

# MICROESTRUCTURA Y DINÁMICA DE LA TASA DE CAMBIO NOMINAL EN COLOMBIA: UNA APROXIMACIÓN CON REDES NEURONALES ARTIFICIALES Y SISTEMAS NEURODIFUSOS\*

*Jhon Alexander Méndez Sayago\*\**

---

\* Este artículo se basa en el documento de investigación registrado como propiedad intelectual ante la Facultad de Ciencias Sociales y Económicas de la Universidad del Valle, Cali, Colombia. Fue desarrollado entre febrero de 2006 y junio de 2007. El artículo se recibió el 03-10-2007 y se aprobó el 18-03-2008.

\*\* Magíster en Economía, Pontificia Universidad Javeriana, Bogotá, Colombia (2006); Magíster en Economía del Medio Ambiente y de los Recursos Naturales, Universidad de los Andes, Bogotá, Colombia (2003); profesor del Departamento de Economía, Facultad de Ciencias Sociales y Económicas, Universidad del Valle, Cali, Colombia. Miembro del grupo de investigación Economía Regional y Ambiental (GERA), Universidad del Valle.  
Correo electrónico: [jhmendez@univalle.edu.co](mailto:jhmendez@univalle.edu.co)

***Microestructura y dinámica de la tasa de cambio nominal en Colombia: una aproximación con redes neuronales artificiales y sistemas neurodifusos***

RESUMEN

Este artículo presenta los resultados de evaluaciones empíricas de pronósticos de la tasa de cambio nominal en Colombia realizados a partir de especificaciones lineales (modelo de rezagos distribuidos) y no lineales (redes neuronales artificiales y sistemas neurodifusos). Todas las especificaciones econométricas emplean el *flujo de órdenes* como determinante del tipo de cambio, la variable más representativa de los modelos que tratan de explicar la tasa de cambio nominal, apoyados en la microestructura del mercado cambiario. Las medidas de error de pronóstico utilizadas para evaluar la habilidad predictiva de los modelos evidenciaron las ventajas de los modelos lineales y no lineales sobre la caminata aleatoria, atribuible al flujo de órdenes. Los modelos no lineales mostraron alguna superioridad sobre el modelo lineal en términos de la raíz del error cuadrado medio, el coeficiente de Theil y el porcentaje de cambios correctos predichos, tan leve que no permite confirmar la hipótesis de no linealidad del proceso generador de la tasa de cambio.

**Palabras clave:** tasa de cambio, flujo de órdenes, redes neuronales artificiales, sistemas neurodifusos.

*Clasificación JEL:* F31, C45, C53.

***Nominal Exchange Rate Microstructure and Dynamics in Colombia: An Approach Using Artificial Neural Networks and Neural Diffusion Systems***

ABSTRACT

This article presents the results of empirical assessments of the nominal exchange rate in Colombia, conducted using linear specifications (distributed lag model) and non-linear specifications (artificial neural networks and neural diffusion systems). All of the econometric specifications employ *order flow* as the determinant for the exchange rate, the most representative variable for models that attempt to explain the nominal exchange rate, supported by the exchange market microstructure. The forecast error measurements used to assess the predictive ability of the models evidenced that linear and non-linear models had an advantage over a random path models, attributable to the order flow. Non-linear models were somewhat superior to the linear model in terms of the root mean square error, Theil's coefficient, and the percentage of predicted correct changes, but the difference was so slight that it does not enable confirming the hypothesis that the exchange rate generation process is non-linear.

**Key Words:** exchange rate, order flow, artificial neural networks, neural diffusion systems.

*JEL Classification:* F31, C45, C53

## Introducción

Investigar las causas que determinan el comportamiento de la tasa de cambio y predecir su evolución es un tema fundamental para la política cambiaria que provee información útil a los inversionistas para la distribución de sus recursos y la cobertura del riesgo cambiario de las firmas. El tratamiento más común es que los modelos que buscan explicar el nivel que alcanza la cotización de una divisa en un momento dado lo hagan con base en las relaciones entre variables macroeconómicas. Este enfoque macroeconómico ha experimentado diversas etapas, desde las *teorías tradicionales de la paridad del poder adquisitivo y la teoría de la balanza de pagos* hasta la *teoría de activos*, que hace hincapié en el movimiento de los flujos financieros para explicar el movimiento del tipo de cambio. La *teoría de la paridad descubierta de interés*, los *modelos monetarios de precios rígidos y flexibles*, y el *modelo de equilibrio de cartera* forman parte de dicha corriente.

A pesar de lo intuitiva, la teoría económica sobre la tasa de cambio entró en crisis a partir de los años ochenta, cuando algunos economistas sugirieron que los modelos macroeconómicos fallaban empíricamente; se planteó así la pregunta respecto a si las principales determinantes de la tasa de cambio eran las variables macroeconómicas. Jalil y Misas (2007, p. 143) advierten que las dinámicas de los tipos de cambio en los mercados de divisas más importantes tienen una baja correlación con los balances comerciales, debido a que el mercado de bienes y servicios es una fracción muy pequeña del total de transacciones de moneda. La teoría de los mercados

de activos tampoco se ha podido corroborar empíricamente. Meese & Rogoff (1983) concluyeron que el paseo aleatorio tenía mejor capacidad que los modelos monetarios y de equilibrio de cartera para explicar el comportamiento de diferentes tasas de cambio nominales, aun utilizando información contemporánea de las variables explicativas en sus pronósticos.

Una alternativa de modelación reciente se fundamenta en los modelos de microestructura, los cuales analizan cómo la estructura de negociación del mercado modela las reglas del comportamiento de los precios. El mayor sustento de esta propuesta se encuentra en los hallazgos de Evans & Lyons (1999), quienes modelaron la tasa de cambio incorporando el flujo de órdenes, una medida de la presión a la compra-venta de divisas en el mercado cambiario. Evans y Lyons probaron empíricamente que el tipo de cambio nominal de corto plazo y el flujo de órdenes están fuertemente correlacionados.

Otra posibilidad es que la inconsistencia empírica se deba a la no linealidad del proceso generador de la tasa de cambio, el cual por lo común es aproximado mediante especificaciones lineales, lo que causaría los pronósticos incorrectos (Baillie y McMahon 1989, citados por Kuan y Liu, 1995).

Gradojevic & Yang (2000) sintetizaron las dos últimas propuestas en un modelo en el que utilizaron el flujo de órdenes como variable determinante para el pronóstico de la tasa de cambio diaria del dólar canadiense y del americano en un modelo de redes neuronales artificiales para atacar las no linealida-

des. Ellos encontraron que su modelo podía predecir mejor los movimientos de la tasa de cambio que la caminata aleatoria.

Jalil y Misas (2007) compararon pronósticos provenientes de un modelo de redes neuronales con los de un modelo lineal tradicional ARIMA. Los pronósticos se evalúan con respecto a medidas tanto simétricas como asimétricas (estas últimas permiten diferenciar los errores dependiendo no solamente de su magnitud sino de su signo). Los autores concluyeron que de acuerdo con las funciones de pérdida asimétricas, los modelos no lineales tienen una mejora considerable en capacidad de pronóstico con respecto a los modelos lineales.

Todo lo anterior motiva a que en este artículo se trate de explicar el comportamiento de corto plazo de la tasa de cambio nominal mediante modelos no lineales de redes neuronales y sistemas neurodifusos, en los cuales lo que determina la tasa de cambio es un híbrido entre la aproximación macroeconómica tradicional y el flujo de órdenes, la variable más representativa de los modelos que tratan de explicar la tasa de cambio nominal, apoyados en la microestructura del mercado cambiario. La competencia de los modelos no lineales se evalúa comparando sus medidas de error de pronóstico con las de un modelo dinámico lineal y el modelo de caminata aleatoria.

## 1. Modelos para la predicción del tipo de cambio

Siguiendo a Evans y Lyons (1999), los modelos para la determinación tipo de cam-

bio pueden ser agrupados en dos grandes enfoques: los modelos *macroeconómicos tradicionales* y los modelos más recientes con *fundamento en la microestructura del mercado cambiario*. Esta clasificación se apoya en el papel que tiene la dinámica de la negociación en la determinación de la tasa de cambio.

### 1.1 Modelos macroeconómicos tradicionales

La teoría económica para la determinación de la tasa de cambio tiene su origen en la *paridad del poder adquisitivo* atribuida originalmente a David Ricardo, pero desarrollada por Gustav Cassel (1916). Esta teoría se basa en la ley del precio único, que sostiene que los bienes similares deben tener el mismo precio en todos los mercados, lo que implica que es indiferente adquirir el bien en un país u otro cualquiera. Por tanto, el tipo de cambio corresponde a la razón entre los precios domésticos ( $P$ ) e internacionales ( $P^*$ ), expresado en su moneda respectiva para la misma cesta de bienes.

En el registro histórico a continuación aparece el enfoque de la *balanza de pagos* que tiene origen en Meade (1951), en el cual el tipo de cambio se ajusta para equilibrar los ingresos y pagos resultantes del comercio internacional de bienes, servicios y activos. Las predicciones del modelo son que la moneda interna se deprecie (aprecie) si disminuye (aumenta) el precio relativo de los bienes y servicios externos ( $P^*/P$ ), si aumenta (disminuye) la renta interna o si disminuye (aumenta) la renta externa. Otros desarrollos que forman parte de esta misma línea son los de

Mundell (1962) y Fleming (1963), que extienden el modelo con la incorporación de los tipos de interés interno  $i$  y externo  $i^*$ .

La fuerte volatilidad de las monedas después de la caída del sistema de Bretton Woods y el incremento en los movimientos de capitales provocó que el enfoque teórico de la balanza de pagos dejara de ser eficaz y se buscaran nuevas alternativas. Así, surgen los *modelos de activos*, que hacen hincapié en el movimiento de los flujos financieros para explicar el movimiento del tipo de cambio. Forman parte de esta corriente la *teoría de la paridad descubierta de los tipos de interés*, los *modelos monetarios de precios rígidos y flexibles* y el *modelo de equilibrio de cartera*.

La teoría de la paridad descubierta de interés de Fisher (1930) plantea que si se tiene en cuenta el tipo de cambio, la rentabilidad del inversor internacional estará formada por dos componentes: el tipo de interés y las variaciones del tipo de cambio. Esta rentabilidad debe ser igual entre los dos países, y aquel que ofrezca un menor tipo de interés nominal verá elevado el valor de su moneda, para dar al inversor un beneficio que compense la menor tasa de interés.

En los modelos monetarios el tipo de cambio es el precio del dinero de una economía expresado en términos de la divisa de otro país, y se determina para equilibrar las ofertas y demandas monetarias de ambos países. En el modelo monetario de precios flexibles los aumentos del tipo de interés interno son originados por un aumento de la tasa de inflación interna esperada. Por la paridad del poder adquisitivo, el aumento de la tasa de

inflación interna esperada se traduce también en un aumento de la depreciación esperada de la moneda interna. La depreciación esperada de la moneda provoca una disminución de la demanda de la moneda en el mercado de divisas, lo que conduce a la depreciación inmediata de la moneda doméstica. En el modelo de precios rígidos los aumentos del tipo de interés interno son originados por un descenso de la oferta monetaria nominal (si los precios son fijos, la oferta real disminuye); el aumento de la tasa de interés atrae capital externo que aprecia la moneda. Las mayores aportaciones a los modelos monetarios de precios flexibles vienen de Frenkel (1976) y Bilson (1978), y la principal referencia de los modelos monetarios de precios rígidos es Dornbusch (1976).

Los modelos de equilibrio de cartera se fundamentan en que en la determinación del tipo de cambio no solo incide el mercado monetario, también el mercado de bonos y el mercado de bienes. El tipo de cambio (dada la renta y la restricción de la riqueza financiera) se determina en función del equilibrio de los mercados de dinero y bonos. Los cambios en las preferencias de los agentes, respecto a las tenencias de activos en diferentes monedas, provocan una redistribución de la riqueza entre países, modificando las demandas relativas de activos y alterando así el tipo de cambio, lo que se conoce como efecto cartera. Una referencia de este modelo es Branson (1976).

Generalizando, los modelos macroeconómicos toman la forma:

$$\Delta P_t = f(\Delta i, \Delta m, \dots) + e_t \quad (1)$$

Donde “ $\Delta P_t$ ” es el cambio en logaritmo de la tasa de cambio nominal entre dos periodos. Las variables explicativas en (1) incluyen los cambios en las tasas de interés nominal doméstica y extranjera “ $i$ ”; en la oferta monetaria “ $m$ ”, y otros determinantes macroeconómicos. Los cambios en estas variables son información pública que incide en la cotización de la tasa de cambio.

### 1.2 Modelos con fundamento en la microestructura del mercado cambiario

La teoría de la microestructura de los mercados de activos financieros es el estudio de los procesos y resultados de la negociación de activos financieros bajo reglas de negociación explícitas en sus sesiones diarias (Marín & Rubio, 2001). Ya que la divisa puede ser considerada un activo financiero, la microestructura del mercado cambiario tiene que ver con la configuración del mercado cambiario, la asimetría de la información, la heterogeneidad de los participantes y las reglas de negociación de dicho mercado.

Según la microestructura del mercado el origen inmediato de los cambios de corto plazo<sup>1</sup> de la tasa de cambio son los cambios a corto plazo de las demandas y expectativas de los agentes. Esta información sobre quién está demandando en la economía y cuánto, de un día a otro, se descubre observando el mercado. Mucha información está dispersa por toda la economía, y cada agente conoce sólo

su pequeño pedazo de la demanda agregada. El papel del mercado es agregar esa información que existe entre los agentes y así formar un agregado que refleja las necesidades y los cambios en las expectativas de la economía a corto plazo.

La contribución fundamental de los modelos de microestructura en la comprensión de la dinámica de la tasa de cambio es señalar el papel del mercado y la presión que los operadores del mercado sienten cuando están fijando precios (Barreiro, 2004). Esta presión se mide a través de la variable llamada flujo de órdenes, una variable que cuantifica tanto la cantidad que es transada en el mercado como la dirección en la cual esta cantidad se está moviendo. El signo lo dicta la dirección en la cual presiona el agente que inicia la transacción. Esta dirección es importante, porque el precio responderá en esta dirección a medida que el precio fijado obtiene información de la orden.

La mayoría de los modelos de microestructura son variaciones de la siguiente especificación:

$$\Delta P_t = g(\Delta x, \Delta I, \dots) + n_t \quad (2)$$

Donde “ $\Delta P_t$ ” es la tasa de variación del tipo de cambio entre dos períodos. Las variables independientes en la función  $g(\Delta x, \Delta I, \dots)$  incluyen el flujo de órdenes “ $\Delta x$ ”, el cambio en la posición neta de los operadores del mercado (o inventario) “ $\Delta I$ ” y otros determinantes macroeconómicos denotados por “ $n_t$ ”.

El flujo de órdenes puede tomar valores positivos o negativos, ya sea por compras del

<sup>1</sup> Los modelos con fundamento en la microestructura del mercado tienen una periodicidad más alta que los modelos macroeconómicos tradicionales, usualmente diaria.

público (+) a la oferta de los operadores, o ventas a la demanda de los mismos (-). Se utiliza la convención que un  $\Delta x$  positivo es una compra neta, haciendo que la relación teórica entre  $\Delta P_t$  y  $\Delta x$  sea positiva, ya que las compras netas de moneda extranjera hacen subir el tipo de cambio. La relación entre  $\Delta P_t$  y  $\Delta I$  recoge el efecto en los precios por el manejo de inventario. Este efecto surge cuando un operador ajusta sus precios para controlar la fluctuación de su inventario respecto a un valor determinado. Por ejemplo, si un operador tiene una posición más larga que la deseada, este puede ajustar su demanda y ofrecer a la baja para inducir la compra de los clientes.

La dinámica entre el tipo de cambio y el flujo de órdenes puede abordarse desde tres hipótesis diferentes, dependiendo de la sincronización entre el flujo de órdenes y el ajuste del tipo de cambio. Si en el modelo existe concurrencia en la realización del tipo de cambio y el flujo de órdenes, pero la causalidad ocurre del flujo de órdenes al tipo de cambio, entonces el modelo obedece a la hipótesis *de ajuste del tipo de cambio por presión del mercado*. Los modelos en los cuales el flujo de órdenes precede al tipo de cambio están basados en la hipótesis *de ajuste del tipo de cambio por anticipación*. Si el flujo de órdenes contiene rezagos de la tasa de cambio, entonces el modelo está basado en la hipótesis *de ajuste del tipo de cambio por retroalimentación*. La sustentación de las diferentes hipótesis puede ser consultada en Azañero (2003).

## 2. Redes neuronales artificiales

Las redes neuronales artificiales (RNA) son en esencia un mecanismo de inferencia esta-

dística no paramétrica inspirado en el sistema nervioso biológico y que tratan de simular el proceso de aprendizaje humano (Montenegro, 2001). Swanson y White (1995) definen las RNA como una clase de modelos no lineales flexibles desarrollados por científicos cognitivos.

En el contexto de la regresión en estadística las RNA pueden asociarse al siguiente planteamiento: *existe una función desconocida a priori, “ $f(x)$ ” con cierto componente estocástico. El proceso de aprendizaje del modelo de RNA consiste en calcular un estimador de la función desconocida, “ $f(x; \Theta) \equiv \hat{f}(x)$ ”, siendo “ $\Theta$ ” el vector de parámetros desconocidos y “ $x$ ” el conjunto de datos observados. El modelo neural definido es por lo tanto un estimador no paramétrico de la esperanza matemática de “ $y$ ” condicionada a “ $x$ ”, es decir  $E(y|x)$ .*

“Una red neural se configura para una aplicación específica de forma tal que el reconocimiento de patrones y la clasificación de información se alcanza a través de un proceso de aprendizaje durante el cual la red modifique sus conexiones “ $\Theta$ ” en respuesta a la información de entrada o estímulo” (Jalil y Misas, 2007). Los algoritmos de aprendizaje se clasifican como supervisados y no supervisados.

El aprendizaje supervisado se puede describir como un algoritmo que modifica las conexiones de la red, a través de la orientación de un maestro externo o supervisor que ajusta de forma iterativa los pesos de la red, en proporción a la diferencia existente entre la salida actual de la red y la salida deseada,

con el objetivo de minimizar el error actual de la red. Por tanto, en todo algoritmo supervisado se requiere conocer la salida deseada. En este artículo se emplea el algoritmo de aprendizaje supervisado denominado *Backpropagation*,<sup>2</sup> desarrollado en Rumelhart, Hinton y Williams (1986), para la estimación de las conexiones de la red perceptrón multicapa y del sistema neurodifuso.

El objetivo del aprendizaje no supervisado (o auto-organizado) es categorizar los datos que se introducen en la red. De esta forma, las informaciones similares son clasificadas en la misma categoría y por tanto activan la misma neurona de salida. Las clases o categorías son creadas por la propia red a través de las correlaciones entre los datos de entrada.

A diferencia de lo que sucede en el aprendizaje supervisado, en el no supervisado no existe ningún maestro externo que indique si la red neuronal está operando correcta o incorrectamente, pues no se dispone de ninguna salida objetivo hacia la cual la red neuronal deba tender. Así, durante el proceso de aprendizaje la red debe descubrir por sí misma rasgos comunes, regularidades, correlaciones o categorías en los datos de entrada, e incorporarlos a su estructura interna de conexiones (Montaño, 2002b). El algoritmo de aprendizaje *Kohonen's Feature Maps* de Kohonen (1988) es usado aquí para encontrar el número de funciones miembro del sistema

neurodifuso y sus respectivas medias y varianzas. Dicho algoritmo está asociado a la red denominada mapas auto-organizados o SOM (*Self-Organizing Maps*) propuesta por Kohonen (1982a, 1982b).

## 2.1 Selección de modelos neuronales

El procedimiento de selección de un modelo neural involucra tres pasos: escoger la forma funcional idónea (especificación del modelo), estimar los parámetros y establecer los criterios de ajuste para la evaluación modelo. La especificación consiste en definir los componentes del modelo: conjunto de modelos neuronales disponibles y arquitectura (número de *inputs* “*p*” y número de neuronas en la capa oculta “*Q*”). En cuanto al conjunto de modelos neuronales disponibles, este trabajo se concentra en dos tipos: la *red perceptrón multicapa* (MLP), la red más común y con mayor éxito en el pronóstico de series de tiempo, y los *sistemas neurodifusos (NFS)*, un híbrido entre las RNA y la lógica difusa.

Los aspectos restantes de la identificación del modelo y la estimación de los parámetros se tratarán en los siguientes numerales; sin embargo, vale la pena anotar que las RNA presentan tres limitaciones (véase Min Qi, 1996, pp. 537-538). En primer lugar, no existe ninguna teoría formal para determinar la estructura óptima de un modelo neuronal; así, aspectos como la determinación del número adecuado de capas, el número de neuronas en cada capa oculta, etcétera, deben decidirse en muchos casos de manera heurística. En segundo lugar, no hay un algoritmo óptimo que garantice el mínimo global en la superficie de

<sup>2</sup> La técnica fue desarrollada inicialmente por Paul Werbos (1974) a mediados de los años setenta, y después independientemente redescubierta por varios grupos de investigadores (Le Cun, 1985; Parker, 1985; Rumelhart, Hinton y Williams, 1986). Es por tanto un caso de descubrimiento múltiple (véase Montaño, 2002a).

error cuando esta presenta mínimos locales. Por último, las propiedades estadísticas de las RNA generalmente no están disponibles y por tanto no se puede llevar a cabo ninguna inferencia estadística con garantías.

El ajuste de los modelos se determina mediante medidas de error de pronóstico. Aquí se emplean la *raíz del error cuadrado medio (RMSE)* y el *índice de Theil*.

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2} \quad (3)$$

$$THEIL = \sqrt{\frac{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2}{\frac{\sum_{t=1}^T y_t^2}{T} + \frac{\sum_{t=1}^T \hat{y}_t^2}{T}}} \quad (4)$$

“*T*” es el número de observaciones, “*y<sub>t</sub>*” el valor estimado por el modelo y “*y<sub>t</sub>*” es el valor observado.

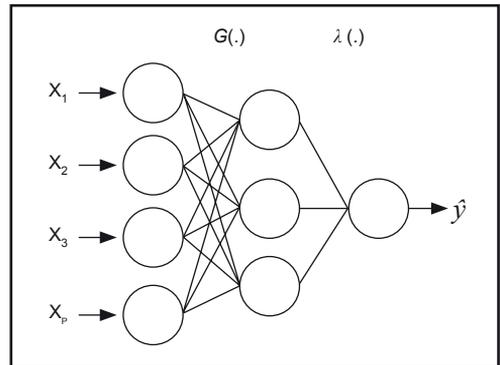
El otro criterio que permite determinar la bondad de los modelos propuestos tiene que ver con su capacidad para determinar el signo correcto de la dirección del cambio diario de la tasa de cambio. Si se considera la divisa un activo financiero, puede ser de tanto interés para los inversores o el Gobierno predecir su revaluación o devaluación como obtener pronósticos próximos a los valores observados.

### 2.2 Red neural perceptrón multicapa (MLP)

Esta es una red de propagación hacia adelante construida por capas que tiene como unidad

básica el Perceptrón; este nombre se debe a que su función de activación es una función sigmoide. La arquitectura de la red comprende una capa de entrada, una o más capas ocultas y una capa de salida. No obstante, por el teorema de aproximaciones universales (Funahashi, 1989) se sabe que una red de una sola capa oculta, con el suficiente número de neuronas, puede actuar como una función de aproximación de cualquier función continua con exactitud satisfactoria. Este caso especial de la red MLP recibe el nombre de *Single Hidden Layer Network (SHLN)* y se muestra el gráfico 1.

Gráfico 1  
Red Neural SHLN



Fuente: elaboración propia.

La forma funcional de la red es:

$$\hat{y} = \lambda \left[ \sum_{h=1}^Q c_h G \left( \sum_{i=1}^P w_{hi} x_{ti} + \theta_h \right) + d \right] \quad (5)$$

Donde “*y<sub>t</sub>*” es el valor de salida de la red en el periodo *t*, y “*G(.)*” y “*λ(.)*” representan la función de activación.

La función de activación comúnmente más usada es la logística, por sus características

de derivación y se recomienda para problemas de predicción, aunque se puede trabajar con cualquier otra función de activación que sea diferenciable (Llano, Hoyos, Arias, Velásquez, 2007). “De cualquier forma no existe un criterio estándar para la elección de la función y comúnmente la selección de funciones de activación se realiza de acuerdo con el problema y a criterio del investigador, en ocasiones por ensayo y error” (Llano *et al.*, 2007, p. 1). Siguiendo la tendencia común, y con las salvedades reseñadas, en esta aplicación se emplea la función logística:

$$f(z) = \frac{1}{1+e^{-z}} \quad (6)$$

Esta función limita la señal de salida al intervalo [0,1], ya que  $f(-\infty) = 0$  y  $f(\infty)=1$ .

Los parámetros  $\{w,c\}$  y  $\{\theta,d\}$  corresponden a los pesos y umbrales de la red respectivamente. Dichos parámetros son ajustados mediante el algoritmo de aprendizaje.

### 2.2.1 Estimación de los parámetros con Backpropagation (BP)

El algoritmo BP es en la práctica el procedimiento más utilizado en el entrenamiento de las redes neurales MLP. Como en el caso de un modelo de regresión lineal, la estimación de los parámetros está vinculada a alguna función objetivo, como la minimización de suma de los cuadrados de los errores “S”:

$$\min_{\Theta} S(\Theta) = \sum_{t=1}^T [y_t - f(x_t, \Theta)]^2 \quad (7)$$

“T” es el tamaño del set de entrenamiento en (7).

La regla de aprendizaje del algoritmo BP tiene como objetivo minimizar el error de salida de la red al cuadrado<sup>3</sup>:  $E_t = \frac{1}{2} * (y_t - \hat{y}_t)^2$  pero en este proceso también se logra la minimización planteada en la ecuación (7).

Para empezar, se requiere inicializar el vector de pesos “ $\Theta \equiv \{w, \theta, d, c\}$ ” llenando sus posiciones con valores aleatorios. En cada iteración “t” el algoritmo actualiza el vector de pesos en alguna dirección, obedeciendo una regla de aprendizaje “ $\Delta^{(t)}$ ”:

$$\hat{\Theta}^{(t+1)} = \phi \hat{\Theta}^{(t)} + (1-\phi)\eta \Delta^{(t)} \quad (8)$$

Los métodos de descenso como BP generan una dirección de búsqueda  $\Delta^{(t)}$  tal que un ligero movimiento en esta dirección, desde un punto inicial  $\hat{\Theta}^{(t)}$  hace decaer el valor de la función objetivo. Del cálculo se sabe que la dirección más rápida de descenso es la del negativo del gradiente:  $\Delta^{(t)} = -\nabla E_t(\hat{\Theta}^{(t)})$ .

En la expresión (8) el parámetro “ $\eta$ ” corresponde a la tasa de aprendizaje<sup>4</sup> y “ $\phi$ ” es el factor de momentum.<sup>5</sup>

El algoritmo que se implementó en este trabajo permite que la tasa de aprendizaje varíe de acuerdo con el rendimiento que presente la red para cada dato de entrenamiento; si el error disminuye, la regla de aprendizaje sigue la dirección correcta y se puede ir más rápido incrementando la tasa de aprendizaje;

<sup>3</sup> El factor 1/2 aparece sólo por conveniencia.

<sup>4</sup> Es un valor en el intervalo [0,1] y tiene como propósito atenuar el cambio en el vector de pesos para evitar saltarse el mínimo de la función objetivo.

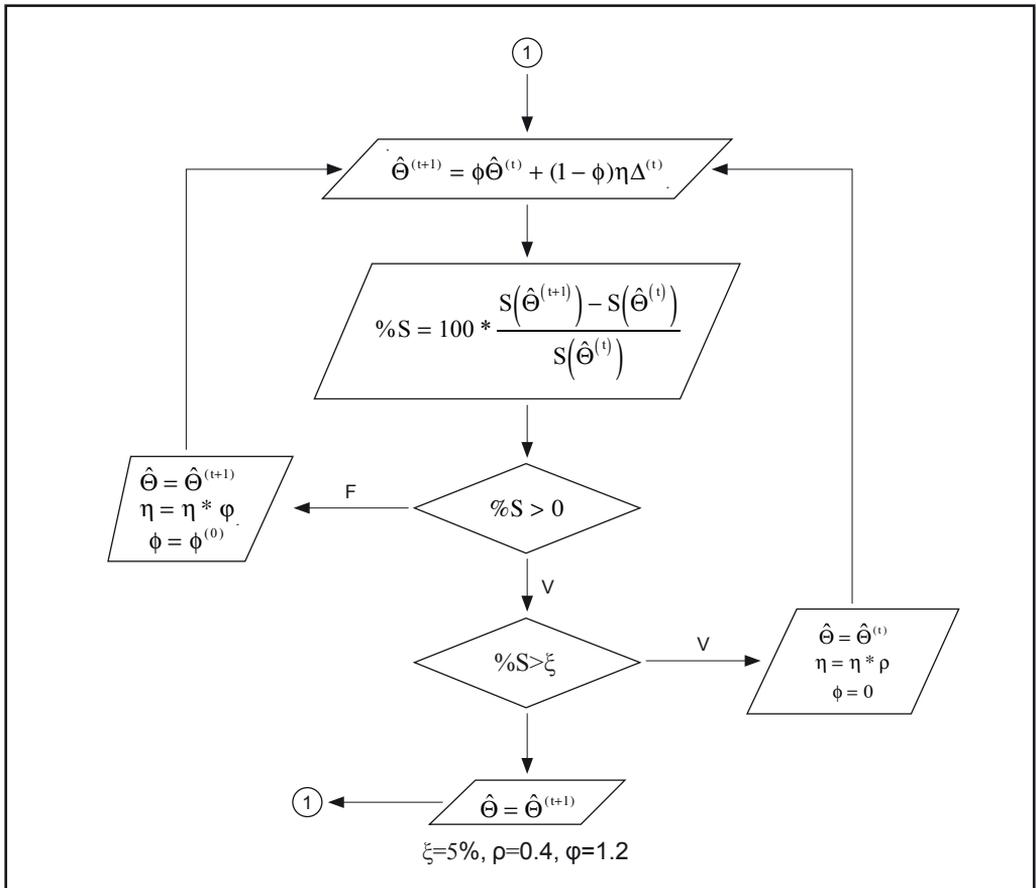
<sup>5</sup> Trata de evitar oscilaciones no deseadas en el proceso de búsqueda del mínimo global.

si el error aumenta, es necesario disminuir la tasa de aprendizaje, esto de acuerdo con las reglas ilustradas el gráfico 2.

La práctica común es detener el algoritmo BP cuando la función objetivo alcanza su mínimo o su cambio durante varios ciclos es nulo. En los métodos de descenso se recomienda ejecutar el algoritmo repetidas veces con distintos valores iniciales para “ $\hat{\theta}$ ” y quedarse

con aquellos que conducen a un menor valor de  $S(\hat{\theta})$ . Esto se debe a que el algoritmo puede quedar atrapado en un mínimo local. Para tratar de superar este inconveniente se empleó el algoritmo BP con tasa de aprendizaje y momentum variable, ya que esta modificación le permite a BP escapar de los mínimos locales y continuar su camino hacia el mínimo global.

Gráfico 2  
Diagrama de flujo algoritmo BP



Fuente: elaboración propia.

El proceso de aprendizaje pretende lograr la generalización de la red, es decir, obtener los *outputs* correctos para un conjunto de *inputs* que no han sido incluidos en el set de entrenamiento. Un problema común en la modelación de RNA que atenta contra la generalización de la red es la memorización. Esto consiste en que la red graba la entrada-salida del set de entrenamiento, pero la relación no se entiende, por lo que no queda almacenada en los pesos sinápticos. Para asegurar la generalización de la red el modelo debe ser validado. Esto se puede lograr con datos que no forman parte del set de aprendizaje, pero que sirven para medir el ajuste del modelo. A dicho set de datos se les conoce como set de validación.

En este artículo se recurrió a la estrategia de realizar el proceso de validación durante la estimación de los parámetros del modelo para optimizar su generalización, por lo que el vector de pesos resultado del algoritmo de BP corresponde a aquel en el cual se alcanza el mínimo de la medida de ajuste del set de validación.

### 2.2.2 Selección de los inputs y dimensión de la capa oculta

Los mecanismos de elección de las variables de entrada a la red pueden ser de diferente índole. A continuación se presentan tres posibilidades: mediante una regresión de la variable dependiente contra un grupo extenso de variables independientes; la utilización de técnicas multivariantes, como el análisis de componentes principales, y la regresión *stepwise*, que supone una selección secuencial mediante criterios estadísticos.

En cuanto al número de neuronas de la capa oculta, esto representa un problema, ya que si es demasiado grande se puede producir la memorización de la red. Refenes (1995) ha tratado algunos métodos para este propósito, pero lo más frecuente es la reducción paulatina de los parámetros del modelo, eliminando las conexiones redundantes o con menor sensibilidad. Los procedimientos adoptados en este artículo se pueden incluir en esta línea y son tomados de Kaashoek & Van Dijk (1999).

En principio, la regla de Thumb puede servir de guía para fijar un valor inicial para el número de neuronas de la capa oculta ( $H$ ) que evite la sobreparametrización:

$$H \approx \frac{T}{5 * (p + o)} \quad (9)$$

En (9) “ $o$ ” es la dimensión de la capa de salida.

Partiendo de una preselección inicial de los *inputs* y de la dimensión de la capa oculta según la regla de Thumb, se procede a recortar la red eliminando *inputs* y neuronas de la capa oculta con fundamento en los siguientes criterios: contribución incremental de los nodos de la red y el análisis del componente principal de los residuales.

#### 2.2.2.1 Contribución incremental de los nodos de la red

Esta medida se determina mediante la expresión:

$$R^2_{inc} = \frac{(\hat{y}y)^2}{(y'y)(\hat{y}\hat{y})} - \frac{(\hat{y}_{-h}y)^2}{(y'y)(\hat{y}_{-h}\hat{y}_{-h})} \quad (10)$$

Donde “ $y$ ” y “ $\hat{y}$ ” son respectivamente los vectores formados por los valores observados y estimados de la variable dependiente del modelo completo. “ $\hat{y}_{-h}$ ” es el vector de valores estimados de la variable dependiente sin incluir la  $h$ -ésima neurona en la capa de salida (o algún *input*).

La técnica consiste en computar “ $\hat{y}$ ” con el número máximo de neuronas y luego calcular “ $\hat{y}_{-h}$ ” eliminando neuronas de la capa oculta (o *inputs*), pero conservando el mismo vector de pesos, excepto el parámetro “ $d$ ”, el cual debe ser re-ajustado por medio de un nuevo proceso de optimización. Después se calcula “ $R_{inc}^2$ ” de cada una de las neuronas y se eliminan las de menor aporte. El proceso se repite hasta que las neuronas que quedan tengan una contribución incremental significativa.

### 2.2.2.2 Análisis del componente principal de los residuos de la red

Para la capa oculta se define la matriz:

$$E_{-Q} = (e_{-1}, e_{-2}, \dots, e_{-H}, e_{-Q}),$$

$$e_{-h} = y - \hat{y}_{-h} \quad (11)$$

Los componentes principales de  $E_{-H}$  son los eigenvectores ortogonales “ $v$ ” obtenidos a partir de los eigenvalores de la matriz simétrica y definida positiva  $E_{-H}'E_{-H}$ .

El uso de componentes principales es motivado por la siguiente premisa: la cantidad relativa de varianza de cada componente principal es proporcional a su correspondiente eigenvalor. El eigenvector “ $v_{max}$ ” producto del eigenvalor más grande “ $\lambda_{max}$ ” define la combinación lineal de los elementos de “ $e_{-h}$ ”

con mayor varianza. Los elementos de este vector “ $v_{max}$ ” revelan cual nodo puede ser excluido. El nodo “ $h$ ” para el cual su elemento correspondiente en el componente principal “ $v_{max}$ ” es mínimo en valor absoluto puede ser excluido, porque su exclusión del modelo no contribuye mucho al peor caso.

Para decidir si un nodo puede excluirse o no de la red con fundamento en el componente principal “ $v_{max}$ ”, este componente debe tener un peso alto con relación a los demás. El peso “ $w_k$ ” del  $k$ -ésimo componente se determina por la magnitud relativa del correspondiente eigenvalor  $\lambda_{max}$ :

$$W_k = \frac{\lambda_k}{\sum_{k=1}^H \lambda_k} \quad (12)$$

Esta técnica también se emplea para recortar los *inputs* de la red.

## 2.3 Sistema neurodifuso (NFS)

Este sistema predictor inteligente es básicamente una red neural multicapa, que integra los elementos básicos de la lógica difusa tradicional en una red neural. La ventaja radica en que el sistema híbrido resultante reúne la capacidad de aprendizaje de las redes neuronales, con la facilidad de trabajo con información incierta, y su traducción a reglas de los sistemas difusos (Li, Ang, Gray, 1999, p. 181).

A diferencia de las redes neuronales, la información contenida en un sistema neurodifuso es interpretable, y la velocidad de aprendizaje del algoritmo de entrenamiento

es más veloz comparado con el de las redes neuronales (Pan, Liu, Mejabi, 1997, p. 1). Las reglas borrosas pueden extraerse del sistema neurodifuso, modificarse y aplicarse a la predicción (Li *et al.*, 1999, p. 181). El sistema ha probado su mejor desempeño sobre la red neural simple, en algunas aplicaciones como control de procesos, reconocimiento del lenguaje, clasificación de datos, diagnóstico de procesos, aprendizaje de maquinas y en el pronóstico de series de tiempo financieras.

**2.3.1 Especificación de NFS**

La especificación y estimación del sistema neurodifuso desarrollado sigue a Li *et al.* (1999). Como lo muestra el gráfico 3, el sistema esta conformado por cinco capas:

**Capa-1:** los nodos de esta primera capa transmiten los valores de los *inputs*  $u_i^1$  a la segunda capa, sin ninguna transformación. Para el nodo  $i$  en la *capa-1* se tiene lo siguiente:

$$o_i^1 = net_i^1 = u_i^1 \quad \forall_i = 1, 2, \dots, p \quad (13)$$

**Capa-2:** en esta capa a cada *input* se le define cierto número de conjuntos difusos, que pueden ser identificados por etiquetas, por ejemplo valor {alto, medio, bajo} de la variable, y para cada conjunto difuso se define una función de pertenencia al conjunto  $\{\mu_{bajo}(\cdot), \mu_{medio}(\cdot), \mu_{alto}(\cdot)\}$ . El objetivo de los conjuntos difusos es poder agrupar los *inputs* y establecer su grado de pertenencia al grupo de acuerdo con la magnitud de la observación. Las funciones de pertenencia utilizadas son distribuciones normales con rango {0,1}. Para el  $j$ -ésimo nodo se tiene:

$$u_{ij}^2 = u_i^1 \quad (14)$$

$$net_j^2 = - \left( \frac{u_{ij}^2 - m_{ij}^2}{\sigma_{ij}^2} \right)^2 \quad (15)$$

$$o_j^2 = f(net_j^2) = e^{net_j^2} \quad (16)$$

$$\forall_i = 1, 2, \dots, n$$

$m_{ij}^2$  y  $\sigma_{ij}^2$  son respectivamente, la media y la varianza de la función miembro del nodo  $j$ .

**Capa-3:** en esta capa se realiza la operación lógica *AND* sobre las señales que se reciben de la capa-2. Esto consiste en filtrar las señales que se reciben de esa capa y transmitir a la capa-4 únicamente las señales mínimas. Para un nodo en esta capa se tiene:

$$u_{ij}^3 = f(net_i^2) \quad (17)$$

$$net_j^3 = \min_i \{u_{ij}^3\} \quad (18)$$

$$o_j^3 = f(net_j^3) = net_j^3 \quad (19)$$

Para  $i = 1, 2, \dots, n; j = 1, 2, \dots, h$ .

**Capa-4:** los nodos de esta capa realizan la operación lógica *OR* sobre sus señales de entrada. Para un nodo en esta capa se tiene:

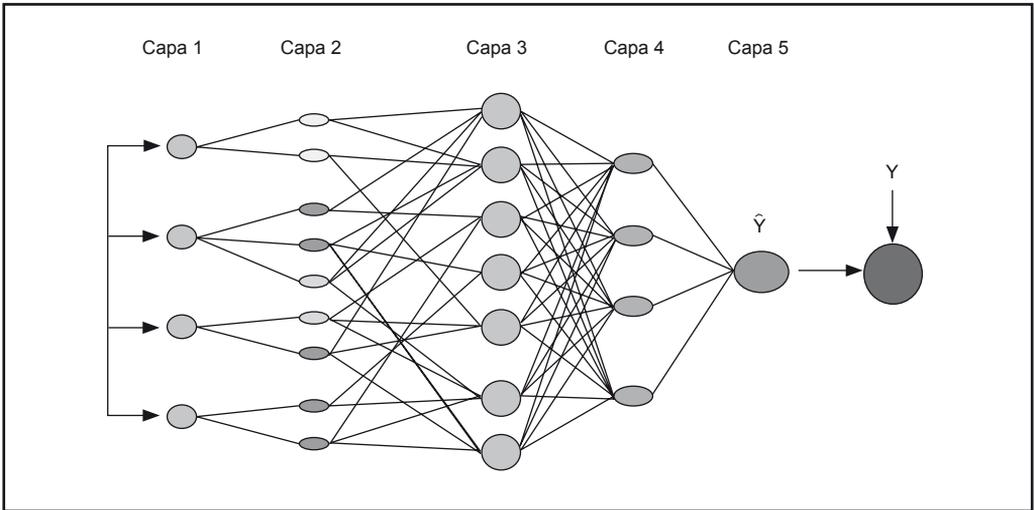
$$U_{ij}^4 = f(net_i^3) \quad (20)$$

$$U_{cj}^4 = \max \{f(net_1^3), f(net_2^3), \dots, f(net_h^3)\} \quad (21)$$

$$o_j^4 = f(net_j^4) = net_j^4 = u_{cj}^4 * r_{ij} \quad (22)$$

Para  $i = 1, 2, \dots, h; j = 1, 2, \dots, m; y c \in [1, 2, \dots, h]$ .

Gráfico 3  
Arquitectura sistema neurodifuso



Fuente: elaboración propia.

El nodo “c” en la capa-3 es el nodo ganador de la operación lógica mín-máx. Los “ $r_{ij}$ ” son los valores de las reglas, que son ajustados mediante un algoritmo de aprendizaje supervisado.

$$\hat{y}_t = \frac{\text{net}^5}{\sum_{j=1}^m \sigma_j^5 * u_j^5} \quad (25)$$

### 2.3.2 Estimación de NFS

La estimación de la red comprende las siguientes subrutinas:

- Algoritmo de aprendizaje auto-organizativo.
- Identificación de las reglas difusas y ajuste de los pesos de las reglas mediante el algoritmo BP.

#### 2.3.2.1 Algoritmo auto-organizativo

El algoritmo de Kohonen (1988) tiene como objetivo agrupar las observaciones de las dis-

**Capa-5:** en esta capa realiza la defusificación de la señal para producir la salida de la red. Para esto se utiliza el método del centro de gravedad de Kosko, en el cual la señal de salida es el centróide de la función miembro. Así, si  $m_j^5$  y  $\sigma_j^5$  son respectivamente la media y la varianza de las funciones miembro de esta capa.

La salida defusificada se calcula de la siguiente manera:

$$U_j^5 = f(\text{net}_j^4); w_j^5 = \sigma_j^5 * m_j^5 \quad \forall_j = 1, 2, \dots, m \quad (23)$$

$$\text{net}^5 = \sum_{j=1}^m w_j^5 * u_j^5 = \sum_{j=1}^m \sigma_j^5 * m_j^5 * u_j^5 \quad (24)$$

tintas variables en alguna de las funciones de pertenencia asignadas a la variable; en este proceso se determinan las medias y las varianzas de dichas funciones de pertenencia. El algoritmo opera de la siguiente manera: para un conjunto de datos  $X_i=(x_1, \dots, x_n)$  de la  $i$ -ésima variable de entrada se asignan arbitrariamente  $k$  valores iniciales a las medias, de forma tal que:

$$\min(x_1, x_2, \dots, x_n) < m_i < \max(x_1, x_2, \dots, x_n) \quad (26)$$

A continuación los datos son agrupados alrededor de las medias iniciales de forma tal que:

$$|x_j - m_c| = \min_i \{x_j - m_c\} \quad (27)$$

$$\forall_i = 1, 2, \dots, k \wedge \forall_j = 1, 2, \dots, n$$

“ $m_c$ ” es la media asociada al dato “ $x_j$ ” y “ $k$ ” es el número de funciones miembro de la  $i$ -ésima variable.

Las medias “ $m_c$ ” son optimizadas mediante el algoritmo:

$$m_c^{(t+1)} = m_c^{(t)} + \alpha^{(t)} [x_j^{(t)} - m_c^{(t)}] \quad (28)$$

$$\text{si } x_j \in m_c$$

$$m_c^{(t+1)} = m_c^{(t)}, \text{ si } x_j \notin m_c \quad (29)$$

“ $\alpha^{(t)}$ ” es una tasa de aprendizaje monotónicamente decreciente  $\in [0, 1]$ .

La varianza de la función de pertenencia puede ser determinada por:

$$\sigma_i = \frac{1}{R} \sqrt{\frac{1}{p_i} \sum_{j=1}^{p_i} (x_j - m_i)^2} \quad (30)$$

$$\forall_i = 1, 2, \dots, k$$

Donde:

“ $\sigma_i$ ” y “ $m_i$ ” son respectivamente la varianza y la media de la  $i$ -ésima función de pertenencia; “ $p_i$ ” es el número de datos agrupados alrededor de la  $i$ -ésima función de pertenencia; “ $R$ ” es el parámetro de traslapo.

El número de funciones de pertenencia de cada variable se determina a prueba y error.

### 2.3.2.2 Identificación de las reglas difusas y ajuste de los pesos

El máximo número posible de reglas de la red esta limitado por  $T(x_1) * \dots * T(x_i) * \dots * T(x_p)$ , donde “ $T(x_i)$ ” es el número de funciones de pertenencia de la  $i$ -ésima variable de entrada. Cada observación del conjunto de datos de entrenamiento quedará asociada a una regla particular del conjunto factible, que será la regla ganadora (regla difusa IF-AND-THEN) de las operaciones lógicas AND y OR realizadas en las capas 3 y 4.

Posteriormente se afinan los pesos de las reglas ganadoras asociados a las distintas observaciones de la muestra de entrenamiento mediante el algoritmo de aprendizaje supervisado BP, para minimizar el error de salida de la red. La expresión para el ajuste iterativo del peso “ $r_{ij}$ ” asociado con la regla representada por la conexión del *nodo*  $i$  de la capa-3 con el *nodo*  $j$  en la capa-4 es:

$$r_{ij}^{(t+1)} = r_{ij}^{(t)} + \eta u_{cj}^4 [y_i - \hat{y}_i]$$

$$\left\{ \frac{\left( \sum_{j=1}^m \sigma_j^5 u_j^5 \right) \sigma_j^5 m_j^5 - \left( \sum_{j=1}^m \sigma_j^5 m_j^5 u_j^5 \right) \sigma_j^5}{\left( \sum_{j=1}^m \sigma_j^5 u_j^5 \right)^2} \right\} \quad (31)$$

Si los pesos de las reglas difusas de algunos valores de las variables de entrada de los pronósticos no han sido ajustadas durante el proceso de aprendizaje (porque a ninguna de las observaciones del conjunto de entrenamiento correspondió esta regla), se debe escoger de forma aleatoria un reemplazo de la regla entre las reglas que tengan un mayor número de funciones miembro en común.

### 3. Modelo para la determinación de la tasa de cambio

El pronóstico del tipo de cambio nominal de corto plazo (un día adelante) propósito de este artículo se realizó bajo la *hipótesis de ajuste del tipo de cambio por anticipación*. La manera de abordar esta hipótesis es a través de un modelo dinámico lineal y de modelos no lineales MLP y NFS. El modelo propuesto incluye variables de los dos enfoques: el macroeconómico tradicional y el flujo de órdenes. Su representación matemática es:

$$\Delta LTCN_t = f(\Delta LTCN, ODF, DIRTER, \Delta EMBI, \Delta LPPE, \Delta LRINT) + \varepsilon_t \quad (32)$$

$$\varepsilon_t \sim iid(0, \sigma^2)$$

Variable dependiente:

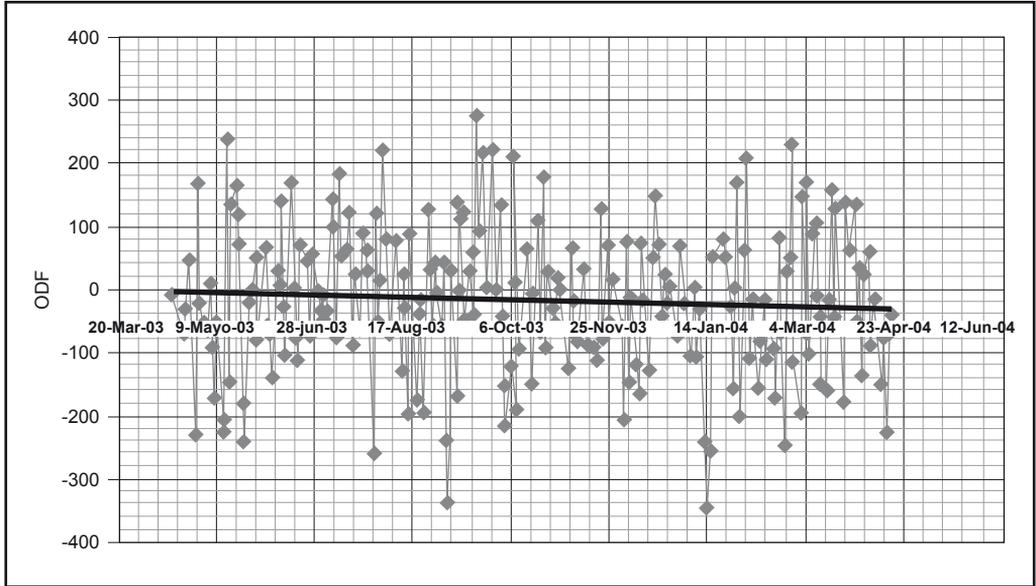
**$\Delta LTCN$ :** Es la primera diferencia del logaritmo de la tasa de cambio nominal. La fuente es el Banco de la República.

Variables explicativas:

- ***ODF***: esta variable se construye mediante la diferencia entre el volumen diario de operaciones propuestas por compradores y el volumen diario de operaciones propuestas por vendedores en el mercado de divisas SET FX<sup>6</sup>. Esto corresponde al exceso de demanda, que puede ser positivo o negativo, lo cual en teoría es lo que presiona los movimientos de la tasa de cambio. El gráfico 4 muestra las magnitudes de los excesos de demanda para el período de estimación.
- ***DINTER***: diferencia entre las tasas de interés interbancarias nacional y extranjera. Esta variable busca recoger el efecto en el tipo de cambio por el manejo de portafolio de los agentes. Corresponden a las tasas de interés interbancaria *overnight* diarias en moneda nacional y moneda extranjera. La fuente de la tasa de interés interbancaria en Colombia es el Banco de la República. Para Estados Unidos, la

<sup>6</sup> Gran parte de las operaciones legales de compra y venta de dólares en Colombia correspondientes al mercado regulado se desarrollan a través del Sistema Electrónico de Transacciones e Información del mercado de divisas SET FX de la Bolsa de Valores de Colombia. Este es un mecanismo electrónico mediante el cual, en las sesiones de negociación, sus asociados pueden ingresar ofertas y demandas, cotizar y/o celebrar entre ellas operaciones, contratos y transacciones relacionadas con la divisa. Las negociaciones se realizan todos los días hábiles de lunes a viernes de 8:00 a.m. a 1:00 p.m.

Gráfico 4  
**Flujo de órdenes en el período de estimación**



Fuente: elaboración propia.

tasa interés *overnight* diaria proviene de los *Fed Funds*.

- $\Delta EMBI$ : esta variable intenta recoger la percepción de los inversionistas del riesgo soberano por títulos financieros emitidos por países emergentes. Como variable proxy se toma el cambio diario del EMBI<sup>7</sup>. Fuente: JP Morgan.
- $\Delta LPPETRO$ : es el diferencial del logaritmo del precio diario del barril de petróleo

en dólares. La fuente es Corfinsura. La inclusión del precio del petróleo se fundamenta en la teoría avanzada de economistas tales como Krugman (1983) que muestra que la volatilidad en la tasa de cambio puede ser capturada por los cambios en el precio del petróleo.

- $\Delta LRINT$ : es el cambio diario del logaritmo de las reservas internacionales. La fuente es el Banco de la República. La inclusión de la variable sigue el enfoque de la balanza de pagos para la determinación del tipo de cambio.

<sup>7</sup> Mide el riesgo país de quince naciones: Argentina, Brasil, Colombia, Ecuador, México, Panamá, Perú, Venezuela, Bulgaria, Corea del Sur, Marruecos, Nigeria, Filipinas, Polonia y Rusia. El riesgo país es un concepto utilizado con frecuencia para expresar la relación riesgo-rendimiento asociada a una inversión en un país determinado.

## 4. Estimación

La muestra<sup>8</sup> consta de un año completo entre el 16 de abril del 2003 y el 15 de abril del 2004. El período de evaluación de los pronósticos comprende las cuarenta últimas observaciones y las restantes son usadas para entrenamiento y validación de los parámetros de las redes.

En un modelo de series de tiempo una muestra pequeña puede generar reservas sobre la capacidad de los parámetros estimados de los modelos para colegir ciertas regularidades como la estacionalidad o el ciclo económico que afectan la demanda de dólares en el mercado cambiario. Es evidente que la posibilidad de involucrar estos eventos, y los impactos que tienen sobre los parámetros estimados de los modelos, depende fundamentalmente del número de repeticiones que se tengan del mismo día a través de los años.

Con la incorporación del flujo de órdenes se puede esperar que los parámetros estimados logren recoger parte de estas regularidades, mitigando así el problema del tamaño de la muestra. Esto se debe a que con la presencia de esta variable se establece una relación directa entre el mercado representado por el flujo de órdenes y la tasa de cambio, por tanto, los cambios en la demanda de dólares son transmitidos directamente al precio de la divisa a través de los parámetros asociados

con el flujo de órdenes, con un menor papel para la repetición del evento.

### 4.1 Estimación modelo dinámico lineal

Con el propósito de preseleccionar las variables a emplear en los modelos no lineales (RNA y NFS) y para comparar sus pronósticos, se estimó un modelo dinámico lineal denominado modelo autorregresivo de rezagos distribuidos (ADL), que tiene forma:

$$\alpha(L)(1-L)^d y_t = \delta + \beta(L) x_t + \varepsilon_t \quad (33)$$

Donde “ $x_t$ ” representa a las variables exógenas incluidas como explicativas que aparecen en la expresión (32). La estimación en e-views del modelo ADL<sup>9</sup> se presenta en el cuadro 1.

Se puede observar en la salida que los signos de los coeficientes de las variables que recogen el flujo de órdenes son los esperados<sup>10</sup> y su significancia en la explicación de la tasa de cambio es evidente.

<sup>8</sup> El tamaño de la muestra fue limitado por el flujo de órdenes. A diferencia de las variables macroeconómicas, la información disponible para crear el flujo de órdenes no está disponible fácilmente, porque no es de dominio público.

<sup>9</sup> Bajo la hipótesis del tipo de cambio por retroalimentación los estimadores del modelo ADL serían sesgados.

<sup>10</sup> Los signos positivos implican que a mayor demanda neta de dólares, la variación del precio de la tasa de cambio es positiva.

Cuadro 1  
**Estimación ADL**

Dependent Variable: ΔLTRM Method: Least Squares Sample(adjusted): 7 198 Convergence achieved after 5 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Intercepto	9.05E-05	8.19E-05	1.104883	0.2706
ODF(-1)	4.85E-06	8.32E-07	5.833572	0.0000
ODF(-2)	6.69E-06	7.99E-07	8.375596	0.0000
ΔLPETRO(-1)	-0.019350	0.009791	-1.976382	0.0496
ΔLPETRO(-5)	0.033091	0.009410	3.516682	0.0005
ΔEMBI(-1)	2.40E-05	1.05E-05	2.292271	0.0230
ΔLTRM(-1)	-0.128701	0.074330	-1.731482	0.0850
R-squared	0.449371	Mean dependent var		-0.000141
Adjusted R-squared	0.431512	S.D. dependent var		0.001650
S.E. of regression	0.001244	Akaike info criterion		-10.50510
Sum squared resid	0.000286	Schwarz criterion		-10.38634
Log likelihood	1015.490	F-statistic		25.16318
Durbin-Watson stat	2.006261	Prob(F-statistic)		0.000000
Inverted AR Roots	-.13			

Fuente: elaboración propia.

### 4.2 Paseo aleatorio (RW)

A partir de los descubrimientos de Meese y Rogoff (1983), RW se ha convertido en un modelo de referencia para la comparación de pronósticos del tipo de cambio. El proceso generador de datos de un paseo aleatorio sin deriva es:

$$y_t = y_{t-1} + \varepsilon_t \text{ donde } \varepsilon_t \sim iid(0, \sigma^2) \quad (34)$$

RW debe su nombre a que avanza dando pasos al azar. A partir del punto donde se encuentre el proceso, digamos “ $y_{t-1}$ ”, dará un paso “ $\varepsilon_t$ ” para llegar a “ $y_t$ ”; allí dará un paso

“ $\varepsilon_{t+1}$ ” para llegar a “ $y_{t+1}$ ” y así sucesivamente, vagando sin rumbo y con muy pocos o ningún cruce por el eje cero.

Si bien el paseo aleatorio no es estacionario, su primera diferencia sí lo es:

$$\Delta y_t = y_t - y_{t-1} = \varepsilon_t \quad (35)$$

Bajo la hipótesis de que el proceso generador del nivel de la tasa de cambio es la caminata aleatoria, el mejor pronóstico de  $\Delta LTCN$  es cero, ya que:

$$\Delta LTCN_t = \varepsilon_t \Rightarrow E[\Delta LTCN_t] = 0 \quad (36)$$

### 4.3 Estimación de modelos no lineales

De forma preliminar, los datos son normalizados en el intervalo [0,1] mediante (37) para evitar problemas de convergencia (ver Riaz, 2000) y atenuar la sensibilidad de las redes neurales a datos que no han hecho parte del proceso de aprendizaje.

$$\text{valor normal} = \frac{\text{valor actual} - \text{valor mínimo}}{\text{valor máximo} - \text{valor mínimo}} \quad (37)$$

#### 4.3.1 Estimación red MLP

Se parte de la especificación  $MLP(7, 4, 1)$ , en donde “7” es el número preliminar de *inputs* según el modelo ADL, la cantidad inicial de neuronas de la capa oculta responde a la aplicación de la regla de Thumb =  $150/(5*8) \approx 4$ , y 1, por la única neurona de la capa de salida. La estimación del vector de pesos de la red se efectuó mediante BP.

En seguida se muestra la aplicación de las técnicas de contribución incremental y análisis del componente principal de los residuos de los nodos de la red, para la selección definitiva de sus *inputs* y la determinación de la dimensión de la capa oculta. Los resultados son los que se muestran en los cuadros 2, 3, 4 y 5.

Cuadro 2

#### Componentes principales - neuronas capa oculta

EIGENVECTORES DE $E'_{-H}E_{-H}$				
	0,3787	0,2240	0,7234	-0,5321
	0,8335	-0,3806	-0,3889	-0,0957

Continúa

EIGENVECTORES DE $E'_{-H}E_{-H}$				
	0,1389	0,8437	-0,4794	-0,1978
	0,3776	0,3053	0,3092	0,8177
$\lambda$ :	19,5546	1,5011	0,1769	0,0245
$W_k$ :	91,99%	7,06%	0,83%	0,11%

Fuente: elaboración propia.

Cuadro 3

#### Contribución incremental - neuronas capa oculta

Neurona excluida	$R^2_{inc}$	SIC*	$V_{Max}$
$H_1$	0,00160	-3,9189	0,3787
$H_2$	0,00540	-2,3838	0,8335
$H_3$	0,00700	-4,6137	0,1389
$H_4$	0,00050	-3,9251	0,3776

\* Criterio de información de Schwartz

Fuente: elaboración propia.

Cuadro 4

#### Componentes principales - inputs

EIGENVALORES DE $E'_{-I}E_{-I}$						
	0,0987	0,5746	-0,1233	0,3530	0,7212	-0,0127
	0,5683	-0,0742	0,7675	0,2837	-0,0258	0,0363
	0,7101	-0,3407	-0,6038	0,1216	0,0117	0,0109
	0,2438	0,3641	9,44e-004	-0,3074	-0,1874	-0,8236
	0,0929	0,5838	-0,1643	0,3110	-0,6526	0,3178
	0,3080	0,2737	0,0644	-0,7674	0,1347	0,4681
$\lambda$ :	14,0912	2,8254	0,6076	0,0969	0,0159	7,93e-003
$W_k$ :	79,86%	16,01%	3,44%	0,54%	0,09%	0,04%

Fuente: elaboración propia.

Cuadro 5  
**Contribución incremental - inputs**

INPUT excluido	$R^2_{inc}$	SIC	$V_{Max}$
$\Delta LTCN_{t-1}$	0,00050	-4,9155	0,0987
$ODF_{t-1}$	0,00840	-3,4147	0,5683
$ODF_{t-2}$	0,01530	-2,975	0,7101
$\Delta EMBI_{t-1}$	0,00010	-4,8057	0,2438
$\Delta LPPETRO_{t-1}$	0,00060	-4,8992	0,0929
$\Delta LPPETRO_{t-5}$	0,00110	-4,5342	0,308

Fuente: elaboración propia.

4.3.1.1 *Análisis*

1. El cuadro 3 muestra que la contribución incremental de la neurona “H<sub>4</sub>” es la más baja. Sin embargo, “H<sub>3</sub>” tiene el menor valor absoluto en  $v_{max}$  y su ausencia tiene el menor efecto sobre SIC. A pesar de la ambigüedad, se puede deducir que sobra una neurona de la capa oculta.
2. Según el cuadro 5 la contribución incremental más alta la tienen las variables de flujo de órdenes  $ODF_{t-1}$  y  $ODF_{t-2}$ . Lo contrario ocurre con las variables  $\Delta LTCN_{t-1}$ ,  $\Delta EMBI_{t-1}$ , y  $\Delta LPPETRO_{t-1}$  que tienen los aportes incrementales más bajos, los menores valores absolutos en  $v_{max}$  y son los *inputs* cuyas ausencias tienen menor efecto sobre SIC. Por tanto, estas últimas variables son excluidas de la red.
3. Posteriormente, una nueva corrida de BP mostró la necesidad de excluir otra neu-

rona de la capa oculta y eliminar el *input*  $\Delta LPPETRO_{t-5}$ . En definitiva, el análisis confirma la importancia de las variables rezagadas del flujo de órdenes en la determinación de la tasa de cambio.

4.3.2 *Estimación (NFS)*

Ya que NFS es un modelo no lineal tal como MLP, parece lógico emplear los mismos *inputs* en ambos casos. Definidos los *inputs*, se procede a asignar el número de funciones de pertenencia a cada variable, para lo que se recomienda la inspección visual de cada serie de tiempo. Los centros se pueden asignar tomando de forma aleatoria valores del conjunto de datos de entrenamiento de cada variable. En la práctica se usaron cuatro funciones de pertenencia y se repartieron los centros entre el intervalo [0,1] (cuadros 6 y 7). En seguida el algoritmo auto-organizativo las perfeccionó (cuadros 8 y 9). Los gráficos 5 y 6 muestran las funciones de pertenencia para “ $\Delta LTCN$ ” antes y después de su optimización.

Cuadro 6  
**Centros no optimizados  
 funciones de pertenencia**

$\mu(\cdot)$	Centros		
	$ODF_{t-1}$	$ODF_{t-2}$	$\Delta LTCN_t$
Bajo	0,25	0,25	0,25
Medio	0,45	0,45	0,45
Alto	0,65	0,65	0,65
Muy alto	0,80	0,80	0,80

Fuente: elaboración propia.

Cuadro 7

**Desviaciones estándar  
funciones de pertenencia no optimizadas**

$\mu(\cdot)$	Desviaciones estándar		
	$ODF_{t-1}$	$ODF_{t-2}$	$\Delta LTCN_t$
<i>Bajo</i>	0,1808	0,1808	0,3536
<i>Medio</i>	0,0886	0,0882	0,0963
<i>Alto</i>	0,0949	0,0957	0,1278
<i>Muy Alto</i>	0,1749	0,1749	0,1728

Fuente: elaboración propia.

Cuadro 9

**Desviación estándar  
funciones de pertenencia optimizadas**

$\mu(\cdot)$	Desviaciones estándar		
	$ODF_{t-1}$	$ODF_{t-2}$	$\Delta LTCN_t$
<i>Bajo</i>	0,3117	0,3117	0,2988
<i>Medio</i>	0,0823	0,1008	0,1333
<i>Alto</i>	0,1038	0,1049	0,0968
<i>Muy Alto</i>	0,1839	0,1839	0,132

Fuente: elaboración propia.

Cuadro 8

**Centros optimizados  
funciones de pertenencia**

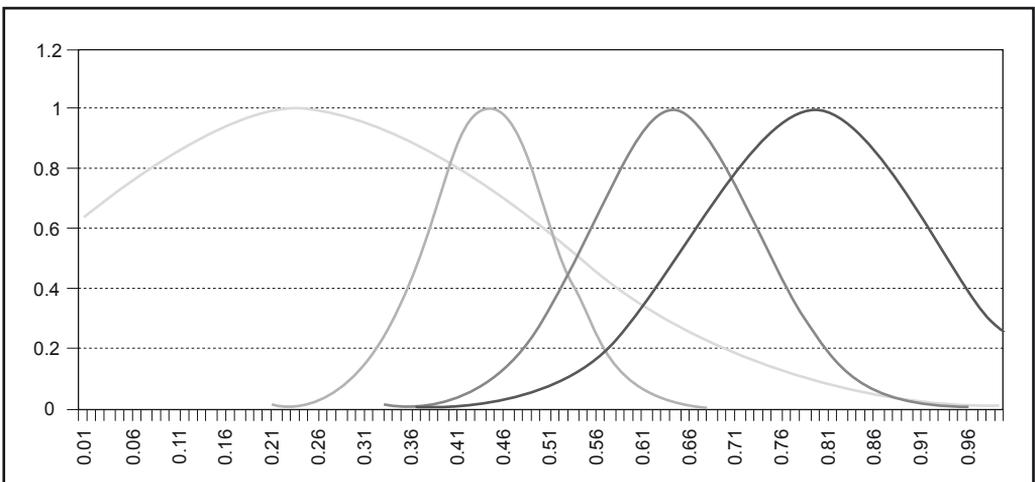
$\mu(\cdot)$	Centros		
	$ODF_{t-1}$	$ODF_{t-2}$	$\Delta LTCN_t$
<i>Bajo</i>	0,1218	0,1218	0,043
<i>Medio</i>	0,4381	0,4629	0,3863
<i>Alto</i>	0,5409	0,5408	0,604
<i>Muy Alto</i>	0,7327	0,7327	0,7189

Fuente: elaboración propia.

La dimensión del conjunto de reglas corresponde al producto del número de funciones de pertenencia de cada *input*:  $4 \times 4 = 16$ . Cuando la red es alimentada, a cada dato se le asigna una regla como resultado de la operación IF. Los pesos de estas reglas son en principio generados a partir de una distribución uniforme  $[0,1]$  y posteriormente optimizados con el algoritmo BP según (31). Parte de los resultados se observan en el cuadro 10.

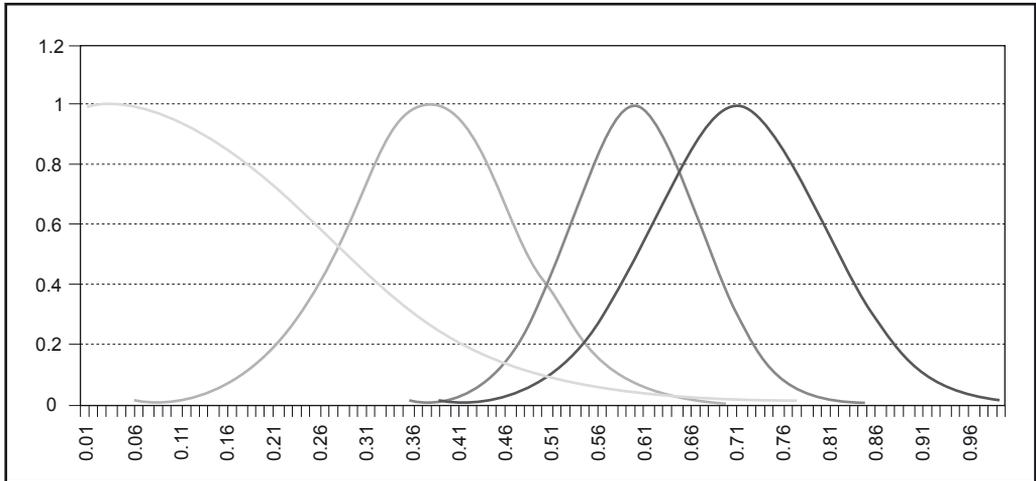
Gráfico 5

**Funciones de pertenencia no optimizadas - $\Delta LTCN$**



Fuente: elaboración propia.

Gráfico 6  
**Funciones de pertenencia optimizadas -  $\Delta LTCN$**



Fuente: elaboración propia.

Cuadro 10  
**Reglas y pesos de las reglas *training set***

TRAINING SET	REGLA	R				R-OPTIMIZADO			
		j=1	j=2	j=3	j=4	j=1	j=2	j=3	j=4
1	1	0,8951	0,1469	0,5486	0,2448	0,0645	1,2102	2,6071	3,4444
2	1	0,8951	0,1469	0,5486	0,2448	0,0645	1,2102	2,6071	3,4444
3	1	0,8951	0,1469	0,5486	0,2448	0,0645	1,2102	2,6071	3,4444
4	1	0,8951	0,1469	0,5486	0,2448	0,0645	1,2102	2,6071	3,4444
5	1	0,8951	0,1469	0,5486	0,2448	0,0645	1,2102	2,6071	3,4444
-	-	-	-	-	-	-	-	-	-
186	1	0,8951	0,1469	0,5486	0,2448	0,0645	1,2102	2,6071	3,4444
187	4	0,7414	0,7886	0,3208	0,4327	0,1344	1,0216	0,9077	1,3967
188	13	0,0848	0,4206	0,5476	0,2076	-2,758	-13,881	-22,662	-34,165
189	1	0,8951	0,1469	0,5486	0,2448	0,0645	1,2102	2,6071	3,4444
190	1	0,8951	0,1469	0,5486	0,2448	0,0645	1,2102	2,6071	3,4444

Fuente: elaboración propia.

El cuadro 11 muestra las funciones de pertenencia que le corresponden a los *inputs* de

los datos a pronosticar y las reglas que se le asignan a los mismos.

Cuadro 11

**Reglas para pronósticos**

Dato de predicción	Regla	Asignaciones lingüísticas funciones de pertenencia		Dato de predicción	Regla	Asignaciones lingüísticas funciones de pertenencia	
		ODF <sub>t-1</sub>	ODF <sub>t-2</sub>			ODF <sub>t-1</sub>	ODF <sub>t-2</sub>
1	1	MUY_BAJO	MUY_BAJO	21	1	MUY_BAJO	MUY_BAJO
2	5	BAJO	MUY_BAJO	22	1	MUY_BAJO	MUY_BAJO
3	1	MUY_BAJO	MUY_BAJO	23	1	MUY_BAJO	MUY_BAJO
4	1	MUY_BAJO	MUY_BAJO	24	4	MUY_BAJO	MEDIO_ALTO
5	1	MUY_BAJO	MUY_BAJO	25	16	MEDIO_ALTO	MEDIO_ALTO
6	4	MUY_BAJO	MEDIO_ALTO	26	13	MEDIO_ALTO	MUY_BAJO
7	16	MEDIO_ALTO	MEDIO_ALTO	27	3	MUY_BAJO	MEDIO
8	13	MEDIO_ALTO	MUY_BAJO	28	9	MEDIO	MUY_BAJO
9	4	MUY_BAJO	MEDIO_ALTO	29	4	MUY_BAJO	MEDIO_ALTO
10	13	MEDIO_ALTO	MUY_BAJO	30	13	MEDIO_ALTO	MUY_BAJO
11	4	MUY_BAJO	MEDIO_ALTO	31	5	BAJO	MUY_BAJO
12	16	MEDIO_ALTO	MEDIO_ALTO	32	3	MUY_BAJO	MEDIO
13	16	MEDIO_ALTO	MEDIO_ALTO	33	12	MEDIO	MEDIO_ALTO
14	16	MEDIO_ALTO	MEDIO_ALTO	34	16	MEDIO_ALTO	MEDIO_ALTO
15	13	MEDIO_ALTO	MUY_BAJO	35	13	MEDIO_ALTO	MUY_BAJO
16	4	MUY_BAJO	MEDIO_ALTO	36	5	BAJO	MUY_BAJO
17	13	MEDIO_ALTO	MUY_BAJO	37	4	MUY_BAJO	MEDIO_ALTO
19	9	MEDIO	MUY_BAJO	39	1	MUY_BAJO	MUY_BAJO
20	1	MUY_BAJO	MUY_BAJO	40	1	MUY_BAJO	MUY_BAJO

Fuente: elaboración propia.

**5. Resultados**

El gráfico 7 muestra los pronósticos de los modelos ADL, MLP y NFS. El cuadro 12 resume los errores de pronóstico de cada modelo. El mismo cuadro muestra los resultados de la regresión del logaritmo de la

tasa de cambio contra el pronóstico de cada modelo.

Según RMSE y el índice de Theil, el modelo MLP muestra una pequeña ventaja sobre NFS y el modelo ADL. El paseo aleatorio presenta pronósticos significativamente me-

nos precisos que los demás, lo que sin lugar a dudas se atribuye al flujo de órdenes.

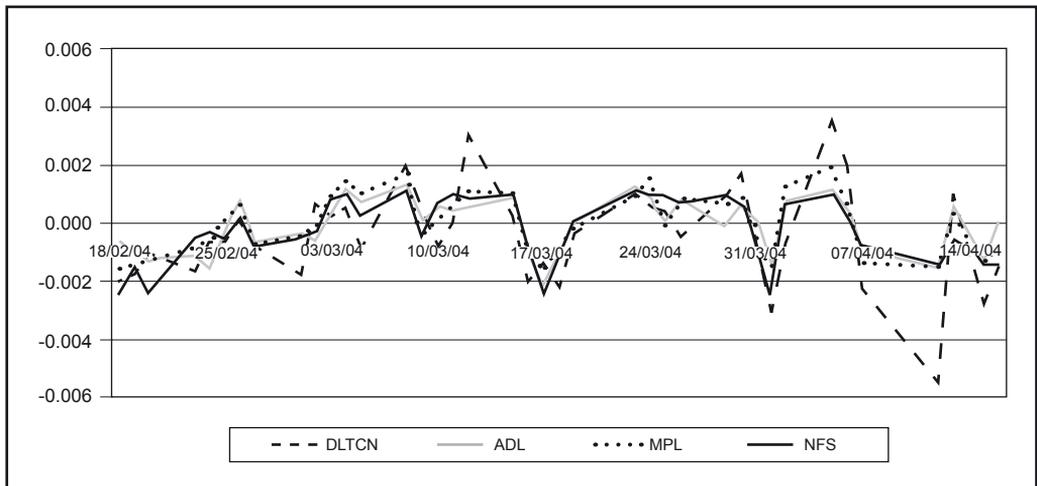
La medida MAPE marca una ligera ventaja del modelo lineal sobre los modelos no lineales NFS y MLP; sin embargo, esta medida no es muy confiable cuando la variable es cercana a 0, como en esta aplicación.

En el cuadro 12 se observa que los modelos no lineales logran predecir correctamente el

signo del 80% de los cambios de la tasa de cambio, mientras que ADL lo hace un 75% de las veces.

Según el “ $p_{valor}$ ” de la prueba “ $\alpha=0;\beta=1$ ” para la regresión planteada en (38), los pronósticos del sistema neurodifuso presentan una clara aceptación la hipótesis nula. En Montenegro (2005, p. 129) se afirma que aceptar “ $H_0$ ” implica un buen pronóstico.

Gráfico 7  
 **$\Delta LTCN$  y pronósticos**



Fuente: elaboración propia.

Cuadro 12  
**Errores de pronóstico**

Modelo	ERROR DE PRONÓSTICO				Regresión: $\Delta LTCN_t = \alpha + \beta \widehat{LTCN}_t + e_t$ (38)			
	RMSE	Theil	MAPE	% Pron. Correctos	$\alpha$	$\beta$	$H_0: \alpha=0; \beta=1$	Pvalor
RW	0,00176	0,158	-	-	-	-	-	-
ADL	0,00126	0,100	1,28	75%	-0,000319	1.3197	$F=2.840$	0.0708
MLP	0,00116	0,090	1,35	80%	-0.000403	1.2781	$F=4.341$	0.0200
NFS	0,00124	0,094	1,39	80%	-0.000184	1.0667	$F=0.572$	0.5690

Fuente: elaboración propia.

## Conclusiones

Este artículo compara pronósticos de la tasa de cambio nominal en Colombia provenientes de modelos no lineales (red neuronal MLP y sistemas neurodifusos) con los de un modelo lineal ADL. La especificación de los modelos para pronóstico se fundamenta en Evans y Lyons (1999) al incorporar como variable explicativa el flujo de órdenes, la variable más representativa de la microestructura del mercado cambiario. Dicha variable fue generada como la diferencia entre volumen de las operaciones propuestas por compradores y vendedores en el mercado de divisas SET FX cada día.

Los p-valor de los dos primeros rezagos del flujo de órdenes de la salida del modelo de rezagos distribuidos prueban la importancia del flujo de órdenes en la determinación de la tasa de cambio. Más contundentes aún son los resultados de los análisis de contribución incremental de las variables y de componentes principales, de los que se concluye que estos rezagos no son solamente las variables más importantes en la determinación del tipo de cambio en Colombia, sino que en su presencia la información macroeconómica disponible no tiene ningún aporte significativo para la explicación de su dinámica.

Las medidas de error de pronóstico RMSE y el índice de Theil revelan una ventaja significativa de los modelos de pronóstico MLP, NFS y ADL sobre el paseo aleatorio, atribuible sin duda a los rezagos del flujo de órdenes. Este hallazgo es importante, porque implica que las posiciones especulativas de los agentes (capturadas en el modelo por el

flujo de órdenes) influyen los movimientos futuros del tipo de cambio, lo que prueba que el mercado cambiario colombiano no es eficiente.

Las medidas RMSE y el índice de Theil también muestran una pequeña ventaja de MLP sobre NFS y ADL en el pronóstico del tipo de cambio. En términos del porcentaje de aciertos correctos, los modelos no lineales logran predecir correctamente el signo del 80% de los cambios de la tasa de cambio, mientras que el modelo ADL lo hace un 75% de las veces. Por el contrario, la medida MAPE señala alguna ventaja del modelo lineal sobre los modelos no lineales.

La regresión de los valores observados de la tasa de cambio sobre sus pronósticos permitió concluir que solo el pronóstico de NFS cumple con las condiciones especificadas para ser catalogado como un buen pronóstico.

A la luz de los resultados, la ligera ventaja del modelo de redes neuronales y el sistema neurodifuso no parece ser evidencia suficiente para apoyar la hipótesis de no linealidad del proceso generador de los datos, sin que pueda ser descartado tampoco.

## Lista de referencias

- Azañero, J. (2003). *Dinámica del tipo de cambio: una aproximación desde la teoría de la microestructura del mercado*. Documento de trabajo, Banco Central de la Reserva del Perú. Recuperado el 15 de diciembre del 2006, de [http://200.121.66.36/bcr/dmdocuments/Publicaciones/Revista/Rev09/04\\_Azanero.pdf](http://200.121.66.36/bcr/dmdocuments/Publicaciones/Revista/Rev09/04_Azanero.pdf)

- Barreiro, R. (2004). *Microestructura del mercado cambiario de Venezuela. Temas de política cambiaria en Venezuela*. Colección: Economía y finanzas. Caracas: Banco Central de Venezuela. Recuperado el 4 de noviembre del 2006, de <http://www.bcv.org.ve/upload/publicaciones/temaspolcambven.pdf>
- Bilson, J. (1978). *The monetary approach to the rate exchange – some empirical evidence*. International Monetary fund Staff Papers, 25, 48-75.
- Branson, W. (1976). *Asset markets and relative prices in exchange rate determination*. Stockholm, Institute for International Economic Studies, Seminar Paper No. 66.
- Cassel, G. (1916). The present situation in the foreign exchanges. *Economic Journal*, 62-65.
- Dornbusch, R. (1976). Expectations and exchange rate dynamics. *Journal of Political Economy*, 84: 1161-1176.
- Evans M., Lyons R. (1999). *Order Flow and exchange rate dynamics*. NBER Working Paper No. 7317. Recuperado el 11 de abril del 2006, de <http://www.nber.org/papers/w7317.pdf>
- Fisher, I. (1930). *The Theory of Interest*. Nueva York: Macmillan.
- Fleming, J. (1962). *Domestic financial policies under fixed and under floating exchange rates*. International Monetary fund Staff Papers, 9, 369-379.
- Frenkel, J. (1976). A monetary approach to the exchange rate. Doctrinal aspects and empirical evidence. *Scandinavian Journal of economics*, 78, 200-224.
- Funahashi, K. (1989). On the approximate realization of continuous mappings by neural networks. *Neural networks*, 2, 183-192.
- Gradojevic, N., Yang J. (2000). *The application of artificial networks to exchange rate forecasting: the role of market microstructure variables*. Bank of Canada, Working Paper 2000-23. Recuperado el 18 de abril del 2007, de <http://www.bankofcanada.ca/en/res/wp/2000/wp00-23.pdf>
- Jalil, M., Misas M. (2007). Evaluación de pronósticos del tipo de cambio utilizando redes neuronales y funciones de pérdida asimétricas. *Revista Colombiana de Estadística*, 30(1), 143-161.
- Kaashoek, J. Van Dijk H. (1999). *A simple strategy to prune neural networks with an application to economic time series*. Econometric Institute, Erasmus University Rotterdam. Working Paper No. 103. Recuperado el 22 de julio del 2007, <http://www.tinbergen.nl/discussionpapers/97123.pdf>
- Kohonen, T. (1982a). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59-69.
- Kohonen, T. (1982b). Analysis of a simple self-organizing process. *Biological Cybernetics*, 44, 135-140.
- Kohonen, T. (1988). *Self-organization and associative memory*. Berlin: Springer-Verlag.
- Kuan, C. Liu, T. (1995). Forecasting exchange rates using feedforward and recurrent neural

- Networks. *Journal of Applied Econometrics* 10, 347-64.
- Le Cun, Y. (1985). A learning procedure for asymmetric threshold network. *Proceedings of Cognitive*, 85, 599-604.
- Llano, L., Hoyos, A., Arias, F., Velásquez, J. (2007). Comparación del desempeño de funciones de activación en redes feedforward para aproximar funciones de datos con y sin ruido. *Revista Avances en Sistemas e Informática*, 4(2).
- Min Qi (1996). *Financial applications of artificial neural Networks*. Handbook of Statistic, 14, 537-538 (Edited by G.S. Maddala and C.R. Rao) Elsevier.
- Meade, J. (1951). *The theory of international economic policy*. vol. I: The balance of payments. London: Oxford University Press.
- Marín, C., Rubio, G. (2001). *Economía financiera*. Barcelona: Antoni Bosch.
- Meese, R., and Rogoff K. (1983). Empirical exchange rate models of the seventies. *Journal of International Economics*, 14, 3-24.
- Montaño, J. (2002a). Tutorial sobre redes neuronales artificiales: los mapas autoorganizados de Kohonen. *Revista Electrónica de Psicología*, 6(1), 273-297.
- Montaño, J. (2002b). *Redes neuronales artificiales aplicadas al análisis de datos*. Tesis doctoral no publicada. Universitat de Les Illes Balears. Recuperado el 8 de agosto del 2007, de [www.tdx.cesca.es/TESIS\\_UIB/AVAILABLE/TDX-0713104-100204/tjmm1de1.pdf](http://www.tdx.cesca.es/TESIS_UIB/AVAILABLE/TDX-0713104-100204/tjmm1de1.pdf)
- Montenegro, A. (2001). *Redes neuronales artificiales*. Documentos de investigación, Borradores de Investigación 2001-11, Departamento de Economía, Pontificia Universidad Javeriana.
- Mundell, R. (1962). *The appropriate use of monetary and fiscal policy for internal and external stability*. IMF Staff Papers, 9: 70-79.
- Parker, D. (1985). *Learning logic*. Informe técnico N° TR-87. Cambridge: Center for Computational Research in Economics and Management Science. Research in Economics and Management Science, MIT.
- Refenes, A. (1995). *Neural Networks in Capital Markets*. England: John Wiley and Sons.
- Rumelhart, D., Hinton, G., Williams, R. (1986). *Learning internal representations by error propagation*. En D.E. Rumelhart y J.L. McClelland (eds.), *Parallel distributed processing* (pp. 318-362). Cambridge, MA: MIT Press.
- Swanson, N., White, H. (1995). A Model-Selection Approach to assessing the Information in the term structure using linear models and artificial neural networks. *Journal of Business & Economic Statistics*, 13(3).
- Werbos, P.J. (1974). *Beyond regression: new tools for prediction an analysis in behavioral sciences*. Tesis doctoral no publicada. Harvard University.
- Li, X., L., Ang L., Gray R. (1999). An intelligent Business Forecaster. *Journal Forecast*, 18, 181-204.

