



Estudios de Economía Aplicada

ISSN: 1133-3197

secretaria.tecnica@revista-eea.net

Asociación Internacional de Economía

Aplicada

España

OLMEDO, F.; VELASCO, F.; VALDERAS, J. M.

Caracterización no Lineal y Predicción no Paramétrica en el IBEX35

Estudios de Economía Aplicada, vol. 25, núm. 3, diciembre, 2007, pp. 815-842

Asociación Internacional de Economía Aplicada

Valladolid, España

Disponible en: <http://www.redalyc.org/articulo.oa?id=30113818012>

- ▶ Cómo citar el artículo
- ▶ Número completo
- ▶ Más información del artículo
- ▶ Página de la revista en [redalyc.org](http://redalyc.org)

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto

# **sección abierta**

**Estudios de  
Economía Aplicada**

## Caracterización no Lineal y Predicción no Paramétrica en el IBEX35

**OLMEDO, E., VELASCO, F. Y VALDERAS, J.M.**

*Dep. Economía Aplicada I. Universidad de Sevilla. Avda. Ramón y Cajal, nº 1, 41018 Sevilla SPAIN*

E-mail: [{olmedo.velasco.valderas}@us.es](mailto:{olmedo.velasco.valderas}@us.es)

### RESUMEN

La modelización no lineal ha experimentado un resurgimiento de la mano de la teoría del caos, que ha puesto de manifiesto la posibilidad de conseguir comportamientos complejos producidos por la dinámica endógena del modelo, sin la necesidad de incluir perturbaciones aleatorias exógenas al mismo. Por otro lado, es indiscutible la importancia de la capacidad de predicción en economía. En el presente trabajo se aplican diferentes técnicas desarrolladas dentro de lo que se denomina Econometría Compleja No Lineal buscando mejorar la capacidad de predicción en los mercados financieros.

*Palabras clave:* No linealidad, predicción, contraste BDS, contraste de Kaplan, modelos ARIMA, vecinos próximos, redes neuronales.

### Nonlinear Characterization And Prediction Of IBEX 35

### ABSTRACT

The nonlinear modelization has experimented a great resurgence of the hand of Chaos Theory, which shown the possibility of obtaining complex behaviors produced endogenously by the dynamics of the model, without the necessity to include exogenous random shocks. On the other hand, the importance of the forecasting ability in Economics is principal. In the work, different techniques developed within Complex Econometrics are applied to improve forecasting in stock markets.

*Keywords:* Nonlinearity, forecasting, BDS test, Kaplan test, ARIMA, near neighbours, neural networks.

Clasificación JEL: C14, C22, C45, C52, C53.

---

Artículo recibido en diciembre de 2005 y aceptado para su publicación en septiembre de 2006.

Artículo disponible en versión electrónica en la página [www.revista-eea.net](http://www.revista-eea.net), ref.: e-25304.

## 1. INTRODUCCIÓN

Uno de los principales problemas de la Ciencia Económica es la predicción del comportamiento futuro de las variables económicas, es decir, a partir de su comportamiento pasado recogido en series temporales, se trata de utilizar esta información y los modelos de teoría económica disponible para predecir su evolución futura. En el enfoque tradicional de series temporales usualmente se trabaja usando la modelización ARIMA o, de manera más generalizada, un enfoque lineal basado en los procesos estocásticos en los que la introducción de perturbaciones aleatorias se hace necesaria para emular el comportamiento complejo observado en la realidad económica. Sin embargo, muchas de las propiedades intrínsecas del comportamiento de las variables económicas parecen no adaptarse correctamente a los modelos lineales. Como ejemplos pueden citarse los distintos comportamientos observados en recesiones y expansiones, la presencia de dos o más regímenes, la posibilidad de comportamientos explosivos o la relevancia de ciertas variables sólo bajo determinadas condiciones. Todo ello conduce al uso de la modelización no lineal tanto en macroeconomía como en finanzas.

Aunque tradicionalmente no se ha utilizado el enfoque no lineal debido a su dificultad y aparentes limitaciones de comportamiento, ambas cuestiones están superadas hoy en día de la mano, por un lado, de la disponibilidad de herramientas informáticas para su tratamiento y, por otro lado, del impulso metodológico proporcionado por la Teoría del Caos. La Teoría del Caos establece que sistemas deterministas con la propiedad de dependencia sensible a las condiciones iniciales son capaces de generar dinámicas aparentemente tan complejas como las generadas por un proceso estocástico. Por tanto, se pone de manifiesto que no es necesaria la presencia de ruidos exógenos aleatorios para generar comportamientos complejos. Dentro de este nuevo enfoque se han desarrollado, por un lado, nuevos modelos no lineales y, por otro lado, nuevas herramientas para trabajar con series temporales. En concreto, se han desarrollado contrastes para determinar si la serie procede o no de un modelo generador de datos no lineal, técnicas para la cuantificación de la complejidad de la serie, ajustes de diversos modelos paramétricos no lineales así como nuevas técnicas no paramétricas de predicción.

La presencia de no linealidad en los índices bursátiles (Sarantis, 2001) ha sido justificada desde un punto de vista tanto teórico como empírico. De hecho, la teoría económica apunta diversas fuentes para justificar la presencia de no linealidad en los índices bursátiles, como son la diversidad en las creencias de los agentes, la heterogeneidad en los objetivos de los inversores que surgen a partir de la variabilidad de los horizontes de inversión y aversión al riesgo, el comportamiento por bloques y los cambios ocasionados por variaciones en el crecimiento económico. Empíricamente

existe evidencia a favor de la existencia de una estructura no lineal en los mercados financieros (Scheinkman y Le Baron, 1989; Hinich y Patterson, 1985; Hsieh, 1989 y 1991; Bajo-Rubio et al., 1992; Brock et al., 1996; Guarda y Salmon, 1996; Brooks, 1996; Cecen y Erkal, 1996; Abhyankar et al., 1997 y Pérez-Rodríguez et al., 2005; Bonilla et al., 2006). Además, existen trabajos en los que se proporciona una base teórica para estos resultados empíricos, principalmente mediante la descripción de la dinámica de precios en los mercados financieros como el resultado de la interacción entre grupos de agentes con información heterogénea (Cutler et al., 1990; De Long et al., 1990; Day and Huang, 1990; Chiarella, 1992; Lux, 1997 y 1998; Lux y Marchesi, 1999a y 1999b; LeBaron, 2000; Chen, Lux y Marchesi, 2001). Por otro lado, también se ha justificado la existencia de caos (dependencia sensible) en el mundo financiero, tanto desde un punto de vista teórico (Baptista y Caldas, 2002; Farmer y Joshi, 2002 o Chen y Song, 2003) como empírico (Nakajima, 2002; Bask, 2002; Wang et al., 2004 o Shintani y Linton, 2004).

El paso siguiente, si existen indicios fundamentados para pensar que el modelo generador de datos es no lineal, es el aprovechamiento de esta no linealidad para mejorar la capacidad predictiva. Sin embargo, es frecuente (Clements et al., 2004) que la bondad de predicción fuera de la muestra de estos modelos no lineales no compense la mayor dificultad de su ajuste frente al del enfoque lineal tradicional. El problema reside en que, aunque los modelos no lineales capturan mejor la naturaleza de la dinámica de la serie temporal, esto no tiene que traducirse en que estos modelos funcionen mejor para la predicción, ya que muchas veces los sistemas económicos son demasiado complejos como para ajustar una simple generalización de un modelo lineal. Es necesario, por tanto, la búsqueda de alternativas mejores para la predicción utilizando modelos no lineales.

Por estos motivos, y centrándonos en la predicción de series temporales financieras, en los últimos años se ha pasado del ajuste de modelos no lineales paramétricos relativamente sencillos (como ejemplos pueden citarse trabajos que utilizan modelos de varianza condicional (Hsieh, 1997; Baillie y Bollerslev, 1991; Baillie et al., 2000; Balaban y Bayar, 2005), el modelo de Markov (Turner et al., 1989 y Kanas, 2003), el modelo no lineal por piezas (Oh y Kim, 2002) o el modelo umbral autorregresivo (Sarantis, 2001 y Bradley y Janse, 2004), a generalizaciones más complejas de estos modelos. Entre ellas pueden citarse diversas generalizaciones del modelo umbral, como la exponencial (McMillan, 2003 y Brooks, 2001) o la generalización a modelos GARCH (Marcucci, 2005), generalizaciones de modelos GARCH, como los modelos GARCH no lineales (Gokcan, 2000 y Wei, 2002) y los fraccionales (Beine et al., 2002 y Jin y Frechet, 2004).

Dentro de esta misma tendencia, en los últimos años se han considerado distintos modelos no paramétricos, dentro de los cuales destacan por su potencia los modelos de redes neuronales. Con la utilización de estas técnicas no paramétricas se entra dentro de lo que se denomina “econometría caótica o no lineal”. Estas técnicas presentan además la ventaja, como señalan Satchell y Timmermann (1996), de funcionar tanto para series caóticas como para series estocásticas.

En este trabajo comparamos dos enfoques distintos aplicándolos a la predicción de los rendimientos del Ibex 35 (del 1 de abril de 1998 hasta el 10 de agosto de 2002), y verificamos si existe o no diferente comportamiento dependiendo del horizonte de predicción que interesa. Por un lado, si el sistema es determinista de dimensión baja, el espacio reconstruido nos permite hacer predicciones posiblemente mejores que las conseguidas con las metodologías tradicionales, ya que recobramos la topología y la dinámica del sistema original desconocido. En este caso se habla de enfoque de comprensión por reconstrucción. Por otro lado, si el sistema es complejo (de dimensión elevada, o con un nivel alto de ruido), parece incuestionable la potencia mostrada para la predicción de las redes neuronales (Zhang et al., 1998).

Así, en la siguiente sección se comprueba la presencia de no linealidad en la serie de rendimientos del Ibex 35. Posteriormente, en la sección tercera se introducen las técnicas de predicción por comprensión y las referidas a la predicción por aprendizaje utilizando redes neuronales. En la sección cuarta se realiza una aproximación a la serie de rendimientos mediante un modelo ARIMA, comprobando su incapacidad para predecir de manera útil el futuro de la serie. Por último, en la sección quinta se muestran los resultados obtenidos por los métodos no lineales aplicados a la serie de rendimientos del Ibex 35, trabajando con diferentes horizontes de predicción. Los resultados obtenidos son claros a favor de estas nuevas técnicas no paramétricas, confirmándose los resultados del trabajo de Pérez-Rodríguez et al., 2005, obtenidos únicamente para predicciones de un paso hacia delante. Nosotros proponemos la utilización de redes neuronales para horizontes cortos de predicción (de cinco o menos pasos hacia delante) y la de modelos locales por media simple o ponderada de vecinos para horizontes más largos.

## 2. CONTRASTES DE NO LINEALIDAD: CONTRASTE DE KAPLAN

Dado que se pretende medir la bondad de técnicas de predicción no paramétricas, es necesario determinar si existe no linealidad sin especificar el tipo concreto de no linealidad. Por este motivo, se utiliza el contraste de Kaplan (1994 y 1995), que a pesar de su simplicidad ha demostrado empíricamente (Barnett et al., 1997) detectar un amplio espectro de clases de no linealidad. Este contraste ha comenzado ya a ser utilizado en series financieras (Chen et al., 2001 y Belaire-Franch et al., 2002).

El contraste de Kaplan se basa en la idea de continuidad y originalmente fue formulado para la detección de determinismo en la dinámica de una serie temporal. Kaplan trata de encontrar la presencia de una estructura determinista en la órbita descrita con la reconstrucción por el método de los retardos, basándose en la continuidad de dicha órbita. Para ello utiliza una propiedad que diferencia a un sistema determinista de otro que no lo es: si el modelo subyacente es determinista, dos puntos inicialmente próximos se mantendrán próximos tras una iteración mientras que, si el modelo subyacente es estocástico dos puntos inicialmente próximos pueden tener imágenes muy alejadas entre sí. Kaplan trabaja con un estadístico  $K$  que calcula para la serie original ( $K_{test}$ ) y para un modelo generador lineal estocástico con las mismas propiedades (con el mismo histograma y función de autocorrelación) que el sistema original ( $KS$ ). Si el valor del estadístico de la serie original excede el de las series bajo la hipótesis alternativa de linealidad estocástica, se acepta la hipótesis nula de linealidad.

La distribución del estadístico no está tabulada, pero Kaplan propone dos cotas máximas para el contraste de la hipótesis nula de linealidad. La primera es el mínimo  $KS$  estimado a partir de las series subrogadas, y la segunda es la media menos dos o tres veces la desviación típica de todos los  $KS$  estimados. Se rechaza la hipótesis nula de linealidad cuando el valor de  $K$  calculado para la serie original reconstruida (que llamamos  $K_{test}$ ) es mayor que, al menos, una de las dos cotas (que llamamos, para simplificar  $KS$  y  $KS_{min}$ ).

### 3. PREDICCIÓN EN EL MARCO DE LA ECONOMETRÍA COMPLEJA

#### 3.1. Técnicas de Predicción en el Enfoque de Comprensión mediante la Reconstrucción del Espacio De Estados

Casdagli (1989) trata el problema de la predicción no lineal como un “problema inverso” en la teoría de sistemas dinámicos. El problema usual en este campo es, conociendo el modelo, describir su evolución conforme pasa el tiempo; el problema inverso es el contrario, es decir, dada la evolución del sistema (que es lo que se manifiesta en la serie de datos recogida a lo largo del tiempo) construir un modelo que pueda haberla originado (modelo generador de datos). Si se localiza la función (necesariamente no lineal si estamos ante un sistema caótico) que puede dar origen a esta evolución, dispondremos de un posible modelo predictivo.

Para conocer esta función se separan los datos en dos grupos; el primero se utiliza para llevar a cabo una interpolación o aproximación de la función desconocida (conjunto de entrenamiento) y el segundo (conjunto de verificación) para evaluar la bondad del ajuste mediante los errores de predicción.

Para comprobar si nuestra predicción es buena, calculamos el error del predictor comenzando con una dimensión de absorción baja y repitiendo el cálculo hasta que alcanzamos un valor del error suficientemente bajo. Por último, debemos tener en cuenta que existen diversas técnicas de interpolación, por lo que elegiremos aquella que proporcione el menor valor para el error, sin aumentar demasiado los cálculos necesarios.

Cuando la dinámica observada es compleja, se suelen utilizar, por el buen ajuste conseguido, los métodos locales, que se basan en ajustar una función predictora para cada intervalo local. Estos modelos se pueden considerar como linealizaciones de modelos globales considerando puntos del entorno. Estos modelos globales funcionan mejor si las funciones a ajustar son suaves, cuestión que es difícil en la práctica. Para el ajuste de estos modelos locales es necesario localizar puntos vecinos al cual se quiere predecir y ajustar la función elegida minimizando el error cuadrático medio y estudiando su evolución (Farmer y Sidorowich, 1987). En finanzas estos métodos han sido relativamente poco utilizados, aunque como ejemplos pueden citarse los trabajos de Diebold y Nason (1990), Sosvilla-Rivero y Bajo-Rubio (1992 y 1998), Lisi y Medio (1997), Cao y Soofi (1999), Soofi y Cao (1999) y Agnon et al. (1999).

El método más sencillo es el método del vecino más próximo, propuesto por el meteorólogo E.N. Lorenz (1963). Lorenz propuso en este trabajo el concepto de análogo, aunque no lo consideró para la predicción. Según él, cualquier sistema acotado de dimensión finita debe pasar por algún estado que pase arbitrariamente próximo a otro estado previo. Si el sistema es estable, el desarrollo futuro de ambos estados permanecerá próximo, mientras que si el sistema es inestable las sucesiones de estados sucesivos a partir de los dos considerados diferirán a partir de un determinado momento. Suponemos que la ley de nuestro sistema dinámico responde a la dinámica de una cierta función continua  $F$  que relaciona el estado presente y el futuro de dicho sistema, es decir, tal que  $x_n = F(x_{n-1})$ . Para predecir el valor  $x_{N+1}$  dado el valor  $x_N$  simplemente se localiza aquel valor más próximo a  $x_N$  (en la terminología de Lorenz, el análogo, cuya existencia está garantizada por la propia naturaleza del atractor), que denotamos por  $x_{n_0}$ . Entonces, debido a la continuidad de la función  $F$  tenemos garantizado que si en el tiempo  $n_0$  el estado del sistema es similar al del tiempo  $N$ ,  $x_{n_0+1}$  está asimismo próximo a  $x_{N+1}$ . De esta manera,  $\hat{x}_{N+1} = x_{n_0+1}$ .

El problema es que los datos reales no se corresponden a los diferentes estados del sistema, sino a sus mediciones utilizando una función desconocida de medida  $h$ . Esta es la razón por la cual es fundamental para estos métodos el Teorema de

Reconstrucción del Espacio de Estados de Takens (1985), que garantiza la misma topología y las mismas propiedades dinámicas en el sistema dinámico original y en la reconstrucción  $m$ -dimensional del mismo, obtenida utilizando los datos de la serie temporal generada. Esta reconstrucción se obtiene mediante la denominada función de retardos, mediante la cual los datos unidimensionales de la serie temporal se transforman en datos  $m$ -dimensionales que, en cada instante del tiempo se obtienen mediante los retardos sucesivos. Es decir, la  $m$ -historia en el instante  $t$  es:

$$a_m(t) = (x(t), x(t-\tau), \dots, x(t-(m-1)\tau))$$

siendo  $\tau$  el tiempo de retardo y  $m$  la dimensión del sistema reconstruido, denominada dimensión de absorción. La evolución dinámica de estas  $m$ -historias está marcada por una función que combina la dinámica del sistema original, la función de medida y la función de retardos de Takens, de manera que  $a_m(t+T) = \psi_T(a_m(t))$ . Como señala Wan (1994), esto implica que existe una autorregresión no lineal del tipo

$$x_{(n+m)\tau} = f(x_{(n+m-1)\tau}, \dots, x_{n\tau})$$

y lo que se pretende en la predicción práctica es encontrar una función  $\tilde{f}_N : \mathbb{R}^m \rightarrow \mathbb{R}$  que aproxime la ecuación anterior para  $1 \leq n \leq N_m^\tau = N - (m-1)\tau$ .

Aplicando el método de los vecinos, para predecir  $x(t+T)$  primero se impone una métrica en el espacio de estados, que denotamos por  $\|\cdot\|$  y se localizan las  $k$   $m$ -historias más próximas a la  $m$ -historia  $a_m(t)$ , es decir, las  $k$   $a_m(t')$  con  $t' < t$  cuya distancia  $\|a_m(t) - a_m(t')\|$  es más pequeña. A partir de aquí se construye un predictor local, considerando cada una de las  $m$ -historias vecinas como un punto en el conjunto origen y el correspondiente punto  $x(t'+T)$  como un punto en el conjunto imagen. Es decir, se busca una función que ajuste los  $k$  pares de puntos  $(a_m(t'), x(t'+T))$ . El enfoque más sencillo es considerar únicamente el vecino más próximo. En este caso, simplemente tenemos  $\hat{x}(t+T) = x(t'+T)$ .

Como señalan Kantz y Schreiber (1994), la función de medida de una cantidad continua es sólo válida hasta un cierto nivel de resolución. Hay que tener en cuenta el problema de la discretización, así como los posibles errores de la medición. Esto nos lleva a que simplemente localizar el estado más próximo puede conducir a errores, ya que las distancias pueden estar contaminadas por los errores de la función de medida.

Por este motivo en la práctica no se suele trabajar con un único punto análogo, sino los análogos en un determinado radio. Así, se localizan los  $k$  puntos más cercanos a  $a_m(t)$ , que se denominan vecinos próximos y que son los  $k$  puntos  $a_m(t')$  con  $t' < t$  con la distancia  $\|a_m(t) - a_m(t')\|^1$  más pequeña. Después se construye un predictor local haciéndole corresponder a cada punto  $a_m(t')$  el correspondiente  $a_m(t'+T)$ . El predictor una función lineal que se ajusta a estos pares de puntos. En este trabajo utilizamos la media aritmética, la media ponderada por las distancias y la regresión lineal.

### 3.2. Redes neuronales: un Enfoque Global

Cuando el sistema no es suficientemente simple es más apropiado el enfoque del aprendizaje por medio de redes neuronales ya que, debido a su generalidad, es capaz de ajustar el comportamiento de cualquier función. Las redes neuronales (Nychka et al., 1992, Jungeilges, 1996 y White, 1989) son una clase de modelos no lineales inspirados en la arquitectura neuronal del cerebro. Fueron desarrollados en un primer momento en el campo de la Inteligencia Artificial para intentar modelizar la capacidad de aprendizaje de los sistemas biológicos neuronales mediante la modelización de la estructura del cerebro. Las redes neuronales son modelos tipo input-output que asignan

al vector input  $x$  un vector output  $y = g(x)$ , donde la función  $g$  ha sido determinada por un proceso de aprendizaje recursivo. En definitivo, al igual que las técnicas de comprensión, las redes neuronales intentan resolver el problema inverso de ajustar una función no lineal utilizando los valores sucesivos de una serie temporal.

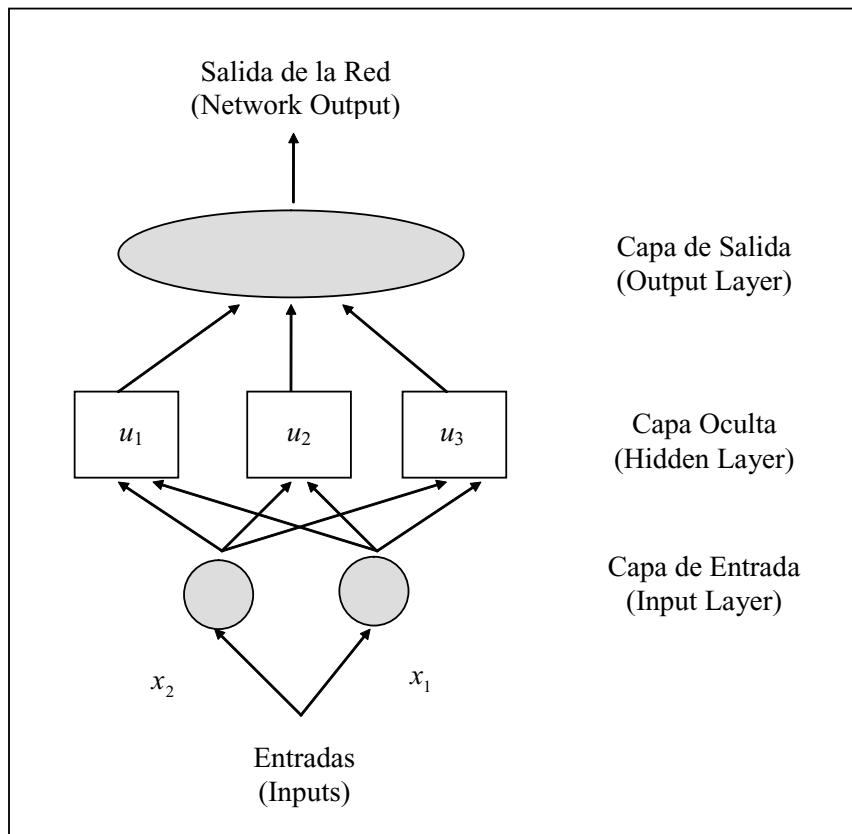
Su principal ventaja es su capacidad de ajuste, demostrada empíricamente, y el crecimiento lineal de los parámetros con el orden de aproximación. Se trata de una estructura de interconexiones, compuesta por una serie de capas, cada una de ellas formada por neuronas, de manera que cada una de ellas alimenta a todas las neuronas de la capa siguiente. Cada una de las neuronas proporciona una salida u output que es el resultado que proporciona una función, llamada función de activación, a la suma de sus inputs multiplicados por unas ponderaciones o pesos. Estas ponderaciones se ajustan conforme a un algoritmo de aprendizaje especificado con el que se pretende minimizar la función de coste calculada a partir del error cometido si se compara la salida de la red con el valor real.

---

1 Generalmente se trabaja con la norma del máximo.

Existe una amplia aplicación de las redes neuronales en el campo de la predicción de series temporales económicas. En concreto, se pueden citar ejemplos en mercados financieros (Trippi y Turban, 1993, Azoff, 1994, Refenes, 1995 y Gately, 1996), predicción en mercados bursátiles (White, 1988, Kimono et al., 1990, Schoneburg, 1990, Yoon y Swales, 1991, Grudnitsky y Osburn, 1993, Chen et al., 2003 y Jasic y Wood, 2004), predicción de bancarrota y fallos de mercados (Coleman et al., 1991, Salchenberger et al., 1992, Tam y Kian, 1992, Fletcher y Goss, 1993, Wilson y Sharda, 1994 y O'Leary, 1998), predicción de tipos de cambio (Weigend et al., 1992, Refenes, 1993, Kuan y Liu, 1995, Hann y Steurer, 1996, Zhang y Hu, 1998, Frances y Griensven, 1998 y Giles et al., 2001). Además también se están desarrollando aplicaciones macroeconómicas, como la predicción de recesiones (Qi, 2001) o del comportamiento de magnitudes macroeconómicas (GDP en Tkacz, 2001), e incluso existe alguna aplicación en el mercado de la vivienda (Caridad y Ceular, 2001).

La red neuronal que se va a utilizar en este trabajo es del tipo Multi Layer FeedForward Network (MLFFN), un tipo de red muy popular, ampliamente utilizado con buenos resultados en el campo de la predicción por su habilidad para ajustar cualquier tipo de relación input-output. En la estructura de la red es fundamental la elección del número de capas ocultas, del número de nodos o neuronas de dichas capas y del número de unidades de entrada. En cuanto al número de capas ocultas, en general es suficiente trabajar con una (Zhang et al., 1998) y, en cuanto a la determinación del número de nodos de la capa oculta se utilizan argumentos heurísticos que lo relacionan con el número de inputs, nunca superando el doble de ellos. De todos modos, como muestran algunos estudios (Lachtermacher y Fuller, 1995 y Tang y Fishwith, 1993), el número de nodos de la capa oculta tiene un efecto sobre la predicción, aunque no verdaderamente significativo, siendo ésta mucho más sensible a la elección del número de nodos de entrada, ya que dicho número contiene información importante acerca de la complejidad de la estructura de la red. Este número suele ser escogido en función de los resultados de diseños experimentales, escogiéndose normalmente el mejor diseño en función del Criterio de Información de Akaike o de Schwartz. En el presente trabajo se utiliza una red con una capa oculta y se elige el número de unidades de la capa oculta buscando minimizar el error cometido. En la Figura 1 se muestra el esquema de la red para una capa oculta y dos valores de entrada,  $x_1$  y  $x_2$ .

**Figura 1. Esquema de una red SLFF con una capa oculta**

Generalizando para una red neuronal con  $n$  inputs  $x_t = (x_{1,t}, \dots, x_{n,t})$  y  $q$  unidades en la capa oculta, para un determinado instante  $t$ , la salida de la red es:

$$o_t = \Phi \left( \beta_0 + \sum_{i=1}^q \beta_i \Psi \left( \gamma_{i0} + \sum_{j=1}^n \gamma_{ij} x_{j,t} \right) \right) \equiv f(x_t, \theta)$$

donde  $\theta = (\beta_0, \dots, \beta_q, \gamma_1, \dots, \gamma_q)$  y  $\gamma_j = (\gamma_{j0}, \dots, \gamma_{jn})$  son los parámetros a estimar,  $\Psi$  es una función de activación en forma de S;  $o_t$  se puede considerar como un estimador para la variable target o deseada  $y_t$ , y  $x_t$  es el vector de inputs. Las unidades ocultas de las redes feedforward no son dinámicas pues no dependen de los valores pasados generados por la red.

Para predecir con una red neuronal, se utiliza el conjunto de entrenamiento para ajustar los pesos o ponderaciones utilizando los errores cometidos entre la salida de la red y el valor real de la serie. Estos pesos se utilizan para predecir en el conjunto de verificación retroalimentando la red con los valores predichos, en lugar de con los reales.

#### 4. CARACTERIZACIÓN DE LA SERIE RENDIMIENTOS DEL IBEX MEDIANTE UN MODELO LINEAL

En primer lugar realizamos un primer análisis de la series de rendimiento del IBEX mediante un modelo ARIMA siguiendo la metodología Box-Jenkins. Consideramos que la serie original de rendimientos del IBEX entre el uno de abril de 1990 hasta el 10 de agosto de 2002 es estacionaria, ya que al tomar primeras diferencias salen claros signos de una raíz unitaria que nos hacen considerar que no debemos diferenciar. El correlograma de la serie original parece indicar que la serie es un ruido blanco, pero el estadístico de Ljung-Box nos indica que el primer retardo es significativamente distinto de cero. Nos inclinamos por tanto, por un modelo MA(1) tras una validación del mismo mediante el análisis de sus residuos. El modelo estimado para los 3100 primeros datos de la serie es igual

$$x_t = 0.080897x_{t-1} + \varepsilon_t \\ [t = 4.517]$$

El hecho de que el modelo sea un MA(1) hace que sólo tengamos una predicción distinta a la media un periodo adelante, pues los rendimientos están incorrelados linealmente para retardos superiores a 2. Para predicciones a horizonte mayor que 1 debemos tomar como predictor óptimo la esperanza incondicional de la serie que en nuestro caso es nula. Por tanto para predicciones con horizonte mayor a 1 las predicciones de nuestro MA(1) son una constante por lo que la varianza de los residuos coincide con la varianza total del modelo con lo que la capacidad predictiva del modelo lineal es nula. Utilizando los primeros 3100 datos como conjunto de entrenamiento y estimando el modelo de forma iterativa entre el dato 3100 y hasta el dato final del periodo muestral considerado para obtener las predicciones un periodo por delante obtenemos que la capacidad predictiva del modelo no es muy buena, como muestra el hecho de que la suma de los cuadrados de los residuos es superior a la suma de los cuadrados de los valores de la variable.

## 5. RESULTADOS OBTENIDOS MEDIANTE MÉTODOS NO LINEALES

En la Tabla 1 se muestran los resultados del contraste de Kaplan, que proporciona evidencia, sin ningún género de duda, a favor de la presencia de no linealidad.

**Tabla 1. Resultados del contraste de Kaplan**

Dimensión de absorción (m)	KSmean	KSstd	KS	Ksmin	Ktest
1	0.0147	0.0001	0.0144	0.0146	<b>0.0142</b>
2	0.0146	0.0002	0.0140	0.0142	<b>0.0117</b>
3	0.0147	0.0003	0.0138	0.0142	<b>0.0096</b>
4	0.0147	0.0006	0.0129	0.0136	<b>0.0080</b>
5	0.0147	0.0007	0.0126	0.0131	<b>0.0072</b>
6	0.0144	0.0008	0.0120	0.0128	<b>0.0066</b>
7	0.0145	0.0009	0.0118	0.0126	<b>0.0065</b>
8	0.0144	0.0011	0.0111	0.0128	<b>0.0058</b>
9	0.0146	0.0019	0.0089	0.0104	<b>0.0062</b>
10	0.0147	0.0021	0.0084	0.0108	<b>0.0061</b>

En este contraste, es necesario calcular el estadístico  $K$  ( $Ktest$ ) para la serie original. La hipótesis nula de determinismo implica que este valor  $K$  es próximo a cero. Como el estadístico no está tabulado, y además se quiere trabajar con la hipótesis nula de no linealidad, se calcula este estadístico  $K$  para una serie de series que verifican que el modelo generador de datos es estocástico y no lineal, y que son susceptibles de haber generado la serie temporal (por tanto, deben tener el mismo histograma y la misma función de autocorrelación). Se considera que se verifica la hipótesis nula de no linealidad si el valor del estadístico  $K$  calculado para la serie original ( $Ktest$ ) es menor que los valores obtenidos para los modelos lineales de los cuales, como medida resumen, se toma el mínimo ( $Kmin$ ) y la media menos tres desviaciones típicas ( $KS$ ). Para un amplio rango de valores de la dimensión de absorción, en todos los casos  $Ktest$  es menor que ambos valores  $Kmin$  y  $KS$ .

A continuación se muestran los resultados para la predicción hacia dos, cinco, diez y veinte períodos hacia delante. Primero (de la Tabla 2 a la Tabla 13) se consideran los métodos de aproximación local, en concreto, la media de vecinos, la media ponderada de vecinos y la regresión lineal de vecinos, para un único vecino, para dos, cinco, diez, veinticinco, cincuenta y cien vecinos. Lo que se proporciona en las tablas es el cociente entre la suma de los cuadrados de los residuos y la suma de cuadrados total de la serie que se está prediciendo. Existirán, por tanto, evidencias a favor del método

utilizado cuando este cociente sea menor que la unidad (valores que se resaltan en negrita). Por último, en la Tabla 14 se muestran los resultados (también como cociente entre suma de cuadrados de residuos y suma total de cuadrados) de la aplicación de una red neuronal SLFFN, siendo  $\Psi$  la función logística y  $\Phi$  la función identidad. El algoritmo de aprendizaje utilizado para el ajuste de los pesos es el de retropropagación. Este algoritmo fue ideado a principios de los 70 por Werbos (1974), y completado con las contribuciones de Parker (1985) y Le Cun (1985), siendo popularizado a partir del trabajo de Rumelhart y McClelland (1986). Este método básicamente consiste en actualizar las ponderaciones del modelo en la dirección en la que el gradiente desciende más rápidamente, es decir,  $\hat{\theta}_{t+1} = \hat{\theta}_t + \eta \nabla f(x_t, \hat{\theta}_t) [y_t - f(x_t, \hat{\theta}_t)]$ , donde  $\nabla f(x, \theta)$  es el vector gradiente de la función  $f$  con respecto a  $\theta$  y  $\eta$  es la tasa de aprendizaje (en nuestro caso, 0.01).

En los dos casos se realizan predicciones para un amplio rango de valores de la dimensión de absorción, desde dos hasta diez. En el caso de las aproximaciones locales, simplificamos el paso previo de la reconstrucción aplicando el teorema de Takens, considerando que la dimensión óptima de reconstrucción es la que proporciona una mejor predicción. En el caso de las redes neuronales, por simplificar notación, hablamos de dimensión de absorción cuando en realidad nos estamos refiriendo al número de unidades de la capa oculta.

En la Tabla 2 se ha aplicado un ajuste local por media aritmética de vecinos próximos. Las filas de la tabla indican las dimensiones de absorción consideradas, que van desde 2 hasta 10 (más de 10 indica la existencia de un sistema excesivamente complejo para intentar su ajuste mediante la reconstrucción del espacio de estados). Si los mejores ajustes se consiguen para dimensiones bajas se indicaría que el espacio de estados que recobra la dinámica original es de baja dimensión. Las columnas de la tabla indican el número de vecinos considerados para el ajuste. Se ha considerado el caso de un único vecino (en definitiva, el método original propuesto por Lorenz), y posteriormente 2, 5, 10, 25, 50 y 100 vecinos. Si el ajuste mejora conforme disminuye el número de vecinos se tendrían indicios a favor de una dinámica compleja a pequeña escala y, por tanto, a favor de los ajustes locales.

**Tabla 2. Predicción por media de vecinos. Horizonte de predicción 2**

Predicción por media de vecinos. Horizonte de predicción 2							
Número de vecinos utilizado							
m	1	2	5	10	25	50	100
2	4861,377000	2402,5965	2316,0465	2155,6583	1872,5234	2120,9173	2135,7082
3	73,887114	29,699716	19,338249	23,142967	25,122942	24,160113	20,955284
4	<b>0,360812</b>	<b>0,68624448</b>	<b>0,60815506</b>	<b>0,91060011</b>	<b>0,86539983</b>	<b>0,98525434</b>	1,0679671
5	<b>0,039171</b>	<b>0,53125025</b>	1,1894442	1,1800209	1,6561842	1,4844268	1,3647658
6	4,733492	4,0459409	3,1647425	3,2393435	2,5387511	2,237478	2,071987
7	1,762123	<b>0,25571725</b>	<b>0,67733305</b>	1,2057553	1,7549747	1,6682619	1,7760011
8	21,626915	3,2967731	4,3887144	4,790599	4,9667914	5,4865793	5,5214688
9	<b>0,844019</b>	1,0272019	<b>0,9662819</b>	<b>0,8220721</b>	<b>0,97672201</b>	1,1890866	1,2238415
10	10,049772	14,507426	13,68425	13,164745	13,091444	12,628214	12,184612

Tanto en esta tabla como en las siguientes se proporciona, para cada dimensión y número de vecinos considerados, el cociente entre la suma de cuadrados de los residuos y la suma total de cuadrados con respecto a la media de los valores reales que presenta la variable que se pretende predecir para los horizontes considerados (en este caso, horizonte 2). Por tanto, siempre que este cociente sea menor que la unidad la predicción obtenida será mejor que trabajar con la media aritmética de la serie. Estos resultados se marcan en negrita. El mejor método será aquel que proporcione un menor cociente.

**Tabla 3. Predicción por media de vecinos. Horizonte de predicción 5**

Predicción por media de vecinos. Horizonte de predicción 5							
Número de vecinos utilizado							
m	1	2	5	10	25	50	100
2	2,954863	1,7500455	2,239362	1,9885716	2,1368807	2,1219317	2,1666577
3	3,702098	2,656832	1,7080742	1,9379616	1,9221709	1,8290327	1,7528378
4	<b>0,390663</b>	1,1724267	<b>0,9347626</b>	1,030132	<b>0,99354249</b>	1,0001732	1,0019379
5	<b>0,738669</b>	<b>0,79915058</b>	1,0415091	<b>0,96581005</b>	1,1896559	1,0666142	<b>0,99823471</b>
6	1,577892	1,2287852	1,179586	1,235331	1,2024409	1,1131596	1,1115472
7	2,062638	<b>0,68697177</b>	<b>0,79996806</b>	<b>0,8992903</b>	<b>0,94945426</b>	<b>0,97074921</b>	<b>0,96049229</b>
8	1,874481	<b>0,80824998</b>	<b>0,76935725</b>	1,1136349	1,0187101	1,0047522	1,0171414
9	2,058703	2,1430933	1,906635	1,7731901	1,7582426	1,9239846	1,9220692
10	1,719074	1,2303796	1,1891305	1,2612042	1,2649776	1,2130856	1,1929691

Los mejores resultados se obtienen (ver negrita), en general, para una dimensión de absorción igual a 4. Además, el ajuste conseguido empeora conforme aumenta el número de vecinos, de manera que el modelo local es más apropiado que el global. El

mejor resultado puntual se obtiene para un único vecino y un espacio de dimensión 5, seguido de dos vecinos y un espacio de dimensión siete.

Siguiendo la misma manera de trabajar pero aumentando la longitud de la ventana de predicción se observa, en primer lugar, un empeoramiento generalizado del ajuste conseguido. Los dos mejores resultados se obtienen para dimensiones y número de vecinos similares a los obtenidos en el caso anterior. Sin embargo, en general, la dimensión óptima de reconstrucción ha aumentado a siete.

Aumentando el horizonte de predicción se ve acentuada la tendencia que acabamos de mencionar (Tablas 4 y 5). Para longitudes de ventana de diez y veinte los valores óptimos de predicción corresponden a cinco vecinos y una dimensión de reconstrucción igual a ocho y la tendencia general es que se necesite una mayor dimensión de absorción para conseguir predicciones aceptables aunque, en cualquier caso, éstas empeoran conforme aumenta el horizonte considerado.

**Tabla 4. Predicción por media de vecinos. Horizonte de predicción 10**

Predicción por media de vecinos. Horizonte de predicción 10							
Número de vecinos utilizado							
m	1	2	5	10	25	50	100
2	1,408486	1,569514	1,294840	1,281017	1,237891	1,228355	1,246996
3	1,834795	1,524594	1,418587	1,454925	1,364760	1,341661	1,277461
4	<b>0,859893</b>	1,424268	1,374619	1,208967	1,280575	1,277214	1,272569
5	<b>0,933871</b>	1,063910	1,111611	1,019414	1,068798	1,055405	1,030246
6	2,211685	1,097602	1,113870	1,105276	1,097771	1,053650	1,059768
7	1,418126	1,022583	<b>0,857795</b>	1,006641	1,026332	1,004416	<b>0,984651</b>
8	1,343327	<b>0,919501</b>	<b>0,732788</b>	<b>0,997553</b>	<b>0,980407</b>	<b>0,992204</b>	<b>0,990217</b>
9	<b>0,886320</b>	1,073303	1,016384	1,018558	<b>0,980083</b>	1,025287	1,034966
10	1,292191	1,244271	1,189551	1,127389	1,156633	1,113187	1,105013

**Tabla 5. Predicción por media de vecinos. Horizonte de predicción 20**

Predicción por media de vecinos. Horizonte de predicción 20							
Número de vecinos utilizado							
m	1	2	5	10	25	50	100
2	1,170636	1,330165	1,164399	1,180346	1,044638	1,084771	1,086594
3	1,434289	1,175310	1,075470	1,167170	1,092520	1,072269	1,058587
4	<b>0,888330</b>	1,081899	1,133278	1,112050	1,044740	1,046682	1,041974
5	<b>0,929501</b>	1,080466	1,051832	<b>0,974741</b>	1,010978	1,007684	<b>0,999981</b>
6	1,549406	1,074064	1,068981	1,031930	1,020051	1,005360	1,014252
7	1,377798	1,024201	1,006968	<b>0,983841</b>	1,025283	<b>0,998378</b>	<b>0,990515</b>
8	1,178065	<b>0,861774</b>	<b>0,862260</b>	1,025849	1,007804	1,000316	<b>0,994479</b>
9	1,035410	1,037228	1,009417	1,003125	<b>0,985247</b>	1,024270	1,026921
10	1,450772	1,159263	1,057219	1,031795	1,044518	1,010879	1,015271

Exactamente los mismos comentarios son válidos si se trabaja utilizando la predicción por vecinos ponderados en lugar de la media simple de vecinos (Tablas 6 a 9): las predicciones empeoran conforme aumenta el horizonte de predicción considerado, a la vez que se necesita una mayor dimensión en el espacio de reconstrucción. A partir del horizonte igual a diez el óptimo de predicción se obtiene para cinco vecinos y un espacio reconstruido en dimensión ocho.

**Tabla 6. Predicción por media ponderada de vecinos. Horizonte de predicción 2**

Predicción por vecinos ponderados. Horizonte de predicción 2						
Número de vecinos utilizados						
m	2	5	10	25	50	100
2	1915,4445	605,78301	3811,372	3092,2013	1012,8762	2206,6737
3	25,235654	7,747302	10,819454	82,250565	282,95161	16,054877
4	<b>0,08986219</b>	<b>0,32926327</b>	<b>0,53181869</b>	<b>0,68036682</b>	<b>0,60662505</b>	<b>0,79481554</b>
5	<b>0,71754488</b>	1,1799548	1,1908592	1,3303489	1,3777037	1,3395448
6	4,1410658	3,3018611	3,3815246	2,6702955	2,281146	2,0388049
7	<b>0,60398198</b>	<b>0,8568158</b>	1,2892736	1,729269	1,5760744	1,7614641
8	3,7948657	4,2718824	4,2740666	4,7586903	5,3389994	5,4855863
9	1,0340557	<b>0,9686591</b>	<b>0,7446262</b>	<b>0,98637258</b>	1,1807322	1,2220073
10	15,316658	14,226045	13,368227	13,187236	12,686639	12,229447

**Tabla 7. Predicción por media ponderada de vecinos. Horizonte de predicción 5**

Predicción por vecinos ponderados. Horizonte de predicción 5						
m	Número de vecinos utilizados					
	2	5	10	25	50	100
2	2,850224	1,8037373	2,6943432	2,339598	1,8692542	2,1264197
3	1,9707289	1,3976372	1,5890842	4,021185	11,073979	1,5750085
4	<b>0,89819172</b>	<b>0,786105</b>	1,2067906	1,0597447	<b>0,9509386</b>	<b>0,98601459</b>
5	<b>0,83290935</b>	<b>0,98156913</b>	<b>0,95015468</b>	1,0089008	1,0075282	<b>0,97806259</b>
6	1,2204887	1,2027591	1,2117109	1,2490485	1,1303935	1,1001009
7	1,087494	<b>0,77750654</b>	<b>0,97991496</b>	<b>0,98215712</b>	<b>0,97114911</b>	<b>0,96157798</b>
8	<b>0,83719313</b>	<b>0,75257802</b>	1,0551966	<b>0,95923708</b>	<b>0,99969359</b>	1,0040755
9	2,1505372	1,9045303	1,6785401	1,7518934	1,8951669	1,9111334
10	1,2891152	1,2550623	1,2898206	1,2827325	1,2218646	1,2015084

Existe, sin embargo, una diferencia interesante: si comparamos la Tabla 6 con la Tabla 2 se observa la general mejora que se obtiene, para un horizonte igual a dos, si se trabaja con vecinos ponderados en lugar de con la simple media aritmética, que no se observa para horizontes de predicción superiores.

**Tabla 8. Predicción por media ponderada de vecinos. Horizonte de predicción 10**

Predicción por vecinos ponderados. Horizonte de predicción 10						
m	Número de vecinos utilizados					
	2	5	10	25	50	100
2	<b>0,989756</b>	1,170193	1,408359	1,362501	1,146378	1,228860
3	1,179274	1,024328	1,284926	1,834805	4,304137	1,396068
4	1,491860	1,130150	1,466617	1,299509	1,228799	1,256047
5	<b>0,996844</b>	1,112781	<b>0,971029</b>	1,030105	1,044384	1,013502
6	1,086404	1,112699	1,104997	1,129725	1,067981	1,049760
7	1,265689	1,041210	<b>0,976731</b>	1,055136	1,002522	<b>0,992906</b>
8	<b>0,868254</b>	<b>0,756976</b>	<b>0,887287</b>	<b>0,979404</b>	<b>0,961793</b>	<b>0,988567</b>
9	1,079246	1,057758	1,013580	<b>0,966009</b>	1,013617	1,033322
10	1,266969	1,198104	1,127224	1,142795	1,113968	1,111128

**Tabla 9. Predicción por media ponderada de vecinos. Horizonte de predicción 20**

m	Predicción por vecinos ponderados. Horizonte de predicción 20					
	Número de vecinos utilizados					
2	5	10	25	50	100	
2	<b>0,949662</b>	<b>0,971817</b>	1,159299	1,115260	1,697983	<b>0,980781</b>
3	<b>0,951401</b>	1,113162	1,170163	1,702558	2,187151	1,276763
4	1,272200	1,023056	1,113479	1,055670	1,030810	1,043649
5	<b>0,964355</b>	1,061379	1,005634	1,006310	1,016600	<b>0,990200</b>
6	1,055056	1,122492	1,075852	1,063854	<b>0,997889</b>	1,009809
7	1,019162	<b>0,982598</b>	<b>0,946635</b>	1,033238	<b>0,997314</b>	<b>0,996759</b>
8	<b>0,919703</b>	<b>0,832727</b>	<b>0,920587</b>	1,007141	<b>0,980534</b>	<b>0,995166</b>
9	1,015642	1,036927	1,014100	<b>0,980725</b>	1,018123	1,024685
10	<b>0,999936</b>	1,055074	1,026273	1,032668	1,010498	1,019561

Si aplicamos la metodología de la regresión lineal de los vecinos próximos (Tablas 10 a 13) observamos la necesidad de incrementar el número de vecinos utilizados, es decir, que se necesitan modelos menos locales. Por el contrario, es interesante comprobar que, utilizando un número suficientemente elevado de vecinos, los resultados obtenidos son substancialmente mejores que los obtenidos para los casos equivalentes en los métodos anteriores. Comparemos, por ejemplo, las Tablas 4 y 8 con la Tabla 12 (las tres para un horizonte de predicción igual a diez). En el caso de la media simple de vecinos el mejor resultado se obtiene para una dimensión de 8 y 5 vecinos (0.7327) seguido de una dimensión de 7 el mismo número de vecinos (0.8577). En el caso de la media ponderada de vecinos el mejor resultado se obtiene de nuevo para una dimensión de 8 y 5 vecinos (0.7569) y posteriormente para la misma dimensión y 2 vecinos (0.8682). Por último, en el caso de la regresión simple el mejor resultado se obtiene para 25 vecinos y una dimensión de 7 (0.6437) seguido por el obtenido para la misma dimensión y 50 vecinos (0.7019).

**Tabla 10. Predicción por regresión lineal de vecinos. Horizonte de predicción 2****\*sólo se producen resultados fiables a partir de 25 vecinos**

Predicción por regresión lineal. Horizonte de predicción 2*						
m	Número de vecinos utilizados					
	2	5	10	25	50	100
2	73152,826	2772,3149	1559,5169	1697,6352	1684,1822	1571,2123
3	45,044293	21,69296	25,924258	22,809692	23,408161	20,920315
4	7,2120047	2,3992686	<b>0,56801166</b>	<b>0,39269678</b>	<b>0,65409474</b>	<b>0,73036549</b>
5	<b>0,25745308</b>	<b>0,94707248</b>	1,5479199	1,9070512	1,5180742	1,2696538
6	746,21504	11,267934	6,5617973	4,4857995	3,0761346	2,1035626
7	61,891831	1,617677	<b>0,05924043</b>	1,0605847	1,5235622	1,2771518
8	19,023101	17,091714	12,553665	1,7607262	3,399953	3,5518376
9	34,700027	6,4308301	4,4124617	<b>0,52013298</b>	1,066788	1,136757
10	192,46353	374,05923	35,101283	27,446229	17,499885	14,543563

**Tabla 11. Predicción por regresión lineal de vecinos. Horizonte de predicción 5****\*sólo se producen resultados fiables a partir de 25 vecinos**

Predicción por regresión lineal. Horizonte de predicción 5*						
m	Número de vecinos utilizados					
	2	5	10	25	50	100
2	21,05631	2,6577045	2,1685721	1,7744063	2,0398388	2,0439055
3	130,24798	1,9499628	1,6475446	1,6082504	1,761414	1,856597
4	13,517146	3,5684806	<b>0,91309072</b>	<b>0,95798084</b>	1,127042	1,060194
5	4,5247012	8,8952538	1,1405991	1,2803565	1,1341091	<b>0,9499</b>
6	147,0883	6,5609917	1,5777291	1,1416289	<b>0,97813037</b>	<b>0,89870765</b>
7	127,2801	1,9316136	<b>0,7484455</b>	<b>0,36976359</b>	<b>0,6209879</b>	<b>0,68981323</b>
8	62,06919	6,0526438	2,5704901	<b>0,3911036</b>	<b>0,62437246</b>	<b>0,69385889</b>
9	1690,0264	85,386863	625,01618	1,4634748	1,6752579	1,4371965
10	60,581569	79,570637	10,93199	2,8286899	1,4413257	1,290336

**Tabla 12. Predicción por regresión lineal de vecinos. Horizonte de predicción 10**

\*sólo se producen resultados fiables a partir de 25 vecinos

Predicción por regresión lineal. Horizonte de predicción 10*						
Número de vecinos utilizados						
m	2	5	10	25	50	100
2	39,927632	1,706885	1,346876	1,291023	1,258348	1,232592
3	46,331447	2,058138	1,688223	1,333874	1,281253	1,302965
4	36,148280	4,669215	1,023205	1,005044	1,147587	1,140171
5	260,079990	12,368107	<b>0,996240</b>	1,049028	1,044923	<b>0,940182</b>
6	Inf	20,669122	1,499241	1,045523	<b>0,946370</b>	<b>0,915213</b>
7	55,781188	8,819547	1,452950	<b>0,643745</b>	<b>0,701921</b>	<b>0,814672</b>
8	26,161622	10,360659	2,420771	1,002752	<b>0,977754</b>	<b>0,863180</b>
9	335,481910	20,748187	111,825850	1,285024	1,152218	<b>0,929264</b>
10	878,588710	40,489010	8,367323	2,167016	1,384698	1,063328

**Tabla 13. Predicción por regresión lineal de vecinos. Horizonte de predicción 20**

\*sólo se producen resultados fiables a partir de 25 vecinos

Predicción por regresión lineal. Horizonte de predicción 20*						
Número de vecinos utilizados						
m	2	5	10	25	50	100
2	19,025141	1,680618	1,274879	1,253552	1,183840	1,143251
3	70,087589	1,729643	1,233565	1,055606	1,037418	1,036616
4	19,021543	3,529759	1,168369	<b>0,964929</b>	<b>0,995292</b>	<b>0,936777</b>
5	386,619580	9,296794	<b>0,839540</b>	1,040375	1,055065	<b>0,934816</b>
6	Inf	26,657163	1,247283	1,165249	1,037409	<b>0,950577</b>
7	796,642390	173,157180	1,876039	<b>0,936870</b>	1,005234	1,035469
8	70,681665	10,329088	3,848928	1,098098	1,019909	<b>0,989133</b>
9	90377,546000	28,702737	74,751687	1,352896	1,185956	1,053392
10	726,740240	32,036714	5,616995	1,880550	1,365035	1,077304

Por último, la metodología de redes neuronales, cuyos resultados se muestran en la Tabla 14, muestra dos ventajas sobre la de vecinos próximos. En primer lugar, los resultados son más estables, obteniéndose resultados más que aceptables para un amplio rango de posibles dimensiones consideradas. En segundo lugar, los resultados son claramente mejores para horizontes cortos de predicción, especialmente para cinco pasos hacia delante, y además para este caso y dos pasos hacia delante los resultados mejores se obtienen para una dimensión muy baja, concretamente igual a

tres. Sin embargo, los resultados obtenidos empeoran rápidamente para horizontes mayores de predicción, perdiéndose toda capacidad de predecir para veinte pasos hacia delante.

**Tabla 14. Predicción por red SLFFN**

m	Predicción por red neuronal SLFFN			
	Horizonte de predicción			
	2	5	10	20
2	<b>0,3588552</b>	<b>0,44360971</b>	<b>0,91954897</b>	1,1002246
3	<b>0,05816525</b>	<b>0,26885222</b>	2,8929046	5,7085676
4	<b>0,2685982</b>	<b>0,88940063</b>	1,2243348	2,6895771
5	<b>0,77538077</b>	<b>0,78369351</b>	1,6795167	2,0061941
6	<b>0,59195401</b>	2,756293	2,4040838	2,1893963
7	<b>0,38563461</b>	<b>0,46712192</b>	<b>0,92551115</b>	1,2218446
8	1,353494	1,7328756	2,5329785	2,4507502
9	7,7573383	4,6954728	7,5405156	6,2600321
10	1,2902691	5,0498519	6,5476984	6,5807438

## 6. CONCLUSIONES

En cuanto a la presencia de no linealidad en el modelo generador de la serie temporal de rendimientos del Ibex 35 para el período considerado, el contraste de Kaplan muestra claras evidencias a su favor. En cuanto a la predicción, aplicando la metodología ARIMA se obtiene que la serie se corresponde con un modelo media móvil de orden 1, de manera que sólo se consiguen predicciones para un período hacia delante.

Las predicciones basadas en ajustes locales muestran un ligero empeoramiento en los resultados conforme se aumenta el horizonte de predicción, lo cual se corresponde con la posible presencia de sensibilidad a las condiciones iniciales. Los mejores resultados, dentro de estos ajustes locales, se obtienen, en general, para un número de vecinos que oscila entre 5 y 25, aunque conviene resaltar que el número de vecinos que conduce a las mejores predicciones es de 5 vecinos en el caso de un horizonte de predicción 20. Esto puede considerarse como un indicio a favor del ajuste global por redes neuronales para pequeños horizontes de predicción. En cuanto a los distintos comportamientos de cada uno de los métodos propuestos, para un número de vecinos suficientemente grande como para que funcione la regresión lineal, este método proporciona resultados ligeramente mejores.

En cuanto a los resultados proporcionados por la red neuronal, resalta en primer lugar que presenta menos variaciones que en el caso de los ajustes locales. Se obtiene

el mejor resultado de predicción para un horizonte de predicción de 5. En segundo lugar, en promedio se obtienen resultados mejores que en el enfoque local para horizontes de predicción cortos, perdiéndose esta propiedad para los largos, lo cual estaría en concordancia con lo mencionado en el párrafo anterior.

En definitiva, existen claras evidencias a favor de la utilización de estos nuevos métodos de predicción no paramétrica para la predicción del comportamiento del Ibex 35 (se pueden establecer relaciones comparativas con los resultados obtenidos por Sarantis (2001) para los índices del G7 utilizando modelos paramétricos no lineales, a favor de los resultados obtenidos en este trabajo). Dentro de los métodos propuestos, es mejor la utilización del ajuste global mediante una red neuronal SL-FFN para horizontes de predicción cortos, y el de los ajustes locales para mayores horizontes de predicción.

## 7. REFERENCIAS BIBLIOGRÁFICAS

- ABHYANKAR, A., COPELAND, L.S. y WONG, W. (1997): Uncovering nonlinear structure in real-time stock-market indexes: the S&P 500, the DAX, the Nikkei 225 and the FTSE-100. *Journal of Business and Economic Studies* 15, 1: 1-14
- AZNAR, A. Y TRÍVEZ, F.J: (1993): *Métodos de predicción en Economía (vol. II): Análisis de series temporales*. Barcelona: Ariel
- AGNON, Y., GOLAN, A. y SHEARER, M. (1999). Nonparametric, nonlinear, short-term forecasting: theory and evidence for nonlinearities in the commodity markets. *Economic Letters* 65: 293-299
- AZOFF, E.M. (1994). *Neural network time series forecasting of financial markets*. Chichester: John Wiley and Sons
- BAILLIE, T.R. y BOLLERSLEV, T. (1991): Intra-day and Intermarket Volatility in Foreign Exchange Rates, *Review of Economic Studies* 58, 3: 565-85
- BAILLIE, T.R., CECEN, A., ERKAL, C. y HAN, Y.W. (2000): High Frequency Deutsche Mark-US Dollar Returns: FIGARCH Representations and Nonlinearities, *Multinational Finance Journal* 4, 3/4: 247-267, 2000.
- BAJO-RUBIO, O., FERNÁNDEZ-RODRÍGUEZ, F. y SOSVILLA-RIVERO, S. (1992): Chaotic behavior in exchange-rate series: results for the peseta-US dollar case. *Economic Letters* 39: 207-211
- BALABAN, E. y BAYAR, A. (2005): Stock returns and volatility: empirical evidence from fourteen countries. *Applied Economic Letters* 12, 10: 603-611

- BAPTISTA, M.S. y CALDAS, I.L. (2002): Stock market dynamics. *Physica A* 312, 3/4: 539-564
- BARNETT, W.A., GALLANT, A.R., HINICH, M.J., JUNGEILGES, J.A. KAPLAN, D.T. y JENSEN, M.J. (1997): A singled blind controlled competition among test for nonlinearity and chaos. *Journal of Econometrics* 77: 157-192
- BASK, M. (2002): A positive Lyapunov exponent in Swedish exchange rates? *Chaos, Solitons and Fractals* 14: 1295-3004
- BEINE, M., BENASSY-QUERE, A. y LECOURT, C. (2002): Central bank intervention and foreign exchange rates: new evidence from FIGARCH estimations. *Journal of International Money and Finance* 21, 1: 115-144
- BELAIRE-FRANCH, J., CONTRERAS, D. y TORDERA-LLEDÓ, L. (2002): Assesing nonlinear structures in real exchange rates using recurrence plot strategies. *Physica D* 171: 249-264
- BONILLA, C.A., ROMERO-MEZA, R. y HINICH, M.J. (2006): Episodic nonlinearity in Latin American stock market indices. *Applied Economic Letters* 16: 193-199
- BRADLEY, M.D. y JANSEN, D.W. (2004): Forecasting with a nonlinear dynamic model of stock returns and industrial production. *International Journal of Forecasting* 20: 321-342
- BROCK, W.A., DECHERT, D., SCHEINKMAN, J. y LEBARON, B. (1996): A test for independence based on the correlation dimension. *Econometric Reviews* 15: 197-235
- BROOKS, C. (1996): Testing for nonlinearity in daily pound exchange rates. *Applied Financial Economics* 6: 307-317
- BROOKS, C. (2001): A double-threshold GARCH model for the French Franc/Deutschmark exchange rate. *Journal of Forecasting* 20, 2: 135-143
- CAO, L. y SOOFI, A.S. (1999). Nonlinear deterministic forecasting of daily dollar exchange rates. *International Journal of Forecasting* 15: 421-430
- CARIDAD, J.M. y CEULAR, N. (2001): Un análisis del mercado de la vivienda a través de redes neuronales artificiales. *Revista de Estudios de Economía Aplicada* 18: 67-81
- CASDAGLI, M. (1989). Nonlinear prediction of chaotic time series. *Physica D*, 35: 35-56
- CECEN, A. y ERKAL, C. (1996): Distinguishing between stochastic and deterministic behavior in high frequency foreign exchange rate returns: can non-linear dynamics help forecasting? *International Journal of Forecasting* 12: 465-473
- CHEN, A.-S, LEUNG, M.T. y DAOUK, H. (2003). Application of neural networks to an emerging financial market. Forecasting and trading the Taiwan Stock Index. *Computers and Operations Research* 30, 6: 901-923
- CHEN, S., LUX, T., y MARCHESI, M. (2001): Testing for non-linear structure in an artificial market. *Journal of Economic Behavior and Organization* 46: 327-342

- CHEN, Y. y SONG, X.F. (2003): Study on the chaos model of liquidity in the stock market. *Systems Research and Behavioral Science* 20, 5: 419-425
- CHIARELLA, C. (1992): The dynamics of speculative behaviour. *Annals of Operations Research* 37: 101-123
- CLEMENTS, M.P., FRANCES, P.H. y SWANSON, N.R. (2004). Economic and financial time-series with non-linear models. *International Journal of Forecasting* 20: 169-183
- COLEMAN, K.G., GRAETTINGER, T.J., y LAWRENCE, W.F. (1991). Neural network for bankruptcy prediction. the power to solve financial problems. *AI Review* 5, julio/agosto: 48-50
- DAY, R. Y HUANG, W. (1990): Bulls, bears and market sheep. *Journal of Economic Behavior and Organization* 14: 299-329
- DE LONG, J.B., SHELIFER, A., SUMMERS, L.H. y WALDMANN, R.J. (1990): Noise trader risk in financial markets. *Journal of Political Economy* 98, 4: 703-738
- DIEBOLD, F. y NASON, J.A. (1990). Nonparametric exchange rate predictions. *Journal of International Economics* 28, 3/4: 315-332
- FARMER, J.D. y JOSHI, S. (2002): The price dynamics of common trading strategies. *Journal of Economic Behavior and Organization* 49, 2: 149-171
- FARMER, J.D. y SIDOROWICH, J.J. (1987). Predicting chaotic time series. *Physical Review Letters* 59: 845-848
- FERNÁNDEZ-RODRÍGUEZ, F. y SOSVILLA-RIVERO, S. (1992). Chaotic behaviour in exchange-rate series: first results for the Peseta-U.S. Dollar case. *Economic Letters* 39: 207-211
- FERNÁNDEZ-RODRÍGUEZ, F. y SOSVILLA-RIVERO, S. (1998). Testing nonlinear forecastability in time series: theory and evidence from the EMS. *Economic Letters* 59: 49-63
- FLETCHER, D. y GOSS, E. (1993). Forecasting with neural networks. An application using bankruptcy data. *Information and Management* 24: 159-167
- FRANCES, P.H. y GRIENSVEN K. (1998). Forecasting exchange rates using neural networks for technical trading rules. *Studies in Nonlinear Dynamics and Econometrics* 2, 4: 109-114
- GATELY, E. (1996). *Neural networks for financial forecasting*. New York: John Wiley and Sons
- GILES, C.L., LAWRENCE, S. y TSOI, A.C. (2001). Noisy edition using a recurrent neural network and grammatical inference. *Machine Learning* 44, 1/2: 161-183
- GOKCAN, S. (2000). Forecasting volatility of emerging stock markets: linear versus non-linear GARCH models. *Journal of Forecasting* 19, 6: 499-504
- GRUDNITSKI, G. y OSBURN, L. (1993). Forecasting S and P and gold futures prices. An application of neural networks. *The Journal of Future Markets* 13, 6: 631-643

- GUARDA, P. y SALMON, M. (1996). Detection of nonlinearity in foreign-exchange rate data. En *Nonlinear Dynamics and Economics* (Eds. W.A. Barnett, A.P. Kirman y M. Salmon); Cambridge: Cambridge University Press
- HANN, T.H. y STEURER, E. (1996). Much ado about nothing? Exchange rate forecasting. neural networks vs. linear models using monthly and weekly data. *Neurocomputing* 10: 323-339
- HINICH, M.J. y PATTERSON, D.M. (1985). Evidence of nonlinearity in daily stock returns. *Journal of Business and Economic Statistics* 3, 1: 69-77
- HSIEH, D.A. (1989). Testing for nonlinear dependence in daily foreign exchange rates. *Journal of Business* 62: 339-368
- HSIEH, D.A. (1991). Chaos and Nonlinear Dynamics: Application to Financial markets, *Journal of Finance* 46, 5: 1839-1877
- JASIC, T. y WOOD, T. (2004). The profitability of daily stock market indices trades based on neural network predictions: case study for the S&P 500, the DAX, the TOPIX and the FTSE in the period 1965-1999. *Applied Financial Economics*, 14, 4: 285-297
- JIN, H.J. y FRECHETTE, D.L. (2004). Fractional integration in agricultural future price volatilities. *American Journal of Agricultural Economics* 86, 2: 432-443
- JUNGEILGES, J.A. (1996). Operational characteristics of White's test. En *Nonlinear dynamics and economics*. (Eds: W.A. Barnett, A.P. Kirman y M. Salmon); Cambridge: Cambridge University Press
- KANTZ, H. y SCHREIBER, T. (1994). Nonlinear time series analysis. Cambridge: Cambridge University Press
- KAPLAN, D.T. (1994). Exceptional Events as evidence for Determinism, *Physica D*, 73: 38-48
- KAPLAN, D.T. (1995). Nonlinearity and Nonstationarity: The Use of Surrogate Data in Interpreting Fluctuations in Interpreting Fluctuations. Center for Nonlinear Dynamics and Dept. of Physiology. McGill University, Montreal, Québec, Canada
- KIMOTO, T., ASAKAWA, K., YODA, M., y TAKEOKA, M. (1990). *Stock market prediction system with modular neural networks*. En *Proceedings of the IEEE International Joint Conference on Neural Networks*. San Diego, California, 2: 11-16
- KUAN, C.-M. y LIU, T. (1995). Forecasting exchange rates using feedforward and recurrent neural networks. *Journal of Applied Econometrics* 10: 347-364
- LACHTERMACHER, G. y FULLER, G.D. (1995). Backpropagation in time-series forecasting. *Journal of Forecasting* 14: 381-393
- LE CUN, Y. (1985). *Une Procédure d'Apprentissage pour Réseau à Seuil Asymétrique*, Proceedings of Cognitiva

- LEBARON, B. (2000). Agent based computational finance : suggested readings and early research. *Journal of Economic Dynamics and Control* 24: 679-702
- LISI, F. y MEDIO. A. (1997). Is a random walk the best exchange rate predictor? *International Journal of Forecasting* 13: 255-267
- LORENZ, E.N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* 20, 2: 130-141
- LUX, T. (1997). Time variation of second moments from a noise trader/infection model. *Journal of Economic Dynamics and Control* 22: 1-38
- LUX, T. (1998). The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions. *Journal of Economic Behavior and Organization* 22: 143-168
- LUX, T. y MARCHESI, M. (1999a). Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature* 397, febrero 1999: 498-500
- LUX, T. y MARCHESI, M. (1999b). Volatility clustering in financial markets: a micro-simulation of interacting agents. *International Journal of Theoretical and Applied Finance* 3, 4: 675-702
- MARCUCCI, J. (2005). Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics and Econometrics* 9, 4: 1-53
- McMILLAN, D.G. (2003). Non-linear predictability of UK stock market returns. *Oxford Bulletin of Economics and Statistics* 65, 5: 557-573
- MOZER, M.C. (1994). Neural net architectures for temporal sequence processing. En *Time series prediction: forecasting the future and understanding the past*. (Eds: A.S. Weigend y N.A. Gershenfeld); Reading, MA: Addison-Wesley
- NAKAJIMA, Y. (2002). Nonlinear properties found in actual markets. *Journal of the Korean Physical Society* 40, 6: 1096-1099
- NYCHKA, D., ELLNER, S., GALLANT, A.R. y MCCAFFREY, D. (1992). Finding chaos in noisy systems. *Journal of the Royal Statistical Society B* 54, 2: 399-426
- OH, K.J. y KIM, K.J. (2001). Analyzing stock market tick data using piecewise nonlinear model. *Expert Systems with Applications* 22, 3: 249-255
- O'LEARY, D.E. (1998). Using neural networks to predict corporate failure. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 7, 3: 187-197
- PARKER, D.B. (1985). *Learning Logic*. MIT Center for Computational Research in Economics and Management science, T.R. 47.
- PÉREZ-RODRÍGUEZ, J.V., TORRA, S. y ANDRADA-FÉLIX, J. (2005): are Spanish Ibex35 stock future index returns forecasted with non-linear models? *Applied Financial Economics* 15, 14: 963-975

- Qi, M. (2001). Predicting US recessions with leading indicators via neural network models. *International Journal of Forecasting* 17: 383-401
- REFENES, A.N. (1993). Constructive learning and its application to currency exchange rate forecasting. En *Neural Networks in Finance and Investing. Using Artificial Intelligence to Improve Real-World Performance*. (Eds. R.R. Trippi y E. Turban); Chicago: Probus Publishing Company
- REFENES, A.N. (1995). *Neural networks in the capital markets*. Chichester, John Wiley
- RUMELHART, D. E., MC CLELLAND, J.L. y PDP RESEARCH GROUP (Eds.) (1986). *Parallel Distributed Processing. Explorations on the Microstructure of Cognition, vol.I. Foundations*, Cambridge MA: MIT Press
- SALCHENKERGER, L.M., CINAR, E.M. y LASH, N.A. (1992). Neural networks. a new tool for predicting thrift failures. *Decision Science* 23, 4: 899-916
- SARANTIS, N. (2001). Nonlinearities, cyclical behaviour and predictability in stock markets: international evidence. *International Journal of Forecasting* 17: 459-482
- SATCHELL, S. y A. TIMMERMANN (1996). Daily returns in international stock markets. En *Nonlinear dynamics and economics*. (Eds. W.A. Barnett, A.P. Kirman y M. Salmon); Cambridge: Cambridge University Press
- SCHEINKMAN, J.A. y LEBARON, B. (1989): Nonlinear dynamics and stock returns. *Journal of Business* 62: 311-337
- SCHONEBURG, E. (1990). Stock price prediction using neural networks. a project report. *Neurocomputing* 2: 17-27
- SHINTANI, M. y LINTON, O. (2004). Nonparametric neural network estimation of Lyapunov exponents and a direct test for chaos. *Journal of Econometrics* 120: 1-33
- SOOFI, A.S. y CAO, L. (1999). Nonlinear deterministic forecasting of daily Peseta-Dollar exchange rate. *Economic Letters* 62: 175-180
- TANG, Z. y FISHWICK, P.A. (1993). Feedforward neural nets as models for time series forecasting. *ORSA Journal of Computing* 5, 4: 374-385
- TAKENS, F. (1986). Detecting strange attractors in turbulence. En *Lecture Notes in Mathematics: Dynamical Systems and Turbulence* (Eds: D.A. Rand y L.-S. Young); New York: Springer-Verlag
- TAM, K.Y. y KIANG, M.Y. (1992). Managerial applications of neural networks. the case of bank failure predictions. *Management Science* 38: 7. 926-947
- TKACZ, G. (2001). Neural network forecasting of Canadian GDP growth. *International Journal of Forecasting* 17: 57-69

- TRIPPI, R.R. y TURBAN, E. (1993). *Neural networks in finance and investment. using artificial intelligence to improve real-world performance*. Chicago: Probus Publishing Company
- TURNER, C.M., STARTZ, R. y NELSON, C.R. (1989). A Markov model of heteroscedasticity. Risk and learning in the stock market. *Journal of Financial Economics* 25: 3-22
- URIEL, E. (1992). Análisis de series temporales: Modelos ARIMA. Madrid: Paraninfo.
- WAN, E. (1994). Time series prediction by using a connectionist network with internal delay lines. En *Time series prediction: forecasting the future and understanding the past*. (Eds. A.S. Weigend y N.A. Gershenfeld); Reading, MA: Addison-Wesley
- WANG, H., CHEN, G. Y LÜ, J. (2004). Complex dynamical behaviours of daily data series in stock exchange. *Physics Letters A* 333: 246-255
- WEI, W.X. (2002). Forecasting stock market volatility with non-linear GARCH models: a case for China. *Applied Economic Letters* 9, 3: 163-166
- WEIGEND, A.S., HUBERMAN, B.A. y RUMELHART, D.E. (1992). Predicting sunspots and exchange rates with connectionist networks. En *Nonlinear Modeling and Forecasting*. (Eds. M. Casdagli y S. Eubank); Redwood City CA: Addison-Wesley
- WERBOS, P.J. (1974). *Beyond Regression. New tools for Prediction and Analysis in the Behavioral Sciences*. Ph.Th. Harvard University. Department of Applied Mathematics
- WHITE, H. (1988). Economic prediction using neural networks. the case os IBM daily stock returns. En *Proceeding of the IEEE International Conference on Neural Networks*, San Diego, California, 2: 451-458
- WHITE, A. (1989). Some asymptotic results for learning in single hidden-layer feedforward network models. *Journal of the American Statistical Association* 84, 408: 1003-1013
- WILSON, R. y SHARDA, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems* 11: 545-557
- YOON, Y. y SWALES, G. (1991). Predicting stock price performance; A neural network approach. *Proceedings of the 24th Hawaii International Conference on System Sciences* 4: 156-162.
- ZHANG, G. y HU, M.Y. (1998). Neural network forecasting of the British Pound/US Dollar exchange rate. *Omega. International Journal of Management Science* 26, 4: 495-506