

ESTUDIOS SOBRE LA ECONOMIA ESPAÑOLA

Diversas Formas de Dependencia No Lineal y Contrastes de Selección de Modelos en la Predicción de los Rendimientos del Ibex35

**Jorge V. Pérez-Rodríguez
Salvador Torra Porras**

EEE 94

Marzo, 2001



<http://www.fedea.es/hojas/publicado.html>

Diversas Formas de Dependencia No Lineal y Contrastes de Selección de Modelos en la Predicción de los Rendimientos del Ibex35

Jorge V. Pérez-Rodríguez*
Universidad de Las Palmas de Gran Canaria

y

Salvador Torra Porras
Universidad de Barcelona

Resumen:

El objetivo de este trabajo es analizar y evaluar la existencia de distintos tipos de dependencia no lineal, que están presentes en la conducta dinámica de las rendimientos diarios del índice del mercado continuo español desde el 30/12/1989. En primer lugar, proporcionamos una visión amplia de las propiedades estadísticas de dicho índice: estacionariedad, reversión a la media, memoria a largo plazo y dependencias no lineales en los rendimientos. En segundo lugar, utilizamos diversos modelos no lineales estocásticos uniecuacionales: los modelos ARCH, los modelos de cambios de régimen (STAR y procesos markovianos), y las redes neuronales; para realizar una evaluación estadística exhaustiva de las predicciones de los rendimientos diarios mediante estadísticos de bondad de la predicción, contrastes de predicción direccional y de igualdad de las predicciones. En general, los resultados obtenidos indican que existe una desviación transitoria de la senda del camino aleatorio y que los rendimientos son más predecibles cuando se utilizan modelos no lineales; en los que las redes neuronales son ligeramente preferibles.

Palabras Clave: Ibex35, GARCH, STAR, cambios de régimen markovianos, RNA.

JEL: C22, C45, C52.

* Agradecemos a Julián Andrada los valiosos comentarios realizados, así como la ayuda prestada en versiones preliminares de este documento. También, agradecemos al Programa INNOVA de Mecenazgo Universitario de la Caja Insular Ahorros de Canarias el apoyo financiero prestado. Los resultados de este trabajo han sido obtenidos utilizando diversos programas informáticos tales como GAUSS, MATLAB, RATS y BDS-STATS escrito por William Dechert. Todos los posibles errores son de la entera responsabilidad de los autores.

1. Introducción

La naturaleza y conducta dinámica de las rentabilidades de los mercados de activos ha sido, y continúa siendo, de interés para los académicos, reguladores de mercado, operadores e inversores, quienes aún suponiendo que el mercado es eficiente¹, están también interesados en la detección y explotación de cualquier patrón de la conducta seguida por los precios de los activos en su evolución temporal.

Estudios recientes sobre los mercados de activos han cuestionado que la conducta de las rentabilidades de los activos sea totalmente aleatoria, y que las técnicas de modelización lineal sean adecuadas para captar algunos patrones complejos que los chartistas observan en la evolución del precio de los activos [véase Fama (1965, pág.80)]. Así, en las últimas dos décadas, algunos investigadores como Hinich y Paterson (1985), Cochrane (1988), Fama y French (1988), Lo y MacKinlay (1988), White (1988), Sheinkamn y LeBaron (1989), Hsieh (1991), Granger (1992), Gençay (1996), Campbell, Lo y MacKinlay (1997), De Lima (1998), Fernández, García y Sosvilla (1999), García y Gençay (2000), o investigadores sobre el mercado de activos español como Ratner (1996), Fernández, Sosvilla y García (1997) y Fernández, González y Sosvilla (2000), han cuestionado y puesto en duda la hipótesis de los mercados eficientes (HME); que el modelo de camino aleatorio sea una descripción razonable del movimiento del precio de los activos; así como, que los modelos lineales describan acertadamente la evolución de los precios de los mismos. En este sentido, muchos de estos trabajos han contribuido a una mejora en la predicción del futuro y la toma de decisiones, y como dice Haugen (1999), han contribuido a que las Finanzas académicas estén preocupadas por el análisis y estudio de los mercados ineficientes en la actualidad².

¹ La creencia es que el precio actual de un activo refleja el valor de mercado de los *cash-flows* futuros de la empresa, y además, cada activo se valora correctamente utilizando toda la información disponible. Consecuentemente, cualquier cambio en el precio de los activos será el resultado de nueva información sobre la empresa. La hipótesis débil de los mercados eficientes (HDME) viene a decirnos que los movimientos de los precios son impredecibles. La mejor predicción para el precio es el valor corriente, y el precio actual sigue el denominado paseo aleatorio.

² Para analizar sus implicaciones, Haugen dice que el fundamento de tal análisis se encuentra en la estadística, econometría y psicología. Al respecto, podríamos concretar que un mercado ineficiente es aquel que tiene errores en la valoración de los activos. Estos errores resultan de tendencias en los precios, y los activos que posean características particulares y propias tienden a producir *premiums* de rentabilidades que pueden ser aprovechados por los agentes. Factores como el número de participantes en el mercado; el número de activos negociados; la capitalización del mercado; las leyes institucionales y reguladoras; así

Existe una creencia extendida entre los analistas de que los movimientos irregulares y persistentes de muchas series financieras y macroeconómicas resultan de no linealidades inherentes a la negociación en el mercado. Las correcciones que hace el mercado sobre las desviaciones del precio de mercado y su valor teórico pueden no ser proporcionales a la cantidad por la que el precio se desvía de su verdadero valor. Por ello, cabría esperar correcciones no lineales. El mecanismo no lineal puede relacionarse al estudio de la psicología de mercado, donde es comprensible que los mercados sobre-reaccionen ante la llegada de malas noticias. De hecho, la conducta de muchas variables económicas es asimétrica sobre el ciclo económico y esto se demuestra porque los *shocks* negativos tienen un mayor impacto que los que son positivos de igual magnitud. Por esto, tal y como se demuestra en un número cada vez mayor de estudios en economía, se presta una menor atención a los modelos lineales en favor de los modelos estocásticos no lineales de series temporales (tanto uniecuacionales como multiecuacionales). Esta mayor atención y creciente interés entre los economistas en los últimos años se justifica, además, por la riqueza de estructuras no lineales que pueden utilizarse para describir y predecir las series económicas. Podríamos citar, brevemente, los modelos GARCH y sus extensiones, los modelos autorregresivos umbral (TAR) y sus extensiones STAR; los modelos de cambios de régimen markovianos (MS); los *wavelets*, la predicción por analogías (NN) o las redes neuronales artificiales (RNA), que se muestran capaces de resolver aceptablemente la difícil tarea de la predicción.

Pues bien, en este artículo pretendemos contribuir al análisis de la naturaleza y conducta temporal de los rendimientos de los activos, concretamente del índice Ibex35³; a través de la modelización no lineal de los efectos asimétricos que se producen en los precios, y además, a través de la evaluación del comportamiento predictivo de los modelos no lineales considerados. Esta última perspectiva, que es una comparación competitiva de las predicciones, se adopta para evaluar la eficacia de dichos modelos. Para ello, la comparación será realizada mediante el cálculo de varios estadísticos de

como otras barreras de entrada, que podrían afectar la eficiencia del mercado, y producir beneficios, al menos en el corto plazo.

³ La importancia del Ibex35, como índice compuesto por los treinta y cinco valores más líquidos que cotizan en el SIBE, radica en que está diseñado para servir de subyacente en la contratación de productos derivados sobre índices: opciones y futuros. Así, la posibilidad de adquirir o vender opciones hace que sea posible para el inversor determinar con una precisión mucho mayor cuáles son los riesgos que desea asumir en función de sus expectativas. Dado que en la evaluación de una opción se tienen en cuenta el precio de ejercicio, el plazo al vencimiento, el precio del activo subyacente, la volatilidad y el tipo de interés a corto plazo; juega un importante papel el conocimiento de la dinámica a corto plazo de los precios: tanto en media como en varianza.

bondad de ajuste o mediante procedimientos de contraste de hipótesis formales. Destacamos como medidas de bondad de la predicción a los estadísticos error absoluto medio (EAM), error porcentual absoluto medio (EPAM), raíz del error cuadrático medio (RECM), U-Theil, *confusion matrix* [véase Swanson y White (1995)]; y como contrastes de hipótesis los contrastes de abarcamiento [Chong y Hendry (1986)], los de igualdad de los ECM de predicción [Diebold y Mariano (1995)] y los de predicción direccional [Schnader y Steckler (1990), Pesaran y Timmermann (1992)].

El trabajo se organiza de la siguiente manera. En la sección 2, se realiza una sinopsis sobre la no estacionariedad y no linealidad de los rendimientos de los activos. En la sección 3 se introducen los modelos no lineales asimétricos que se utilizan en este estudio. En la sección cuarta, se analizan exhaustivamente las propiedades de los rendimientos del Ibex35 mediante los contrastes de raíces unitarias y de reversión a la media, así como varios contrastes de no linealidad. Seguidamente, pasamos revista a diferentes especificaciones no lineales clásicas dentro de la literatura econométrica y empírica. Y, finalmente analizamos la capacidad predictiva de algunos de los modelos planteados. Por último, en la sección cinco destacamos las conclusiones más relevantes de este estudio.

2. No estacionariedad y no linealidad en la evolución de los precios de los activos

Un nuevo examen de la conducta y naturaleza de los rendimientos de los activos ha contribuido a poner en duda la hipótesis débil de los mercados eficientes desarrollada por Fama (1965) y, de esta forma, que el movimiento de los precios sea adecuadamente descrito por el modelo de paseo aleatorio. Estudios recientes sobre los precios de los activos destacan dos cuestiones fundamentales que admiten una conducta predecible de las rentabilidades de los activos. En primer lugar, la presencia de la reversión a la media y la existencia de autocorrelaciones negativas en los precios. La significación de la reversión a la media en los precios de los activos puede contrastarse, en principio, mediante la descomposición del precio de los activos en dos componentes. Muchos autores han propuesto métodos de descomposición de las series de precios en componentes transitorios y permanentes, basándose en el contexto de los contrastes de persistencia o no estacionariedad, que en la literatura son paralelos a los contrastes de raíces unitarias y que justifican la presencia de componentes similares a los del paseo aleatorio. Así, por ejemplo, algunos autores como Cochrane (1988), Poterba y Summers

(1988), Lo y MacKinlay (1988), Kim, Nelson y Startz (1989) empleando el contraste de razón de las varianzas, argumentan que los precios poseen reversión a la media en el largo plazo; y autores como Fama y French (1988) muestran que la existencia de un componente de reversión a la media en los precios de los activos tiende a inducir autocorrelaciones negativas en las rentabilidades que son más fuertes en el largo plazo. En este sentido, desarrollan un contraste de la hipótesis de mercados eficientes que está basado en que la conducta de las rentabilidades a largo plazo pueden proporcionar una clara impresión de la importancia de la reversión a la media de los componentes de los precios. Específicamente, un componente que decae lentamente induce autocorrelaciones negativas. En general, todos estos autores consideran que de esta forma puede justificarse la existencia de ciertas desviaciones transitorias de la senda de paseo aleatorio.

Por otro lado, también la utilización de modelos no lineales ha puesto en duda la hipótesis débil de mercados eficientes. Algunos contrastes de linealidad, como el exponente de Hurst o el contraste BDS, tal y como demuestran Fernández, Sosvilla y García (1997), De Lima (1998) y Opong, Mulholland, Fox y Farahmand (1999), muestran la existencia de no linealidades y, por lo tanto, de patrones complejos en los datos analizando distintos índices⁴. Las razones que podríamos argumentar para justificar la no linealidad son variadas. Por un lado, podríamos admitir la posibilidad de que no todos los agentes alcanzasen simultáneamente toda la información; que se provocasen importantes diferencias en los objetivos y en el tiempo de negociación; o que aquellos agentes con algoritmos más sofisticados fuesen capaces de hacer mejor uso de la información disponible. Por ejemplo, Hinich y Paterson (1985) han mostrado que las rentabilidades son realizaciones muestrales de procesos estocásticos no lineales estacionarios; y Granger (1992) argumenta que ampliando el horizonte temporal, utilizando datos desagregados, tratando adecuadamente los eventos excepcionales y

⁴ La no linealidad es un concepto que resulta más difícil definir. No obstante, Franses (1998, pág.174) dice que: “una serie temporal económica puede decirse que es no lineal cuando grandes *shocks* tienen un impacto diferente al producido por *shocks* pequeños, en el sentido de que el impacto de un *shock* no es proporcional a su tamaño”. En cierto sentido, podríamos decir que no linealidad podría implicar que el impacto de los *shocks* depende del signo de estos, es decir, de si son positivos o negativos. Precisamente, esta característica aparece en los mercados de activos, donde existe un predominio más frecuente de *shocks* negativos que de *shocks* similares de tamaño positivo.

outliers, y especialmente considerando la no linealidad, pueden obtenerse mayores beneficios⁵.

Sin embargo, el tratamiento de la no linealidad no resulta fácil por varios motivos. En primer lugar, porque existe un número elevado de alternativas: modelos bilineales, modelos de heterocedasticidad autorregresiva condicional (ARCH y sus extensiones), modelos *smooth transition autoregressive* (STAR), los modelos de redes neuronales artificiales (RNA), los algoritmos genéticos, los *wavelets* e incluso la dinámica caótica, los cuales no sólo hacen difícil la selección de los modelos sino que también pueden favorecer el ***data mining***. En segundo lugar, porque la flexibilidad existente en su utilización puede crear ajustes espurios [Granger y Teräsvirta (1993)]; y en tercer lugar, porque cuando se considera un periodo de tiempo muy grande puede aparecer el problema del cambio de régimen y la existencia de más *outliers* que dificultan la estimación de los modelos [De Lima (1998)].

La aplicación de los modelos no lineales a las series de rendimientos de los activos ha tenido distinto éxito. Así como Lamoreaux y Lastrapes (1990) argumentan que la presencia de cambios en la varianza no condicional pueden explicarse por la persistencia de la varianza condicional, es decir la integrabilidad de la varianza; Hsieh (1991) destaca que la no linealidad de las rentabilidades de los activos se debe principalmente a la existencia de efectos ARCH y no tanto a la existencia de cambios de estructura. Algunos estudios como Granger y Teräsvirta (1993), Swanson y White (1995), Gençay (1996) o Clements y Smith (1999) indican que los modelos no lineales del tipo STAR y RNA, aplicados a las series financieras, y cuando se les compara con modelos lineales, algunas veces mejoran la predicción. Pero, también autores como Episcopos y Davis (1995), Giles, Lawrence y Tsoi (1997), Pérez-Rodríguez, Torra y Borrell (2000) y García y Gençay (2000), entre otros, demuestran que las RNA predicen algo mejor que los modelos lineales y los modelos no lineales de tipo ARCH. En el caso concreto de los rendimientos de los activos, White (1988), Collard (1991) o Donaldson y Kamstra (1996) destacan la aplicación de las RNA. Particularmente, los últimos autores investigan el uso de la RNA para predecir los rendimientos de diferentes índices de los mercados de activos internacionales (USA, Canadá, Japón y UK), mostraron que en términos predictivos estas técnicas dominaban a los procedimientos lineales, porque

⁵ Ahora bien, si no se encuentra una regla de beneficio y esta no produce beneficios durante un periodo extenso, entonces la HDME no podría ser rechazada.

su flexibilidad tenía en cuenta relaciones no lineales complejas que no eran capturadas fácilmente por los modelos lineales.

En España, el análisis de la capacidad predictiva de los rendimientos bursátiles mediante modelos no lineales se ha centrado en el estudio de la volatilidad condicional de los rendimientos de un índice compuesto [Alonso (1995), para el que la evidencia empírica sobre las regularidades que caracterizan la volatilidad del mercado español de renta variable es la ausencia de respuesta asimétrica de su volatilidad ante movimientos alcistas y bajistas de los precios]; y los rendimientos del Ibex35 [León y Mora (1999), entre otros, quienes demuestran la relevancia de efectos asimétricos de la volatilidad ante *shocks* negativos que positivos de igual magnitud, y no tanto los denominados efectos en media]; en el estudio del índice general de la Bolsa de Madrid por Fernández, Sosvilla y García (1997) utilizando predictores locales inspirados en la literatura de predicción en sistemas no lineales; y Fernández, González y Sosvilla (2000) utilizando un modelo simple de redes neuronales artificiales para evaluar los beneficios de una regla técnica de negocio, que es superior a la del paseo aleatorio.

3. *Dinámica no lineal en los rendimientos de los activos*

Consideremos un activo que proporciona una rentabilidad diaria por peseta invertida igual a r_t , $t=1, \dots, T$. Consideremos al mercado de activos como un sistema de procesamiento de información. El conjunto de información cambia continuamente, y el proceso de información del mercado produce un ajuste de los precios hacia el valor de mercado que considera apropiado. Podríamos considerar que el mercado forma una expectativa para el próximo periodo, condicionada a la información existente, que matemáticamente podría escribirse como: $E_{t-1}[r_t] = f(\psi_{t-1})$, siendo ψ_t el conjunto de información en el periodo t . La función $f(\cdot)$ podría ser lineal o estar caracterizada por funciones no lineales complejas. Algunos autores dicen que el mercado de activos tiene la propiedad de ser un sistema dinámico no lineal que produce una conducta caótica.

Considerado este entorno, podríamos escribir que la rentabilidad en el periodo t fuese igual a: $r_t = f(\psi_{t-1}) + \varepsilon_t$, siendo ε_t un error de predicción. Dado que está bien documentado que las rentabilidades son a menudo predecibles; se asume, por ejemplo, que la secuencia $\{\varepsilon_t\}$ es independiente con media nula, y posiblemente variable en el tiempo, tal que: $E_{t-1}[\varepsilon_t^2] = h_t$. A continuación, vamos a describir algunos modelos no

lineales que utilizaremos en este estudio, que afectan a los dos términos de la ecuación definida para las rentabilidades.

3.1. Modelos de volatilidad ARCH

La evidencia empírica presentada hasta el momento - dentro de la literatura econométrico-financiera - sugiere que la volatilidad de la rentabilidad de mercado no es únicamente variable temporalmente, sino que está asimétricamente relacionada con las innovaciones pasadas, esto es, las rentabilidades inesperadas que son negativas afectan mayormente a la volatilidad futura que las que son inesperadas y positivas de igual magnitud. Este hecho se justificaría, por ejemplo, porque las acciones de los agentes están basadas en la distribución condicional a la información disponible de las rentabilidades, y en el hecho de que los datos reflejan el resultado de la negociación entre compradores y vendedores. En este sentido, podríamos decir que ε_t suele poseer una distribución cuya varianza es heterocedástica y está generada por un proceso dependiente temporalmente en los cuadrados. Dentro de los modelos no lineales que más se han popularizado en los últimos años, y que además son extensamente utilizados en finanzas, están los modelos ARCH. La especificación de tales modelos es amplia, y muchos han sido investigados por autores como Engle y Ng (1993). No obstante, los modelos que emplearemos son los siguientes. Suponiendo que la varianza condicionada es h_t , el modelo GARCH(1,1) se escribe del modo siguiente:

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \quad [1]$$

siendo ε_t los errores estimados del modelo ARMA(p,q) para la rentabilidad; los parámetros son ω , α y β , y donde $\alpha + \beta < 1$ es la condición de existencia de la varianza incondicional. Este modelo es claramente simétrico, es decir, los *shocks* negativos y positivos de igual magnitud incrementan la volatilidad. En este sentido, para considerar la respuesta asimétrica de la volatilidad ante *shocks* negativos y positivos se introducen dos modelos. Por un lado, el modelo de Glosten, Jagannathan y Runkle (GJR, 1993), denominado GARCH umbral, que puede escribirse para un GJR(1,1) como:

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} + \gamma \varepsilon_{t-1}^2 D_{t-1} \quad [2]$$

donde $\gamma > 0$ es el coeficiente del efecto asimetría y $D_t = 1$ si $\varepsilon_{t-1}^2 < 0$ y cero en caso contrario. Por otro lado, está el modelo EGARCH. Este es un modelo de

heterocedasticidad multiplicativa que no anida al modelo ARCH. De esta forma, bajo la normalidad del error podríamos escribir el modelo EGARCH(1,1) tal como;

$$\log h_t = \omega + \beta \log h_{t-1} + \gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \alpha \left[\frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} - \sqrt{\frac{2}{\pi}} \right] \quad [3]$$

donde, $\omega > 0$, $0 < \beta < 1$, $0 < \alpha < 1$, y $\gamma < 0$. β mide la persistencia de la volatilidad, mientras que γ mide el efecto “apalancamiento” o asimetría. La estimación se realiza por máxima verosimilitud, donde la función del logaritmo de verosimilitud es igual a:

$$\text{Log } L(\theta) = -\sum_{t=1}^T \left[\log(h_t) + \frac{\varepsilon_t^2}{h_t} \right], \text{ siendo } \theta \text{ el vector de parámetros.}$$

3.2. Modelos de cambios de régimen

Un camino alternativo para introducir la asimetría es considerar modelos de cambio de régimen, tanto en la rentabilidad como en su volatilidad⁶. Hemos visto que la no linealidad puede representarse mediante diverso tipos de modelos de tipo ARCH como [1], [2] y [3]. Sin embargo, en este epígrafe vamos a especificar otros modelos no lineales que también pueden representar la conducta del Ibex35. Los modelos de cambios de régimen que consideramos son los conocidos como *smooth transition regression* (STAR) y los cambios de régimen markovianos. La diferencia entre ambos es que en los primeros se permite una transición más alisada de un régimen a otro mientras que los segundos se caracterizan por cambios más abruptos.

Los modelos STAR(p;d) pueden representarse, siguiendo a Teräsvirta, Tjøstheim y Granger (1994), según:

$$r_t = \phi_{10}r_{t-1} + \dots + \phi_{1p}r_{t-p} + [\phi_{20}r_{t-1} + \dots + \phi_{2p}r_{t-p}]F_{t,d}(\gamma, c) + \varepsilon_t \quad [4]$$

donde r_t son las rentabilidades, ϕ_{ij} , $i=1,2$; $j=0,1,2,\dots,p$ son los parámetros desconocidos que corresponden a cada uno de los dos regímenes; $F_{t,d}(\gamma, c)$ es una función que varía con t , depende de los parámetros: γ , el parámetro umbral c , que representa el cambio de

⁶ La inestabilidad de un modelo se define a veces como un cambio de un régimen a otro en la ecuación de regresión. Sin embargo, en muchas ocasiones los investigadores tienen poca información sobre los momentos del tiempo en que los parámetros cambian, y de esta manera se necesitan hacer inferencias sobre tales acontecimientos desconocidos de cambio así como la significación de los parámetros cambiantes. Por ejemplo, Goldfeldt y Quandt (1973) permiten cambios independientes del estado, e introducen la dependencia mediante cambios markovianos. Hamilton (1989) desarrolla un modelo que es una extensión de Goldfeld y Quandt (1973), basándose en un cambio estructural de los parámetros de un proceso autorregresivo, el cual sigue un régimen cambiante markoviano dependiente del estado.

un régimen a otro, y el número de retardos de la variable, d . Esta función introduce el cambio de régimen en los parámetros del modelo y la no linealidad. El cambio de régimen puede especificarse mediante una función logística: $F_{t,d}(\gamma, c) = \{1 + \exp[-\gamma(r_{t-d} - c)]\}^{-1}, \gamma > 0$, en cuyo caso el modelo se denomina LSTAR(p;d); o mediante la distribución exponencial, para lo cual $F_{t,d}(\gamma, c) = \{1 - \exp[-\gamma(r_{t-d} - c)^2]\}, \gamma > 0$, y en cuyo caso, el modelo es denominado como ESTAR(p;d). En ambos casos, la variable de transición puede ser cualquier variable contenida en el conjunto de información ψ_{t-1} . La estimación de este tipo de modelos suele realizarse mediante mínimos cuadrados no lineales (MCNL) o máxima verosimilitud, pues proporciona resultados equivalentes; y después de un gran número de iteraciones suele alcanzarse el valor óptimo de la función objetivo. Adicionalmente, podría considerarse que el término de perturbación tuviese varianza condicional definida por [1], [2] y [3], como otra vía para introducir la no linealidad y para estar en consonancia con los argumentos de la sección 3.1. En este caso, la estimación se realizaría por máxima verosimilitud suponiendo, por ejemplo, normalidad condicional utilizando la expresión definida en la sección 3.1.

En otra situación están los modelos de cambios de régimen markovianos. En estos modelos la dinámica de la serie depende de la realización de un número finito de estados discretos, y en él, los cambios dependen del estado en que se encuentra el sistema económico. Los modelos de cambios markovianos se han popularizado en los estudios de la producción industrial, los tipos de interés, los precios de los activos o en las tasas de desempleo. Estos modelos se adoptan genéricamente para representar características específicas de las series temporales económicas, tales como la asimetría de la actividad económica [Hamilton (1989)], la existencia de distribuciones mixtas con colas gruesas [Nefçi (1984)], el agrupamiento de la volatilidad y la reversión a la media en los precios de los activos [Turner, Startz y Nelson (1989)] o en los tipos de interés [Hamilton (1988)]. En general, el proceso de Markov supone que el proceso puede cambiar aleatoriamente y abruptamente de un régimen a otro. En nuestro caso, se pretende la modelización de los *shocks* transitorios mediante un modelo empírico que represente la dinámica a corto plazo en los precios de los activos tanto en la media como en la varianza. Por un lado, consideramos el modelo de cambios de régimen de Hamilton, que es un AR(p) donde la media sigue un proceso de Markov de primer orden tal que:

$$r_t - \mu_{s_t} = \phi_1(r_{t-1} - \mu_{s_{t-1}}) + \dots + \phi_p(r_{t-p} - \mu_{s_{t-p}}) + u_t, \quad u_t \sim iid(0, \sigma^2)$$

tal que la media del proceso es igual a: $\mu_{s_t} = \mu_1(1 - S_t) + \mu_2 S_t$, distinguiendo dos estados de la economía que están definidos por la variable no observable S_t , que está generada por un proceso de Markov de primer orden y que toma dos valores: el cero y el uno. La transición de un estado a otro puede describirse por la probabilidades de transición, que son iguales a: $P[S_t = 1/S_{t-1} = 1] = p_{11}$ y $P[S_t = 0/S_{t-1} = 0] = p_{22}$. Sin embargo, este proceso no recoge los cambios que experimenta la varianza de dichos estados. En este sentido, consideramos el modelo empírico para las rentabilidades con heterocedasticidad markoviana para dos regímenes o estados, definido por Turner, Startz y Nelson (1989, TSN). Estos autores introducen un modelo para el mercado de activos en el que los rendimientos (por ejemplo, los excesos de rentabilidad) están caracterizados por una mezcla de dos densidades normales. Suponiendo que el mercado cambia entre dos estados; que estos están caracterizados por las varianzas de sus densidades, las cuales representan a un estado de baja y elevada varianza; que el estado es generado por un proceso de Markov de primer orden; su modelo permite que la varianza condicional sea una función estocástica de la varianza de periodos previos. Su modelo se escribe de la siguiente manera: $r_t = \mu + \varepsilon_t$, $\varepsilon_t \sim N(0, \sigma_t^2)$, siendo μ una constante (pero que podría ser modelizable de forma similar al modelo de Hamilton presentado anteriormente), y donde la varianza es igual a: $\sigma_t^2 = \sigma_1^2 S_{1t} + \sigma_2^2 S_{2t}$ siendo σ_i^2 la varianza del régimen $i=1,2$, donde $\sigma_1^2 < \sigma_2^2$; y $S_{kt} = 1$ si $S_t = k$, $k=1,2$ y $S_{kt} = 0$ en caso contrario. Además, $P[S_t = j/S_{t-1} = i] = p_{ij}$, con $i,j=1,2$ y $\sum_{j=1}^2 p_{ij} = 1$. En este modelo, la volatilidad depende del estado en que se encuentra la economía, y p_{ij} son las probabilidades de transición del estado i al estado j . Este modelo permite decrecer más rápidamente a la volatilidad. La estimación de ambos modelos se realiza por máxima verosimilitud. El logaritmo de verosimilitud está definido por:

$$\log L(\theta) = \sum_{t=1}^T \ln \left\{ \sum_{S_t=1}^J \sum_{S_{t-1}=1}^J f(r_t/S_t, S_{t-1}, \psi_{t-1}) \Pr[S_t, S_{t-1}/\psi_{t-1}] \right\}$$

donde J es el número de estados de la naturaleza, $f(\cdot)$ es la función de densidad condicional de las rentabilidades a los estados S_t, S_{t-1} y al conjunto de información ψ_{t-1} y $\Pr[S_t, S_{t-1}/\psi_{t-1}]$ son las ponderaciones o probabilidades. Precisamente, la obtención de

Pr[.] necesita de la aplicación de un filtro que calcula las probabilidades y luego las actualiza. Este filtro puede ser bien el derivado por Hamilton o el definido por Kim⁷.

3.3. Modelos de redes neuronales

Aunque no exenta de críticas, el uso de las redes neuronales artificiales (RNA) se ha popularizado en los últimos años, dada la mayor comprensión de estos modelos. Especialmente, estos modelos son aproximadores universales; permiten representar automáticamente complejas relaciones no lineales; describe varias formas de cambios de régimen, por lo tanto distintos efectos asimétricos, lo que permite sugerir que existen subperiodos más predecibles que otros; y son buenos predictores. Sus aplicaciones en el ámbito de la economía se han centrado fundamentalmente en la gestión. Por ejemplo, en el ámbito de los datos de corte transversal; la predicción de bancarrotas [Tam y Kiang (1992)] y los *rating* de los bonos corporativos [Surkan y Singleton (1990), Moody y Utans(1995)]; y en el ámbito de la predicción de series temporales, a parte del estudio de los rendimientos de los activos (citados en el epígrafe 2), los relacionados con la decisión [Sharda y Patil (1992) y Hill, Márquez, O'Connor y Remus (1994), entre otros]. En general, casi todos los trabajos analizan la capacidad predictiva de las redes comparando diversos modelos, tanto lineales como no lineales. Sin embargo, los resultados determinan una moderada ventaja de la predicción RNA frente a cualquiera de los modelos lineales ARIMA y no lineales GARCH de series temporales que son estudiados.

Una RNA es una colección de funciones de transferencia que relacionan a la variable output de interés, r_t , con algunas variables inputs, z_t , las cuales pueden ser funciones incluso de otras variables explicativas. En este sentido, las RNA son una clase de modelos de regresión no lineales y esencialmente dispositivos para la inferencia estadística no paramétrica. Estas se caracterizan por dos aspectos básicos: una especificación paramétrica o *topología de red*, y un mecanismo para la estimación o *entrenamiento de la red*. Desde el punto de vista estadístico, tienen una interpretación simple. Supongamos una muestra $D_T = \{z_t, r_t\}_{t=1}^T$, formada por un conjunto de observaciones que es finito de las variables explicativas contenidas en el vector z_t (por ejemplo, $r_{t-1}, \dots, r_{t-p}$) así como la variable r_t . Supongamos que esta muestra puede

⁷ Para una panorámica de la estimación de este tipo de modelos véase Kim y Nelson (1999), págs 65-66.

generarse mediante una función desconocida, $f(z)$, y la suma de un componente residual estocástico, ε . Entonces, la tarea del aprendizaje neuronal se basa en construir un estimador $G(z, \theta; D_T) \equiv \hat{f}(z)$ de $f(z)$, donde $\theta = (\theta_1, \dots, \theta_p)$ es un conjunto de parámetros libres. Ya que no existen supuestos a priori sobre la forma funcional de $f(z)$, el modelo de redes $G(z, \theta)$ es un estimador no paramétrico de la densidad condicional $E[r/z]$, en oposición al estimador paramétrico que conocemos habitualmente y donde la forma funcional es conocida a priori.

Un RNA puede ser tan simple como un simple nodo con un simple input. No obstante, pueden existir muchos nodos iniciales que filtran los datos para producir outputs intermedios; luego estos outputs intermedios son usados como inputs en una segunda capa de nodos, produciendo nuevos outputs que quizás puedan ser utilizados en una tercera capa de nodos y así sucesivamente, hasta que el output final sea producido.

Algunos modelos de red neuronal que utilizaremos en este trabajo son los siguientes. En primer lugar, la red *multilayer perceptron* o MLP, la cual es el modelo neuronal más conocido y nombrado en la literatura especializada. En general, el modelo de regresión no lineal que representa la RNA de tipo MLP(p;q) tiene la forma:

$$r_t = \mu + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \dots + \phi_p r_{t-p} + \sum_{j=1}^q \beta_j g(\mu_j + \phi_{1j} r_{t-1} + \dots + \phi_{pj} r_{t-p}) + \varepsilon_t \quad [5]$$

donde $\mu, \phi_1, \dots, \phi_p, \beta_j, j=1, \dots, q$, son parámetros; p es el número de los términos autorregresivos; q es un entero; $\lambda_j = (\mu \phi_{1j} \dots \phi_{pj})$ son factores de escala necesariamente seleccionados para alcanzar la identificación de los parámetros β_j , y donde están generados mediante una distribución uniforme, tal que $0.5\lambda_j \sim U[-1,1]$; $g(\cdot)$ es una función de transferencia no lineal de tipo sigmoidal que determina las conexiones entre los nodos de la capa oculta⁸; ε_t es un residuo iid; y finalmente r_t es la rentabilidad en t o output del sistema⁹. En su versión más compleja, la topología permite

⁸ La función $g(\cdot)$ es sigmoidal si $g: \mathbb{R} \rightarrow [0,1]$; $g(a) \rightarrow 0$ cuando $a \rightarrow -\infty$; $g(a) \rightarrow 1$ cuando $a \rightarrow \infty$. Por ejemplo, g puede ser una función logística: $g(a) = [1 + \exp(-a)]^{-1}$. También podría ser una función bipolar: $h(a) = 2g(a) - 1$. O podría estar definida por la función tangente hiperbólica: $\tanh(a) = [\exp(a) - \exp(-a)] / [\exp(a) + \exp(-a)]$, tal y como hemos considerado en este trabajo.

⁹ Hornick, Stinchcombe y White (1990) demostraron que las RNA de la forma definida por [5] son aproximadores universales en una amplia variedad de espacios de funciones de interés práctico. Esto significa que muchas funciones de uso habitual pueden aproximarse (generalmente en compactos) con precisión arbitraria por medio de estructuras del tipo [5], con tal de que se elijan adecuadamente la complejidad del modelo, indexada por q, y el vector de parámetros o pesos.

la introducción de una o más capas ocultas entre el output y los inputs. Este modelo tiene como principal ventaja, la habilidad para actuar como aproximador de relaciones complejas no lineales. Sin embargo, su desventaja principal es la naturaleza estática, la cual es superada por otras topologías que incorporan la dinámica de las relaciones entre inputs-outputs a lo largo del tiempo.

En segundo lugar, utilizamos también la red con conexiones directas entre los inputs y el output, denominada *jump connection nets* o JCN(p;q). Siguiendo a Kuan y White (1994), la expresión del output del modelo es:

$$r_t = F \left[\mu + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \dots + \phi_p r_{t-p} + \sum_{j=1}^q \beta_j g(\mu_j + \phi_{1j} r_{t-1} + \dots + \phi_{pj} r_{t-p}) \right] + \varepsilon_t \quad [6]$$

donde F(.) indica el tipo de función ajustable para el output. Generalmente es una función de distribución acumulada. Así, por ejemplo, F(.) podría ser una distribución logística o normal. Si F(a)=a, por ejemplo, tenemos el modelo lineal estándar aumentado por términos no lineales [5]. En nuestro caso, F(.)=g(.).

Y finalmente, utilizaremos las redes parcialmente recurrentes, donde destacan los esquemas de Jordan(p;q) y Elman(p;q). Estas poseen la propiedad de reconocer y, a veces, reproducir secuencias. En este sentido, estas clases de RNA son algo más complejas que la RNA unidireccional definida por [5] e incluso por [6]. En el caso concreto de la red recurrente de tipo Elman [véase Elman (1990)], esta se caracteriza por una estructura de tipo dinámico¹⁰, que puede adoptar la forma:

$$\begin{aligned} r_t &= \mu + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + \sum_{j=1}^q \beta_j c_{tj} + \varepsilon_t \\ c_{tj} &= g(\mu + \phi_1 r_{t-1} + \dots + \phi_p r_{t-p} + c_{t-1,j} \delta_j) \end{aligned} \quad [7]$$

donde c_t es un vector de variables de estado, también llamado contexto. Este tipo de redes posee algunas propiedades que las hacen especialmente adecuadas para la modelización de series temporales. Así, desde la perspectiva econométrica, la estructura de RNA adoptada aquí es una clase particular de modelo dinámico con variables latentes. Dicha representación es ventajosa porque vincula las redes recurrentes a una clase de modelos conocidos en nuestro ámbito como son los modelos de espacio de estados. Por último, existe una variante de redes parcialmente recurrentes elaborada por

¹⁰ Para una detallada discusión sobre las RNA y sus aplicaciones econométricas, véase Kuan y White (1994).

Jordan, donde, a diferencia de Elman, donde la conexión de las variables de estado se produce desde la capa oculta, la conexión en este caso se produce desde el output.

La estimación de la RNA puede realizarse minimizando la siguiente función:

$$\min_{\theta \in \Theta} T^{-1} \sum_{t=1}^T [r_t - f(z_t, \theta)]^2 \quad [8]$$

El mínimo, si existe, converge casi seguramente - bajo una condición general cuando $T \rightarrow \infty$ - hacia θ^* , que es la solución de problema. El estimador m-recursivo sigue un procedimiento Robbins-Monro tal como: $\hat{\theta}_{t+1} = \hat{\theta}_t + \Delta_t = \hat{\theta}_t + \delta_t(r_t, z_t'; \hat{\theta}_t)$. Este procedimiento, también conocido como regla delta, puede aplicarse a una especificación de red como [5], [6] o [7] y permite, dado un $\hat{\theta}_t$, el conocimiento de $\hat{\theta}_{t+1}$. El término Δ_t recoge el aprendizaje de la red. Una generalización de la regla delta – también denominada método *backpropagation* – cuando, por ejemplo, el output de la red $f(z, \theta)$ es no lineal puede realizarse atendiendo a la siguiente expresión que apuntó White (1987):

$$\hat{\theta}_{t+1} = \hat{\theta}_t + a_t \nabla f(z_t, \hat{\theta}_t) [r_t - f(z_t, \hat{\theta}_t)]$$

donde $\nabla f(z_t, \hat{\theta}_t)$ es el gradiente de $f(z, \hat{\theta})$. Esta es una aproximación estocástica de los mínimos cuadrados no lineales recursivos, y supone que la tasa de aprendizaje, a_t , no es constante¹¹.

4. Datos y resultados

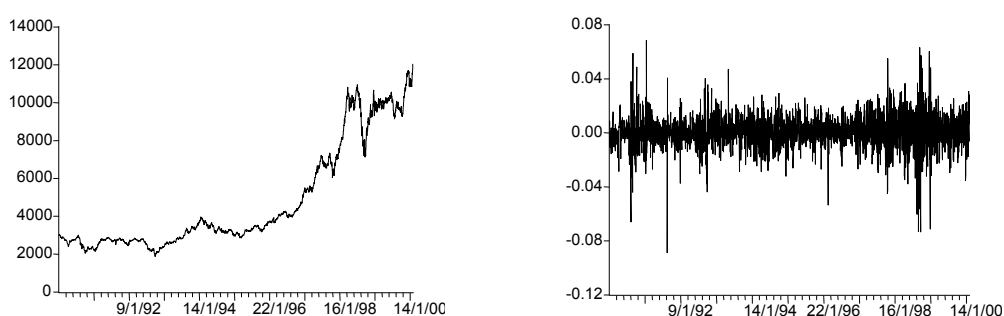
En este estudio utilizamos datos diarios del índice Ibex35 durante el periodo que abarca desde el 29 de diciembre de 1989 hasta el 10 de febrero de 2000, haciendo un total de 2520 observaciones diarias. El índice Ibex35 (I_t) se compone de los 35 valores más líquidos negociados en el sistema continuo o de intermediación bursátil de las

¹¹ Este aspecto hace que los resultados estándar de la teoría estocástica puedan aplicarse para establecer la casi segura convergencia de $\hat{\theta}_t$ hacia θ^* , solución local del problema de mínimos cuadrados [8]. De esta forma, bajo condiciones generales $\sqrt{T}[\hat{\theta}_T - \theta^*]$ converge en distribución, cuando $T \rightarrow \infty$, a una distribución normal multivariante con media cero y matriz de covarianzas estimable consistentemente. La teoría de convergencia de White para estimadores m-recursivos se basa en observaciones iid de x e y , sin embargo, cuando existen procesos heterogéneos y dependientes puede verse Gallant y White (1988). Según Moody y Utans (1992) y Refenes y Zapranis (1999), los requerimientos esenciales para que el estimador RNA posea la propiedad asintótica de consistencia son satisfacer las propiedades de convergencia y unicidad, las cuales suponen que el hessiano no contiene valores propios negativos o que la distribución de las ponderaciones o pesos es normal multivariante.

cuatro bolsas españolas (SIBE), que durante el periodo de control cuentan con el mayor volumen de contratación en pesetas efectivas. Este es un índice compuesto que posee una elevada representatividad y que está ajustado por la capitalización y los dividendos de los activos incluidos, pero no por las ampliaciones de capital. La serie es transformada en logaritmos para calcular los rendimientos continuos, obedeciendo a la siguiente expresión: $r_t = \log\left(\frac{I_t}{I_{t-1}}\right)$, siendo log el logaritmo natural.

En la figura 1 se observa la evolución del índice y sus rendimientos diarios. El crecimiento espectacular del índice a partir de 1996 se debe a la caída de los tipos de interés libres de riesgo y el traslado secuencial de los inversores hacia el mercado de renta variable. El periodo que podemos considerar de interés en la evolución de los precios y rendimientos es la crisis asiática de 27-29 de Octubre de 1997, donde se observa una fuerte caída de las rentabilidades.

Figura 1. Evolución del índice y su rentabilidad continua diaria.



A continuación, se describen las propiedades estadísticas del Ibex35, se estiman los modelos presentados en la sección 2 y se realiza una evaluación de la capacidad predictiva¹².

4.1. *Propiedades estadísticas del Ibex35 durante todo el periodo muestral*

4.1.1. *Contrastes de raíces unitarias y reversión a la media de los rendimientos*

¹² Al igual que León y Mora (2000), no encontramos significativos los efectos del día de la semana por lo que todo el análisis de los resultados se circunscribe a los rendimientos sin filtrar de tales efectos.

El orden de integrabilidad es un determinante crucial de las propiedades que exhibe una serie temporal. Con la finalidad de analizar la no estacionariedad de dicho índice, aplicamos diversos contrastes de raíces unitarias convencionales, que analizan la existencia de integrabilidad en las variables. Concretamente, utilizamos el contraste Dickey y Fuller Aumentado (ADF) y el contraste de Phillips y Perron (PP) que corrige de forma no paramétrica las estimaciones de los errores estándar de los parámetros en el modelo de Dickey y Fuller, abandonando el supuesto de perturbaciones independientes e idénticamente distribuidas. Adicionalmente, y derivado de la baja potencia de estos contrastes, el estudio de la existencia de integrabilidad se complementa con un test cuya hipótesis nula es la estacionariedad. Un contraste de este tipo es el de Kwiatkowski, Phillips, Schmidt y Shin (1992, KPSS). Los resultados de la aplicación de los diversos contrastes de existencia de raíces unitarias aparecen en el cuadro 1.

Cuadro 1. Contrastes de raíces unitarias convencionales.

Series	ADF(p=4)		PP(l=8)		KPSS(l=8)	
	Constante	Tendencia	Constante	Tendencia	η_{μ}	η_{τ}
$\log I_t$	0.7974	-2.2433	-0.8302	-2.2154	24.517	5.0945
r_t	-22.055 ^a	-22.147 ^a	-44.901 ^a	-44.948 ^a	0.4111 ^b	0.0475 ^b
1%	-3.4360	-3.9671	-3.4360	-3.9671	0.739	0.216
5%	-2.8632	-3.4142	-2.8632	-3.4142	0.463	0.146

Nota: En las últimas dos líneas aparecen los valores críticos de los test ADF y PP obtenidos de MacKinnon, y los valores críticos del test KPSS obtenidos de Kwiatkowski, Phillips, Schmidt y Shin (1992). El número de retardos p para el test ADF se elige en base al criterio de información de Akaike, mientras que el retardo de truncamiento que se ha seleccionado - para estimar la varianza de Newey y

West - es igual a la expresión $l = \text{int} \left[4(T/100)^{2/9} \right]$, siendo *int* la parte entera del número que no excede el argumento. En los casos de la rentabilidad del Ibex35, se denota por **a** al rechazo de la hipótesis nula en los test ADF y PP. La letra **b** denota no rechazar la hipótesis nula de estacionariedad en el test KPSS.

En particular, se muestran los resultados del test ADF y PP, quienes contrastan la hipótesis nula de I(1) frente a I(0). En todos los contrastes se distingue la presencia - en la parte determinista de la ecuación - de una constante y de una constante y tendencia determinística. Por otro lado, se ha elegido un número de retardos igual a 4 en el contraste ADF. Sin embargo, para el test PP y KPSS se utiliza como punto de truncamiento el valor 6. Las conclusiones obtenidas muestran que para la serie del índice ibex35, los contrastes ADF y PP no rechazan la hipótesis nula I(1) mientras que en el contraste KPSS se rechaza que las series sean estacionarias. De forma contraria sucede con la serie de rendimientos, donde se observa que esta es I(0).

Puesto que los cambios de estructura pueden tener importantes efectos sobre las implicaciones de los contrastes de raíz unitaria aplicados a series económicas, realizamos el contraste realizamos el contraste de Perron (1997), que introduce la posible ruptura de forma endógena¹³. Así, para el estudio del logaritmo del Ibex35, empleando el contraste de Perron (1997) muestra que existe una raíz unitaria aún cuando existe cambio de estructura en la fecha $T_B=31/10/1996$, con un valor del estadístico $t_{\alpha=1}=-3.7483$, cuyo valor es mayor que el valor crítico tabulado al 5% e igual a -5.08 .

La atención a las propiedades de largo plazo de una serie financiera tiene, también, nuestro interés. Una gran cantidad de estudios sobre la conducta de los mercados de activos obtienen que la predicción de las rentabilidades de los activos, lejos de una conducta impredecible, indican la presencia de autocorrelaciones negativas en un horizonte temporal amplio, lo que se ha denominado como reversión a la media o persistencia. De esta manera, se han desarrollado técnicas para medir y contrastar la reversión a la media, como medida del grado de persistencia en una serie temporal, paralela al estudio de la presencia de raíces unitarias o componentes de camino aleatorio. Cochrane (1988) propuso una medida no paramétrica de la persistencia denominada estadístico de razón de las varianzas (VR), y explota el hecho de que si el logaritmo del precio de los activos, incluyendo dividendos acumulados, sigue un camino aleatorio, la varianza de la rentabilidad sería proporcional al horizonte de rentabilidad. El contraste que utilizamos es la versión de Cochrane (1988), Diebold (1987) y Lo y McKinley (1988). El estadístico utilizado es el siguiente: $VR(q) = \frac{\sigma_b^2(q)}{\sigma_a^2}$,

donde $\sigma_b^2(q) = \frac{1}{T} \sum_{k=q}^T [x_k - x_{k-p} - q\hat{\mu}]^2$, $\hat{\mu} \equiv \frac{1}{T} [x_T - x_0]$, $\sigma_a^2 = \frac{1}{T} \sum_{k=1}^T [x_k - x_{k-1} - \hat{\mu}]^2$, q es un entero natural; T es el tamaño muestral; y x es el logaritmo del índice. Lo y MacKinlay (1988) derivaron la distribución asintótica del estadístico centrado para el caso en el que la hipótesis nula es un paseo aleatorio y las innovaciones son iid normales, bajo los supuestos adicionales donde q es fijo y T tiende a infinito. Estos autores demostraron que la distribución del estadístico es asintóticamente normal, tal

¹³ El test de Perron (1997) es: $y_t = \mu + \beta_1 DU_t + \beta_2 D(T_B)_t + \beta_3 t + \beta_4 DT_t + \alpha y_{t-1} + \sum_{i=1}^p c_i \Delta y_{t-i} + u_t$,

siendo y_t , tanto el logaritmo del índice como la rentabilidad, $DU_t=1$ si $t>T_B$, 0 en caso contrario; $DT_t=1(t>T_B)t$ si $t>T_B$, 0 en caso contrario, $1(\cdot)$ es una función indicador; y $D(T_B)_t=1$ si $t=T_B+1$, y 0 en caso contrario.

que $\sqrt{T}M_q \overset{a}{\sim} N\left(0, \frac{2(2q-1)(q-1)}{3q}\right)$, siendo $M_q \equiv VR(q)-1$. La varianza de dicha

distribución puede corregirse ante la existencia de heterocedasticidad [véase Lo y MacKinlay (1988)]. Bajo la hipótesis nula, el estadístico VR(q) es igual a la unidad. Si VR(q) es mayor que la unidad existe *mean averting* mientras que si VR(q) es menor que la unidad existe reversión a la media para valores de q elevados. Dado que la varianza es conocida, el estadístico de prueba es normal estándar (z), de tal forma que

$$z(q) = \frac{\sqrt{T}M_q}{\sqrt{\sigma^2}}, \sigma^2 = \frac{2(2q-1)(q-1)}{3q}.$$

Los resultados pueden verse en el cuadro 2. Este muestra los valores de q (días) para los que se calcula la razón de varianzas.

Cuadro 2. Valores del estadístico de razón de varianzas para diversos q. T=2520 observaciones.

Valores de q									
	2	5	10	15	30	45	60	90	180
VR(q)	0.9043 (-4.80)	0.8839 (-2.66)	0.8748 (-1.86)	0.8233 (-2.09)	0.7314 (-2.19)	0.7177 (-1.86)	0.7212 (-1.58)	0.7364 (-1.22)	0.8464 (-0.50)

Nota: Entre paréntesis aparece el valor del estadístico z(q). Este resultados se compara con los valores críticos de la distribución N(0,1), que al 5% y 1% son 1.96 y 2.576, respectivamente.

Estos resultados muestran que en el caso del logaritmo del Ibex35, el valor VR(q) cae por debajo de la unidad, lo que implica que las rentabilidades están negativamente autocorrelacionadas en el tiempo y existe reversión a la media para el índice Ibex35. Además, no se rechaza la proximidad a la unidad de dicho valor, empleando el estadístico z(q) y el 95% de nivel de confianza.

4.1.2. Contrastes de linealidad: Exponente de Hurst, BDS estándar y BDS recursivo

En esta sección abordaremos el estudio de la posible existencia de dependencias no lineales en la serie de rendimientos. Un contraste de hipótesis que analiza si una serie está distribuida de forma independiente e idéntica es el exponente de Hurst. Este contraste remueve las dependencias a corto plazo mientras que al mismo tiempo permite detectar dependencias a largo plazo o dinámica fractal¹⁴. Siguiendo a Peters (1994) ,

¹⁴ Otros modelos que permiten analizar la memoria a largo plazo son los modelos ARFIMA o modelos integrados fraccionalmente, que tienen la propiedad de capturar la dependencia de muy largo plazo al ser

después de transformar la serie en rentabilidad logarítmica, $\log r_t$, se preblanquea con un AR(1) para minimizar la influencia lineal y no estacionariedad y, seguidamente se divide la serie en A subperiodos contiguos de longitud n, tal que $A \cdot n = N$ o longitud de la serie. Cada subperiodo se denomina I_a , $a=1,2,\dots,A$, y cada elemento del intervalo se denomina $N_{k,a}$, $k=1,2,\dots,n$. En cada intervalo de longitud n, el valor medio se define por

e_a , siendo: $e_a = \frac{1}{n} \sum_{k=1}^n N_{k,a}$. A continuación se define el rango de cada subperiodo, de tal

forma que: $R_{I_a} = \max(x_{k,a}) - \min(x_{k,a})$, $x_{k,a} = \sum_{i=1}^k (N_{i,a} - e_a)$ siendo $x_{k,a}$ la desviación

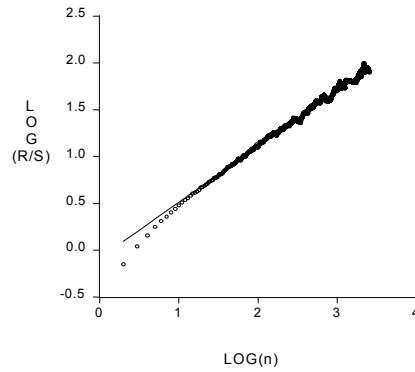
acumulada del valor medio para cada subperiodo. Cada rango se normaliza dividiéndolo

por la desviación estándar muestral, S, de tal forma que: $S_{I_a} = \left[\frac{1}{n} \sum_{k=1}^n (N_{k,a} - e_a)^2 \right]^{-0.5}$.

Así, para cada longitud n, el valor medio R/S será: $\left(\frac{R}{S} \right)_n = \left(\frac{1}{A} \sum_{a=1}^A (R_{I_a} / S_{I_a}) \right)$. El paso

final será realizar la regresión por mínimos cuadrados ordinarios de $\log(R/S)$ sobre $\log(n)$.

Figura 2. Exponente de Hurst. Diagrama *scatter* con línea de regresión.



El coeficiente de la pendiente de la regresión es el exponente de Hurst, H; que está comprendido entre 0 y 1. Un valor de H igual a 0.5 indica que la serie temporal es similar al paseo aleatorio, y por lo tanto los incrementos son independientes, la correlación es cero y no existe memoria a largo plazo; por lo que dicho proceso podría representar cualquier distribución de probabilidad independiente gaussiana o no

procesos estocásticos que se caracterizan por poseer una dependencia fuerte y ser cíclicamente no periódico. Estos aspectos podrían también extenderse a la varianza, por ejemplo, los modelos FIGARCH.

gausiana. Los valores de H que son distintos a 0.5 indican la existencia de dependencia en el proceso y por lo tanto que el proceso no tiene un comportamiento aleatorio. Concretamente, $H > 0.5$ indica persistencia, y por lo tanto que la correlación es positiva; mientras que $H < 0.5$ indica antipersistencia o un sistema con correlación negativa. El rechazo de la hipótesis nula puede apuntar hacia un proceso caótico, un proceso no lineal estocástico o a una dependencia estocástica lineal en la conducta de la serie. La aplicación de esta metodología a los rendimientos del Ibex35 muestra que la estimación de H es 0.5966, con un error estándar de 0.001288. El valor es mayor a 0.5, y estadísticamente bajo la hipótesis de que $H=0.5$, la t-Student es igual a 75.06, valor que claramente rechaza dicha hipótesis. La figura 2 muestra la relación entre $\log(R/S)$ y $\log(n)$, y la recta de regresión ajustada a la nube de puntos.

Otra forma de evaluar la posible no linealidad de las series de los rendimientos es mediante el contraste de Brock, Dechert y Scheinkman (1987), en adelante BDS. Utilizamos el BDS porque es un contraste que posee una potencia relativamente buena para contrastar las no linealidades inducidas, por ejemplo, por la autocorrelación de los segundos momentos así como de otras fuentes. En comparación con otros tests estadísticos de no linealidades – como el contraste del bi-espectro de Hinich – el BDS ofrece la ventaja de no depender de la existencia de momentos elevados. El test BDS está basado en la siguiente reconstrucción del tiempo-demora del espacio de fases, que es topológicamente equivalente al espacio de fases original [Takens (1981)]: $Y = \{Y_t, Y_{t+\tau}, \dots, Y_{t+\tau(m-1)}\}$, $t = 1, 2, \dots, T - (m-1)\tau$, donde, Y es una rentabilidad en t y $\tau \in \mathbb{N}$ es un retardo temporal fijado. Para asegurar la independencia lineal de las variables reconstruidas, este retardo se fija haciéndolo igual al número de retardos en que la primera autocorrelación se hace cero. Este contraste está basado en el hecho de que si $\{Y_t\}$ es un proceso lineal iid, entonces la correlación integral

$$C_{\varepsilon, m} = \frac{2}{T_N(T_N - 1)} \sum_{1 \leq s < t < T} I_{\varepsilon}(Y_t^m, Y_s^m) \text{ donde } \varepsilon \text{ es la distancia relativa a la desviación}$$

estándar de los datos; m es la dimensión de inmersión o historias; $T_N = T - m + 1$ es el número de m-historias que pueden construirse del tamaño muestral T;

$Y_t^m \stackrel{\text{def}}{=} (y_t, y_{t+1}, \dots, y_{t+m-1})$ define a un proceso de m-historias; $I_{\varepsilon}(Y_t^m, Y_s^m)$ es una función indicador igual a 1 si $\|Y_t^m - Y_s^m\| < \varepsilon$, y cero en otro caso, siendo $\|\cdot\|$ la norma máxima.

Brock, Dechert y Scheinkman (1987) mostraron que $V_{\varepsilon, m} = \sqrt{T}(C_{\varepsilon, m} - C_{\varepsilon, 1}^m) / \sigma_{\varepsilon, m}$ tiene

una distribución límite que es normal para todo $\varepsilon > 0$ y $m=2,3, \dots$ bajo la hipótesis nula de proceso lineal iid; $\sigma_{\varepsilon,m}$ es la desviación estándar asintótica de $\sqrt{T}(C_{\varepsilon,m} - C_{\varepsilon,1}^m)$. Nótese que bajo la hipótesis nula de iid, $C_{\varepsilon,m} - C_{\varepsilon,1}^m = 0$, $\forall \varepsilon > 0$ y $m = 2,3,\dots$.

La selección de ε y m es crucial. Para evitar los posibles rechazos de la hipótesis nula debidos a la dependencia lineal, el test BDS o $V_{\varepsilon,m}$, se aplica comúnmente a los residuos estimados del proceso ARMA. La distribución asintótica del BDS no está afectada cuando se aplican filtros lineales a los datos¹⁵.

Cuadro 3. Test $V_{\varepsilon,m}$ aplicado a los residuos del modelo ARIMA(0,1,1). Ecuación [9].

	$\varepsilon=0.5\sigma$	$\varepsilon=0.75\sigma$	$\varepsilon=\sigma$	$\varepsilon=1.25\sigma$	$\varepsilon=1.5\sigma$	$\varepsilon=1.75\sigma$	$\varepsilon=2\sigma$
$V_{\varepsilon,m}$ Residuos							
m=2	7.8404	8.2212	8.8818	9.7587	10.881	11.943	12.950
m=3	11.177	11.474	12.013	12.805	13.811	14.755	15.647
m=4	15.024	14.847	14.978	15.471	16.117	16.745	17.369
m=5	18.654	17.903	17.429	17.483	17.773	18.136	18.565
m=6	24.188	22.037	20.520	19.759	19.420	19.364	19.523
m=7	30.773	26.367	23.708	21.955	10.885	20.343	20.204
m=8	39.572	31.219	27.211	24.280	22.371	21.307	20.843
$V_{\varepsilon,m}$ Residuos “barajados”							
m=2	-0.4842	-0.7235	-1.0446	-1.0478	-1.0067	-0.8652	-0.4988
m=3	-0.3970	-0.0621	-0.9401	-0.9087	-0.8017	-0.6099	-0.2802
m=4	-0.0027	-0.2828	-0.5809	-0.4720	-0.3162	-0.1315	0.1551
m=5	0.1409	-0.0571	-0.2599	-0.1457	0.0006	0.1641	0.3998
m=6	0.0466	0.0494	-0.0091	0.2312	0.4054	0.5754	0.7738
m=7	-0.3157	-0.0593	-0.1203	0.2543	0.5042	0.7394	0.9674
m=8	-0.8812	-0.3359	-0.0064	0.4339	0.7234	0.9813	1.2159

Nota: Se ha utilizado el valor de $\sigma=0.0794$. Los valores críticos de la distribución $N(0,1)$ son 1.645, 1.96 y 2.576 al 10%, 5%, y 1%, respectivamente.

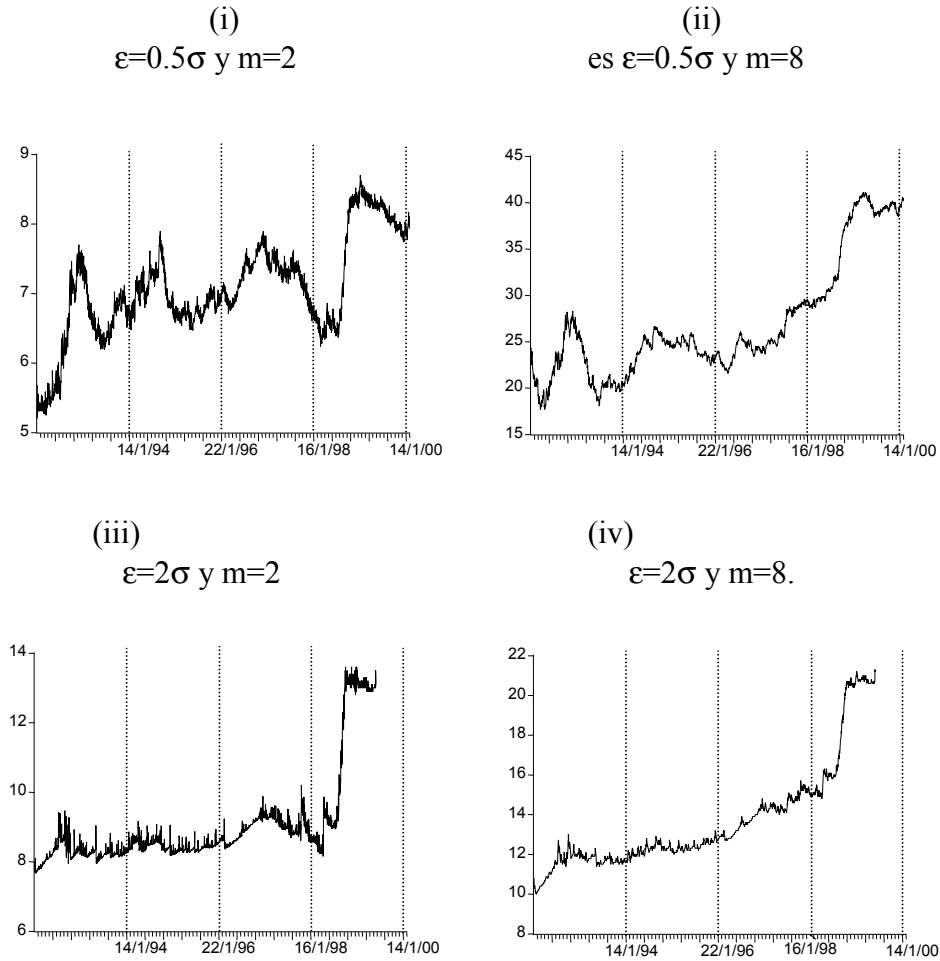
De esta manera, llevando a cabo este procedimiento, un rechazo de la hipótesis nula se podría atribuir a dos factores. Por un lado la dependencia no lineal [determinista (por ejemplo, el caos) o estocástica], o por otro lado, a un proceso no estacionario. Debido a estas razones, y para evitar sobre todo el segundo aspecto, filtramos la serie haciéndola estacionaria primero y luego obteniendo los residuos del proceso ARMA estacionario e invertible. En este sentido, modelizamos el logaritmo del índice Ibex35 como un proceso ARIMA(p,1,q). Concretamente, el modelo estimado es igual a:

¹⁵ Estudios de simulación han mostrado que el test tiene potencia ante una extensa variedad de alternativas, incluyendo los modelos no lineales ARCH [véase Bollerslev, Engle y Nelson (1994)]. Sin embargo, como han apuntado Granger y Teräsvirta (1993), el hecho de que no se utilice un contraste tipo

$$\begin{array}{lcl} (1-L)\log I_t = (1+0.1154L)u_t, & \sigma = 0.00055, & R^2 = 0.02 \\ t: & (5.83) & (1.98) \end{array} \quad [9]$$

donde los órdenes de los polinomios de retardos media móvil son $p=0$ y $q=1$. En el cuadro 3 se resumen los resultados del test BDS aplicados a los residuos del modelo [9]. También aparecen los residuos “barajados”, es decir, recreados como si fueran datos “muestreados” aleatoriamente sin reemplazamiento. En este trabajo, utilizamos $m=2,3,4,5,6,7,8$ y un ε comprendido entre $[0.5\sigma, 0.75\sigma, \sigma, 1.25\sigma, 1.5\sigma, 1.75\sigma, 2\sigma]$, para el valor de σ correspondiente a $\sigma=0.1088$.

Figura 3. Evolución del valor del estadístico $V_{\varepsilon,m}$ recursivo. Residuos ARIMA(0,1,1).



Estas series son completamente aleatorias pero tienen la propiedad de que su distribución es igual a la original. Utilizamos esta técnica siguiendo a Scheinkman y LeBaron (1989) con la finalidad de reforzar los resultados, de tal manera que en este

multiplicador de Lagrange dificulta la determinación de qué tipo de modelo alternativo podría considerarse cuando se rechazase la hipótesis nula.

caso deberíamos no rechazar la hipótesis nula de proceso lineal iid. Así, podremos evidenciar que existe alguna estructura no lineal en los datos originales que ha sido removida por el “barajamiento”. Los resultados muestran que existen estructuras no lineales en los datos del logaritmo del índice Ibex35, pues tanto el test aplicado a los residuos como a los residuos “barajados” muestran el rechazo de la hipótesis nula en el primero de los casos y el no rechazo de la hipótesis nula en el segundo de los casos.

Finalmente, también ilustramos el test $V_{\varepsilon,m}$ de forma recursiva para determinados ε y m , con la finalidad de comprobar si el test es sensible a la posible existencia de inestabilidades puntuales o cambios de régimen durante el periodo, en una línea similar a De Lima (1998). Este autor desarrolla una metodología de contraste que intenta discriminar entre los rechazos de la hipótesis nula de linealidad debido a la no linealidad intrínseca y rechazos que son debidos a la no estacionariedad de los datos. Se basa en el Teorema Funcional Central de Límite (TFCL), y demuestra que las sumas parciales del test estadístico $V_{\varepsilon,m}$ convergen a un movimiento browniano. Este test - que es una generalización de BDS - es sensible a los cambios en la varianza y a los cambios en la distribución de los datos. La figura 3 muestra la evolución del estadístico $V_{\varepsilon,m}$ recursivo, comenzando las recursiones a partir de la fecha 14 de enero de 1994 y dejando las quinientas primeras observaciones como valores iniciales para la realización de las recursiones. Se observa cómo los valores de $V_{\varepsilon,m}$ son claramente no estacionarios, pues crecen a lo largo del tiempo, rechazando aún más la hipótesis nula de linealidad. Además, se observa que durante el año 1997, y aproximadamente sobre el 27 de octubre una caída del valor de dicho estadístico, y sobre agosto - septiembre del año 1998 una ruptura estructural importante debido a variaciones negativas del índice con caídas del 21.24% respecto a julio de ese mismo año, lo que todavía acrecienta el alejamiento y mayor rechazo de la hipótesis nula de linealidad. Por lo tanto, una importante conclusión sería que el valor del estadístico BDS varía en el tiempo, y no modifica los resultados que se han calculado y aparecen en el cuadro 3, salvo para los primeros periodos donde no se rechazaría dicha hipótesis.

4.2. Estimación de algunos modelos no lineales

Los contrastes de no estacionariedad muestran la presencia de componentes similares al camino aleatorio, así como la existencia de autocorrelaciones negativas o

reversión a la media. Pero, si los contrastes de linealidad indican la presencia de componentes no lineales y la teoría económica no sugiere un modelo completamente especificado, entonces la estructura del modelo tiene que ser especificada por los datos. A continuación exponemos los resultados muestrales de algunos de los modelos no lineales estimados, que podemos considerar representan ciertos hechos estilizados de la dinámica a corto plazo de los rendimientos de los activos: los modelos de volatilidad condicional y los modelos de cambios de régimen. Adicionalmente, analizamos distintos esquemas RNA con diferentes capas de nodos.

4.2.1. Modelos de volatilidad y cambio de régimen

El estudio de la variación temporal de la volatilidad del mercado de activos español es importante, está bien documentada y extensamente reconocida [véase Alonso (1995) y León y Mora (1999), por ejemplo].

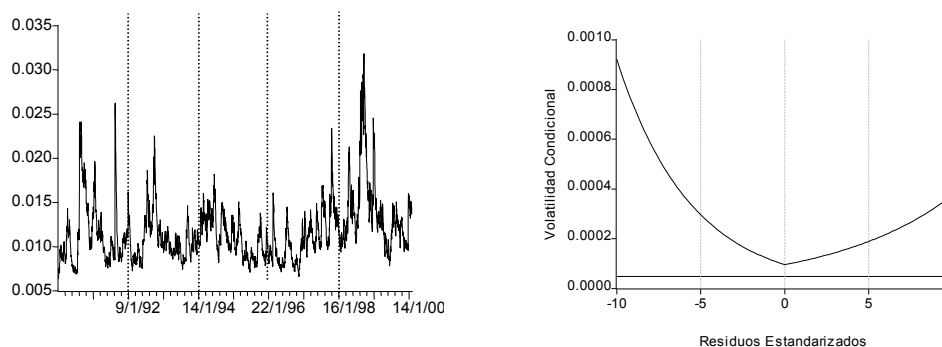
Cuadro 4. Estimación máximo verosímil. Modelos no lineales para las rentabilidades.

	Modelos ARCH				Modelos STAR			Modelos Markovianos	
	GARCH	GJR	EGARCH		ESTAR	LSTAR		Hamilton	TSN
μ	0.00073 (3.01)	0.00055 (2.15)	0.00061 (2.45)	β_{10}	0.026 (0.21)	0.040 (1.52)	ϕ_1	0.1132 [0.027]	
θ_1	0.1388 (5.72)	0.1389 (5.78)	0.1457 (5.95)	β_{11}	0.399 (3.84)	0.113 (8.65)	ϕ_2	-0.026 [0.098]	
ω	0.000004 (3.64)	0.000004 (3.82)	-0.4258 (-5.17)	β_{12}	-0.105 (-1.30)	-0.065 (-4.28)	μ		0.0007 [0.0001]
α	0.1028 (6.83)	0.0587 (3.90)	0.1817 (7.31)	β_{20}	0.026 (0.20)	0.271 (2.03)	μ_1	-5.2087 [0.388]	
β	0.8672 (47.78)	0.8789 (55.02)	0.9687 (120.3)	β_{21}	-0.303 (-2.91)	0.023 (1.31)	μ_2	0.1008 [0.057]	
γ	--	0.0672 (2.98)	-0.0450 (-2.91)	β_{22}	0.066 (1.80)	0.326 (3.58)	p_{11}		0.987 [0.004]
				γ	22.28 (1.07)	5.482 (1.17)	p_{12}	0.9933 [0.002]	0.013 [0.003]
				c	1.294 (29.9)	2.362 (10.4)	p_{21}		0.0593 [0.018]
							p_{22}	0.2272 [0.152]	0.9407 [0.021]
							σ_1^2	1.153 [0.02]	0.0005 [0.001]
							σ_2^2		0.0053 [0.001]
LogL	-7884.3	-7863.2	-7862.4	LogL	-1812.4	-1807.1	LogL	-4028.8	-7711.5

Nota1: Los modelos de heterocedasticidad condicional son GARCH(1,1), EGARCH(1,1) y GJR(1,1). Entre paréntesis aparecen los valores t-Student robustos ante la existencia de heterocedasticidad condicional, utilizando el procedimiento de Bollerslev y Wooldridge. El criterio de convergencia utilizado es 1×10^{-4} y el procedimiento de optimización de Berndt, Hall, Hall y Hausman. **Nota2:** β_{10} y β_{20} representan las constantes de la ecuación mientras que β_{11} , β_{12} , β_{21} y β_{22} representan los coeficientes que acompañan a x_{t-1} y x_{t-2} . Entre paréntesis aparecen los valores t-Student bajo la hipótesis nula de parámetro igual a cero. Entre corchetes aparecen los errores estándar de la estimación. Hamilton es el modelo de Hamilton estimado como AR(2) y cambios de régimen en la media.

Sin realizar un exhaustivo ejercicio de selección y comparación de modelos ARCH - tal y como puede verse en León y Mora (1999) para la serie de rendimientos del Ibex35 - hemos estimado los modelos [1], [2] y [3]. El cuadro 4 resume los resultados de la estimación de los diferentes modelos no lineales: ARCH, STAR y cambios de régimen markovianos. Todos los modelos que aparecen en el cuadro son estimados por máxima verosimilitud, suponiendo normalidad del error. Con respecto a los modelos de tipo ARCH - donde la estimación se realizó utilizando el algoritmo de Berndt, Hall, Hall y Hausman (BHHH) y el procedimiento de Bollerslev y Wooldridge que permite estimar de forma robusta los errores estándar de los parámetros - observamos, en general, una estructura regular casi nula en la media, un *clustering* importante y el denominado efecto asimetría. En términos del logaritmo de verosimilitud, la estimación que parece aceptable estadísticamente es la perteneciente al modelo MA(1)-EGARCH(1,1), cuya volatilidad condicional y curva de impacto de las noticias [según Engle y Ng (1993)] pueden verse en la figura 3. El parámetro que mide la persistencia de la volatilidad es igual a 0.9687, mientras que el efecto que mide el efecto “apalancamiento” o asimetría es igual a -0.045 . En este caso, dado que es estrictamente negativo, los *shocks* positivos de la rentabilidad generan menor volatilidad que los *shocks* negativos, manteniéndose constante todo lo demás.

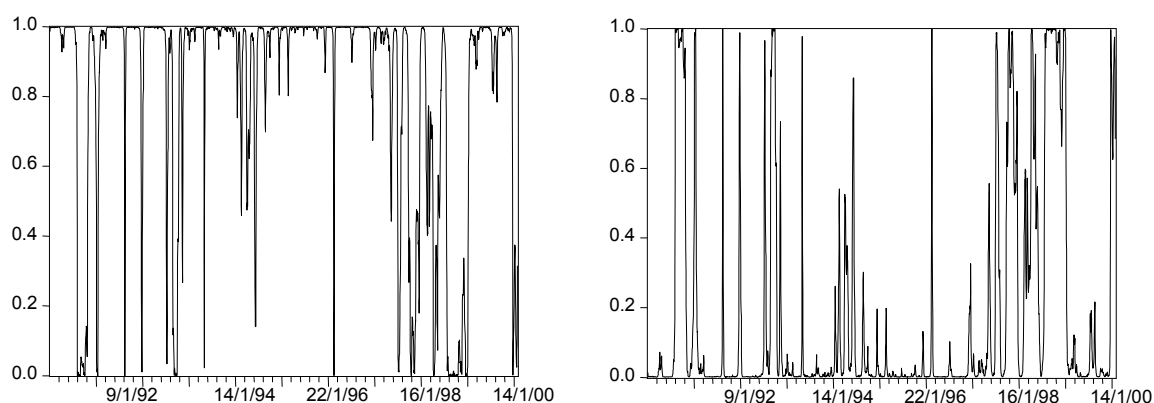
Figura 4. Desviación estándar condicional del proceso EGARCH(1,1) y curva de impacto de las noticias.



En otras palabras, en periodos volátiles existe más incertidumbre sobre la próxima rentabilidad que la que existe en periodos menos volátiles, y el efecto de los *shocks* negativos es mayor que el de los positivos de igual tamaño. Por lo tanto, podría advertirse que dado que existen asimetrías, tenemos una interpretación para sean las rentabilidades “sorprendentemente” bajas (negativas) – y no las “sorprendentemente” altas (positivas) – quienes generan una mayor volatilidad.

En segundo término, hemos estimado distintos modelos que reflejan los cambios de régimen, por ejemplo, la ecuación [4] con función de tipo ESTAR y LSTAR, y el modelo de cambios de régimen de Hamilton así como el modelo de volatilidad cambiante markoviana (TSN)⁽¹⁶⁾. Por un lado, observamos que las estimaciones máximo verosímiles de los parámetros de los modelos de tipo STAR de dos regímenes son parecidas. En ambos, el orden de la parte autorregresiva es $p=2$, mientras que el retardo $d=3$. El parámetro de cambio de régimen o umbral (c) es significativo en las dos ecuaciones estimadas. No obstante, el parámetro de alisamiento (γ) ha sido estimado de forma muy imprecisa por su elevada varianza, aspecto que Teräsvirta (1994) reconoce sucede con cierta frecuencia. En general, los resultados muestran que ambas estimaciones son factibles, si bien LSTAR(2) parece mejor que ESTAR(2) estadísticamente, en términos del logaritmo de verosimilitud.

Figura 5. Evolución de las probabilidades de los regímenes de baja y elevada volatilidad en el modelo TSN. Periodo Completo.



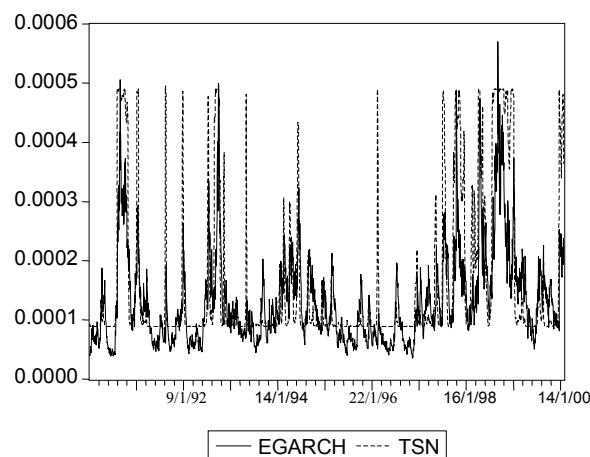
Por otro lado, y en cuanto a los modelos de cambio de régimen markovianos, las estimaciones máximo verosímiles muestran la significación de dos regímenes, tanto

¹⁶ Se han estimado modelos STAR suponiendo heterocedasticidad condicional, sin embargo, los resultados empeoran a los obtenidos sin considerar tal supuesto. También, se han estimado otras especificaciones de modelos markovianos, tanto para los cambios de régimen en la volatilidad como en la media, encontrándose no significativos.

para el modelo de Hamilton como para TSN. En el caso del modelo Hamilton, los cambios en la media son significativos para la probabilidad de paso del régimen 1 al régimen 1; $p_{11}=0.99$, mientras que es menos significativa para el paso del régimen 2 al 2, $p_{22}=0.22$. Quizás, este modelo es menos adecuado para representar la conducta de las rentabilidades.

En el caso del modelo TSN, las probabilidades de transición indican que el estado de baja volatilidad domina al estado de alta volatilidad; y además, existe una mayor probabilidad en la transición del régimen de baja volatilidad a otro de baja volatilidad [$p_{11}=0.987$] y también en el paso de un régimen de alta volatilidad a otro de alta volatilidad [$p_{22}=0.9407$], siendo menores en las transiciones cruzadas de los estados. La figura 5 dibuja las probabilidades de cada régimen o estado. Concretamente, el primer régimen caracteriza a un estado de baja volatilidad mientras que el segundo estado caracteriza a un régimen de alta volatilidad. Como puede observarse, el régimen de baja volatilidad de las rentabilidades predomina frente al régimen de elevada volatilidad para las mismas.

Figura 6. Evolución de las volatilidades estimadas EGARCH y TSN.



A continuación, vamos a comparar las volatilidades EGARCH y las obtenidas por el modelo TSN. La figura 6 muestra una comparación de las volatilidades heterocedásticas de tipo EGARCH frente a las estimadas mediante un proceso markoviano con dos estados (TSN). En dicha figura puede observarse los cambios abruptos en TSN frente a la volatilidad EGARCH. Siguiendo a Bollerslev, Engle y Nelson (1994), ya que un modelo ARCH asume que los residuos estandarizados z_t son iid, parece razonable utilizar el BDS como test de especificación, aplicándolo a tales

residuos. No obstante, aunque el contraste BDS tiene una considerable potencia para detectar la mala especificación de los modelos ARCH; su distribución asintótica está fuertemente afectada por el ajuste del modelo ARCH. Brock, Hsieh y LeBaron (1991) dicen que la aplicación del test BDS sobre z_t puede rechazar la hipótesis nula. El cuadro 5 muestra los resultados de dicho contraste, con la finalidad de comprobar la independencia de tales residuos. Los resultados indican que, en el caso del modelo EGARCH, siempre se rechaza la hipótesis alternativa de no linealidad. Por lo tanto, puede considerarse que dichos residuos son iid y el modelo está correctamente especificado. Sin embargo, en el caso del modelo TSN, no puede decirse lo mismo para $m=4$ en todos los casos y al 10%, y para todo los valores de $m=8$ a cualquier nivel de significación. No obstante, el contraste de Jarque y Bera indica que existe normalidad para ambos residuos estandarizados, puesto que el valor del estadístico es igual a 1.57 para el EGARCH [p-valor igual a 0.46] y 5.92 [p-valor=0.0517] para el TSN.

Cuadro 5. Test BDS aplicado a los residuos estandarizados del modelo EGARCH y TSN. Periodo completo.

	$\varepsilon=0.5\sigma$	$\varepsilon=0.75\sigma$	$\varepsilon=\sigma$	$\varepsilon=1.25\sigma$	$\varepsilon=1.5\sigma$	$\varepsilon=1.75\sigma$	$\varepsilon=2\sigma$
Residuos estandarizados TSN							
m=2	1.5485	0.8290	0.8010	0.6480	0.5584	0.5070	0.6596
m=4	4.8630	3.8829	3.3207	2.8105	2.4264	2.2143	2.1097
m=8	11.520	7.8400	2.8105	5.5970	4.9600	4.5260	4.1797
Residuos estandarizados EGARCH							
m=2	-0.5623	-1.1798	-1.1598	-1.1332	-1.0466	-0.9100	-0.6830
m=4	0.4769	0.2389	0.2969	0.1397	0.2638	0.2257	0.3478
m=8	2.9735	1.3936	1.3880	1.0182	1.1078	0.9272	0.8729

Nota: Se ha utilizado el valor de $\sigma=0.0981$ para los residuos estandarizados pertenecientes al EGARCH mientras que $\sigma=0.1310$ para los del modelo TSN. Los valores críticos de la distribución $N(0,1)$ son 1.645, 1.96 y 2.576 al 10%, 5%, y 1%, respectivamente.

4.2.2. Modelos de redes neuronales artificiales

Finalmente, las estimaciones de los esquemas de red atienden a los siguientes aspectos. Primero, se han escalado los inputs suponiendo una distribución uniforme entre $[-1;1]$. Segundo, los modelos estimados son del tipo $RNA(p;q_1,q_2,q_3)$ siendo p la parte autoterregresiva y q_i el número de elementos de la i -ésima capa oculta, $i=1,2,3$. Tercero, la elección del número de neuronas en la capa oculta se determina mediante una de las fórmulas habituales de cálculo: $1/2(k+i)+\sqrt{T}$, siendo k el número de inputs, i el número de outputs y T el tamaño de la muestra. Así, por ejemplo: MLP1 o MLP(1;44) posee una sola capa oculta; MLP (1;22,22) posee dos capas ocultas; y MLP

(1;15,15,15) posee tres capas ocultas; Elman (1;44) realiza la recurrencia en la capa oculta; Jordan (1;44) realiza la recurrencia en la capa de outputs; JCN1 es MLP con conexiones directas entre output-inputs, y posee una sola capa oculta, JCN(1;44); JCN2 es MLP con conexiones directas entre output-inputs y dos capas ocultas, JCN(1;22,22); y finalmente MLP con conexiones directas entre output-inputs, que posee tres capas ocultas, JCN(1;15,15,15). Cuarto, la fase de aprendizaje de la RNA se realiza con 1855 observaciones, mientras que 463 observaciones se utilizan en la fase de test. Ambos tamaños se han obtenido al azar. Las 200 observaciones finales se han reservado para hacer predicciones. Quinto, se ha comprobado el decrecimiento del error de la fase de aprendizaje y test, verificando que en 4000 iteraciones no existía movimiento de dichos errores. Este aspecto se tiene en cuenta para evitar los problemas de sobreajuste, ya que después de un cierto número de ciclos de entrenamiento, el error de test podría comenzar a incrementarse, mientras el error de aprendizaje continuase decreciendo.

Cuadro 6. Estadísticos de los modelos RNA en las fases de aprendizaje y test.

Modelos RNA								
	MLP1	MLP2	MLP3	Elman	Jordan	JCN1	JCN2	JCN3
EAM1	0.9022	0.9025	0.9039	0.9016	0.9041	0.9022	0.9045	0.9197
EAM2	0.8839	0.8786	0.8813	0.8888	0.8858	0.8838	0.8790	0.8866

Nota: EAM1 es el error absoluto medio de la fase de aprendizaje, y EAM2 es el correspondiente a la fase de test.

A título ilustrativo, el cuadro 6 muestra los errores cometidos en la fase de aprendizaje y test en la última iteración. Puede comprobarse como los errores son parecidos en todos los esquemas de red planteados, si bien en la fase de aprendizaje estos son menores en la RNA de tipo Elman, mientras que en la fase de test en la RNA de tipo MLP2.

4.3. Evaluación estadística de las predicciones extramuestrales sobre los rendimientos

Una herramienta útil para valorar la adecuación del modelo no lineal de series temporales es la evaluación de la capacidad predictiva de los mismos¹⁷. La práctica de

¹⁷ Según Swanson y White (1995), la selección de modelos es un complemento a las versiones tradicionales de predicción porque permite analizar el cumplimiento de la predicción extramuestral y no requiere de la especificación de un modelo correcto para su aplicación, como en el contraste de hipótesis tradicional; y porque la probabilidad de elegir el mejor modelo se aproxima a uno a medida que el tamaño muestral (T) se incrementa, en contraste con la práctica tradicional de fijar un tamaño del test y rechazar la hipótesis nula para un T dado. Una limitación, sin embargo, es que puede dificultar la valoración del

la predicción es fundamental en el análisis económico, y su importancia se justifica en ocasiones por el éxito de la misma. Como consecuencia, la evolución detallada de las predicciones económicas es crucial. Hemos visto que las rentabilidades pueden ser predecibles, y además, pueden ser representadas estadística y económicamente mediante diferentes tipos de modelos no lineales. En este epígrafe realizaremos una comparación competitiva de las predicciones de algunos de los modelos lineales y no lineales que pueden utilizarse para la predicción de los rendimientos. El periodo de predicción elegido al azar corresponde a los últimos 200 periodos de la muestra. Dicho periodo abarca desde el 3 de mayo de 1999 hasta el 10 de febrero de 2000. Las predicciones se obtienen con horizonte de un periodo (*one-step-ahead*).

Para evaluar la capacidad predictiva de los distintos modelos empleamos varios estadísticos de la bondad de la predicción. Las medidas de bondad que utilizamos en este trabajo se basan en $h=1, \dots, H$ periodos de predicción para r_t , que denominaremos por $\hat{r}_t$. En primer lugar, los estadísticos del error absoluto medio: $EAM = \frac{1}{H} \sum_{h=1}^H |r_h - \hat{r}_h|$;

el error porcentual absoluto medio: $EPAM = \frac{1}{H} \sum_{h=1}^H |r_h - \hat{r}_h| / r_h \times 100$; la raíz cuadrada del

error cuadrático medio: $RECM = \sqrt{\frac{1}{H} \sum_{h=1}^H (r_h - \hat{r}_h)^2}$; el estadístico:

$U - Theil = RECM / \left[\sqrt{\frac{1}{H} \sum_{h=1}^H r_h^2} + \sqrt{\frac{1}{H} \sum_{h=1}^H \hat{r}_h^2} \right]$. Por otro lado, y como contrastes de

hipótesis destacamos los siguientes. Los contrastes de predicción direccional de los cambios o de independencia de las predicciones y de los valores actuales de los precios. Construyendo la siguiente tabla de contingencia, denominada *confusion matrix*:

$$\begin{array}{c} r_h \text{ predicha} \\ \geq 0 < 0 \\ r_h \text{ real} \begin{array}{c} \geq 0 \\ < 0 \end{array} \begin{bmatrix} n_{00} & n_{01} \\ n_{10} & n_{11} \end{bmatrix} \end{array}, \text{ donde } n_{ij} \text{ son las frecuencias absolutas pueden elaborarse dos}$$

contrastos. El primero es el estadístico denominado *confusion rate* $= (n_{01} + n_{10}) / H$, que se distribuye como hipergeométrica bajo la hipótesis nula de independencia. El segundo es

un contraste de independencia de tipo χ^2 . Así, $\chi^2 = \sum_{i=0}^1 \sum_{j=0}^1 [n_{ij} - n_{i.} n_{.j} / H]^2 / [n_{i.} n_{.j} / H]$,

error de Tipo I asociado con el contraste cuando no existe distinción explícita entre dos modelos en términos predictivos.

siendo n_i y n_j las frecuencias marginales. Rechazar la hipótesis nula significa que existe una asociación entre los signos de las predicciones y los valores reales. Por último, podríamos destacar otro contraste de predicción direccional como es el de Pesaran y Timmermann (DA, 1992). Bajo la hipótesis nula los valores reales y predichos son independientes. En este caso, la distribución del estadístico es $N(0,1)$. El procedimiento de contraste es el siguiente: $DA = [\text{var}(SR) - \text{var}(SRI)]^{0.5}(SR - SRI)$, donde

$$SR = H^{-1} \sum_{h=1}^H I_i[y_h \cdot \hat{y}_h > 0], \quad SRI = p\hat{p} + (1-p)(1-\hat{p}), \quad \text{siendo } SRI \text{ es el ratio de éxito en el caso de independencia entre los valores reales y los predichos bajo la hipótesis nula. El resto de elementos son iguales a: } p = H^{-1} \sum_{h=1}^H I_i[y_h > 0], \hat{p} = H^{-1} \sum_{h=1}^H I_i[\hat{y}_h > 0],$$

$$\text{var}(SR) = H^{-1} [SRI(1 - SRI)] \quad y$$

$$\text{var}(SRI) = H^{-2} [H(2\hat{p} - 1)^2 p(1 - p) + (2p - 1)^2 \hat{p}(1 - \hat{p}) + 4p\hat{p}(1 - p)(1 - \hat{p})].$$

Otro tipo de contraste es el de Chong y Hendry (1986). Este explora la posibilidad de abarcamiento de las predicciones. Muy sintéticamente, la ecuación a estimar es: $e_{ih} = \lambda(e_{ih} - e_{rwh}) + u_h$, donde e_i es el error de predicción del modelo i , y e_{rw} es el error de predicción del camino aleatorio. La hipótesis nula que se contrasta es $\lambda=0$, y si no se rechaza entonces la predicción del modelo i -ésimo abarca o es condicionalmente más eficiente que la predicción del paseo aleatorio. Por último, también utilizaremos el contraste de los ECM de predicción de Diebold y Mariano (1995). Estos autores proponen un método para contrastar la igualdad de predictores competitivos. Dados dos predictores *h-step ahead*, denotando por e_{1h} y e_{2h} a los correspondientes errores de predicción, entonces se contrasta la hipótesis nula, donde $E[d_h] = 0$, $d_h = f(e_{1h}, e_{2h})$, siendo d_h una función de los errores de predicción. Estos autores se concentran en el caso $f(\cdot)$ es del tipo $d_h = g(e_{1h}) - g(e_{2h})$. De esta forma, no rechazar la hipótesis nula supondría que ambos predictores tuviesen la misma hipótesis nula. El test estadístico está basado en la media muestral $\bar{d} : S = [\hat{V}(\bar{d})]^{-1/2} \bar{d}$, donde $\hat{V}(\bar{d})$ es un estimador consistente de la varianza de la distribución muestral de $\bar{d}$. El estadístico tiene una distribución asintótica normal estándar bajo la hipótesis nula. El estimador consistente de es igual a:

$$\hat{V}(\bar{d}) = 2\pi \hat{f}_d(0) = \sum_{\tau=-(k+1)}^{k-1} \hat{\gamma}_d(\tau), \quad \hat{\gamma}_d(\tau) = H^{-1} \sum_{h=|\tau|+1}^H (d_h - \bar{d})(d_{h-|\tau|} - \bar{d})$$

La comparación de las predicciones se realiza entre los resultados obtenidos de los modelos ARIMA(0,1,1), ARIMA(1,1)-GARCH(1,1), ARIMA(1,1)-TARCH(1,1), ARIMA(1,1)-EGARCH(1,1), LSTAR(2;3), ESTAR(2;3), MLP1, MLP2, MLP3, Elman, Jordan, JCN1, JCN2 y JCN3; estrictamente para el periodo de predicción, y sin incluir a los modelos markovianos porque su predicción es peor. Los resultados de los estadísticos de bondad de ajuste, predicción direccional y abarcamiento de las predicciones se resumen en el cuadro 7. Cabe resaltar, que los modelos RNA estimados predicen ligeramente mejor que el resto de modelos no lineales y lineales considerados. Así, por ejemplo, en términos de EAM (MLP1), EPAM (JCN1), RECM (MLP1, JCN1), U-Theil (JCN3), CR (JCN1), χ^2 (JCN2), DA (Elman), CH (JCN1). En el caso concreto del cociente CR, este muestra el porcentaje de predicciones incorrectas, el cual es menor en el modelo JCN1. En lo que respecta al test χ^2 , el menor valor de dicho estadístico estimado es el correspondiente al modelo JCN2. En cuanto al test de Chong y Hendry (CH), éste se obtiene comparándolo con la predicción del camino aleatorio. Los resultados concluyen en casi todos los casos, mediante el p-valor - excepto en el caso del modelo GARCH y GJR - que la predicción del modelo es condicionalmente más eficiente respecto al paseo aleatorio. No se ha obtenido una matriz de dichos estadísticos al considerarse el camino aleatorio como única referencia cruzada.

Finalmente, en el cuadro 8 aparecen los resultados del contraste de Diebold y Mariano para el supuesto en que $d_t = |e_{1t}| - |e_{2t}|$. También, se ha considerado el caso en que $d_t = e_{1t}^2 - e_{2t}^2$, pero encontramos resultados similares a los presentados en el cuadro 8, por lo cual se han omitido. A partir de los resultados de dicho estadístico, y teniendo en cuenta su valor, podemos decir que para cada par de predicciones casi siempre se rechaza la hipótesis nula de diferencias iguales en los ECM de predicción, salvo en algunos modelos donde la naturaleza de tales especificaciones es similar: por ejemplo modelos de tipo ARCH frente a modelos de tipo ARCH, o modelos de tipo RNA frente a modelos de tipos RNA. Por ejemplo, el modelo ARIMA, GARCH, GJR contra EGARCH, presenta resultados iguales a -1.026, 1.632, y 1.46, respectivamente. No obstante, este aspecto no se cumple en todos los casos. En cuanto al signo, podemos decir que en muchos casos pondera más el error del error del modelo considerado como e_{2h} que el considerado como e_{1h} , por cuanto el signo de d_h es negativo.

Cuadro 7. Estadísticos de bondad de la capacidad predictiva para los diferentes modelos considerados en el estudio. H=200.

	ARIMA	GARCH	GJR	EGARCH	ESTAR	LSTAR	MLP1	MLP2	MLP3	Elman	Jordan	JCN1	JCN2	JCN3
EAM	0.00905	0.01360	0.01362	0.009077	0.00910	0.009041	0.00904 ^a	0.009156	0.009071	0.009086	0.009085	0.009041	0.009173	0.009276
EPAM	111.76	378.05	378.57	116.69	116.23	115.11	109.61	123.10	111.43	120.26	120.08	109.55 ^a	124.94	143.66
RECM	0.011367	0.01678	0.01679	0.011388	0.01142	0.011345	0.01136 ^a	0.01147	0.00405	0.01138	0.011409	0.01136 ^a	0.011513	0.011655
U-Theil	0.88440	0.8889	0.8889	0.865232	0.8652	0.8781	0.90053	0.876527	0.885686	0.862205	0.857665	0.900121	0.860018	0.80597 ^a
CM	$\begin{bmatrix} 73 & 38 \\ 60 & 29 \end{bmatrix}$	$\begin{bmatrix} 62 & 49 \\ 44 & 45 \end{bmatrix}$	$\begin{bmatrix} 61 & 50 \\ 44 & 45 \end{bmatrix}$	$\begin{bmatrix} 72 & 39 \\ 60 & 29 \end{bmatrix}$	$\begin{bmatrix} 69 & 42 \\ 57 & 32 \end{bmatrix}$	$\begin{bmatrix} 68 & 43 \\ 54 & 35 \end{bmatrix}$	$\begin{bmatrix} 76 & 35 \\ 65 & 24 \end{bmatrix}$	$\begin{bmatrix} 69 & 42 \\ 55 & 34 \end{bmatrix}$	$\begin{bmatrix} 74 & 37 \\ 58 & 31 \end{bmatrix}$	$\begin{bmatrix} 73 & 38 \\ 64 & 25 \end{bmatrix}$	$\begin{bmatrix} 77 & 34 \\ 63 & 25 \end{bmatrix}$	$\begin{bmatrix} 76 & 35 \\ 55 & 34 \end{bmatrix}$	$\begin{bmatrix} 69 & 42 \\ 55 & 34 \end{bmatrix}$	$\begin{bmatrix} 79 & 32 \\ 67 & 22 \end{bmatrix}$
CR	0.49	0.615	0.47	0.495	0.495	0.485	0.5	0.485	0.475	0.51	0.485	0.45 ^a	0.485	0.495
χ^2	0.0603	2.7745	0.6028	0.1432	0.0751	0.0072	0.495	0.0028	0.0494	0.8642	0.1161	0.9727	0.00278 ^a	0.4233
DA	-0.3472	0.7708	0.6420	-0.4813	-0.3837	-0.0292	-0.7968	-0.0587	0.1203	-1.0270 ^a	-0.4843	-0.4488	-0.059	-0.7382
CH	0.5972 [0.43]	235.11 [0.00]	235.81 [0.00]	1.3263 [0.24]	1.7370 [0.18]	0.4370 [0.51]	0.2753 [0.60]	3.3140 [0.07]	1.3073 [0.25]	1.2531 [0.27]	1.9409 [0.16]	0.2737 ^a [0.60]	4.7156 [0.03]	10.114 [0.00]

Nota: El superíndice a indica el mejor de los resultados para cada estadístico. CM es el estadístico confusion matrix y CR es confusion rate. En el estadístico CH aparece el valor del contraste de Wald p-valor entre corchetes para la hipótesis de que el parámetro es estadísticamente no significativo.

Cuadro 8. Contraste de Diebold y Mariano. H=200.

$e_1 \backslash e_2$	ARIMA	GARCH	GJR	EGARCH	ESTAR	LSTAR	MLP1	MLP2	MLP3	Elman	Jordan	JCN1	JCN2	JCN3
ARIMA	--	-2.2073 ^b	-2.1499 ^b	-1.0269	-0.9257	0.9889	-2.3383 ^b	-2.3365 ^b	-2.3378 ^b	-2.3377 ^b	-2.3378 ^b	-0.2991	-2.3362 ^b	-2.3348 ^b
GARCH		--	1.0041	1.6207	1.8530 ^a	2.6788 ^c	-7.9017 ^c	-8.2048 ^c	-8.1012 ^c	-7.7576 ^c	-8.2222 ^c	1.8246 ^a	-8.3806 ^c	-8.5768 ^c
TARCH			--	1.4642	1.6573 ^a	2.5949 ^c	-1.6492 ^a	-1.6345	-1.6456 ^a	-1.6299	-1.6412	1.7091 ^a	-1.6326	-1.6179
EGARCH				--	-0.0654	1.5889	-1.7136 ^a	-1.7112 ^a	-1.7130 ^a	-1.7127 ^a	-1.7127 ^a	0.4877	-1.7108 ^a	-1.7086 ^a
ESTAR					--	1.6795 ^a	-1.9562 ^a	-1.9542 ^a	-1.9557 ^a	-1.9555 ^a	-1.9555 ^a	0.5656	-1.9537 ^a	-1.9515 ^a
LSTAR						--	-2.7284 ^c	-2.7272 ^c	-2.7280 ^c	-2.7279 ^c	-2.7279 ^b	-1.1121	-2.7268 ^c	-2.7260 ^c
MLP1							--	2.0668 ^b	1.1659	0.4673	0.8358	1.8900 ^a	1.9672 ^b	2.1588 ^b
MLP2								--	-2.0965 ^b	-0.6269	-1.2912	1.8885 ^a	0.6472	1.3725
MLP3									--	0.1552	0.3403	1.8895 ^a	2.2167 ^b	2.1459 ^b
Elman										--	-0.0126	1.8893 ^a	0.7622	1.3475
Jordan											--	1.8893 ^a	1.6606 ^a	2.6416 ^c
JCN1												--	-1.8882 ^a	-1.8869 ^a
JCN2													--	1.3550
JCN3														--

Nota: a indica rechazo al 10%, b indica rechazo al 5% y c indica rechazo al 1%. Valores críticos distribución N(0,1), 1.645, 1.96, 2.576 al 10%, 5%, y 1%, respectivamente.

También, se han computado los resultados para el paseo aleatorio, rechazándose en casi todos los casos la hipótesis nula, salvo para los modelos TARCH y EGARCH, a cualquier nivel de significación utilizado.

En resumen, la predicción de RNA mejora ligeramente los resultados de algunos modelos no lineales de tipo ARCH y cambios de régimen de tipo STAR, encontrándose que las predicciones son en general independientes de sus valores reales, mediante los contrastes de predicción direccional, y los ECM no son en general iguales.

5. Conclusiones

El objetivo de este trabajo ha sido mostrar algunas evidencias alternativas sobre las características que poseen las rentabilidades diarias del Ibex35. Estas evidencias alternativas atienden a las características estadísticas de no estacionariedad y no linealidad, fundamentalmente. En los términos de la no estacionariedad, permiten ilustrar la posibilidad de que los rendimientos son predecibles, ya que se demuestra la existencia de reversión a la media y, por lo tanto, se obtienen autocorrelaciones negativas en el largo plazo. Esta conducta, implica que los precios se apartan ligeramente de la conducta de camino aleatorio y, de esta forma, el argumento induciría a rechazar la hipótesis débil de los mercados eficientes. Precisamente, como primera conclusión de este trabajo está que los resultados de los contrastes de raíces unitarias - con o sin cambio de estructura - muestran que existen componentes similares al camino aleatorio en los precios; pero, el contraste de la razón de las varianzas muestra que existe reversión a la media, evidenciando que las rentabilidades son predecibles y por lo tanto que existen desviaciones de dicha senda.

Por otro lado, otra propiedad de las rentabilidades del Ibex35, a parte de la de ser ligeramente predecibles, es la conducta no lineal que las caracteriza. Esta conducta permite representar ciertos hechos estilizados, como la existencia de autocorrelaciones distintas de cero en los cuadrados de las rentabilidades o la existencia de asimetrías y estados de la economía que diferencian a las rentabilidades positivas y negativas, así como a las volatilidades de las mismas que son bajas y elevadas. En este sentido, los modelos planteados evidencian que en las distintas posibilidades existentes para la modelización de las no linealidades, pueden tenerse en cuenta ciertos hechos estilizados.

Sin embargo, la presencia de diversas alternativas de modelización dificulta la selección de los modelos planteados. Por este motivo, recurrimos a la comparación de la predicción que realizan los modelos. Así, comparamos los modelos lineales y no lineales mediante diversos estadísticos descriptivos y contrastes de hipótesis, que permiten concluir que los modelos lineales predicen peor las rentabilidades que los

modelos no lineales, y que dentro de los modelos no lineales algunos esquemas de redes neuronales artificiales presentan ciertas ventajas en la capacidad predictiva frente al resto de modelos no lineales considerados en este trabajo. En este sentido, también deberíamos advertir que las RNA tienen algunas desventajas como la carencia de contrastes de hipótesis, la imposible interpretación de los parámetros y los problemas derivados del sobreajuste producido por la inclusión de demasiados coeficientes, aspecto este último que hemos intentado evitar. Además, también es un aspecto a tener en cuenta, que dado que el residuo del modelo se supone generalmente iid, podría ser interesante determinar las propiedades estadísticas cuando el residuo es, por ejemplo, condicionalmente heterocedástico.

Referencias Bibliográficas

Alonso, F. (1995): “La Modelización de la Volatilidad del Mercado Bursátil Español”. Banco de España. Servicio de Estudios. Documento de Trabajo nº. 9507.

Bollerslev, T., Engle, R. and Nelson, D. (1994): ARCH Models. En *Handbook of Econometrics*, Vol. IV, Cap. 49, 2961-3037. Eds. R.F. Engle and D.L. McFadden. Elsevier Science.

Brock, W., Dechert, W. and Scheinkman, J. (1987): “A Test for Independence based on the Correlation Dimension”. University of Wisconsin, Dept. of Economics. Revisado en Brock, W., Dechert, W., Scheinkman, J. and LeBaron, B. (1996), *Econometric Reviews*, 15, 197-235.

Brock, W., Hsieh, D., and LeBaron, B. (1991): *Nonlinear Dynamics, Chaos and Instability: Statistical Theory and Economic Evidence*, MIT Press, Cambridge, MA.

Campbell, J., Lo, A. and MacKinley, A. (1997): *The Econometric of Financial Markets*. Princeton University Press, Princeton, New Jersey.

Chong, Y. and Hendry, D. (1986): “Econometric Evaluation of Linear Macroeconomic Models”. *Review of Economic Studies*, 53, 671-690.

Clements, M. and Smith, J. (1999): “A Monte Carlo Study of the Forecasting Performance of Empirical SETAR Models”. *Journal of Applied Econometrics*, 14, 123-141.

Cochrane, J. (1988): “How Big is the Random Walk in GNP?”. *Journal of Political Economy*, 96, 893-920.

Collard, J. (1991): “A B-P ANN Commodity Trader”. In R.P. Lippmann et al. (eds). *Advances in Neural Information Processing Systems* (Morgan Kaufmann, San Mateo, CA), 551-556.

De Lima, P. (1998): “Non Linearities and Nonstationarities in Stock Return”. *Journal of Business and Economic Statistics*, 16(2), 227-236.

Diebold, F. (1988): “Deviation from Random-Walk Behaviour Tests Based on the Variance-Time Function”. Special Studies Paper 224, Federal Reserve Board, Washington DC.

- Diebold, F. and Mariano, R. (1995): "Comparing Predictive Accuracy". *Journal of Business & Economic Statistics*, 13(3), 253-263.
- Donaldson, R. and Kamstra, M. (1996): "Forecast Combining with Neural Networks". *Journal of Forecasting*, 15, 49-61.
- Elman, J.L. (1990): "Finding Structure in Time". *Cognitive Science* 14, 179-211.
- Episcopos, A. and Davis, J. (1995): "Predicting Returns on Canadian Exchange Rates with Artificial Neural Networks and Egarch-M Models". Working Paper. Clarkson University.
- Engle, R.F. and Ng, V. (1993): "Measuring and Testing the impact of News on Volatility". *Journal of Finance*, 48, 1749-1778.
- Fama, E. (1965): "The Behaviour of Stock Market Prices". *Journal of Business*, 38, 34-105.
- Fama, E. and French, K. (1988): "Permanent and Temporary Components of Stock Prices". *Journal of Political Economy*, 96(2), 246-273.
- Fernández, F., Sosvilla, S. y García, D. (1997): "Using Nearest-Neighbour Predictors to Forecast the Spanish Stock Market". *Investigaciones Económicas*, 21, 75-91.
- Fernández, F., García, D. and Sosvilla, S. (1999): "Dancing with Bulls and Bears: Nearest-Neighbour Forecast for the Nikkei Index". *Japan and the World Economy*, 11, 395-413.
- Fernández, F., González, Ch. y Sosvilla, S. (2000): "On the Profitability of Technical Trading Rules based on Artificial Neural Networks: Evidence from the Madrid Stock Market". *Economic Letters*. Pendiente de publicación.
- Franses, P.H. (1998): *Time Series Model for Business and Economic Forecasting*. Cambridge University Press.
- Franses, P.H. and Van Homelen, P. (1998): "On Forecasting Exchange Rates Using Neural Networks". *Applied Financial Economics*, 8, 589-596.
- Gallant, A.R. and White, H. (1988): *A Unified Theory of Estimation and Inference for Non-Linear Dynamic Models*. Oxford: Basic Blackwell.
- García, R. and Gençay, R. (2000): "Pricing and Hedging Derivative Securities with Neural Networks and a Homogeneity Hint". *Journal of Econometrics*, 94, 93-115.
- Gençay, T. (1996): "Non-Linear Prediction of Security Return with Moving Average Rules". *Journal of Forecasting*, 15, 165-174.
- Giles, C., Lawrence, S. and Tsoi, A. (1997): "Rule Inference for Financial Prediction Using Recurrent Networks". *Proceedings of IEEE/IAFE Conference on Computational Intelligence for Financial Engineering (CIFER)*, IEEE, Piscataway, NJ, 253-259.
- Glosten, L, Jagannathan, R. and Runkle, D. (1993): "Relationship Between the Expected Value and the Volatility of the Nominal Excess Return of Stocks". *Journal of Finance*, 48, 1779-1801.
- Goldfeld, S. and Quandt, R. (1973): "A Markov Model for Switching Regression". *Journal of Econometrics*, 1, 3-16.

- Granger, C.W.J. (1992): "Forecasting Stock Market Proces: Lessons for Forecasters". *International Journal of Forecasting*, 8, 3-13.
- Granger, C.W.J. and Teräsvirta, T. (1993): *Modelling Nonlinear Economic Relationships*. Oxford, UK: Oxford University Press.
- Hallman, J. and Kamstra, M. (1989): "Combining Algorithms based on Robust Estimation Techniques and Cointegration Restrictions". *Journal of Forecasting*, 8, 189-198.
- Hamilton, J. (1988): "Rational ExpectationsEconometric Analysis of Changes in Regime: An Investigation of the term Structure of Interest Rates". *Journal of Economic Dynamic and Control*, 12, 385-432.
- Hamilton, J. (1989): "Analysis of Time Series Subject to changes in Regime". *Journal of Econometrics*, 45, 39-70.
- Haugen, R. (1999): *The Inefficient Stock Market. What Pays Off and Why?*. Prentice Hall, Upper Saddle River, New Jersey.
- Hill, T., Marquez, L., O'Connor, L. and Remus, W. (1994): "Artificial Neural Network Models for Forecasting and Decision Making". *International Journal of Forecasting*, 10, 5-15.
- Hinich, M. and Paterson, D. (1985): "Evidence of Nonlinearity in Stock Returns". *Journal of Business and Economic Statistics*, 3, 69-77.
- Hornik, K., Stinchcombe, M. and White, H. (1990): "Universal Approximation of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks". *Neural Networks*, 3, 551-560.
- Hsieh, D. (1991): "Chaos and Nonlinear Dynamics: Applications to Financial Markets". *Journal of Finance*, 46, 1839-1878.
- Kim, C. and Nelson, C. (1999): *State-Space Models with Regime Switching. Classical and Gibbs-Sampling Approaches with Applications*. MIT Press, Cambridge. Massachusetts.
- Kuan, C., Hornik, K. and White, H. (1990): "Some Convergence Results for Learning in Recurrent Neural Networks". UCSD Department of Economics Discussion Paper.
- Kuan, C. and White, H. (1994): "Artificial Neural Networks: An Econometric Perspective". *Econometric Reviews*, 13, 1-91.
- Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992): "Testing the Null Hypothesis of Stationary Against the Alternative of a Unit Root", *Journal of Econometrics*, 54, 159-178.
- Lamoureux, C.G and Lastrapes, W.D. (1990): "Persistence in Variance, Structural Change, and the GARCH Model". *Journal of Business and Economic Statistics*, 8, 225-234.
- León, A. y Mora, J. (1999): "Modelling Conditional Heterokedasticity: Application to the IBEX35 Stock Return Index". *Spanish Economic Review*, 1(3), 215-238.
- Lo, A. and MacKinley, A. (1988): "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test". *Review of Financial Studies*, 1, 41-66.

Moody, J. and Utans, J. (1995): "Architecture Selection Strategies for Neural Networks: Application to Bond Rating Prediction". In Refenes, A.-P.N. (ed.), *Neural Networks in the Capital Markets*, New York, Wiley.

Neftçi, S. (1984): "Are Economic Time Series Asymmetric Over the Business Cycle". *Journal of Political Economy*, 92, 307-328.

Pérez-Rodríguez, J., Torra, S. y Borrell, M. (2000). "Predicción de Tipos de Interés en el Mercado Interbancario: Estructuras Lineales, GARCH y Redes Neuronales Artificiales". *Revista Asturiana de Economía*. Pendiente de publicación.

Perron, P. (1997): "Further Evidence on Breaking Trend Functions in Macroeconomic Variables", *Journal of Econometrics*, 80, 355-386.

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

Peters, E. (1994): *Fractal Market Analysis: Applying Chaos Theory*. Wiley, New York.

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

Ratner, M. (1996): "Investigating the Behaviour and characteristics of the Madrid Stock Exchange". *Journal of Banking & Finance*, 20, 135-149.

Refenes, A.-P. and Zapranis, A. (1999): "Neural Model Identification, Variable Selection and Model Adequacy". *Journal of Forecasting*, 18, 299-332.

Schnader, M. and Stekler, H. (1990): "Evaluating Predictions of Change". *Journal of Business*, 63, 99-107.

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

Scheinkman, J. and LeBaron, B. (1989): "Nonlinear Dynamics and Stock Returns". *Journal of Business*, 62, 311-338.

Surkan, A. and Singleton, J. (1990): "Neural Networks for Bond Rating Improved by Multiple Hidden Layers". *International Joint Conference on Neural Networks*, Vol.2, San Diego, CA(IEEE, New York), 157-162.

Swanson, N. and White, H (1995): "A Model-Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks". *Journal of Business & Economic Statistics*, 13(3), 265-275.

Takens, F. (1981): "Detecting Strange Attractor in Turbulence". In Rand, D. and Young, L. (eds.). *Dynamical Systems and Turbulence*. Springer Verlag, Berlin.

Tam, K. and Kiang, M. (1992): "Managerial Applications of Neural Networks: The Case of Bank Failure Predictions". *Management Science*, 38, 926-947.

Teräsvirta, T. (1994): "Specification, Estimation and Evaluation of Smooth Transition Autoregressive Models". *Journal of the American Statistical Association*, 89, 208-218.

Teräsvirta, T., Tjøstheim, D. and Granger, C.W.J. (1994): "Aspects of Modelling Nonlinear Time Series". En *Handbook of Econometrics*, Vol. IV, Cap. 48, 2919-2957. Eds. R.F. Engle and D.L. McFadden. Elsevier Science.

White, H. (1987): "Some Asymptotic Results for Back-Propagation". Proceedings of IEEE First International Conference on Neural Networks, San Diego, New York: IEEE, Press, III, 261-266.

White, H. (1988): "Economic Prediction Using Neural Networks: The Case of IBM Daily Stock Returns". Proceedings of the IEEE International Conference on Neural Networks, San Diego, California.

Abstract

This paper studies the nonstationary and nonlinear dynamics of the stock returns on the Ibex35 Stock Exchange Market. We analyze the nonstationary and nonlinear path of time series using unit root tests, mean reversion test, long-term memory test and BDS standard and recursive test. Also, we employ different class of uniequational nonlinear specifications: ARCH models, STAR models, Markov switching and Artificial Neural Networks models. These are fitted and forecasts the stock returns during the period 30/12/1989 to 10/2/2000. The results of the nonstationary and nonlinearity tests imply that we may have a good chance to fit and forecast more accurately the daily stock returns by using nonlinear models. Also, we evaluate the accuracy of forecast using some statistics of goodness of forecast, and encompassing tests, directional accuracy tests and the equality of competing forecast test. The results show that artificial neural network forecast some better, when we compare with others predictors.