MODELOS PREDICTIVOS DE REDES NEURONALES EN ÍNDICES BURSÁTILES*

*Antonino Parisi F.,** Franco Parisi F.****
*y José Luis Guerrero C.*****

RESUMEN

Este estudio analiza la capacidad de los modelos de redes neuronales para predecir el signo de las variaciones semanales de los índices bursátiles CAC40, Hang Seng, KLSE, MMX, STI, Dow Jones Industry, S&P500, GDAX, Bovespa, Nikkei225 y FTSE100, entendiendo que la predicción de la dirección del movimiento del índice accionario es pertinente para desarrollar estrategias de transacción efectivas (Leung, Daouk y Chen, 2000). Se usaron modelos de redes neuronales de algoritmo de aprendizaje supervisado de propagación hacia atrás: el perceptrón multicapa, la red recurrente Jordan-Elman y la red *Ward*. Por otra parte, el proceso de evaluación se hizo sobre la base de 51 conjuntos extramuestrales, cada uno compuesto por 50 observaciones semanales. En esta etapa, el desempeño relativo de los modelos fue medido por el número de predicciones correctas (*hits*) del signo de la variación del índice, aplicando para ello la prueba de acierto direccional de Pesaran y Timmermann (1992). Los resultados muestran que la capacidad predictiva de los modelos varía a lo largo del tiempo, por lo que es necesario no sólo recalcular los coeficientes de la ecuación periodo a periodo sino también reelaborar el modelo en sí, por lo que no existiría un único modelo explicativo de la evolución de los índices bursátiles. Finalmente, al comparar los resultados de la red *Ward* y del modelo ARIMA con los de una estrategia *buy and hold* se observó que, independientemente de la significancia estadística de la capacidad predictiva, ambas técnicas permitieron aumentar la rentabilidad o reducir las pérdidas.

* *Palabras clave*: redes neuronales, red *Ward*, *test directional accuracy*, capacidad predictiva. *Clasificación JEL*: G10, G14, G15. Artículo recibido el 10 de septiembre de 2002 y aceptado el 17 de febrero de 2003. Los autores desean agradecer al director y a un dictaminador anónimo de EL TRIMESTRE ECONÓMICO sus importantes comentarios que claramente enriquecieron el artículo; también a Edinson Cornejo, por su labor como asistente de investigación, y los comentarios realizados por Jorge Gregoire, Fernando Bravo, Jorge Niño, Alex Silemberger y a los participantes en el seminario realizado en el Departamento de Administración de la Universidad de Chile. Premio al mejor artículo de investigación en la III Conferencia Internacional de Finanzas organizado por la Universidad de Santiago de Chile, enero de 2003. Todos los errores y omisiones que persistan son de exclusiva responsabilidad de los autores.
** Profesor asistente, Departamento de Administración, Universidad de Chile.
*** Profesor asistente, Management Department, Jesse Jones Graduate School of Management, Rice University (correo electrónico: fparisi@rice.edu).
**** Profesor asociado, MacDonough School of Business, Georgetown University.

Abstract

This study analices the neuronal networks models' capacity to predict the sign of the weekly variations of CAC40, Hang Seng, KLSE, MMX, STI, Dow Jones Industry, S&P500, GDAX, Bovespa, Nikkei225 and FTSE100. The relative performance of the models was measured by the number of correct predictions of the index's variation sign, on the base of 51 joint out-samples, each one made up of 50 weekly observations. The results proved that the predictive capacity of the models change through the time, then it is necessary to reestimate the weight of the equation and reconstruct the model period after period. This suggest that an exclusive and unique explanatory model of the stock-exchange indice's evolution does not exist.

Introducción

Este estudio analiza la capacidad de los modelos de redes neuronales para predecir el signo de las variaciones semanales de los índices bursátiles CAC40, Hang Seng, KLSE, MMX, STI, Dow Jones Industry, S&P500, GDAX, Bovespa, Nikkei225 y FTSE100, comparado con la capacidad predictiva de un modelo ARIMA (1,1,1). Lo anterior, entendiendo que la predicción de la dirección del movimiento del índice accionario es pertinente para desarrollar estrategias de transacción efectivas, las cuales pueden arrojar mejores resultados que las basadas en la proyección del valor de la variable observada (Leung, Daouk y Chen, 2000).

El análisis se centrará en el empleo de las redes neuronales multicapas con aprendizaje supervisado, específicamente el perceptrón multicapa y algunas de sus variantes. Se utilizarán estos modelos de redes ya que actúan como funciones "mapeadoras" universales, desempeñándose muy bien con las series de tiempo. A su vez, se emplearán distintas arquitecturas de redes neuronales y un modelo ARIMA (1,1,1), a fin de evaluar y comparar la capacidad predictiva de cada uno de ellos.

Los operadores financieros que usan el análisis técnico creen que los precios y volumen accionarios históricos contienen implícitamente información útil relativa a los movimientos futuros de precios y que, examinando estos datos y con una cierta dosis de sentido común, es posible detectar pautas de comportamiento en las series de precios accionarios que permitan realizar pronósticos de su evolución futura. En palabras de Malkiel (1981), "el análisis técnico es un anatema para el mundo académico". No obstante, éste goza de gran popularidad entre los agentes del mercado al ser, por sus características, un instrumento dirigido sobre todo a inversionistas de corto plazo y especuladores.

Varios estudios han concluido que existe evidencia significativa de que los precios accionarios no siguen un camino aleatorio y muestran que los rendimientos accionarios son predecibles en algún grado. Por ejemplo, Lo y MacKinley (1988), empleando datos de mercados bursátiles desarrollados, como los de los Estados Unidos, la Europa Occidental y Japón, registraron una correlación serial positiva entre los rendimientos semanales; Conrad y Kaul (1988, 1989) también encontraron evidencia de predictibilidad de los rendimientos en el corto plazo; DeBondt y Thaler (1985), Fama y French (1988), Poterba y Summers (1988) y Chopra, Lakonishok y Ritter (1992) hallaron una correlación serial negativa en los rendimientos de los activos individuales y varias carteras de intervalos de tres a diez años, es decir, en el largo plazo. Por su parte, Jegadeesh (1990) examinó la predictibilidad de los rendimientos mensuales de activos individuales y encontró una correlación serial negativa de primer orden altamente significativa para rezagos de dos meses y una correlación serial positiva para rezagos mayores. Blume, Easley y O'Hara (1994) sugieren que existe una relación significativa entre los rezagos del volumen transado y los rendimientos actuales de los activos individuales.

Al momento de explicar la predictibilidad de las variaciones de los rendimientos accionarios se postulan dos argumentos: *i*) los mercados son ineficientes y los precios de los activos se mueven alrededor de su valor fundamental, y *ii*) los mercados son eficientes y la predictibilidad de las variaciones puede explicarse por un equilibrio en los rendimientos *time-varying*. Al respecto Ferson y Harvey (1991) muestran que la predictibilidad de los rendimientos accionarios no se debe forzosamente a ineficiencias del mercado o a una exageración de los inversionistas irracionales, sino a la predictibilidad que presentan algunas variables agregadas que son parte del conjunto de la información que explica la rentabilidad de los activos. Por su parte, Leung, Daouk y Chen (2000) señalan que la predicción de los rendimientos accionarios, dadas las variables agregadas en el conjunto de información de los inversionistas, es un hecho aceptado en la bibliografía de las finanzas empíricas, y las preguntas apuntan hacia cómo usar la información de manera óptima para predecir y comerciar en los mercados.

La mayoría de las prácticas de transacción adoptadas por los analistas financieros confían en la predicción precisa de los precios de los instrumentos financieros. No obstante, estudios recientes han sugerido que las estrategias de transacción basadas en proyecciones de la dirección del cambio de precios son más efectivas y pueden generar ganancias más altas que las basadas en una predicción puntual de los precios de los instrumentos financieros. En esta línea, Leung, Daouk y Chen (2000) compararon la capacidad predictiva de los modelos de clasificación[1] con los de estimación de nivel[2] y concluyeron que los primeros se desempeñan mejor que los segundos, en términos de su tasa de acierto, es decir, son capaces de generar beneficios más altos. Maberly (1986) analizó la relación entre la dirección de los cambios de precio interdía e intradía, y O'Connor, Remus y Griggs (1997) apoyan la utilidad de proyectar la dirección del cambio en los precios más que el nivel de precios en sí. Lo anterior resulta pertinente para la comunidad financiera, en el sentido de que deben centrar sus esfuerzos en predecir la dirección de los movimientos o el signo de la variación de los precios bursátiles, en vez de minimizar la desviación de las estimaciones de los valores observados. Más aún, Hodgson y Nicholls (1991) sugieren evaluar la significación económica de predecir la dirección de los cambios en los precios de los activos y no su nivel.

Los modelos de redes neuronales han sido desarrollados para predecir valores de índices bursátiles y de activos individuales, situándose la mayoría de las primeras investigaciones y aplicaciones en mercados establecidos en los Estados Unidos (Bosarge, 1993; Tsibouris y Zeidenberg, 1995; White, 1993), la Gran Bretaña (Tsibouris y Zeidenberg, 1995) y Japón (Yoda, 1994). Dichos modelos han sido usados para predecir los rendimientos de índices bursáti-

[1] *Linear discriminant analysis, logit model, probit model* y *probabilistic neural network*.

[2] *Adaptive exponential smoothing, vector autoregression with Kalman filter, multivariate transfer function* y *multilayered feedforward neural network*.

les, entre otras aplicaciones relacionadas con la toma de decisiones en los ámbitos de finanzas e inversión (Hawley, Johnson y Raina, 1990; Refenes, 1995). Cabe señalar que aun cuando existen estudios que muestran la fortaleza de las redes neuronales en la predicción de series de tiempo no lineales con respecto a los métodos estadísticos tradicionales, se requieren estudios adicionales para evaluar la aplicación de estos modelos a la proyección y a la toma de decisiones, con objeto de hacerlos instrumentos confiables para los analistas encargados de realizar pronósticos, carencia que intentamos cubrir en parte con esta investigación.

Es por lo anterior que estudiamos la capacidad predictiva del cambio en el signo de los rendimientos accionarios a partir de modelos de redes neuronales. Es interesante notar que las pruebas realizadas permiten concluir que estas técnicas aumentan la rentabilidad o reducen las pérdidas en la administración activa de carteras indizadas. Por otra parte, se pudo observar que una mayor capacidad predictiva no siempre se traduce en mayores rendimientos. También se observó que la capacidad predictiva de los modelos varía a lo largo del tiempo, por lo que es necesario no sólo recalcular los coeficientes de la ecuación periodo a periodo, sino también el modelo en sí.

Este artículo se divide en dos secciones: la I explica la metodología empleada en la investigación y la sección II aborda el análisis de los resultados. Al final se presenta las conclusiones y recomendaciones del estudio.

I. Metodología y datos

Los índices bursátiles CAC40, Hang Seng, KLSE, MMX, STI, Dow Jones Industry, S&P500, GDAX, Bovespa, Nikkei225 y FTSE100 fueron seleccionados por su importancia internacional y la liquidez de sus mercados. Se utilizaron datos de cierre semanal de las variables incluidas en cada modelo, correspondientes al periodo comprendido entre el 25 de octubre de 1993 y el 22 de julio de 2002. En los modelos formulados la variable de salida está dada por la variación del índice bursátil que se desea proyectar, correspondiente al periodo t, mientras que las variables de entrada se refieren a las variaciones de un conjunto de variables que se detallan más adelante. Los modelos incluyen rezagos de las variables de entrada, para así considerar el desfase lógico que existe entre el momento en que se realiza la predicción, el periodo actual o $(t-1)$, y el momento futuro para el cual la proyección es válida, al que nos referimos como periodo (t). Así, debido a que la información de hoy está desfasada con respecto a la proyección, se utilizan modelos dinámicos con rezago en las variables independientes.

Las redes neuronales pueden entenderse como modelos multiecuacionales o multietapas, en los que el *output* de unos constituye el *input* de otros. En el caso de las redes multicapas, existen etapas en las cuales las ecuaciones operan de manera paralela. Los modelos de redes neuronales, al igual que, por ejemplo, los modelos de suavizamiento exponencial y de análisis de regresión, utilizan *inputs* para generar un *output* a manera de una proyección. La diferencia radica en que las redes neuronales incorporan inteligencia artificial en el proceso que conecta los *inputs* con los *outputs* (Kuo y Reitsch, 1995-1996).

Según Martín del Brío y Sanz (1997) las redes neuronales artificiales (RNA) "son sistemas de procesamiento que copian esquemáticamente la estructura neuronal del cerebro para intentar reproducir sus capacidades". En consecuencia, son una clase de modelos no lineales flexibles que se caracterizan por ser sistemas *paralelos*,[3] *distribuidos*[4] y *adaptativos*.[5] Según estos autores, estos tres conceptos se traducen en un mejor rendimiento y en una mayor velocidad de procesamiento. Por su parte, Herbrich, Keilbach, Graepel, Bollmann-Sdorra y Obermayer (2000) señalan que la característica más importante de las redes neuronales es su capacidad para aprender dependencias basadas en un número finito de observaciones, en las que el término "aprendizaje" significa que el conocimiento adquirido a partir de las muestras puede ser empleado para proporcionar una respuesta correcta ante datos no utilizados en el entrenamiento de la red. La bibliografía sugiere que las redes neuronales poseen varias ventajas potenciales sobre los métodos estadísticos tradicionales, destacándose que éstas pueden ser aproximaciones de funciones

[3] Cuentan con una gran cantidad de neuronas o procesadores elementales (PE), cada uno de los cuales trabaja paralelamente con una pequeña parte de un problema mayor.

[4] Cuentan con muchos PE a través de los cuales distribuyen su memoria.

[5] Tienen la capacidad de adaptarse al entorno modificando sus pesos y sinapsis para encontrar una solución aceptable al problema.

universales aun para funciones no lineales (Hornik, Stinchcombe y White, 1989), lo que significa que ellas pueden aproximar automáticamente cualquier forma funcional (lineal o no lineal) que mejor caracterice los datos, permitiéndole a la red extraer más señales a partir de formas funcionales subyacentes complejas (Hill, Marquez, O'Connor y Remus, 1994).

En la realización de nuestro estudio se emplearon diversos modelos de redes neuronales que usan el algoritmo de aprendizaje supervisado de propagación hacia atrás, con objeto de predecir el signo de la variación de los índices bursátiles analizados. Estos modelos se caracterizan porque en ellos el resultado es conocido y la red se entrena a sí misma hasta que es capaz de predecir el resultado asociado con los datos de entrada (Dasgupta, Dispensa y Ghose, 1994). Los modelos analizados son: el perceptrón multicapa, la red recurrente Jordan-Elman y la red *Ward*.

El programa de redes neuronales utilizado es el Neuroshell 2, el cual requiere para su funcionamiento que la base de datos se divida en tres conjuntos diferentes: *i*) conjunto de entrenamiento[6] (conjunto de datos empleados para que la red aprenda el problema); *ii*) conjunto de prueba (conjunto de datos utilizados para evitar el sobreaprendizaje[7] de la red y que, unido a las observaciones del conjunto de entrenamiento, se extiende desde el 25 de octubre de 1993 al 1º de agosto de 2000),[8] y *iii*) conjunto de producción o extramuestral (datos no incorporados anteriormente, que fueron usados para probar la capacidad de predicción de la red ante datos que nunca ha visto, y que pertenecen al subperiodo que abarca desde el 7 de agosto de 2000 al 22 de julio de 2002, contando así con 100 datos

6 En una red neural hay pares de *inputs* y *outputs* que son usados para entrenar la red. Puede haber múltiples *inputs* (variables explicativas) y múltiples *outputs* (proyecciones de diferentes variables). Entre los *inputs* y los *outputs* hay una capa (o múltiples capas) de procesamiento que imitan el trabajo del cerebro humano. Luego, dado un nuevo conjunto de *inputs*, la red puede producir un nuevo *output* (proyección) sobre la base de lo que aprendió de los pares de *inputs* y *outputs* que le fueron provistos. El analista puede controlar algunos aspectos del proceso, como la tasa de aprendizaje y la precisión deseada del *output* (Wilson y Keating, 1998).

7 El sobreajuste a los datos de la muestra o sobreaprendizaje de la red se produce cuando el sistema se ajusta demasiado a los datos de entrenamiento, aprendiendo incluso el ruido presente en ellos, por lo que crece el error ante pautas diferentes de los empleados en el entrenamiento y disminuye la precisión de la proyección).

8 El periodo comprendido por los datos del conjunto de entrenamiento y del conjunto de pruebas se extiende, en algunos casos, hasta el 7 de agosto de 2000 y, en otros, hasta el 14 de agosto de 2000, como consecuencia de las diferencias de calendario que caracteriza a los países analizados.

para constituir 51 periodos de prueba, de 50 observaciones semanales cada uno). Cabe señalar que otros autores dividen el periodo de datos en dos segmentos, en los que el primero es utilizado para entrenar la red mientras que el segundo es usado para determinar y validar su arquitectura y la especificación del modelo (Chen y Leung, 1998; Leung, Daouk y Chen, 2000). Respecto a las limitaciones que presenta el programa, éstas se refieren principalmente a *i*) la gran cantidad de datos que se requiere para conformar los conjuntos de entrenamiento, de prueba y extramuestral, y *ii*) la falta de una función que permita realizar el proceso de recursividad de manera automática.

Luego de probar numerosas arquitecturas y parámetros, los mejores resultados en términos de capacidad predictiva se obtuvieron con la red *Ward* de tres capas: una de entrada, una oculta y otra de salida. En términos generales, la capa de entrada (*Slab* 1) posee una neurona por cada *i*-ésima variable de entrada, la capa oculta posee *j* neuronas (las cuales se dividen en partes iguales entre los *Slab* 2 y 3), y la capa de salida (*Slab* 4) posee una neurona. De esta manera, la salida neta de la capa oculta viene dada por la ecuación (1):

$$i_{pj}^{h} = f_{j}^{h}(Neta_{pj}^{h}), \qquad \text{con } j = 1, 2, \ldots, n \tag{1}$$

en la expresión (2) se emplea para los *Slabs* 2 y 3,[9] respectivamente:

$$Neta_{pj}^{h} = \begin{bmatrix} w_{11} & w_{12} & \Lambda\, w_{1i} \\ w_{21} & w_{22} & \Lambda\, w_{2i} \\ M & M & \Lambda\, M \\ w_{\left[\frac{j}{2}\right]1} & w_{\left[\frac{j}{2}\right]2} & \Lambda\; w_{\left[\frac{j}{2}\right]i} \end{bmatrix}^{h}_{\left(\left[\frac{j}{2}\right]\times i\right)} \times \begin{bmatrix} x_1 \\ x_2 \\ M \\ x_i \end{bmatrix}_{(i\times 1)} + \begin{bmatrix} \theta_1 \\ \theta_2 \\ M \\ \theta_{\left[\frac{j}{2}\right]} \end{bmatrix}^{h}_{\left(\left[\frac{j}{2}\right]\times 1\right)} = \begin{bmatrix} Neta_1 \\ Neta_2 \\ M \\ Neta_{\left[\frac{j}{2}\right]} \end{bmatrix}^{h}_{\left(\left[\frac{j}{2}\right]\times 1\right)} \tag{2}$$

La salida neta de la capa de salida queda expresada de acuerdo con la ecuación (3), para los *Slabs* 2 y 3:

$$o_{pk} = f_{k}^{o}(Neta_{pk}^{o}) \tag{3}$$

en la que:

[9] En los modelos usados para predecir los índices Hang Seng, MMX, STI, Dow Jones Industry, GDAX y FTSE100, la capa oculta posee tres *Slabs*, por lo que las ecuaciones (2) y (4) se emplean para los *Slabs* 2, 3 y 4. Para ello, la expresión $[j/2]$ debe ser remplazada por $[j/3]$.

$$Neta^{o}_{pk} = \left[w_{11} \;\; w_{12} \;\Lambda\; w_{1\left[\frac{j}{2}\right]} \right]^{o}_{\left(1\times\left[\frac{j}{2}\right]\right)} \times \begin{bmatrix} i_1 \\ i_2 \\ M \\ i_{\left[\frac{j}{2}\right]} \end{bmatrix}^{h}_{\left(\left[\frac{j}{2}\right]\times 1\right)} + [\theta_1]^{o}_{(1\times 1)} = [Neta_1]^{o}_{(1\times 1)} \qquad (4)$$

Véase una explicación más detallada en el apéndice, el cual también contiene una descripción de las formas funcionales de los modelos de redes neuronales usados para predecir el signo de la variación de los índices bursátiles. Las características de las redes *Wards* usadas para proyectar se presentan en el cuadro 1, mientras que las reglas de aprendizaje de propagación hacia atrás se presentan en el cuadro 2.

Al momento de predecir el signo de las variaciones que componen el conjunto extramuestral se empleó un proceso recursivo, metodología que ha sido empleada para medir el desempeño de modelos de redes neuronales que buscan predecir periodos de recesión en los Estados Unidos (Qi, 2001; Estrella y Mishkin, 1998). De esta manera, si consideramos que el conjunto extramuestral tiene m datos, al evaluar el funcionamiento de la red se consideró sólo la predicción del primer valor. Posteriormente, el dato analizado sale del conjunto extramuestral (quedando con $m-1$ datos) y pasa a formar parte del conjunto de entrenamiento, siguiendo un proceso aleatorio, por lo que la muestra de n datos que contiene los valores de entrada se incrementa a $n+1$. Luego se realizó una nueva iteración, lo que implicó reelaborar los pesos del modelo para cada una de las $m-1$ proyecciones, permitiéndole a la red aprender del error cometido en la predicción y ajustar de nuevo los pesos estimados para los valores de entrada de las capas de salida y oculta, w_{jk} y w_{ij}, respectivamente. Este proceso se repitió hasta que en el conjunto extramuestral quedó sólo un dato, por lo que la red recalculó los pesos w_{jk} y w_{ij} hasta el momento en que la última observación (correspondiente a $t-1$) es empleada para proyectar el valor que la variable de salida podría alcanzar en el momento t, el cual representa el futuro inmediato.

El proceso de evaluación empírica se hizo sobre la base de los datos de los 51 conjuntos extramuestrales, en los que cada uno de ellos se compone de 50 observaciones semanales. En esta etapa, el desem-

CUADRO 1. *Arquitectura de las redes neuronales utilizadas para proyectar las variaciones de los índices bursátiles. Funciones de activación*

Índice	*Red*	*Slab 1*	*Slab 2*	*Slab 3*	*Slab 4*	*Slab 5*
CAC40	*Ward*	Tanh	Gaussiana	Gauss. complemento	Logistic	
Hang Seng	*Ward*	Tanh	Tanh	Tanh	Gauss. complemento	Tanh
KLSE	*Ward*	Tanh	Gaussiana	Gauss. complemento	Logistic	
MMX	*Ward*	Tanh	Gaussiana	Tanh	Gauss. complemento	Logistic
STI	*Ward*	Tanh	Gaussiana	Tanh	Gauss. complemento	Logistic
DJI	*Ward*	Tanh	Gaussiana	Tanh	Gauss. complemento	Logistic
GDAX	*Ward*	Linear (-1,1)	Gaussiana	Tanh	Gauss. complemento	Logistic
Bovespa	*Ward*	Tanh	Gauss. complemento	Gaussiana	Logistic	
S&P500	*Ward*	Tanh	Gaussiana	Gauss. complemento	Logistic	
Nikkei225	*Ward*	Tanh	Gaussiana	Gauss. complemento	Logistic	
FTSE	*Ward*	Tanh	Tanh	Tanh	Gauss. complemento	Logistic

CUADRO 2. *Arquitectura de las redes neuronales utilizadas para proyectar las variaciones de los índices bursátiles. Tasas de aprendizaje y de momento*

Índice	*Link 1*		*Link 2*		*Link 3*		*Link 4*		*Link 5*		*Link 6*	
	Apren-dizaje	*Momento*	*Apren-dizaje*	*Momento*	*Apren-dizaje*	*Momento*	*Apren-dizaje*	*Momento*	*Apren-dizaje*	*Momento*	*Apren-dizaje*	*Momento*
CAC40	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01				
Hang Seng	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
KLSE	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01				
MMX	0.01	0.1	0.01	0.1	0.01	0.1	0.01	0.1	0.01	0.1	0.01	0.1
STI	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
DJI	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
GDAX	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Bovespa	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01				
S&P500	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01				
Nikkei225	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01				
FTSE	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01

peño relativo de los modelos fue medido por el número de predicciones correctas (*hits*) del signo de la variación del índice. Así, al igual que Kanas (2001), se aplicó la prueba de certeza direccional (DA) de Pesaran y Timmermann (1992) a fin de medir la precisión direccional de los modelos de proyección. La prueba compara el signo de la proyección, $\hat{y}_{n+i}$, con el del valor observado, y_{n+i}, para cada i-ésima observación del conjunto extramuestral ($i = 1, 2, \ldots, m$), en el que el signo indica la dirección en la que se moverá el mercado accionario: al alza, si es positivo, o a la baja, si es negativo. Si los signos coinciden, aumentan la efectividad de la red, y en caso de no existir coincidencia, aumenta el error de predicción de la red. La proposición de éxito (SR) se define como:

$$SR = m^{-1} \sum_{i=1}^{m} I_i[y_{n+i}, \hat{y}_{n+i} > 0] \tag{5}$$

en el que $I_i[\cdot]$ es una función indicador que toma el valor de 1 cuando su argumento es cierto y 0 en otro caso. Además:

$$P = m^{-1} \sum_{i=1}^{m} I_i[y_{n+i} > 0] \tag{6}$$

y

$$\hat{P} = m^{-1} \sum_{i=1}^{m} I_i[\hat{y}_{n+i} > 0] \tag{7}$$

La proporción de éxito en el caso de independencia de $\hat{y}_{n+i}$ y y_{n+i}, SRI, está dada por:

$$SRI = P \cdot \hat{P} + (1 - P)(1 - \hat{P}) \tag{8}$$

cuya varianza es:

$$VAR[SRI] = m^{-2}[m(2 \cdot \hat{P} - 1)^2 P(1 - P) + m(2 \cdot P - 1)^2 \hat{P}(1 - \hat{P}) + \\ + 4 \cdot P \cdot \hat{P}(1 - P)(1 - \hat{P})] \tag{9}$$

Por su parte, la varianza de SR se define como:

$$VAR[SR] = m^{-1} \cdot SRI(1 - SRI) \tag{10}$$

Finalmente, la prueba DA de Pesaran y Timmermann (1992) está dado por:

$$DA = (VAR[SR] - VAR[SRI])^{-1/2}(SR - SRI) \tag{11}$$

Pesaran y Timmermann (1992) mostraron que, según la hipótesis nula de que $\hat{y}_{n+i}$ y y_{n+i} están independientemente distribuidos, dicha prueba sigue una distribución normal estándar. La prueba fue aplicada en los resultados de cada uno de los modelos de proyección.

La capacidad de los modelos de redes neuronales para predecir el signo de la variación de un índice bursátil determinado, junto con la rentabilidad que un inversionista habría obtenido de haber seguido las recomendaciones de compra y venta arrojadas por el modelo, fue contrastada con la capacidad predictiva y con el rendimiento de un modelo ARIMA. A su vez, los resultados de ambos fueron comparados con la rentabilidad generada por una estrategia *buy and hold*.[10] Los modelos ARIMA, calculados recursivamente y usados para predecir el signo de la variación de los índices bursátiles, fueron los siguientes:

$$\Delta Index_{it} = \alpha_1 \cdot \Delta Index_{it-1} + \alpha_2 \cdot \varepsilon_{t-1}, \tag{12}$$

en el que Δ representa la primera diferencia entre los precios de cierre semanales de cada uno de los índices bursátiles abordados en el estudio; α_1 y α_2 corresponden a los coeficientes que acompañan a las variables independientes y son estimados por el modelo, y ε_{t-1} representa el término de error del periodo anterior, el cual es incluido en el modelo como variable explicativa.

II. Análisis de los resultados

El cuadro 3 resume la capacidad predictiva promedio de los 51 periodos extramuestrales analizados[11] y el número de periodos en los cuales ésta resultó estadísticamente significativa.[12] Además, presenta la desviación estándar promedio de los índices bursátiles y la rentabilidad acumulada promedio que había logrado un inversio-

[10] La estrategia *buy and hold* es una estrategia de inversión pasiva que no aplica gestión alguna. Consiste en comprar el activo de que se trate y mantenerlo durante todo el periodo de inversión, para venderlo al final de éste. La rentabilidad de esta estrategia está dada por la diferencia entre el valor de venta y el de compra del activo analizado.

[11] Cabe señalar que para cada uno de los periodos extramuestrales se calcularon la prueba DA, la desviación estándar de los índices bursátiles, el rendimiento, la frecuencia del modelo ARIMA y de la red *Ward*, y la rentabilidad de la estrategia *buy and hold*, los cuales están disponibles para los interesados. Para ello comunicarse con alguno de los autores.

[12] Se usó un nivel de significación de 10 por ciento.

CUADRO 3. *Resumen de medidas estadísticas. Promedio de los resultados de las 51 muestras*

(Porcentaje)

Índice	*Modelo*	*Índice promedio*	*Frecuencia, promedio*	*Número de periodos significativos*[a]	*Rendimiento promedio*	*Rendimiento promedio, B&H*
CAC40	Red *Ward*	22.94	58.7	11	−0.24	−23.53
	ARIMA		58.4	9	4.63	
FTSE100	Red *Ward*	16.43	51.4	5	−11.42	−15.68
	ARIMA		53.0	0	−2.42	
Hang Seng	Red *Ward*	28.97	67.2	51	32.24	−24.44
	ARIMA		55.7	10	9.44	
KLSE	Red *Ward*	20.98	62.0	20	11.57	4.80
	ARIMA		49.0	0	16.20	
MMX	Red *Ward*	25.79	53.0	1	11.51	4.83
	ARIMA		51.3	0	15.41	
Bovespa	Red *Ward*	31.54	52.1	5	−14.37	−18.54
	ARIMA		49.7	0	−2.24	
DJI	Red *Ward*	22.61	71.0	51	29.28	−7.74
	ARIMA		60.1	26	28.75	
GDAX	Red *Ward*	32.62	54.6	4	−10.85	−23.49
	ARIMA		59.5	12	7.44	
Nikkei	Red *Ward*	22.98	61.8	40	9.20	−22.55
	ARIMA		53.2	10	−1.89	
S&P500	Red *Ward*	21.04	55.8	9	6.80	−15.99
	ARIMA		57.1	7	6.32	
STI	Red *Ward*	24.08	64.6	41	−2.15	−12.38
	ARIMA		55.5	0	2.77	

[a] Número de periodos extramuestrales (de un total de 51) en los cuales la capacidad predictiva de los modelos resultó ser estadísticamente significativa (usando $\alpha = 10\%$), de acuerdo con la prueba DA.

nista de haber seguido las recomendaciones de compra-venta de la red *Ward*, del modelo ARIMA, así como la de una estrategia *buy and hold*.

Al analizar la frecuencia promedio (calculada sobre los 51 periodos extramuestrales) encontramos que, para la red *Ward*, ésta fue superior a 50% en todos los índices estudiados. En el caso del modelo ARIMA, la frecuencia promedio de predicción del signo fue superior a 50%, excepto en los índices KLSE y Bovespa. Luego, al comparar la capacidad predictiva de ambos modelos, tenemos que la red *Ward* superó al modelo ARIMA en 8 de los 11 índices analizados: CAC40, Hang Seng, KLSE, MMX, Bovespa, DJI, Nikkei y STI. Por otra parte, al considerar la estabilidad de la capacidad predictiva de los modelos,

CUADRO 4. *Frecuencia promedio y desviación estándar de la frecuencia*
(Porcentaje)

	ARIMA		Red Ward	
	Media		*Media*	
CAC40	58.4	2.8	58.7	3.0
FTSE100	53.0	2.9	51.4	5.1
Hang Seng	55.7	2.0	67.2	2.3
KLSE	49.0	3.0	62.0	3.1
MMX	51.3	2.6	53.0	3.8
Bovespa	49.7	3.7	52.1	3.1
DJI	60.1	4.3	71.0	2.0
GDAX	59.5	3.0	54.6	1.9
Nikkei	53.2	7.2	61.8	2.3
S&P500	57.1	3.2	55.8	6.2
STI	55.5	2.2	64.6	2.4

se tiene que la desviación estándar[13] de la frecuencia con que la red *Ward* predijo correctamente el signo de la variación del índice resultó mayor a la del modelo ARIMA en 7 de los 11 índices. En otras palabras, la capacidad predictiva de la red *Ward* evidencia mayor estabilidad que la del modelo ARIMA únicamente en los índices Bovespa, DJI, GDAX y Nikkei (véase cuadro 4).

Al aplicar la prueba DA en los 51 periodos extramuestrales encontramos que la capacidad predictiva de la red *Ward* resultó significativa en cada uno de ellos sólo en el caso de los índices Hang Seng y DJI. En el resto de los índices bursátiles tal conclusión no es clara, ya que tanto la capacidad predictiva de la red *Ward* como del modelo ARIMA pasa de significativa en un periodo a no serlo en el siguiente (véase cuadro 3). Lo anterior muestra que la capacidad predictiva de los modelos varía a lo largo del tiempo, por lo que es necesario no sólo recalcular los coeficientes periodo a periodo, sino también reelaborar el modelo en sí, por lo que no existiría un único modelo explicativo de la evolución de los índices bursátiles.

Se analizó la correlación entre capacidad predictiva (medida por la frecuencia de acierto) y volatilidad del índice (medida por su desviación estándar), para lo cual se utilizó el modelo presentado en la ecuación (13).

[13] Esta desviación estándar fue calculada sobre la frecuencia de los 51 periodos extramuestrales.

$$Frec_{Index\ i} = \alpha_0 + \alpha_1 \cdot \sigma_{Index\ i} + \varepsilon_i \qquad (13)$$

Como se observa en el cuadro 5 el coeficiente α_1, que mide la sensibilidad de la capacidad predictiva del modelo ARIMA a la volatilidad de índice bursátil, resultó significativo a 5% en los índices CAC40, Hang Seng, KLSE, MMX, Bovespa, Nikkei y STI. En estos casos, el signo del coeficiente fue negativo, por lo que un aumento en la variabilidad del índice disminuiría la capacidad predictiva del modelo ARIMA, y viceversa. La excepción estuvo dada por el índice STI, para el cual se encontró evidencia significativa de una correlación positiva entre variabilidad y capacidad de predicción.

Luego, al analizar la relación entre variabilidad del índice y capacidad predictiva de la red *Ward* (véase cuadro 5), se encontró que para los índices CAC40, Hang Seng, DJI, GDAX, Nikkei y S&P500 existe una correlación positiva y estadísticamente significativa entre dichas variables. Así, en los casos señalados, la capacidad predictiva de la red *Ward* aumentaría frente a panoramas de mayor volatilidad. Por otra parte, en el caso de los índices KLSE, MMX y STI el coeficiente α_1 resultó significativo y negativo, mientras que para los índices FTSE100 y Bovespa no se encontró evidencia de una relación significativa.

El cuadro 6 muestra la frecuencia con que la red *Ward* superó en rentabilidad al modelo ARIMA, junto con la frecuencia con que los modelos generaron un rendimiento mayor al de la estrategia *buy and hold*, considerando en ambos casos los 51 periodos extramuestrales. Se observó que la red *Ward* superó la rentabilidad del modelo ARIMA, con una frecuencia superior a 50%, sólo en el caso de los índices Hang Seng y Nikkei (100 y 86.3%, respectivamente). Esto, a pesar de que la red *Ward* superó en términos de capacidad predictiva al modelo ARIMA en 8 de los 11 índices analizados. Puede concluirse entonces que una alta capacidad predictiva, acompañada por una elevada desviación estándar, puede generar, como en este caso, menor rentabilidad promedio.

Por otra parte, considerando de nuevo los 51 periodos extramuestrales, la frecuencia con que los modelos de proyección superaron la rentabilidad de la estrategia *buy and hold* fluctuó entre 84.3 y 100%. Así, puede observarse que, independientemente de si la ca-

CUADRO 5. *Estadísticos modelos ARIMA y red Ward*[a]

	CAC40	*FTSE100*	*Hang Seng*	*KLSE*	*MMX*	*Bovespa*	*DJI*	*GDAX*	*Nikkei*	*S&P500*	*STI*
ARIMA											
R^2	0.4888	0.023	0.209	0.227	0.096	0.099	0.071	0.002	0.370	0.0360	0.3344
α_0	0.7785	0.602	0.774	0.678	0.667	0.685	0.485	0.585	2.126	0.649	0.4542
	(27.303)	(8.845)	(12.81)	(13.659)	(9.804)	(8.437)	(8.137)	(22.002)	(7.154)	(11.213)	(22.289)
α_1	−0.8465	−0.444	−0.750	−0.897	−0.601	−0.599	0.510	0.027	−6.939	−0.371	0.416
	(6.845)	(−1.072)	*(−3.601)*	*(−3.801)*	*(−2.282)*	*(−2.329)*	(1.944)	(0.345)	*(−5.367)*	(−1.352)	*(4.960)*
Prueba F	46.858	1.149	12.96	14.454	5.207	5.428	3.780	0.119	28.806	1.830	24.608
Red *Ward*											
R^2	0.439	0.001	0.261	0.469	0.110	0.000	0.126	0.2464	0.155	0.142	0.180
α_0	0.389	0.542	0.396	0.893	0.770	0.521	0.637	0.487	0.288	0.254	0.728
	(12.161)	(4.401)	(5.998)	(21.494)	(7.904)	(7.242)	(23.269)	(32.906)	(2.633)	(2.393)	(28.8609)
α_1	0.860	−0.171	0.950	−1.301	−0.931	−0.002	0.320	0.179	1.433	1.438	−0.341
	(6.196)	(−0.229)	*(4.168)*	*(−6.5873)*	*(−2.470)*	(−0.011)	*(2.662)*	*(4.003)*	*(3.003)*	*(2.849)*	*(−3.284)*
Prueba F	38.393	0.0525	17.374	43.392	6.102	0.000	7.089	16.023	9.022	8.117	10.786

[a] R^2 y valores de los coeficientes para los modelos ARIMA y red *Ward* para el periodo de estimación. En paréntesis se presenta el valor del test-*t*. Las cifras en cursivas son significativas en términos estadísticos a 5 por ciento.

CUADRO 6. *Frecuencia con que la red* Ward *superó en rentabilidad al modelo* ARIMA *y frecuencia con que los modelos superaron en rentabilidad a la estrategia* buy and hold, *considerando los 51 periodos extramuestrales*

(Porcentaje)

	Red vs ARIMA	*Modelo vs B&H*		*Red vs ARIMA*	*Modelo vs B&H*
CAC40	15.7	100.0	DJI	43.1	100.0
FTSE100	2.0	100.0	GDAX	0.0	100.0
Hang Seng	100.0	100.0	Nikkei	86.3	100.0
KLSE	15.7	88.2	S&P500	49.0	100.0
MMX	21.6	84.3	STI	23.5	98.0
Bovespa	13.7	100.0			

pacidad predictiva de los modelos es significativa o no, ambas técnicas permiten aumentar la rentabilidad o reducir las pérdidas asociadas a la inversión en índices bursátiles o carteras indizadas.

Conclusiones

De las diversas arquitecturas de redes neuronales con funcionamiento recursivo utilizadas en este estudio, la red *Ward* de tres capas fue la que obtuvo el mejor rendimiento en la predicción del signo de las variaciones de los índices bursátiles CAC40, FTSE100, Hang Seng, KLSE, MMX, Bovespa, DJI, GDAX, Nikkei225, S&P500 y STI.

En términos de capacidad predictiva, la red *Ward* superó al modelo ARIMA en 8 de los 11 índices analizados: CAC40, Hang Seng, KLSE, MMX, Bovespa, DJI, Nikkei y STI. Cabe señalar al respecto que Sharda y Patil (1992) encontraron que, para modelos de series de tiempo de largo plazo, los modelos de redes neuronales y los modelos de Box-Jenkins producen resultados similares, mientras que Tang *et al* (1991) encontraron que, para modelos de series de tiempo de largo plazo, las redes neuronales arrojan mejores resultados que las técnicas de Box-Jenkins. Por otra parte, se encontró que la capacidad predictiva de la red *Ward* evidencia mayor estabilidad que la del modelo ARIMA únicamente en los índices Bovespa, DJI, GDAX y Nikkei.

Nuestros resultados muestran que la capacidad predictiva de los modelos varía a lo largo del tiempo, por lo que es menester no sólo recalcular los coeficientes periodo a periodo, sino reconstruir el modelo en sí, por lo que no existiría un único modelo explicativo de la evolución de los índices bursátiles.

Luego, al analizar la relación entre variabilidad del índice y capacidad predictiva de la red *Ward*, se encontró que para los índices CAC40, Hang Seng, DJI, GDAX, Nikkei y S&P500 existe una correlación positiva y estadísticamente significativa entre dichas variables. Así, en los casos señalados, la capacidad predictiva de la red *Ward* aumentaría frente a panoramas de mayor volatilidad. No obstante, los resultados apuntan a que una alta capacidad predictiva, acompañada por una gran desviación estándar, puede generar una rentabilidad promedio relativamente menor.

También se observó que, al comparar los resultados de la red

Ward y del modelo ARIMA con los de la estrategia *buy and hold*, y de manera independiente de si la capacidad predictiva de los modelos resulta significativa en términos estadísticos, ambas técnicas permiten aumentar la rentabilidad o reducir las pérdidas asociadas a la inversión en índices bursátiles o carteras indizadas.

Finalmente, el estudio mostró que la capacidad predictiva de los modelos de redes neuronales resulta pertinente, lo que los sitúa como una opción al análisis técnico al momento de tomar decisiones de inversión en activos bursátiles. Así, el empleo de redes neuronales en la proyección semanal del signo de la variación de un índice bursátil y, por tanto, de la evolución del mercado accionario podría facilitar la conformación de carteras de inversión. Por medio de la generación de pronósticos semanales, los modelos de redes neuronales entregarían información que orientaría a los administradores de carteras de inversión respecto al riesgo por asumir y la estrategia de inversión por seguir, dadas las expectativas de hechos futuros y la relación riesgo-rendimiento que se espera obtener.

Apéndice

1. *Algoritmo de aprendizaje de repropagación del error*

A continuación se presenta el proceso mediante el cual el perceptrón multicapa estándar obtiene la salida (predicción) y actualiza los pesos de conexión entre las capas. De acuerdo con Freeman y Skapura (1993), la aplicación del algoritmo de aprendizaje de repropagación del error se realiza de la siguiente manera:

a) Se aplica el vector de datos $x_p = [x_{p1}, x_{p2}, \ldots, x_{pN}]^t$ a las unidades de entrada, el cual contiene las variables que se emplearán para predecir el signo de las variaciones del IPSA. x_p corresponde al vector de datos de entrada relativos al periodo p, x_{pi} representa el valor de la i-ésima variable explicativa en el periodo p de la serie de tiempo, y el superíndice t indica traspuesta.

b) Cada p-ésimo vector se emplea para calcular los valores de entrada neta para las unidades de la capa oculta. Así, cada neurona de la capa oculta calcula el valor de entrada neto como la sumatoria de los valores de entrada ponderados por sus respectivos pesos, como se indica en la ecuación (A1):

$$Neta^h_{pj} = \sum_{i=1}^{N} w^h_{ji} \cdot x_{ij} + 0^h_j \tag{A1}$$

en la que $Neta^h_{pi}$ es el valor de entrada neto para la j-ésima neurona de la capa oculta[14] relacionada al periodo p, w_{ji}^{h} es la ponderación que recibe la i-ésima variable explicativa asociada a la j-ésima neurona de la capa oculta, la cual puede ser excitatoria (pesos positivos) o inhibitoria (pesos negativos), x_{ij} corresponde al valor de la i-ésima variable de entrada asociada a la j-ésima neurona de la capa oculta, y θ_{j}^{h} es el sesgo (*bias*), el cual representa los grados de libertad y se comporta como otro ponderador. El superíndice h acompaña a los valores relacionados con la capa oculta.

De esta manera se tendrá una matriz de pesos, como se muestra en la expresión (A2):

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} & \Lambda & w_{1n} \\ w_{21} & w_{22} & w_{23} & \Lambda & w_{2n} \\ M & M & M & M & M \\ w_{m1} & w_{m2} & w_{m3} & \Lambda & w_{mn} \end{bmatrix}_{(i)\times(j)}^{h}, \qquad \text{con } j = 1, \ldots, n;\ i = 1, \ldots, m \qquad \text{(A2)}$$

c) Una vez que la sumatoria ha sido calculada, ésta debe ser procesada por una función de activación (lineal, tangente, *logistic* o gaussiana). En consecuencia, se aplica una función de activación en el valor de entrada neto de cada una de las neuronas que conforman la capa oculta. Así se obtiene la salida de la capa oculta, i_{pj}^{h}, como se muestra en la ecuación (A3):

$$i_{pj}^{h} = f_{j}^{h}(Neta_{pj}^{h}) \qquad \text{(A3)}$$

i_{pj}^{h} representa el valor de salida de la j-ésima neurona perteneciente a la capa oculta y f_{j}^{h} es la función de activación aplicada en el valor de entrada neto de la j-ésima neurona de la mencionada capa.

d) Luego, i_{pj}^{h} pasa a ser la entrada de la capa de salida. De nuevo se calculan los valores netos de las entradas para cada unidad o neurona de la capa de salida, como lo señala la ecuación (A4):

$$Neta_{pk}^{o} = \sum_{j=1}^{L} w_{kj}^{o} \cdot i_{pj}^{h} + \theta_{k}^{o} \qquad \text{(A4)}$$

en la que $Neta_{pk}^{o}$ es el valor de entrada neto (relacionado con el periodo p) para la k-ésima neurona de la capa de salida, w_{kj}^{o} es la ponderación que reci-

[14] Un punto importante en la elaboración de una red neuronal consiste en determinar el número apropiado de neuronas en la capa oculta, el cual está relacionado con la capacidad de mapeo de la red. Mientras más grande es su número, mayor será la capacidad del modelo para memorizar el conjunto de entrenamiento. Sin embargo, si se continúa aumentando el tamaño de la red, hay un punto en el que la generalización (la capacidad de proporcionar una respuesta correcta ante datos no utilizados en su entrenamiento) empeorará, debido a que puede ocurrir un sobreajuste al conjunto de entrenamiento. Infortunadamente, no hay una respuesta congruente a esta cuestión, y la manera más eficiente para determinar el tamaño óptimo de la red es por medio de la experimentación. No obstante, Salchenberger, Cinar y Lash (1992) recomiendan que el número de neuronas de la capa oculta debe corresponder a aproximadamente 75% de las unidades de entrada.

be el j-ésimo valor de salida de la capa oculta asociado a la k-ésima neurona de la capa de salida. θ^o_k es el sesgo, el cual representa los grados de libertad y se comporta como otro ponderador, y el superíndice o acompaña a los valores relacionados con la capa de salida.

De esta manera se tendrá una matriz de pesos, como se muestra en la expresión (A5):

$$\begin{bmatrix} w_{11} & w_{12} & w_{13} & \Lambda & w_{1n} \\ w_{21} & w_{22} & w_{23} & \Lambda & w_{2n} \\ M & M & M & M & M \\ w_{l1} & w_{l2} & w_{l3} & \Lambda & w_{ln} \end{bmatrix}^o_{(k)\times(j)}, \quad \text{con } j=1,\ldots,n;\ k=1,\ldots,l \tag{A5}$$

e) Posteriormente se obtienen las salidas de la capa oculta, o_{pk} en la ecuación (A6), aplicándose la función de activación en el valor neto anterior, $Neta^o_{pk} \cdot f^o_k$ es la función de activación aplicada en el valor de entrada neto de la k-ésima neurona de la capa de salida:

$$o_{pk} = f^o_j(Neta^o_{pk}) \tag{A6}$$

La función de activación en esta fase no forzosamente será la misma que la aplicada en la capa oculta.

f) Se calculan los términos de error para las unidades de salida, δ^o_{pk}, como se muestra en la ecuación (A7). y_{pk} representa los valores observados de la variable de salida en el momento p, o_{pk} significa el valor de la predicción para el periodo p generado por la k-ésima neurona de la capa de salida, y $f^{o\prime}_k(Neta^o_{pk})$ corresponde a la primera derivada de la función de activación aplicada en la capa de salida:

$$\delta^o_{pk} = (y_{pk} - o_{pk}) f^{o\prime}_k(Neta^o_{pk}) \tag{A7}$$

g) Se calculan los términos de error para las neuronas de la capa oculta: el error total se distribuye entre las neuronas de la capa oculta en proporción a la aportación que cada una de ellas realizó al error total, como se señala en la ecuación (A8):

$$\delta^h_{pj} = f^{h\prime}_j(Neta^h_{pj}) \sum_k \delta^o_{pk} \cdot w^o_{kj} \tag{A8}$$

El término δ^h_{pj} representa la fracción del error total de la predicción realizada para el periodo p, asociado a la j-ésima neurona de la capa oculta.

h) Una vez que se ha distribuido el error entre las neuronas de la capa oculta, se actualizan los pesos que recibe el j-ésimo valor de salida de la capa oculta asociado a la k-ésima neurona de la capa de salida, $w^o_{kj}(t)$. Esta actualización se hace en función de la tasa de aprendizaje, η, del error estimado para cada neurona de la capa de salida, δ^{o}_{pk}, y del valor de salida de la j-ésima neurona perteneciente a la capa oculta, i^{h}_{pj}. El peso actualizado, $w^{o}_{kj}(t+1)$, se expresa por medio de la ecuación (A9):

$$w^{o}_{kj}(t+1) = w^{o}_{kj}(t) + \eta \cdot \delta^{o}_{pk} \cdot i^{h}_{pj} \tag{A9}$$

i) Se actualizan los pesos asociados a la j-ésima neurona de la capa oculta, w^{h}_{ji}, en función de la tasa de aprendizaje, η, del error obtenido para cada neurona de la capa oculta, δ^{h}_{pj}, y del valor de la i-ésima variable de entrada, x_i, como se indica en la ecuación (A10):

$$w^{h}_{ji}(t+1) = w^{h}_{ji}(t) + \eta \cdot \delta^{h}_{pj} \cdot x_i \tag{A10}$$

j) Finalmente, se calcula el término de error por medio de la expresión (A11):

$$E_p = \frac{1}{2} \cdot \sum_{k=1}^{M} (\delta^{o}_{pk})^2 \tag{A11}$$

k) El proceso descrito a partir del paso *a*) se repite para un nuevo vector de datos de entrada, $x_p = [x_{p1}, x_{p2}, \ldots, x_{pN}]^{T}$, y finaliza con el cálculo del error de predicción explicado en el paso *j*). Cuando el proceso se haya realizado para cada uno de los vectores[15] que componen el conjunto de datos de entrenamiento, se habrá concluido una época. A su vez, pueden completarse tantas épocas como se considere necesario durante la experimentación.

2. *Formas funcionales*

A continuación se describe la forma funcional de los modelos de redes neuronales usados para predecir el signo de la variación de los índices bursátiles. Las letras *c*, *h*, *k*, *m*, *s*, *d*, *d*65, *g*, *b*, *n*, *f*, *nq*, *sp* y *sp*100 representan las variaciones porcentuales de los valores de cierre semanales de los índices CAC40, Hang Seng, KLSE, MMX, STI, Dow Jones Industry (DJI), Dow Jones 65, GDAX, Bovespa, Nikkei225, FTSE100, Nasdaq, S&P500 y S&P100, respectivamente. Por su parte, las siglas *B*13, *B*5, *B*10 y *B*30, corresponden a las tasas de interés de los bonos del gobierno estadunidense a 13 semanas, a 5 años, a 10 años y a 30 años. Las variables se utilizaron con uno (−1) o dos (−2) rezagos:

$$c_t = z(c_{-1}, d_{-1}, nq_{-1}, sp_{-1}, g_{-1}, f_{-1}) \tag{A12}$$

$$\begin{aligned} f_t = z(&f_{-1}, f_{-2}, d_{-1}, d_{-2}, d65_{-1}, d65_{-2}, nq_{-1}, nq_{-2}, B13_{-1}, B13_{-2} B5_{-1}, B5_{-2}, \\ &B10_{-1}, B10_{-2}, B30_{-1}, B30_{-2}, sp_{-1}, sp_{-2}, t_{-1}, t_{-2}, c_{-1}, c_{-2}, g_{-1}, g_{-2}) \end{aligned} \tag{A13}$$

$$h_t = z(h_{-1}, n_{-1}, d_{-1}, sp_{-1}, f_{-1}, k_{-1}, s_{-1}) \tag{A14}$$

[15] El número de vectores será igual al número de datos que contengan las series de tiempo que forman parte del conjunto de entrenamiento.

$$k_t = z(k_{-1}, h_{-1}, n_{-1}, d_{-1}, sp_{-1}, f_{-1}, s_{-1}) \tag{A15}$$

$$m_t = z(m_{-1}, d_{-1}, nq_{-1}, sp_{-1}, f_{-1}) \tag{A16}$$

$$b_t = z(b_{-1}, n_{-1}, d_{-1}, sp_{-1}, f_{-1}, h_{-1}) \tag{A17}$$

$$\begin{aligned} d_t = z(&d_{-1}, d_{-2}, d65_{-1}, d65_{-2}, nq_{-1}, nq_{-2}, B13_{-1}, B13_{-2}, B5_{-1}, B5_{-2}, B10_{-1}, \\ &B10_{-2}, B30_{-1}, B30_{-2}, sp_{-1}, sp_{-2}, sp100_{-1}, sp100_{-2}, t_{-1}, t_{-2}, c_{-1}, c_{-2}, \\ &g_{-1}, g_{-2}, f_{-1}, f_{-2}) \end{aligned} \tag{A18}$$

$$g_t = z(g_{-1}, n_{-1}, d_{-1}, sp_{-1}, f_{-1}, h_{-1}, c_{-1}) \tag{A19}$$

$$n_t = z(n_{-1}, d_{-1}, sp_{-1}, f_{-1}, h_{-1}) \tag{A20}$$

$$\begin{aligned} sp_t = z(&d_{-1}, d_{-2}, nq_{-1}, nq_{-2}, B13_{-1}, B13_{-2}, B5_{-1}, B5_{-2}, B10_{-1}, B10_{-2}, \\ &B30_{-1}, B30_{-2}, sp_{-1}, sp_{-2}, sp100_{-1}, sp100_{-2}, t_{-1}, t_{-2}, c_{-1}, c_{-2}, g_{-1}, \\ &g_{-2}, f_{-1}, f_{-2}) \end{aligned} \tag{A21}$$

$$s_t = z(h_{-1}, n_{-1}, d_{-1}, sp_{-1}, f_{-1}, k_{-1}, s_{-1}) \tag{A22}$$

Referencias bibliográficas

Blume, L., D. Easley y M. O'Hara (1994), "Market Statistics and Technical Analysis: The Role of Volume", *Journal of Finance*, 49, pp. 153-182.

Bosarge, W. E. (1993), "Adaptive Processes to Exploit the Nonlinear Structure of Financial Markets", R. R. Trippi y E. Turban (comps.), *Neural Networks in Finance and Investing*, Nueva York, Irwin, pp. 371-402.

Chen, A., y M. Leung (1998), "Dynamic Foreign Currency Trading Guided by Adaptive Forecasting", *Review of Pacific Basin Financial Markets and Policies* 1, pp. 383-418.

Chopra, N., J. Lakonishok y J. R. Ritter (1992), "Measuring Abnormal Returns: Do Stocks Overreact?", *Journal of Financial Economics* 31, pp. 235-268.

Conrad, J., y G. Kaul (1988), "Time-Variation in Expected Returns", *Journal of Business* 61, pp. 409-425.

—— y —— (1989), "Mean Reversion in Short-Horizon Expected Returns", *Review of Financial Studies* 2, pp. 225-240.

Dasgupta, C. G., G. S. Dispensa y S. Ghose (1994), "Comparing the Predictive Performance of a Neural Network Model with Some Traditional Market Response Models", *International Journal of Forecasting* 10, pp. 235-244.

DeBondt, W. F. M., y R. Thaler (1985), "Does the Stock Market Overreact?", *Journal of Finance* 40, pp. 793-805.

Estrella, A., y F. S. Mishkin (1998), "Predicting US Recessions: Financial Variables as Leading Indicators", *The Review of Economics and Statistics* 80(1), pp. 45-61.

Fama, E., y K. R. French (1988), "Permanent and Temporary Components of Stock Prices", *Journal of Political Economy* 98, pp. 247-273.

Ferson, W., y C. Harvey (1991), "The Variation of Economic Risk Premiums", *Journal of Political Economy* 99, pp. 385-415.

Hawley, D., J. Johnson y D. Raina (1990), "Artificial Neural Systems: A New Tool for Financial Decision-Making", *Financial Analysts Journal* 23, páginas 63-72.

Herbrich, R., M. Keilbach, T. Graepel, P. Bollmann-Sdorra y K. Obermayer (2000), "Neural Networks in Economics: Background, Applications and New Developments", T. Brenner (comp.), *Advances in Computational Economics: Computational Techniques for Modelling Learning in Economics* 11, pp. 169-196, Kluwer Academics.

Hill, T., L. Marquez, M. O'Connor y W. Remus (1994), "Artificial Neural Network Models for Forecasting and Decision Making", *International Journal of Forecasting* 10, pp. 5-15.

Hodgson, A., y D. Nicholls (1991), "The Impact of Index Futures Markets on Australian Share Market Volatility", *Journal of Business Finance and Accounting* 18, pp. 267-280.

Hornik, K., M. Stinchcombe y H. White (1989), "Multilayer Feedforward Networks are Universal Approximators", *Neural Networks* 5, pp. 359-366.

Jegadeesh, N. (1990), "Evidence of Predictable Behavior of Security Returns", *The Journal of Finance* 45, pp. 881-898.

Kanas, A. (2001), "Neural Networks Linear Forecasts for Stocks Returns", *International Journal of Finance and Economics* 6, pp. 245-254.

Kuo, C., y A. Reitsch (1995-1996), "Neural Networks vs. Conventional Methods of Forecasting", *Journal of Business Forecasting* 14, pp. 17-22.

Leung, Mark T., Hazem Daouk y An-Sing Chen (2000), "Forecasting Stock Indices: A Comparison of Classification and Level Estimation Models", *International Journal of Forecasting* 16, pp. 173-190.

Lo, A., y A. C. MacKinley (1988), "Stock Market Price do not Follow Random Walk: Evidence From a Simple Specification Test", *Review of Financial Studies* 1, pp. 41-66.

Maberly, E. D. (1986), "The Informational Content of the Interday Price Change with Respect to Stock Index Futures", *Journal of Futures Markets* 6, pp. 385-395.

Malkiel, B. (1981), *A Random Walk Down Wall Street*, 2a. ed., Nueva York, Norton.

Martín del Brío, B., y A. Sanz (1997), *Redes neuronales y sistemas borrosos: Introducción, teórica y práctica*, primera edición, Ra-ma.

O'Connor, M., W. Remus y K. Griggs (1997), "Going Up-Going Down: How Good Are People at Forecasting Trends and Changes in Trends?", *Journal of Forecasting* 16, pp. 165-176.

Pesaran, M. H., y A. Timmermann (1992), "A Simple Nonparametric Test of Predictive Performance", *Journal of Business and Economic Statistics* 10, pp. 461-465.

Poterba, J. M., y L. H. Summers (1988), "Mean Reversion in Stock Prices: Evidence and Implications", *Journal of Financial Economics* 22, pp. 27-59.

Qi, M. (2001), "Predicting US Recessions with Leading Indicators Via Neural Network Models", *International Journal of Forecasting* 17, pp. 383-401.

Refenes, A. P. (1995), *Neural Networks in the Capital Markets*, Nueva York, Wiley.

Sharda, R., y R. Patil (1992), "Connectionist Approach to Time Series Prediction: An Empirical Test", *Journal of Intelligent Manufacturing*, vol. 3, pp. 317-323.

Tang, Z., C. de Almieda y P. Fishwick (1991), "Time-Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology", *Simulation* 57, pp. 303-310.

Tsibouris, G., y M. Zeidenberg (1995), "Testing the Efficient Markets Hypothesis with Gradient Descent Algorithms", A. P. Refenes (comp.), *Neural Networks in the Capital Markets*, Chichester, Wiley, pp. 127-136.

White, H. (1993), "Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns", R. R. Trippi y E. Turban (comps.), *Neural Networks in Finance and Investing*, Nueva York, Irwin, pp. 315-328.

Wilson, J. H., y B. Keating (1998), *Business Forcasting*, tercera edición, U. Irwin/McGraw-Hill.

Yoda, M. (1994), "Predicting the Tokyo Stock Market", G. J. Deboeck (comp.), *Trading on the Edge: Neural, Genetic, and the Fuzzy Systems for Chaotic Financial Markets*, Nueva York, Wiley, pp. 66-79.