar a cabo investigaciones que permitan unificar, clasificar, ordenar, organizar y sintetizar este volumen de información, con el fin de extraer nuevo conocimiento a partir de los estudios realizados.

Por otra parte, y entrando ya en materia, Santana (2006), en el caso colombiano, aplicó los perceptrones multicapa para predecir la variación porcentual del Índice de Precios al Consumidor (IPC) y concluyó que las redes neuronales artificiales permiten obtener pronósticos más precisos que los calculados con el suavizado exponencial o el modelo autorregresivo integrado de media móvil estacional (SARIMA). Existen dos puntos que hacen pensar que los resultados obtenidos por Santana (2006) pueden mejorarse:

- En dicho trabajo se utiliza como algoritmo de optimización la regla delta generalizada o *backpropagation*. Los problemas de convergencia de esta técnica de optimización son bien conocidos, por lo que se han propuesto muchas heurísticas que permitan encontrar mejores parámetros para la red neuronal (LeCun, Bottou, Orr y Muller, 1998), y también se han desarrollado nuevos algoritmos como RPROP (Riedmiller y Braun, 1993; Riedmiller 1994; Igel y Hüsken, 2000), que permiten obtener modelos que se ajustan mucho mejor a los datos.

- Santana (2006) utiliza como entradas a la red neuronal artificial, los valores pasados de la serie y un grupo de variables ficticias o *dummy* para facilitar que dicho modelo capture el patrón estacional asociado al año. En este sentido, existen tanto argumentos a favor como en contra de esta práctica; pero el principal argumento es que un modelo no lineal está en capacidad de aprender directamente el patrón estacional, y que la introducción de variables *dummy* o la eliminación del ciclo estacional cambia la dinámica de la serie.

Consecuentemente, el objetivo de esta investigación es determinar si un perceptrón multicapa puede capturar directamente la dinámica de la serie y generar pronósticos más precisos que los obtenidos por Santana (2006). La originalidad y relevancia de esta investigación está sustentada en los siguientes aspectos:

- Se demuestra que un perceptrón multicapa que usa como entradas es capaz de capturar mejor la dinámica de la serie que el modelo propuesto por Santana (2006). Ello implica que no sería necesario el uso de variables *dummy* para capturar la naturaleza cíclica de la serie.
- Se contribuye a la creación de un conjunto de series *benchmark* que exhiben características no lineales, y que pueden usarse para contrastar las bondades de otros modelos no lineales.

- Se presenta una discusión sobre la temática de predicción de series de tiempo usando redes neuronales, que está contextualizada sobre el caso de aplicación analizado. Ello permite exponer aspectos que, aun cuando son conocidos, no se usan en muchos casos de aplicación.

Para lograr el objetivo propuesto, el resto de este artículo está organizado como sigue. En la sección 1 se discute la información utilizada y la metodología empleada; entre tanto, en la sección 2 se discuten los resultados obtenidos. Finalmente, se concluye.

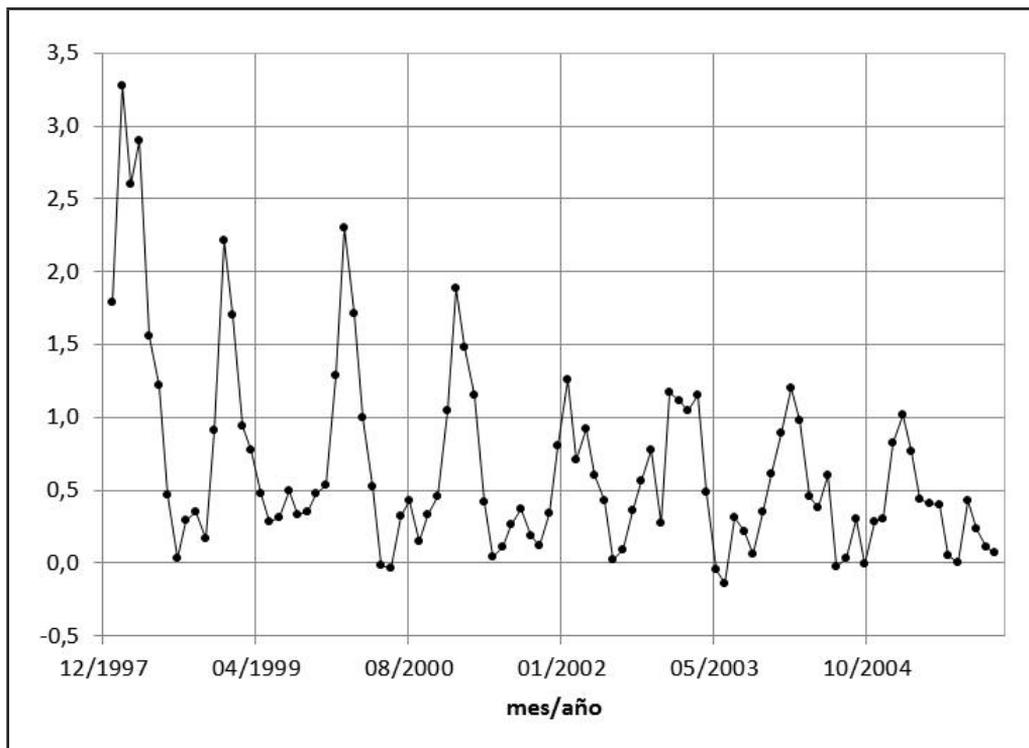
1. Datos y metodología

1.1 Información utilizada

La información utilizada en este estudio corresponde a la variación porcentual del IPC entre 1998:1 y 2005:12, publicada en la página web del Departamento Administrativo Nacional de Estadística (DANE). Su evolución en el tiempo se presenta en el Gráfico 1. La serie evidencia fuertes indicios de no estacionareidad y estacionalidad.

Gráfico 1

Variación porcentual del Índice de Precios al Consumidor



Fuente: DANE.

1.2 Investigación de Santana

Los aspectos fundamentales de la investigación realizada por Santana (2006) se resumen en los siguientes puntos:

- Los datos fueron preprocesados usando la función logaritmo natural, con el ánimo de controlar la varianza de la serie. No obstante, todos los resultados son presentados en la escala real de los datos.
- La serie se pronosticó usando la técnica de suavizado exponencial de Holts-Winters, un modelo SARIMA $(1,1,1) \times (0,1,1)_{12}$ y diferentes configuraciones de redes neuronales tipo perceptrón multicapa.
- Los parámetros de los modelos se calcularon usando la información disponible entre 1998:1 y 2005:6. Con los seis datos restantes se evaluó la capacidad de generalización de los modelos.

Un perceptrón multicapa es un modelo no paramétrico y no lineal de regresión inspirado en la estructura de los circuitos de neuronas del cerebro. Una representación pictórica de un perceptrón multicapa (con dos entradas, cuatro neuronas en la capa oculta y una neurona de salida) se presenta en el Gráfico 2. En este tipo de red neuronal, la información se propaga de la siguiente manera:

- Se presenta una señal de entrada a la red neuronal, la cual corresponde al vector $[y_{t-1}, y_{t-2}]'$.
- En cada neurona de la capa oculta se calcula la entrada neta que corresponde a la

suma ponderada de la señal de entrada por el peso α_{ih} , asociado a cada conexión, con $i = \{1,2\}$ y $h = \{1,2,3,4\}$; más un pulso unitario (+1) multiplicado por el peso asociado a cada conexión que transmite su señal a cada neurona de la capa oculta, denotado como ω_h . La señal neta de entrada a cada neurona de la capa oculta es transformada por la función de activación $G()$. Así se obtiene la señal de salida de la capa oculta. En el Gráfico 2, el bloque con el símbolo de sumatoria realiza la agregación de las entradas a cada neurona, y el bloque que le sigue representa la función $G()$.

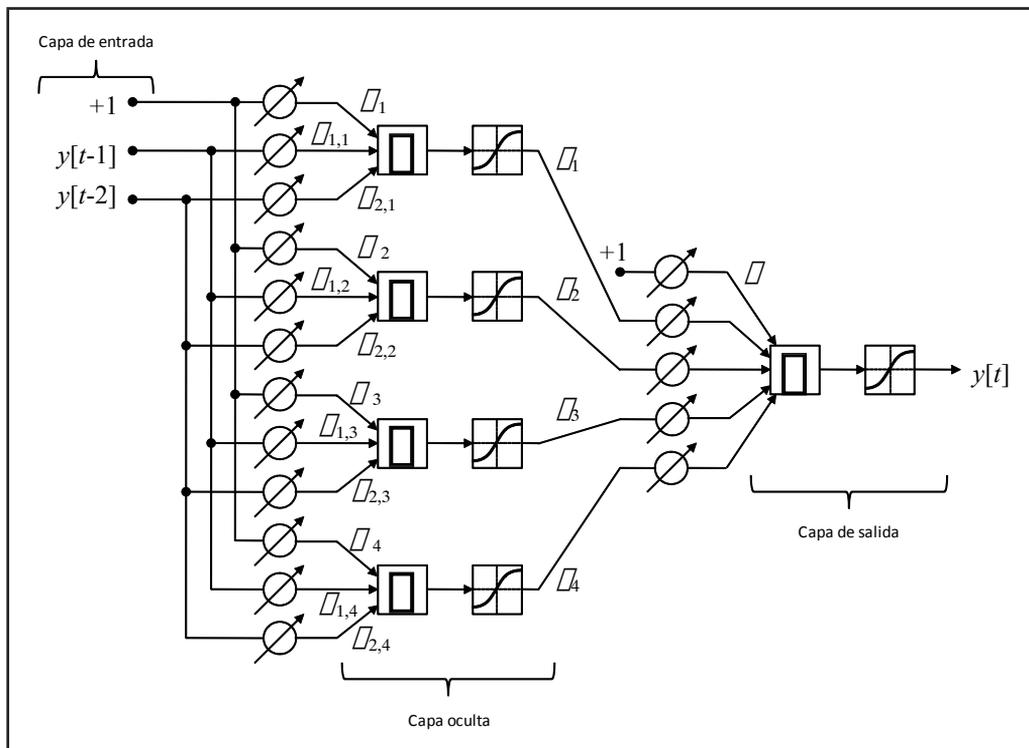
- Para la capa de salida, la propagación de la señal se realiza de forma idéntica. En este caso, la señal de salida de la capa oculta se propaga hacia la neurona de salida a través de las conexiones β_h . En el Gráfico 2, η representa la conexión entre una neurona que genera un pulso unitario y la neurona de salida. Al igual que para la capa oculta, la entrada neta es transformada usando una función no lineal y de este modo se obtiene el valor calculado por la red para y_t . La salida de la red neuronal descrita puede expresarse como:

$$y_t = G\left(n + \sum_h \left\{ \beta_h \times G\left(w_h + \sum_i \alpha_{i,h} y_{t-i} \right) \right\}\right)$$

Santana (2006) usa la sigla ANN para referirse a los perceptrones multicapa y especifica que la función de activación es $G(u) = \tanh u$. La cantidad de capas y la cantidad de neuronas por capa es especificada entre paréntesis. Así, un modelo ANN(17, 5, 1) representa un

Gráfico 2

Estructura de un perceptrón multicapa con dos entradas, cuatro neuronas en la capa oculta y una neurona en la capa de salida



Fuente: elaboración propia.

perceptrón multicapa con 17 entradas, una capa oculta con cinco neuronas y una capa de salida con una neurona; mientras que un perceptrón multicapa con 17 entradas, dos capas ocultas con tres y dos neuronas en cada capa, y una neurona en la capa de salida es representado como ANN(17, 3, 2, 1).

El modelo ANN(17, 5, 1) utiliza como entradas los rezagos 1, 2, 8, 9 y 12 más 12 variables *dummy* que representan los meses del año; tiene cinco neuronas en la capa oculta y una neurona de salida. El modelo ANN(15,

4, 1) usa los rezagos 1, 12 y 13 más 12 variables *dummy*. El modelo ANN(17, 2, 2, 1) difiere del modelo ANN(17,5,1) en que tiene dos capas ocultas con dos neuronas cada una. Para todos los modelos de redes neuronales la serie fue escalada al intervalo [-1, 1]. Se utilizó el algoritmo de regla delta generalizada para la estimación de los parámetros de cada red neuronal artificial.

En el Cuadro 1 se resumen los estadísticos de ajuste a la muestra de predicción calculados por Santana (2006). Estos fueron: el

Cuadro 1
Estadísticos obtenidos

Modelo	Entrenamiento		Predicción			
	MSE	MAD	MSE	MAD	MSE	MAD
			(1S)	(1S)	(MS)	(MS)
Resultados de Santana (2006)						
Holt-Winters aditivo	-	-	0,037	0,183	0,035	0,178
SARIMA	-	-	0,033	0,152	0,036	0,158
ANN(17, 5, 1)	-	-	0,014	0,096	0,023	0,125
ANN(15, 4, 1)	-	-	0,018	0,123	0,018	0,124
ANN(17,2,2,1)	-	-	0,022	0,120	0,019	0,108
Este estudio						
MLP ($P=13, H=1$)	0,056	0,172	0,050	0,221	0,085	0,251
MLP ($P=13, H=2$)	0,024	0,116	0,071	0,240	0,083	0,245
MLP ($P=13, H=3$)	0,013	0,078	0,075	0,250	0,083	0,244
MLP ($P=25, H=1$)	0,024	0,109	0,030	0,154	0,033	0,164
MLP ($P=25, H=2$)	0,006	0,055	0,011	0,095	0,007	0,075
MLP ($P=25, H=3$)	0,006	0,054	0,013	0,100	0,008	0,078

MSE: error cuadrático medio; MAD: desviación media absoluta; 1S: predicción un mes adelante; MS: extrapolativo hacia adelante.
 Fuente: elaboración propia.

error cuadrático medio (MSE) y la desviación media absoluta (MAD). Las columnas con la cabecera 1S indican que el estadístico se calculó usando la predicción un mes adelante; esto es, se toman los datos reales hasta el mes actual y se genera el pronóstico para el siguiente mes. La cabecera MS indica que se realizó el pronóstico extrapolativo hacia delante (o multipaso), para el cual el último dato conocido corresponde al mes 2005:6. Entre tanto, los pronósticos para los meses siguientes se obtienen de la siguiente forma: para 2005:7 se utiliza la información histórica; para 2005:8, la información histórica

hasta 2005:6 y el pronóstico para 2005:7; para 2005:9, la información hasta 2005:6 y los pronósticos para 2005:6, 2005:7 y 2005:8, y así sucesivamente.

1.3 Metodología empleada en este estudio

En este estudio se utilizó una red neuronal tipo perceptrón multicapa. A continuación se discuten los aspectos en que difiere la configuración utilizada del modelo aplicado por Santana (2006), así como las razones para ello.

1.3.1 Número de capas ocultas

Se consideró un perceptrón multicapa con una sola capa oculta, tal como el presentado en el Gráfico 2. Masters (1993 y 1995) indica que una sola capa oculta es suficiente para aproximar funciones continuas en el sentido matemático y que su primera derivada también es continua. Cuando una función es continua, pero presenta una discontinuidad en su primera derivada (por ejemplo, una función definida por intervalos) se requieren, a lo sumo, dos capas ocultas.

1.3.2 Función de activación

La función *squash* $G(u) = u(1 + |u|)^{-1}$ es utilizada para activar las neuronas de la capa oculta. En la práctica se ha encontrado que esta función facilita el entrenamiento de la red neuronal, aunque teóricamente no se ha demostrado que una función de activación es mejor que otra.

La función identidad $G(u) = u$ se usa para activar la neurona de la capa de salida; no existen motivos teóricos que justifiquen emplear obligatoriamente una función no lineal a fin de activar las neuronas de esta capa. Adicionalmente, el uso de la función identidad evita tener que escalar los datos al intervalo $[-1,+1]$ cuando se utiliza la función de activación $G(u) = \tanh u$, tal como es el caso de Santana (2006).

Un proceso inadecuado de escalamiento de los datos es una de las causas de que el pronóstico con redes neuronales sea pobre (Masters, 1993); por ejemplo, si los datos son escalados en el intervalo $[-1,+1]$ y se usa la fun-

ción $G(u) = \tanh u$, se requiere que la entrada neta u a la neurona de la capa salida varíe en el intervalo $[-3,+3]$, con lo cual el proceso de estimación de los parámetros del modelo es mucho más difícil que si se hubieran escalado los datos al intervalo $[-0,9,+0,9]$.

1.3.3 Selección de entradas a la red neuronal

Por otra parte, la selección de los rezagos que el modelo de redes neuronales artificiales debe emplear es un problema todavía no resuelto de forma completa y satisfactoria. En esta investigación se optó por utilizar la propuesta de Teräsvirta (1994), desarrollada para los modelos autorregresivos de transición suave. En este caso, se considera que el modelo no lineal usa los mismos rezagos que un modelo lineal autorregresivo, cuyo orden P se obtiene al minimizar algún criterio de información.

1.3.4 Tendencia y componente estacional

En relación con la presencia de componentes como el ciclo periódico estacional o la tendencia, existen argumentos a favor y en contra de utilizar variables *dummy* o los operadores de diferenciación simple y estacional para modelar explícitamente características como el ciclo estacional o la tendencia (Clements, Frances y Swanson, 2004). Entre los puntos más importantes se encuentran:

- Una red neuronal puede capturar la dinámica de un proceso ARMA y realizar predicciones precisas (Zhang, 2001); sin embargo, Nelson, Hill, Remus y O'Connor

(1999) y Zhang y Qi (2005) presentan resultados experimentales que favorecen el preprocesamiento de la serie usando operadores de diferenciación simple y estacional, cuando los datos provienen de un proceso SARIMA. Esta misma recomendación es dada por Masters (1993), argumentando que si una serie de tiempo presenta una tendencia y un patrón cíclico estacional, y estos no son removidos, la red neuronal se concentra en aprender dichas componentes, despreciando otros patrones determinísticos en los datos que podrían mejorar la precisión del pronóstico. Los autores desconocen si existen estudios sobre el uso de variables dummy para representar la tendencia y los ciclos cuando se usan redes neuronales artificiales.

- Los conceptos de integración (simple y estacional) y de estacionariedad en series de tiempo constituyen el origen de los operadores de diferenciación (simple y estacional) y de diversos tipos de contrastes estadísticos de integración; su derivación matemática es realizada partiendo del supuesto de que el proceso es lineal (véase Granger y Teräsvirta, 1993). Por lo tanto, ya que las redes neuronales artificiales son modelos no lineales, el supuesto de linealidad se incumple y, consecuentemente, el uso de los contrastes de integración y de los operadores de diferenciación pierde su sustento teórico; queda sólo como una recomendación empírica.
- Finalmente, muchos modelos no lineales de series de tiempo pueden generar se-

ries de tiempo con patrones estacionales cíclicos y tendencias (Kantz y Schreiber, 1999), de tal forma que no se requeriría usar los operadores de diferenciación; esto equivale a afirmar, por ejemplo, que una red neuronal artificial podría capturar dicho comportamiento directamente a partir de los datos mismos. Más aún, se argumenta que el uso de la diferenciación afecta las propiedades de la serie, lo que causaría que el modelo no capture la dinámica real de los datos.

1.3.5 Algoritmo de optimización

Santana (2006) utiliza el algoritmo de *backpropagation* para estimar los parámetros de la red neuronal. En este algoritmo, el error cuadrático para la observación en el tiempo t es:

$$e_t = 0,5(y_t - f_t)^2$$

Siendo f_t el pronóstico de la red neuronal para el período t . Sea $\Omega = [\alpha_{ih}, \omega_h, \beta_{ih}, \eta]$, con $i = 1, \dots, P$ y $h = 1, \dots, H$, el vector de parámetros de la red neuronal. En el algoritmo de *backpropagation*, el vector de parámetros en la k -ésima iteración es ajustado como:

$$\Omega(k) = \Omega(k-1) - k \frac{\partial e_t}{\partial \Omega} + \lambda \times \Delta \Omega(k-1)$$

El desempeño final de la red neuronal depende de los valores de los parámetros k y λ , especificados por el usuario y controlan el desempeño del algoritmo, y del valor del vector inicial $\Omega(0)$; esto se debe a la complejidad de la superficie de la función de error $E = \sum_t e_t$, plagada de puntos de mínima local

y planicies que dificultan el proceso de optimización. Esto obliga a que la optimización se realice para muchas combinaciones de los valores con los parámetros k y λ y de pesos aleatorios iniciales, con el fin de escapar de los mínimos locales. Estos problemas de convergencia son bien conocidos desde hace más de una década y se han propuesto muchas heurísticas que permitan encontrar mejores parámetros para la red neuronal (véase, por ejemplo, el trabajo de LeCun et al., 1998).

A continuación se describen las consideraciones sobre el algoritmo de optimización utilizado en esta investigación. Teniendo en cuenta que la función identidad $G(u) = u$ es usada para la activación de la neurona de salida, la propagación de la señal de entrada a través de la red neuronal puede realizarse por capas. Así, la salida de h -ésima neurona de la capa oculta O_h es:

$$O_h = G\left(\omega_h + \sum_i \alpha_{i,h} y_{i-1}\right)$$

Y la salida de la red puede calcularse en términos de dicha salida:

$$y_t = n + \sum_h \beta_h \times O_h$$

La ventaja de esta representación es que si los pesos de la capa de entrada a la capa oculta son conocidos (α_{ih}, ω_h), los valores de η y β_h pueden calcularse por mínimos cuadrados. El nuevo vector de parámetros se reduce a $\Omega = [\alpha_{ih}, \omega_h]$, que reduce la complejidad del problema de optimización. Esta idea fue planteada por Masters (1993) en el contex-

to de las redes neuronales, aunque es de uso común en la optimización de otros modelos econométricos no lineales (véase, por ejemplo, a Teräsvirta, 1994).

No obstante, el proceso puede mejorarse. Si $O_h(t)$ representa la salida de la h -ésima neurona de la capa oculta para el instante t , la serie de tiempo tiene T observaciones y se considera la arquitectura del Gráfico 2, entonces, la salida de la red neuronal puede escribirse como:

$$\eta + \beta_1 O_1(3) + \beta_2 O_2(3) + \beta_3 O_3(3) + \beta_4 O_4(3) = y_3$$

$$\eta + \beta_1 O_1(4) + \beta_2 O_2(4) + \beta_3 O_3(4) + \beta_4 O_4(4) = y_4$$

:

$$\eta + \beta_1 O_1(T) + \beta_2 O_2(T) + \beta_3 O_3(T) + \beta_4 O_4(T) = y_T$$

Equivalente a un sistema lineal sobre-determinado de la forma $\mathbf{Ax} = \mathbf{B}$, donde $\mathbf{x} = [\eta, \beta_1, \beta_2, \beta_3, \beta_4]'$, $\mathbf{B} = [y_3, \dots, y_T]'$ y \mathbf{A} es una matriz cuyas filas son de la forma $[1, O_1(t), O_2(t), O_3(t), O_4(t)]$ para $t = 3, \dots, T$. Así, el vector de parámetros puede ser calculado por mínimos cuadrados ordinarios como $\mathbf{x} = (\mathbf{A}'\mathbf{A})^{-1}\mathbf{A}'\mathbf{B}$. No obstante, las salidas de las neuronas de la capa oculta pueden ser altamente colineales, por lo que se prefiere el uso de la técnica *ridge regression*, que es el mecanismo utilizado en esta investigación para calcular los pesos de las conexiones entrantes a la neurona de salida. En este caso, el algoritmo de optimización no lineal (*backpropagation* o cualquier otro) se emplea para ajustar los valores de los parámetros $\Omega = [\alpha_{ih}, \omega_h]$, mientras que los valores de $[\eta, \beta_h]$, para $h = 1, \dots, H$ son calculados usan-

do mínimos cuadrados ordinarios o *ridge regression*.

Otro punto que se debe considerar es el uso de algoritmos constructivos en la optimización de la red neuronal, como un mecanismo para evadir mínimos locales; en esta clase de algoritmos, se empieza optimizando el modelo más simple, que corresponde a un perceptrón multicapa con una sola neurona en la capa oculta. Luego se adiciona al modelo optimizado una segunda neurona cuyas conexiones (entrantes y salientes) son iniciadas usando pesos aleatorios. El nuevo modelo es optimizado y el proceso continúa hasta que se cumpla algún criterio de parada. Este concepto fue desarrollado originalmente por Fahlman y Lebiere (1990), pero hoy en día existen muchos algoritmos basados en este principio (véase, por ejemplo, los trabajos de Bahi, Contassot-Vivier y Sauget, 2009; Lehtokangas, 1999).

En esta investigación se utilizó el algoritmo constructivo básico ya descrito. Para evitar los problemas asociados con el uso de la regla delta generalizada, se optó por el método de Broyden-Fletcher-Goldfarb-Shanno, que es una técnica de gradiente de segundo orden, donde la modificación de los parámetros de la red neuronal se calcula directamente por el algoritmo. La descripción de la racionalidad que inspira esta metodología cae por fuera del alcance de este artículo, por lo que se sugiere la consulta del trabajo de Fletcher (1987).

1.4 Metodologías alternas posibles

Esta investigación podría extenderse fácilmente al considerar otros tipos de modelos

de redes neuronales más precisos que un perceptrón multicapa en la predicción de series de tiempo. Esto se debe a que en la realidad se desconoce el mecanismo matemático exacto que genera la serie de tiempo, y este es aproximado por la representación matemática equivalente de la red neuronal. Así, el mejor modelo es aquel más cercano al mecanismo matemático que genera los datos. Entre los modelos alternativos se incluyen:

- Las máquinas de vectores de soporte, que son un tipo de red neuronal diseñada específicamente para resolver problemas de clasificación. Detalles sobre su implementación y su aplicación específica a la predicción de series de tiempo pueden consultarse en los trabajos de Velásquez, Franco y Olaya (2010) y Velásquez, Olaya y Franco (2010).
- La red neuronal DAN2 de Ghiassi, Saidane y Zimbra (2005), que representa la no linealidad de los datos como una sumatoria de funciones de seno y coseno, de una forma similar a la transformada de Fourier.
- La red neuronal adaptativa de Wong, Xia y Chu (2010), que está basada en una métrica adaptativa de las entradas inspirada en la técnica del “vecino más cercano”, y la mezcla de su salida usando una función predefinida.

Adicionalmente, sería necesario considerar la combinación de modelos usando máquinas de comité. Entre las metodologías de combinación se encuentran los ensambles, el *boosting*, la mezcla de expertos y la mezcla

jerárquica de expertos; una introducción es presentada por Haykin (1999).

2. Resultados obtenidos y discusión

A continuación se describen los resultados obtenidos al pronosticar la variación porcentual del IPC utilizando la metodología propuesta; igualmente, se comparan los resultados obtenidos con los reportados por Santana (2006).

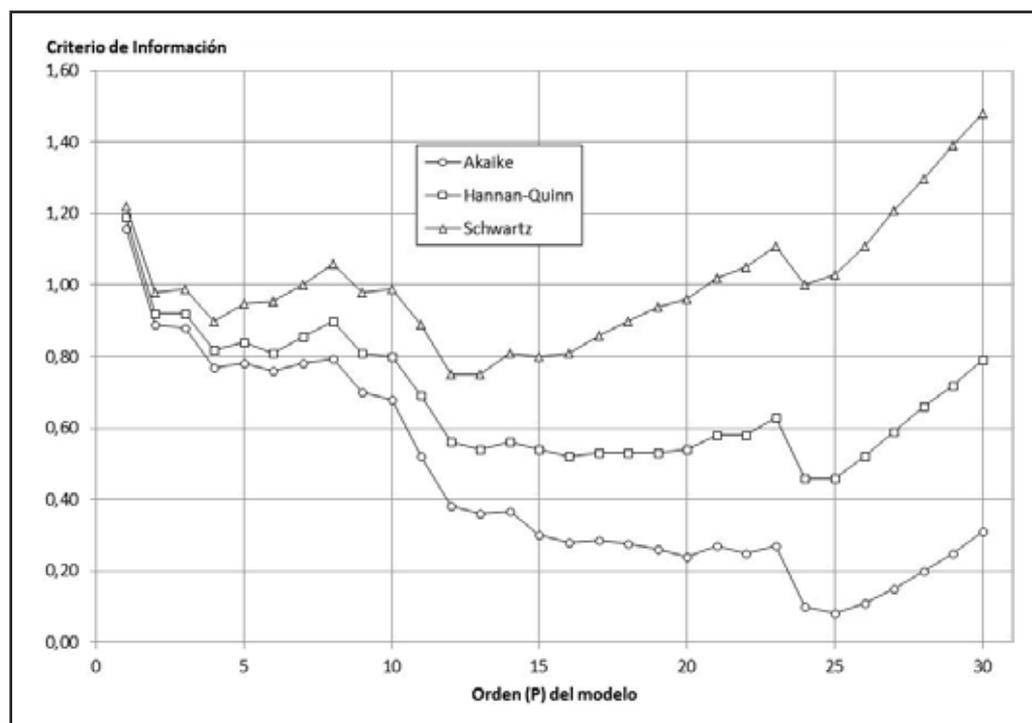
En primer lugar, se estimaron los valores de los criterios de información de Akaike

(1973), Hannan-Quinn (1979) y Schwartz (1978), con el fin de determinar el orden óptimo P de un modelo autorregresivo que capture la dinámica de la serie estudiada. Los valores estimados se presentan en el Gráfico 3. Los criterios de Akaike (1973) y Hannan-Quinn (1979) presentan un mínimo en $P = 13$, mientras que $P = 25$, según el criterio de Schwartz (1978). Así, se consideraron modelos de redes neuronales artificiales que utilizan como entradas los rezagos 1-13 o 1-25.

La estimación de los parámetros se inicia para una red neuronal artificial con una neurona

Gráfico 3

Valor de los criterios de información considerados para diferentes órdenes de un modelo autorregresivo



Fuente: elaboración propia.

en la capa oculta. Cuando la red neuronal ha sido optimizada, se agrega una nueva neurona a la capa oculta y se asignan valores aleatorios a las nuevas conexiones. El proceso se repite para un máximo de tres neuronas en la capa oculta. Ya que el proceso no garantiza encontrar la mejor red neuronal, el proceso completo se repitió 50 veces y se escogió el modelo con la mejor generalización de entre todas las redes neuronales estimadas para 1, 2 y 3 neuronas en la capa oculta.

Los estadísticos de ajuste a las muestras de entrenamiento y pronóstico para los mejores modelos obtenidos son presentados en el Cuadro 1. P indica la cantidad de rezagos usados en el modelo. H es la cantidad de neuronas en la capa oculta. Cuando se considera un máximo de 13 rezagos, todas las redes neuronales artificiales presentan un desempeño muy inferior a los modelos desarrollados por Santana (2006). Cuando se tienen en cuenta un máximo de 25 rezagos, se encontró que el número óptimo de neuronas en la capa oculta es dos. Este es notado como MLP ($P = 25, H = 2$) en el Cuadro 1.

El modelo encontrado es mejor que los demás modelos estimados en esta investigación, tanto en la predicción un mes adelante como en la predicción varios meses adelante. Respecto al modelo ANN(17, 5, 1), el MSE se reduce al 79% y al 30% para las predicciones un mes y varios meses hacia adelante; para el MAD las reducciones correspondientes son del 99% y del 60%. Si se considera el modelo ANN(15, 4, 1), que presenta el menor MSE para la predicción varios pasos adelante, se da una reducción del 60%. El modelo ANN(15, 4, 1) tiene un total de

$15 \cdot 4 + 4 + 4 + 1 = 69$ parámetros, mientras que el modelo MLP ($P = 25, H = 2$) tiene $25 \cdot 2 + 2 + 2 + 1 = 55$ parámetros. Así, el modelo seleccionado en esta investigación es más parsimonioso que el modelo ANN(15, 4, 1) de Santana (2006). La predicción un mes adelante para la muestra de calibración y de seis meses adelante para la muestra de predicción es presentada en el Gráfico 4.

Conclusiones

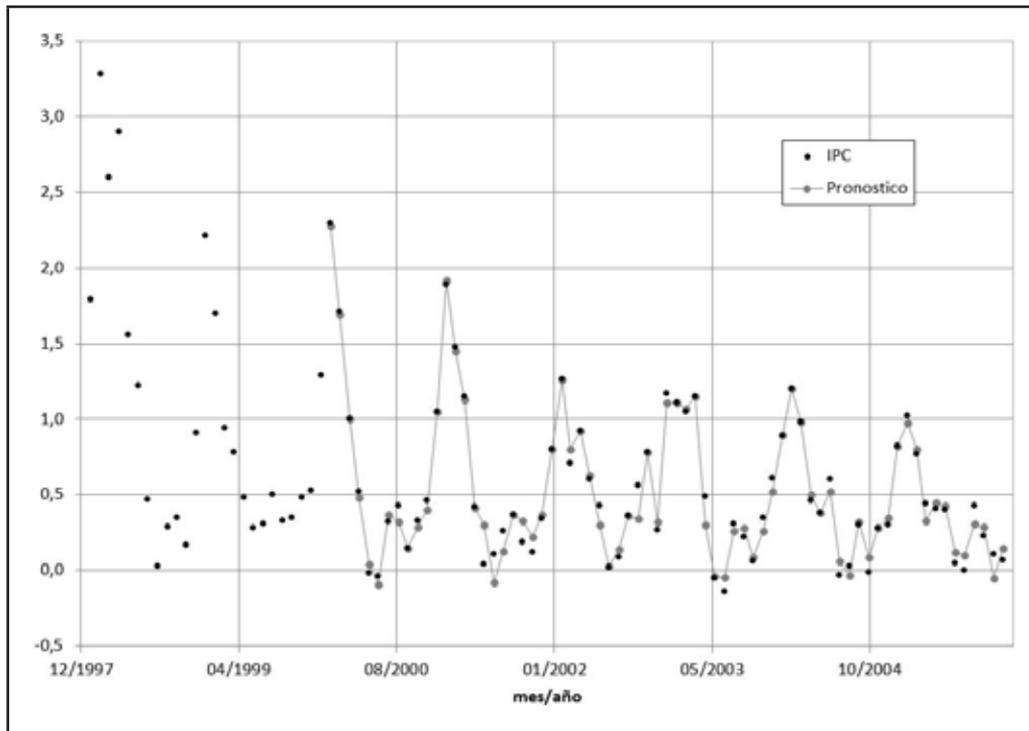
En este artículo se desarrolla una red neuronal tipo perceptrón multicapa para la predicción de la variación mensual del IPC en Colombia. La red considerada utiliza únicamente como entradas los valores pasados de la serie, sin que se represente explícitamente la componente estacional existente en los datos. Los resultados muestran que la red neuronal propuesta es capaz de pronosticar con mayor precisión los datos por fuera de la muestra de calibración, tanto si se considera únicamente la predicción para el siguiente mes, así como la predicción varios meses adelante, como otros modelos propuestos en la literatura.

Lista de referencias

- Aiken, M. (1999). Using a neural network to forecast inflation. *Industrial Management and Data Systems*, 99 (7), 296-301.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrov and F. Csaki (Eds.), 2nd International Symposium on Information Theory (pp. 267-281). Budapest: Akademia Kiado.

Gráfico 4

Predicción obtenida usando el modelo MLP (P = 25, H = 2)



Fuente: elaboración propia.

Anders, U. and Korn, O. (1999). Model selection in neural networks. *Neural Networks*, 12, 309-323.

Bahi, J. M.; Contassot-Vivier, S. and Sauget, M. (2009). An incremental learning algorithm for function approximation. *Advances in Engineering Software*, 40 (8), 725-730.

Binner, J. M.; Bissoondeal, R. K.; Elger, T.; Gazely, A. M. and Mullineux, A. W. (2005). A comparison of linear forecasting models and neural networks: An application to Euro inflation and Euro Divisia. *Applied Economics*, 37 (6), 665-680.

Binner, J. M.; Elger, T.; Nilsson, B. and Tepper, J. A. (2004). Tools for non-linear time series forecasting in economics: an empirical comparison of regime switching vector autoregressive models and recurrent neural networks. *Advances in Econometrics*, 19, 71-91.

Binner, J. M.; Elger, C. T.; Nilsson, B. and Tepper, J. A. (2006). Predictable non-linearities in U.S. inflation. *Economics Letters*, 93 (3), 323-328.

Binner, J. M.; Jones, B.; Kendall, G.; Tepper, J. and Tino, P. (2006). *Does money matter?: An arti-*

- ficial intelligence approach*. Documento procedente de 9th Joint Conference on Information Sciences, JCIS 2006 CIEF-129.
- Chen, X.; Racine, J. and Swanson, N. (2001). Semiparametric ARX neural network models with an application to forecasting inflation. *IEEE Transactions on Neural Networks*, 12, 674-683.
- Clements, M. P.; Frances, P. H. and Swanson, N. R. (2004). Forecasting economic and financial time-series with non-linear models. *International Journal of Forecasting*, 20, 168-183.
- Düzgün, R. (2010). Generalized regression neural networks for inflation forecasting. *International Research Journal of Finance and Economics*, 51, 59-70.
- Fahlman S. E. and Lebiere C. (1990). The Cascade-Correlation learning architecture. *Advances in Neural Information Processing Systems*. 2, 524-532.
- Fletcher, R. (1987). *Practical methods of optimization*. New York: Wiley-Interscience.
- Friedman, J. (1991). Multivariate adaptive regression splines (with discussion). *Annals of Statistics*, 19, 1-141.
- Ghiassi, M.; Saidane, H. and Zimbra, D. K. (2005). A dynamic artificial neural network model for forecasting time series events. *International Journal of Forecasting*, 21, 341-362.
- Granger, C. and Teräsvirta, T. (1993). *Modeling non-linear economic relationships*. Oxford: Oxford University Press.
- Gungor, C. and Berk, A. (2006). Money supply and inflation relationship in the Turkish Economy. *Journal of Applied Sciences*, 6 (9), 2083-2087.
- Haykin, S. (1999). *Neural networks: a comprehensive foundation*. New York: Pearson.
- Hannan, E. and Quinn, B. (1979). The determination of the order of an autoregression. *Journal of Royal Statistical Society, Series B*, 41, 190-195.
- Heravi, S.; Osborn, D. and Birchenhall, C. (2004). Linear versus neural network forecasts for European industrial production series. *International Journal of Forecasting*, 20, 435-446.
- Igel, C. and Hüsken, M. (2000). *Improving the RPROP learning algorithm*. Documento procedente de Second International Symposium on Neural Computation, NC2000, ICSC Academic Press.
- Kaasra, I. and Boyd, M. (1996). Designing a neural network for forecasting financial and economic series. *Neurocomputing*, 10, 215-236.
- Kantz, H. and Schreiber, T. (1999). *Non-linear time series analysis*. Cambridge, UK: Cambridge University Press.
- LeCun, Y.; Bottou, L.; Orr, G. B. and Muller, K.-R. (1998). Efficient backprop. En *Neural Networks: Tricks of the Trade* (pp. 5-50). s. l.: Springer Lecture Notes in Computer Sciences 1524.
- Lehtokangas, M. (1999). Modelling with constructive backpropagation. *Neural Networks*, 12 (4-5), 707-716.

- Masters, T. (1993). *Practical neural network recipes in C++*. New York: Academic Press.
- (1995). *Neural, novel and hybrid algorithms for time series prediction*. New York: John Wiley and Sons.
- Nakamura, E. (2005). Inflation forecasting using a neural network. *Economics Letters*, 86 (3), 373-378.
- McAdam, P. and McNelis, P. (2005). Forecasting inflation with thick models and neural networks. *Economic Modelling*, 22 (5), 848-867.
- McNelis, P. D. (2002). Nonlinear Phillips curves in the Euro Area and USA?: Evidence from linear and neural network models. *Proceedings of the International Joint Conference on Neural Networks*, 3, 2521-2526.
- Moshiri, S. and Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of Forecasting*, 19 (3), 201-217.
- Moshiri, S.; Cameron, N. E. and Scuse, D. (1999). Static, dynamic and hybrid neural networks in forecasting inflation. *Computational Economics*, 14 (3), 219-235.
- Nelson, M.; Hill, T.; Remus, W. and O'Connor, M. (1999). Time series forecasting using neural networks: should the data be deseasonalized first? *Journal of Forecasting*, 18, 359-367.
- Rodríguez, N. y Siado, P. (2003). Un pronóstico no paramétrico de la inflación colombiana. *Revista Colombiana de Estadística*, 26 (2), 89-128.
- Riedmiller, M. and Braun, H. (1993). A direct adaptive method for faster backpropagation learning: The RPROP algorithm. *Proceedings of the IEEE International Conference on Neural Networks*, 86-591.
- Riedmiller, M. (1994). Advanced supervised learning in multi-layer perceptrons: from backpropagation to adaptive learning algorithms. *Computer Standards and Interfaces*, 16, 265-278.
- Santana, J. C. (2006). Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana. *Revista Colombiana de Estadística*, 29 (1), 77-92.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Stock, J. H. and Watson, M. W. (1998). A Comparison of Linear and Nonlinear Univariate Models for Forecasting Macroeconomic Time Series. *JBES*, 14 (1), 11-30.
- (1999). Forecasting inflation. *Journal of Monetary Economics*, 44, 293-335.
- Swanson, N. and White, H. (1997a). Forecasting economic time series using adaptive versus nonadaptive and linear versus non-linear econometric models. *International Journal of Forecasting*, 13, 439-461.
- (1997b). A model selection approach to real time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, 39, 540-550.
- Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive

- models. *Journal of the American Statistical Association*, 89, 208-218.
- Tseng, F. M.; Tzeng, G. H.; Yu, H. C. and Yuan, B. J. C. (2001). Fuzzy ARIMA model for forecasting the foreign exchange market. *Fuzzy Sets and Systems*, 118, 9-19.
- Van Djck, D. (1999). *Smooth transition models: extensions and outlier robust inference*. Tesis de PhD no publicada, Erasmus University, Rotterdam.
- Velásquez, J. D.; Franco, C. J. y Olaya, Y. (2010). Predicción de los precios promedios mensuales de contratos despachados en la Bolsa de Energía de Colombia usando máquinas de vectores de soporte. *Cuadernos de Administración*, 23 (40), 321-337.
- Velásquez, J. D.; Olaya, Y. y Franco, C. J. (2010). Predicción de series temporales usando máquinas de vectores de soporte. *Ingeniare. Revista Chilena de Ingeniería*, 18 (1), 64-75.
- Weng, D. (2010). The consumer price index forecast based on ARIMA model. *Proceedings of WASE International Conference on Information Engineering*, 5571115, 307-310.
- Wong, W. K.; Xia, M. and Chu, W.C. (2010). Adaptive neural network model for time-series forecasting. *European Journal of Operational Research*, 207 (2), 807-816.
- Zhang, G. P. (2001). An investigation of neural networks for linear time-series forecasting. *Computers & Operations Research*, 28 (12), 1183-1202.
- ; Patuwo, B. and Hu, M. (1998). Forecasting with artificial neural networks: the state of the art. *International Journal of Forecasting*, 14, 35-62.
- Zhang, G. P. and Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160 (2), 501-514.