

PREDICCIÓN DEL COMPORTAMIENTO DIARIO DE LA ACCIÓN DE SURAMINV: REDES NEURONALES Y MODELOS ECONOMETRÍCOS

JAIME ENRIQUE ARRIETA BECHARA*

JUAN CAMILO TORRES CRUZ*

HERMILSON VELASQUEZ CEBALLOS*

Resumen:

El presente estudio presenta evidencia de que con la utilización de modelos estadísticos, econométricos y de inteligencia artificial es posible predecir el comportamiento diario de la acción de SURAMINV, contrastando la hipótesis de la teoría de eficiencia débil de mercado. El trabajo va más allá que otros del tema en el sentido en que más que lograr un buen pronóstico *in sample* busca obtener resultados *out of sample*, controlando de esta manera la existencia de *data snooping* y por tanto evaluando la verdadera factibilidad de aprovechar estos resultados. Además, a partir de estas predicciones y usando sistemas de negociación se evalúa la posibilidad de obtener rendimientos extraordinarios sobre la estrategia *Buy & Hold*.

Palabras claves: *Red neuronal artificial (RNA), modelos econométricos, análisis de componentes principales (ACP), predicción, SURAMINV, sistemas de negociación.*

Abstract:

The current study shows evidence that using statistical, econometric and artificial intelligence models it is possible to predict the daily stock price fluctuations of SURAMINV, in contrast with the weak form of efficient-market hypothesis. This study goes beyond similar ones in that besides achieving a good *in sample* forecast, it also aims at evaluating *out of sample* results. Additionally it controls the possibility of *data snooping* and therefore evaluates the true potential of taking advantage of the results. Furthermore, the forecasts obtained are used to analyze through negotiation systems the possibility of gaining extraordinary returns with regard to the *Buy & Hold* strategy.

Key words: *Artificial Neural Network, econometric models, prediction, SURAMINV, negotiation systems.*

* Jaime Enrique Arrieta Bechara, Universidad EAFIT. jarriet1@eafit.edu.co. Medellín, Colombia.

* Juan Camilo Torres Cruz, Universidad EAFIT. ltorres7@eafit.edu.co. Medellín, Colombia.

* Ermilson Velasquez Ceballos. Universidad EAFIT. evelas@eafit.edu.co. Medellín, Colombia.

1. Introducción

La Teoría de Eficiencia de Mercado es uno de los pilares del desarrollo moderno de las finanzas. Según esta teoría, y como lo explica Fama (1970), los precios de las acciones son el reflejo de toda la información disponible por lo que ningún agente puede obtener retornos extraordinarios en los mercados de manera consistente.

Roberts (1967) distingue tres diferentes tipos de eficiencia en los mercados: débil, semifuerte y fuerte. La hipótesis de la primera es que las series históricas del mercado, incluyendo precios, volúmenes y demás datos referentes a las transacciones, no poseen ningún tipo de información que pueda ser aprovechable por los agentes para obtener rendimientos extraordinarios de forma consistente. Según este planteamiento, no es posible predecir precios ni rendimientos futuros de los activos. La eficiencia semifuerte tiene en cuenta además de la información ya mencionada toda aquella que tenga que ver con la empresa y que sea de público conocimiento, como por ejemplo los resultados financieros, la información del sector o de la economía y la eficiencia fuerte tiene en cuenta además de la información de las anteriores la que no es de público conocimiento, exponiendo que incluso si se cuenta con información privilegiada, los agentes no pueden obtener retornos extraordinarios de manera consistente.

Son muchos los trabajos realizados en los que se busca evaluar la eficiencia de los mercados, obteniendo conclusiones a favor y en contra de dicha teoría. Lo cierto es que en todo el mundo tiene mucha aceptación entre operadores de mercados financieros la implementación de estrategias que pretenden predecir comportamientos futuros a través de indicadores técnicos o modelos (matemáticos, estadísticos, econométricos o de inteligencia artificial) contruidos a partir de las series históricas de las transacciones. Según la firma Aite Group LLC “una tercera parte de todos los negocios en acciones americanas en el 2006 fue conducida por programas automáticos o algoritmos”. Según esta misma firma, para el 2010 se espera que la participación de las negociaciones basadas en este tipo de herramientas sea cercana al 50%¹.

¹ KELLY, Jason. The Ultimate Money Machine. Revista Bloomberg, pág. 38. Junio de 2007

El auge de estrategias de negociación basadas en predicciones se basa en la necesidad de las firmas financieras de contar con métodos de operación ajenos o menos dependientes de las emociones humanas; criterio que se ha demostrado que puede conducir a grandes pérdidas en los mercados financieros. Además, tienen a favor de su gran aceptación el hecho de no presentar grandes dificultades de comprensión e implementación para los agentes.

La divergencia existente entre la teoría y la práctica justifica la realización de trabajos de este tipo, en los que se busque evaluar la eficiencia de mercado en la negociación de activos específicos. Concretamente, consideramos que es de gran relevancia evaluar la predictibilidad del comportamiento de la acción de Suramericana de Inversiones S.A. SURAMINV por ser la acción que lideró durante el periodo de estudio las negociaciones en la Bolsa de Valores de Colombia, al ser la de mayor participación en el Índice General de la Bolsa de Colombia IGBC y a la vez la de mayor volumen transado. La relevancia desde el punto de vista académico radica en el avance en la discusión de la teoría de eficiencia de mercado a partir de un caso colombiano y desde el punto de vista de los actores de mercados por contar con un mayor número de criterios para la toma de decisiones.

Es importante resaltar que para la construcción de los modelos estadísticos, econométricos y de redes neuronales empleados en el trabajo se aplicó toda la rigurosidad que exige el desarrollo teórico de cada uno de ellos, teniendo especial cuidado en evitar que los resultados sean el reflejo de relaciones *espureas*. El trabajo pretende ir más allá que muchos otros de predicción de comportamiento de mercados en el sentido en que más que lograr un buen pronóstico *in sample* se busca evaluar resultados *out of sample*, controlando de esta manera la posibilidad de *data snooping* y por tanto evaluando la verdadera posibilidad de aprovechamiento de los resultados. Además, el estudio no se queda sólo en la predictibilidad de las series, sino que regresa al campo financiero en la medida en que evalúa la posibilidad de aprovechar esa predictibilidad desde el punto de rendimientos de exceso.

El trabajo se compone de ocho apartados, siendo el primero esta introducción y el segundo una síntesis de los principales trabajos realizados sobre predicciones de mercados financieros. En la tercera sección se abordan desde el punto de vista teórico

y matemático las redes neuronales por ser uno de los fundamentos de este estudio y además se hace una breve mención de los modelos de componentes principales, econométricos y de series de tiempo. Luego, en el apartado cuatro se hace referencia a los criterios que se emplean para la evaluación de los modelos y en el cinco se describen los datos utilizados y el tratamiento realizado a los mismos. El sexto apartado presenta los resultados del problema de predicción planteado, para luego, en el séptimo, evaluar la posibilidad de aprovechar esas predicciones desde el punto de vista financiero. Por último, la octava parte del trabajo recoge las conclusiones del mismo, a la vez que muestra las puertas que a criterio de los autores quedan abiertas después de la realización de este estudio para futuras investigaciones.

2. Antecedentes

El estudio de la predictibilidad de los activos financieros se podría decir que data de la misma fecha en la que empezaron los mercados en los que ellos se negocian. Sin embargo, se ha presentado un proceso de perfeccionamiento y sofisticación de las herramientas empleadas para dichas predicciones, teniendo hoy en día un amplio abanico de posibilidades para este fin. La literatura existente sobre predictibilidad en los mercados financieros y evaluación de estrategias de negociación es muy amplia, pero se pueden destacar en el campo internacional los estudios mencionados a continuación.

Hua y Yann (1996) desarrollan un modelo de redes neuronales recurrentes para hacer predicciones a mediano plazo del mercado accionario de Taiwan, usando entradas obtenidas del análisis ARIMA de la serie de estudio. Los resultados permitieron observar que entrenando la red con datos semanales con cuatro años de historia, se pueden obtener buenas predicciones del mercado en un horizonte de seis meses.

Fernández, González y Sosvilla (2000) desarrollan una regla de negociación para el Índice General de la Bolsa de Madrid utilizando redes neuronales y análisis técnico. Concluyen que dicha regla obtiene mejores rendimientos de forma consistente frente a una estrategia *Buy & Hold* mientras el mercado se encuentra estable o a la baja, pero que cuando el mercado se encuentra en tendencia alcista, la mejor estrategia es la

pasiva. Por su parte, Chen, Leung y Daouk (2003) buscan modelar y predecir la dirección de los retornos del Índice de la Bolsa de Acciones de Taiwán a través de estrategias de negociación con base en redes neuronales probabilísticas, obteniendo mejores rendimientos que la estrategia *Buy & Hold*.

Parisi, Parisi y Guerrero (2003) a través del uso de redes neuronales artificiales hacen predicciones de los cambios de signo semanales de los índices bursátiles Bovespa (Brasil), CAC40 (Francia), Dow Jones Industrial (Estados Unidos), FTSE100 (Reino Unido), GDAX(Alemania), Hang Seng (Hong Kong), KLS (Malasia), Nikkei225 (Tokio), STI (Tailandia) y S&P500 (USA), encontrando que esta técnica presenta mayor poder predictivo sobre las series estudiadas frente a modelos ARIMA.

Álvarez y Álvarez (2003) utilizan Algoritmos genéticos y redes neuronales para predecir los precios del Yen y la Libra Esterlina. Adicionalmente, hacen una fusión de datos a través de la cual combinan los resultados obtenidos por ambos métodos con el fin de mirar la existencia de sinergias para mejorar los resultados en la predicción. Pese a lo sofisticado de las herramientas no se ve una mejora considerable en los resultados de predicción de tipos de cambio frente a trabajos anteriores.

Sallehuddin et al. (2007) propone un modelo híbrido para la predicción de diferentes índices económicos y financieros asiáticos. El modelo propuesto es una combinación de redes neuronales tipo GRANN ("Grey Relational Artificial Neural Networks") y un modelo lineal ARIMA. Obtienen como resultado que el modelo híbrido propuesto presenta una mejor predicción que los demás modelos con los que fue comparado.

La literatura sobre el tema con aplicación a mercados colombianos es también muy amplia, destacándose trabajos como Pantoja (2000), donde se realiza una comparación de resultados de predicción a partir de modelos ARIMA y redes neuronales para once series de precios de mercados colombianos. Se obtiene evidencia de mejor desempeño en las redes, pero se detecta que en la presencia de estacionalidad los modelos ARIMA son muy potentes. Gil y Pérez (2005) por su parte realizan una comparación entre un modelo tipo GARCH y uno de redes neuronales para la predicción de la acción de Acerías Paz del Río. Concluyen que por ser las redes modelos más flexibles y adaptables, se pueden considerar como una buena

alternativa para la predicción de series de tiempo, pero a su vez los consideran más como un complemento a las técnicas tradicionales que como un reemplazo.

Pertuz (2005) estima y analiza modelos de series de tiempo y redes neuronales para la acción de Bavaria entre enero de 2002 y abril de 2005. Se utilizan modelos ARIMA y de redes neuronales con retardos y se concluye que estas últimas presentan un mejor desempeño. Jalil y Misas (2007) hacen una comparación en la predicción del tipo de cambio a través del uso de modelos de redes neuronales y de tipo lineal, obteniendo mejores resultados de predicción con los primeros.

Villada et al. (2008) proponen un modelo de redes neuronales para el pronóstico del precio de la energía eléctrica colombiana. Las estructuras de redes utilizadas tienen como entradas las series de precios diarios y el nivel medio de los embalses. Se hace una comparación con un modelo tipo GARCH el cual muestra un mejor desempeño de pronóstico dentro de la muestra, pero se observa que fuera de la muestra las redes tienen mejor predicción.

Como se puede observar, existe una gran variedad de literatura sobre predicción de mercados a través de diferentes técnicas, llegando a concluir a favor o en contra de la teoría de eficiencia de mercado en forma débil. Este estudio pretende ampliar la discusión sobre el tema y otorgar nuevos criterios que permitan contrastar la teoría financiera en mercados colombianos.

3. Marco teórico

La predicción de series financieras se puede realizar por medio de diferentes clases de modelos que abarcan desde sencillos sistemas de reglas técnicas o de mercado hasta avanzadas herramientas matemáticas, estadísticas, econométricas o de inteligencia artificial. En este trabajo se emplea Análisis de Componentes Principales (ACP) como herramienta para mejorar las características de las series iniciales y modelos ARIMA con estructuras ARCH y GARCH, de regresión múltiple y de redes neuronales artificiales para realizar las predicciones.

3.1 Análisis de Componentes Principales (ACP)

El ACP es una herramienta estadística ampliamente utilizada para manejar conjuntos de series que sirvan de entrada a modelos econométricos o de inteligencia artificial y que se presume que presentan altas correlaciones. El objetivo de este análisis es obtener a partir de una muestra de p variables $X_1, X_2, \dots, X_p$ un número $k \leq p$ de nuevas variables o Componentes Principales $Z_1, Z_2, \dots, Z_k$ que estén incorrelacionadas y expliquen la mayor parte de la variabilidad de las variables de entrada iniciales.

El proceso para obtener la primera componente principal parte de la expresión de la misma como combinación lineal de las variables originales:

$$Z_{1i} = u_{11} X_{1i} + u_{12} X_{2i} + \dots + u_{1p} X_{pi} \quad (1)$$

La información del sistema de ecuaciones al representar la información para cada observación $i = 1, 2, \dots, n$ se puede recoger en forma matricial como:

$$Z_1 = XU_1. \quad X_{n \times p}, \quad U_{p \times 1} \quad (2)$$

Procediendo de manera similar al caso univariado tenemos que:

$$V(Z_1) = E(Z_1 - E(Z_1))^2 = E(Z_1^2) = E(Z_1' Z_1) = E[(XU_1)' XU_1] = E(U_1' X' X U) \quad (3)$$

Por el principio de analogía:

$$\widehat{V(Z_1)} = \frac{1}{n} (U_1' X' X U_1) = U_1' \frac{X' X}{n} U_1 = U_1' V U_1, \quad V = \frac{X' X}{n} \quad (4)$$

El problema de optimización a resolver en cada componente principal es:

$$\max_{U_1} \widehat{V(Z_1)} \quad s. a \quad U_1' U_1 = 1 \quad (5)$$

Sea λ el multiplicador de lagrange y consideremos:

$$L = U_1' V U_1 - \lambda (U_1' U_1 - 1) \quad (6)$$

$$\frac{\partial(L)}{\partial U_1} = 2 V U_1 - 2 \lambda U_1 \quad (7)$$

La ecuación (7) es la derivada de una forma cuadrática.

Utilizando las condiciones de primer orden para la optimización:

$$2 V U_1 - 2 \lambda U_1 = 0 \Leftrightarrow V U_1 - \lambda U_1 = 0 \Leftrightarrow (V - \lambda I) U_1 = 0 \quad (8)$$

Si descartamos la solución trivial $U_1 = 0$ se requiere que $\det(V - \lambda I) = 0$, siendo entonces un problema de determinación de las raíces características, tomando como solución aquella que corresponda al mayor valor propio de acuerdo con el objetivo planteado en el problema de optimización. Para encontrar las otras componentes se sigue el mismo proceso descrito. La forma de establecer el número de componentes a retener y de obtener las puntuaciones es abordada ampliamente en la literatura de análisis multivariado.

3.2 Modelos ARIMA con estructuras ARCH y GARCH y modelos de regresión múltiple

Los modelos ARIMA para el manejo de series de tiempo y los de regresión múltiple para la identificación de patrones o relaciones entre conjuntos de datos han sido durante mucho tiempo las herramientas más empleadas en trabajos de predicción debido a su amplia aceptación científica. Es mucha la literatura existente sobre el desarrollo teórico y matemático de este tipo de modelos e igualmente sobre su aplicación en series financieras, siendo por estas razones muy conocidos en el medio académico. Dadas estas condiciones, consideramos que no es necesario entrar a detallar nuevamente la teoría de estos modelos, como si lo haremos con las redes neuronales artificiales, y más bien consideraremos su conocimiento como un requisito previo para lograr una plena comprensión del presente estudio. Sin embargo es importante mencionar que la investigación fue muy rigurosa en cuanto al cumplimiento

de todos los supuestos teóricos que exigen los modelos econométricos. En la bibliografía se reseñan algunas de las fuentes sobre este tipo de modelos utilizadas en el desarrollo del trabajo.

3.3 Redes Neuronales Artificiales (RNA)²

Como los métodos de aproximación lineal y polinomial, una red neuronal es un método que por medio de un conjunto de algoritmos busca establecer la relación entre un conjunto de variables de entrada, $\{x_i\}$, $i = 1, \dots, n$ con un conjunto de una o más variables de salida, $\{y_j\}$, $j = 1, \dots, k$. La diferencia de la red con los otros métodos de aproximación consiste en que esta usa una o más capas ocultas, en las que las variables de entrada van siendo transformadas y comprimidas por una función especial. La red representa la forma en la que el cerebro humano procesa datos de entrada sensoriales, recibidos como neuronas de entrada, hasta el reconocimiento como neuronas de salida.

3.3.1 Redes *feedforward*

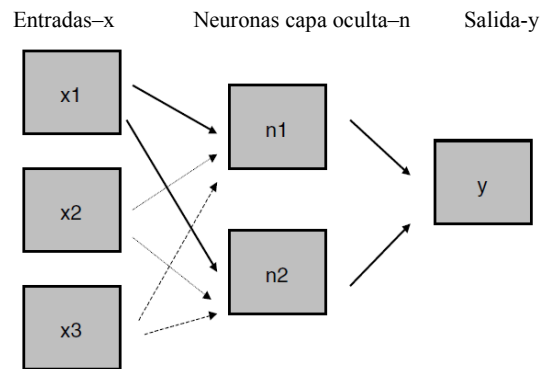
En una red neuronal las conexiones entre las neuronas pueden ser hacia adelante o *feedforward*, hacia atrás o *backforward* o laterales; siendo el primer tipo el más usado en aplicaciones financieras y económicas.

Las redes *feedforward* se caracterizan porque los datos se transmiten de las neuronas de capas anteriores a las de capas posteriores, o lo que es lo mismo, existe una conexión hacia delante y es en este sentido en el que va viajando la información: un conjunto de información de entrada va siendo procesado por las diferentes neuronas de la red para establecer la relación con la variable de salida. El modelo lineal es un caso especial de red *feedforward* en el que la única neurona de la capa oculta es una función de activación lineal que conecta la única capa de salida con un parámetro de peso.

² McNELIS, Paul D. Neural Networks in Finance: Gaining Predictive Edge in the Market. Elsevier. Reino Unido. 2005.

La *Gráfica 1* ilustra la arquitectura de una red neuronal *feedforward* con una capa oculta de dos neuronas, tres variables de entrada $\{x_i\}$, $i = 1,2,3$, y una salida y . Esta es la estructura más básica que puede tener una red *feedforward* y es la más usada, cambiando el número de variables de entrada, en problemas económicos y financieros.

Gráfica 1
Red Neuronal *Feedforward*



En cierto tipo de problemas en los que exista un mayor nivel de complejidad, puede ser deseable la implementación de redes con más de una capa oculta. Sin embargo, esto implica la estimación de un mayor número de parámetros, el sacrificio de grados de libertad, un aumento de la complejidad de la red y el costo de tener mayor tiempo de entrenamiento. Además, en la medida en que se estén usando más parámetros, se está aumentando la probabilidad de que la convergencia sea a mínimos locales y no globales. Por estas razones muchos autores argumentan que los beneficios que se podrían obtener de la utilización de una red con estas características podrían no compensar los costos derivados.

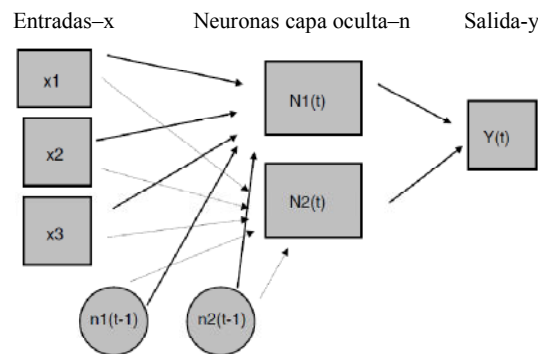
Desde el punto de vista de la economía y las finanzas, las capas ocultas se pueden interpretar como el procesamiento que realizan los agentes de los valores actuales y pasados de las tasas de interés, tasas de cambio de monedas, cambios en el producto de los países y de toda la información económica, financiera y de cualquier otro tipo que conocen, con el fin de formar sus expectativas hacia el futuro. Este procesamiento en el cerebro de los agentes, representado en las transformaciones realizadas en las capas ocultas de la red, es el que conlleva a reacciones y a tomar decisiones en los

mercados en los que realizarán compras o ventas. Estas acciones provienen entonces de predicciones basadas en el procesamiento paralelo de neuronas interconectadas.

3.3.2 Redes recurrentes de *Elman*

Las redes recurrentes de *Elman* son un tipo de arquitectura alternativo comúnmente utilizado. Estas redes permiten que las neuronas dependan no solo de las variables de entrada x , sino también de sus propios valores rezagados, teniendo así “memoria” en la evolución de las neuronas. La *Gráfica 2* presenta la arquitectura de una red neuronal recurrente de *Elman* con una capa oculta de dos neuronas, tres variables de entrada $\{x_i\}$, $i = 1,2,3$ y una salida y . Esta es la estructura más básica e igual que en el caso de la *feedforward* es la más usada en problemas económicos y financieros cambiando el número de variables de entrada.

Gráfica 2
Red Neuronal Recurrente de *Elman*



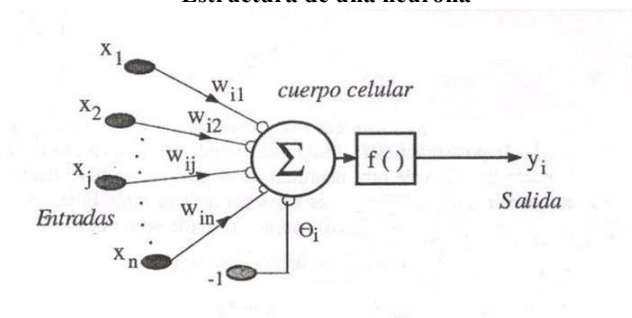
Las redes de *Elman* hacen uso de los valores actuales y rezagados de las neuronas no comprimidas y no observables de la capa oculta. El proceso empieza partiendo de valores iniciales del vector de neuronas rezagadas con *proxies* de la neurona rezagada de una simple red *feedforward*. Luego se estiman sus coeficientes y se recalcula el vector de neuronas rezagadas, reestimando los valores de los parámetros de forma recursiva hasta lograr la convergencia.

Este tipo de redes se constituyen en una forma de capturar memoria en los mercados financieros, particularmente para predicción en datos de alta frecuencia como retornos diarios, intradiarios o en tiempo real de mercados de acciones o monedas. Mientras el uso de rezagos ciertamente es una forma de capturar memoria, la memoria podría revelarse en forma de cambios de estructura no lineal a través del tiempo. El uso de las redes de *Elman* es una forma muy práctica de modelar este tipo de estructura de memoria en la que la capa oculta por si misma cambia a través del tiempo. Las redes de *Elman* son explícitamente un tipo de redes dinámicas, mientras que las redes *feedforward* son usualmente reconocidas como estáticas.

3.3.3 Funciones de transferencia

Como se puede observar en la *Gráfica 3*, las neuronas procesan los datos de entrada de dos formas: primero haciendo combinaciones lineales de los datos de entrada y luego procesando esas combinaciones por medio de funciones de activación o transferencia. Normalmente se reconocen tres tipos de funciones de transferencia: las lineales, cuya salida es un valor proporcional a la entrada, las de umbral, que reportan una salida de valores discretos como por ejemplo 0 o 1 dependiendo si la entrada supera o no un determinado umbral, y las funciones no lineales en las que la salida no es proporcional a la entrada.

Gráfica 3
Estructura de una neurona



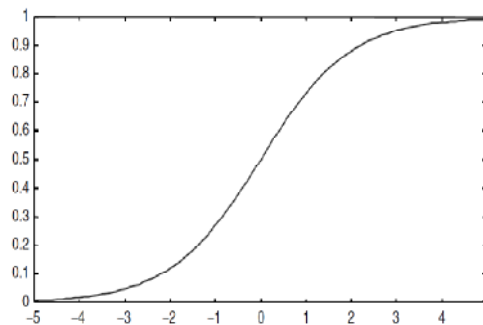
Normalmente las funciones no lineales conocidas como compresoras son las que mejor representan el comportamiento económico de los agentes. Además, como lo señalan Mandic y Chambers (2001) tienen buenas propiedades de aproximación y

pueden simular otros tipos de funciones como la de Base Radial. Es por estas razones que se constituyen en las más usadas para las capas ocultas de redes en las que se trabajan aplicaciones de economía y finanzas.

Por su parte, para la capa de salida la función de transferencia más usada es la lineal por ser continua para todos los reales, arrojando como resultado exactamente el mismo valor de la entrada neta que le llega y por tanto siendo equivalente a no aplicar función de transferencia. La continuidad es una característica deseada para la función de esta capa en el sentido que la red tendrá libertad para arrojar cualquier valor de salida.

La *Gráfica 4* ilustra la operación de la típica función compresora de activación *logsigmoid*, la cual transforma las entradas en el rango entre 0 y +1 antes de transmitir sus efectos en la salida.

Gráfica 4
Función *logsigmoid*



La ventaja de la función de transformación *logsigmoid* proviene de su umbral de comportamiento, el cual caracteriza muchos tipos de respuestas económicas a cambios en variables fundamentales. Por ejemplo, si las tasas de interés son ya muy altas o muy bajas, pequeños cambios de esta variable tendrán un efecto muy bajo en la decisión de compra de bienes durables. Sin embargo, cuando la tasa este en un rango de valores críticos entre los dos extremos, pequeños cambios pueden generar grandes efectos en la demanda de bienes durables.

La forma de la función *logsigmoid* refleja además una forma de comportamiento de aprendizaje. Siguiendo con el ejemplo de las tasas de interés, cuando las tasas están en niveles muy bajos y comienzan a aumentar, los consumidores analizarán la probabilidad de movimientos hacia arriba o abajo en la tasa basados en la información financiera disponible. Entre más experiencia tengan, hasta algún nivel, más acertados serán al interpretar esa señal en el tiempo para tomar ventaja de la tasa de interés actual o de posponer una compra. Los resultados serán muy diferentes dependiendo del punto de la historia temporal de las tasas de interés. Por esta razón, la función no lineal *logsigmoid* captura una respuesta en un umbral que caracteriza la racionalidad o el proceso de aprendizaje en la formación de expectativas.

La red *feedforward* descrita e ilustrada en la *Gráfica 1*, usando la función *logsigmoid*, está representada en las siguientes ecuaciones:

$$n_{k,t} = \omega_{k,0} + \sum_{i=1}^{i^*} \omega_{k,i} x_{i,t} \quad (9)$$

$$N_{k,t} = L(n_{k,t}) \quad (10)$$

$$= \frac{1}{1+e^{-n_{k,t}}} \quad (11)$$

$$y_t = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k N_{k,t} \quad (12)$$

donde $L(n_{k,t})$ representa la función de activación *logsigmoid* con la forma $\frac{1}{1+e^{-n_{k,t}}}$. En este sistema hay i^* variables de entrada $\{x\}$ y k^* neuronas. Una combinación lineal de estas variables de entrada observadas en el momento t , $\{x_{i,t}\}$, $i = 1, \dots, i^*$, con el vector de coeficientes o conjuntos de pesos de entrada $\{\omega_{k,i}\}$, $i = 1, \dots, i^*$, y con la constante $\omega_{k,0}$, forman la variable $n_{k,t}$. Esta variable es comprimida por la función *logsigmoid* y se convierte en una neurona $N_{k,t}$ en el momento de la observación t . El conjunto de k^* neuronas en el momento t están combinadas en una forma lineal con el vector de coeficientes $\{\gamma_k\}$, $k = 1, \dots, k^*$, y con un término constante γ_0 , para formar la predicción $\hat{y}_t$ en el momento t . Normalmente esta es la primera red que prueban los investigadores como alternativa a los modelos de predicción lineal.

La red recurrente de *Elman*, ilustrada en la *Gráfica 2* y usando la función de transferencia *logsigmoid*, estaría representada por:

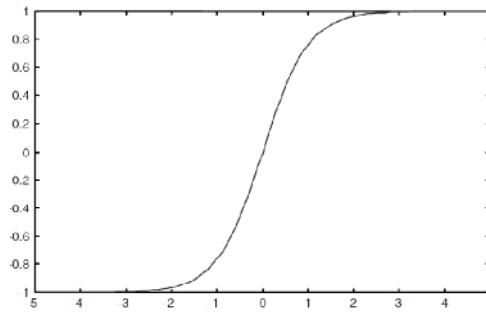
$$n_{k,t} = +\omega_{k,0} + \sum_{i=1}^{i^*} \omega_{k,i} x_{i,t} + \sum_{k=1}^{k^*} \phi_k n_{k,t-1} \quad (13)$$

$$N_{k,t} = \frac{1}{1+e^{-n_{k,t}}} \quad (14)$$

$$y_t = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k N_{k,t} \quad (15)$$

Una función de activación compresora alternativa para las neuronas en una red neuronal es la función tangente hiperbólica, también conocida como la función *tansig* o *tanh*. Esta comprime las combinaciones lineales de las entradas en el rango comprendido entre -1 y +1, en vez de entre 0 y +1 como lo hace la *logsigmoid*. La *Gráfica 5* presenta el comportamiento de esta función.

Gráfica 5
Función tansig



La representación matemática de la red *feedforward* de la *Gráfica 5* con la función de activación *tansig* está dada por:

$$n_{k,t} = \omega_{k,0} + \sum_{i=1}^{i^*} \omega_{k,i} x_{i,t} \quad (16)$$

$$N_{k,t} = T(n_{k,t}) \quad (17)$$

$$= \frac{e^{n_{k,t}} - e^{-n_{k,t}}}{e^{n_{k,t}} + e^{-n_{k,t}}} \quad (18)$$

$$y_t = \gamma_0 + \sum_{k=1}^{k^*} \gamma_k N_{k,t} \quad (19)$$

donde $T(n_{k,t})$ es la función de activación *tansig* para la neurona de entrada $n_{k,t}$.

3.3.4 Backpropagation³

El aprendizaje o proceso de entrenamiento de una red se puede realizar por medio de un método supervisado o por uno no supervisado. El primero consiste en proporcionarle a la red la información de entrada y la de salida buscando establecer la mejor relación entre los dos conjuntos de información. En el segundo, la red tiene el conjunto de datos de entrada, pero no los de salida, y debe por si misma encontrar patrones y clasificarlos sin conocer el objetivo al que debe llegar.

En los problemas de predicción económica o financiera, se utiliza normalmente el algoritmo de entrenamiento *backpropagation*⁴ por las virtudes que posee, ya que es capaz de resolver problemas que no son linealmente separables, reduce el tiempo de entrenamiento de la red por la utilización del procesamiento paralelo para encontrar la relación entre patrones dados y no necesita la especificación de un algoritmo por anticipado para llegar al algoritmo correcto.

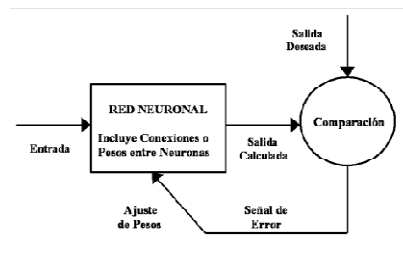
Backpropagation es un tipo de aprendizaje supervisado con un mecanismo de propagación - adaptación en dos fases. Tal como se muestra en la *Gráfica 6*, la primera consiste en entrar un determinado patrón y propagarlo por todas las capas de la red hasta llegar a una salida, la cual es comparada con la salida que se desea obtener, determinando un término de error. Este término de error inicia la segunda fase, ya que se convierte en una señal que se propaga hacia atrás, desde la capa de salida hacia la capa oculta previa a la salida. La señal de error propagada será distribuida entre todas las neuronas de la capa oculta de acuerdo a la contribución relativa que cada una tuvo en la salida original y es con esta fracción de la señal que llega a cada uno de los pesos con la que se realiza la actualización de los mismos. Este mismo proceso se repite para todas las capas hasta que cada una de las

³ Acosta, M., Salazar, H. y Zuluaga, C. (2000). Tutorial de Redes Neuronales. Universidad Tecnológica de Pereira, Facultad de Ingeniería Eléctrica. <http://ohm.utp.edu.co/neuronales/>

⁴ Muchos autores han documentado la superioridad de este tipo de red en el campo de economía y finanzas. Algunos de ellos son: Gately (1996), Rumelhart y McClelland (1986), Wong (1995) y Yao, Li y Tan (1997).

neuronas haya recibido la fracción de la señal del error acorde con su contribución relativa al error original y haya actualizado sus pesos. Al ir repitiendo este proceso, la red se va entrenando en el sentido en que converge hacia unos valores en los que relacione correctamente los patrones de entrenamiento.

Gráfica 6
Mecanismo de propagación – adaptación del
entrenamiento *backpropagation*



La bondad de este método radica en que después que la red ha sido entrenada, es capaz de responder adecuadamente ante patrones de entrada incompletos o con ruidos, lo cual se origina de la capacidad que desarrollan las neuronas de conocer las características de los patrones con los que fueron entrenadas. De esta manera, cuando identifiquen que el nuevo patrón no tiene las características que conoce (ruido), no se generará una salida para el mismo.

Como cualquier método de aprendizaje supervisado, el algoritmo *backpropagation* necesita tener un conjunto de entrenamiento que consiste en la asociación de cada salida con el valor deseado de esa salida. Antes del entrenamiento se debe tener definida la topología de la red y asignar unos valores iniciales a cada uno de los parámetros que tenga.

Asumiendo una red de una sola capa oculta para efectos de ilustración, el proceso de entrenamiento parte de ingresar a la red un patrón de entrenamiento de componentes:

$$P = \begin{bmatrix} P_1 \\ P_2 \\ \cdot \\ \cdot \\ P_i \\ \cdot \\ \cdot \\ \cdot \\ P_q \end{bmatrix} \quad (20)$$

Este patrón se propagará a través de los pesos iniciales hasta la capa oculta y producirá una entrada neta n en cada una de las neuronas de dicha capa. La entrada neta (antes de pasar por la función de transferencia o de activación) a la neurona j de la capa oculta estará dada por:

$$n_j^o = \sum_{i=1}^q W_{ji}^o P_i + b_j^o \quad (21)$$

donde W_{ji}^o es el peso que tiene la componente i en la neurona j de la capa oculta, P_i es el componente i del patrón de entrenamiento y b_j^o es la ganancia de la neurona j de la capa oculta. (o) denota que el parámetro pertenece a la capa oculta.

La salida a_j^o de la neurona j de la capa oculta estará dada por:

$$a_j^o = f^o(\sum_{i=1}^q W_{ji}^o P_i + b_j^o) \quad (22)$$

donde f^o denota la función de transferencia o activación de la capa oculta.

Las salidas de las neuronas de la capa oculta se propagan a través de los pesos de la capa de salida, produciendo una entrada neta para la capa de salida denotada por:

$$n_k^s = \sum_{j=1}^m W_{kj}^s a_j^o + b_k^s \quad (23)$$

donde W_{kj}^s es el peso que tiene la neurona j de la capa oculta en la neurona k de la capa de salida, a_j^o es la salida de la neurona j de la capa oculta, b_k^s es la ganancia de la neurona k de la capa de salida. (s) denota que el parámetro pertenece a la capa salida.

La salida final de la red estará dada por:

$$a_k^s = f^s(n_k^s) \quad (24)$$

donde f^s denota la función de transferencia o activación de la capa de salida.

Al reemplazar (23) en (24) se puede obtener la salida final en función de la entrada neta y de los pesos de la última capa:

$$a_k^s = f^s\left(\sum_{j=1}^m W_{kj}^s a_j^o + b_k^s\right) \quad (25)$$

La salida final de la red se comparará con su valor deseado t_k para determinar el error de cada unidad de salida:

$$\delta_k = (t_k - a_k^s) \quad (26)$$

El error de cada patrón ingresado a la red estará dado entonces por:

$$ep^2 = \frac{1}{2} \sum_{k=1}^s (\delta_k)^2 \quad (27)$$

donde ep^2 es el error medio cuadrático para cada patrón de entrada P y δ_k es el error en la neurona k de la capa de salida con l neuronas.

El proceso que se acaba de describir debe ser repetido para el número total de patrones de entrenamiento (r) y luego minimizar el error medio cuadrático total descrito por:

$$e^2 = \sum_{p=1}^r ep^2 \quad (28)$$

Donde e^2 es el error en el proceso de aprendizaje en una iteración, después de haber ingresado los r patrones a la red.

El error generado en función de los pesos genera un espacio de n dimensiones, donde n se refiere al número de pesos de la red. Al tratar de minimizar este error en el proceso de aprendizaje, se tomará la dirección del gradiente negativo del error en esta superficie, es decir, aquella dirección en la que el error decrece, ya que este es el criterio para la actualización de los pesos:

$$W_{k+1} = W_k - \alpha \nabla e p^2 \quad (29)$$

El gradiente negativo de $e p^2$ se denotará como $-\nabla e p^2$ y se calcula como la derivada del error respecto a los pesos de conexión de toda la red.

Se puede demostrar que los términos de error para la capa de salida estarán dados por:

$$\delta_k^s = (t_k - a_k^s) \times f'^s(n_k^s) \quad (30)$$

Este algoritmo es llamado *Backpropagation* o de propagación inversa por el hecho de que el error se propaga hacia atrás en la red. Por su parte, el error en la capa oculta estaría dado por:

$$\delta_j^o = f'^o(n_j^o) \times \sum_{k=1}^l \delta_k^s W_{kj}^s \quad (31)$$

Con el valor del gradiente del error se actualizan todos los pesos de la red empezando por los de la capa de salida:

$$W_{kj}(t+1) = W_{kj}(t) - 2\alpha \delta_k^s \quad (32)$$

$$b_k(t+1) = b_k(t) - 2\alpha \delta_k^s \quad (33)$$

Donde α es la tasa de aprendizaje que varía entre 0 y +1.

Después de actualizar los pesos y ganancias de la capa de salida se hace el mismo proceso con la capa oculta, terminado así el proceso de actualización de pesos con base en los errores en una red de tres capas:

$$W_{ji}(t + 1) = W_{ji}(t) - 2\alpha\delta_j^o P_i \quad (34)$$

$$b_j(t + 1) = b_j(t) - 2\alpha\delta_j^o \quad (35)$$

Las expresiones para generalizar este proceso a una red de más capas se pueden deducir fácilmente de la ecuación (31).

Como se mencionó, el término α es la tasa de aprendizaje. Este es el valor que determina la magnitud de los cambios al hacer las actualizaciones de los pesos. A pesar de ser un proceso más demandante desde el punto de vista de tiempo, es conveniente que esta tasa sea baja, ya que no se sabe qué tan lejos o cerca se está del punto mínimo del error y por tanto con valores altos se podría correr el riesgo de pasar por encima. Un valor pequeño de α implica que la red tendrá que hacer un número muy grande de iteraciones, mientras que un valor grande implicará lo opuesto.

Es importante tener en cuenta la posibilidad de que la red converja hacia un mínimo local y no global del error, problema que no tiene solución garantizada con el mecanismo de *backpropagation*. Sin embargo, más que el hecho de que el mínimo alcanzado sea local o global, lo importante es que sea admisible desde el punto de vista de evaluación de los resultados.

El entrenamiento *backpropagation* puede ser configurado de diferentes maneras por medio de la variación de las funciones de entrenamiento para el ajuste de los pesos y de aprendizaje para el proceso de convergencia. La selección depende del tipo de problema y los resultados de estimación. La *Tabla 1* presenta un resumen de estas funciones.

Tabla 1
Funciones de entrenamiento y aprendizaje del
entrenamiento *backpropagation*

Funciones de Entrenamiento - Backpropagation	
TRAINLM	Levenberg - Marquardt
TRAINGDX	Gradiente descendiente con momento y regla de aprendizaje adaptativa
TRAINGDM	Gradiente descendiente con momento y regla de aprendizaje adaptativa
TRAINGDA	Gradiente descendiente con regla de aprendizaje adaptativa
TRAINGD	Gradiente descendiente

Funciones de aprendizaje	
LEARNGDM	Gradiente descendiente con momento peso/sesgo
LEARGD	Gradiente descendiente peso/sesgo

Una de las funciones de entrenamiento más comúnmente usadas es *TRAINLM*, la cual utiliza el algoritmo de *Levenberg – Marquardt* para el ajuste de los pesos. Este algoritmo “es una modificación del método de Newton, diseñado para minimizar funciones que sean la suma de los cuadrados de otras funciones no lineales, por lo que tiene un excelente desempeño en el entrenamiento de redes neuronales donde el rendimiento de la red esté determinado por el error medio cuadrático”⁵.

La forma en la que trabaja el algoritmo de Levenberg - Marquardt puede resumirse en los siguientes pasos:

1. Teniendo en cuenta las salidas calculadas y observadas, se calculan los errores definidos en la ecuación (26), a partir de los cuales se calcula la suma de los errores cuadrados para una de las entradas, tal como se expresa en la ecuación (27).
2. Luego, “se calculan las sensibilidades individuales y la matriz sensibilidad total y con estas, se calculan los elementos de la matriz Jacobiana $J(x)$ ”.
3. Se obtiene Δx_k de la siguiente manera

$$\Delta x_k = -[J^T(x_k)J(x_k) + \mu_k I]^{-1} J^T(x_k) v(x_k) \quad (36)$$

4. Se recalcula la suma de los errores cuadrados usando $x_k + \Delta x_k$. Si esta nueva suma es más pequeña que el valor calculado en el paso 1 entonces se divide μ

⁵ Ibid.

por ϑ ⁶, donde μ es la nueva constante luego de ajustar el método de Newton que determina la tendencia del algoritmo y ϑ es un factor de escala. Luego, con este nuevo valor se calcula $x_{k+1} = x_k + \Delta x_k$ y se empieza nuevamente desde el paso 1. En caso de no presentarse una disminución de la suma, se multiplica μ por ϑ y se regresa al paso 3.

“El algoritmo debe alcanzar convergencia cuando la norma del gradiente de $\nabla e(x) = 2J^T(x)v(x)$ sea menor que algún valor predeterminado, o cuando la suma de los errores cuadrados ha sido reducida a un error que se haya trazado como meta”⁷. La ventaja de este algoritmo es su rapidez cuando el número de parámetros de la red es moderado, ya que normalmente para converger requiere un número menor de iteraciones que otros métodos.

En cuanto a las funciones de aprendizaje la más empleada es la *LEARN GDM*, la cual trabaja con un gradiente descendiente con *momentum* peso/sesgo cuyo objetivo es la incorporación de un filtro con *momentum* que busca suavizar la parte final del proceso típico de convergencia de una red *backpropagation*, en la que se nota una caída del error muy brusca en la iteración en la que logra alcanzar la convergencia, abriendo la posibilidad a que se presenten oscilaciones no deseadas.

El algoritmo de ajuste de pesos del *backpropagation* tradicional se puede ilustrar de la siguiente manera:

$$\Delta W^m(k) = -as^m(a^{m-1})^T \quad (37)$$

$$\Delta b^m(k) = -as^m \quad (38)$$

El algoritmo resultante al agregar el filtro con *momentum* permite una convergencia mucho más estable, la cual se denota de la siguiente manera:

$$\Delta W^m(k) = \gamma \Delta W^m(k-1) - (1-\gamma)as^m(a^{m-1})^T \quad (39)$$

$$\Delta b^m(k) = \gamma \Delta b^m(k-1) - (1-\gamma) - as^m \quad (40)$$

⁶ Ibid.

⁷ Ibid.

4. Criterios de evaluación de los resultados de los modelos

El proceso de evaluación de un modelo de estimación empieza por analizar el ajuste de sus resultados de predicción a los mismos datos utilizados para su estimación. Sin embargo, la verdadera prueba ácida para comparar modelos alternativos es el desempeño que presenten en la predicción fuera de la muestra. Estas pruebas evalúan la capacidad de modelos que compiten para generalizar fuera de los datos usados para sus estimaciones. Un modelo que tiene buen ajuste dentro de la muestra puede no tenerlo fuera de la muestra debido a que puede estar capturando la dinámica de la muestra particular pero no es capaz de generalizar para la población.

A continuación se hace una descripción de los criterios empleados en el trabajo para la evaluación de los modelos:

- R-Cuadrado: es la medida de bondad del ajuste más usada. Se refiere al cociente entre la varianza de los datos pronosticados por el modelo y la de los datos observados y que arroja como resultado un valor dentro del intervalo $[0,1]$, interpretándose 1 como el ajuste perfecto:

$$R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y}_t)^2} \quad (41)$$

- Estadístico del error medio cuadrático (MSE): mide el promedio del error cuadrático. Sirve para comparar modelos usados para realizar una misma estimación, favoreciendo al que tenga un menor valor para este estadístico:

$$MSE = \frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2 \quad (42)$$

- Estadístico de raíz del error medio cuadrático (RMSE): se calcula como la raíz cuadrada del MSE, buscando reexpresarlo en las mismas unidades de la variable estimada.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (43)$$

- Estadístico de raíz del error porcentual medio cuadrático (RMSPE): es la reexpresión del MSE en forma porcentual respecto a la variable observada.

$$RMSPE = \sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{\hat{y}_t - y_t}{y_t} \right)^2} \quad (44)$$

- Estadístico del error medio absoluto (MAE): se calcula como el promedio de los valores absolutos de los errores de estimación.

$$MAE = \frac{1}{n} \sum_{t=1}^n |\hat{y}_t - y_t| \quad (45)$$

- Estadístico del error porcentual medio absoluto (MAPE): es la reexpresión del MAE en forma porcentual respecto a la variable observada.

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{y}_t - y_t}{y_t} \right| \quad (46)$$

- Criterio de información *Hannan – Quinn (HQIF)*: a medida que se incremente el número de parámetros de un modelo es posible obtener Coeficientes R-Cuadrado mayores. Una forma de modificar este estadístico es con el Criterio de Información *Hannan-Quinn* (1979), el cual castiga el desempeño del modelo de acuerdo al número de parámetros, k , que usa:

$$hqif = \left[\ln \left(\sum_{t=1}^n \frac{(y_t - \hat{y}_t)^2}{n} \right) \right] + \frac{k \{ \ln[\ln(n)] \}}{n} \quad (47)$$

donde n es el número de observaciones.

El criterio es seleccionar el modelo con el menor valor. El factor de castigo al modelo esta dado por $k \{ \ln[\ln(n)] \} / n$. El *Criterio de Akaike* usa como factor de castigo $2k/n$, mientras que el de *Schwartz* usa $k[\ln(n)]/n$. La ventaja del Criterio de Información *Hannan-Quinn* es que castiga a un modelo por el número de parámetros más que el de *Akaike*, pero no tan severamente como lo hace el de *Schwartz*.

5. Series de datos

La selección de las series de datos históricos empleadas para el estudio fue realizada teniendo en cuenta tres criterios:

- Los fundamentales macroeconómicos y las relaciones entre mercados financieros que han sido ampliamente documentadas en diversos trabajos como los de Eum y Shim (1989), Kock y Kock (1991), Kasa (1992), Phylaktis y Ravazzolo (2001), Chen y Ruy (2002) y Phylaktis y Ravazzolo (2005).
- La disponibilidad de los datos en la periodicidad requerida.
- El conocimiento de los investigadores sobre el funcionamiento del mercado financiero colombiano.

El conjunto inicial de datos se presenta en el *Anexo 1* y consta de 77 variables, dentro de las que se incluyen: tasas de interés, precios de mercados accionarios, monedas, *commodities* y otros indicadores, tanto de Colombia como del extranjero⁸, expresadas en términos de precios y de rendimientos diarios⁹ desde el 21 de junio de 2002 hasta 11 de marzo de 2008 (1493 datos). Este fue un periodo caracterizado por una fase expansiva de la economía colombiana sustentada, entre otros factores, por un fuerte flujo de inversión en el país y por un contexto de economía mundial favorable. Consecuentemente con el crecimiento de la economía se logró una reducción significativa de la tasa de desempleo en el país, logrando llegar a cifras de un dígito en el año 2007. Además hay que resaltar que de forma paralela, a través de la política monetaria se continuó trabajando en la convergencia de la inflación hacia la meta de largo plazo, logrando reducciones importantes en las variaciones de precios para varios de los años considerados.

El contexto económico favorable, catalizado por una confianza muy alta en el gobierno colombiano y unas buenas condiciones de liquidez nacional e internacional, conllevaron a un *rally* muy importante en el mercado accionario local hasta el primer semestre de 2006, logrando alcanzar niveles de precios nunca antes vistos y logrando valorizaciones excepcionalmente altas. Sin embargo, a mediados de 2006 se presenta

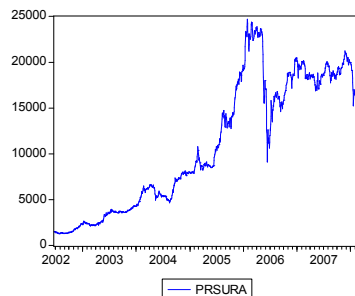
⁸ Series históricas obtenidas de Bloomberg.

⁹ Los rendimientos fueron calculados como variaciones logarítmicas de los precios.

una gran caída de los precios de las acciones interpretado por muchos como la explosión de la burbuja que se había venido formando. Un factor agravante en este cambio de tendencia fue claramente el alto nivel de apalancamiento que tenía el mercado y que conllevaba a liquidaciones en cadena ante la imposibilidad de atender nuevos llamados de margen por parte de los inversionistas. Durante la parte restante del periodo de estudio se observa una recuperación de los precios luego de la crisis y un posterior movimiento lateral de los mismos.

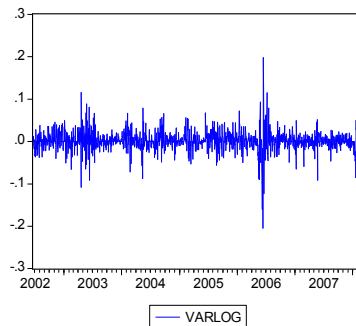
Es importante resaltar que la predicción se está realizando para $t+1$, por lo que las variables independientes de los modelos pueden ser incluidas como máximo hasta t , buscando que la modelación sea consistente con la información que se tendría disponible al hacer la implementación del modelo en la realidad. En la *Gráfica 7*, que presenta la serie de precios de la acción de SURAMINV para el periodo de estudio, se puede observar que la muestra disponible para el estudio imprime cierta dificultad en la modelación con fines de predicción en la medida en que parece ser una realización de un proceso mucho más amplio, mostrando una larga y fuerte tendencia al alza para casi todo el periodo de estudio, un fuerte y puntual movimiento a la baja y un posterior movimiento aparentemente lateral.

Gráfica 7
Precio acción SURAMINV



Los rendimientos de la acción de SURAMINV en el periodo de estudio se presentan en la *Gráfica 8*, serie que también será foco de estudio a pesar de las dificultades conocidas que implica la modelación de rendimientos.

Gráfica 8
Rendimiento acción SURAMINV



Para cumplir con los supuestos formales de algunos de los procesos utilizados en la construcción de los modelos, las 77 variables inicialmente seleccionadas se someten a un proceso de estandarización por el hecho de tener diferentes escalas. El nuevo conjunto de datos estandarizado, fue sometido al Análisis de Componentes Principales (ACP), logrando reducir el conjunto de entrada a 5 componentes principales en el caso de los precios, los cuales explican el 93,89% de la variabilidad total de los datos. En general se observó que cada uno de estas componentes refleja el comportamiento de alguno de los grupos en los que fueron clasificadas las 77 variables originales. Por su parte, al realizar el mismo proceso para las 77 series de rendimiento, se reduce el conjunto de información a 21 componentes principales con una explicación de la variabilidad inicial de 65,53%.

El paso siguiente en la organización de los datos es la división de la muestra, obteniendo dos subconjuntos: uno para la estimación del modelo (*in sample*) y otro que no interviene en la estimación y que se utiliza para la prueba (*out of sample*). Aunque normalmente se acostumbra realizar un ejercicio de partición, en el presente trabajo se realizaron tres diferentes divisiones buscando contar con herramientas para realizar un análisis más robusto de los resultados:

- Inicialmente se usan 68% para estimación *in sample*, dejando el 32% final de los datos para efectos de pruebas *out of sample*. Esta división corresponde al punto medio aproximado de la caída más significativa de precios de la acción de SURAMINV, presentada aproximadamente entre mayo y junio de 2006. Con

esto se busca probar si utilizando la primera parte de dicha caída para la estimación, los modelos son capaces de predecir *out of sample* la segunda parte. Esta partición fue usada en los modelos ARIMA, de regresión múltiple y redes neuronales.

- La segunda división usa 83% de los datos para estimación y el 17% final para pruebas. Con esta división se busca probar la capacidad predictiva de los modelos luego del cambio estructural que parece haberse presentado en el mercado a partir de la caída grande de precio ya mencionada, pero dejando parte de los datos de la nueva dinámica de los precios dentro del conjunto de entrenamiento. Igual que en el caso anterior, esta partición fue probada con los tres tipos de modelos.
- Para los modelos de redes neuronales se empleó una tercera partición realizada de forma intercalada, en la cual cada cuarto dato de la serie fue reservado para efectos de prueba y los restantes se usaron para la estimación. Con esto se pretendía probar si existía alguna mejoría en la capacidad de generalización de la red en la medida en que fuera entrenada con base en datos de representativos de toda la muestra.

Por último, siguiendo las recomendaciones existentes en la literatura sobre RNA, fue necesario subdividir cada una de las muestras para estimación de estos modelos *in sample* en tres grupos: uno de entrenamiento, que es donde el modelo va realizar el proceso de aprendizaje, otro de validación, con el que se busca optimizar el tiempo computacional en el entrenamiento de la red, y un tercero de prueba con el que se busca validar la capacidad de generalización del modelo¹⁰.

El procedimiento más usado para dividir los datos consiste en asignar al primer grupo el 50% de la muestra y a cada uno de los otros el 25%. Buscando que cada una de estas fases (entrenamiento, validación y prueba) se realice en un grupo de datos representativo del 100% de la serie, se acostumbra realizar la partición de manera discontinua. Así, la partición se realizó de la siguiente manera:

¹⁰ Esta metodología es ampliamente recomendada en la literatura de RNA's. Para mayor profundización ver Yao y Tan (2000).

- Grupo de entrenamiento: datos impares de la muestra partiendo de la primera posición (datos 1, 3, 5, 7, etc.),
- Grupo de validación: datos múltiplos de 4 (datos 4, 8, 12, 16, etc.),
- Grupo de prueba: datos restantes, o sea los que responden a la fórmula $2 + 4n$, donde $n = 0, 1, 2, 3$, etc. (datos 2, 6, 10, 14, etc.).

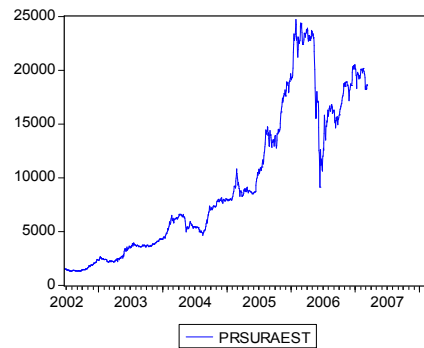
6. Resultados de predicción

El problema de predicción planteado se abordó utilizando diferentes herramientas y modelos estadísticos, econométricos y de inteligencia artificial. Con estos modelos, más que lograr un buen ajuste y pronóstico *in sample* que es el objetivo de la mayoría de trabajos de este tipo, se pretende evaluar la capacidad de predicción fuera de muestra en un horizonte de tiempo amplio. Este horizonte se relaciona con la parte de las series reservada para pruebas fuera de muestra en cada una de las particiones: 32% final, 17% final y 25% intercalado.

Al observar los resultados, se obtiene mejor desempeño en las estimaciones cuando los modelos son estimados en el 83% inicial de los datos que cuando se hace sobre el 68% inicial. Esto puede estar explicado por el hecho de que se estén incluyendo para el entrenamiento mayor número de datos con la nueva dinámica que parece existir en el mercado a partir de la gran caída de los precios a mediados de 2006 y que finalmente es la misma dinámica sobre la que se va a realizar la predicción *out of sample*. A continuación se presentan los resultados obtenidos bajo esta partición.

Inicialmente la predicción fue abordada como un problema de series de tiempo para los precios de SURAMINV. La *Gráfica 8* muestra la parte de la serie de precios usada para el ajuste del modelo.

Gráfica 9
Serie *in sample* de precios de la acción de SURAMINV



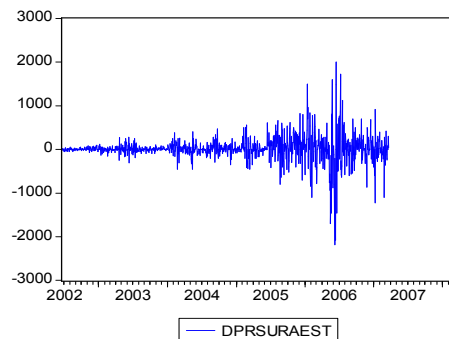
Se observa que la serie tiene una clara tendencia por lo que se puede suponer que no es estacionaria en media, lo cual se contrastó utilizando pruebas de raíces unitarias, además el correlograma de la *Tabla 2* muestra evidencia a favor de la no estacionariedad.

Tabla 2
Correlograma serie *in sample* de precios de la acción de SURAMINV

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
*****	*****	1	0.998	0.998	1229.9	0.000
*****	*	2	0.995	-0.114	2454.4	0.000
*****		3	0.993	-0.003	3673.5	0.000
*****		4	0.990	0.018	4887.4	0.000
*****		5	0.988	-0.028	6095.8	0.000
*****		6	0.985	-0.006	7298.6	0.000
*****		7	0.982	0.012	8496.0	0.000
*****		8	0.980	0.022	9688.2	0.000
*****		9	0.977	0.004	10875.	0.000
*****		10	0.975	0.003	12057.	0.000
*****	*	11	0.972	-0.078	13234.	0.000
*****		12	0.969	-0.007	14404.	0.000
*****		13	0.966	0.010	15569.	0.000
*****		14	0.963	-0.025	16728.	0.000
*****		15	0.960	-0.035	17880.	0.000
*****		16	0.957	-0.022	19025.	0.000
*****		17	0.954	-0.019	20164.	0.000
*****		18	0.951	-0.013	21295.	0.000
*****		19	0.947	-0.013	22419.	0.000
*****		20	0.944	-0.000	23536.	0.000
*****		21	0.940	0.013	24646.	0.000
*****		22	0.937	-0.008	25749.	0.000
*****		23	0.934	0.022	26846.	0.000
*****		24	0.930	-0.011	27935.	0.000
*****		25	0.927	-0.015	29018.	0.000

Por esta razón, siguiendo la metodología de series de tiempo, se procedió a la diferenciación de la serie, resolviendo el problema de falta de estacionariedad en media, pero observando la existencia de problemas de volatilidad condicionada que serán retomados más adelante. La serie diferenciada se presenta en la *Gráfica 10*.

Gráfica 10
Serie *in sample* de primera diferencia de los precios
de la acción de SURAMINV



La estimación del modelo para la serie en primeras diferencias exigió la inclusión de muchos retardos con mucha historia y con poca interpretación desde el punto de vista financiero. Los resultados de esta estimación se presentan en la *Tabla 3*, resaltando que todos los coeficientes resultan ser estadísticamente significativos.

Tabla 3
Estimación modelo ARIMA

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.288188	0.027254	10.57398	0.0000
AR(4)	0.056090	0.027306	2.054099	0.0402
AR(6)	-0.084524	0.027400	-3.084807	0.0021
AR(10)	0.092857	0.027360	3.393868	0.0007
AR(12)	-0.067896	0.028480	-2.383958	0.0173
AR(13)	0.077574	0.028471	2.724692	0.0065
AR(29)	-0.064173	0.028746	-2.232411	0.0258
AR(30)	0.116846	0.028749	4.064374	0.0001
R-squared	0.110871	Mean dependent var		14.21358
Adjusted R-squared	0.105685	S.D. dependent var		278.9645
S.E. of regression	263.8118	Akaike info criterion		13.99495
Sum squared resid	83516022	Schwarz criterion		14.02870
Log likelihood	-8444.950	Durbin-Watson stat		1.982776

La *Tabla 4* presenta el correlograma de los residuales de esta estimación, siendo importante resaltar que con un nivel de confianza de 95% los residuales resultan ser

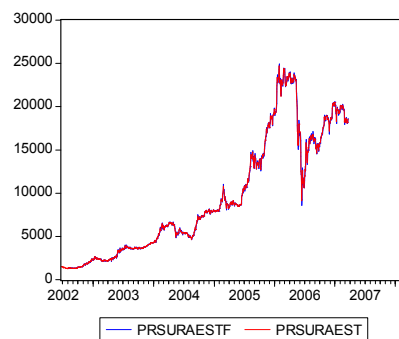
ruido blanco y que por tanto en estos datos no está quedando información relevante o aprovechable y que no existen problemas en la estimación.

Tabla 4
Correlograma residuales modelo ARIMA

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
.	.	1	0.008	0.008	0.0692
.	.	2	-0.028	-0.028	0.9955
.	.	3	-0.031	-0.030	2.1422
.	.	4	-0.010	-0.011	2.2729
.	.	5	0.023	0.021	2.9078
.	.	6	0.004	0.002	2.9292
.	.	7	-0.015	-0.015	3.2213
.	.	8	-0.009	-0.007	3.3165
.	.	9	-0.010	-0.010	3.4347
.	.	10	0.003	0.002	3.4489
.	.	11	-0.002	-0.003	3.4535
.	.	12	0.002	0.002	3.4578
.	.	13	-0.013	-0.013	3.6581
.	.	14	0.016	0.016	3.9659
.	.	15	0.030	0.029	5.0384
.	.	16	0.030	0.030	6.1407
.	.	17	-0.002	-0.000	6.1445
.	.	18	0.025	0.029	6.9293
.	.	19	0.020	0.021	7.4128
.	.	20	-0.017	-0.017	7.7602
.	.	21	-0.009	-0.008	7.8684
.	.	22	-0.057	-0.055	11.840
.	.	23	0.014	0.014	12.088
.	.	24	0.047	0.043	14.863
.	.	25	0.023	0.023	15.543

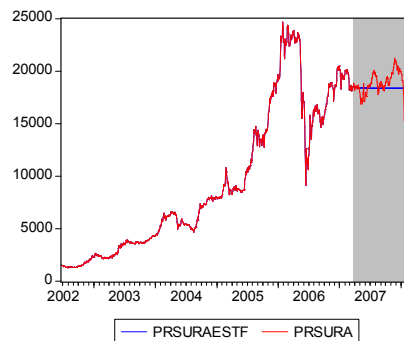
Los modelos ARIMA se caracterizan por presentar un buen ajuste del pronóstico *in sample* a los datos reales. Para el modelo especificado, esto se evidencia en la *Gráfica 11*, donde se puede observar que se captura satisfactoriamente la dinámica de dichos datos.

Gráfica 11
Pronóstico *in sample* modelo ARIMA



Sin embargo, al realizar la prueba de la estimación con datos *out of sample* se observa que el modelo predice precios prácticamente constantes para casi todo el horizonte de tiempo y por tanto no hay ajuste ni logra capturar la dinámica de la serie real. Este resultado era el esperado *a priori* si se tiene en cuenta que los modelos ARIMA pueden tener un buen desempeño cuando son implementados en predicciones fuera de muestra en horizontes de tiempo cortos (pocos periodos de predicción). En la *Gráfica 12* se contrasta la predicción del modelo con los datos reales, correspondiendo el área sombreada a la parte de la serie *out of sample*.

Gráfica 12
Predicción *out of sample* modelo ARIMA



En vista de que inicialmente se observaron algunos problemas de volatilidad condicionada, se procedió a la realización de la prueba de multiplicadores de Lagrange para contrastar la presencia de heteroscedasticidad condicionada, obteniendo que la probabilidad asociada al estadístico del contraste se sitúa en la región de rechazo de la hipótesis nula de no existencia de efectos ARCH y por lo que se procedió a la modelación incluyendo dichos efectos. Los resultados de esta nueva estimación se presentan en la *Tabla 5*, resaltando que en el pronóstico *in sample* y en la predicción *out of sample* no se observaron cambios significativos.

Tabla 5
Estimación modelo ARIMA con estructura ARCH

	Coefficient	Std. Error	z-Statistic	Prob.
AR(1)	0.103051	0.030623	3.365146	0.0008
AR(4)	-0.152983	0.018344	-8.339763	0.0000
AR(6)	-0.133095	0.016347	-8.141761	0.0000
AR(10)	0.092072	0.013145	7.004094	0.0000
AR(12)	0.067894	0.015417	4.403866	0.0000
AR(13)	0.026233	0.012492	2.100069	0.0357
AR(29)	0.056778	0.010424	5.447061	0.0000
AR(30)	0.109644	0.009878	11.09976	0.0000
Variance Equation				
C	1230.424	145.4949	8.456821	0.0000
RESID(-1)^2	0.292637	0.034049	8.594561	0.0000
RESID(-2)^2	0.553308	0.047533	11.64045	0.0000
RESID(-3)^2	0.455523	0.049351	9.230288	0.0000
RESID(-4)^2	0.362672	0.040978	8.850428	0.0000
R-squared	0.000876	Mean dependent var	14.21358	
Adjusted R-squared	-0.009157	S.D. dependent var	278.9645	
S.E. of regression	280.2389	Akaike info criterion	12.92383	
Sum squared resid	93847922	Schwarz criterion	12.97868	
Log likelihood	-7792.993	Durbin-Watson stat	1.611702	

Los resultados de la estimación de modelos ARIMA con precios fueron superiores a los obtenidos al trabajar con variaciones logarítmicas, ya que con estas últimas, aparte de no lograr un buen desempeño fuera de muestra, se deteriora significativamente la capacidad de pronóstico dentro de la muestra.

Buscando mejorar la capacidad de predicción fuera de muestra se estimaron modelos de regresión múltiple a partir de las componentes principales de los precios. Como se mencionó previamente, el hecho de que los modelos estén siendo estimados sobre componentes principales garantiza que las variables de entrada cumplan con las condiciones exigidas por los supuestos teóricos sobre los que se construyen este tipo de modelos, además de evitar que ese ingrese información redundante a los mismos. Se emplearon las dos mismas divisiones de la muestra que en el caso de los ARIMA, obteniendo nuevamente mejores resultados para el caso de la partición 83%-17%.

La *Tabla 6* presenta los resultados de la estimación con Componentes Principales. Es importante resaltar que todas las componentes resultan ser estadísticamente significativos y además que se incluyen procesos AR(1), MA(1) y MA(6). Los dos primeros pueden ser interpretados como la influencia de la sesión anterior de mercado en la sesión actual y el MA(6) por su parte podría asociarse al “efecto día de la semana” que se presenta en los mercados.

Tabla 6
Estimación regresión múltiple de precios
de SURAMINV contra CP

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11660.41	996.5804	11.70042	0.0000
COMP1	1172.376	51.23022	22.88446	0.0000
COMP2	337.9347	77.37566	4.367455	0.0000
COMP3	-553.7213	48.25325	-11.47531	0.0000
COMP4	217.3527	58.84818	3.693448	0.0002
COMP5	543.3387	40.22689	13.50685	0.0000
AR(1)	0.994264	0.003562	279.1549	0.0000
MA(1)	0.159197	0.028394	5.606643	0.0000
MA(6)	-0.071996	0.028355	-2.539094	0.0112
R-squared	0.999292	Mean dependent var	9465.373	
Adjusted R-squared	0.999288	S.D. dependent var	6844.275	
S.E. of regression	182.6637	Akaike info criterion	13.26041	
Sum squared resid	41006860	Schwarz criterion	13.29764	
Log likelihood	-8199.196	F-statistic	216931.5	
Durbin-Watson stat	1.993945	Prob(F-statistic)	0.000000	

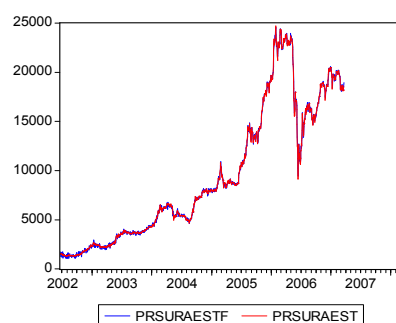
El R-cuadrado obtenido en la estimación es 0.999292, lo cual está indicando muy buen ajuste dentro de la muestra. Es importante mencionar que a pesar de que parece existir un cambio de régimen en el mercado a partir de la caída de los precios observada a mediados del 2006, el modelo logra capturar satisfactoriamente la dinámica que presenta la serie dentro de la muestra. Además, hay que resaltar que los residuales de la estimación, tal como se muestra en el correlograma presentado en la *Tabla 7*, son ruido blanco, por lo que se puede descartar la posibilidad de relaciones espúreas o problemas de estimación, a pesar de estar trabajando con precios.

Tabla 7
Correlograma residuales modelo regresión múltiple

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1	0.000	0.000	0.0003
		2	-0.012	-0.012	0.1875
		3	0.020	0.020	0.6790
		4	0.024	0.024	1.3850
		5	-0.012	-0.011	1.5607
		6	-0.007	-0.007	1.6216
		7	-0.044	-0.046	4.0726
		8	-0.001	-0.001	4.0731
		9	-0.003	-0.004	4.0878
		10	0.003	0.005	4.0965
		11	-0.035	-0.033	5.6376
		12	0.021	0.020	6.1804
		13	0.044	0.043	8.6159
		14	0.010	0.010	8.7495
		15	0.058	0.060	12.925
*	*	16	0.069	0.066	18.820
		17	-0.001	-0.001	18.820
		18	-0.014	-0.017	19.066
		19	-0.034	-0.038	20.532
		20	-0.026	-0.026	21.412
		21	-0.019	-0.017	21.879
		22	-0.056	-0.051	25.877
		23	-0.012	-0.004	26.074
		24	-0.009	-0.007	26.187
		25	0.049	0.050	29.258

El análisis del ajuste *in sample* permite concluir que igual que en el caso de los modelos ARIMA, la estimación por regresión múltiple tiene un muy buen desempeño dentro de esta parte de los datos, ya que como se observa en la *Gráfica 13*, captura adecuadamente la dinámica de la serie.

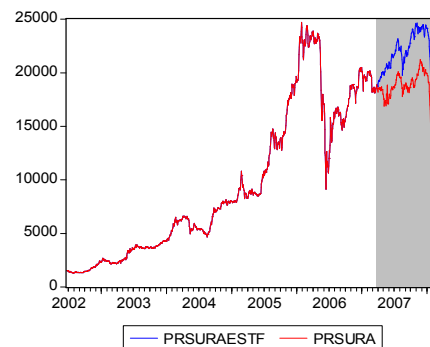
Gráfica 13
Pronóstico *in sample* modelo regresión múltiple



Cuando se extiende el análisis anterior a los datos *out of sample* se puede observar que se logra cierto avance respecto a la capacidad predictiva de los modelos ARIMA

en esta parte de los datos en la medida en que a pesar de no existir un buen ajuste, por lo menos se logra capturar algo de las tendencias de corto plazo de la serie de precios. El área sombreada de la *Gráfica 14* contrasta la predicción *out of sample* del modelo con los datos reales.

Gráfica 14
Predicción *out of sample* modelo regresión múltiple



Al estimar modelos de regresión múltiple de los rendimientos de SURAMINV con respecto a las 21 Componentes Principales obtenidos a partir de las series de variaciones, se observan resultados de estimación con una capacidad mucho menor de pronóstico *in sample* y predicción *out of sample*.

Un tercer grupo de modelos empleado fueron las redes neuronales, las cuales exigen la optimización del error de estimación a través de la comprobación de diferentes estructuras, estando esto relacionado con el tipo de red, número de capas, número de neuronas por capa, tipo de funciones de transferencia, aprendizaje y entrenamiento. La realización de este proceso implicó el entrenamiento de más de 4000 redes, por lo que tiene gran importancia dentro de la estructura de la red la utilización de algoritmos que sean eficientes desde el punto de vista de los tiempos computacionales. Además, dado que el proceso de optimización es no lineal, es necesario correr diferentes simulaciones para cada arquitectura variando los valores iniciales del vector de parámetros, los cuales en el trabajo son obtenidos aleatoriamente y respondiendo a una distribución uniforme entre (0,1). Se realizaron estimaciones con 20 variaciones de los valores de dicho vector.

Es importante mencionar que se entrenaron redes alimentadas con cuatro diferentes conjuntos de datos: las 77 series de variaciones logarítmicas de precios, las 21 series de sus Componentes Principales, las 77 series de precios y las 5 series de sus Componentes Principales. Para cada uno los grupos de datos se entrenaron redes *feedforward* y recurrentes de *Elman*. Potencialmente, este último tipo de redes es más poderoso que las *feedforward* debido a que permiten una retroalimentación en las neuronas de las capas ocultas y por tanto permiten construir historia de las neuronas en su evolución. La *Tabla 8* presenta los estadísticos del error de estimación para ambas clases de redes entrenadas a partir de las 77 series de rendimientos y de los 21 Componentes Principales de estas 77 series, resaltando con rojo la red que para cada uno de los estadísticos de evaluación presenta el mejor valor y en azul la segunda mejor.

Tabla 8
Estadísticos de error de estimación de las redes neuronales
entrenadas con rendimientos

		Redes <i>feedforward</i>		Redes recurrentes de <i>Elman</i>	
		77 rend	21 CP de 77 rend	77 rend	21 CP de 77 rend
<i>In sample</i>	R ²	<u>0.98605</u>	<u>0.97928</u>	0.97155	0.94008
	MSE	<u>0.00008</u>	<u>0.00013</u>	0.00018	0.00036
	RMSE	<u>0.00909</u>	<u>0.01123</u>	0.01353	0.01902
	RMSPE	<u>0.56020</u>	<u>0.99255</u>	1.53656	1.64843
	MAE	<u>0.00032</u>	<u>0.00026</u>	0.00060	0.00065
	MAPE	<u>0.21899</u>	<u>0.32735</u>	0.39105	0.47616
	HQIF	-4.47758	<u>-7.65493</u>	-1.80618	<u>-6.40394</u>
<i>Out of sample</i>	R ²	<u>0.00331</u>	0.00297	<u>0.00375</u>	0.00250
	MSE	<u>0.00034</u>	<u>0.00031</u>	0.00048	0.00039
	RMSE	<u>0.01839</u>	<u>0.01759</u>	0.02184	0.01983
	RMSPE	<u>1.90591</u>	<u>1.87247</u>	4.02150	2.70646
	MAE	<u>0.01210</u>	<u>0.01174</u>	0.01464	0.01373
	MAPE	<u>1.24261</u>	<u>1.25368</u>	1.96297	1.61311

El entrenamiento de redes alimentadas con rendimientos permitió llegar a modelos con muy buen ajuste *in sample* en la medida en que los coeficientes R-Cuadrado para todos los casos reportan valores superiores a 0.94. Sin embargo, se observa un claro deterioro del ajuste cuando se evalúan los resultados *out of sample*, donde el mejor coeficiente R-Cuadrado obtenido tiene un valor de 0.00375, lo cual es una clara muestra de que este conjunto de redes no logra capturar la dinámica de la serie estudiada más allá de los datos usados para el entrenamiento. Esta situación deja en evidencia la dificultad existente en las predicciones de rendimientos. Contrario a lo

esperado, se observa que en general las redes *feedforward* tienen un mejor desempeño que las recurrentes de *Elman*.

Al abordar la predicción desde el punto de vista de las 77 series de precios y de sus 5 Componentes Principales se observa una notable mejoría en el desempeño de los modelos. La *Tabla 9* resume los estadísticos de evaluación de estos modelos de manera similar que como se realizó con los rendimientos.

Tabla 9
Estadísticos de error de estimación de las redes neuronales
entrenadas con precios

		Redes <i>feedforward</i>		Redes recurrentes de <i>Elman</i>	
		77 precios	5 CP de 77 precios	77 precios	5 CP de 77 precios
In sample	R ²	<u>0.9988</u>	0.9981	0.9979	<u>0.9982</u>
	MSE	<u>53984.8926</u>	86882.7934	97825.9517	<u>85896.8661</u>
	RMSE	<u>232.3465</u>	294.7589	312.7714	<u>293.0817</u>
	RMSPE	<u>0.0324</u>	0.0354	0.0358	<u>0.0350</u>
	MAE	<u>148.9150</u>	197.8466	202.4942	<u>196.2940</u>
	MAPE	<u>0.0217</u>	0.0262	0.0264	<u>0.0261</u>
	HQIF	<u>13.4739</u>	18.9071	18.1701	<u>11.9041</u>
Out of sample	R ²	<u>0.8358</u>	<u>0.5493</u>	0.4980	0.4237
	MSE	<u>307155.4567</u>	<u>843874.0118</u>	897376.9724	1055005.6654
	RMSE	<u>554.2161</u>	<u>918.6262</u>	947.2998	1027.1347
	RMSPE	<u>0.0288</u>	<u>0.0505</u>	0.0528	0.0513
	MAE	<u>406.6969</u>	746.4488	<u>680.8780</u>	728.9252
	MAPE	<u>0.0223</u>	0.0408	<u>0.0375</u>	0.0393

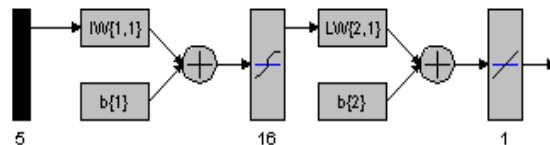
Aunque el ajuste dentro de la muestra al trabajar con rendimientos es bueno, cuando se hace el ejercicio con los precios se observan coeficientes R-Cuadrado superiores a 0.997 para todos los casos, valores que muestran que para este conjunto de datos los modelos logran capturar casi la totalidad de la dinámica de la serie. Es importante resaltar que en todos los estadísticos de error, a excepción del *HQIF*, la red que muestra mejor desempeño en la estimación es la *feedforward* alimentada con las series de los 77 precios. Es de esperar que para el *HQIF* esta misma red no reporte el mejor valor debido a que posee 77 neuronas en su capa de entrada, lo cual implica una gran cantidad de conexiones con la capa oculta.

Como se mencionó antes, la verdadera prueba acida para los modelos está en la capacidad de seguir presentando un buen ajuste sobre datos no usados para el

entrenamiento. Al observar los estadísticos de evaluación de los modelos *out of sample* se observa claramente que este grupo de redes tiene un alto poder predictivo si se compara con los otros modelos hasta ahora analizados. Es importante resaltar que contrario a lo esperado, las redes *feedforward* superan de manera importante a las recurrentes de *Elman* en esta parte de los datos. Concretamente, se observa que los coeficientes R-Cuadrado son 0.5493 y 0.8358 para las redes *feedforward* alimentadas con 5 CP y con 77 variables respectivamente, frente a valores de 0.4237 y 0.4980 observados en el caso de las redes de *Elman*. Esta superioridad se observa en la mayoría de los estadísticos de error.

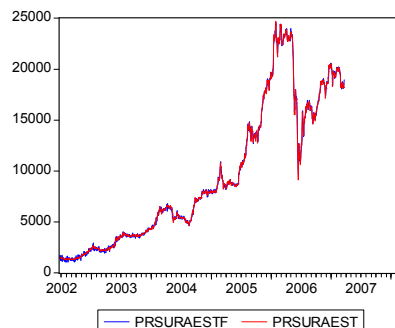
La *Gráfica 15* presenta la estructura de la red *feedforward* alimentada con los 5 CP de los precios que presentó mejor desempeño, resaltando que se emplea una capa oculta con 16 neuronas. Esta red trabaja con la función de transferencia *tansig* en su única capa oculta, *purelin* en la de salida, *LEARNGDM* como función de aprendizaje y *TRAINLM* como función de entrenamiento.

Gráfica 15
Estructura red neuronal artificial alimentada con 5 CP de los precios



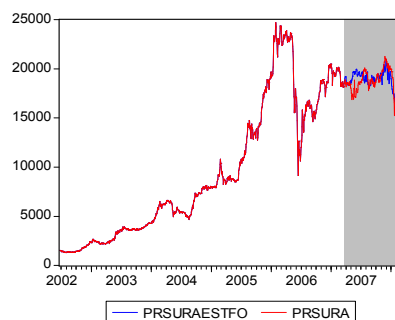
Acorde con lo que se observa en los estadísticos de error, la *Gráfica 16* muestra que la red captura totalmente la dinámica de los datos dentro de la muestra.

Gráfica 16
Pronóstico *in sample* red neuronal artificial
alimentada con 5 CP de los precios



Al analizar el desempeño fuera de la muestra, presentado en la *Gráfica 17*, se encontró que con esta red se logra un avance importante frente al modelo de regresión múltiple previamente presentado en la medida en que se sigue capturando la tendencia de corto plazo de la serie pero con un ajuste que supera ampliamente al del modelo mencionado.

Gráfica 17
Predicción *out of sample* red neuronal artificial
alimentada con 5 CP de los precios

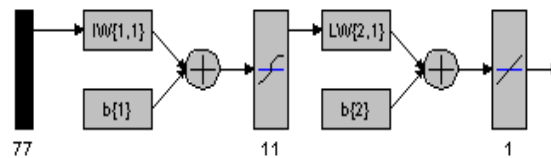


Es importante verificar las características de los residuales de la red, por lo que se realizaron diferentes pruebas estadísticas que nos permitieran mostrar que los residuales eran una variable aleatoria ruido blanco.

Hay que destacar que si bien las redes *feedforward* superan en desempeño a las recurrentes, los resultados obtenidos entre las redes del primer tipo alimentadas con

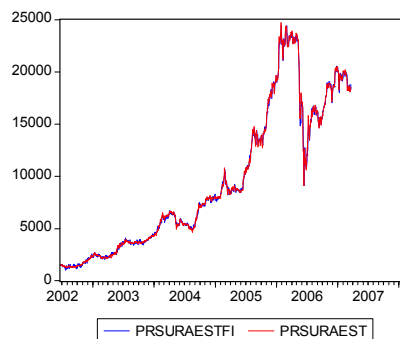
los 5 CP y con las 77 variables de precios difieren de forma importante, ya que esta última, de manera consistente con los resultados *in sample*, es la que presenta los mejores resultados en los estadísticos de error *out of sample*, resaltando la gran ventaja que tiene frente a las otras. La *Gráfica 18* presenta la estructura de esta red, donde se puede observar que posee una sola capa oculta con 11 neuronas. Además, emplea la función de transferencia *tansig* en su única capa oculta y *purelin* en la capa de salida, a la vez que trabaja con *LEARNGDM* como función de aprendizaje y *TRAINLM* como función de entrenamiento.

Gráfica 18
Estructura red neuronal artificial alimentada con 77 series de precios



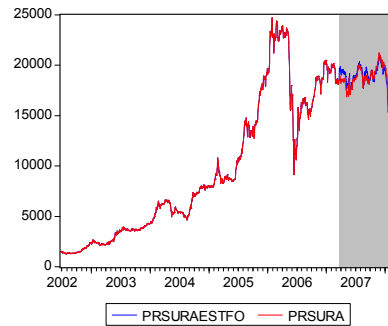
En la *Gráfica 19* se puede observar el pronóstico de esta red contra los datos observados.

Gráfica 19
Pronóstico *in sample* red neuronal artificial
alimentada con 77 series de precios



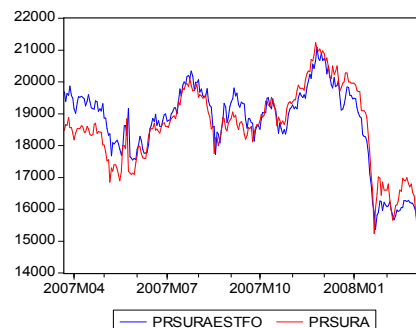
Igualmente, en la *Gráfica 20* se presenta en el área sombreada la predicción fuera de muestra del modelo y los datos verdaderamente observados, resaltando que tal como lo muestran los estadísticos de evaluación, se logra capturar gran parte de la dinámica de la serie.

Gráfica 20
Predicción *out of sample* red neuronal artificial
alimentada con 77 series de precios



La *Gráfica 21* es una extracción de la *Gráfica 20* en la que sólo se muestra la predicción y los datos reales para la parte *out of sample* de los datos.

Gráfica 21
Extracción predicción *out of sample* red neuronal artificial
alimentada con 77 series de precios



De manera similar que con la red anterior, al someter los residuales del modelo a pruebas estadísticas formales, se encontró que los mismos son ruido blanco.

La *Tabla 10* presenta un resumen de los coeficientes R-Cuadrado de las estimaciones realizadas con base en los 77 precios y en los 5 CP de los 77 precios para el modelo de regresión múltiple y los dos tipos de redes. Se observa claramente la buena capacidad de pronóstico de todos los modelos *in sample* sin mostrar grandes

diferencias, pero salta a la vista la superioridad de las redes en los datos *out of sample*, principalmente las *feedforward* y resaltando nuevamente dentro de ellas el desempeño de la que es alimentada con los 77 precios. Se presentan en rojo los modelos con mejor desempeño dentro de cada conjunto de datos y en azul los de segundo mejor desempeño.

Tabla 10
Coefficientes R^2 diferentes modelos

	<i>In sample</i>	<i>Out of sample</i>
Regresión múltiple 5 CP, AR(1), MA(1) y MA(6)	0.9993	0.3593
Red recurrentes de <i>Elman</i> 5 CP de 77 precios	0.9979	0.4237
Red recurrentes de <i>Elman</i> 77 precios	0.9979	0.4980
Red feedforward 5 CP de 77 precios	0.9981	0.5493
Red feedforward 77 precios	0.9988	0.8358

Como se mencionó previamente, además de las particiones de la muestra 83% - 17% y 68% - 32%, en el entrenamiento de las redes se empleó una división *in sample* – *out of sample* intercalada, en la que se reservaron todos los datos que en el ordenamiento total de la serie tienen posiciones que sean múltiplos de 4 (4, 8, 12, 16, etc.) para efectos de prueba del modelo y se usaron los restantes en el entrenamiento. La finalidad de esta partición era determinar si al ser las redes entrenadas con datos representativos del 100% de la muestra total se podía llegar a mejor capacidad de generalización en datos no conocidos. Se encontró, tal como se esperaba, que estas redes presentan un muy buen ajuste *in sample* y *out of sample*, pero que en términos prácticos tendrían poca aplicación por la misma forma en la que se está haciendo la separación de los datos.

Dado que la red *feedforward* alimentada con las 77 series de precios presenta resultados tan sobresalientes, se realizó a manera de ejercicio la transformación de los precios que predice el modelo en rendimientos. Con esto se busca determinar si de esta manera indirecta (a través de los precios) se logra capturar la dinámica de los rendimientos de la acción, objetivo no alcanzado de forma satisfactoria cuando se trataron de predecir directamente con los modelos. Se encontró que se presenta un claro deterioro del pronóstico *in sample* (R -Cuadrado = 0.2098) y de la predicción *out of sample* (R -Cuadrado = 0.0003). Este resultado es predecible si se tiene en cuenta que una buena predicción de precios no tiene por qué traducirse en una buena

predicción de rendimientos. Además, hay que resaltar que en términos financieros esta situación no desmerita las predicciones de precios en el sentido de que pueden ser igualmente aprovechables desde el punto de vista de retornos de exceso independientemente de que al devolverlos a rendimientos no se obtengan buenos resultados de predicción.

Por último, en la *Tabla 11* se presenta un resumen de los diferentes estadísticos de error utilizados para cada uno de los modelos estimados con base en las series de precios o en las componentes principales de las mismas.

Tabla 11
Estadísticos de error de estimación de los modelos entrenados con precios

		ARIMA		RegMult	Redes <i>feedforward</i>		Redes recurrentes de <i>Elman</i>	
		Sin ARCH	Con ARCH	5 CP de 77 precios	77 precios	5 CP de 77 precios	77 precios	5 CP de 77 precios
<i>In sample</i>	R ²	0,9985	0,9983	0,9993	0,99880	0,99810	0,9979	0,9982
	MSE	67459,8293	75805,48	33123,0295	53984,89260	86882,79340	978259,9517	85896,8661
	RMSE	259,7303	275,328	181,997	232,34650	294,75890	312,7714	293,0817
	RMSPE	0,0239	0,0245	0,0385	0,03240	0,03540	0,0358	0,035
	MAE	144,1493	147,3529	125,1301	148,91500	197,84660	202,4942	196,294
	MAPE	0,0153	0,0152	0,0225	0,02170	0,02620	0,0264	0,0261
	HQIF	11,1568	11,2734	10,4223	13,47390	18,90710	18,1701	11,9041
<i>Out of sample</i>	R ²	0,0002	0,0017	0,3593	0,83580	0,54930	0,498	0,4237
	MSE	1694641	1806636	13533654	307155,45670	843874,01180	897376,9724	1055005,665
	RMSE	1301,7838	1344,112	3678,811	554,21610	918,62620	947,2998	1027,1347
	RMSPE	0,0726	0,0732	0,2039	0,02880	0,05050	0,0528	0,0513
	MAE	1022,1072	1095,485	3431,8611	406,69690	746,44880	680,878	728,9252
	MAPE	0,0559	0,0592	0,1875	0,02230	0,04080	0,0375	0,0393

7. Evaluación financiera de las predicciones

Luego de tener predicciones del comportamiento de la acción de SURAMINV en el mercado, es importante determinar si por medio de sistemas simples de negociación¹¹ elaborados sobre dichas predicciones hubiera sido posible obtener rendimientos extraordinarios con respecto a una estrategia pasiva representada en comprar y mantener o *Buy and Hold (B&H)* como comúnmente se le conoce. El objetivo final que se busca con este ejercicio es evaluar la existencia de uno de los pilares de la teoría

¹¹ Un sistema de negociación hace referencia a un conjunto de condiciones programadas para arrojar señales de compra o venta en el mercado.

de eficiencia de mercado en forma débil, representado en la hipótesis de que los agentes, a través de información obtenida con base en las series históricas de precios, no pueden obtener rendimientos extraordinarios con respecto a una estrategia pasiva.

La comparación de los sistemas de negociación frente a la estrategia *B&H* se realizó con base en la predicción de la red *feedforward* alimentada con las 77 series de precio, para todo el periodo de estudio (junio 2002 – marzo 2008), cada uno de los años de dicho periodo y para la parte *in sample* y *out of sample* de los datos con la separación 83% - 17%. El ejercicio fue realizado con base en los siguientes supuestos:

- Sólo se permiten posiciones *largas*¹², lo cual es acorde con la regulación del mercado accionario colombiano para el periodo de estudio.
- La estrategia *B&H* y los sistemas alternativos siempre inician con un valor de portafolio correspondiente al valor de mercado de 100 acciones de SURAMINV en el primer día del periodo considerado para la comparación. Esta es una suma en efectivo que busca actualizar el disponible que tiene cada estrategia para empezar a negociar con base en los precios de la acción y de esta manera disminuir los efectos del apalancamiento, los cuales podrían distorsionar los resultados de la comparación.
- Todas las compras son de 100 acciones. De esta manera, la estrategia *B&H* compra 100 acciones el primer día de cada uno de los periodos de análisis, las cuales mantendrá hasta el último día de dicho periodo. Por su parte, los sistemas alternativos realizan compras y ventas de 100 acciones de acuerdo a las señales resultantes de las condiciones bajo las cuales están programados. Se prefirió esta forma de comparación por encima de una estrategia en la que se invierte siempre el 100% de los recursos disponibles por el hecho de que en esta última, por las características de la serie estudiada, los montos de inversión se disparan mucho más pronto en la medida en que los sistemas logran acumular ganancias muy rápido, las cuales se traducen en crecimientos del tamaño de la posición en cuanto al número de acciones, produciendo

¹² Nombre que reciben las posiciones de compra en el mercado cuando se está hablando de apuestas direccionales sobre los precios.

resultados que para efectos de comparación no son convenientes en la medida en que la estrategia *B&H* siempre sigue invertida en las mismas 100 acciones.

- Los faltantes o sobrantes de liquidez del portafolio son financiados a la tasa mínima de expansión del Banco de la República¹³, con lo cual se busca que las estrategias tengan en cuenta el costo de los recursos desde el punto de vista de la posición propia de una entidad financiera.
- Se tiene en cuenta un costo de transacción que representa la diferencia entre el *Bid* y el *Ask*¹⁴ del mercado en el momento de realizar las operaciones. Este costo es calculado como el margen proporcional efectivo ilustrado en la ecuación (48).

$$\text{Margen Prop. Efectivo} = 2 \times \frac{|\text{Precio transacción} - \text{Precio medio}|}{\text{Precio medio}} \quad (48)$$

Agudelo (2009) realiza un estudio de los “costos de transacción y medidas de liquidez intradiaria en BVC”, estimando que el margen promedio proporcional efectivo para la acción de SURAMINV en el periodo comprendido entre el 24 de abril y el 3 de agosto es 0.271%. Suponiendo que el 50% de las veces es necesario agredir al mercado por lo que se asume el costo de transacción y que el otro 50% asume el costo la contraparte, se decidió trabajar con un margen promedio correspondiente a la mitad del estimado en el estudio mencionado, lo cual corresponde a un valor de 0.1355% por operación. En el ejercicio se trabajó desde el punto de vista de un inversionista representado en la posición propia de una entidad financiera por lo que no tienen en cuenta otros costos de transacción que podrían ser relevantes, como es el caso de las comisiones.

- Se asume que los modelos se corren cada día 5 minutos antes del cierre del mercado, teniendo la posibilidad incluso de realizar operaciones al cierre del

¹³ En la página web del Banco de la República esta tasa es definida como “la tasa que las entidades financieras deben pagar por tener acceso a la liquidez temporal a través de los repos que se ofrecen por subasta, es decir, con un monto limitado. La tasa a la que se da liquidez a través de la subasta oscila entre esta tasa mínima y la tasa máxima o lombarda de expansión”.

¹⁴ *Bid* y *Ask* son los nombres que reciben el mejor precio de demanda y de oferta del mercado en un momento determinado.

mismo día. Para efectos de simplificación, en el ejercicio se supone que los precios existentes 5 minutos antes son *proxies* que responden exactamente a los datos verdaderos de cierre.

Es importante resaltar que los resultados de la estrategia *B&H* se ven fuertemente influenciados por la tendencia de los precios. En una fase puramente alcista como la que se presenta desde junio de 2002 hasta finales de 2005, la estrategia *B&H* por estar siempre comprada tiene ventaja frente a cualquier sistema que entre y salga del mercado en la medida en que este último puede perder parte de la tendencia alcista. Caso contrario es lo que sucede en la segunda parte de la serie tomada desde principio de 2006 hasta marzo de 2008, periodo en el cual los precios se movieron principalmente de forma lateral con algunos saltos importantes a la baja, condiciones que ponen en desventaja a la estrategia *B&H* por nunca poder salir del mercado y por tanto no tener la posibilidad de evitar estas caídas, como si lo pueden hacer los sistemas alternativos. Por esta razón es de esperar que en la primera parte de la serie la estrategia *B&H* tenga ventaja frente a cualquier otro sistema, mientras que en la segunda parte la situación sea contraria. Sin embargo, un buen sistema de negociación para mercados con fuertes tendencias debería seguir muy de cerca la estrategia *B&H* durante épocas alcistas de precios y ganarle de forma importante en movimientos laterales o bajistas.

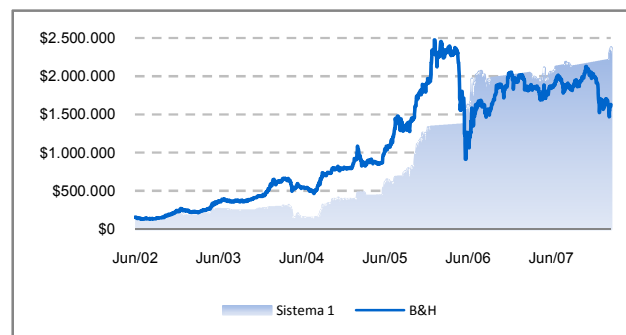
Las estrategias de negociación están basadas de manera importante en las predicciones, evitando la utilización de herramientas de análisis técnico complejas que puedan quitar relevancia a la predicción. De esta manera, el primer sistema considerando, denominado *Sistema 1*, se compone de dos condiciones de negociación muy sencillas y basadas puramente en la predicción obtenida:

- Si no hay posición y el retorno esperado por la RNA es mayor que 1,8% se origina una señal de compra.
- Si hay una posición *larga* en el mercado y el retorno esperado por la RNA es menor que 1,8% se origina una señal de venta.

El nivel del rendimiento esperado que determina las señales de compra y venta fue obtenido por medio de un proceso de optimización en el que se buscaba maximizar el retorno de la estrategia para el periodo de estudio.

El *Sistema 1* alcanza un rendimiento de 61.59% efectivo anual frente a 51.23% que obtiene la estrategia *B&H*. A pesar de este resultado, en la *Gráfica 22* se observa una amplia ventaja del valor del portafolio de la estrategia *B&H* durante la tendencia alcista de los precios y no se evidencia una buena capacidad del *Sistema 1* de mantener las ganancias obtenidas en la fase lateral – bajista del mercado. Aunque es de esperar que el sistema pierda en la primera parte y gane en la segunda, la evolución del valor de su portafolio frente a la estrategia *B&H* hace pensar que no tiene un buen desempeño en la medida que no cumple con la característica esperada para un buen sistema de negociación.

Gráfica 22
Valoración portafolio *Sistema 1* y *B&H*



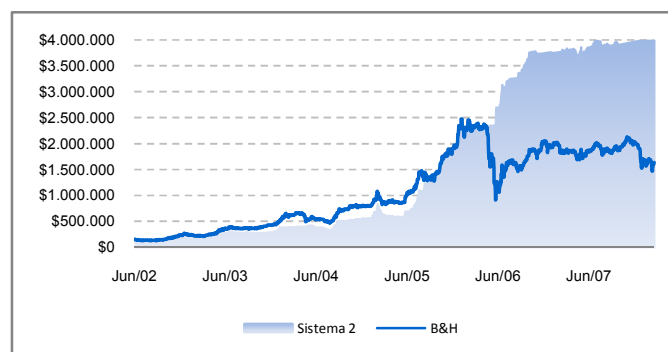
El resultado obtenido con el *Sistema 1* es previsible si se tiene en cuenta que las condiciones de programación arrojan señales con relativa facilidad, las cuales, al ser determinadas sólo por el dato esperado para un día, no tienen en cuenta la tendencia dentro de la que puede estar el mercado, perdiendo de esta manera partes importantes del movimiento al alza. Además, la realización de un número tan alto de operaciones termina traducándose en altos costos de transacción. En vista de estos resultados, se decidió incluir herramientas básicas de análisis técnico para tratar de aprovechar la predicción del precio pero teniendo en cuenta la tendencia. Se probaron promedios móviles de diferentes tipos (simple, ponderado y exponencial) obteniendo

un nuevo sistema de *trading* que denominamos *Sistema 2* y que trabaja bajo las siguientes condiciones:

- Si no hay posición se evalúa si ante un cumplimiento exacto de la predicción para la siguiente jornada, la serie de precios de cierre cruzaría hacia arriba el promedio móvil ponderado¹⁵ de 3 días de dicha serie. En caso de presentarse esta situación, se genera una señal de compra al precio de cierre del día actual.
- Si hay posición abierta y el cruce es en sentido contrario se genera una señal de venta.

El número de días del promedio móvil ponderado es el resultado de un proceso de optimización en el que se buscaba maximizar los rendimientos del *Sistema 2*. La *Gráfica 23* presenta la evolución del valor de los portafolios para este nuevo ejercicio, resaltando que si bien la estrategia *B&H* obtiene mejores resultados en la tendencia al alza, las ganancias del *Sistema 2* logran seguirla muy de cerca. La brecha que se presenta durante cierto periodo de tiempo está explicada en gran parte por los costos de transacción, ya que al hacer simulaciones sin incluirlos se observó un desempeño muy similar de ambas estrategias en la tendencia alcista.

Gráfica 23
Valoración portafolio *Sistema 2* y *B&H*



¹⁵ Los promedios móviles ponderados son ampliamente usados como herramientas de análisis técnico de mercados. Se calculan mediante una ponderación lineal de los precios, multiplicando el primero (más reciente) por n , el segundo por $(n-1)$, el tercero por $(n-2)$ y así sucesivamente, donde n es la periodicidad del promedio móvil. El resultado final es luego dividido por la suma de las ponderaciones.

Este es precisamente el comportamiento que se espera de un buen sistema de trading en épocas de tendencia al alza, ya que sólo pierde partes relativamente pequeñas del movimiento. Es importante mencionar que incluso en el 2005, año de fuerte tendencia positiva, el *Sistema 2* logra superar a la estrategia *B&H*. En términos agregados para todo el periodo de estudio, el sistema de negociación logra obtener un rendimiento anualizado de 78.27% frente a 51.23% que obtiene la estrategia *B&H*.

En el periodo comprendido entre principio de 2006 y marzo de 2008 y caracterizado por movimientos a la baja y laterales, se observa que el *Sistema 2* logra una ventaja muy importante frente a la estrategia *B&H*, la cual es consistente durante todo este tiempo. Inicialmente, al presentarse la caída de los precios a mitad de 2006, el *Sistema 2* logra mantener prácticamente constante el valor del portafolio en la medida que se mantuvo por fuera del mercado. Luego se observa que el sistema es capaz de aprovechar la recuperación de los precios posterior a dicha caída, incrementando de manera significativa el valor del portafolio. La parte final de la serie está caracterizada por movimientos laterales de los precios con una caída relevante en los últimos días, periodo que nuevamente el *Sistema 2* aprovecha para consolidar el valor de su portafolio, ya que antes que disminuirlo, logra conservar una leve pendiente alcista en el mismo. La evolución de la estrategia *B&H* durante esta parte de la serie fue totalmente contraria a la descrita para el sistema de negociación, en la medida en que tuvo que soportar todo el rigor de los movimientos adversos en el precio de la acción.

Se quiso realizar el análisis comparativo igualmente para la parte *out of sample* de la serie, obteniendo los resultados presentados en la *Gráfica 24* en los que se evidencia nuevamente la superioridad del *Sistema 2* frente a la estrategia *B&H*, en la medida en que es capaz de mantener el valor del portafolio prácticamente constante en una época de movimientos laterales y bajistas del mercado.

Gráfica 24
Valoración portafolio *Sistema 2* y *B&H out of sample*



Es conveniente detallar la evolución del valor de portafolio en diferentes subperiodos que permitan evaluar si la ventaja que presenta el *Sistema 2* frente a La estrategia *B&H* se mantiene de forma consistente. En este orden de ideas, la *Tabla 12* presenta un resumen de los rendimientos anualizados para todo el periodo de estudio, para cada uno de los años que lo componen y para la parte *in sample* y *out of sample* de los datos. Se observa nuevamente que a pesar de que en los tres primeros años el sistema de negociación es superado, logra mantener buenos niveles de rendimiento en la época de tendencia alcista.

En el año 2005, caracterizado también por una fuerte tendencia al alza, el sistema es capaz de superar a la estrategia *B&H*, lo cual evidencia las bondades que posee para capturar los movimientos positivos del mercado. En los años restantes, tal como fue descrito, se observan rendimientos muy superiores del sistema. Estos resultados, además de los generales de todo el periodo de estudio, llevan a pensar que el *Sistema 2* tiene buena consistencia para obtener rendimientos extraordinarios en el mercado.

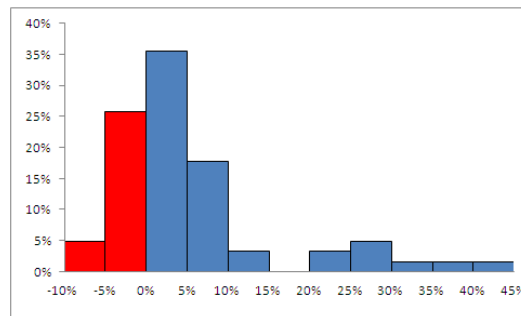
Tabla 12
Rendimientos anualizados Sistema 2 y estrategia B&H

Año	Sistema 2	B&H
2002	65,98%	118,35%
2003	27,85%	87,14%
2004	59,05%	86,17%
2005	<u>176,68%</u>	146,35%
2006	<u>94,05%</u>	4,17%
2007	<u>3,77%</u>	-2,56%
2008	<u>50,99%</u>	-65,23%
<i>In Sample</i>	<u>96,82%</u>	69,20%
<i>Out of Sample</i>	<u>2,80%</u>	-14,20%
jun 2002 - mar 2008	<u>78,27%</u>	51,23%

En términos agregados se observa que el desempeño promedio del sistema de negociación, si se tienen en cuenta las diferentes tendencias del mercado, es superior al de la estrategia B&H, ya que en las épocas de subida tiene un comportamiento ligeramente inferior al de la estrategia pasiva, pero en épocas de movimientos laterales o bajista obtiene rendimientos muy superiores. Es de esperar que al contar con mayor historia de los precios de la acción y tener en cuenta mayor cantidad de movimientos alcistas, laterales y bajistas, la ventaja del Sistema 2 sea aun más marcada, ya que la muestra actual está caracterizada por una fuerte tendencia al alza en la mayor parte del tiempo.

En la evaluación del Sistema 2 es importante analizar la calidad de las señales que genera a través del histograma del retorno porcentual de sus *trades*. El eje vertical de la Gráfica 25 muestra la frecuencia relativa para cada uno de los rangos de rendimientos que se muestran en el eje horizontal, observando claramente que de los 62 *trades* realizados por el sistema, más del 70% tuvieron una variación porcentual positiva entre el precio de salida y el de entrada con una alta concentración entre 0% y 10%. Esta es una característica deseable en un sistema de negociación en la medida en que indica que su retorno total no es el resultado de unos pocos negocios con rentabilidades extraordinarias, sino que por el contrario se deriva de muchos *trades* que en términos agregados ganan más de lo que pierden.

Gráfica 25
Histograma del retorno porcentual de los *trades* del Sistema 2



Por último, se realizó un análisis de sensibilidad de los resultados de la comparación con respecto a los costos de transacción buscando determinar el deterioro de los resultados del *Sistema 2* a medida que se tuvieran mayores costos. Se realizaron simulaciones duplicando e incluso triplicando dicho costo, obteniendo respuestas muy similares a las iniciales. En el primero de los casos se obtuvo un rendimiento anualizado de 76.93% para todo el periodo de estudio, mientras que en el segundo se consiguió 75.53%, valores que siguen siendo significativamente superiores al 51.23% alcanzado por la estrategia *B&H*.

8. Conclusiones

La investigación permitió contrastar la teoría de eficiencia débil de mercado en la acción de Suramericana de Inversiones S.A. SURAMINV, acción líder en el mercado bursátil colombiano dentro del periodo considerado para el estudio: junio 2002 – marzo 2008. En términos generales se obtuvo que en este mercado, contrario a la hipótesis defendida por esta teoría, si es posible hacer predicciones confiables y obtener rendimientos extraordinarios con respecto a una estrategia *Buy & Hold*. El trabajo fue más allá que muchos otros de su estilo en el sentido en que además de lograr un buen pronóstico *in sample*, demuestra que es posible obtener buenas estimaciones del comportamiento de la acción *out of sample* y por tanto evalúa de manera más confiable la verdadera posibilidad de aprovechamiento financiero de los resultados.

Partiendo de 77 variables expresadas en términos de precios y rendimientos logarítmicos, se empleó el análisis de componentes principales para reducir el tamaño de la muestra, evitar la utilización de información redundante y garantizar el cumplimiento de las condiciones exigidas por los supuestos teóricos de algunos de los modelos trabajados. Los datos disponibles fueron divididos en dos grupos buscando controlar la posibilidad de *data snooping*: el primero, *in sample*, se usa para el ajuste de los modelos y el segundo, *out of sample*, que no interviene en el entrenamiento, se emplea para evaluar la capacidad predictiva de los modelos. Se usaron diferentes particiones para este fin buscando determinar la sensibilidad de la capacidad de generalización de los modelos a la muestra en la que eran entrenados. Se obtuvo que efectivamente este aspecto era de gran importancia en los resultados, destacándose la división 83% *in sample* – 17% *out of sample* por generar entrenamientos más robustos es cuanto a capacidad de pronóstico y predicción.

La evidencia empírica del trabajo dejó claramente expuesta la superioridad de los modelos construidos para la predicción de precios sobre los de rendimientos. Aunque este es un resultado esperado, es importante resaltar que la diferencia en cuanto a poder de predicción es muy significativa en la medida en que los modelos construidos sobre retornos no capturan de forma adecuada la dinámica de los datos *out of sample*. El trabajo con precios no representa una dificultad si se tiene en cuenta que a partir de una buena predicción de los mismos, se pueden obtener retornos aprovechables desde el punto de vista financiero. Además, al trabajar con precios es importante mencionar que para la construcción de los modelos se aplicó toda la rigurosidad que exige el desarrollo teórico de cada uno de ellos, teniendo especial cuidado en evitar que los resultados sean el reflejo de relaciones *espureas*.

La modelación se abordó inicialmente como un problema de series de tiempo, empleando estructuras ARIMA con efectos ARCH y GARCH. Se obtuvo un buen ajuste del pronóstico *in sample* a los datos reales, pero un desempeño inadecuado en la parte *out of sample* en la medida en que los modelos no fueron capaces de ajustarse a los precios observados fuera de muestra, ni de capturar su dinámica. Este resultado fue acorde a lo esperado si se tiene en cuenta que los modelos ARIMA suelen no tener buen desempeño en un horizonte largo de predicción fuera de

muestra, periodo para el cual generalmente termina proyectando un dato casi constante. En vista de lo anterior, se optó por explorar la predicción a través de modelos de regresión múltiple de los precios de la acción contra sus componentes principales, estimación que arrojó como resultado un buen ajuste *in sample*. Al hacer la evaluación de la predicción *out of sample*, se observó que se logró un avance importante respecto a los modelos ARIMA en lo que tiene que ver con la captura de la dinámica de la serie real, pero sigue existiendo una brecha de ajuste muy importante.

Las redes neuronales artificiales son una herramienta de la inteligencia artificial que ha venido ganando importancia en problemas de predicción de series financieras por ciertas bondades que poseen en la estimación de patrones no lineales. Por este motivo se decidió modelar las series de precios y rendimientos de SURAMINV por medio de redes *feedforward* y recurrentes de *Elman*, empleando en ambos casos el algoritmo de entrenamiento *backpropagation*. Para estas estimaciones se emplearon cuatro conjuntos de datos para alimentar los modelos, a saber: las 77 series de precios, las 77 de rendimientos y las componentes principales de cada uno de estos grupos.

La evaluación de los resultados de las redes muestran que al igual que en el caso de los modelos econométricos, se logra capturar adecuadamente la dinámica de los datos *in sample*. Sin embargo, en las predicciones de precios se observa un notorio avance respecto al desempeño de los modelos fuera de la muestra en el sentido en que se logra capturar la dinámica de la serie real mucho mejor que con los modelos econométricos y a la vez se mejora considerablemente el ajuste. Contrario a lo esperado, las redes *feedforward* presentaron mejor desempeño fuera de muestra que las recurrentes; además, como ya se mencionó, los resultados en cuanto a poder de predicción fueron considerablemente superiores al trabajar con precios que con rendimientos. Las redes con mejor desempeño fueron en su orden la *feedforward* alimenta con las 77 series de precios y la *feedforward* alimenta con las 5 componentes principales de los precios, cuyos resultados superan ampliamente a los de todos los otros modelos empleados en el estudio.

Tal como se planteó desde el principio, el trabajo no se queda solamente en un ejercicio de predicción, sino que regresa al campo financiero en la medida en que

evalúa la posibilidad de aprovechar esa predictibilidad desde el punto de rendimientos de exceso, con miras a contrastar la teoría de eficiencia débil de mercado. Esta evaluación se realizó por medio de la comparación en diferentes periodos entre sistemas de negociación basados en las predicciones de la red de mejor resultado y la estrategia pasiva de comprar y mantener.

En la evaluación financiera de la predicción se obtuvo que, contrario a lo que defiende la teoría de eficiencia débil de mercado, por medio de los sistemas de negociación es posible obtener en el largo plazo rendimientos muy superiores a los de la estrategia pasiva con un alto nivel de confianza. Es importante hacer énfasis en que en ciertos subperiodos de alta tendencia alcista de los precios la estrategia pasiva resulta tener rendimientos superiores. Sin embargo, al evaluar a nivel agregado las diferentes tendencias presentadas durante el periodo de estudio, el desempeño promedio del sistema de negociación resulta ser superior. Es de esperar que al contar con una serie más representativa de los diferentes movimientos de precios que puede tener este activo en el mercado, este resultado tienda a consolidarse de manera más importante e incluso tienda a ser aun más superior la ventaja del sistema.

Luego de la realización de este trabajo y de los resultados descritos, es importante resaltar la importancia de profundizar en las herramientas de inteligencia artificial para el estudio de los mercados financieros colombianos. En este orden de ideas, proponemos los siguientes puntos como motivación para futuros trabajos:

- El empleo de redes neuronales artificiales buscando realizar nuevos análisis, tales como predicciones de diferentes plazos (intradía y de largo plazo), identificación de tendencias y generación de señales de negociación por medio de la implementación de funciones de transferencia de umbral.
- La implementación de otras técnicas de Inteligencia Artificial como la *Lógica Difusa*, las cuales fusionadas con las redes podrían arrojar mejores resultados.

Referencias bibliográficas

Acosta, M., Salazar, H. y Zuluaga, C. (2000). Tutorial de Redes Neuronales. Universidad Tecnológica de Pereira, Facultad de Ingeniería Eléctrica. <http://ohm.utp.edu.co/neuronales/>

Agudelo, D. y Uribe, J. (2007). ¿Realidad o Sofisma?: Poniendo a Prueba el Análisis Técnico en las Acciones Colombianas. Universidad EAFIT, Medellín.

Agudelo, D. (2009). Costos de transacción y medidas de liquidez intradiaria en BVC. Trabajo presentado en el VI Simposio Nacional y III Internacional de Docentes en Finanzas, Bogotá, Colombia.

Alvarez, M. y Alvarez, A. (2003). Predicción No Lineal de Tipos de Cambio: Algoritmos Genéticos, Redes Neuronales y Fusión de Datos. Documentos de trabajo, Universidad de Vigo. N.1.

Aristizabal, M. (2006). Evaluación asimétrica de una red neuronal artificial: aplicación al caso de inflación en Colombia. Borradores de economía N° 377, Banco de la República de Colombia.

Ayala, M. (2006). Usando redes neuronales en espacios no estándar para la predicción en series de tiempo financieras. UPM.

Azoff, E. M. (1996). Neural Network: Time Series Forecasting of Financial Markets, Wiley, A Wiley Finance Edition.

Beltratti, A. & Terna, M. P. (1995). Neural Networks for Economics and Financial Networks. University of Turin, International Thompson Computer Press, Italy.

Chen, G., Firth, M. & Rui, O. (2002). Stock Market Linkages: Evidence from Latin America. Journal of Banking & Finance, N.26, 1113-1141.

Chen, A.S., Leung, M.T. & Daouk, H. (2003). Application of Neural Networks to an Emerging Financial Market: Forecasting and Trading the Taiwan Stock Index. Computers & Operations Research, Vol. 30, No. 6.

Enders, W. (1995). Applied Econometric Time Series. Ed. Wiley, 1995. 448 p.

Eum, C. & Shim, S. (1989). International transmission of stock market movements. Journal of Financial and Qualitative Analysis, N.24, 241-256.

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, American Finance Association, vol. 25(2), 383-417.

Fernandez, F., González, C. & Sosvilla, S. (2000). On the profitability of technical trading rules based on artificial neural networks:: Evidence from the Madrid stock market," *Economics Letters*, Elsevier, vol. 69(1), 89-94.

Garcia, I. (2003). Análisis y predicción de la serie de tiempo del precio externo del café colombiano utilizando redes neuronales artificiales. *Universitas Scientiarum*, revista de la facultad de ciencias. Pontificia Universidad Javeriana. Vol.8, 45-50.

Gately, E. (1996). *Neural networks for financial forecasting*, Wiley Trader's Advantage, John Wiley & Sons.

Gil, M. y Pérez, F. (2005). Análisis y predicción de la acción de la empresa Acerías Paz del Río utilizando un modelo GARCH y redes neuronales artificiales. *Revista Ingenierías*. Universidad de Medellín v.4, N.7, 83 – 97.

Green, H. William (1999). *Análisis econométrico*. Prentice Hall.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Segunda Edición. Prentice Hall. Estados Unidos.

Hannan, E. J. & Quinn, B. (1979). The determination of the order of an autoregression', *Journal of the Royal Statistical Society, Series B* 41, 190–191.

Hua, J., Yann, Y. (1996). Stock market trend prediction using ARIMA-based neural networks. *International Conference on IEEE*. Volume 4, Issue , 3-6, 2160 – 2165.

Jalil, A. y Misas, M. (2007). Evaluación de pronósticos del tipo de cambio utilizando redes neuronales y funciones de pérdida asimétricas. *Revista Colombiana de Estadística*. Vol. 30 N. 1, 143 – 161.

Johnston, J. (1992). *Métodos de econometría*. Vicens Vives.

Kasa, K. (1992). Common stochastic trends in international stock markets. *Journal of Monetary Economics* N.29, 95-124.

Kelly, J. (2007). The Ultimate Money Machine. *Revista Bloomberg*, pág. 38, Junio de 2007.

Kock, P. y Kock, T. (1991). Evolution in dynamic linkages across daily national Indexes. *Journal of International Money and Finance*.

Lawrence, R. (1997). Using Neural Networks to Forecast Stock Market Prices. University of Manitoa.

Lutkepohl, H. (2006). New Introduction to multiple time series analysis. Springer, 764p.

Mandic, D. P. & Chambers, J. A. (2001) Recurrent Neural Networks for Prediction, John Wiley and Sons.

MATLAB. Neural Network Toolbox User's Guide.

McNELIS, P. D. (2005). Neural Networks in Finance: Gaining Predictive Edge in the Market. Elsevier. Reino Unido.

Medina, S., Jaramillo, J. y Mejía, M. (2008). Pronóstico del valor de un fondo común ordinario colombiano mediante un sistema de redes neuronales. E.I.O, facultad de Minas, Universidad Nacional de Colombia, Medellín.

Pantoja, María. (2006). Analysis of time series forecasting with neuronal networks, ARIMA models and GARCH process for non stationary time series. Department of Industrial Engineering, Andes University, Bogotá D.C.

Parisi, A., Parisi, F. y Guerrero, J. L. (2003). Redes Neuronales en la Predicción de Indices Accionarios Internacionales. El trimestre Económico 70, 721-744.

Phylaktis, K. & Ravazzolo, F. (2001). Measuring financial and economic integration with equity prices in emerging markets. Mimeo, City University Business School.

Phylaktis, K., & Ravazzolo, F. (2005). Stock market linkages in emerging markets: implications for international portfolio diversification. Journal of International Financial Markets, Institutions and Money, Vol. 15, N. 2, Págs 91-106.

Pindyck, R. y Rubinfeld, D. (1998). Econometría modelos y pronósticos. Mc graw Hill.

Pertuz, J. (2005). Modelos para Pronosticar Series de Tiempo Financieras con Redes Neuronales Backpropagation Dependiente del Tiempo. Seminario Día Matlab, Ponencia #2, UDES. Noviembre.

Quintero, L. (2002). Las redes neuronales como herramienta para la predicción. Revista Ingenierías Universidad de Medellín, 121 – 136, Medellín.

Roberts, H. (1967). Statistical versus clinical predictions of the stock markets. Unpublished manuscript, Center for research in Security Prices, University of Chicago, May.

Rumelhart, D. E. & McClelland, J.L. (1986). Parallel Distributed Processing. Vol. 1 and 2, MIT Press.

Sallehuddin, R. Hj., Shamsuddin, S. M., Mohd. Hashim, S. Z. & Abraham, Ajith. (2009). Forecasting Time Series Data Using Hybrid Grey Relational Artificial Neural Network and Auto Regressive Integrated Moving Average. Applied Artificial Intelligence, v.23 n.5, 443-486.

Shachmurove, Y. (2000). Utilizing Artificial Neural Network Model to Predict Stock Markets. CARESS Working Paper #00-11.

Uriel, E. y Aldás, J. (2005). Análisis Multivariante Aplicado. Thomson. 531p.

Villada, F., Cadavid, D. y Molina, J. (2008). Pronóstico del precio de la energía eléctrica usando redes neuronales artificiales. Revista facultad de ingeniería, Universidad de Antioquia N. 44, 111-118. Junio.

Wong, B. K. (1995). A bibliography of neural network business application research: 1988-September 1994, Expert Systems 12, 3.

Yao J., Li, Y. & Tan, C. (1997). Forecasting the exchange rates of CHF vs USD using neural networks. Journal of Computational Intelligence in Finance, 5, 2, 7-13.

Yao, J. & Tan, C. (2000). A case study on using neural networks to perform technical forecasting of Forex. Neurocomputing N.34, 79-98.

ANEXO 1. Series de datos de entrada para los modelos

Acciones Colombianas	Suramericana Acerías Paz del Río Banco de Bogotá Bancolombia O Bancolombia P ADR – Bancolombia Colombiana de Inversiones Coltejer Chocolates Éxito Fabricato Grupo Aval ISA Tablemac Valorem Cemargos Inverargos	Monedas	TRM Euro Yen Libra Franco suizo Real Peso mexicano Peso chileno Peso argentino Won surcoreano Dólar taiwan Rublo ruso Yuan chino Dólar canadiense Dólar australiano Lira turca
Indices Accionarios	IGBC (Colombia) Dow Jones (EUA) S&P 500 (EUA) NASDAQ COMPOSITE (EUA) S&P EUROPE 350 (Europa) FTSE 100 (Reino Unido) CAC 40 (Francia) DAX (Alemania) NIKKEI 225 (Japón) SHANGHAI COMPOSITE (China) HANG SENG (China) SENSEX 30 (India) BOVESPA (Brasil) BMV (México) MERVAL (Argentina) IPSA (Chile) IGBVL (Perú)	Tasas	Tasa de la FED Tasa BanRep Tasa Banco de Japón Tasa del Banco Central Europeo Tasa del Banco del Reino Unido Tasa de Brasil Tasa genérica tesoros EUA 10 años LIBOR 1 Semana LIBOR 1 Mes LIBOR 3 Meses LIBOR 6 Meses ITES DTF
Futuros	Futuro Genérico Dow Jones Futuro Genérico S&P Futuro Genérico Nasdaq	Commodities	Crudo de petróleo WTI Crudo de petróleo Brent Precio del café colombiano Oro E-mini crudo de petróleo Futuros del precio del café
		Otros	VIX EMBI Global EMBI+ Colombia Spread deuda colombiana UVR