

# Estudio empírico sobre la capacidad predictiva de las redes neuronales en el pronóstico de la inflación colombiana: una metodología alternativa

Juan Camilo Santana Contreras\*, Álvaro Andrés Camaro\*\*,  
Arnoldo Casas Henao\*\*\* & Édgar Jiménez Méndez\*\*\*\*

## resumen

Evaluar la capacidad de predicción de las metodologías de redes neuronales SARIMA de Box-Jenkins, suavizamiento exponencial y modelos de regresión con coeficientes variando en el tiempo es de interés en el pronóstico de la inflación colombiana, cuyo conocimiento es primordial para el diseño de políticas económicas y programas estratégicos de inversión tanto en el sector público como en el privado. Una aplicación que pronostica valores futuros de la serie de inflación colombiana permite mostrar que las redes neuronales con ayuda de componentes no observables pueden ser más precisas si se comparan con las metodologías tradicionales de Box-Jenkins, el suavizamiento exponencial y los mínimos cuadrados flexibles. Además, los resultados revelan que combinaciones de pronósticos haciendo uso de las redes neuronales tienden a proporcionar mejores predicciones.

**Palabras clave:** perceptron multicapas, modelos SARIMA, suavizamiento exponencial, mínimos cuadrados flexibles, combinación de pronósticos, componentes no observables.

## abstract

An empirical study of neuronal networks' predictive ability in forecasting colombian inflation: an alternative methodology

Evaluating the prediction ability of neuronal networks (Box-Jenkins' SARIMA, exponential smoothing and varying coefficient regression models) is interesting in forecasting Colombian inflation. Such knowledge is fundamental in designing economic policy and strategic investment programmes in both the public and private sectors. An application forecasting future values from a series of Colombian inflation shows that neuronal networks supported, by non-observable components, could give more precise forecasting compared to traditional Box-Jenkins', exponential smoothing and flexible square minimum methodologies. The results also revealed that forecasting combinations making use of neuronal networks tended to provide better predictions.

**Key words:** multi-layer perception, SARIMA models, exponential smoothing, flexible square minimums, forecasting combination, non-observable components.

## résumé

Étude empirique sur la capacité de prédiction des réseaux neuronaux dans le pronostic de l'inflation colombienne : une méthodologie alternative.

Évaluer la capacité de prédiction des méthodologies de réseaux neuronaux, SARIMA de Box-Jenkins, d'atténuation exponentielle et des modèles de régression avec des coefficients variant dans le temps est intéressant pour le pronostic de l'inflation colombienne, dont la connaissance est primordiale dans l'élaboration de politiques économiques et de programmes stratégiques d'investissement tant dans le secteur public que dans le privé. Une application qui pronostique des données futures de la série d'inflation colombienne permet de montrer que les réseaux neuronaux avec l'aide d'éléments non observables, peut-être plus précise en comparaison avec les méthodologies traditionnelles de Box-Jenkins, l'atténuation exponentielle et les cadres flexibles minimums. De plus, les résultats révèlent que les combinaisons de pronostics utilisant des réseaux neuronaux, tendent à fournir de meilleures prédictions.

**Mots clés :** «perceptron» multicouches, modèles SARIMA, atténuation exponentielle, cadres flexibles minimums, combinaisons de pronostics, composantes non observables.

## resumo

Estudo empírico sobre a capacidade de previsão das redes de neurônios no prognóstico da inflação colombiana: uma metodologia alternativa.

Avaliar a capacidade de predição das metodologias de redes de neurônios, SARIMA de Box-Jenkins, suavização exponencial e modelos de regressão com coeficientes variando no tempo é interessante no prognóstico da inflação colombiana, cujo conhecimento é primordial para o desenho de políticas econômicas e programas estratégicos de investimentos tanto no setor público como no privado. Uma aplicação prognosticando valores futuros da série de inflação colombiana nos permite visualizar que as redes de neurônios com ajuda de componentes não observáveis, podem ser mais precisas comparadas com as metodologias tradicionais de Box-Jenkins, a suavização exponencial e os mínimos quadrados flexíveis. Além disso, os resultados revelam que combinações de prognósticos utilizando-se as redes neurônios, têm uma tendência a proporcionar melhores predições.

**Palavras Chave:** perceptron multicamadas, modelos SARIMA, suavização exponencial, mínimos quadrados flexíveis, combinação de prognósticos, componentes não observáveis.

Clasificación JEL: C45, C61.

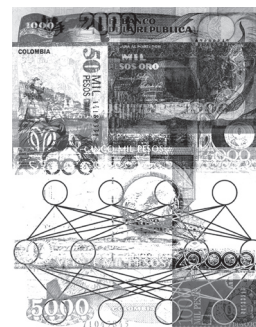
Recibido: 23/03/2006. Aprobado: 18/07/2006.

\* Máster en Estadística de la Universidad Federal de Pernambuco U.F.P.E.  
- Brasil. Estadístico de la Universidad Nacional de Colombia, sede Bogotá.  
Correo electrónico:  
csantana@promotorabursatil.com

\*\* Candidato a máster en Administración de la Universidad Nacional de Colombia, sede Bogotá. Economista de la Universidad del Rosario.  
Correo electrónico:  
acamaro@promotorabursatil.com

\*\*\* Candidato a máster en Economía de la Universidad de los Andes. Economista de la Universidad de los Andes.  
Correo electrónico:  
acasas@promotorabursatil.com

\*\*\*\* Especialista en finanzas, Universidad de los Andes. Contador Público de la Universidad Nacional de Colombia, sede Bogotá.  
Correo electrónico:  
ejimenez@promotorabursatil.com



Santana, J.C., Camaro, Á.A., Casas, A. & Jiménez, É. (2006). *Estudio empírico sobre la capacidad predictiva de las redes neuronales en el pronóstico de la inflación colombiana: una metodología alternativa. Innovar 16(28), 187-198.*

## 1. Introducción

Una metodología alternativa utilizando redes neuronales será considerada en el presente documento con el objetivo de elaborar pronósticos sobre la serie de inflación colombiana, el cual suele medirse como las variaciones sobre el índice de precios al consumidor, IPC. El desarrollo de metodologías que permitan pronosticar y comprender el comportamiento de la inflación son de interés para muchos sectores de la población y la economía. De esta forma, su cuantificación resulta necesaria para la toma de decisiones dentro de los contextos financiero, económico y social.

Uno de los fenómenos económicos que más ha centrado la atención de los economistas en general y los diseñadores de políticas de las diferentes naciones es la inflación. Los efectos negativos que produce un aumento desmesurado en el nivel general de precios sobre la asignación de los factores de producción, el ingreso de los agentes, el valor de las inversiones y la estabilidad macroeconómica de un país en general, han generado un constante afán por mejorar la manera como se estiman las trayectorias futuras de las cotizaciones de bienes y servicios de una economía, a fin de poder adoptar decisiones trascendentales en cuanto a instrumentos de política monetaria como tasas de interés de intervención, encajes bancarios u operaciones de diversa índole que aumenten o contraigan la disponibilidad de recursos en la economía.

Para los mercados de capitales, el contar con políticas de los bancos centrales enfocadas a mantener bajo control la inflación es fundamental, en la medida que se aminora la incertidumbre de los agentes, las tasas de interés pueden reducirse y, por ende, se estimula la actividad económica al disminuirse el costo de oportunidad del consumo e incentivar proyectos de inversión cuya tasa de retorno esperada supera con creces los rendimientos que con el tiempo reconocen bonos gubernamentales o corporativos.

En Colombia la estimación de la inflación entre los equipos de investigación de instituciones financieras, sociedades comisionistas de bolsa y agentes interesados en proyectar la trayectoria de activos financieros como los TES, las acciones o los títulos de empresas del sector privado, ha cobrado especial relevancia en los últimos años. Se observa con beneplácito cómo los instrumentos utilizados y explorados para aumentar el nivel de acierto vienen ganando terreno tanto en número como en sofisticación. Por esto los autores consideran pertinente evaluar metodologías alternativas dentro de una de las herramientas que a la luz de los estudios existentes parece tener el mejor récord de desempeño en cuanto a la predicción de la inflación se refiere; éstas son las redes neuronales.

Investigaciones y aplicaciones de las redes neuronales a nivel económico, en el ámbito colombiano, han sido realizadas por Sanín (2001), Misas, López y Borrero (2002), Misas, López, Arango y Hernández (2003) y recientemente Jalil y Misas (2006) y Aristizábal y Misas (2006), evidenciando las bondades de las redes neuronales en la predicción, comparado con otras metodologías tradicionales. Motivados por las características de las redes neuronales, nuestra idea principal es describir una metodología alternativa, referente al mecanismo de modelamiento de las redes neuronales, diferente a la expuesta en artículos preliminares en el contexto colombiano, que permita unirse al abanico de técnicas ya existentes en esta línea.

Métodos alternativos a las redes neuronales, como los mínimos cuadrados flexibles (FLS) o de parámetros variando en el tiempo –ver Kalaba y Tesfatsion (1996)–, han ganado importancia en los últimos años debido a su capacidad para captar cambios estructurales, en comparación con técnicas tradicionales como los mínimos cuadrados ordinarios (OLS), ampliamente utilizados en el área econométrica. Misas y Melo (2004) hacen una comparación entre estas dos técnicas, aplicada a la inflación colombiana, destacando las ventajas del FLS en controlar *shocks* o variaciones que puedan generar un efecto negativo sobre los pronósticos.

Hornik, Stinchcombe y White (1989), Cybenko (1989), entre otros, han demostrado que las redes neuronales son aproximadores universales, y el perceptron multicapas una de las arquitecturas más utilizadas en la solución de problemas debido a su fácil uso y aplicabilidad; ver, Narendra y Parthasaranty (1990), Wiegand *et al.* (1990), Cohen *et al.* (1993) y Santana (2005) para aplicaciones con redes neuronales. Dentro del área estadística, las redes neuronales son consideradas como métodos no paramétricos no lineales multivariados; ver Zhang, Patuwo y Hu (1998).

Para evaluar el desempeño de las redes neuronales en el pronóstico, serán utilizadas las metodologías SARIMA de Box-Jenkins, los mínimos cuadrados flexibles y el suavizamiento exponencial, al igual que la combinación de pronósticos. Será supuesto que el lector tiene un conocimiento básico con relación a cada tópico. No obstante, el desarrollo metodológico de la teoría de Box-Jenkins puede ser estudiada más ampliamente en Box y Jenkins (1976), Box, Jenkins y Reinsel (1994) y Morettin y Tolo (2004) entre otros; para más detalles sobre los algoritmos de suavizamiento exponencial, ver Montgomery y Johnson (1976) y Morettin y Tolo (2004); para el desarrollo metodológico sobre los mínimos cuadrados flexibles se puede consultar Kalaba y Tesfatsion (1989, 1996); con relación a la combinación de pronósticos, Barnard (1963) y Hendry y Clements (2004) hacen desarrollos completos en

este sentido. Igualmente, el conocimiento respecto a la obtención de componentes no observables, como tendencia y estacionalidad, a través de la metodología SARIMA de Box-Jenkins, serán importantes en el desarrollo de este artículo. El lector podrá referirse a Maravall y Kaiser (2000) para una discusión más amplia sobre este tema.

El presente artículo se encuentra organizado de la siguiente forma: en la sección 2, se elabora una descripción histórica referente a las investigaciones sobre redes neuronales en Colombia; en la sección 3, se describen las principales características de la red neuronal perceptron multicapas y las condiciones propuestas para el modelamiento; en la sección 4, se calculan pronósticos para la inflación, y se realizan comparaciones con las metodologías clásicas y la combinación de pronósticos. Por último, en la sección 5, se sintetizan las conclusiones.

## 2. Revisión histórica sobre la aplicación de redes neuronales en investigaciones colombianas

El desarrollo investigativo sobre redes neuronales a nivel mundial, como modelos computacionales que emulan el comportamiento del cerebro humano, comenzó desde los años sesenta del siglo pasado, con abundantes aplicaciones en diferentes áreas de la ciencia. Su desenvolvimiento teórico y práctico ha estado sujeto principalmente a los avances computacionales, los cuales han permitido comprender su capacidad funcional para captar relaciones no lineales.

En Colombia, los primeros trabajos públicos conocidos sobre aplicaciones de redes neuronales en el ámbito económico fueron los realizados por el Banco de la República, citando a Misas *et al.* (2002, 2003)<sup>1</sup>, Jalil y Misas (2006) y Aristizábal y Misas (2006). Adicionalmente, una aplicación en el contexto de las ciencias sociales fue dada por Sanín (2001) (Departamento de Planeación Nacional), en donde las redes neuronales son utilizadas para un análisis en el que son clasificados candidatos políticos en categorías que determinan su preferencia política (liberal o conservador, político nuevo o en transición partidista), dadas ciertas descripciones sobre su carrera política, además de su perfil

educativo y social. Aunque considera que en ciertos casos los resultados obtenidos en la clasificación no son completamente satisfactorios, el autor no hace una descripción sobre el proceso de modelamiento con las redes neuronales que permita analizar en qué circunstancias la red neuronal clasifica erróneamente.

El trabajo de Misas *et al.* (2002) se centra en el análisis econométrico de las relaciones entre la inflación y el dinero en la predicción del nivel de precios de la economía, sustentando el uso de las redes neuronales en el conocimiento sobre las relaciones asimétricas latentes entre la política monetaria y los precios. La arquitectura de red utilizada contiene una única capa oculta, y utiliza datos trimestrales del IPC, M1 y M3<sup>2</sup> en conjunto para predecir la inflación. La capacidad de predicción de las mejores redes neuronales son comparadas con metodologías tradicionales de pronóstico de la inflación (C phillips,  $P^*$ , Escandinavo, precios relativos y ARIMA), demostrando empíricamente el potencial de las redes neuronales en la predicción. Posteriormente, Misas *et al.* (2003) centran sus esfuerzos en pronosticar la demanda de dinero a través de las redes neuronales, basados fundamentalmente en que los modelos lineales no consiguen captar mejores predicciones y que la demanda de dinero exhibe un comportamiento del tipo no lineal. Para determinar la efectividad de los modelos no lineales, como las redes neuronales en la predicción, se realizan comparaciones con los pronósticos obtenidos a través de metodologías lineales como los ARIMA, ARIMA con transferencia y vector autorregresivo cointegrado (VEC). A través del modelamiento con redes neuronales, se definen dos tipos de redes: i) la autorregresiva, en la que la serie del efectivo real se utiliza en rezagos para predecir la demanda de efectivo, y ii) la ampliada, en la cual, adicional a la serie de efectivo real, la serie en rezagos de la inflación anual se incluye en el proceso de estimación del modelo.

La comparación entre las dos redes neuronales en la capacidad de pronóstico fuera de la muestra deja con un mejor desempeño en promedio la red ampliada contra la red autorregresiva. Igualmente, la red ampliada se desempeña considerablemente bien en la predicción fuera de la muestra, al compararse con las metodologías ARIMA (con función y sin función de transferencia) y el VEC.

<sup>1</sup> A partir de los noventa, un desarrollo investigativo a nivel universitario relativo a la aplicación práctica de las redes neuronales en diferentes áreas como la ingeniería y la física ha conducido a interesantes documentos de grado que pueden ser consultados a través del catálogo de la Universidad Nacional de Colombia.

<sup>2</sup> El IPC es un indicador que mide el comportamiento de los precios de un conjunto de bienes y servicios de consumo final; M1 mide la cantidad de dinero en efectivo circulante de la economía, mientras que M3 mide el nivel de liquidez total de la economía.

Tres años después, Jalil y Misas (2006) publicaron un trabajo en la misma línea, intentando analizar la capacidad de pronóstico de las redes neuronales, esta vez en la predicción del tipo de cambio. En este trabajo es importante señalar que el desempeño en pronóstico es evaluado sobre funciones de pérdida simétricas (frecuentemente utilizadas en determinar si un método es mejor que otro en pronóstico), y funciones asimétricas (las cuales ponderan los errores de pronóstico dependiendo de su magnitud y signo). Los resultados experimentales revelan que las redes neuronales tienen una mayor capacidad predictiva sobre cada una de las funciones de pérdida simétrica y asimétrica, comparado con métodos lineales como el ARIMA y la caminata aleatoria.

Por último, tenemos el trabajo de Aristizábal y Misas (2006), en el que se retoma el estudio sobre la relación entre dinero e inflación. En este caso, un proceso dinámico de pronóstico fuera de la muestra es realizado a través del método conocido como *rolling*, que consiste en actualizar o reestimar los parámetros de la red cuando se ingresa nueva información en la red, antes de realizar las predicciones fuera de la muestra.

Posteriormente, la medida de exactitud de la predicción se evalúa con relación a los verdaderos valores a través de la función de pérdida asimétrica LINLIN. Los resultados finales exhiben las ventajas de utilizar las redes neuronales en la predicción, en comparación con la metodología tradicional de Box-Jenkins.

### 3. Redes neuronales artificiales

La arquitectura de redes neuronales más ampliamente utilizada es la conocida con el nombre de perceptron multicapas, la cual se caracteriza por el hecho de que sus neuronas son agrupadas en capas de diferentes niveles. Cada una de estas capas está constituida por un conjunto de neuronas. Hay tres tipos de capas diferentes: la capa de entrada, las capas ocultas y la capa de salida, como se observa en la figura 1. Las neuronas de la capa de entrada se encargan únicamente de recibir señales o patrones que vienen del exterior y propagan tales señales a todas las neuronas de la capa siguiente. La última capa actúa como salida de la red, proporcionando al exterior la respuesta de la red para cada uno de los patrones de entrada. Las neuronas de las capas ocultas realizan un procesamiento no lineal de los patrones recibidos. Como puede observarse en la figura 1, las conexiones del perceptron multicapas están siempre dirigidas hacia adelante, i.e., las neuronas de una capa se conectan con las neuronas de la capa siguiente; por tal motivo, reciben el nombre de redes alimentadas hacia adelante o redes *feedforward*. A las

conexiones se les asocia un número real llamado peso de la conexión, y a las neuronas de la red, un umbral, que en el caso del perceptron multicapas es tratado como una conexión adicional a la neurona.

#### 3.1 Propagación de los patrones de entrada

Una vez descrita la forma como fluye la información a través de la arquitectura del perceptron multicapas, presentaremos a continuación las expresiones para el cálculo de las activaciones de las neuronas de la red.

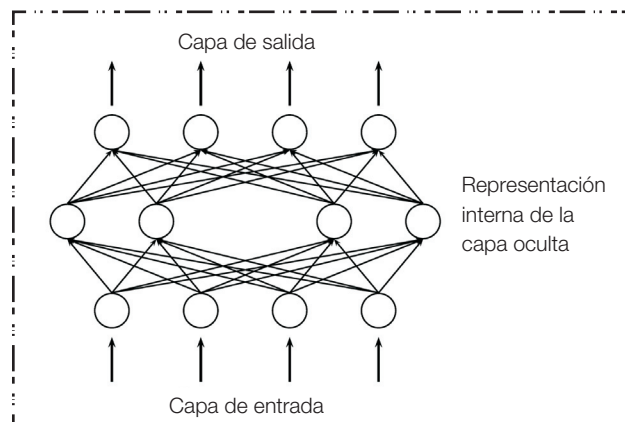
Considere un perceptron multicapas con  $C$  capas ( $C-2$  capas ocultas) y  $n_c$  neuronas en la capa  $c$ , para  $c = 1; 2; \dots; C$ . Sea  $W_c = (w_{ij}^c)$  la matriz de pesos asociada a las conexiones de la capa  $c$  a la capa  $c+1$ , para  $c = 1, 2, \dots, C-1$ , en que  $w_{ij}^c$  representa el peso de la conexión de la neurona  $i$  de la capa  $c$  a la neurona  $j$  de la capa  $c+1$ ; además, sea  $U^c = (u_i^c)$  el vector de umbrales de las neuronas de la capa  $c$  para  $c = 2, \dots, C$ . Se denota por  $a_i^c$  la activación de la neurona  $i$  de la capa  $c$ ; estas activaciones se calculan de la siguiente forma:

$$a_i^c = f\left(\sum_{j=1}^{n_{c-1}} w_{ij}^{c-1} a_j^{c-1} + u_i^c\right),$$

para  $i = 1, 2, \dots, n_c$  y  $c = 2, 3, \dots, C$ . Las activaciones para la primera capa corresponden simplemente con las observaciones de entrada a la red.

La función  $f(\cdot)$  es llamada *función de activación o transferencia*. Para el perceptron multicapas, las funciones de activación más utilizadas son la logística o sigmoide y la tangente hiperbólica; sin embargo, también se utilizan otras funciones de activación; ver Gately (1996). El propósito de la función de activación o transferencia es no permitir la salida de valores muy grandes, que puedan retrasar el proceso de convergencia del algoritmo de entrenamiento o aprendizaje, el cual será descrito a continuación.

FIGURA 1. Red Neuronal *Feedforward*



### 3.2 Algoritmo de retropropagación.

El algoritmo de aprendizaje es el mecanismo mediante el cual se van adaptando y modificando todos los parámetros de la red. El problema de aprendizaje de la red es un problema de minimización de la siguiente forma:

$$\min_w E$$

siendo  $W$  el conjunto de parámetros de la red (pesos y umbrales) y  $E$  una función del error que evalúa la diferencia entre la salida de la red y la salida deseada. En la mayoría de los casos la función del error se define por

$$E = \frac{1}{N} \sum_{n=1}^N e(n), \quad (1)$$

donde  $N$  es el número de observaciones o patrones y  $e(n)$  es el error cometido por la red para el  $n$ -ésimo patrón, que es dada por

$$e(n) = \frac{1}{2} \sum_{i=1}^{n_c} (s_i(n) - y_i(n))^2, \quad (2)$$

con  $Y_n = (y_1(n), y_2(n), \dots, y_{n_c}(n))$  y

$S_n = (s_1(n), s_2(n), \dots, s_{n_c}(n))$ , los vectores de salida de la red y salida deseada para el  $n$ -ésimo patrón, respectivamente.

Si  $W^*$  es un mínimo de la función (1), i.e., el punto donde el error es el menor posible y la salida de la red es próxima de la deseada, se obtiene el fin del proceso de aprendizaje.

Para el perceptron multicapas, el método de optimización no lineal más utilizado es el *steepest descent* sobre la función  $E$ . De esta forma, cada parámetro  $w$  de la red es modificado para cada patrón de entrada  $n$  de acuerdo con la siguiente ley de aprendizaje:

$$w(n) = w(n-1) - \eta \frac{de(n)}{dw} \quad (3)$$

con  $e(n)$  definido en (2) y  $\eta$  siendo la tasa de aprendizaje que influye en la magnitud de desplazamiento sobre la superficie del error. El método del gradiente puede ser aplicado de forma eficiente, resultando en el conocido algoritmo de retropropagación o *regla delta generalizada*; ver Rumelhart *et al.* (1986a) e Isasi y Galván (2004) para una descripción más detallada sobre esta regla.

Como puede ser observado en (3), el cambio en un peso es proporcional al gradiente del error, con la proporcionalidad siendo dada por el parámetro  $\eta$ . Valores altos de la tasa de aprendizaje en principio podrían favorecer una convergencia más rápida, pues permite

avanzar rápidamente en la superficie del error. Sin embargo, tasas de aprendizaje altas pueden tener consecuencias negativas sobre el aprendizaje, haciendo que el método salte u oscile alrededor del mínimo. Valores pequeños de las tasas de aprendizaje pueden evitar estos problemas, aunque lleve posiblemente a una convergencia más lenta del algoritmo de aprendizaje, debido a que la magnitud del desplazamiento sobre la superficie del error es menor.

Un método simple para evitar la inestabilidad en el algoritmo de aprendizaje, debido a la tasa de aprendizaje, consiste en modificar la ecuación (3) a través de la inclusión de un segundo término llamado *momento*, obteniendo de esta forma la siguiente ley:

$$w(n) = w(n-1) - \eta \frac{de(n)}{dw} + \alpha \Delta w(n-1), \quad (4)$$

donde  $\alpha$  es un número positivo actuando como ponderador. Esta regla fue propuesta por Rumelhart *et al.* (1986b) y preserva las propiedades de la regla definida en (3), en el sentido que modifica los parámetros de la red para minimizar la función del error (1). El nuevo término,  $\alpha \Delta w(n-1)$ , incorpora al método alguna inercia, haciendo que la modificación actual del parámetro dependa sólo de la dirección de la modificación anterior, consiguiendo evitar oscilaciones. Haciendo cálculos sucesivos sobre  $\Delta w(n-1)$ , Isasi y Galván (2004) exhiben una expresión más general de (4):

$$w(n) = w(n-1) - \eta \sum_{t=0}^{n-1} \alpha^{n-t} \frac{de(t)}{dw}, \quad (5)$$

El proceso de aprendizaje del perceptron multicapas debe ser finalizado cuando  $\frac{dE}{dw} \approx 0$ ,

momento en el cual los parámetros de la red no cambian de forma perceptible entre iteraciones consecutivas.

### 3.3 Capacidad de generalización

A la hora de evaluar el comportamiento de la red y en particular del perceptron multicapas no importa saber únicamente si la red aprendió con éxito los patrones utilizados durante el aprendizaje, sino también conocer el comportamiento de la red frente a patrones que no fueron utilizados durante el entrenamiento.

Para tal fin, es necesario disponer de dos conjuntos de patrones: el *conjunto de entrenamiento* que entrena y modifica los pesos y umbrales de la red, y el *conjunto de validación* o *test*, que mide la capacidad de la red para responder correctamente a los patrones que no fueron ingresados durante el entrenamiento. Cuando la red aproxima correctamente los patrones de aprendizaje,

pero no responde bien a los patrones de validación, se dice que hubo sub-aprendizaje de la red, posiblemente ocasionado por varios factores, como el uso de un número excesivo de neuronas o capas ocultas, implicando un aumento en el número de parámetros por ser estimados; ver Isasi y Galván (2004), Kaastra y Boyd (1996) y Zhang *et al.* (1998) para una discusión mayor sobre el tema.

### 3.4 Metodología de modelamiento

La propuesta de Varfis y Versino (1990) para la determinación de las variables de entrada a la red es considerada. Su utilidad recae principalmente en incluir como entradas a la red componentes adicionales a las variables exógenas, tal como la tendencia o la estacionalidad (cuando sea el caso), que ayuden a la red a captar mejor la dinámica de la serie, debido a que las redes neuronales muchas veces son incapaces de modelar adecuadamente estructuras como la tendencia o la estacionalidad, teniendo efectos posteriores en la etapa de pronóstico; ver Nelson *et al.* (1994) para profundizar más en este tema.

## 4. Modelamiento y pronóstico de la inflación con redes neuronales: comparación con metodologías tradicionales

El análisis referente al modelamiento y predicción presentado a continuación se realiza sobre las variaciones del índice de precios al consumidor colombiano. El índice de precios al consumidor, al igual que las variaciones mensuales (inflación) son publicadas por el DANE<sup>3</sup> mensualmente. Bases de datos históricas se pueden consultar a través de la página [www.dane.gov.co](http://www.dane.gov.co). Se utiliza la serie mensual entre enero de 1998 y diciembre de 2005 para el análisis.

Se seleccionó el periodo de enero de 1998 a junio de 2005 para el modelamiento de la serie con el fin de obtener las predicciones del periodo de julio a diciembre de 2005. Nuestro principal interés consistió en observar cómo cada modelo captura la estructura dinámica de la serie y la refleja sobre el periodo de predicción. La razón por la que son consideradas seis observaciones en la predicción y no un periodo más largo, tal como un año, radica en la creencia de que la existencia de cambios estructurales puedan hacer que el modelo no recoja adecuadamente la dinámica que exhibiría la serie verdadera y, por ende, se podrían presentar elevados errores de pronóstico. De esta forma, conocer

la dinámica de la inflación del primer trimestre del año puede mejorar la capacidad predictiva del modelo y hace posible una mejor predicción para el segundo semestre.

Dos formas de medición del error de predicción fueron adoptadas para establecer las comparaciones, respecto a las predicciones realizadas por cada metodología:

- **Error cuadrático medio (MSE).** Si  $Z_1, Z_2, \dots, Z_t$  son las observaciones de la serie temporal y  $Z_{t+1}, Z_{t+2}, \dots, Z_{t+h}$  sus  $h$  predicciones respectivas, entonces

$$MSE = \frac{1}{h} \sum_{k=1}^h e_{t+k}^2,$$

donde  $e_{t+k}$  es el error de predicción de  $Z_{t+k}$ , el valor de la serie en el instante  $t+k$ .

- **Error absoluto medio (MAE).** Aquí,

$$MAE = \frac{1}{h} \sum_{k=1}^h |e_{t+k}|$$

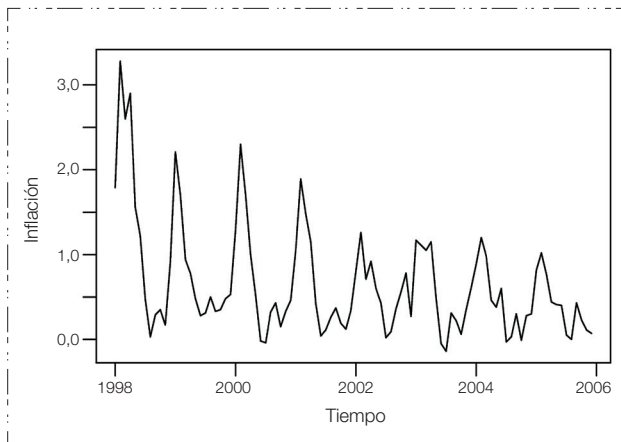
donde  $e_{t+k}$  es el error de predicción en el instante  $t+k$ ;  $k = 1, 2, \dots, h$ .

La serie de inflación para el periodo comprendido entre enero de 1998 y diciembre de 2005 se exhibe en la figura 2. Es posible observar a través de esta figura fuertes indicios de no estacionariedad y estacionalidad, es decir, tanto la media como la varianza cambian con el tiempo, y cada año parecen presentarse comportamientos semejantes, respectivamente. La transformación logaritmo se aplica para el análisis a la serie de inflación, con el fin de estabilizar la varianza.

**Modelamiento y pronóstico.** Los algoritmos de suavizamiento aditivo y multiplicativo de Holt-Winters fueron utilizados para obtener, inicialmente, predicciones referentes al periodo que se extiende de julio a diciembre de 2005. Los valores optimizados de las constantes de suavizamiento para la serie fueron semejantes para los modelos aditivo y multiplicativo; en este caso, resultado de utilizar la transformación logaritmo sobre la serie. Igualmente, las predicciones obtenidas a través de estos dos modelos resultaron semejantes, así que el modelo aditivo fue escogido para la generación de las predicciones. La tabla 1 señala las respectivas predicciones, un paso ( $h = 1$ ) y seis ( $h = 6$ ) pasos al frente, sobre la escala original.

<sup>3</sup> Departamento Administrativo Nacional de Estadística.

**FIGURA 2. Serie de inflación. Ene./98-Dic./05.**



**TABLA 1. Predicción de la inflación a través del suavizamiento exponencial aditivo.**

Periodo	h = 1	h = 6
Jul. -05	-0,113	-0,113
Ago. -05	0,067	0,057
Sep. -05	0,230	0,224
Oct. -05	0,037	0,020
Nov. -05	0,316	0,286
Dic. -05	0,340	0,323

Para el proceso de identificación del modelo a través de la metodología SARIMA de Box-Jenkins fue utilizado el criterio de información bayesiano BIC. El programa TRAMO-SEATS se empleó para estimar los parámetros del modelo SARIMA, y puede ser obtenido gratuitamente a través del portal del Banco de España: [www.bde.es](http://www.bde.es)

**TABLA 2. Estadísticas de diagnóstico.**

Serie	SARIMA(1,1,1)x(0,1,1) <sub>12</sub>
SE(Res)	0,17
BIC	-3,33
Q-Val	34,11
N-Test	1,32
SK(t)	1,06
KUR(t)	-0,44
Q2	10,96
RUNS	-0,23

El modelo SARIMA(1,1,1)x(0,1,1)<sub>12</sub> se escogió para la serie de la inflación. El análisis de diagnóstico permite observar que no hay evidencias contra la hipótesis de ausencia de autocorrelación de los residuos, como tampoco contra la hipótesis de normalidad de los residuos al nivel de significancia del 1%.

La tabla 2 muestra las estadísticas del análisis de diagnóstico sobre los residuales del modelo seleccionado, donde SE(Res) es el error estándar de los residuos; Q-Val es la estadística de Ljung-Box para probar la hipótesis de correlación serial, calculada sobre 24 autocorrelaciones (en todos los casos se utiliza la distribución asintótica  $\chi^2$ , con 21 grados de libertad); N-test es la estadística de Bowman-Shenton para probar la hipótesis de normalidad (se aplica la distribución asintótica  $\chi^2$  (2 g.l.)); SK(t) es la estadística empleada para probar si la asimetría es cero contra si es diferente de cero; KUR(t) es la estadística utilizada para probar si el exceso de curtosis es cero contra si es diferente de cero; Q2 es la estadística de McLeod y Li (1983) para probar la linealidad del proceso (se emplea la distribución asintótica  $\chi^2$  (24 g.l.)); por último, RUNS es la estadística utilizada para probar la hipótesis nula sobre aleatoriedad del conjunto de residuos. Todas las pruebas de hipótesis se realizan a un nivel de significancia de 1%. Los cuantiles de la ji-cuadrado con 2, 21 y 24 grados de libertad, al 1%, son dados por 9,21, 38,93 y 42,98, respectivamente.

La ecuación del modelo estimado para la serie transformada de la inflación, denotada por  $Y_t$ , es dada por

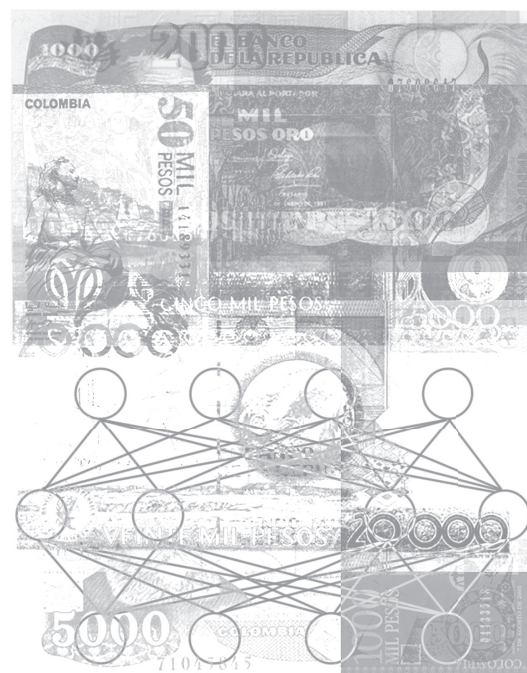
$$(1 + 0,429B) Y_t = (1 + 0,991B)(1 + 0,769 B^{12}) a_t; \\ [t = 3,4] \quad [t = 66] \quad [t = -3.9]$$

donde

$$Y_t = (1-B)(1-B^{12})Z_t,$$

y

$$a_t \sim N(0, (0,1736)^2).$$



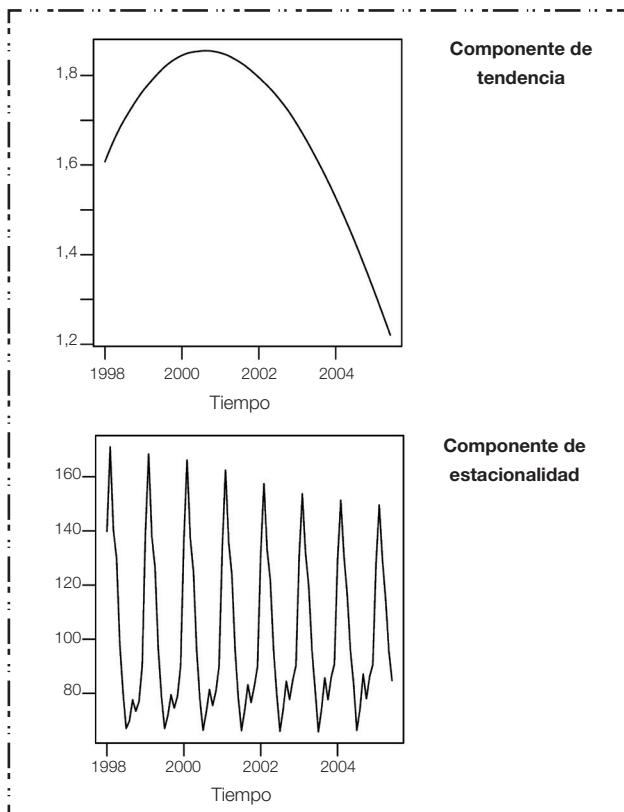
Las componentes de tendencia y estacionalidad sirven para el modelamiento con redes neuronales y fueron extraídas del modelo SARIMA; ver Gómez y Maravall (1996) para una discusión mayor sobre el tema. La figura 3 exhibe las dos componentes no observables entre enero de 1998 y junio de 2005. Nótese que la tendencia exhibida por la inflación sigue una trayectoria decreciente desde finales de 2000 hasta mitad de 2005. Con la componente estacional podemos observar que en los meses de febrero se registran los datos de inflación más altos del año, mientras que en julio están los más bajos.

La tabla 3 contiene las predicciones un paso ( $h = 1$ ) y seis pasos ( $h = 6$ ) al frente, utilizando el modelo SARIMA estimado anteriormente.

**TABLA 3. Predicción de la inflación a través de la metodología SARIMA de Box-Jenkins.**

Periodo	$h = 1$	$h = 6$
Jul. -05	0,044	0,044
Ago. -05	0,170	0,164
Sep. -05	0,200	0,281
Oct. -05	0,208	0,150
Nov. -05	0,356	0,318
Dic. -05	0,309	0,410

**FIGURA 3. Componentes no observables del IPC.**



Se realizó un análisis preliminar basado en regresión en busca de quiebres estructurales sobre la serie mensual de inflación, utilizando los procesos CUSUM, MOSUM y los basados en estimaciones (ME, *Fluctuation test*); ver Zeileis *et al.* (2002) para una descripción más detallada sobre estas pruebas con implementaciones computacionales en el paquete estadístico R. Del conjunto de pruebas realizado sobre un modelo de regresión con las variables en rezago de primer, segundo, octavo, noveno y duodécimo orden, la única prueba significativa fue la realizada con el proceso MOSUM, exhibiendo un quiebre estructural a principios de 1999 sobre la variable con el duodécimo rezago. La figura 4 señala el resultado comentado anteriormente.

El modelo de regresión con parámetros variando en el tiempo se utiliza para pronosticar los datos de inflación del periodo de julio a diciembre de 2005. Una aproximación a través de espacio de estados, suponiendo que los coeficientes varían en el tiempo de forma recurrente, se utiliza para obtener los pronósticos. La tabla 4 muestra las predicciones respectivas. El modelo estimado registró un valor de la log-verosimilitud igual a  $-10.823$ . Otras representaciones en espacio de estados fueron consideradas; sin embargo, la descrita exhibió los mejores pronósticos posibles con relación a los verdaderos valores de la inflación en el periodo de julio a diciembre de 2005.

**TABLA 4. Predicción de la inflación a través de FLS**

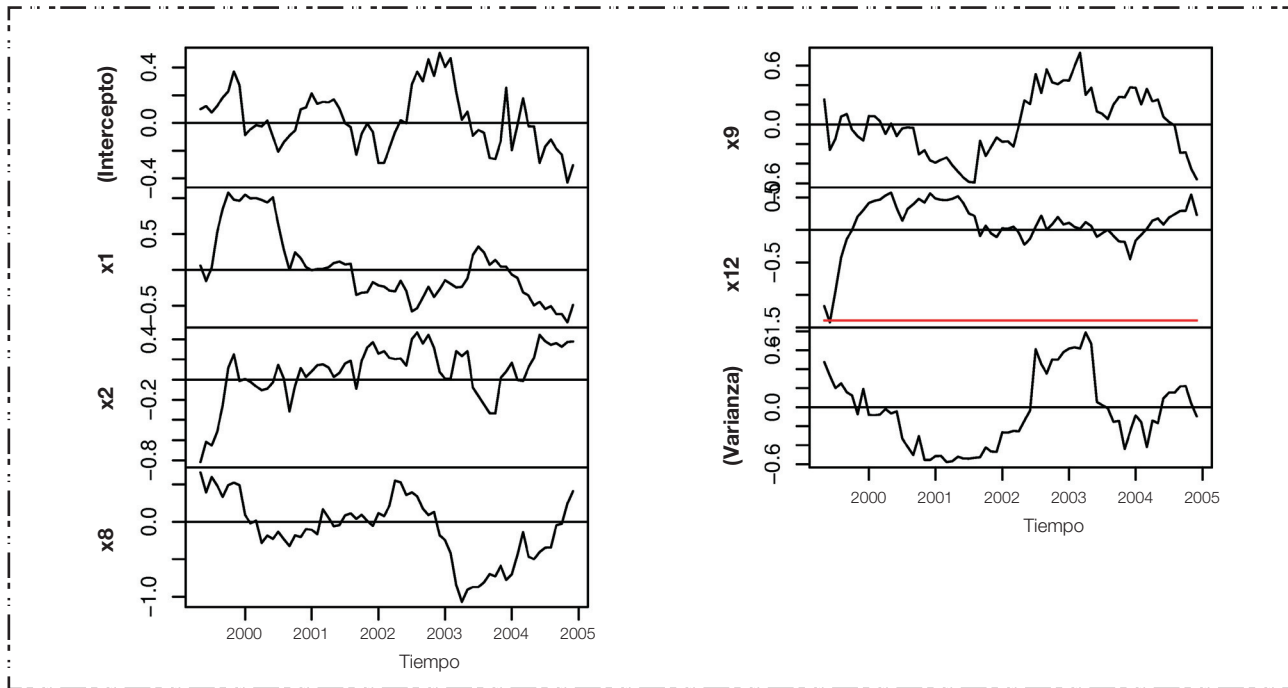
Periodo	$h = 1$	$h = 6$
Jul. -05	0,189	0,189
Ago. -05	0,057	0,062
Sep. -05	0,130	0,136
Oct. -05	0,388	0,394
Nov. -05	0,273	0,277
Dic. -05	0,295	0,302

Para el modelamiento de la inflación a través de redes neuronales se utilizó la propuesta de Varfis y Versino (1990). Cada serie es re-escalada en el intervalo  $[-1, 1]$  antes de incluir variables rezagadas, componentes de tendencia y estacionalidad como neuronas de entrada. Posteriormente, dos capas ocultas son consideradas, dos parámetros de aprendizaje, dos parámetros de momento, 10.000 *epochs* y un máximo de seis neuronas por capa oculta, siguiendo las sugerencias de Kaastra y Boyd (1996) y Zhang *et al.* (1998). La función de activación es la función tangente hiperbólica anti-simétrica descrita en Haykin (1994). Se definen tres clases de arquitectura de redes neuronales:

- a) Red 1. Una capa oculta es considerada, con un número máximo de seis neuronas.



**FIGURA 4. Prueba Score-MOSUM sobre inflación.**



- b) Red 2. Dos capas ocultas son consideradas, con igual número de neuronas en cada capa oculta, con máximo seis neuronas.
- c) Red 3. Dos capas ocultas son consideradas, con una neurona adicional en la segunda capa y cada una con un número máximo de seis neuronas.

El objetivo principal fue obtener una red con buen ajuste y la mejor predicción posible. Para bautizar la red neuronal identificada se utilizó la notación de Souza y Zandonade (1993), dada por ANN( $n_1, n_2, \dots, n_c$ ), donde  $n_c, c = 1, 2, \dots, C$  es el número de neuronas en la capa  $c$ . De esta forma, dado un conjunto fijo de neuronas de entrada, 72 redes neuronales fueron simuladas. Consideramos adicionalmente la simulación de un conjunto de redes neuronales eximiendo la componente de tendencia, de tal forma que pudiéramos evaluar, empíricamente, la ventaja de introducir tal componente.

Un programa escrito en lenguaje R se utilizó para el proceso de simulación y predicción con redes neuronales. En promedio, cada red neuronal empleó diez minutos durante el proceso de entrenamiento para un total de doce horas de simulación. Las variables de entrada a la red fueron determinadas por medio de un modelo autorregresivo.

Resultado de las simulaciones, la red ANN(17; 5; 1) fue identificada con 17 neuronas de entrada, definidas por el primero, segundo, octavo, noveno y duodécimo

rezagos de la serie y doce variables “dummy” correspondientes a cada mes del año. Esta red es del tipo 1 con  $\eta=0,1, \alpha=0,1$ , y es la que proporcionó la mejor predicción un paso al frente tanto en MSE como en MAE. La red ANN(15; 4; 1), con variables rezagadas de primero, duodécimo, décimotercer orden y doce variables dummy, exhibió las mejores predicciones múltiples pasos según MSE. Esta red es del tipo 1, con  $\eta=0,01, \alpha=0,5$ . En función del MAE las mejores predicciones múltiples pasos fueron proporcionadas por la red ANN(17; 2; 2; 1) con rezagos de primero, segundo, octavo, noveno y duodécimo orden, junto con doce variables dummy como antes. Ésta es una red tipo 2 con  $\eta=0,01, \alpha=0,1$ . La tabla 5 exhibe las predicciones ANN y múltiples pasos al frente para estas tres redes. Obsérvese que las tres redes encontradas anteriormente no incluyen la tendencia entre sus neuronas de entrada.

**TABLA 5. Predicciones ANN: uno y seis pasos al frente.**

Periodo	ANN(17,5,1)		ANN(15,4,1)		ANN(17,2,2,1)	
	h=1	h=6	h=1	h=6	h=1	h=6
Jul. -05	0,008	0,008	-0,013	-0,013	-0,063	-0,063
Ago. -05	-0,116	-0,098	0,151	0,133	0,028	0,021
Sep. -05	0,379	0,379	0,226	0,217	0,257	0,293
Oct. -05	0,43	0,513	0,285	0,132	0,34	0,201
Nov. -05	0,107	0,288	0,209	0,184	0,131	0,186
Dic. -05	0,236	-0,03	0,238	0,234	0,344	0,345

Una vez obtenidas las predicciones a través de cada metodología, se inspeccionaron las medidas de error de predicción MSE y MAE para determinar cuál metodología, de forma individual, es la que proporciona los mejores resultados. La tabla 6 exhibe estos resultados, en los que se puede destacar que las redes neuronales presentan las mejores predicciones tanto en un paso como en múltiples pasos, comparado con metodologías tradicionales como el suavizamiento exponencial de Holt-Winters, SARIMA de Box-Jenkins y FLS.

**TABLA 6. Medidas de error de predicción.**

Modelo	MSE		MAE	
	h=1	h=6	h=1	h=6
HW	0,037	0,035	0,183	0,178
SARIMA	0,033	0,036	0,152	0,158
FLS	0,035	0,036	0,173	0,176
ANN(17,4,1)	<b>0,014</b>	0,023	<b>0,096</b>	0,125
ANN(15,3,1)	0,018	<b>0,018</b>	0,123	0,124
ANN(17,2,2,1)	0,022	0,019	0,12	<b>0,108</b>

La media ponderada de las predicciones individuales proporcionadas por el suavizamiento exponencial, por el método SARIMA de Box-Jenkins y por las redes neuronales se utilizaron para obtener una predicción mejorada. Estos pesos fueron definidos como siendo inversamente proporcionales al error de predicción. Es importante resaltar que sólo se discutirán las combinaciones que proporcionaron las mejores predicciones según el MSE, es decir, aquellas combinaciones que proporcionaron el mínimo MSE. Para el estudio de estas combinaciones, un conjunto adicional de 16 redes neuronales con las mejores predicciones según el MSE se obtuvieron a través de simulaciones, evaluando diferentes configuraciones de la capa de entrada.

Poseemos en total 22 modelos: el mejor modelo SARIMA según el BIC, el mejor modelo FLS en pronóstico, el mejor modelo de Holt-Winters aditivo y 19 redes neuronales con el mejor ajuste y predicción posible. Fueron considerados, entonces,  $2^{22}$  posibles combinaciones, y para cada combinación fue calculado el MSE. El proceso de simulación para esta fase tuvo una duración de ocho minutos.

En la predicción un paso al frente, la combinación de las redes neuronales ANN(18; 2; 3; 1), ANN(17; 5; 1), ANN(17; 5; 5; 1) y ANN(17; 3; 1), que llamaremos COMB1, exhibió las mejores predicciones con MSE = 0,005 y MAE = 0,058. Estos resultados fueron mejores que los registrados en la tabla 6. Para la predicción seis pasos al frente, las redes neuronales ANN(18; 2; 3; 1), ANN(14; 6; 6; 1), ANN(15; 4; 1) y ANN(17; 3; 1), que llamaremos COMB2, presentaron los mejo-

res resultados con MSE = 0,009 y MAE = 0,072. La tabla 7 contiene las predicciones uno y seis pasos al frente, resultado de estas combinaciones, y la tabla 8 contiene los errores de predicción, los cuales pueden ser contrastados con los exhibidos en la tabla 6. Las variables rezagadas y los coeficientes  $\eta$  y  $\alpha$  de las redes neuronales que hicieron parte de estas combinaciones se describen en la tabla 9.

**TABLA 7. Combinación de predicciones.**

Periodo	COMB1		COMB2	
	h=1	h=6	h=1	h=6
Jul. -05	0,008	0,008	0,019	0,019
Ago. -05	0,05	0,032	0,045	0,03
Sep. -05	0,313	0,329	0,221	0,275
Oct. -05	0,203	-0,007	0,199	0,275
Nov. -05	0,098	0,097	-2,251	0,096
Dic. -05	0,171	0,623	0,194	0,224

**TABLA 8. Medidas de error de predicción.**

Combinación	MSE		MAE	
	h=1	h=6	h=1	h=6
COMB1	0,005	0,063	0,058	0,163
COMB2	0,940	0,009	0,467	0,072

**TABLA 9. Redes neuronales utilizadas.**

Red	Rezagos	$\eta$	$\alpha$
ANN(17,5,1)	$z1, z2, z8, z9, z12, 12D$	0,1	0,1
ANN(15,4,1)	$z1, z12, z13, 12D$	0,01	0,5
ANN(17,2,2,1)	$z1, z2, z8, z9, z12, 12D$	0,01	0,1
ANN(18,2,3,1)	$z1, z2, z8, z9, z12, Ten, 12D$	0,1	0,1
ANN(17,5,5,1)	$z1, z2, z8, z9, z12, 12D$	0,01	0,1
ANN(14,6,6,1)	$Z1, Ten, 12D$	0,01	0,5
ANN <sub>2</sub> (15,4,1)	$z1, z12, z13, 12D$	0,01	0,1
ANN(17,3,1)	$z1, z2, z12, z13, Ten, 12D$	0,01	0,5

La combinación de pronósticos con redes neuronales, incluidos las predicciones del modelo SARIMA, FLS y el suavizamiento de Holt-Winters, deja a la combinación de las predicciones de los modelos SARIMA, HW, FLS ANN(18; 2; 3; 1), ANN(17; 5; 1) y ANN(17; 5; 5; 1) (COMB3) con la mejor predicción un paso al frente según el MSE y a la combinación de las predicciones de los modelos SARIMA, HW, FLS, ANN(18; 2; 3; 1), ANN(14; 6; 6; 1), ANN<sub>2</sub>(15; 4; 1) -diferente a la red ANN(15; 4; 1) descrita anteriormente- y ANN(17; 3; 1) (COMB4) con la mejor predicción múltiples pasos al frente según el MSE. Nótese que las redes neurona-

les consideradas en estas combinaciones se analizaron en combinación antes, proporcionando las mejores predicciones.

Adicionalmente, la combinación de las predicciones de los modelos SARIMA, HW y FLS con las redes neuronales, redujo el MSE y el MAE para COMB3 en la predicción múltiples pasos, comparado con los resultados obtenidos por COMB1. Lo inverso ocurrió con COMB4 en beneficio de las predicciones un paso al frente en comparación con COMB2. La tabla 10 exhibe los resultados comentados anteriormente, mientras que las predicciones obtenidas por COMB3 y COMB4 se presentan en la tabla 11.

**TABLA 10. Medidas de error de predicción.**

Combinación	MSE		MAE	
	h=1	h=6	h=1	h=6
COMB3	0,008	0,038	0,071	0,142
COMB4	0,059	0,013	0,17	0,09

**TABLA 11. Combinación de predicciones.**

Periodo	COMB3		COMB4	
	h=1	h=6	h=1	h=6
Jul. -05	0,023	0,023	0,026	0,026
Ago. -05	0,059	0,043	0,055	0,041
Sep. -05	0,281	0,301	0,21	0,259
Oct. -05	0,194	-0,02	0,189	0,319
Nov. -05	0,089	0,087	0,634	0,085
Dic. -05	0,205	0,45	0,227	0,258

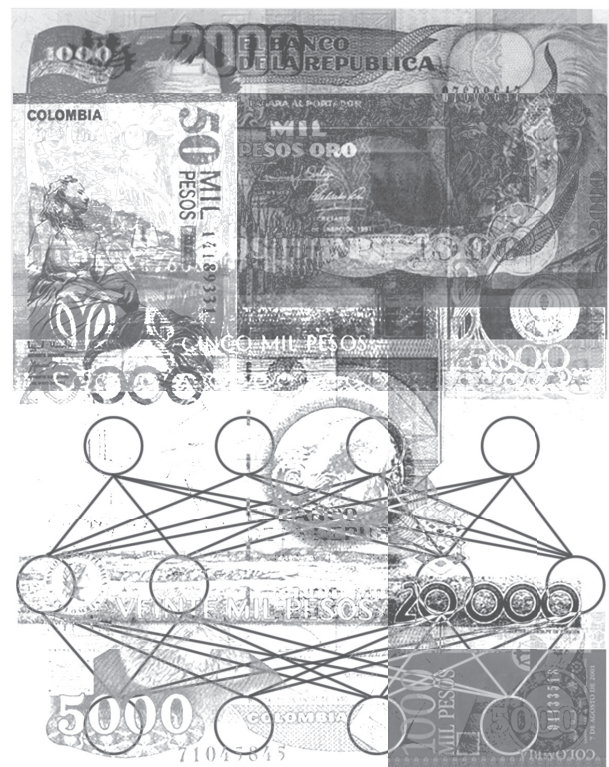
## 5. Conclusiones

Los resultados obtenidos ilustraron el uso de las redes neuronales en la predicción de series temporales. Un ejemplo aplicado sobre las variaciones en el índice de precios al consumidor (IPC) permitió observar que las predicciones obtenidas a través de nuestra metodología de redes neuronales tienden a ser más precisas que aquellas originadas en metodologías tradicionales, como el suavizamiento exponencial de Holt-Winters, el método SARIMA de Box-Jenkins y los mínimos cuadrados flexibles, siendo coherente con muchos de los estudios descritos en esta línea. Adicionalmente, la utilidad de nuestra red neuronal en combinación con otras redes de la misma clase o metodologías tradicionales se mostró efectiva en el proceso de predicción. Finalmente, las redes neuronales sin la neurona de tendencia, en la mayoría de los casos, exhibieron las menores medidas de error en la predicción, si se

comparan con la metodología utilizada que propone la tendencia como una neurona de entrada; es decir, empíricamente, nuestros modelos de redes neuronales pueden conseguir describir apropiadamente el nivel de la serie sin necesitar de una componente de tendencia. Las conclusiones descritas antes nos permiten afirmar que la metodología propuesta exhibe una alta capacidad predictiva en comparación con las técnicas lineales tradicionales, siendo una herramienta alternativa a las ya existentes, de la cual puede disponer cualquier agente económico para la toma de decisiones.

En todos los casos se estimaron y seleccionaron redes que tuvieran el mejor ajuste posible sobre el conjunto de entrenamiento y la menor medida de error de predicción sobre el conjunto de validación o test, para evitar el problema de sub-aprendizaje.

El desarrollo histórico de las redes neuronales en Colombia ha tenido un proceso lento, determinado fundamentalmente por el desconocimiento y la poca difusión de la metodología. No obstante, las puertas están aún abiertas para el desarrollo y las aplicaciones en esta línea, tal como la evaluación de la capacidad predictiva de las redes neuronales recurrentes, las redes eliminando conexiones, los algoritmos genéticos y las redes dinámicas DAN2, cuya metodología es descrita más ampliamente en Ghiassi y Saidane (2005) y Ghiassi, Saidane y Zimbra (2005).



## Referencias bibliográficas

- Aristizábal, M. & Misas, M. (2006). Evaluación asimétrica de una red neuronal artificial: Una aplicación al caso de la inflación en Colombia. Reporte técnico, *Working Paper* 377. Subgerencia de estudios económicos. Banco Central de Colombia.
- Barnard, G. A. (1963). New Methods of Quality Control. *Journal of the Royal Statistical Society A* 126, 255-259.
- Box, G. E. P. & Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- Box, G. E. P., Jenkins, G. M. & Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control* (3rd ed.). Englewood Cliffs: Prentice Hall.
- Cohen, M., Franco, H., Morgan, N., Rumelhart, D. & Abrash, V. (1993). *Advances in Neural Information Processing Systems* (pp. 649-657). Morgan Kaufmann, chapter Context-dependent multiple distribution phonetic modeling with MLPs.
- Cybenko, M. (1989). Aproximation by Superposition of a Sigmoidal Function. *Mathematics of Control, Signals and Systems* 2, 303-314.
- Gately, E. (1996). *Neural Networks for Financial Forecasting*. New York: John Wiley and Sons.
- Ghiassi, M. & Saidane, H. (2005). A Dynamic Architecture for Artificial Neural Networks. *Neurocomputing* 63, 397-413.
- Ghiassi, M., Saidane, H. & Zimbra, D. K. (2005). A Dynamic Artificial Neural Network Model for Forecasting Time Series Events. *International Journal of Forecasting* 21, 341-362.
- Gómez, V. & Maravall, A. (1996). Programs TRAMO (Time series Regression with Arima noise, Missing observations, and Outliers) and SEATS (Signal Extraction in Arima Time Series). Instructions for the User. Reporte técnico, *Working Paper* 9628. Servicio de Estudios. Banco de España.
- Haykin, S. (1994). *Neural Networks*. New York: McMillan College Publishing Company.
- Hendry, D. F. & Clements, M. P. (2004). Pooling of Forecast. *Econometrics Journal* 7, 1-31.
- Hornik, K., Stinchcombe, M. & White, H. (1989). Multilayer Feedforward Networks and Universal Approximations. *Neural Networks* 2, 359-366.
- Isasi, P. & Galván, I. (2004). *Redes neuronales artificiales. Un enfoque práctico*. Madrid: Pearson-Prentice Hall.
- Jalil, M. & Misas, M. (2006). Evaluación de pronósticos del tipo de cambio utilizando redes neuronales y funciones de pérdida asimétricas. Reporte técnico, *Working Paper* 376. Subgerencia de estudios económicos. Banco Central de Colombia.
- Kaasta, I. & Boyd, M. (1996). Design a Neural Network for Forecasting Financial and Economic Time Series. *Neurocomputing* 10, 215-236.
- Kalaba, R. & Tesfatsion, L. (1989). Time-Varing Linear Regression Via Flexible Least Squares. *Computers and Mathematics with Applications* 17, 1215-1245.
- Kalaba, R. & Tesfatsion, L. (1996). A Multicriteria Approach to Model Specification and Estimation. *Computational Statistics & Data Analysis* 21, 193-214.
- Maravall, A. & Kaiser, R. (2000). Notes on Time Series Analysis, ARIMA models and Signal Extraction. Departamento de Investigaciones. Banco de España. [www.bde.es/servicio/software/trabajos.htm](http://www.bde.es/servicio/software/trabajos.htm)
- McLeod, A. I. & Li, W. K. (1983). Diagnostic Checking ARMA Time Series Models Using Squared-Residuals Autocorrelation. *Journal of the Time Series Analysis* 4, 269-273.
- Misas, M. & Melo, F. (2004). Modelos estructurales de inflación en Colombia: estimación a través de mínimos cuadrados flexibles. Reporte técnico, *Working Paper* 283. Subgerencia de estudios económicos. Banco Central de Colombia.
- Misas, M., López, E. & Borrero, P. (2002). La inflación en Colombia: una aproximación desde redes neuronales. *Ensayos sobre Política Económica* 41- 42, 143-214.
- Misas, M., López, E., Arango, C. & Hernández, J. (2003). La demanda de efectivo en Colombia: una caja negra a la luz de las redes neuronales. Reporte técnico, *Working Paper* 268. Subgerencia de estudios económicos. Banco Central de Colombia.
- Montgomery, D. C. & Johnson, L. A. (1976). *Forecasting and Time Series Analysis*. New York: McGraw-Hill.
- Morettin, P. A. & Toloí, C. M. (2004). *Análise de Séries Temporais*. ABE – Projeto Fisher. São Paulo: Edgard Bläucher.
- Narendra, K. & Parthasaranty, K. (1990). Identification and Control of Dynamical Systems Using Neural Networks. *IEEE Transactions on Neural Networks* 1, 4-27.
- Nelson, M., Hill, T., Remus, B. & O'Connor, M. (1994). Can Neural Networks be applied to Time Series forecasting and learn seasonal patterns: An empirical investigation (pp. 649-655). En: Proceedings of the twenty seventh Annual Hawaii International Conference on Systems Sciences.
- Rumelhart, D., Hilton, G. & Williams, R. (1986a). *Parallel Distributed Processing*. Cambridge: The MIT Press, chapter Learning representations by backpropagating errors.
- Rumelhart, D., Hilton, G. & Williams, R. (1986b). Learning Representations by Backpropagating Errors. *Nature* 323, 533-536.
- Sanín, F. G. (2001). Agregando votos en un sistema altamente desinstitucionalizado. Reporte técnico, *Working Paper* 157. Dirección de estudios económicos. Departamento Nacional de Planeación.
- Santana, J. C. (2005). Previsão da Arrecadação do ICMS Atraves de Redes Neurais no Brasil. Tese de Mestrado. Departamento de Estatística. Universidade Federal de Pernambuco, Brasil.
- Souza, R. C. & Zandonade, E. (1993). Forecasting Via Neural Networks: Comparative Study. Reporte técnico. Department of Electrical Engineering, Catholic University of Rio de Janeiro.
- Varfis, A. & Versino, C. (1990). *Univariate Economic Time Series Forecasting*. Cambridge University Press.
- Wiegand, A., Huberman, B. & Rumelhart, D. (1990). Predicting the Future: a Connectionist Approach. Reporte técnico. PARC.
- Zeileis, A., Leisch, F., Hornik, K. & Kleiber, C. (2002). An R Package for Testing for Structural Change in Linear Regression Models. *Journal of Statistical Software* 7, 1-38.
- Zhang, G., Patuwo, B. & Hu, Y. (1998). Forecasting with Artificial Neural Networks: The State of Art. *International Journal of Forecasting* 14, 35-62.