



UNIVERSIDAD  
POLITECNICA  
DE VALENCIA



**UNIVERSIDAD POLITÉCNICA DE VALENCIA  
FACULTAD DE ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS**

**TRABAJO FINAL DE MÁSTER DIRECCIÓN FINANCIERA Y FISCAL**

**MODELIZACIÓN DE LA VOLATILIDAD CONDICIONAL DEL  
ÍNDICE IBEX-35:**

**COMPARATIVA MODELOS GARCH VERSUS REDES  
NEURONALES**

**Autor: Javier Oliver Muncharaz**

**Director: D. Francisco Guijarro Martínez**

**Valencia Septiembre 2010**

## **AGRADECIMIENTOS**

En primer lugar quiero agradecer a mi director D. Francisco Guijarro Martínez por su magnífica dirección y gran ayuda con sus aportaciones que, sin duda alguna, he tomado en gran consideración.

También quiero agradecer a toda mi familia su apoyo durante toda mi vida tanto personal, académica y profesional.

Y, por encima de todo, a mi mujer M<sup>a</sup> Carmen por estar en todo momento dándome ánimo para realizar el presente trabajo.

# INDICE

Índice de figuras y tablas. ....	4
1. Introducción y Objetivos .....	6
2. Modelización de la Volatilidad Mediante Modelos GARCH .....	9
2.1. Concepto de Volatilidad .....	9
2.2. Medidas de Volatilidad.....	11
2.3. Modelos Univariantes.....	13
2.4. Modelos Multivariantes.....	22
2.4.1. Modelo de corrección constante .....	23
2.4.2. Modelo Vech .....	23
2.4.3. Modelo BEKK.....	24
2.5. Estimación de los modelos ARCH .....	25
2.6. Estudio de la relación volumen-volatilidad .....	26
3. Las Redes Neuronales Artificiales.....	30
3.1. Introducción.....	30
3.2. Fundamentos Biológicos .....	32
3.3. Estructura y Elementos de las RNA .....	35
3.3.1. Las neuronas artificiales .....	35
3.3.2. Estado de Activación .....	37
3.3.3. Función de Activación o de Transferencia .....	37
3.3.4. Función de salida .....	38
3.3.5. Mecanismos o Reglas de Aprendizaje .....	41
3.3.6. Aprendizaje supervisado.....	43
3.3.7. Aprendizaje no supervisado.....	45
3.3.8. Relación entre la información de entrada y salida de la RNA .....	48
3.4. RNA con conexiones hacia adelante: Backpropagation .....	49
3.4.1. El Aprendizaje en la Red Backpropagation: La Regla Delta Generalizada.....	49
3.4.2. Estructura y Aprendizaje de la Red Backpropagation .....	51
4. Estudio comparativo entre modelo ARMA-GARCH-M y la Red Neuronal Backpropagation: Aplicación sobre la volatilidad condicional y rentabilidad del índice Ibex-35 .....	57
4.1. Modelización volatilidad condicional y la rentabilidad del Ibex-35 mediante modelo ARMA- GARCH-M.....	57
4.2. Estimación y predicción de la volatilidad condicional y la rentabilidad del Ibex-35 mediante la RNA Backpropagation .....	84
5. Conclusiones .....	93
6. Bibliografía.....	95

# Índice de figuras y tablas.

## FIGURAS

Figura 3-1: Gráfico neurona .....	33
Figura 3-2: Esquema estructura red neuronal.....	35
Figura 3-3: Esquema funcionamiento transmisión .....	36
Figura 3-4: Función de activación escalón .....	39
Figura 3-5: Función de activación lineal o identidad .....	39
Figura 3-6: Función de activación lineal-mixta.....	40
Figura 3-7: Función de activación sigmoïdal .....	41
Figura 3-8: Estructura Red Backpropagation .....	52
Figura 4-1: Gráfico cierres diarios del Ibex-35 del periodo muestral.....	58
Figura 4-2: Gráfico rentabilidad del Ibex-35.....	58
Figura 4-3: Histograma de la rentabilidad del Ibex-35.....	59
Figura 4-4: Histograma residuos modelo ARMA(1,1).....	68
Figura 4-5: Histograma residuos modelo ARMA(1,1)-GARCH(2,1).....	71
Figura 4-6: Histograma residuos modelo ARMA(1,1)-GARCH(2,1)-M .....	80
Figura 4-7: Histograma residuos modelo ARMA(1,1)-GARCH(2,1)-M sin outliers.....	83
Figura 4-8: Histograma residuos modelo ARMA(1,1)-GARCH(4,2) con todas las observaciones .....	83
Figura 4-9: Listado Red Backpropagation: criterio Average Absolute Error .....	87
Figura 4-10: Datos Red Backpropagation seleccionada: criterio Average Absolute Error .....	88
Figura 4-11: Listado Red Backpropagation: criterio Mean Squared Error.....	89
Figura 4-12: Datos Red Backpropagation seleccionada: criterio Mean Squared Error .....	90
Figura 4-13: Listado Red Backpropagation: criterio Root Mean Squared Error .....	91
Figura 4-14: Datos Red Backpropagation seleccionada: criterio Root Mean Squared Error .....	91

## TABLAS

Tabla 4-1: FAC y FACP de la rentabilidad Ibex-35.....	59
Tabla 4-2: Test DFA serie rentabilidad.....	60
Tabla 4-3: Test Phillip-Perron serie rentabilidad.....	61
Tabla 4-4: Test D-W con 1 retardo serie rentabilidad.....	62
Tabla 4-5: Test D-W con 2 retardos serie rentabilidad.....	62
Tabla 4-6: Test D-W con 3 retardos serie rentabilidad.....	63
Tabla 4-7: Test D-W con 4 retardos serie rentabilidad.....	63
Tabla 4-8: Test D-W con 5 retardos serie rentabilidad.....	64
Tabla 4-9: Estimación modelo ARMA(1,0).....	64
Tabla 4-10: Estimación modelo ARMA(1,1).....	65
Tabla 4-11: Estimación modelo ARMA(2,2).....	65
Tabla 4-12: Estimación modelo ARMA(1,2).....	66
Tabla 4-13: Estimación modelo ARMA(0,1).....	66
Tabla 4-14: Estimación modelo ARMA(0,2).....	67
Tabla 4-15: FAC y FACP modelo ARMA(1,1) serie rentabilidad.....	68
Tabla 4-16: Test autocorrelación Breusch-Godfrey residuos modelo ARMA(1,1).....	69
Tabla 4-17: Test Heterocedasticidad residuos modelo ARMA(1,1).....	70
Tabla 4-18: Estimación modelo ARMA(1,1)-GARCH(2,1).....	71
Tabla 4-19: Test estabilidad Chow modelo ARMA(1,1)-GARCH(2,1).....	72
Tabla 4-20: Estimación modelo ARMA(1,1)-GARCH(2,1)-M.....	72
Tabla 4-21: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando volumen con un retardo en la varianza condicional.....	73
Tabla 4-22: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando logaritmo del volumen con un retardo en la varianza condicional.....	74
Tabla 4-23: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando el volumen con un retardo en la varianza condicional e incorporando el logaritmo varianza condicional en la ecuación de la media.....	75
Tabla 4-24: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando el logaritmo del volumen con un retardo en la varianza condicional e incorporando el logaritmo varianza condicional en la ecuación de la media.....	76
Tabla 4-25: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando el volumen con un retardo en la varianza condicional e incorporando la desviación de la varianza condicional en la ecuación de la media.....	77
Tabla 4-26: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando el logaritmo del volumen con un retardo en la varianza condicional e incorporando la desviación de la varianza condicional en la ecuación de la media.....	78
Tabla 4-27: Estimación modelo ARMA(1,1)-GARCH(2,1)-M con desviación estándar de la varianza condicional en la ecuación de la media y volumen con un retardo en la varianza condicional.....	79
Tabla 4-28: FAC y FACP del modelo ARMA(1,1)-GARCH(2,1)-M con desviación estándar de la varianza condicional en la ecuación de la media y volumen con un retardo en la varianza condicional.....	80
Tabla 4-29: Estimación modelo ARMA(1,1)-GARCH(4,2)-M con desviación estándar de la varianza condicional en la ecuación de la media y volumen con un retardo en la varianza condicional.....	81
Tabla 4-30: FAC y FACP del modelo ARMA(1,1)-GARCH(4,2)-M con desviación estándar de la varianza condicional en la ecuación de la media y volumen con un retardo en la varianza condicional.....	82
Tabla 4-31: Resultados del modelo ARMA(1,1)-GARCH(4,2)-M con desviación estándar de la varianza condicional en la ecuación de la media y volumen con un retardo en la varianza condicional.....	84
Tabla 4-32: Resultados Red Neuronal Backpropagation: criterio Average Absolute Error.....	88
Tabla 4-33: Resultados Red Neuronal Backpropagation: criterio Mean Squared Error.....	89
Tabla 4-34: Resultados Red Neuronal Backpropagation: criterio Root Mean Squared Error.....	92
Tabla 5-1: Cuadro resumen resultados para la rentabilidad del índice Ibex-35.....	93
Tabla 5-2: Cuadro resumen resultados para la volatilidad condicional del índice Ibex-35.....	94

# 1. Introducción y Objetivos

En las últimas décadas hemos asistido a un gran desarrollo y expansión de los mercados financieros y de los activos que se negocian en ellos. Todo ello derivado, por un lado de la globalización internacional de los mercados, destacando la desregularización de los mercados domésticos, y la liberación de flujos de capital a nivel internacional. Por otro lado, está el rápido y continuo desarrollo a nivel de telecomunicaciones y de sistemas de información, con los que se ha logrado dicha globalización. Todo esto da la posibilidad a los inversores, en general, de realizar una correcta y más eficiente colocación de su inversión, dado que aumentan las posibilidades de colocación del capital, así como una reducción de costes de transacción.

Dada todas estas nuevas circunstancias que se han ido dando a lo largo de los últimos años en los mercados financieros internacionales, se puede decir que la volatilidad se ha incrementado considerablemente respecto a periodos anteriores. Como ejemplos de este suceso se pueden citar el crack bursátil de 1987, crisis del sistema monetario Europeo en 1992-93, entre otros Ayuso, et al. (1995). Es a partir del crack de 1987 cuando se intensifican los estudios sobre la volatilidad y su modelización.

Con la aparición y posterior desarrollo de los mercados de derivados financieros, y en concreto, sobre índices bursátiles, se ha dado una serie de posibilidades al inversor que antes o eran imposibles de realizar o los grandes costes de transacción lo impedían. Con estos nuevos activos derivados, se puede realizar coberturas en las carteras de los inversores de una forma más rápida y menos costosa, reduciendo así el riesgo de éstas, ya que realizar múltiples operaciones directas en contado para ajustar las carteras resultaba muy costoso; ha incrementado el flujo de información en la formación de precios en el mercado de contado. La aparición de los primeros mercados de derivados sobre índices bursátiles en 1982 (el primer contrato de futuros sobre índice bursátil fue implantado en el mercado Kansas City Board of Trade teniendo como subyacente el Value Line Composite Index) tuvieron críticas muy importantes, debido a que su implantación coincidió con periodos de alta volatilidad en los mercados de contado. A partir de este momento se estableció una discusión importante sobre si la creación de mercados derivados desestabilizan o no el mercado de contado. A pesar de

las críticas suscitadas, es clara las ventajas, entre otras, de liquidez y cobertura, que proporcionan los mercados derivados a los inversores. Aunque en la práctica en determinados vencimientos y para activos concretos puede haber cierto movimiento provocado por la transmisión de la volatilidad de un mercado a otro, no por ello se desestabiliza.

Por tanto, se hace necesaria la modelización de la volatilidad de los activos financieros e índices bursátiles para obtener una valoración de los mismos más adecuada para realizar tanto inversiones como coberturas en los mercados.

En el presente trabajo se plasma la comparativa entre un modelo econométrico y las redes neuronales para la estimación y predicción simultánea de la rentabilidad y la varianza condicional sobre el índice Ibex-35. El objetivo es demostrar que, a priori, las redes neuronales tienen la ventaja de ser capaces de encontrar relaciones más complejas entre las variables y, por tanto, tienen mayor capacidad de predicción frente a los modelos econométricos.

En el capítulo 2 del presente trabajo se realiza una revisión de los principales modelos GARCH que se utilizan para la modelización de la volatilidad condicional. Se detallan tanto los modelos univariantes como los multivariantes, que se emplean generalmente para estudios de transmisión de volatilidad.

En el capítulo 3 se realiza una descripción de los elementos principales de las redes neuronales así como sus estructuras más importantes. Se estudiará los mecanismos y tipos de aprendizaje de las mismas. Para finalizar este capítulo se realizará un estudio en profundidad sobre la red neuronal Backpropagation que será la utilizada en nuestro estudio comparativo.

En el capítulo 4 realizaremos el estudio comparativo en cuestión. Para ello se tomará una muestra de las rentabilidades del índice Ibex-35. En primer lugar se analizará la estacionariedad de la serie para, después, encontrar el proceso generador ARMA de la rentabilidad del índice. A este modelo se le incorporará el modelo GARCH para la volatilidad condicional, en concreto, GARCH-M o garch en media, que relaciona la rentabilidad (ecuación de la media) con la volatilidad condicional. Sobre

este modelo se realiza un estudio sobre la relación entre volumen negociado y volatilidad y su incorporación al modelo. Una vez finalizado el modelo se muestran los errores de estimación y predicción para la muestra seleccionada. Por otro lado se procede a determinar los inputs que alimentarán la red Backpropagation, así como los parámetros configurables sobre la misma para el proceso de aprendizaje. Se establecerán tres criterios de selección de red y por tanto, se realizarán tres entrenamientos diferentes. A continuación se presentarán los errores cometidos por la red, en los tres casos, en la predicción de la rentabilidad y volatilidad del índice Ibex-35.

Por último, en el capítulo 5 se establecen las principales conclusiones tanto en la comparativa entre el modelo ARMA-GARCH-M y la red neuronal, destacando la ventaja que supone la segunda sobre la primera a tenor de los resultados obtenidos. También se indicará algunas sugerencias para mejorar y consolidar estos resultados.

## **2. Modelización de la Volatilidad Mediante Modelos GARCH**

### **2.1. Concepto de Volatilidad**

En los últimos años se han incrementado notablemente los estudios, tanto teóricos como empíricos, sobre la volatilidad y su modelización. En la valoración de acciones se encuentran múltiples trabajos en los que se determina la importancia de una correcta especificación del modelo de volatilidad, así tenemos a Merton (1980) y French, Schwert, y Stambaugh (1987), que exponen que la esperanza del rendimiento del mercado está relacionada con la volatilidad del mercado de contado.

El concepto de volatilidad nombrado anteriormente, se puede definir de forma sencilla como la velocidad de los movimientos de los precios como expresa Lamothe, P. (1993) citando Natemberg (1988) o también es el posible rango de variaciones de precios. Otra definición algo más formal sería aquella que da Peiró (1992) siendo volatilidad la variabilidad del rendimiento de un activo, que en definitiva es una medida de riesgo. Esta variabilidad del rendimiento de los precios se produce por las noticias e información (innovaciones) que aparecen en el mercado, dado que éstas afectan a las expectativas de los inversores. Lógicamente gran parte de la información será económica-financiera. La repercusión que tendrá las noticias vendrá dada por la mayor o menor diferencia entre los datos reales y lo que había estimado los inversores. Existe un tipo de información o noticias que son casi impredecibles y cuyas consecuencias no son fáciles de determinar. Entre estas podemos destacar noticias políticas y fenómenos naturales. Las características principales de la información que afecta a la variabilidad de los precios son tres. En primer lugar la información no siempre llega distribuida de una forma uniforme a lo largo del tiempo, sino que aparece frecuentemente en bloques y no siempre distribuida normalmente. En segundo lugar, según el carácter de las noticias tendrán un efecto diferente sobre la volatilidad (en múltiples estudios se ha visto que las noticias negativas incrementan más la volatilidad que noticias positivas). Por último, el efecto de la llegada de la nueva información al mercado no siempre es inmediato, sino que éste necesita analizarla y valorarla. Por tanto, cuanto más grande sea el impacto de las noticias, más tiempo durarán sus efectos sobre la volatilidad.

Después de haber definido el concepto de volatilidad se hace necesario distinguir varios tipos de volatilidad existentes, volatilidad histórica, implícita y futura. La volatilidad histórica se puede definir como una medida retrospectiva de la volatilidad y refleja el comportamiento de la variabilidad de los precios en el pasado. Para el cálculo de esta volatilidad se analizan, por definición, las series históricas de los precios. Las variables para su cálculo pueden basarse según Lamothe (1993) bien en base de precios de cierre, bien en base al máximo y mínimo de las sesiones de negociación. En función del periodo de tiempo elegido para el cálculo de la volatilidad podríamos obtener volatilidades diarias, semanales anuales, etc. También hay que tener en cuenta que es diferente calcular la volatilidad teniendo en cuenta varios años de una serie histórica que sólo varios meses. A pesar de ello, existe una correlación muy importante entre la volatilidad calculada para diferentes periodos de tiempo.

Para entender el concepto de volatilidad implícita hemos de situarnos en el contexto de los mercados derivados, y más concretamente, en las opciones financieras. En el mercado se utiliza de forma general la fórmula de Black-Scholes para valorar este tipo de instrumento financiero. Dicho modelo de valoración requiere bien la volatilidad, bien la prima de la opción para calcular uno u el otro. La volatilidad implícita podríamos definirla como aquella que se obtiene utilizando como dato de entrada la prima de la opción que se está negociando en el mercado. Así pues, para el cálculo de la volatilidad implícita, con carácter general, es necesario seleccionar un modelo de valoración que sea utilizado por la gran mayoría del mercado. Esta volatilidad reflejará la volatilidad que se está negociando o que subyace en un momento concreto del tiempo.

Por último, podemos definir la volatilidad futura como aquella que cualquier inversor quisiera saber. Es decir, se trata de la volatilidad que habrá de hecho en un momento  $t+1$ . Los diferentes modelos que se construyen para su estimación utilizan bien la volatilidad histórica, bien la volatilidad implícita, ya que son las únicas herramientas posibles. Puede destacarse la relación entre volatilidad implícita e histórica, mediante los conos de volatilidad como instrumento inicial de predicción de la volatilidad futura. A pesar de la utilización de la volatilidad implícita en la predicción de la volatilidad futura, destaca Lamothe (1993), que se trata de un estimador sesgado

cuando se aproxima a alguna de las bandas del cono de volatilidad histórica. En la actualidad, los modelos ARCH y GARCH son los que mejores resultados parecen obtener en la predicción de la volatilidad.

## **2.2. Medidas de Volatilidad**

Una vez acotada de forma más o menos generalizada las definiciones sobre el concepto de volatilidad más comúnmente aceptadas, se presenta el problema de disponer de una medida lo más correcta posible que nos indique la variabilidad de los rendimientos de los activos. La estadística proporciona las herramientas adecuadas para el cálculo de la volatilidad. Dado que existen diferentes formas temporales de cálculo del rendimiento (diarios, mensuales,...) y varias formas de medir la dispersión o variabilidad de dichos rendimientos, surgen, pues, diferentes medidas de volatilidad. Siguiendo a Peiró (1992), se presenta a continuación una de las medias de volatilidad más utilizadas en la literatura financiera. Se trata de la desviación típica muestral de los rendimientos diarios correspondientes a un mes concreto.

$$\sqrt{\frac{\sum_{i=1}^n (R_{it} - \bar{R}_k)^2}{n}}$$

Donde  $R_{ik}$  con  $i = 1, 2, \dots, n$ , corresponde a los diferentes rendimientos diarios del mes  $k$ , que tiene  $n$  días de negociación,  $\bar{R}_k$  es la media muestral de los rendimientos diarios del mes  $k$ . Esta medida se puede utilizar para periodos más largos (rendimientos semanales, mensuales,...).

Fama plantea una desviación típica móvil de 12 rendimientos mensuales en la que se elimina el rendimiento más alejado a la vez que se añade uno nuevo. Esta forma de obtener la serie de datos de la volatilidad es mucho más suave que con la fórmula anterior, favoreciendo así su estudio.

Ahora bien, cuando se deja de lado estas medidas estadísticas muestrales y se plantea que los rendimientos siguen un proceso estocástico, aparece el problema que no es posible calcular la desviación típica de los rendimientos teniendo una sola

observación, dado que cada rendimiento se considera una variable aleatoria independiente. El análisis de la volatilidad, ante estas circunstancias, se deberá realizar mediante mecanismos de estimación. Así tenemos a French, Schwert y Stambaugh (1987) que desarrollan una estimación de la volatilidad mensual, a través de los rendimientos diarios, expresados de forma logarítmica. Utiliza por tanto la varianza del rendimiento, expresado en logaritmos, como medida de volatilidad.

$$Var(R_k) = \sum_{i=1}^n \sum_{j=1}^n Cov(R_{ik}, R_{jk}) = \sum_{i=1}^n Var(R_{ik}) + \sum_{i=1}^n \sum_{j=1}^n Cov(R_{ik}, R_{jk})$$

con  $i \neq j$

Donde  $R_k$  es el rendimiento mensual, expresado en logaritmos, correspondiente al mes  $k$ .

Suponiendo que la esperanza de los rendimientos es cero  $E(R_{ik})=0$ , para todo  $i$ , se tiene que  $Var(R_{ik})=E(R_{ik}^2)$  y  $Cov(R_{ik}, R_{jk})=E(R_{ik} R_{jk})$ . De esta forma se tiene como estimadores de estos, los siguientes:

$$\hat{Var}(R_{ik}) = R_{ik}^2$$

$$\hat{Cov}(R_{ik}, R_{jk}) = R_{ik} R_{jk}$$

Si los rendimientos están incorrelados entre sí, la estimación de la volatilidad queda como sigue:

$$\hat{Var}(R_k) = \sum_{i=1}^n R_{ik}^2$$

Si suponemos que los rendimientos están correlados con orden 1 la estimación de la volatilidad será:

$$\hat{Var}(R_k) = \sum_{i=1}^n R_{i-1k}^2 + 2 \sum_{i=1}^{n-1} R_{ik} R_{i+1k}$$

Expresión que utilizan en su trabajo French, Schwert y Stambaugh (1987). Por otro lado, como apunta Ayuso y otros (1995), el concepto de volatilidad está asociado al movimiento de la variable rendimiento de precios, es decir, se asimila al concepto de riesgo. Éste viene medido por la varianza condicional que se expresa de la siguiente forma:

$$h_t \equiv V_t(X_{t+i}) \equiv E_t(X_{t+i} - E_t(X_{t+i}))^2$$

donde  $h_t$  es la varianza condicional (volatilidad) y  $E_t$  es la esperanza condicional con la información disponible hasta el momento  $t$ . Esta será la expresión que se utilizará de ahora en adelante al referirnos a la volatilidad.

Hemos visto alguna de las formas más habituales en la literatura financiera de medir la volatilidad, se plantea ahora su modelización para poder realizar estudios sobre su valoración y evolución, para un activo concreto o para un mercado.

Dado que el rendimiento se puede descomponer en una parte anticipada (condicional) y otra no anticipada, muchos autores han utilizado un término que refleje la variación condicional de los precios, generalmente expresado como  $h^{1/2}$ . Esto unido a la idea planteada por Mandelbrot (1963) y citada por muchos autores como Engle, Liben y Robins (1987), de que “ después de variaciones grandes en los precios siguen variaciones grandes de cualquier signo, mientras que variaciones pequeñas en precios van seguidas de variaciones pequeñas”; da lugar a una serie de modelos que reflejan en buena medida este comportamiento de la volatilidad. A continuación se va a explicar de forma resumida los modelos más generales y utilizados en la literatura financiera, siguiendo la clasificación realizada por Pagan (1996).

### **2.3. Modelos Univariantes**

En los modelos univariantes, como su nombre indica, sólo se expresará la volatilidad condicional de la única variable dependiente, existiendo, pues, una sola ecuación en medias. Engle (1982) introdujo por primera vez un modelo que representaba la volatilidad condicional de una serie de datos, llamado autorregresivo condicionalmente heterocedástico, Autoregressive Conditional Heteroskedasticity (ARCH).

Con la aparición de los modelos ARCH (Autorregresivos Condicionalmente Heterocedásticos), Engle propone uno nuevo proceso en el que se relaja la suposición de varianza constante, es decir, se plantea un modelo con media cero, que sigue un proceso serialmente incorrelado, en el que las varianzas

condicionales son no constantes y las varianzas no condicionales siguen siendo constantes. Estos modelos contrastan notablemente con los modelos convencionales de series temporales y econométricos que existían hasta entonces. Estos nuevos modelos presentan, pues, una varianza condicional, que depende de la información disponible, y que varía todo el tiempo en función de los residuos pasados, manteniendo constante su varianza incondicional. En Engle (1982) se presenta formalmente por primera vez, de forma general, el modelo ARCH. En Engle y Ng (1993) se puede ver el mismo modelo pero con un planteamiento desarrollado pensando en un mercado de acciones. Siguiendo a estos autores podemos explicar el modelo ARCH de la siguiente forma:

Tomando  $y_t$  como el ratio de rendimiento de un título bursátil o acción (sin perder generalidad, también podría tratarse de un índice bursátil o de cualquier otro activo financiero) del momento  $t-1$  a  $t$ , y siendo  $F_{t-1}$  la información pasada de las variables relevantes hasta el momento  $t$ . La esperanza del rendimiento es el valor de la esperanza condicional con la información disponible hasta ese momento. Es decir, la esperanza condicional del rendimiento ( $m_t$ ) es igual a  $m_t = E(y_t | F_{t-1})$ . Por otro lado, tenemos la volatilidad que será expresada como la varianza condicional del rendimiento,  $h_t = \text{Var}(y_t | F_t)$ . Dado que nos movemos en un ambiente con cierto grado de incertidumbre, puede haber un rendimiento inesperado en el momento  $t$ , que se define como  $\varepsilon_t = y_t - m_t$ , que no es más que el residuo de una estimación. Cuando hacemos referencia al mercado bursátil,  $\varepsilon_t$  se asocia a una medida de la nueva información que se incorpora al mercado en el momento  $t$ . Cuando  $\varepsilon_t$  es positivo (un incremento inesperado del rendimiento) implica que se ha incorporado al mercado información de carácter positivo, esto es, buenas noticias; mientras que cuando  $\varepsilon_t$  es negativo (un decrecimiento inesperado en el rendimiento) implicará la llegada de malas noticias. Un valor elevado de  $|\varepsilon_t|$  implicará que la información que se ha producido es significativa, y por tanto, relevante, ya que produce un gran efecto inesperado en los precios.

Engle (1982) sugiere que la varianza condicional  $h_t$  se puede modelizar como una función de retardos de  $\varepsilon_t$ . Esto significa que la volatilidad dependerá de la información pasada. El modelo ARCH desarrollado para un orden de  $p$  retardos (ARCH ( $p$ )) se puede expresar de la siguiente forma:

$$y_t = \varepsilon_t h^{1/2}$$

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \varepsilon_{t-i}^2$$

donde  $\alpha_0$  y  $\alpha_1 \dots \alpha_p$  son parámetros constantes. El efecto de un shock en el pasado sobre la volatilidad actual está determinado por  $\alpha_i$ , siendo  $i \leq p$ . Generalmente se esperará que la información más antigua tenga un menor efecto sobre la volatilidad actual. Es decir, que  $\alpha_i < \alpha_j$  siendo  $i > j$ .

En Engle, Lilien y Robins (1987) se desarrolla un modelo ARCH en el que la desviación estándar de cada observación afecta a la media. De esta forma cambios en la varianza condicional afectan de forma directa sobre la esperanza del rendimiento de un activo. Este modelo resuelve algunas de las paradojas que se planteaban en estudios empíricos, donde al incluir la heterocedasticidad, los tests estadísticos dejan de ser insesgados, requisito que suele ser preferido en los estimadores para un buen modelo. La consecuencia directa de esta paradoja empírica es la significatividad errónea de algunas variables. Formalmente se puede escribir como :

$$Y_y | X_t, \pi_t \sim N( \beta'X_t + \sigma h_t, h_t^2 )$$

$$h_t^2 = \alpha'W_{\pi} + \gamma Z_t$$

donde  $X_t$  y  $Z_t$  son  $k \times 1$  y  $j \times 1$  vectores de variables exógenas y de variables dependientes retardadas. El vector  $Z_t$  incluye una constante, donde su coeficiente representa el componente constante de la varianza de  $h_t$ .  $W$  es una matriz  $q \times p$  de constantes fijadas que se utilizan para imponer restricciones a la parametrización de la respuesta de la varianza condicional a los retardos de los residuos al cuadrado. El caso menos restringido sería aquel en que la matriz  $W$  sea la matriz identidad. Este modelo resuelve el problema de que variables que aparentemente se utilizan en la predicción de los excesos de rendimientos, son correladas con la prima de riesgo y si se incluye una función de la varianza condicional en el regresor, pierden su significatividad aparente.

En muchos trabajos empíricos en los que se ha aplicado el modelo ARCH, se encuentra el problema que muchos de ellos violaban las restricciones de no negatividad impuestas por el modelo, debido a que se introducía, de forma arbitraria, una estructura lineal fija de retardos en la ecuación de la varianza condicional, con el fin de poder reflejar la larga persistencia detectada en muchas de las series económicas utilizadas. Por este motivo, Bollerslev (1986) desarrolla un modelo Generalizado Autorregresivo Condicionalmente Heterocedástico, Generalizaed Autorregressive Conditional Heteroskedasticity (GARCh) en el que la estructura de retardos es mucho más flexible. A continuación se expone formalmente el modelo general GARCh (p,q):

$$\varepsilon_t | \psi_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} = \alpha_0 + A(L)\varepsilon_t^2 + B(L)h_t$$

donde  $p \geq 0; q > 0$

$$\alpha_0 > 0; \alpha_i \geq 0 \quad i = 1, \dots, q$$

$$\beta_i \geq 0 \quad i = 1, \dots, p$$

siendo  $\varepsilon_t$  el valor de un proceso estocástico de tiempo discreto, que al igual que se dijo en el modelo ARCH, puede representar una medida de la información que llega al mercado de acciones y  $\psi_{t-1}$  será la información disponible hasta el momento t. En este caso se ha asumido, al igual que en el caso ARCH, que la distribución condicional es normal, aunque podría darse otras distribuciones que no invalidarían dichos modelos.

Se puede observar fácilmente que para un orden p=0, el modelo general (GARCh) se convierte en el modelo presentado por Engle, el modelo ARCH(q); y para un orden p=q=0, se trataría de un proceso ruido blanco. La diferencia fundamental entre los dos modelos reside en que mientras en el modelo ARCH(q) la varianza condicional se especifica sólo como una función lineal de varianzas pasadas, en el modelo GARCh (p,q) incorpora en su función a las propias varianzas condicionales desfasadas.

El modelo GARCh descrito anteriormente es uno de los modelos más

utilizados para cualquier estudio relacionado con la volatilidad condicional. Estos modelos, así como sus variantes, están divididos en dos ecuaciones. La primera, la ecuación en medias, describe los datos observados en función de otras variables más un término de error. La segunda ecuación, que es la presentada anteriormente, especifica la evolución de la varianza condicional del término error de la ecuación en medias como una función de retardos de varianzas condicionales y términos de error.

Una de las variantes del modelo GARCH es el GARCH-M o GARCH en media. Fue propuesto por Engle, Lilien y Robins (1987), realmente propusieron el modelo ARCH-M. Este modelo se basa en incorporar a la ecuación de la rentabilidad condicionada la varianza condicional. Es decir, que la rentabilidad esperada dependa también de su varianza condicional. Esto puede ser muy adecuado en modelos en los que se espera de la existencia de una prima de riesgo, de forma que, el rendimiento esperado será mayor conforme aumente su varianza o riesgo. Por tanto, en este tipo de modelos, la varianza condicional se incorpora como variable explicativa en la ecuación de la media, de esta forma puede estudiarse la influencia que tiene la volatilidad (riesgo) que tiene sobre la rentabilidad. Según esto el modelo GARCH-M se presenta como sigue:

$$\text{Ecuación de la media } R_t = \delta + \gamma h_t + h_t^{1/2} \varepsilon_t$$

$$\text{Ecuación de la varianza condicional } h_t = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i h_{t-i} = \alpha_0 + A(L)\varepsilon_t^2 + B(L)h_t$$

donde  $p \geq 0; q > 0$

$$\alpha_0 > 0; \alpha_i \geq 0 \quad i = 1, \dots, q$$

$$\beta_i \geq 0 \quad i = 1, \dots, p$$

Siguiendo  $\varepsilon_t$  una distribución normal con media cero y desviación uno.

La forma de incorporar la influencia de la volatilidad condicional en la determinación del rendimiento puede hacerse de varias formas. Una es como se ha presentado, es decir, incorporar la varianza condicional directamente. Existen otras dos formas, también muy habituales, de esta incorporación bien incorporarlo como logaritmo de la varianza condicional  $\log h_t$  bien como la desviación estándar condicional  $h_t^{1/2}$ .

A continuación vamos a proceder a dar una breve descripción de otros de

los modelos más conocidos de la familia GARCH. Para una descripción detallada de los mismos se puede ver Engle y Ng(1993) y Alonso(1995) que realiza una revisión bastante completa describiendo las ventajas y cualidades de unos y otros. Los diferentes modelos surgen por la diferente forma de tratar la evolución de la volatilidad, así como la incorporación al modelo de medidas que recojan el impacto y posible asimetría de las noticias o innovaciones que llegan al mercado sobre la volatilidad.

Empezaremos por el modelo GARCH Asimétrico, Asymmetric GARCH (AGARCH). Este modelo introduce, aunque de forma sencilla, los posibles efectos asimétricos en la modelización de la varianza condicional, producidos por la diferencia de impacto de la información (noticias) positiva y negativa que se produce en el mercado. En trabajos empíricos es frecuente encontrar que las malas noticias, es decir, rendimientos menores a los esperados, generan mayor volatilidad que las buenas noticias Nelson(1990). En 1976, F. Black, ya argumentó que esta asimetría se debe al llamado “efecto apalancamiento”, en el que cuando una innovación (nueva información) negativa llega al mercado, se eleva el ratio Deuda/Capital de la acción, con el consiguiente incremento del impacto de las innovaciones futuras sobre el rendimiento de las acciones y, por tanto, sobre el riesgo de la inversión del activo. Este tema de las asimetrías se volverá a tratar en un capítulo posterior. Volviendo al modelo AGARCH, este mantiene la forma lineal del modelo GARCH, pero añade un término que permite que tenga respuestas de diferente intensidad a los errores pasados (recordemos que  $\varepsilon_t$  es la información disponible), en función de si los shocks o impactos son positivos o negativos. Así pues, un valor negativo de dicho parámetro provocará una respuesta asimétrica, con lo que una innovación positiva producirá un menor impacto que las negativas sobre la varianza condicional, es decir, la volatilidad.

$$h_t = \alpha_0 + \alpha_1(\varepsilon_t + \gamma)^2 + \beta_1 h_{t-1}$$

donde  $h_t$  es la varianza condicional,  $\varepsilon_t$  es el error e  $\gamma$  es el término que se añade al modelo GARCH(1,1) y será el que recoja la fuerza de los errores pasados. Al igual que en el modelo GARCH los parámetros  $\alpha_0, \alpha_1, \beta_1 \geq 0$  han de restringirse a la no negatividad.

A continuación podemos seguir con el modelo expuesto por Nelson en 1990. Se trata del modelo GARCH Exponencial, en terminología anglosajona Exponential GARCH (EGARCH). Al igual que el anterior modelo, el modelo EGARCH recoge respuestas asimétricas de la varianza condicional de los residuos pasados (innovaciones). Se diferencia del modelo GARCH y del anterior en que estos requerían de la imposición de restricciones de no negatividad en muchos de sus parámetros, con el objetivo de garantizar la positividad de la varianza condicional, mientras que en el modelo EGARCH no se plantean dichas restricciones. En este caso se asegura la positividad de la varianza condicional tomando ésta una forma funcional logarítmica en la expresión de la varianza condicional que, además, facilita la estimación de sus parámetros.

$$\log(h_t) = \alpha_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \left[ \gamma \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} + \left( \frac{|\varepsilon_{t-1}|}{\sqrt{h_{t-1}}} - (2/\pi)^{1/2} \right) \right]$$

donde el segundo término de la ecuación representa la esperanza matemática de  $\frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}}$

suponiendo una distribución Normal(0,1), siendo  $E \left[ \left| \frac{\varepsilon_{t-1}}{\sqrt{h_{t-1}}} \right| \right] = \left( \frac{2}{\pi} \right)^{1/2}$

La asimetría, en este caso, vendrá recogida por un parámetro  $\gamma$  que multiplica al cociente error/varianza condicional, retardado un periodo. El parámetro será generalmente negativo, por tanto, shocks positivos en los rendimientos generarán menor volatilidad que shocks o innovaciones negativas. Dicho de otro modo más formal, el coeficiente de los errores positivo será  $\beta_i(1 + \gamma)$ , y para los negativos tendrá un valor de  $\beta_i(1 - \gamma)$ . Así pues con  $\beta_i > 0$ , un  $\gamma$  positivo implicaría que las innovaciones positivas o noticias “buenas” que llegan al mercado tendrían una influencia mayor que las negativas en la volatilidad de los rendimientos de la serie. Por el contrario, con un  $\gamma$  negativo, implica que los shocks o noticias “malas” que llegan al mercado tienen una influencia superior que los positivos en la volatilidad del rendimiento del activo. Un valor de  $\gamma = 0$  implicaría que los shocks, tanto positivos como negativos, afectarían de la misma forma. Es decir son independientes de su signo.

A continuación vamos a describir el modelo GARCH Umbral, en

terminología anglosajona Threshold GARCH (TGARCH). Este modelo fue propuesto por primera vez por Rabemananjara y Zakoïan en 1992 y al igual que en los modelos anteriores también presenta la posibilidad de recoger las asimetrías sobre la volatilidad. La ventaja de este modelo frente a los anteriores reside en que puede percibir comportamientos cíclicos como respuesta a los diferentes shocks. Una ventaja más directa sobre el modelo EGARCH, es que su especificación es aditiva en vez de exponencial. Se modeliza con desviaciones típicas en vez de varianzas como en el resto de modelos. Es decir, utiliza la desviación estándar condicional como una función lineal de innovaciones y retardos de las propias desviaciones estándar.

$$h_t^{1/2} = \alpha_0 + \sum_{i=1}^q (\beta_i^+ \varepsilon_{t-i}^+ - \beta_i^- \varepsilon_{t-i}^-) + \sum_{j=1}^p \alpha_j \sqrt{h_{t-j}}$$

donde :

$$\varepsilon_t^+ = \text{Max}(\varepsilon_t, 0)$$

$$\varepsilon_t^- = \text{Min}(\varepsilon_t, 0)$$

todos los coeficientes  $\alpha_j$  positivos. En este modelo se observa claramente que en caso de igualdad de coeficientes en el momento  $i$ , es decir,  $\beta_i^+ = \beta_i^-$ , implicaría que la influencia del error cometido (innovaciones) sobre la varianzas condicional en un momento anterior  $t-i$ , no dependería del signo. Es decir, existiría simetría en el efecto de los shocks sobre la varianzas condicional. Dado que este modelo está restringido desde el punto de vista de la positividad de los coeficientes, sólo es capaz de recoger las asimetrías de una forma lineal. Si eliminamos la restricción de no negatividad de los parámetros  $\alpha_j$ , el modelo se planteará de la siguiente forma:

$$h_t^{1/2} = \alpha_0 + \sum_{i=1}^q (\beta_i^+ \varepsilon_{t-i}^+ - \beta_i^- \varepsilon_{t-i}^-) + \sum_{j=1}^p (\alpha_j^+ \sqrt{h_{t-j}^+} - \alpha_j^- \sqrt{h_{t-j}^-})$$

donde:

$$\sqrt{h_{t-j}^+} = \text{Max}(\sqrt{h_{t-j}}, 0)$$

$$\sqrt{h_{t-j}^-} = \text{Min}(\sqrt{h_{t-j}}, 0)$$

En este caso se amplían las posibilidades del modelo para que pueda percibir diferentes

comportamientos de la varianza condicional y no sólo de forma lineal como ocurría en el modelo restringido. Por tanto, este modelo puede percibir respuestas asimétricas de la varianza condicional también cíclicas y de forma más flexible que otros modelos y más sencilla, ya que la asimetría se captura de forma aditiva, mientras que, por ejemplo, la forma exponencial del modelo EGARCH es mucho más compleja. En cambio, como inconveniente se debe destacar que la utilización de la desviación condicional estándar, en vez de la varianza condicional como en el resto del modelo, junto al hecho que puede obtenerse valores negativos de  $h_t^{1/2}$ , puede provocar problemas de su interpretación, ya que llega a desvirtuar el concepto de volatilidad, sobre todo, en el modelo no restringido.

Existe una gran cantidad de variantes y de nuevos modelos Autorregresivos Condicionalmente Heterocedásticos no nombrados anteriormente, como son Modelo ARCH no lineal [Engle y Bollerslev(1986)], ARCH Multiplicativo [Mihoj(1987),Geweke(1986),Pantula(1986)], Modelo GJR [Glosten, Jagannathan y Runkle(1989),Zakoian(1990)], Modelo Autorregresivo de Desviación Standár [Schwert(1990)],... entre otros. De todos ellos vamos a destacar brevemente dos de ellos, el modelo ARCH no lineal y el modelo GJR, dado que son utilizados en bastante bibliografía, sobre todo el modelo GJR.

Empezaremos por el modelo ARCH no lineal, Nonlinear ARCH (NARCH), presentado por Engle y Bollerslev(1986). Su formulación es la siguiente:

$$h_t = \alpha_0 + \alpha_1 |\varepsilon_{t-1}|^\gamma + \beta_1 h_{t-1}$$

Según estudios realizados por Engle y Ng(1993), este modelo posee una curva de impacto de las noticias o innovaciones sobre la varianza condicional que es simétrica. Ahora bien, según los autores, para el caso de  $\gamma < 2$ , la respuesta que tendrá la varianza condicional ante shocks extremos será reducida.

El siguiente modelo, llamado modelo de Glosten, Jagannathan, and Runkle, en nombre a sus autores, añaden una variable dicotómica al modelo GARCH con el fin de que ésta recoja las asimetrías de los impactos de las noticias. La expresión

del modelo es la siguiente:

$$h_t = \alpha_0 + \beta h_{t-1} + \alpha_1 \varepsilon_{t-1}^2 + \gamma S_{t-1}^- \varepsilon_{t-1}^2$$

donde:

$$S_{t-1}^- = 1 \text{ si y solo si } \varepsilon_t < 0$$

$$S_{t-1}^- = 0 \text{ en cualquier otro caso.}$$

En este modelo sólo se recogerá la asimetría negativa en la volatilidad de forma proporcional al cuadrado de las innovaciones negativas, es decir, cuando  $\varepsilon_t < 0$ , de ahí que ésta aparezca al cuadrado .

## **2.4. Modelos Multivariantes**

Como hemos visto en el apartado anterior, los momentos de segundo orden (varianza), que son desconocidos, pueden ser medidos mediante un modelo con una forma funcional para la varianza condicional, de forma simultánea junto con el momento de primer orden, es decir, mediante los modelos ARCH. Posteriormente, Bollerslev (1986) realiza una extensión y generalización del modelo, de una forma similar a la extensión de los modelos ARMA. Del mismo modo que en la literatura econométrica y financiera, se ha realizado una extensión del modelo ARMA para el caso multivariante, pasando a un modelo de vector ARMA (Bollerslev, Engle y Wooldrige, 1988; Engle y Rodriogues, 1989; Engle, Granger y Kraft, 1984; Kaminsky and Peruga, 1990)

En los modelos ARCH, o más general GARCH, multivariantes, la matriz de varianzas-covarianzas dependen de la información presentada en un vector ARMA. Este tipo de modelos son muy utilizados en finanzas, como por ejemplo para la cobertura dinámica de una cartera, para el CAPM,...

La extensión de un modelo GARCH univariante en uno multivariante requiere que la matriz condicional de varianzas-covarianzas de dimensión  $n$ , cuyas variables , es decir, el error o residuo de la estimación, son aleatorias de media cero, que dependen de la información disponible hasta ese momento,  $\psi_{t-1}$ . La característica

principal de los modelos GARCH multivariantes reside, pues, en que el efecto autorregresivo de las varianzas se especifica mediante la matriz de varianzas-covarianzas, que debe ser definida positiva por definición. A continuación, de forma resumida, y siguiendo la clasificación realizada por Navarro y Torró (1998), se procede a la exposición de las parametrizaciones más frecuentes para el modelo Bivariante GARCH, BGARCH (1,1), siendo éste un caso específico de los modelos GARCH multivariantes, fácilmente generalizable a dimensión n. En esta clasificación se destacan tres casos que pasamos a describir brevemente en los siguientes apartados.

#### 2.4.1. Modelo de corrección constante

En este modelo la matriz varianzas-covarianzas será de orden 2 x 2, cuya estructura es la siguiente:

$$\begin{bmatrix} \sigma_{s,t}^2 & \sigma_{sf,t} \\ \sigma_{sf,t} & \sigma_{f,t}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{s,t} & 0 \\ 0 & \sigma_{f,t} \end{bmatrix} \begin{bmatrix} 1 & \rho_{sf,t} \\ \rho_{sf,t} & 1 \end{bmatrix} \begin{bmatrix} \sigma_{s,t} & 0 \\ 0 & \sigma_{f,t} \end{bmatrix} = \begin{bmatrix} \sigma_{s,t}^2 & \rho_{sf,t} \sigma_{f,t} \sigma_{s,t} \\ \rho_{sf,t} \sigma_{f,t} \sigma_{s,t} & \sigma_{f,t}^2 \end{bmatrix}$$

donde  $\rho_{sf,t}$  es el coeficiente de correlación constante,  $\sigma_{sf,t}$  es la covarianza entre el mercado de contado y el de futuro, y  $\sigma_{s,t}^2$  y  $\sigma_{f,t}^2$  son las varianzas del mercado de contado y del de futuro respectivamente (en nuestro caso el Índice IBEX-35 y su correspondiente contrato de futuro). Este modelo se ha utilizado frecuentemente en la literatura financiera para hallar una matriz definida positiva. Para ello se debería realizar previamente un estudio sobre la constancia del coeficiente de correlación, ya que no siempre puede ser factible dicha hipótesis de constancia. En el caso GARCH ( 1, 1 ), el número de parámetros a estimar es muy reducido, siendo de sólo 7.

#### 2.4.2. Modelo Vech

Se trata de un modelo más general que el anterior, puesto que plantea una estructura más dinámica con una matriz varianzas –covarianzas compuesta por una matriz de constantes, otra sobre los residuos o innovaciones y otra sobre las varianzas-covarianzas del periodo anterior. De esta forma se obtiene una matriz varianzas-covarianzas que representa con mayor dinamismo la interacción entre los dos mercados.

El modelo Vech se plantea como sigue:

$$vech(\Sigma_t) = vech(c) + A \times vech(\varepsilon_{t-1}, \varepsilon'_{t-1}) + B \times vech(\Sigma_{t-1})$$

siendo  $vech()$  un operador que extrae de la parte inferior de las matrices simétricas sus

elementos, transformándolos en un vector columna. Ej:  $\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$ . La matriz de

parámetros  $C$  es simétrica y de orden  $2 \times 2$ , mientras que  $A$  y  $B$  son matrices de coeficientes de orden  $3 \times 3$ . Para que en este modelo se obtenga una matriz definida positiva se ha de cumplir que  $A$ ,  $B$  y  $C$  sean semidefinidas positivas, lógicamente esto no tiene porque darse en todos los mercados ni para todos los activos, en cuyo caso el modelo no sería válido. Desarrollando los operadores  $vech()$ , el modelo sigue como sigue:

$$\begin{bmatrix} \Sigma_{11,t} \\ \Sigma_{12,t} \\ \Sigma_{22,t} \end{bmatrix} = \begin{bmatrix} C_{11} \\ C_{12} \\ C_{22} \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} \times \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \times \begin{bmatrix} \Sigma_{11,t-1} \\ \Sigma_{12,t-1} \\ \Sigma_{22,t-1} \end{bmatrix}$$

Como se observa, hay 21 parámetros con lo que se dificulta el proceso de estimación. Para resolver este problema se puede suponer que las matrices  $A$  y  $B$  son diagonales, con lo que al no permitir relaciones cruzadas entre sus elementos estamos asumiendo que cada varianza y covarianza, sólo dependen de sus propios valores y errores de predicción pasados ( ver Bollerslev 1988). En este modelo BGARCH ( 1,1 ), llamado Vech diagonal, sólo se estimarán 9 parámetros obteniendo así una gran reducción de estimaciones.

### 2.4.3. Modelo BEKK

Este modelo es el más general de los tres. Recibe su nombre por las iniciales de los autores que lo propusieron ( Yoshi Baba, Robert Engle, Dennis Kraft y Kenneth Kronen). Para un estudio más completo de dicho modelo puede verse Engle-Kroner (1995), donde se realiza un estudio de los modelos GARCH multivariantes. La

expresión del modelo BEKK, en forma compacta, es la siguiente:

$$\Sigma_t = C'C + A' \varepsilon_{t-1} \varepsilon_{t-1}' A + B' \Sigma_{t-1} B$$

donde C, A y B son matrices de parámetros de orden 2x2 y además C es simétrica,  $\varepsilon_t$  es el vector de residuos  $(\varepsilon_{1t}, \varepsilon_{2t})'$ . En este modelo se requiere la estimación de 11 parámetros. A pesar de ello, presenta unas ventajas claras sobre los modelos anteriores. Por un lado se asegura la obtención de una matriz varianzas-covarianzas definida positiva y por otro, permite que el coeficiente de correlación varíe, dando mayor dinamismo, mientras que en los otros modelos éste permanecía constante.

## **2.5. Estimación de los modelos ARCH**

En toda la literatura financiera aparecen de forma muy frecuente cinco métodos de estimación de los modelos autorregresivos condicionalmente heterocedásticos (ARCH) y sus variantes. Estos métodos de estimación son los siguientes:

- Método de los momentos generalizados
- Máximoverosimilitud
- Cuasimáximoverosimilitud
- Estimación indirecta
- Estimación bayesiana

De todos ellos, sólo los tres primeros son los más utilizados, ya que reducen considerablemente el tiempo de computación. De hecho, son los que suelen estar implementados en los paquetes informáticos comerciales.

La estimación de los momentos generalizados aplicado a los modelos ARCH fue implementada Rich,R., y otros (1991); Glosten,L.R., y otros (1993) Para la estimación de la función máximoverosimilitud y cuasimáximoverosimilitud puede verse Nelson (1990), Engle y Gonzalez-Rivera (1991), Steigerwald (1992) y Linton (1993), entre otros. Para la estimación indirecta, método aplicado a la estimación de la

volatilidad en 1992, puede verse Gallant y Tauchen (1992), Gouriéoux y otros (1993). Por último, para la estimación Bayesiana aplicada a los modelos ARCH puede verse Albert y Chib (1993), Geweke(1994).

## **2.6. Estudio de la relación volumen-volatilidad**

En este capítulo se pretende hacer una revisión de la existencia o no de correlación entre volumen y volatilidad.

En el estudio de Karpoff (1987), en el que resume los principales trabajos empíricos al respecto, llega a la conclusión que si se utiliza los cambios de precios en valores absolutos se encuentra una relación positiva entre volumen y volatilidad, tanto para el mercado continuo como para el de futuro. Sin embargo, si se utiliza los cambios de precios per sé, es decir, sin valor absoluto, esta relación positiva sólo se encuentra para los mercados de contado.

Una explicación de la relación positiva encontrada entre volumen y volatilidad, reside en el hecho de que ambas variables varían como respuesta a la nueva información que va llegando a los mercados. El volumen indicará, pues, la fuerza o debilidad, con la que se mueven los precios del mercado, ya que los agentes / inversores toman sus posiciones, tanto a nivel de precios como en la cuantía de títulos, en función de la información de que disponen.

Para estudiar esta relación se va a seguir la metodología utilizada por Najand y Yung (1991). Ellos utilizan, al igual que en un gran número de estudios sobre volatilidad, el modelo GARCH(1,1), dado que esta especificación es la más adecuada para modelizar una serie de rendimientos que, generalmente, suele ser leptokúrtica y que presenta una larga persistencia en la varianza de la distribución condicional del rendimiento.

Dado que anteriormente en este trabajo ya se ha descrito el modelo GARCH presentado por Bollerslev (1986), aquí describimos directamente el modelo GARCH (1,1) incluyendo la variable volumen en la varianza condicional como realizan, entre otros, Najand y Yung (1991). El modelo queda especificado como sigue:

$$\begin{aligned}
y_t &= \mu + \varepsilon_t \\
\varepsilon_t | \phi_{t-1} &\sim N(0, h_t) \\
h_t &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{t-1} + \delta V_t
\end{aligned}$$

donde  $y_t$  es el rendimiento de los precios expresado como  $\ln(\text{cierre}_t / \text{cierre}_{t-1})$  y  $V_t$  es el volumen. El uso del logaritmo para el cálculo del rendimiento es debido a la no estacionariedad que pueden presentar las series, sobre todo para el contrato de futuro. Todos los coeficientes del modelo se estiman de forma conjunta mediante la función de máximoverosimilitud.

En un estudio de Lamoreux y Lastrapes (1990) para 20 de los títulos más activos del Standard & Poor's, introduciendo la variable volumen en el modelo GARCH (1, 1) como se ha indicado anteriormente, observan que  $\alpha_1$  y  $\beta$  son pequeños y estadísticamente no significantes, ya que el volumen es serialmente correlado. Ellos encuentran que la persistencia del modelo GARCH, medido por la suma ( $\alpha_1 + \beta$ ), no es significativa y, por tanto, los efectos GARCH disminuyen al incorporar el volumen en la regresión. Afirman que el coeficiente positivo y significativo del volumen en todos los títulos de la muestra confirma la hipótesis inicial de una relación positiva entre volumen y volatilidad.

En el estudio de Najand y Yung (1991) para el futuro sobre bonos del tesoro del Chicago Board of Trade (CBOT), al incluir el volumen, los efectos del modelo GARCH permanecen vigentes, siendo éstos fuertemente significativos para todo el periodo muestral que utilizan ( desde 1984 hasta 1989). El hecho que la suma ( $\alpha_1 + \beta$ ) sea muy próxima a la unidad confirma la alta persistencia de la volatilidad pasada en la explicación de la volatilidad actual. En cuanto a la relación positiva entre volumen y la variabilidad de los precios, observan que sólo en dos subperiodos de la muestra confirman dicha relación. Hay que tener en cuenta que, como apuntan Lamoreux y Lastrapes(1990), puede haber problemas de simultaneidad. Éste aparece si el volumen es endógeno del sistema, es decir está correlado con los residuos, en cuyo caso, el proceso de estimación basado en la función de máximoverosimilitud será inconsistente. Para dar solución a este problema de simultaneidad, se utilizará el

volumen retardado considerando el volumen como variable exógena del modelo. Utilizando este método, comentado por Harvey en un estudio en 1989, Najand y Yung(1991) observan que hay un mayor número de coeficientes significativos y positivos de la variable volumen retardada un periodo. Este resultado nos sugiere que sí existe un problema de simultaneidad y que la utilización del volumen retardado es una buena medida para determinar que el volumen es una variable endógena del sistema y que existe una relación positiva entre volumen y volatilidad contemporánea. Sin embargo, en Lamoreux y Lastrapes (1990) la aplicación de esta técnica empeora los resultados, ya que con el volumen contemporáneo ya obtienen que el coeficiente es positivo y significativo para las 20 acciones que utilizan de muestra. Más recientemente, Gallo y Pacini (1997), argumentan que la utilización del volumen retardado no es una buena aproximación del volumen contemporáneo. Presentan otros proxies al volumen como medida de la negociación en un mercado, como por ejemplo, la volatilidad intradía definida como  $IDV_t = \ln \frac{Máximo_t}{Mínimo_t}$  y  $ONI_t = |\log \frac{Apertura_t}{Cierre_{t-1}}|$  que representa la

diferencia de precios ocurrida durante el cierre del mercado. En ambos casos encuentran una reducción de la persistencia del modelo GARCH(1,1) en su estudio para 10 acciones muy negociadas del mercado estadounidense.

Algunos autores como Sharma (1996) realizan un estudio sobre índices bursátiles americanos, evidenciando que la incorporación del volumen al modelo GARCH reduce la persistencia de la volatilidad (aunque en este caso no llega a eliminarla) y por tanto, se demuestra la existencia de relación entre volumen y volatilidad.

Para el caso español, el autor García Montalvo (1999) analiza la persistencia de la volatilidad diaria del futuro sobre el bono nocional e incluye en el modelo tanto el volumen como el número de transacciones. En sus resultados se aprecia que con la inclusión de estas variables se logra reducir la persistencia de la volatilidad.

Como resumen de todos estos trabajos y otros posteriores que relacionan el volumen y la volatilidad indicar que la incorporación del volumen en los modelos GARCH, en la mayoría de casos, elimina la persistencia de la volatilidad y en algunos sólo consigue reducirla, por tanto, la incorporación de esta variable en el modelo, parece

ser indispensable, ya que en cualquier caso, mejora la modelización de la volatilidad y por tanto, podrá valorarse más adecuadamente el activo correspondiente.

## 3. Las Redes Neuronales Artificiales

### 3.1. *Introducción*

Los sistemas neuronales tienen la misión del control de todas las funciones biológicas. Esto es común para todos los animales. Sin embargo sólo en los animales superiores, además cumplen otra función principal relacionada con el comportamiento del individuo y su relación con el entorno a través de la memoria y el pensamiento.

Las redes neuronales artificiales podemos definirlas como una serie de algoritmos matemáticos con la función de encontrar relaciones no lineales entre un conjunto de datos dado. Están basadas en el comportamiento y funcionamiento de las neuronas humanas. También se pueden definir como un nuevo sistema computacional inspirado en fundamentos biológicos (neuronas cerebrales). Las redes neuronales se aplican en diversos campos científicos, y en particular en la economía y las finanzas donde gracias a su capacidad en la estimación no paramétrica, modelización y predicción de series temporales,...han dado muy buenas soluciones a dichos problemas. A pesar del gran boom que tienen en la actualidad las redes neuronales artificiales sus orígenes parten de las primeras décadas del siglo XX. En 1913, S. Rusell planteó un dispositivo hidráulico basado en las redes neuronales, aunque no fue hasta los años cuarenta cuando se plantearon matemáticamente estos modelos. A continuación vamos a realizar una breve reseña histórica de la evolución de las redes neuronales.

1936- Alan Turing. Fue el precursor en el estudio del cerebro humano asimilándolo a la computación.

1943- Warren McCulloch y Walter Pitts, neurofísico y matemático respectivamente. Propusieron por primera vez un modelo matemático de red neuronal basado en una teoría que constituía la forma de trabajar de las neuronas. Uno de los problemas principales de este modelo fue que no tenía capacidad de aprendizaje.

1949- Donald Hebb. Precursor en la explicación del aprendizaje de las neuronas, elemento indispensable para la inteligencia humana. La idea principal es que el

aprendizaje se produce ante cambios que hacen que una neurona se active. Desarrolló un algoritmo de aprendizaje denominado aprendizaje Hebbiano.

1950- Karl Lashley. Realizó diversas pruebas en las que determinó que la información no es almacenada en el cerebro de forma centralizada si no que se distribuye por todo él.

1954- Marvin Minsky. Obtuvo los primeros resultados con las redes neuronales junto con Dean Edmons que crearon una máquina con 40 neuronas, aplicando los modelos y teorías de McCulloch y Pitts, incorporando además el aprendizaje Hebbiano.

1956- Congreso de Dartmouth.

1957- Frank Rosenblatt. A partir de los trabajos de Albert Uttley (1956) sobre su modelo llamado Informon, Rosenblatt desarrolla el Perceptron, uno de los modelos más conocidos a la largo de la historia. Se utiliza principalmente en el reconocimiento de patrones. Este modelo es capaz de generalizar reconociendo patrones similares a los que había aprendido previamente. Sin embargo una de sus grandes limitaciones es su incapacidad para realizar clasificaciones de clases no separables linealmente.

1959- Frank Rosenblatt. Escribe el libro “Principios de Neurodinámica” en el que indicaba que el aprendizaje del Perceptron converge hacia un estado finito, bajo ciertas circunstancias.

1960- Bernard Widroff/Marcian Hoff. Desarrollan el modelo Adaline (Adaptative Linear Elements). Éste fue el primer modelo con una aplicación a un problema real. Fúe usado para la eliminación de los ecos en las líneas telefónicas y se utilizó durante varias décadas.

1969- Marvin Minsky / Seymour Papert. En su libro “Perceptrons”, plantean las limitaciones de las capacidades del Perceptron. Así pues, demostraron matemáticamente que no era capaz de resolver problemas relativamente fáciles como el aprendizaje de una función no lineal, funciones muy empleadas en la computación y en muchos problemas reales.

1974- Paul Werbos. Desarrolla el algoritmo de aprendizaje de propagación hacia atrás (backpropagation).

1977- Stephen Grossberg. Plantea la teoría de resonancia adaptada. Se trata de una arquitectura de red que se diferencia de las planteadas hasta la fecha y es capaz de simular la memoria a largo y corto plazo.

1982-1985- John Hopfield. Escribe su libro “ Computación neuronal de decisiones en problemas de optimización”. Describe un sistema (modelo de Hopfield) basado en sistemas olfativos, obteniendo un modelo construido por elementos interconectados que buscan un estado interno de mínima energía.

1986- David Rumelhart /G. Hinton. Redescubren y desarrollan el algoritmo de propagación hacia atrás (backpropagation).

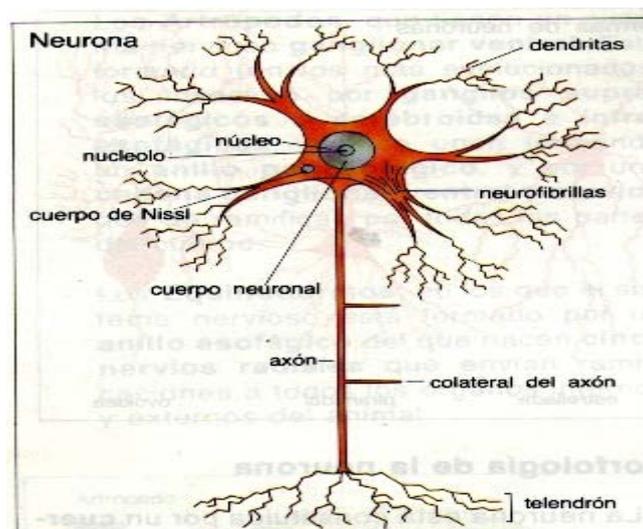
En la actualidad existen numerosos trabajos relacionados con las redes neuronales aunque pueden destacarse: la Teoría de la Resonancia Adaptativa (ART) desarrollada por Gail Carpenter y Stephen Grossberg en 1986, las máquinas de Boltzman y Cauchy estudiadas por Hinton y Sejnowski en 1986, los mapas Autoorganizados de Teuvo Kohonen (1984),...

### **3.2. Fundamentos Biológicos**

Las redes neuronales artificiales dado que se base fundamental parte de la estructura y funcionamiento de las neuronas biológicas, vamos a estudiar algunos de los fundamentos biológicos que existen en las mismas. Una de las características principales que las redes neuronales artificiales intentan simular es su sistema de interconexión de procesos mediante las neuronas, formando redes, en las que dichas conexiones se transforman en otras estructuras y funciones a largo y corto plazo. Dichos cambios son los que provocan el aprendizaje y la memoria. Por tanto, será las neuronas el elemento principal sobre las que las redes neuronales artificiales se fijarán para simular su comportamiento, estructura y modificaciones para poder realizar la función de aprendizaje.

En el cerebro humano existen aproximadamente unas  $10^{11}$  neuronas y se estima que existen unas  $10^{15}$  interconexiones. Tanto el tamaño como la estructura de la red varía a lo largo del tiempo. La mayoría de neuronas biológicas tienen una serie de partes características y principales como son el núcleo celular, axón, dendritas.

El núcleo celular es el responsable de procesar las señales que en forma de impulsos le llega de otras células a través de las dendritas, generando un nuevo impulso hacia otras células a través del axión. Es decir, la neurona obtiene los datos por las dendritas en forma de impulsos eléctricos. Esto hace que la neurona se active y envíe un impulso eléctrico a otras neuronas mediante el axón. Este impulso libera neurotransmisores que llegan a otras células. Las neuronas no se encuentran unidas unas con otras. Existen un pequeño espacio entre ellas, la conexión o sinapsis se realiza pues mediante los neurotransmisores.



**Figura 3-1: Gráfico neurona**

Las transmisiones de una neurona se producen por impulsos eléctricos que modifican la química de los neurotransmisores. En el proceso de aprendizaje se produce con estas modificaciones.

Lógicamente existen diferencias entre los estudios sobre las neuronas biológicas y las artificiales. Las propiedades físicas de las neuronas y sus conexiones son muy complejas en comparación con los modelos de redes neuronales artificiales. Las estructuras de las RNA son biológicamente imposibles, así como las funciones y

algoritmos de aprendizaje. A pesar de estas grandes diferencias, las RNA simulan muy aceptablemente el comportamiento de las neuronas humanas y se obtienen resultados muy parecidos a los procesos biológicos. Es decir, son capaces de aprender de la experiencia, de generalizar a partir de casos anteriores a nuevos casos, discernir entre características principales a partir de datos de entrada, ...

A continuación vemos algunas de las ventajas que poseen las RNA y que comparten con las neuronas biológicas.

Las RNA pueden realizar aprendizaje adaptativo, es decir, tienen la capacidad de aprender mediante un entrenamiento o experiencia, dado que pueden discernir patrones mediante ejemplos, por lo que no es necesario implementar modelos a priori ni hace falta asignar funciones de distribución probabilística. Son sistemas autoadaptativos por la capacidad de las neuronas de adaptarse y ajustarse. Además son dinámicos, es decir, pueden cambiar constantemente para adaptarse a nuevas condiciones. Durante el proceso de aprendizaje las ponderaciones de las conexiones de las neuronas varían hasta obtener un resultado específico.

Las RNA son autoorganizadas, es decir, pueden crear su propia organización de la información que se le presenta durante el proceso de aprendizaje. La autoorganización consiste en la modificación de la red neuronal para conseguir un objetivo determinado. Esto provoca que las redes puedan responder adecuadamente ante patrones o datos sobre los que no tenía experiencia previa (capacidad de generalización).

La tolerancia a fallos de las RNA se basa en que la destrucción parcial de una red lleva a una modificación en su estructura, pero pueden permanecer algunas de sus capacidades intactas. Una red aunque tenga fallos en alguna de sus neuronas, seguirá funcionando, aunque, es verdad, se verá influenciado por este deterioro. Se puede distinguir la tolerancia a fallos relacionada con la distorsión o ruido en los datos de entrada, incluso con datos incompletos, en los que la red es capaz de distinguir patrones.

Las RNA pueden programarse para su ejecución en paralelo pudiendo trabajar, incluso en tiempo real.

### 3.3. Estructura y Elementos de las RNA

#### 3.3.1. Las neuronas artificiales

Como se ha dicho anteriormente, las RNA son modelos que intentan simular el comportamiento de las neuronas, por tanto, serán éstas las principales unidades de proceso, que tendrán como mínimo un canal de entrada de la información y uno de salida. Por tanto, cualquier red neuronal tendrá una o más neuronas que se distribuirán a lo largo de una estructura específica para la resolución de un problema planteado con los datos disponibles.

Existirán tres tipos de neuronas. Las primeras serán las que reciben los estímulos externos y serán las que capten la información de entrada a la red. En segundo lugar, se encuentran el grupo de neuronas que se encuentran en el intermedio de la estructura de la red y se encargan del procesamiento de la información y aprendizaje sobre la misma, siendo este grupo neuronas ocultas. Por último, se presentan las neuronas que, una vez realizado el proceso de aprendizaje, presentan la información resultante dando respuesta al problema planteado.

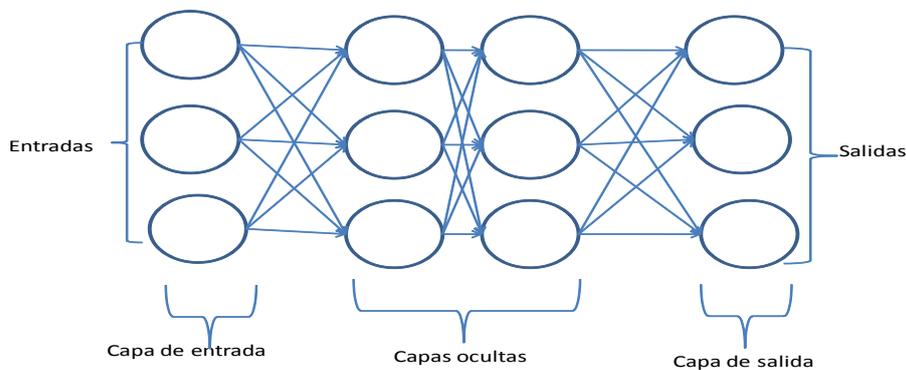


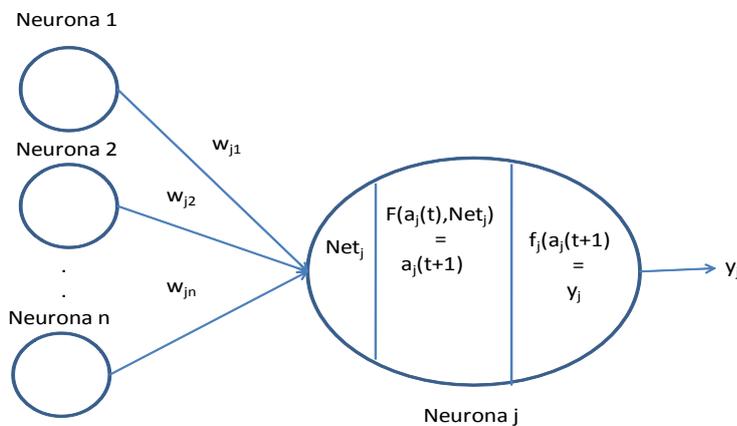
Figura 3-2: Esquema estructura red neuronal

Podemos definir como capa o nivel a un conjunto de neuronas que tienen la misma fuente de entrada de datos (pueden ser otra capa de neuronas) cuyas salidas se encaminan a una misma dirección (puede ser otra capa de neuronas).

Cada neurona está caracterizada por un estado de activación  $a_i(t)$  asociado. Existe una función de salida  $f_i$  (también llamada función de transferencia) que modifica dicho estado y lo transforma en una señal de salida  $y_i$ . Estas señales serán transmitidas a otras neuronas mediante las conexiones o sinapsis mediante ponderaciones  $w_{ij}$  según alguna regla determinada. La agregación de cada señal de salida por sus ponderaciones que han llegado a la neurona  $j$ -ésima se combinan entre ellas de tal forma que constituyen la llamada entrada total  $Net_j$ , o regla de propagación, que se trata del cálculo del valor neto que realiza esta neurona con las diferentes señales que han entrado en la misma.

$$Net_j = \sum_i y_i w_{ij}$$

La función de activación o de transferencia  $F$  estima el nuevo estado de activación  $a_j(t + 1)$  de una neurona en un momento posterior teniendo en cuenta el estado de activación anterior  $a_j(t)$  y la entrada total calculada  $Net_j$ .



**Figura 3-3: Esquema funcionamiento transmisión**

Cuando la regla de propagación o  $Net_j$  supera un cierto número (llamado umbral  $\Phi$ ) entonces la neurona se activa y el número resultante se incorpora a la función de transferencia que dará como resultado las señales de salida  $y_i$ .

### 3.3.2. Estado de Activación

Como se ha indicado anteriormente la neurona biológica puede estar activada (excitada) o permanecer inactiva (no excitada), esto es lo que hemos denominado su estado de activación  $a_i(t)$ . En el caso de las neuronas artificiales también pueden tener estos dos estados de activación, pero algunas pueden tomar cualquier valor dentro de un intervalo determinado. Es decir, los valores de activación pueden ser continuos o discretos. Pueden ser limitados o ilimitados. Si se trata de valores discretos suele tomarse un conjunto pequeño de valores, aunque lo habitual son la elección de valores binarios. Con los números binarios, cuando una neurona se encuentra en un estado activo se representa por 1, mientras que un 0 representa un estado pasivo de la neurona que estará en reposo y no realizará ningún impulso de transmisión de valores o información a otra neurona. En algunos modelos de RNA se consideran conjuntos continuos de valores de estado de activación. Así podemos tener, por ejemplo, el intervalo  $[0,1]$  o también  $[-1,1]$  muy utilizado en funciones sigmoideas.

### 3.3.3. Función de Activación o de Transferencia

Esta función sirve para combinar las entradas de información de la neurona con su estado de activación actual para así provocar un cambio en la misma y dar salida o no a nueva información, en función del grado de activación que se produzca. La función  $F$  de transferencia produce nuevos cambios en la neurona a partir del estado actual de la neurona  $a_j(t)$  y la combinación de los pesos de las conexiones de la neurona con las entradas de información a la misma, lo que hemos llamado  $Net_j$ . Con todo esto se produce el nuevo estado de activación de la neurona que hemos llamado anteriormente  $a_j(t + 1)$ . Por tanto, escribir que  $a_j(t + 1) = F(a_j(t), Net_j)$  tal y como aparece en la figura anterior.

La mayoría de ocasiones la función de activación  $F$  es la llamada función identidad, esto implica que no se tiene en cuenta el estado de activación actual de la neurona y, por tanto, la información que pasa a la función de salida  $f_j$  que se produce en la neurona será exclusivamente el valor de  $Net_j$ . Es decir, en este caso se cumple que  $y_j = f(Net_j) = f(\sum_{i=1}^n w_{ij} y_i)$ . Además esta función de activación o de transferencia

no está centrada en origen dado que existe un ligero desplazamiento provocado por las características internas de cada neurona y que son diferentes a cada una de ellas. Es lo que hemos denominado el umbral de activación  $\Phi$  para una neurona. Por tanto, la función de activación se reescribe como  $y_j = f(Net_j - \Phi) = f(\sum_{i=1}^n w_{ij} y_i - \Phi)$ . Es decir cuando el valor de entrada neta  $Net_j$  sea superior al umbral  $\Phi$  de esa neurona, ésta se activará.

### 3.3.4. Función de salida

Esta función es la última que presenta la neurona y su tarea no es otra que determinar el valor de salida que transfiere a otras neuronas que se encuentren vinculadas con ella. Como se ha indicado anteriormente, si este valor es inferior al umbral, la neurona no se activará y no producirá ningún resultado a transferir. Dado que los datos de entrada a la neurona están limitados a los valores que pueden tomar los estados de activación de la misma, como dijimos anteriormente, pueden ser valores continuos, binarios,...los valores de salida de la neurona estarán también en dichos rangos.

Los diferentes valores de salida que puede ofrecer la neurona vienen en función del tipo de función de activación o de transferencia  $f$  que se trate. Así pues, destacamos los siguientes tipos principales de función de activación:

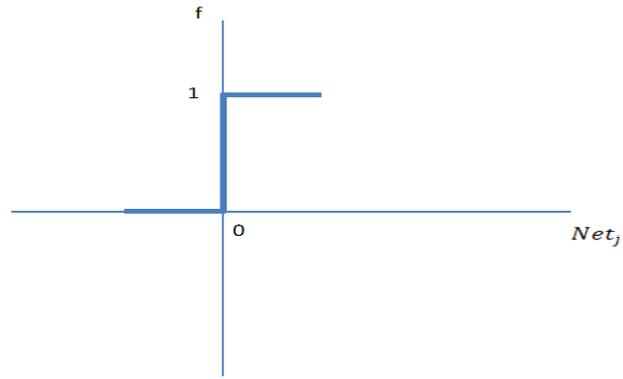
#### a) Función de activación escalón

Este tipo de función se utiliza generalmente cuando las neuronas tienen un estado de activación binario (0,1). La neurona se activará cuando el valor neto de entrada  $Net_j$  sea superior o igual al umbral de la neurona y no se activará en caso contrario

$$y_j = f(Net_j - \Phi) \text{ toma valor 1 cuando } Net_j \geq \Phi$$

$$\text{toma valor 0 cuando } Net_j < \Phi$$

Por ejemplo, en el caso que el umbral fuera cero tendríamos la siguiente figura:

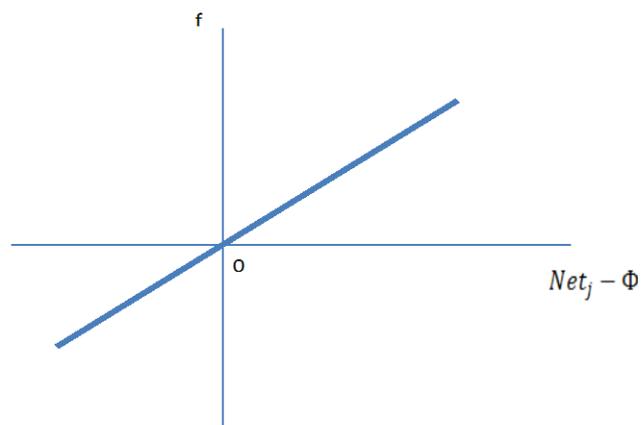


**Figura 3-4: Función de activación escalón**

b) Función de activación lineal o identidad

Esta función puede obtener cualquier número real y el estado de activación coincide con el valor de la entrada neta  $Net_j$ .

$$y_j = f(Net_j - \Phi) = Net_j - \Phi$$



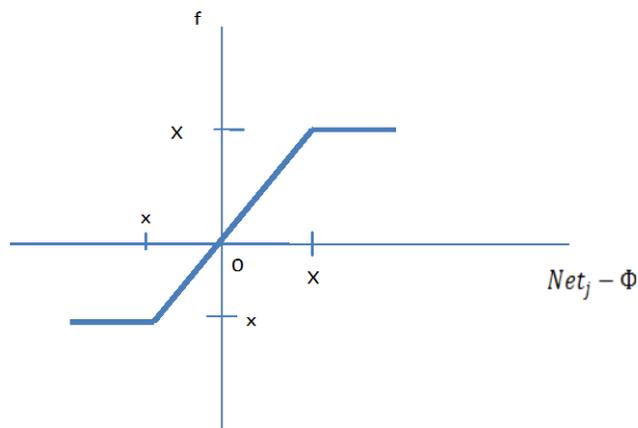
**Figura 3-5: Función de activación lineal o identidad**

c) Función de activación lineal-mixta

Se trata de una variante de la función de activación lineal. En este caso los valores de estado que puede obtener la neurona se encuentran fijados dentro de un intervalo de números reales.

$$y_j = f(\text{Net}_j - \Phi) \text{ toma valor } x \text{ cuando } \text{Net}_j \leq x + \Phi$$
$$\text{toma valor } \text{Net}_j - \Phi \text{ cuando } x + \Phi < \text{Net}_j < X + \Phi$$
$$\text{toma valor } X \text{ cuando } \text{Net}_j \geq X$$

Su gráfica será la siguiente:



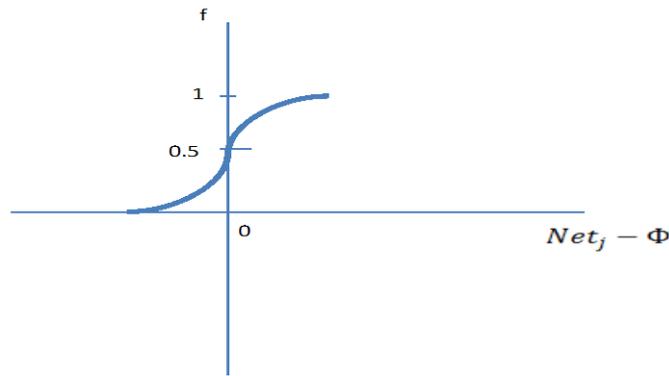
**Figura 3-6: Función de activación lineal-mixta**

d) Función de activación sigmoideal

Se trata de una función de activación continua dentro del intervalo del eje real, con un valor de salida continuo entre cero y uno.

$$y_j = f(\text{Net}_j - \Phi) = \frac{1}{(1 + e^{-(\text{Net}_j - \Phi)})}$$

Teniendo su gráfica la siguiente figura:



**Figura 3-7: Función de activación sigmoide**

### **3.3.5. Mecanismos o Reglas de Aprendizaje**

El proceso de aprendizaje viene determinado por la variación o modificación de los estados de activación de cada neurona. En las neuronas biológicas se establece que el conocimiento se encuentra en la sinapsis, es decir, está relacionado más con los valores sinápticos de las conexiones neuronales que con las propias neuronas. Asimilando esta cuestión a las redes neuronales artificiales, podemos decir que el conocimiento será representado por los valores de cada peso  $w_{ij}$  de las conexiones entre neuronas. Entonces el proceso de aprendizaje provoca cambios en los estados de activación de las neuronas y en sus conexiones, es decir, variando los pesos la neurona va aprendiendo.

La neurona irá aprendiendo realizando cálculos para encontrar una solución correcta para los datos de entrada. Este será el proceso de entrenamiento. Los cambios en el valor de los pesos durante el aprendizaje lo que hace es simular el comportamiento de las neuronas biológicas que están constantemente creando, modificando y destruyendo conexiones entre neuronas. Así, en el caso de las RNA, las neuronas crearán una nueva conexión cuando el valor de su peso sea mayor que cero y se destruirá dicha conexión cuando el peso tome valor cero.

El proceso de aprendizaje finaliza cuando los pesos que tienen cada

conexión de cada neurona permanecen ya constantes sin ninguna variación  $\frac{dw_{ij}}{dt} = 0$ . Por tanto, la clave del aprendizaje residen en qué criterios se sigue para la modificación de cada peso asignado a cada conexión para que la neurona, y la red, aprenda y obtenga nuevas informaciones.

Una de las clasificaciones de los mecanismos de aprendizaje de las RNA son las de RNA con aprendizaje supervisado y RNA con aprendizaje no supervisado. La diferencia fundamental entre ambas es que en el caso de aprendizaje supervisado existe un agente externo (supervisor) con controla el mismo proceso de aprendizaje de las neuronas y de la red.

Otra de las clasificaciones que se hacen es si la red puede o no aprender mientras está en funcionamiento. Es decir, se llama aprendizaje on line si la red puede aprender mientras está en funcionamiento. Es este caso la red no distingue entre la fase de entrenamiento y de operatividad de la misma. Los pesos irán modificándose constantemente de forma dinámica mientras que la neurona consiga nueva información. Cuando la red tiene un mecanismo de aprendizaje off line, la red deja de estar en funcionamiento cuando se encuentra en la fase de entrenamiento, utilizando un conjunto de datos de entrenamiento. Una vez entrenada la red, entrará en funcionamiento y utilizará otro conjunto de datos para realizar test o pruebas sobre el aprendizaje efectuado. En este caso, los pesos de las neuronas quedan fijados definitivamente una vez concluida la fase de entrenamiento.

Con carácter general los cambios en los pesos se producen a partir del peso existente. Es decir, el nuevo peso es igual al existente más la variación o cambio que se produce en ese instante.

$$w_{ij}(t + 1) = w_{ij}(t) + \Delta w_{ij}(t)$$

Donde  $t$  es el tiempo de referencia durante el aprendizaje,  $w_{ij}(t + 1)$  será el nuevo valor del peso y  $w_{ij}(t)$  el valor anterior del peso.

A continuación vamos a describir los diferentes tipos de aprendizaje de RNA siguiendo la distinción entre aprendizaje supervisado y no supervisado.

### 3.3.6. Aprendizaje supervisado

Como se ha indicado anteriormente, el aprendizaje supervisado requiere de un agente externo o supervisor que determina, mediante un entrenamiento controlado, las respuestas que debe dar la RNA a partir de la entrada de datos dada. Este tipo de entrenamiento controlado puede realizarse de tres formas diferentes:

#### A) Aprendizaje por corrección de error

Este primer tipo de entrenamiento supervisado consiste en que el supervisor va ajustando los pesos de las conexiones de las neuronas en función de las diferencias que se van produciendo entre los valores que se desean obtener y los que realmente la RNA ha obtenido.

Un ejemplo simple de este tipo de aprendizaje podría ser el siguiente algoritmo:

$$\Delta w_{ij} = \alpha y_i (d_j - y_j)$$

Donde

$\Delta w_{ij}$ : es la variación del peso (o variación del estado) de la conexión de la neurona i y j

$y_i$ : Valor de salida de la neurona i

$d_j$ : es el valor de salida deseado para la neurona j

$y_j$ : valor de salida de la neurona j obtenido realmente

$\alpha$ : es el factor de aprendizaje que tendrá un valor entre 0 y 1, y que determina la velocidad de aprendizaje de la neurona.

Uno de los primeros modelos que utilizaron este tipo de aprendizaje lo podemos encontrar en la regla de aprendizaje del Perceptron, desarrollado como se indicó en el apartado de referencias históricas, por Rosenblatt en 1958.

Otros algoritmos basados en este tipo de aprendizaje que han evolucionado son, por ejemplo, el propuesto por Widrow y Hoff en 1960, denominado regla delta o del mínimo error cuadrado, que mejora la rapidez de aprendizaje de la RNA. Este algoritmo se aplicó principalmente a las RNA ADALINE y MADALINE, que son redes adaptativas lineales y múltiples, respectivamente.

Uno de los principales algoritmos más extendidos y utilizados

relacionados con este tipo de aprendizaje supervisado es el de la regla delta generalizada o error backpropagation. Se trata de una generalización del algoritmo propuesto por Widrow y Hoff. En este caso, este algoritmo generalizado, a diferencia del anterior, es que puede utilizarse en estructuras de red multicapa, lo que le da un mayor poder de aprendizaje. Se basa en estructuras de conexiones hacia adelante con múltiples capas y capas ocultas intermedias.

#### B) Aprendizaje por refuerzo

El aprendizaje por refuerzo se basa en que el supervisor simplemente indica a la RNA una señal de refuerzo si la salida obtenida es satisfactoria (+1) o en caso contrario (-1) penalizando así dicha salida. En función de estas señales (mecanismo probabilístico) los pesos de las neuronas irán variando. Este tipo de aprendizaje es mucho más lento que el anterior, ya que se basa en que no se dispone de la información completa referente al comportamiento deseado de la red ante una entrada de datos.

Algunos de los algoritmos basados en este tipo de aprendizaje son: algoritmo lineal con recompensa y penalización (Linear Reward-Penalty) propuesto por Narendra y Thathacher en 1974, ampliado posteriormente en 1985 por Barto y Anandan proponiendo el algoritmo asociativo con recompensa y penalización (Associative Reward-Penalty) que se aplica en redes con conexiones hacia adelante de dos capas y que tienen neuronas de salida con función de activación de tipo estocástico.

Otro tipo de algoritmo conocido es el de Barto, Sutton y Anderson, que en 1983 presentan Adaptive Heuristic Critic, utilizado en redes de tipo feedforward con tres capas diseñadas de tal forma que la propia red sea capaz de generar las señales de refuerzo para aplicarlas a las neuronas de salida de la red.

#### C) Aprendizaje estocástico

Este tipo de aprendizaje supervisado se basa principalmente en provocar cambios aleatorios en el valor de los pesos de las conexiones de las neuronas para, posteriormente, realizar una evaluación de los efectos producidos a partir del objetivo deseado y de una distribución de probabilidad determinada.

En el aprendizaje estocástico una vez realizado un cambio en los pesos, se determina si el comportamiento de la red es próximo al deseado se aceptaría dicho

cambio. En caso contrario, se aceptará el cambio en función de una distribución de probabilidad establecida por el supervisor.

Una de las redes más conocidas que utilizan este sistema de aprendizaje supervisado es la denominada Máquina de Boltzmann, propuesta por Hinton, Ackley y Sejnowski en 1984, y combina este aprendizaje con el aprendizaje Hebbiano (aprendizaje no supervisado que veremos posteriormente) o también lo combina con el aprendizaje supervisado por corrección de error. Se trata de un tipo de red que siempre tiene neuronas ocultas que, mediante ajustes probabilísticos, pueden introducir un ruido en la fase de aprendizaje que irá disminuyendo conforme pase el tiempo de aprendizaje con el objetivo de evitar resultados de mínimos relativos locales, intentando buscar así el mínimo global (cuando el resultado de salida se aproxima al resultado deseado).

Otra red basada en este tipo de aprendizaje es la Cauchy Machine propuesta por Szu en 1986. Se trata de una mejora de la máquina de Boltzmann, haciendo más rápido el aprendizaje en la búsqueda de ese mínimo global.

### **3.3.7. Aprendizaje no supervisado**

Este tipo de aprendizaje se basa en que no es necesario un supervisor para ajustar los pesos de las conexiones de las neuronas. Se trata de un aprendizaje autosupervisado y la red no tiene información referente a si las salidas generadas son o no correctas. Con este tipo de aprendizaje, las RNA encuentran características, correlaciones, ... en los datos de entrada. Dado que no se le indica a la red qué salida es la correcta, en función del tipo de estructura y del algoritmo de aprendizaje utilizado variará la interpretación de las salidas que proporcione.

Así pues, podemos encontrarnos con salidas que buscan similitudes con los datos de entrada, otras pueden realizar clustering o establecer categorías indicando las salidas a qué categoría pertenece los datos de entrada, siendo la red la que genera dichas categorías en función de las correlaciones que detecta entre los datos. Otras redes lo que hacen es un mapeo de características (features mapping), siendo las neuronas de salida un mapa topográfico de las características de los datos de entrada.

Se pueden distinguir dos tipos principales de aprendizaje no supervisado o autosupervisado:

#### A) Aprendizaje Hebbiano

Este tipo de aprendizaje fue propuesto por Hebb en 1949 y decía que: *“Cuando un axón de una celda A está suficientemente cerca como para conseguir excitar una celda B y repetida o persistentemente toma parte de su activación, algún proceso de crecimiento o cambio metabólico tiene lugar en una o ambas celdas, de tal forma que la eficiencia de A, cuando la celda a activar es B, aumenta”*. Hebb denomina celdas, a un conjunto de neuronas muy interconectadas con una estructura muy compleja. La eficiencia, en este caso, se identifica con la intensidad de las conexiones de las neuronas, es decir, el peso.

El aprendizaje Hebbiano consiste en el ajuste de pesos de las conexiones de las neuronas conforme a la correlación existente entre los valores de activación (salidas) de las dos neuronas conectadas.

$$\Delta w_{ij} = y_i y_j$$

Donde  $\Delta w_{ij}$  es la variación en el peso en la conexión de la neurona i y j;  $y_i$  es la salida de la neurona i;  $y_j$  salida de la neurona j.

Esto indica que si las dos neuronas con activas, se produce un reforzamiento en la conexión entre las dos neuronas. Por el contrario, cuando una está activa y la otra no, se produce un debilitamiento de la conexión, es decir, se penaliza dicha conexión. Como las modificaciones de los pesos se basan en las salidas (estados de activación) de las neuronas ante la información entrante en cada una de ellas, el aprendizaje es no supervisado.

Este tipo de aprendizaje se utiliza en la Red Hopfield, propuestas por este autor en 1982, red muy conocida e implementada en muchos circuitos integrados. Posteriormente también se ha utilizado en la red Bidirectional Associative Memory propuesta por Kosko 1988,...

Una versión de este tipo de aprendizaje hebbiano es el hebbiano diferencial y se basa en utilizar la correlación entre las derivadas en el tiempo de las funciones de activación de las neuronas. Este tipo de aprendizaje se utiliza en la Adaptive Bidirectional Associative Memory propuesta igual que la anterior, por Kosko.

#### B) Aprendizaje competitivo y cooperativo

En este tipo de aprendizaje las neuronas compiten y cooperan entre ellas para conseguir un objetivo dado. Ante la información de entrada en la red busca que las neuronas de salida de la misma obtengan su máximo valor de respuesta posible. Esta competición y cooperación se da en todas las neuronas y en todas las capas de la red. Cuando sean competitivas, las conexiones entre neuronas tendrán conexiones recurrentes de autoexcitación (Activación) y conexiones de inhibición, cuando sean neuronas próximas. Cuando sean de tipo cooperativo las conexiones serán de activación.

Con este tipo de aprendizaje se intenta realizar categorías (clustering) con los datos de entrada a la red. Cuando los datos sean similares serán clasificados en la misma categoría y serán activadas la misma neurona de salida. Las distintas categorías son creadas por la propia red neuronal, dado que estamos ante un aprendizaje no supervisado, en base a las correlaciones y características existentes entre los datos de entrada.

Uno de los tipos de redes más conocidos que utilizan este tipo de aprendizaje es el desarrollado por Kohonen en 1988, denominado Learning Vector Quantization, que se aplica en redes con dos capas de tipo feedforward.

Otros modelos con tipo de aprendizaje competitivo son los ART1 y ART2 (Adaptive Resonance Theory) desarrollado por Carpenter y Grossberg en 1986. Se trata de una red de dos capas de tipo feedforward/feedback. Este tipo de construye un prototipo con la información que recibe de entrada y genera una salida otro prototipo o ejemplo que aglutina todas las informaciones del mismo tipo o categoría.

### **3.3.8. Relación entre la información de entrada y salida de la RNA**

Como se ha indicado las neuronas conforman un sistema que almacena información aprendida y queda registrada en la misma distribuyéndose por toda la red en función de los pesos asociados las conexiones existentes entre las mismas. Puede decirse, que una red neuronal goza de cierto tipo de memoria de tipo asociativo, en el que la red ofrece una respuesta ante la entrada de información a la misma.

Existen dos mecanismos básicos para la asociación entre la información de entrada y salida de la red. El primero, heteroasociación, la información es aprendida y asociada por pares de datos. Para este tipo de red se necesitan como mínimo dos capas en la estructura, una para obtener la información de entrada y otra para la salida de la información asociada. Este tipo de redes pueden tener las conexiones de las neuronas hacia adelante (feedforward), utilizadas en las redes Perceptron, Backpropagation,..); pueden tener conexiones hacia atrás feedforward/feedback como las redes ART..., también conexiones laterales como las redes Cabam,

El segundo mecanismo de asociación entre la información es la autoasociativa, donde la red aprende algunas informaciones y cuando se le presenta nueva información crea autocorrelaciones entre esta nueva información y la ya aprendida dando como respuesta de salida aquella que más se parezca a las almacenadas. Este tipo de red sólo requiere una capa de neuronas como mínimo en su estructura, en este caso la red comienza a retener y a aprender de los datos de entrada para después ofrece la respuesta autoasociada a la misma. Si se quisiera mantener tanto la información de entrada como la de salida será necesario ir añadiendo capas a la estructura. Las redes más utilizadas con este tipo de conexiones es la de Hoppfield. Suelen utilizarse para filtrar información y reconstrucción de datos, eliminando el ruido o distorsiones existentes en las informaciones de entrada. También pueden utilizarse para problemas de optimización.

### **3.4. RNA con conexiones hacia adelante: Backpropagation**

La red neuronal Backpropagation fue propuesta por Rumelhart, Hinton y Williams en 1986 y trata de construir una red neuronal con más niveles de neuronas que las que se había utilizado con la red Perceptron de Rosenblatt. Se trata de una red que utiliza el aprendizaje supervisado basado en la regla delta o propagación del error hacia atrás, ya visto anteriormente. En el caso de la red Backpropagation este algoritmo de aprendizaje se generaliza para poder utilizarse con una red con más de dos capas.

El funcionamiento de este tipo de red se realiza en dos fases. Inicialmente, la información entra en las neuronas de la primera capa y genera una asociación de los datos entrada-salida a pares. En segundo lugar, se produce una propagación al resto de neuronas del resto de capas y va comparando las diferentes salidas de las neuronas con la salida que se desea obtener, calculando el error que comete en el aprendizaje. Posteriormente, estos errores de cada neurona se transmiten hacia atrás desde las neuronas de salida hasta las neuronas de las capas anteriores y directamente conectadas con la salida de la red. De esta forma se determina la contribución de cada neurona en el error total. Con esta nueva información los pesos de las neuronas implicadas van variando, para que cuando los datos que se presenten de nuevo a la red sean similares el resultado o salida que generen será más próximo al resultado deseado, disminuyendo el error.

Este tipo de red tiene la importancia en su capacidad de autoadaptar los pesos de las neuronas de las capas intermedias que incrementan el potencial de aprendizaje sobre la relación entre los patrones que genera inicialmente la red con las salidas correspondientes. Se trata de la capacidad de generalización, en la que la red es capaz de dar salidas y soluciones muy favorables ante informaciones nuevas sobre las que no había entrenado anteriormente.

#### **3.4.1. El Aprendizaje en la Red Backpropagation: La Regla Delta Generalizada**

Como se indicó en el apartado de los algoritmos de aprendizaje, el algoritmo de aprendizaje que se utiliza en la red Backpropagation, está basado en el propuesto por Widrow en 1960, la llamada regla delta, que para este caso se generaliza

(regla delta generalizada) para poder utilizarse con una red con conexiones de neuronas hacia adelante (feedforward) que, además, sus funciones de activación son de tipo continuo, bien lineal o sigmoideal (como ya vimos),llegando así al algoritmo de retropropagación (backpropagation).

El mecanismo de aprendizaje de la delta generalizada es el mismo que el utilizado en los ajustes de los pesos de las neuronas para las redes Perceptron y Adaline. Es decir, los pesos van variando a proporción de la delta (diferencia entre la salida deseada y la obtenida realmente).

$$\Delta w_{ij}(t + 1) = \alpha \delta_{pj} y_{pi}$$

Donde:

$\Delta w_{ij}(t + 1)$  : es la variación del peso entre la neurona i y j que se produce una vez recibe un impulso que las activa.

$\alpha$ : es el factor de aprendizaje que tendrá un valor entre 0 y 1, y que determina la velocidad de aprendizaje de la neurona. Será un valor constante.

$y_{pi}$ : salida de la neurona i bajo el patrón de aprendizaje p.

$\delta_{pj}$ : es la delta o diferencia entre la salida deseada y la obtenida realmente.

La diferencia entre la delta y la delta generalizada estriba precisamente en el valor concreto de  $\delta_{pj}$  que se obtiene. En el caso de la delta ya vimos que la diferencia era  $(d_j - y_j)$ . Es decir, la diferencia entre el valor deseado de la neurona j y el valor realmente obtenido por la neurona j. Sin embargo, para el caso de la regla delta generalizada, la delta se define como:

$$\delta_{pj} = (d_{pj} - y_{pj})f(net_j)$$

Donde :

$d_{pj}$ : es la salida deseada de la neurona j para el patrón p

$y_{pj}$ : es la salida de la neurona j mediante el patrón p

$f(net_j)$ : es la función de entrada neta de información que recibe la neurona j.

En la delta generalizada, cuando una neurona no sea de salida de la red, es decir, sea intermedia, el error que se genere será función del error cometido por las neuronas que reciban como entrada la salida de dicha neurona. Esto será lo que se ha denominado anteriormente la propagación del error hacia atrás. En este caso la delta

generalizada se expresará como sigue:

$$\delta_{pj} = \left( \sum_k \delta_{pk} w_{kj} \right) f'(net_j)$$

Donde el parámetro k representa todas aquellas neuronas que están conectadas a la salida de la neurona j. Por tanto, el error que se comete en una neurona en una capa oculta (intermedia) será la suma de los errores que se produzcan en las neuronas que están conectadas a la salida de ésta, siempre ponderado por los pesos correspondientes a cada conexión.

Como se ha indicado anteriormente el parámetro  $\alpha$  es el factor de aprendizaje y que determina la velocidad de aprendizaje de la neurona. A mayor tasa de aprendizaje más grandes serán los cambios de los pesos a cada iteración, y por tanto, más rápido será el aprendizaje. Sin embargo, este aprendizaje podría hacer oscilaciones y empeorar los resultados. Para reducir estas oscilaciones, Rumelhart, Hinton y Williams en 1986, proponen una modificación en la variación de incremento de los pesos, incorporando un término denominado momento o  $\beta$  quedando como sigue:

$$w_{ij}(t+1) = w_{ij}(t) + \alpha \delta_{pj} y_{pi} + \beta (w_{ij}(t) - w_{ij}(t-1)) =$$

$$\Delta w_{ij}(t+1) = \alpha \delta_{pj} y_{pi} + \beta \Delta w_{ij}(t)$$

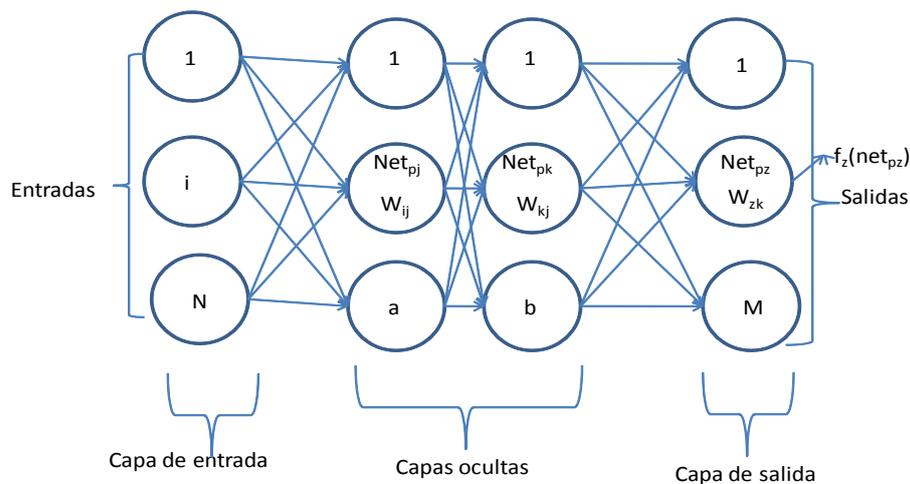
Donde  $\beta$  es una constante que determina el efecto en t+1 del cambio de los pesos en el instante t. Con esta constante lo que se consigue es la convergencia de la red con menos iteraciones, pues si en el momento t el incremento de un peso es positivo y en t+1 también lo es, entonces el error disminuye más rápidamente. En el caso en que en t el incremento es positivo y en t+1 es negativo, el paso de la información y las iteraciones necesarias serán menores ya que se ha pasado por un mínimo de error y para alcanzarlo serán necesarios menos pasos, pues ya ha sido localizado.

### 3.4.2. Estructura y Aprendizaje de la Red Backpropagation

La red backpropagation presenta una capa de entrada con n neuronas y una capa de salida con m neuronas. En cuanto al número de capas intermedias (ocultas), al menos, debe existir una. En las capas intermedias, las neuronas de una capa recibe la

información de las salidas de las neuronas de la capa anterior y éstas envían su salida a las neuronas de la siguiente capa. Por tanto, esta estructura de red no presenta conexiones hacia atrás (feedback) ni laterales entre neuronas de la misma capa.

Como se ha indicado, el algoritmo de aprendizaje backpropagation se aplica en dos fases, una hacia adelante y otra hacia atrás. En la primera, hacia adelante, se presenta un patrón de entrada con la información y se propaga por toda la red hasta alcanzar la salida de la misma. Una vez se obtienen estas salidas entra en juego la segunda fase, en la que se comparan éstas con los valores de salida deseados o esperados y se obtiene el error. Posteriormente se ajustan los pesos de las neuronas de la última capa ajustados a dicho error y se pasa a la capa anterior (hacia atrás) ajustándose en cada paso los pesos de las neuronas de cada capa hasta llegar a la primera capa. Así tenemos toda la red con todos los pesos modificados y adaptados para el ejemplo o patrón presentado.



**Figura 3-8: Estructura Red Backpropagation**

La aplicación del algoritmo backpropagation o regla delta generalizada requiere que las neuronas tengan una función de activación de tipo continuo y diferenciable, generalmente serán de tipo sigmoideal, aunque puede utilizarse del tipo lineal.

A continuación vamos a detallar de forma resumida todo el proceso de aprendizaje y entrenamiento de la red.

### Fase 1

Inicialmente, se dan valores a los pesos de las neuronas. Generalmente serán valores pequeños y obtenidos de forma aleatoria.

### Fase 2

Se presenta a la red el vector o patrón de entrada  $X_p : X_{p1}, X_{p2}, \dots, X_{pN}$ , y se especifica los datos de salida deseados (recordemos que estamos ante una red con aprendizaje supervisado)  $d_1, d_2, \dots, d_M$ . En caso que la red Backpropagation se utilice para obtener clasificaciones de los datos, estas salidas deseadas serán cero excepto aquella a la clase a la que pertenezca el patrón de entrada.

### Fase 3

Se procede a calcular la salida de la red (recordemos primera fase del algoritmo hacia adelante) presentándose todas las entradas y calculando cada salida de cada capa hasta llegar a la salida de la red  $y_1, y_2, \dots, y_M$ .

Para lograr este objetivo, primero se calculan las entradas netas para las neuronas de las capas ocultas.

$$net_{pj}^h = \sum_{i=1}^N w_{ij}^h x_{pi} + \theta_j^h$$

Donde el índice  $h$  representa la magnitud de la capa oculta (hidden); el subíndice  $p$  representa al  $p$ -ésimo vector de entrenamiento; la  $j$  es la  $j$ -ésima neurona oculta. El resto de parámetros ya son conocidos.

A continuación se ha de calcular las salidas de las neuronas ocultas en función de las entradas netas correspondientes.

$$y_{pj} = f_j^h(net_{pj}^h)$$

Por último se realizan los mismos cálculos (entrada y salida), pero para las neuronas que se encuentran en la capa de salida (capa o :output).

$$net_{pk}^o = \sum_{j=1}^L w_{kj}^o y_{pj} + \theta_k^o$$

$$y_{pk} = f_k^o(net_{pk}^o)$$

#### Fase 4

Una vez finalizado el proceso hacia adelante, comienza el proceso para atrás. Para ello, se ha de calcular los errores cometidos por todas las neuronas de la red. En el caso de ser una neurona de la capa de salida, el valor de la delta será:

$$\delta_{pk}^o = (d_{pk} - y_{pk}) f_k^o'(net_{pk}^o)$$

Como ya se ha indicado en otras ocasiones la función  $f$  debe ser derivable, por lo que la función de activación no puede ser de tipo escalón. Por tanto, se pueden utilizar bien la función lineal que como vimos  $f_k(net_{jk}) = net_{jk}$  o bien la función de activación sigmoideal  $f_k(net_{jk}) = \frac{1}{1+e^{-net_{jk}}}$ .

La elección de una función u otra depende de los datos de salida de la red que se deseen. Si queremos que las neuronas de salida sean binarias se utilizará la función sigmoideal, ya que esta función es derivable y biestable. En cualquier otro caso se pueden utilizar cualquiera de las dos indistintamente.

Si la neurona sobre la que se quiere calcular el error, neurona  $j$ , no es una neurona de salida, entonces la derivada parcial del error no puede hallarse directamente. Por esto, a partir del valor de la delta obtenido, que para este caso de neurona de una capa intermedia es,

$$\delta_{pj}^h = f_j^h'(net_{pj}^h) \sum_k \delta_{pk}^o w_{kj}^o$$

Donde el error de las neuronas en las capas ocultas depende de todos los errores de las neuronas de la capa de salida (proceso de propagación hacia atrás como habíamos

visto).

## Fase 5

Una vez determinado los errores de todas las neuronas, se procede al ajuste de los pesos de cada conexión de cada neurona, empezando desde la capa de salida hacia atrás. Los pesos se ajustarán para reducir la delta generalizada, es decir, reducción del error.

Para el caso de los pesos de las neuronas de la capa de salida se ajustarán así:

$$w_{kj}^o(t+1) = w_{kj}^o(t) + \Delta w_{kj}^o(t+1)$$
$$\Delta w_{kj}^o(t+1) = \alpha \delta_{pk}^o y_{pj}$$

Para los pesos de las neuronas de las capas intermedias (ocultas) se ajustarán:

$$w_{ji}^h(t+1) = w_{ji}^h(t) + \Delta w_{ji}^h(t+1)$$
$$\Delta w_{ji}^h(t+1) = \alpha \delta_{pj}^h x_{pi}$$

Para acelerar el proceso de aprendizaje se puede incorporar el término *momento* como ya vimos. Para las neuronas de salida el momento será igual a

$$\beta (w_{kj}^o(t) - w_{kj}^o(t-1))$$

Para el caso de neuronas de capas intermedias u ocultas:

$$\beta (w_{ji}^h(t) - w_{ji}^h(t-1))$$

## Fase 6

Todo este proceso se repite hasta que el término de error se reduce a la mínima expresión para cada patrón aprendido.

$$E_p = \frac{1}{2} \sum_{k=1}^M \delta_{pk}^2$$

Aplicando las fases anteriores, el algoritmo backpropagation busca el mínimo del error. Ahora bien, este mínimo puede ser local o global. Como el algoritmo presenta un gradiente descendente se asegura que con el cambio de los pesos se asegura la reducción del error hasta encontrar el mínimo más cercano, por lo que en algún caso el proceso podría detenerse en algún mínimo local y no global. Ahora bien, según el problema planteado no tiene porque ser necesaria la reducción al mínimo del error global, podría bastar y ser satisfactoria la solución obtenida con un valor de error de mínimo local.

El control del gradiente decreciente viene determinado por el parámetro  $\alpha$  (recordemos que es la tasa de aprendizaje). En función del número que se elija se incrementa o se reduce la velocidad del aprendizaje. A mayor importe, menor tiempo de aprendizaje, pero los saltos que se producen en los pesos son grandes y, por tanto, podría saltarse el mínimo y no poder alcanzarlo. En general suele aplicarse unos valores entre 0.05 a 0.25.

Cuando la red llega a un mínimo, sea local o global, detiene el aprendizaje. Si se trata de un mínimo del error local y la solución que ofrece la red no es satisfactoria, puede plantearse algunas de las siguientes soluciones para alcanzar el mínimo global o alcanzar un nuevo mínimo local menor que el anterior. Estas posibles soluciones pueden ser bien añadiendo más capas ocultas, cambios en los parámetros de aprendizaje, volver a iniciar el proceso de aprendizaje desde el inicio pero con pesos iniciales diferentes. En algunos paquetes informáticos es posible forzar a la red para que continúe con el aprendizaje, a pesar de haber alcanzado un mínimo local.

## **4. Estudio comparativo entre modelo ARMA-GARCH-M y la Red Neuronal Backpropagation: Aplicación sobre la volatilidad condicional y rentabilidad del índice Ibex-35**

### **4.1. *Modelización volatilidad condicional y la rentabilidad del Ibex-35 mediante modelo ARMA-GARCH-M***

En los anteriores apartados se ha realizado una revisión de la familia de los modelos GARCH, utilizados para el estudio y modelización de la volatilidad condicional, entre otros. También se ha revisado los elementos fundamentales y funcionamiento de las redes neuronales en general, y la red BackPropagation en particular. En este apartado se va a realizar un estudio comparativo entre un modelo de la familia GARCH y la red neuronal BackPropagation en la modelización y estimación de la volatilidad condicional del índice Ibex-35. Inicialmente se propuso el estudio únicamente de la volatilidad condicional, pero dada la existencia de algunos modelos de la familia GARCH, en los que puede modelizarse y estimarse de forma simultánea la rentabilidad y volatilidad condicional, se ha optado por realizar este estudio comparativo, pues, además, para el caso de la red neuronal no nos influye dado que es flexible a la hora de plantear qué outputs queremos obtener con la misma.

Se ha escogido la serie de cierres diarios del índice Ibex-35 desde el 03/01/2000 hasta el 14/07/2010. A continuación se procedió al cálculo de la rentabilidad continua diaria del índice.

El siguiente paso fue el estudio de la estacionariedad de la serie de rendimientos para poder aplicar la metodología Box-Jenkins y poder obtener el proceso ARMA generador de la serie de rendimientos.

A continuación se presenta la serie de rendimientos en la que se observa la estacionariedad en media. En cuanto a la estacionariedad en varianza indicar que, general, es estacionaria en gran parte de la muestra, presentando algunas desviaciones importantes en algunas observaciones que pudieran tratarse como outliers. Sobre este tema se hablará más adelante en la estimación del modelo final.



Figura 4-1: Gráfico cierres diarios del Ibx-35 del periodo muestral

### RT

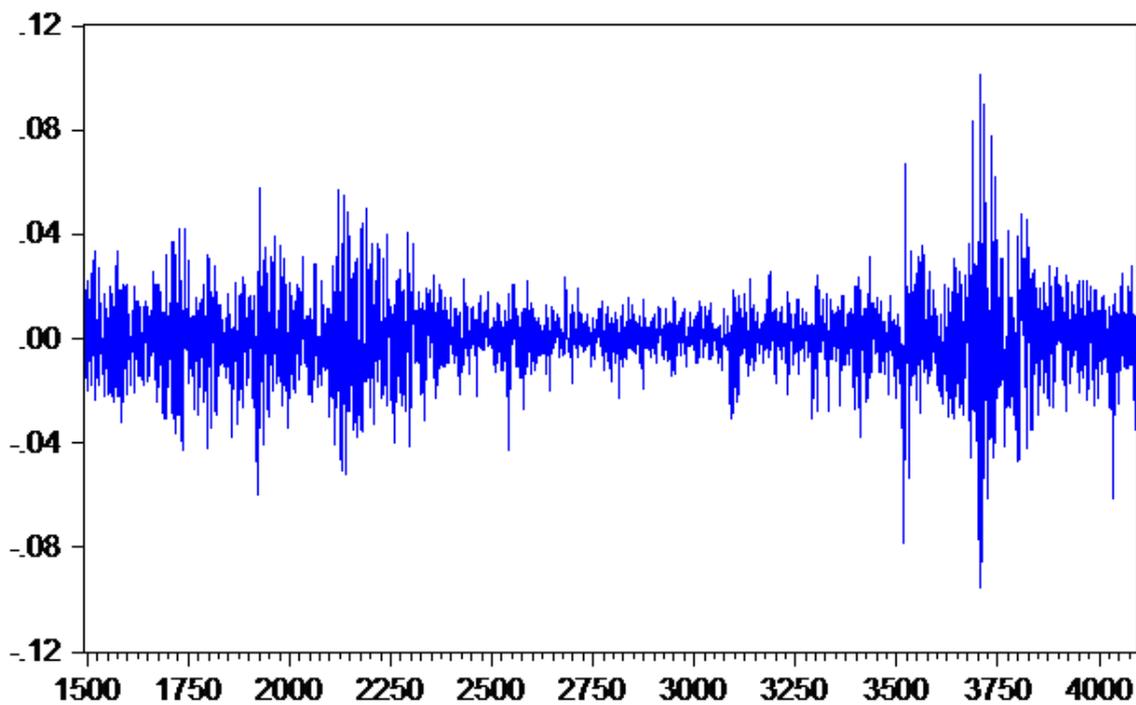


Figura 4-2: Gráfico rentabilidad del Ibx-35

La serie de rendimientos, como la mayoría de series de rendimientos bursátiles, suelen ser series leptocúrticas, con kurtosis elevada y suelen rechazar estadísticamente la hipótesis de normalidad según el estadístico Jarque-Bera.

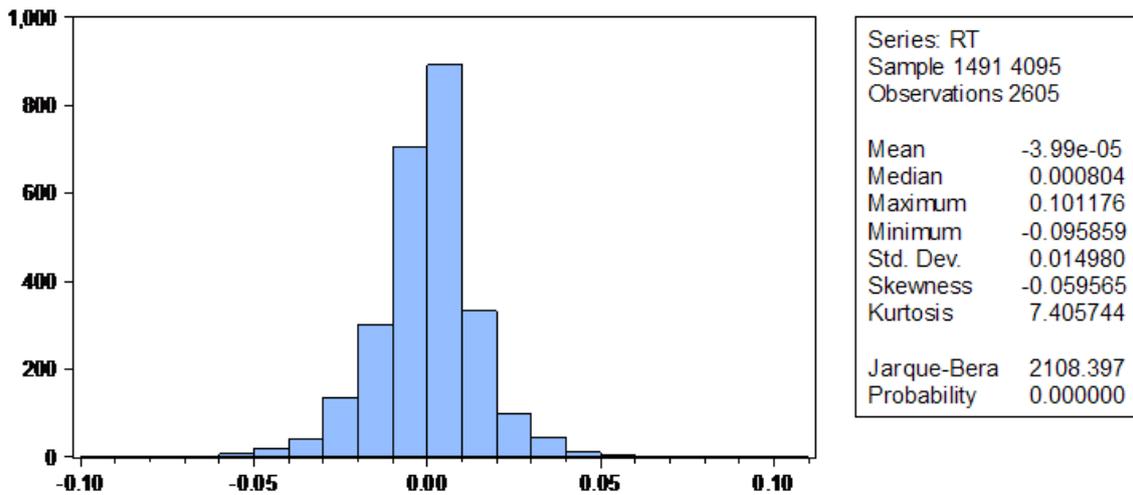


Figura 4-3: Histograma de la rentabilidad del Ibx-35

A continuación se presentan las funciones de autocorrelación y autocorrelación parcial para observar que, al menos visualmente, la serie parece estacionaria.

Tabla 4-1: FAC y FACP de la rentabilidad Ibx-35

Date: 07/16/10 Time: 09:44  
 Sample: 1491 4095  
 Included observations: 2605

Autocorrelation		Partial Correlation		AC	PAC	Q-Stat	Prob	
.		.		1	-0.026	-0.026	1.7489	0.186
*		*		2	-0.044	-0.045	6.8298	0.033
*		*		3	-0.036	-0.038	10.199	0.017
. *		. *		4	0.060	0.056	19.696	0.001
*		*		5	-0.049	-0.050	26.022	0.000
.		.		6	-0.031	-0.030	28.536	0.000
.		.		7	-0.014	-0.016	29.067	0.000
. *		. *		8	0.067	0.057	40.738	0.000
.		.		9	-0.018	-0.013	41.603	0.000
*		*		10	-0.049	-0.045	47.854	0.000
.		.		11	0.019	0.019	48.810	0.000
.		.		12	0.017	0.004	49.572	0.000
.		.		13	0.019	0.024	50.473	0.000
.		.		14	-0.008	0.002	50.650	0.000
.		.		15	0.002	-0.002	50.657	0.000
. *		.		16	0.041	0.037	55.104	0.000
.		.		17	0.009	0.012	55.336	0.000
.		.		18	-0.026	-0.014	57.159	0.000
.		.		19	0.002	0.002	57.166	0.000
.		.		20	-0.005	-0.012	57.222	0.000
.		.		21	0.008	0.008	57.376	0.000
.		.		22	0.011	0.018	57.697	0.000
.		.		23	-0.021	-0.020	58.902	0.000
.		.		24	0.019	0.015	59.873	0.000
.		.		25	0.027	0.026	61.831	0.000
.		.		26	-0.001	0.005	61.832	0.000
*		*		27	-0.064	-0.060	72.788	0.000
.		.		28	0.007	0.002	72.933	0.000
. *		.		29	0.048	0.041	78.901	0.000
.		.		30	0.037	0.036	82.431	0.000
.		.		31	-0.025	-0.007	84.044	0.000
.		.		32	-0.003	-0.007	84.062	0.000
.		.		33	-0.008	-0.019	84.212	0.000
.		.		34	-0.010	-0.009	84.481	0.000
*		*		35	-0.049	-0.036	90.940	0.000
.		.		36	0.011	0.006	91.241	0.000

Para confirmar la estacionariedad de la serie se ha procedido utilizar uno de los tests típicos como el de Dickey-Fuller Aumentado, que se estimó sin constante ni tendencia, pues éstas resultaron no significativas. Para la selección de los retardos en este test se utilizó el criterio Schwartz. El resultado obtenido para los niveles de significatividad del 1%, 5% y 10%, resultaron que se rechazaba la hipótesis nula de existencia de raíces unitarias, siendo, por tanto, la serie estacionaria.

**Tabla 4-2: Test DFA serie rentabilidad**

Null Hypothesis: RT has a unit root

Exogenous: None

Lag Length: 0 (Automatic - based on SIC, maxlag=27)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-52.36938	0.0001
Test critical values: 1% level	-2.565850	
5% level	-1.940945	
10% level	-1.616618	

\*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(RT)

Method: Least Squares

Date: 07/16/10 Time: 13:12

Sample: 1491 4095

Included observations: 2605

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RT(-1)	-1.025912	0.019590	-52.36938	0.0000
R-squared	0.512957	Mean dependent var		-1.06E-07
Adjusted R-squared	0.512957	S.D. dependent var		0.021457
S.E. of regression	0.014975	Akaike info criterion		-5.564522
Sum squared resid	0.583927	Schwarz criterion		-5.562271
Log likelihood	7248.790	Hannan-Quinn criter.		-5.563707
Durbin-Watson stat	2.002229			

Como el valor del contraste DFA (-52.369) es menor que el valor tabulado para los tres niveles de significación indicados se rechaza la hipótesis nula de existencia de raíces unitarias  $H_0 = \phi = 1$ .

A pesar de la estacionariedad de la serie según lo que hemos visto, también puede observarse la posible existencia de autocorrelación, al menos en primer orden, ya que el primer retardo de la serie rendimiento en la función de autocorrelación y autocorrelación parcial resulta ser significativo.

Con el test de Phillips-Perron podemos determinar la estacionariedad de la serie  $y$ , al mismo tiempo, podemos confirmar la existencia de autocorrelación de orden uno.

**Tabla 4-3: Test Phillip-Perron serie rentabilidad**

Null Hypothesis: RT has a unit root

Exogenous: Constant

Bandwidth: 2 (Newey-West automatic) using Bartlett kernel

	Adj. t-Stat	Prob.*
Phillips-Perron test statistic	-52.38846	0.0001
Test critical values:		
1% level	-3.432667	
5% level	-2.862450	
10% level	-2.567299	

\*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.000224
HAC corrected variance (Bartlett kernel)	0.000217

Phillips-Perron Test Equation

Dependent Variable: D(RT)

Method: Least Squares

Date: 07/16/10 Time: 11:51

Sample: 1491 4095

Included observations: 2605

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RT(-1)	-1.025919	0.019594	-52.35970	0.0000
C	-4.09E-05	0.000293	-0.139477	0.8891
R-squared	0.512961	Mean dependent var		-1.06E-07
Adjusted R-squared	0.512774	S.D. dependent var		0.021457
S.E. of regression	0.014978	Akaike info criterion		-5.563762
Sum squared resid	0.583923	Schwarz criterion		-5.559259
Log likelihood	7248.800	Hannan-Quinn criter.		-5.562131
F-statistic	2741.538	Durbin-Watson stat		2.002231
Prob(F-statistic)	0.000000			

El test de Phillip-Perron ha confirmado, al igual que el test DFA, la estacionariedad de la serie. Además ha confirmado la existencia de autocorrelación de orden uno. Recordamos que este test sólo puede confirmar la autocorrelación de orden uno, por lo que habrá que realizar más pruebas para la existencia de órdenes superiores.

Por último, se ha utilizado el test Durbin-Watson, en el que se ha estimado la serie de rentabilidades con varios retardos, desde 1 hasta 5, para determinar que en todos los casos no existe autocorrelación en los residuos, pues este estadístico se

encuentra en torno a 2, y por tanto, podemos concluir que la serie es estacionaria.

**Tabla 4-4: Test D-W con 1 retardo serie rentabilidad**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 12:36  
 Sample: 1491 4095  
 Included observations: 2605

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.09E-05	0.000293	-0.139477	0.8891
RT(-1)	-0.025919	0.019594	-1.322829	0.1860
R-squared	0.000672	Mean dependent var		-3.99E-05
Adjusted R-squared	0.000288	S.D. dependent var		0.014980
S.E. of regression	0.014978	Akaike info criterion		-5.563762
Sum squared resid	0.583923	Schwarz criterion		-5.559259
Log likelihood	7248.800	Hannan-Quinn criter.		-5.562131
F-statistic	1.749877	Durbin-Watson stat		2.002231
Prob(F-statistic)	0.186008			

**Tabla 4-5: Test D-W con 2 retardos serie rentabilidad**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 12:38  
 Sample: 1491 4095  
 Included observations: 2605

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.31E-05	0.000293	-0.147138	0.8830
RT(-1)	-0.027099	0.019584	-1.383742	0.1666
RT(-2)	-0.045230	0.019595	-2.308170	0.0211
R-squared	0.002714	Mean dependent var		-3.99E-05
Adjusted R-squared	0.001947	S.D. dependent var		0.014980
S.E. of regression	0.014965	Akaike info criterion		-5.565040
Sum squared resid	0.582729	Schwarz criterion		-5.558285
Log likelihood	7251.464	Hannan-Quinn criter.		-5.562593
F-statistic	3.540218	Durbin-Watson stat		2.003442
Prob(F-statistic)	0.029147			

**Tabla 4-6: Test D-W con 3 retardos serie rentabilidad**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 12:39  
 Sample: 1491 4095  
 Included observations: 2605

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.48E-05	0.000293	-0.152881	0.8785
RT(-1)	-0.028854	0.019593	-1.472697	0.1410
RT(-2)	-0.046244	0.019591	-2.360483	0.0183
RT(-3)	-0.039028	0.019618	-1.989417	0.0468
R-squared	0.004229	Mean dependent var		-3.99E-05
Adjusted R-squared	0.003080	S.D. dependent var		0.014980
S.E. of regression	0.014957	Akaike info criterion		-5.565793
Sum squared resid	0.581844	Schwarz criterion		-5.556787
Log likelihood	7253.445	Hannan-Quinn criter.		-5.562530
F-statistic	3.682089	Durbin-Watson stat		1.995491
Prob(F-statistic)	0.011591			

**Tabla 4-7: Test D-W con 4 retardos serie rentabilidad**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 12:39  
 Sample: 1491 4095  
 Included observations: 2605

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.31E-05	0.000293	-0.147193	0.8830
RT(-1)	-0.026641	0.019579	-1.360644	0.1737
RT(-2)	-0.043714	0.019582	-2.232351	0.0257
RT(-3)	-0.037321	0.019598	-1.904310	0.0570
RT(-4)	0.057250	0.019634	2.915839	0.0036
R-squared	0.007475	Mean dependent var		-3.99E-05
Adjusted R-squared	0.005948	S.D. dependent var		0.014980
S.E. of regression	0.014935	Akaike info criterion		-5.568290
Sum squared resid	0.579948	Schwarz criterion		-5.557032
Log likelihood	7257.697	Hannan-Quinn criter.		-5.564211
F-statistic	4.895060	Durbin-Watson stat		1.994388
Prob(F-statistic)	0.000623			

Una vez que se ha confirmado la estacionariedad de la serie mediante los tests indicados, se procede a determinar el mejor modelo ARIMA que refleje el proceso generador de la serie de rendimientos. Para ello, se plantean diversos modelos ARMA y mediante la utilización del criterio Schwartz elegiremos el mejor modelo. En este caso se ha obviado el criterio AIC dado que éste suele dimensionar por exceso los órdenes de retardo de autocorrelación. Los modelos planteados han sido ARMA(1,0), ARMA(1,1), ARMA(2,2), ARMA(1,2), ARMA(0,1) y ARMA(0,2).

**Tabla 4-8: Test D-W con 5 retardos serie rentabilidad**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 12:39  
 Sample: 1491 4095  
 Included observations: 2605

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.43E-05	0.000292	-0.151450	0.8796
RT(-1)	-0.023859	0.019592	-1.217774	0.2234
RT(-2)	-0.045518	0.019576	-2.325189	0.0201
RT(-3)	-0.039355	0.019596	-2.008320	0.0447
RT(-4)	0.055900	0.019622	2.848825	0.0044
RT(-5)	-0.048809	0.019646	-2.484470	0.0130
R-squared	0.009826	Mean dependent var		-3.99E-05
Adjusted R-squared	0.007921	S.D. dependent var		0.014980
S.E. of regression	0.014920	Akaike info criterion		-5.569894
Sum squared resid	0.578574	Schwarz criterion		-5.556385
Log likelihood	7260.787	Hannan-Quinn criter.		-5.565000
F-statistic	5.158358	Durbin-Watson stat		2.002950
Prob(F-statistic)	0.000103			

A continuación se describen los resultados principales de dichos modelos estimados:

**Tabla 4-9: Estimación modelo ARMA(1,0)**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 14:20  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.99E-05	0.000286	-0.139477	0.8891
AR(1)	-0.025919	0.019594	-1.322829	0.1860
R-squared	0.000672	Mean dependent var		-3.99E-05
Adjusted R-squared	0.000288	S.D. dependent var		0.014980
S.E. of regression	0.014978	Akaike info criterion		-5.563762
Sum squared resid	0.583923	Schwarz criterion		-5.559259
Log likelihood	7248.800	Hannan-Quinn criter.		-5.562131
F-statistic	1.749877	Durbin-Watson stat		2.002231
Prob(F-statistic)	0.186008			
Inverted AR Roots	-0.03			

**Tabla 4-10: Estimación modelo ARMA(1,1)**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 14:22  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 13 iterations  
 MA Backcast: 1490

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.85E-05	0.000256	-0.072351	0.9423
AR(1)	0.709029	0.157422	4.504000	0.0000
MA(1)	-0.746037	0.148692	-5.017318	0.0000
R-squared	0.003538	Mean dependent var		-3.99E-05
Adjusted R-squared	0.002772	S.D. dependent var		0.014980
S.E. of regression	0.014959	Akaike info criterion		-5.565867
Sum squared resid	0.582248	Schwarz criterion		-5.559113
Log likelihood	7252.542	Hannan-Quinn criter.		-5.563420
F-statistic	4.619822	Durbin-Watson stat		1.984338
Prob(F-statistic)	0.009936			
Inverted AR Roots	.71			
Inverted MA Roots	.75			

**Tabla 4-11: Estimación modelo ARMA(2,2)**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 14:23  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 14 iterations  
 MA Backcast: 1489 1490

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.89E-05	0.000283	-0.137547	0.8906
AR(1)	0.162442	0.020487	7.929080	0.0000
AR(2)	-0.942026	0.019864	-47.42392	0.0000
MA(1)	-0.192286	0.023887	-8.049933	0.0000
MA(2)	0.923333	0.023287	39.65090	0.0000
R-squared	0.018074	Mean dependent var		-3.99E-05
Adjusted R-squared	0.016564	S.D. dependent var		0.014980
S.E. of regression	0.014855	Akaike info criterion		-5.579026
Sum squared resid	0.573754	Schwarz criterion		-5.567769
Log likelihood	7271.682	Hannan-Quinn criter.		-5.574948
F-statistic	11.96447	Durbin-Watson stat		1.996915
Prob(F-statistic)	0.000000			
Inverted AR Roots	.08+.97i	.08-.97i		
Inverted MA Roots	.10+.96i	.10-.96i		

**Tabla 4-12: Estimación modelo ARMA(1,2)**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 14:23  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 14 iterations  
 MA Backcast: 1489 1490

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.20E-05	0.000255	-0.085980	0.9315
AR(1)	0.645456	0.197428	3.269318	0.0011
MA(1)	-0.674350	0.198003	-3.405752	0.0007
MA(2)	-0.017042	0.024319	-0.700769	0.4835
R-squared	0.003712	Mean dependent var		-3.99E-05
Adjusted R-squared	0.002563	S.D. dependent var		0.014980
S.E. of regression	0.014960	Akaike info criterion		-5.565274
Sum squared resid	0.582146	Schwarz criterion		-5.56268
Log likelihood	7252.769	Hannan-Quinn criter.		-5.562011
F-statistic	3.230466	Durbin-Watson stat		2.000062
Prob(F-statistic)	0.021538			
Inverted AR Roots	.65			
Inverted MA Roots	.70	-.02		

**Tabla 4-13: Estimación modelo ARMA(0,1)**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 14:24  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 6 iterations  
 MA Backcast: 1490

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.99E-05	0.000285	-0.140029	0.8886
MA(1)	-0.028519	0.019593	-1.455590	0.1456
R-squared	0.000738	Mean dependent var		-3.99E-05
Adjusted R-squared	0.000354	S.D. dependent var		0.014980
S.E. of regression	0.014977	Akaike info criterion		-5.563828
Sum squared resid	0.583884	Schwarz criterion		-5.559325
Log likelihood	7248.886	Hannan-Quinn criter.		-5.562197
F-statistic	1.921783	Durbin-Watson stat		1.997302
Prob(F-statistic)	0.165779			
Inverted MA Roots	.03			

**Tabla 4-14: Estimación modelo ARMA(0,2)**

Dependent Variable: RT  
 Method: Least Squares  
 Date: 07/16/10 Time: 14:24  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 7 iterations  
 MA Backcast: 1489 1490

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.96E-05	0.000272	-0.145754	0.8841
MA(1)	-0.030321	0.019587	-1.548052	0.1217
MA(2)	-0.041983	0.019594	-2.142633	0.0322
R-squared	0.002669	Mean dependent var		-3.99E-05
Adjusted R-squared	0.001903	S.D. dependent var		0.014980
S.E. of regression	0.014965	Akaike info criterion		-5.564995
Sum squared resid	0.582755	Schwarz criterion		-5.558241
Log likelihood	7251.406	Hannan-Quinn criter.		-5.562548
F-statistic	3.481999	Durbin-Watson stat		1.996802
Prob(F-statistic)	0.030889			
Inverted MA Roots	.22	-.19		

Utilizando el criterio Schwartz, tal y como se ha indicado anteriormente, todos los modelos obtienen un resultado bastante similar. Sin embargo se ha elegido el modelo ARMA(1,1) dado que es el que menor valor presenta con sus términos ARMA significativos, pues el modelo ARMA(1,2), a pesar de tener menor valor en el criterio Schwartz, y por tanto, en principio mejor modelo, presenta el término MA(2) no significativo por lo que rechazamos el modelo.

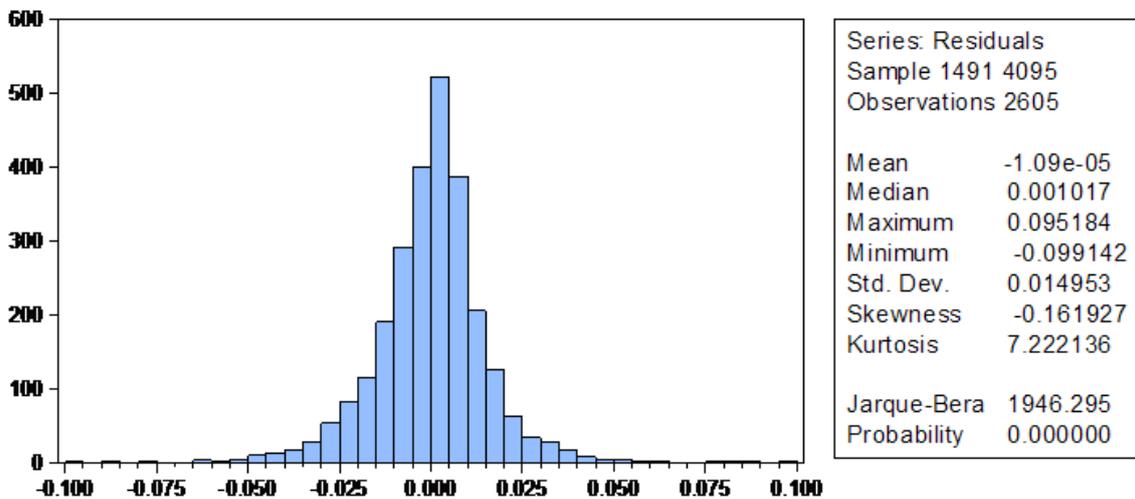
Una vez hemos escogido el modelo generador de la serie de rentabilidades procedemos a realizar una serie de contrastes sobre los residuos de dicho modelo estimado. Así pues, utilizando el contraste de independencia de Box-Pierce se observa la no existencia de autocorrelación en los residuos.

**Tabla 4-15: FAC y FACP modelo ARMA(1,1) serie rentabilidad**

Date: 07/16/10 Time: 14:41  
 Sample: 1491 4095  
 Included observations: 2605  
 Q-statistic probabilities  
 adjusted for 2 ARMA  
 term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.008	0.008	0.1582	
		2	-0.019	-0.019	1.1057	
		3	-0.017	-0.017	1.8823	0.170
*	*	4	0.070	0.070	14.775	0.001
		5	-0.039	-0.041	18.753	0.000
		6	-0.024	-0.021	20.245	0.000
		7	-0.009	-0.007	20.438	0.001
*	*	8	0.069	0.062	32.723	0.000
		9	-0.015	-0.012	33.287	0.000
		10	-0.045	-0.042	38.627	0.000
		11	0.021	0.023	39.737	0.000
		12	0.019	0.007	40.666	0.000
		13	0.020	0.026	41.748	0.000
		14	-0.005	0.003	41.826	0.000
		15	0.004	-0.000	41.871	0.000
		16	0.042	0.037	46.572	0.000
		17	0.011	0.010	46.893	0.000
		18	-0.024	-0.015	48.435	0.000
		19	0.003	0.002	48.455	0.000
		20	-0.003	-0.011	48.476	0.000
		21	0.009	0.010	48.693	0.000
		22	0.012	0.018	49.086	0.000
		23	-0.019	-0.020	50.069	0.000
		24	0.020	0.016	51.112	0.000
		25	0.028	0.025	53.152	0.000
*	*	26	0.000	0.003	53.152	0.001
		27	-0.062	-0.059	63.160	0.000
		28	0.008	0.005	63.333	0.000
		29	0.047	0.042	69.129	0.000
		30	0.036	0.034	72.534	0.000
		31	-0.024	-0.010	73.999	0.000
		32	-0.003	-0.010	74.026	0.000
		33	-0.009	-0.021	74.221	0.000
*	*	34	-0.012	-0.010	74.578	0.000
		35	-0.049	-0.035	81.019	0.000
		36	0.009	0.008	81.247	0.000

En cuanto a la normalidad de los residuos indicar que los mismos no cumplen con las condiciones de normalidad.



**Figura 4-4: Histograma residuos modelo ARMA(1,1)**

A continuación se ha procedido a realizar el contraste de Breusch-Godfrey que también estudia la existencia o no de autocorrelación serial de los residuos del modelo estimado.

**Tabla 4-16: Test autocorrelación Breusch-Godfrey residuos modelo ARMA(1,1)**

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.651119	Prob. F(2,2600)	0.5215
Obs*R-squared	1.302703	Prob. Chi-Square(2)	0.5213

Test Equation:

Dependent Variable: RESID

Method: Least Squares

Date: 07/16/10 Time: 14:49

Sample: 1491 4095

Included observations: 2605

Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.74E-07	0.000256	0.003799	0.9970
AR(1)	0.047811	0.255115	0.187408	0.8514
MA(1)	-0.031228	0.217612	-0.143502	0.8859
RESID(-1)	-0.008582	0.049255	-0.174239	0.8617
RESID(-2)	-0.029777	0.032812	-0.907490	0.3642
R-squared	0.000500	Mean dependent var	-1.09E-05	
Adjusted R-squared	-0.001038	S.D. dependent var	0.014953	
S.E. of regression	0.014961	Akaike info criterion	-5.564832	
Sum squared resid	0.581956	Schwarz criterion	-5.553575	
Log likelihood	7253.194	Hannan-Quinn criter.	-5.560754	
F-statistic	0.325213	Durbin-Watson stat	2.001264	
Prob(F-statistic)	0.861201			

Como se observa, el p-valor es mayor de 0.05 y por tanto, no podemos rechazar la hipótesis nula de no autocorrelación de los residuos. Por tanto queda confirmada la no existencia de autocorrelación con ambos contrastes.

A continuación hemos realizado un contraste para determinar la existencia de Heterocedasticidad de los residuos.

**Tabla 4-17: Test Heterocedasticidad residuos modelo ARMA(1,1)**

Heteroskedasticity Test: ARCH

F-statistic	104.8812	Prob. F(5,2594)	0.0000
Obs*R-squared	437.2282	Prob. Chi-Square(5)	0.0000

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 07/16/10 Time: 14:50

Sample (adjusted): 1496 4095

Included observations: 2600 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.94E-05	1.18E-05	6.720517	0.0000
RESID^2(-1)	0.088781	0.019415	4.572770	0.0000
RESID^2(-2)	0.138318	0.019363	7.143455	0.0000
RESID^2(-3)	0.151940	0.019324	7.862800	0.0000
RESID^2(-4)	0.115030	0.019392	5.931975	0.0000
RESID^2(-5)	0.150576	0.019447	7.742714	0.0000
R-squared	0.168165	Mean dependent var	0.000223	
Adjusted R-squared	0.166561	S.D. dependent var	0.000558	
S.E. of regression	0.000509	Akaike info criterion	-12.32466	
Sum squared resid	0.000673	Schwarz criterion	-12.31113	
Log likelihood	16028.06	Hannan-Quinn criter.	-12.31976	
F-statistic	104.8812	Durbin-Watson stat	2.015033	
Prob(F-statistic)	0.000000			

Como se observa, existen varios residuos que son estadísticamente significativos, por lo que se confirma la existencia de procesos ARCH que podrían ser de orden superior a uno.

A continuación se procedió a incorporar al modelo, la modelización de la varianza condicional de las rentabilidades mediante la incorporación de un modelo GARCH. Para ello se estimaron varios modelos variando el orden de retardos tanto en la parte ARCH como en la parte GARCH, todos ellos se estimaron con constantes. En principio, el modelo más ajustado con todos los coeficientes significativos resultó ser un modelo ARMA(1,1)-GARCH(2,1), es decir, orden 2 para la parte de los residuos (ARCH) y uno para la parte de la varianza condicional GARCH.

**Tabla 4-18: Estimación modelo ARMA(1,1)-GARCH(2,1)**

Dependent Variable: RT  
 Method: ML - ARCH (Marquardt) - Normal distribution  
 Date: 07/21/10 Time: 18:10  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 20 iterations  
 MA Backcast: 1490  
 Presample variance: backcast (parameter = 0.7)  
 GARCH = C(4) + C(5)\*RESID(-1)^2 + C(6)\*RESID(-2)^2 + C(7)  
 \*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.000643	0.000197	3.269651	0.0011
AR(1)	0.785289	0.208928	3.758655	0.0002
MA(1)	-0.805318	0.198564	-4.055707	0.0000

Variance Equation				
C	2.63E-06	4.85E-07	5.420384	0.0000
RESID(-1)^2	0.051052	0.019043	2.680849	0.0073
RESID(-2)^2	0.070457	0.021092	3.340500	0.0008
GARCH(-1)	0.869262	0.012608	68.94723	0.0000

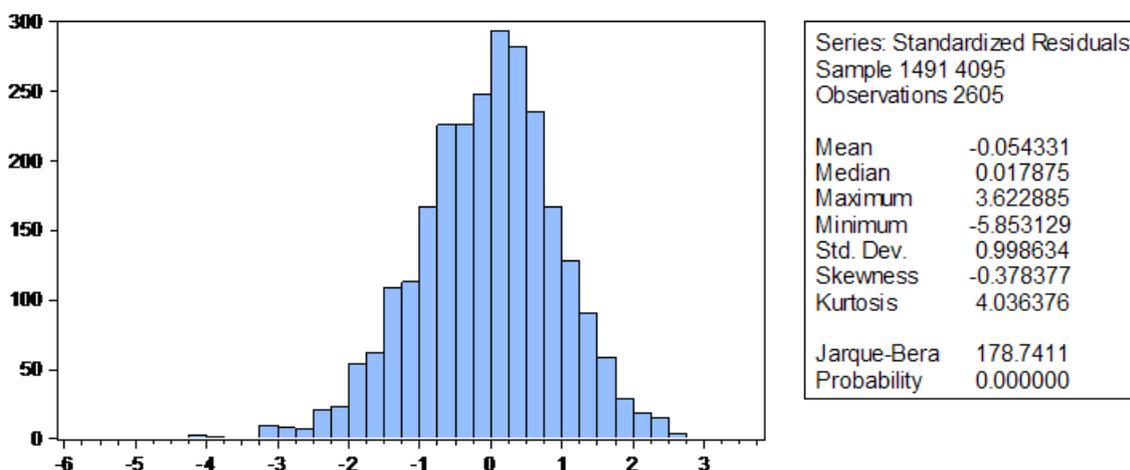
  

R-squared	0.000669	Mean dependent var	-3.99E-05
Adjusted R-squared	-0.000099	S.D. dependent var	0.014980
S.E. of regression	0.014980	Akaike info criterion	-5.938841
Sum squared resid	0.583924	Schwarz criterion	-5.923081
Log likelihood	7742.341	Hannan-Quinn criter.	-5.933131
Durbin-Watson stat	2.012400		

Inverted AR Roots	.79
Inverted MA Roots	.81

Si obtenemos el histograma de los residuos de este modelo estimado observamos que se ha mejorado tanto en los aspectos de kurtosis, resulta menos leptokúrtica, y mejora considerablemente el test de normalidad de Jarque-Bera, respecto a lo obtenido hasta el momento.



**Figura 4-5: Histograma residuos modelo ARMA(1,1)-GARCH(2,1)**

Para verificar la estabilidad estructural de este modelo se procedió a realizar el test estructural de CHOW, en el que se determinó separando un tercio de la muestra como suelen realizar algunos autores. El resultado obtenido es la estabilidad estructural del modelo en para todo el periodo muestral.

**Tabla 4-19: Test estabilidad Chow modelo ARMA(1,1)-GARCH(2,1)**

Chow Breakpoint Test: 3444  
Null Hypothesis: No breaks at specified breakpoints

Equation Sample: 1491 4095

F-statistic	0.446206	Prob. F(2,2601)	0.6401
Log likelihood ratio	0.893631	Prob. Chi-Square(2)	0.6397
Wald Statistic	0.891814	Prob. Chi-Square(2)	0.6402

A partir de este momento se introdujo un modelo GARCH-M introduciendo así la varianza condicional en la ecuación de la rentabilidad, dado que ambas están relacionadas entre sí de alguna forma. Se estimó pues el modelo ARMA(1,1)-GARCH-M(2,1) con y sin constante en la ecuación de la media eligiendo el mejor mediante el criterio de Hannan-Quinn, siendo en este caso el modelo sin constante ya que además obtenía todos sus parámetros significativos.

**Tabla 4-20: Estimación modelo ARMA(1,1)-GARCH(2,1)-M**

Dependent Variable: RT  
Method: ML - ARCH (Marquardt) - Normal distribution  
Date: 07/21/10 Time: 18:46  
Sample: 1491 4095  
Included observations: 2605  
Convergence achieved after 34 iterations  
MA Backcast: 1490  
Presample variance: backcast (parameter = 0.7)  
GARCH = C(4) + C(5)\*RESID(-1)^2 + C(6)\*RESID(-2)^2 + C(7)\*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	2.849229	1.237026	2.303290	0.0213
AR(1)	0.760269	0.261374	2.908738	0.0036
MA(1)	-0.776217	0.253046	-3.067495	0.0022
Variance Equation				
C	2.65E-06	5.00E-07	5.293364	0.0000
RESID(-1)^2	0.048751	0.019266	2.530366	0.0114
RESID(-2)^2	0.070958	0.021076	3.366714	0.0008
GARCH(-1)	0.870561	0.012385	70.29059	0.0000
R-squared	-0.004480	Mean dependent var	-3.99E-05	
Adjusted R-squared	-0.005252	S.D. dependent var	0.014980	
S.E. of regression	0.015019	Akaike info criterion	-5.936734	
Sum squared resid	0.586933	Schwarz criterion	-5.920973	
Log likelihood	7739.596	Hannan-Quinn criter.	-5.931024	
Durbin-Watson stat	2.009897			
Inverted AR Roots	.76			
Inverted MA Roots	.78			

Como ya se indicó en la revisión sobre la relación entre el volumen y la volatilidad, se hace indispensable la incorporación de esta variable como componente importante en la modelización de la volatilidad condicional. Por tanto, y dado que hay autores que incorporan tal cual a la ecuación de la varianza condicional el volumen retardado y otros lo incorporan como el logaritmo neperiano del volumen con un retardo, procedemos a estimar dos modelos con cada una de estas variables.

**Tabla 4-21: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando volumen con un retardo en la varianza condicional**

Dependent Variable: RT  
Method: ML - ARCH (Marquardt) - Normal distribution  
Date: 07/21/10 Time: 19:12  
Sample: 1491 4095  
Included observations: 2605  
Convergence achieved after 19 iterations  
MA Backcast: 1490  
Presample variance: backcast (parameter = 0.7)  
GARCH = C(4) + C(5)\*RESID(-1)^2 + C(6)\*RESID(-2)^2 + C(7)  
\*GARCH(-1) + C(8)\*VOL(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	3.031734	1.300498	2.331211	0.0197
AR(1)	0.739891	0.383357	1.930029	0.0536
MA(1)	-0.743924	0.380766	-1.953755	0.0507

Variance Equation				
C	6.85E-07	9.68E-07	0.707258	0.4794
RESID(-1)^2	0.040785	0.018446	2.211076	0.0270
RESID(-2)^2	0.074550	0.020630	3.613672	0.0003
GARCH(-1)	0.871591	0.012642	68.94485	0.0000
VOL(-1)	1.42E-11	5.36E-12	2.649950	0.0081

R-squared	-0.005729	Mean dependent var	-3.99E-05
Adjusted R-squared	-0.006502	S.D. dependent var	0.014980
S.E. of regression	0.015028	Akaike info criterion	-5.936855
Sum squared resid	0.587662	Schwarz criterion	-5.918843
Log likelihood	7740.753	Hannan-Quinn criter.	-5.930329
Durbin-Watson stat	2.031941		

Inverted AR Roots	.74
Inverted MA Roots	.74

**Tabla 4-22: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando logaritmo del volumen con un retardo en la varianza condicional**

Dependent Variable: RT  
 Method: ML - ARCH (Marquardt) - Normal distribution  
 Date: 07/21/10 Time: 19:12  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 54 iterations  
 MA Backcast: 1490  
 Presample variance: backcast (parameter = 0.7)  
 GARCH = C(4) + C(5)\*RESID(-1)^2 + C(6)\*RESID(-2)^2 + C(7)  
 \*GARCH(-1) + C(8)\*LNVOL(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
GARCH	2.799337	1.245796	2.247026	0.0246
AR(1)	0.760611	0.273791	2.778074	0.0055
MA(1)	-0.775367	0.265889	-2.916131	0.0035
Variance Equation				
C	-1.24E-05	1.01E-05	-1.219370	0.2227
RESID(-1)^2	0.047155	0.019277	2.446150	0.0144
RESID(-2)^2	0.071703	0.021229	3.377617	0.0007
GARCH(-1)	0.871741	0.012545	69.49092	0.0000
LNVOL(-1)	1.26E-06	8.48E-07	1.489267	0.1364
R-squared	-0.004322	Mean dependent var		-3.99E-05
Adjusted R-squared	-0.005094	S.D. dependent var		0.014980
S.E. of regression	0.015018	Akaike info criterion		-5.936482
Sum squared resid	0.586840	Schwarz criterion		-5.918470
Log likelihood	7740.268	Hannan-Quinn criter.		-5.929957
Durbin-Watson stat	2.012712			
Inverted AR Roots	.76			
Inverted MA Roots	.78			

En estos dos modelos nos surgen varios problemas. Por un lado, si elegimos incorporar el volumen con un retardo en la ecuación de la varianza condicional, esta sale significativa, pero sin embargo los parámetros del modelo ARMA para la determinación de la rentabilidad aparecen como no significativos y el impacto de la varianza condicional en la ecuación de la media aparece significativa. Por otro lado, si elegimos el modelo que incorpora el logaritmo neperiano del volumen con un retardo en la ecuación de la varianza condicional, aparece que esta no es significativa y sin embargo todos los parámetros de la ecuación de la media resultan significativos. Para dar solución a este dilema, creemos que en vez de incorporar la varianza condicional en la ecuación de la media, incorporamos el logaritmo de la varianza condicional, tal y como muestran algunos autores o la desviación estándar de la misma. Recordemos que en el modelo GARCH-M la incorporación de la varianza condicional en la ecuación de la media se puede hacer de tres formas, bien incorporando tal cual la varianza condicional (como habíamos hecho), bien el logaritmo de la varianza

condicional o ,por último, la desviación estándar de la varianza condicional.

Así pues volvemos estimaremos 4 modelos. Dos de ellos incorporando el logaritmo de la varianza condicional en la ecuación de la media y otros dos incorporando la desviación estándar de la varianza condicional. En ambos casos se estima el modelo incorporando bien el volumen con un retardo bien el logaritmo del volumen con un retardo en la ecuación de la varianza condicional.

**Tabla 4-23: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando el volumen con un retardo en la varianza condicional e incorporando el logaritmo varianza condicional en la ecuación de la media**

Dependent Variable: RT  
Method: ML - ARCH (Marquardt) - Normal distribution  
Date: 07/21/10 Time: 19:20  
Sample: 1491 4095  
Included observations: 2605  
Convergence achieved after 30 iterations  
MA Backcast: 1490  
Presample variance: backcast (parameter = 0.7)  
GARCH = C(4) + C(5)\*RESID(-1)^2 + C(6)\*RESID(-2)^2 + C(7)  
\*GARCH(-1) + C(8)\*VOL(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
LOG(GARCH)	-6.77E-05	2.12E-05	-3.184960	0.0014
AR(1)	0.790484	0.222989	3.544950	0.0004
MA(1)	-0.808703	0.212950	-3.797625	0.0001
Variance Equation				
C	7.59E-07	9.81E-07	0.773127	0.4394
RESID(-1)^2	0.048491	0.018978	2.555150	0.0106
RESID(-2)^2	0.070687	0.021254	3.325784	0.0009
GARCH(-1)	0.871155	0.012727	68.44687	0.0000
VOL(-1)	1.23E-11	5.49E-12	2.249338	0.0245
R-squared	0.001029	Mean dependent var	-3.99E-05	
Adjusted R-squared	0.000262	S.D. dependent var	0.014980	
S.E. of regression	0.014978	Akaike info criterion	-5.939158	
Sum squared resid	0.583714	Schwarz criterion	-5.921146	
Log likelihood	7743.754	Hannan-Quinn criter.	-5.932633	
Durbin-Watson stat	2.016856			
Inverted AR Roots	.79			
Inverted MA Roots	.81			

**Tabla 4-24: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando el logaritmo del volumen con un retardo en la varianza condicional e incorporando el logaritmo varianza condicional en la ecuación de la media**

Dependent Variable: RT  
 Method: ML - ARCH (Marquardt) - Normal distribution  
 Date: 07/21/10 Time: 19:22  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 28 iterations  
 MA Backcast: 1490  
 Presample variance: backcast (parameter = 0.7)  
 GARCH = C(4) + C(5)\*RESID(-1)^2 + C(6)\*RESID(-2)^2 + C(7)  
 \*GARCH(-1) + C(8)\*LNVOL(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
LOG(GARCH)	-6.86E-05	2.12E-05	-3.240991	0.0012
AR(1)	0.789865	0.213622	3.697485	0.0002
MA(1)	-0.809146	0.203322	-3.979633	0.0001
Variance Equation				
C	-1.06E-05	1.03E-05	-1.030136	0.3029
RESID(-1)^2	0.049619	0.018998	2.611723	0.0090
RESID(-2)^2	0.070692	0.021155	3.341687	0.0008
GARCH(-1)	0.870755	0.012708	68.52097	0.0000
LNVOL(-1)	1.11E-06	8.56E-07	1.294481	0.1955
R-squared	0.001032	Mean dependent var		-3.99E-05
Adjusted R-squared	0.000264	S.D. dependent var		0.014980
S.E. of regression	0.014978	Akaike info criterion		-5.938507
Sum squared resid	0.583712	Schwarz criterion		-5.920495
Log likelihood	7742.906	Hannan-Quinn criter.		-5.931982
Durbin-Watson stat	2.014722			
Inverted AR Roots	.79			
Inverted MA Roots	.81			

**Tabla 4-25: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando el volumen con un retardo en la varianza condicional e incorporando la desviación de la varianza condicional en la ecuación de la media**

Dependent Variable: RT  
 Method: ML - ARCH (Marquardt) - Normal distribution  
 Date: 07/21/10 Time: 19:23  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 21 iterations  
 MA Backcast: 1490  
 Presample variance: backcast (parameter = 0.7)  
 GARCH = C(4) + C(5)\*RESID(-1)^2 + C(6)\*RESID(-2)^2 + C(7)  
 \*GARCH(-1) + C(8)\*VOL(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
@SQRT(GARCH)	0.052947	0.019385	2.731286	0.0063
AR(1)	0.718392	0.320805	2.239341	0.0251
MA(1)	-0.731760	0.314051	-2.330070	0.0198
Variance Equation				
C	7.04E-07	9.98E-07	0.705521	0.4805
RESID(-1)^2	0.046611	0.019244	2.422130	0.0154
RESID(-2)^2	0.075130	0.021551	3.486136	0.0005
GARCH(-1)	0.868221	0.012986	66.85942	0.0000
VOL(-1)	1.34E-11	5.59E-12	2.392653	0.0167
R-squared	-0.001586	Mean dependent var		-3.99E-05
Adjusted R-squared	-0.002355	S.D. dependent var		0.014980
S.E. of regression	0.014997	Akaike info criterion		-5.938399
Sum squared resid	0.585242	Schwarz criterion		-5.920387
Log likelihood	7742.765	Hannan-Quinn criter.		-5.931874
Durbin-Watson stat	2.021024			
Inverted AR Roots	.72			
Inverted MA Roots	.73			

**Tabla 4-26: Estimación modelo ARMA(1,1)-GARCH(2,1)-M incorporando el logaritmo del volumen con un retardo en la varianza condicional e incorporando la desviación de la varianza condicional en la ecuación de la media**

Dependent Variable: RT  
 Method: ML - ARCH (Marquardt) - Normal distribution  
 Date: 07/21/10 Time: 19:24  
 Sample: 1491 4095  
 Included observations: 2605  
 Convergence achieved after 9 iterations  
 MA Backcast: 1490  
 Presample variance: backcast (parameter = 0.7)  
 GARCH = C(4) + C(5)\*RESID(-1)^2 + C(6)\*RESID(-2)^2 + C(7)  
 \*GARCH(-1) + C(8)\*LNVOL(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
@SQRT(GARCH)	0.056222	0.019362	2.903700	0.0037
AR(1)	-0.003223	2.271270	-0.001419	0.9989
MA(1)	-0.006134	2.271897	-0.002700	0.9978
Variance Equation				
C	-6.64E-06	6.36E-07	-10.43364	0.0000
RESID(-1)^2	0.055739	0.019793	2.816035	0.0049
RESID(-2)^2	0.083738	0.022267	3.760722	0.0002
GARCH(-1)	0.846559	0.014447	58.59909	0.0000
LNVOL(-1)	8.32E-07	5.52E-08	15.08751	0.0000
R-squared	-0.003812	Mean dependent var		-3.99E-05
Adjusted R-squared	-0.004584	S.D. dependent var		0.014980
S.E. of regression	0.015014	Akaike info criterion		-5.935493
Sum squared resid	0.586543	Schwarz criterion		-5.917481
Log likelihood	7738.979	Hannan-Quinn criter.		-5.928967
Durbin-Watson stat	2.024902			
Inverted AR Roots	-.00			
Inverted MA Roots	.01			

Sea como sea la incorporación de la volatilidad condicional en la ecuación de la media, el impacto del volumen con un retardo en la varianza condicional parece ser más importante y significativa que si se incorpora como el logaritmo del volumen con un retardo. No obstante el coeficiente del volumen en la ecuación de la varianza condicional es muy bajo, lo que nos lleva a pensar que, a pesar de que existe cierta relación entre ambas, ésta no tiene un peso específico muy destacable.

Para la elección del modelo final entre estos cuatro estimados se utilizó simultáneamente el criterio Schwartz y el criterio Hannan-Quinn, que en este caso coinciden, siendo el modelo elegido un ARMA(1,1)-GARCH-M(2,1) (con desviación de la varianza condicional en la ecuación de la rentabilidad y el volumen con un retardo en la ecuación de la varianza condicional).

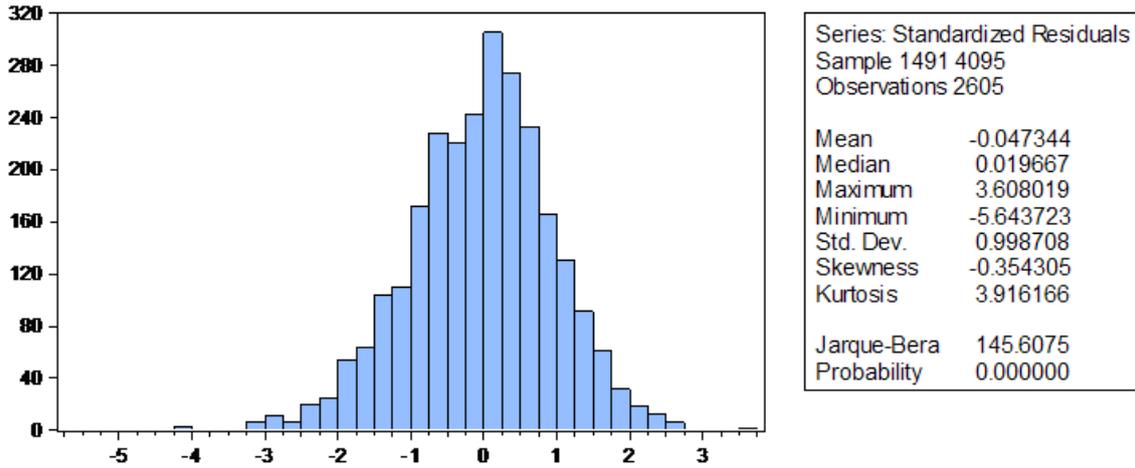
A continuación reproducimos de nuevo los datos principales de la estimación del modelo para realizar después un estudio de sus residuos.

**Tabla 4-27: Estimación modelo ARMA(1,1)-GARCH(2,1)-M con desviación estándar de la varianza condicional en la ecuación de la media y volumen con un retardo en la varianza condicional**

Dependent Variable: RT  
Method: ML - ARCH (Marquardt) - Normal distribution  
Date: 07/21/10 Time: 19:35  
Sample: 1491 4095  
Included observations: 2605  
Convergence achieved after 21 iterations  
MA Backcast: 1490  
Presample variance: backcast (parameter = 0.7)  
GARCH = C(4) + C(5)\*RESID(-1)^2 + C(6)\*RESID(-2)^2 + C(7)  
\*GARCH(-1) + C(8)\*VOL(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
@SQRT(GARCH)	0.052947	0.019385	2.731286	0.0063
AR(1)	0.718392	0.320805	2.239341	0.0251
MA(1)	-0.731760	0.314051	-2.330070	0.0198
Variance Equation				
C	7.04E-07	9.98E-07	0.705521	0.4805
RESID(-1)^2	0.046611	0.019244	2.422130	0.0154
RESID(-2)^2	0.075130	0.021551	3.486136	0.0005
GARCH(-1)	0.868221	0.012986	66.85942	0.0000
VOL(-1)	1.34E-11	5.59E-12	2.392653	0.0167
R-squared	-0.001586	Mean dependent var		-3.99E-05
Adjusted R-squared	-0.002355	S.D. dependent var		0.014980
S.E. of regression	0.014997	Akaike info criterion		-5.938399
Sum squared resid	0.585242	Schwarz criterion		-5.920387
Log likelihood	7742.765	Hannan-Quinn criter.		-5.931874
Durbin-Watson stat	2.021024			
Inverted AR Roots	.72			
Inverted MA Roots	.73			

A continuación se presenta el histograma de los residuos con sus estadísticas habituales. En este caso se observa mejora en todos los aspectos relacionados con la normalidad de los mismos, respecto a los residuos de otros modelos vistos anteriormente en este apartado. Quizás se detecta cierto alargamiento de una de las colas, que podría indicar cierta asimetría.



**Figura 4-6: Histograma residuos modelo ARMA(1,1)-GARCH(2,1)-M**

A continuación se procede a observar la función de autocorrelación y autocorrelación parcial. En ellas observamos que aún sigue existiendo cierta persistencia de la volatilidad. Lo que nos lleva a profundizar en un modelo con más retardos de los que habíamos planteado hasta ahora en la parte del modelo GARCH-M.

**Tabla 4-28: FAC y FACP del modelo ARMA(1,1)-GARCH(2,1)-M con desviación estándar de la varianza condicional en la ecuación de la media y volumen con un retardo en la varianza condicional**

Date: 07/21/10 Time: 19:37  
 Sample: 1491 4095  
 Included observations: 2605  
 Q-statistic  
 probabilities  
 adjusted for 2  
 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.009	0.009	0.1974	
		2	-0.005	-0.005	0.2710	
		3	0.006	0.006	0.3770	0.539
		4	0.041	0.041	4.7216	0.094
		5	-0.010	-0.010	4.9730	0.174
		6	-0.005	-0.004	5.0337	0.284
		7	-0.009	-0.009	5.2328	0.388
		8	0.041	0.040	9.5991	0.143
		9	-0.006	-0.005	9.6785	0.208
		10	-0.035	-0.034	12.838	0.118
		11	0.023	0.024	14.224	0.115
		12	0.003	-0.001	14.243	0.162
		13	0.035	0.037	17.500	0.094
		14	-0.011	-0.009	17.798	0.122
		15	-0.009	-0.010	18.013	0.157
		16	0.023	0.021	19.424	0.149
		17	0.001	-0.003	19.426	0.195
		18	0.003	0.009	19.456	0.246
		19	0.005	0.004	19.534	0.299
		20	0.013	0.011	19.995	0.333

Para determinar el orden definitivo de retardos en el modelo, se han planteado diversos modelos con mayores retardos, tanto en la parte ARCH como en la parte GARCH, hasta orden 6. Se utilizó el criterio BIC y Schwartz para determinar cuál de ellos mejoraba al que teníamos planteado hasta ahora. Finalmente resultó ser el modelo ARMA(1,1)-GARCH-M(4,2), el que mejor resultado dio, teniendo además sus parámetros significativos.

**Tabla 4-29: Estimación modelo ARMA(1,1)-GARCH(4,2)-M con desviación estándar de la varianza condicional en la ecuación de la media y volumen con un retardo en la varianza condicional**

Dependent Variable: RT  
Method: ML - ARCH (Marquardt) - Normal distribution  
Date: 07/27/10 Time: 12:46  
Sample: 1491 4095  
Included observations: 2605  
Convergence achieved after 32 iterations  
MA Backcast: 1490  
Presample variance: backcast (parameter = 0.7)  
GARCH = C(4) + C(5)\*RESID(-1)^2 + C(6)\*RESID(-2)^2 + C(7)\*RESID(-3)^2 + C(8)\*RESID(-4)^2 + C(9)\*GARCH(-1) + C(10)\*GARCH(-2) + C(11)\*VOL(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
@SQRT(GARCH)	0.038438	0.018317	2.098495	0.0359
AR(1)	0.671640	0.282922	2.373939	0.0176
MA(1)	-0.694047	0.274637	-2.527147	0.0115

Variance Equation				
C	9.16E-07	2.26E-06	0.405927	0.6848
RESID(-1)^2	0.051926	0.020564	2.525082	0.0116
RESID(-2)^2	0.103736	0.020112	5.157912	0.0000
RESID(-3)^2	0.063024	0.024331	2.590303	0.0096
RESID(-4)^2	0.053587	0.022002	2.435582	0.0149
GARCH(-1)	-0.087357	0.033284	-2.624559	0.0087
GARCH(-2)	0.795179	0.029842	26.64611	0.0000
VOL(-1)	3.37E-11	1.27E-11	2.661942	0.0078

R-squared	0.000629	Mean dependent var	-3.99E-05
Adjusted R-squared	-0.000139	S.D. dependent var	0.014980
S.E. of regression	0.014981	Akaike info criterion	-5.941658
Sum squared resid	0.583948	Schwarz criterion	-5.916891
Log likelihood	7750.009	Hannan-Quinn criter.	-5.932685
Durbin-Watson stat	2.007001		

Inverted AR Roots	.67
Inverted MA Roots	.69

Si se observa ahora la función de autocorrelación y autocorrelación parcial del modelo estimado, podemos observar que parece haberse reducido considerablemente la persistencia de la volatilidad, siendo eliminada en los retardos más cercanos, hecho que no sucedía antes.

**Tabla 4-30: FAC y FACP del modelo ARMA(1,1)-GARCH(4,2)-M con desviación estándar de la varianza condicional en la ecuación de la media y volumen con un retardo en la varianza condicional**

Date: 07/27/10 Time: 12:46  
 Sample: 1491 4095  
 Included observations: 2605  
 Q-statistic  
 probabilities  
 adjusted for 2  
 ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	-0.001	-0.001	0.0036	
		2	0.002	0.002	0.0147	
		3	0.018	0.018	0.8423	0.359
		4	0.021	0.021	2.0101	0.366
		5	-0.030	-0.030	4.3583	0.225
		6	0.008	0.008	4.5446	0.337
		7	-0.025	-0.026	6.2276	0.285
		8	-0.017	-0.016	6.9519	0.325
		9	-0.012	-0.011	7.3071	0.398
		10	-0.014	-0.014	7.8107	0.452
		11	0.046	0.048	13.303	0.149
		12	-0.025	-0.025	14.927	0.135
		13	-0.015	-0.014	15.490	0.161
		14	0.002	-0.000	15.497	0.215
		15	0.011	0.009	15.834	0.258
		16	-0.001	0.003	15.836	0.323
		17	0.009	0.006	16.035	0.380
		18	-0.024	-0.023	17.573	0.349
		19	-0.040	-0.040	21.681	0.197
		20	0.004	0.004	21.727	0.244
		21	-0.019	-0.018	22.637	0.254
		22	-0.002	-0.002	22.649	0.306
		23	0.026	0.028	24.385	0.275
		24	-0.022	-0.023	25.703	0.265
		25	-0.002	-0.002	25.718	0.314
		26	0.004	-0.001	25.767	0.365
		27	-0.021	-0.022	26.906	0.361
		28	-0.023	-0.023	28.270	0.345
		29	-0.008	-0.009	28.425	0.389
		30	-0.008	-0.004	28.614	0.432
		31	0.013	0.012	29.094	0.460
		32	-0.001	-0.001	29.099	0.512
		33	0.018	0.018	29.954	0.520
		34	-0.001	-0.005	29.955	0.570
		35	-0.005	-0.004	30.013	0.617
		36	0.016	0.015	30.675	0.631

Una vez determinado definitivamente el modelo, nos planteamos, tal y como indicamos al inicio de este apartado, si consideramos algunas observaciones como outliers y las eliminamos de la estimación o, por el contrario, deben formar parte de la muestra pues el modelo es capaz de captar, al menos en parte, estas desviaciones. Se ha considerado como posibles outliers aquellas observaciones que tengan una desviación estándar superior a 0.05. También se ha vuelto a estimar de nuevo el modelo ARMA(1,1)-GARCH-M(2,1), a pesar que ya lo habíamos descartado dado que la

presencia de outliers podría haber provocado precisamente esa mayor persistencia en la volatilidad que nos llevó a su descarte en aras de profundizar en mayores retardos en el modelo GARCH-M que recogieran esa persistencia.

Utilizando el criterio Schwartz y Hanan-Quinn el modelo elegido sigue siendo el modelo ARMA(1,1)-GARCH-M(4,2) teniendo en cuenta todas las observaciones, es decir, sin eliminar los supuestos outliers. En el caso de eliminar las observaciones, para el caso del modelo ARMA(1,1)-GARCH-M(4,2), dejan de ser significativa los parámetros de la estructura ARMA(1,1). El modelo ARMA(1,1)-GARCH-M(2,1) sin outliers sigue descartándose dado que según los criterios Schwartz y Hanan-Quinn se elige el otro modelo, además sus residuos presentan algo más de curtosis y un valor del estadístico Jarque-Bera superior al de los otros modelos y, por tanto, se alejan más de la normalidad de los mismos.

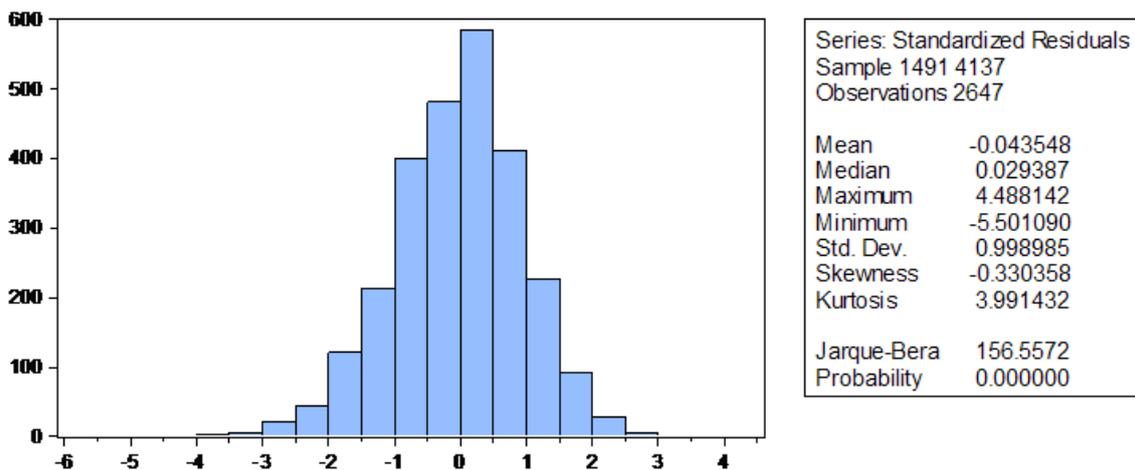


Figura 4-7: Histograma residuos modelo ARMA(1,1)-GARCH(2,1)-M sin outliers

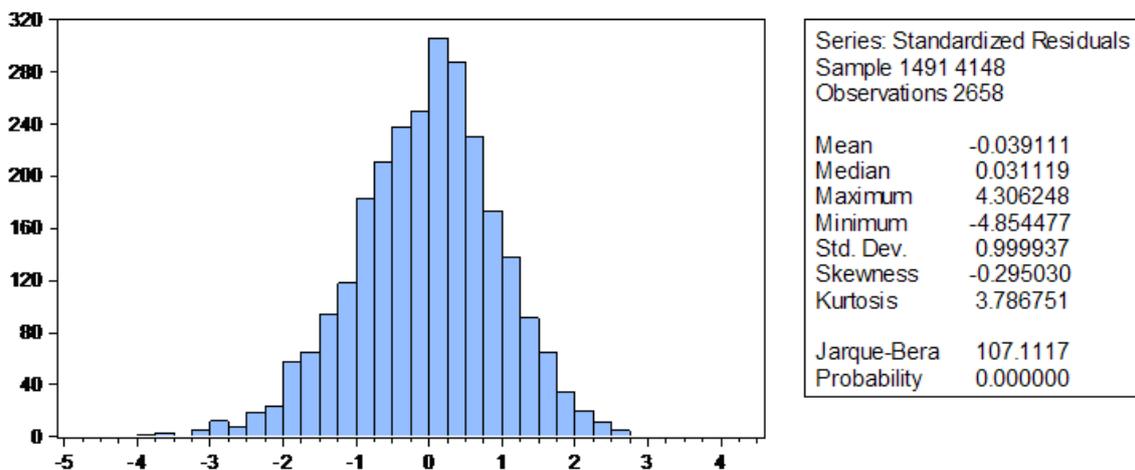


Figura 4-8: Histograma residuos modelo ARMA(1,1)-GARCH(4,2) con todas las observaciones

Una vez elegido el modelo definitivo se ha procedido a la estimación del mismo durante todo el periodo muestral (2658 observaciones) y se han calculado los diferentes tipos de errores obtenidos en la misma tanto para la rentabilidad como para la volatilidad condicional. Los tipos de errores elegidos han sido MAPE (Mean Absolute Percentage Error), MAE (Mean Absolute Error), MSE (Mean Squared Error), MPE (Mean Percentage Error) y RMSE (Root Mean Squared Error).

La estimación del modelo tanto para la rentabilidad (ecuación de la media) como para la volatilidad condicional (ecuación de la varianza) se ha realizado de forma simultánea utilizando el algoritmo Levenberg-Marquardt.

A continuación se presenta un cuadro resumen de los distintos tipos de errores de predicción del modelo ARMA(1,1)-GARCH-M(4,2) (incluyendo la desviación de la varianza condicional en la ecuación de la media e incluyendo el volumen con un retardo en la varianza condicional), tanto para la rentabilidad como para la volatilidad condicional.

**Tabla 4-31: Resultados del modelo ARMA(1,1)-GARCH(4,2)-M con desviación estándar de la varianza condicional en la ecuación de la media y volumen con un retardo en la varianza condicional**

	RT	GARCH
MAPE	1,098356625	2,046341064
MAE	0,010810	0,000234
MSE	0,000238249	0,000000104
MPE	-0,971461761	1,890341047
RMSE	0,015435317	0,000323262

#### **4.2. Estimación y predicción de la volatilidad condicional y la rentabilidad del Ibex-35 mediante la RNA Backpropagation**

Para la estimación y predicción simultánea de la rentabilidad del Ibex-35 y la volatilidad condicional para todo el periodo muestral ya indicado, simplemente hace falta indicar los elementos y estructura que conforman la red.

Se ha optado por utilizar la red Backpropagation (ya vista en el capítulo

3) por su capacidad de adaptación de los pesos de las neuronas ante los errores que se cometen durante el aprendizaje y que dan la posibilidad de dar soluciones muy favorables ante situaciones no aprendidas anteriormente. El aprendizaje de la red está basado, por tanto, en un aprendizaje por corrección del error y es del tipo supervisado, por lo que para su entrenamiento se ha de indicar los inputs de entrada que se tienen y las salidas que se desean se obtengan para dichos inputs.

Para la simulación y entrenamiento de la red neuronal se ha utilizado el programa NeuroGenetic Optimizer de la empresa BioComp Systems Inc, que es una aplicación de 32 bits para Windows que realiza simulación de diversos tipos de redes neuronales.

Generalmente a la red se le indica de todas las observaciones cuales serán utilizadas para el aprendizaje y cuales para testear las diferentes redes que va generándose. En este caso, y dado que esta aplicación es más flexible para esta circunstancia, se le ha indicado que todas las observaciones las utilice tanto para el aprendizaje como para su testeo. Para ello el programa realiza una copia de los datos. Hacer este tipo de mecanismo hace que el aprendizaje sea algo más lento, pero de mayor calidad y por tanto, cometiendo menos errores.

En cuanto a las capas ocultas, este programa permite la existencia de dos (recordemos que para la existencia de la red Backpropagation es necesario como mínimo una capa oculta). Se ha permitido que se establezcan múltiples conexiones entre las diferentes neuronas y se ha escogido el número máximo posible de éstas del programa 256 neuronas para cada capa. Los ratios de aprendizaje establecidos para las dos capas ocultas han sido de 0,1 a 0,4 y para las capas de salida de 0,1 a 0,2. Para el parámetro del momentum se ha establecido para las capas ocultas entre 0,1 a 0,3 y para las capas de salida de 0,1 a 0,2. Estos parámetros se han puesto en forma de rango por lo que la red va optimizando con las redes con diferentes ratios de aprendizaje y momentum.

Las ponderaciones iniciales de los pesos de las conexiones de las neuronas se han establecido por defecto en +/- 0,3 sobre las que el programa realiza diversas pruebas y va optimizando y variando con el aprendizaje.

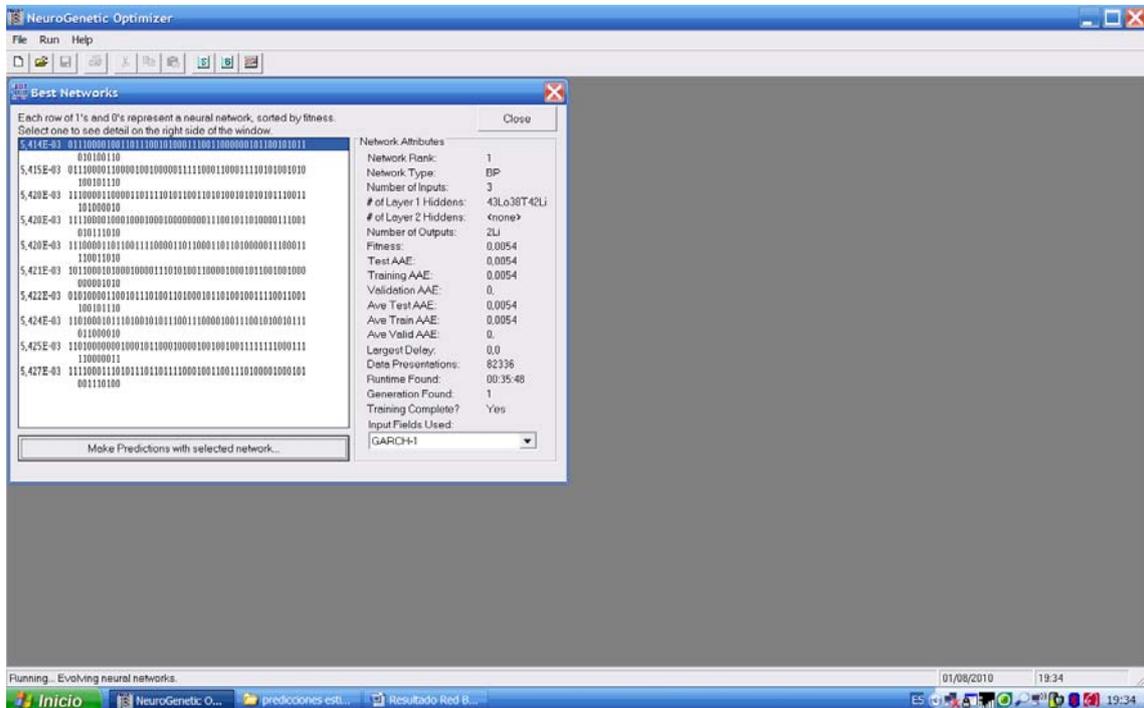
Como se ha indicado el número máximo de neuronas a utilizar en cada capa es el máximo permitido por el programa 256 neuronas por capa. Ahora bien, en cuanto a los parámetros que se establecen en la población de las neuronas son los siguientes: población de 100, la población inicial se establece sin restricciones, el programa localiza aquellas neuronas que son activadas en la primera red que crea para generar el primer patrón, el patrón de mutación de las neuronas se ha establecido de forma aleatoria con una ratio de un 0.25.

Los inputs que se han establecido para el aprendizaje de la red han sido la rentabilidad del índice con un retardo, la varianza condicional con uno y dos retardos y el volumen con un retardo. Para los outputs serán la rentabilidad y la varianza condicional. De esta forma se tienen todos los parámetros utilizados en el modelo econométrico para poderlo comparar. Por un lado la red ya aprende la relación que pueda existir entre la rentabilidad con su retardo (en el modelo econométrico la parte ARMA(1,1)). También al incorporar como input la varianza condicional con uno y dos retardos la red puede establecer la relación existente entre la rentabilidad y la varianza condicional (GARH-M en el modelo econométrico) y también entre la varianza condicional y sus retardos (sería la parte GARCH del modelo econométrico, la parte ARCH como se trata de residuos de estimación la red ya los tiene en cuenta en el aprendizaje ya que se trata de aprendizaje por corrección del error). Con el input volumen con un retardo, se establece la posible relación existente entre este y la varianza condicional. Ahora bien las relaciones que establece la red entre los inputs y outputs son más complejas que las establecidas en el modelo econométrico, por lo que se espera una mejora en los resultados, y por tanto, una reducción en los errores cometidos en la predicción de la rentabilidad del Ibex-35 y la varianza condicional.

Dado que el software NGO, realiza una optimización de los parámetros tanto de aprendizaje, como de los pesos de las neuronas,..etc, se puede establecer que el aprendizaje pretenda conseguir redes que minimicen el Average Absolute Error, Mean Squared Error o Root Mean Squared Error. Se ha procedido a entrenar tres veces la red con cada uno de los criterios de selección de las redes generadas durante el aprendizaje.

## Aprendizaje de la red Backpropagation mediante criterio de selección de red Average Absolute Error

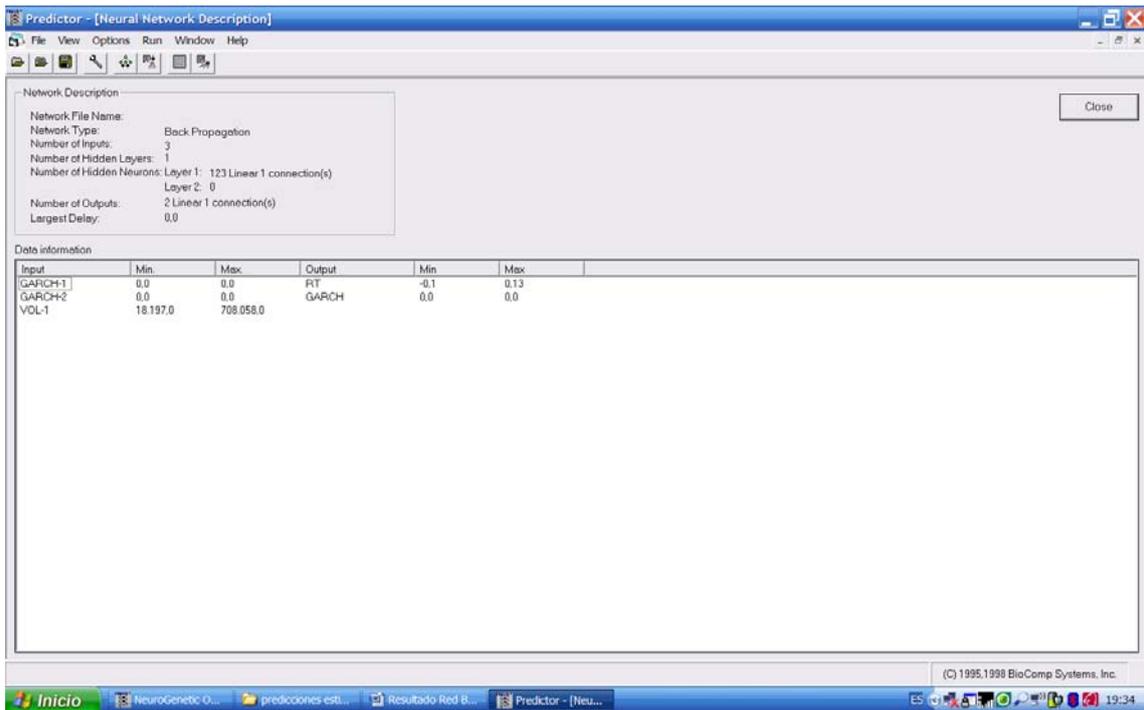
En este caso se generaron diversas redes eligiendo la mejor generada, bajo este criterio, a los 35 minutos y 48 segundos.



**Figura 4-9: Listado Red Backpropagation: criterio Average Absolute Error**

Si bien la red que figura a continuación en el ranking obtiene prácticamente los mismos resultados y además se obtiene en menos de 5 minutos, se ha escogido la primera para ver si mejora más los resultados al estar más tiempo de aprendizaje respecto los otros criterios de selección de redes en los que se han escogido redes generadas en muy pocos segundos.

Los inputs seleccionados por esta red han sido la volatilidad condicional con uno y dos retardos y el volumen con un retardo. Se ha establecido una única capa oculta con 123 neuronas, siendo 43 con funciones de transferencia de tipo logístico sigmoidal, 38 neuronas de tipo tangente hiperbólica y 42 neuronas de tipo lineal. Las neuronas de las capas de salida han utilizado funciones de transferencia de tipo lineal.



**Figura 4-10: Datos Red Backpropagation seleccionada: criterio Average Absolute Error**

A continuación se describen los resultados de los errores de predicción para la rentabilidad y la varianza condicional.

**Tabla 4-32: Resultados Red Neuronal Backpropagation: criterio Average Absolute Error**

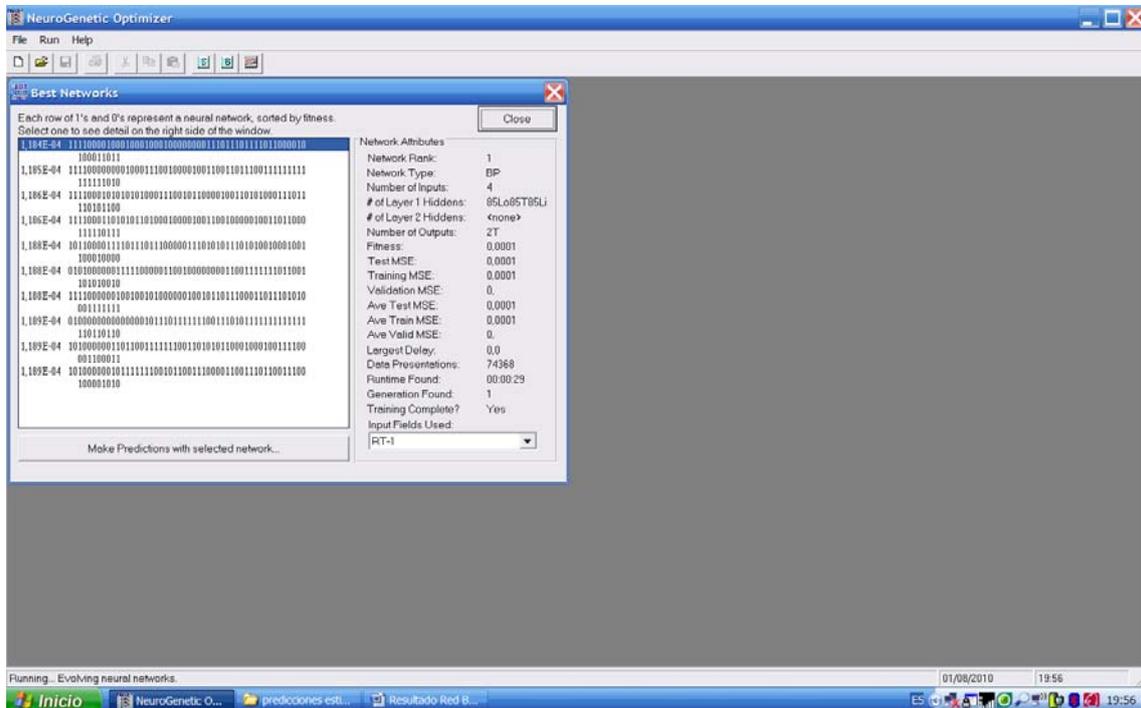
	RT	GARCH
<b>MAPE</b>	<b>1,142042409</b>	<b>0,188839984</b>
<b>MAE</b>	<b>0,010792093</b>	<b>0,0000379</b>
<b>MSE</b>	<b>0,000237482</b>	<b>0,000000006</b>
<b>MPE</b>	<b>-0,99147652</b>	<b>0,104597277</b>
<b>RMSE</b>	<b>0,015410446</b>	<b>0,0000800898</b>

### Aprendizaje de la red Backpropagation mediante criterio de selección de red Mean Squared Error

En este caso se ha elegido una red generada a los 29 segundos, pues se dejó entrenando la red más de media hora sin mejora alguna sobre esta.

Los inputs seleccionados en esta red han sido la rentabilidad con un retardo, la varianza condicional con uno y dos retardos, y el volumen con un retardo.

La red ha utilizado también una sola capa oculta con 255 neuronas, siendo 85 neuronas con función de transferencia de tipo logístico sigmoïdal, 85 neuronas de tipo tangente hiperbólica y 85 neuronas de tipo lineal.



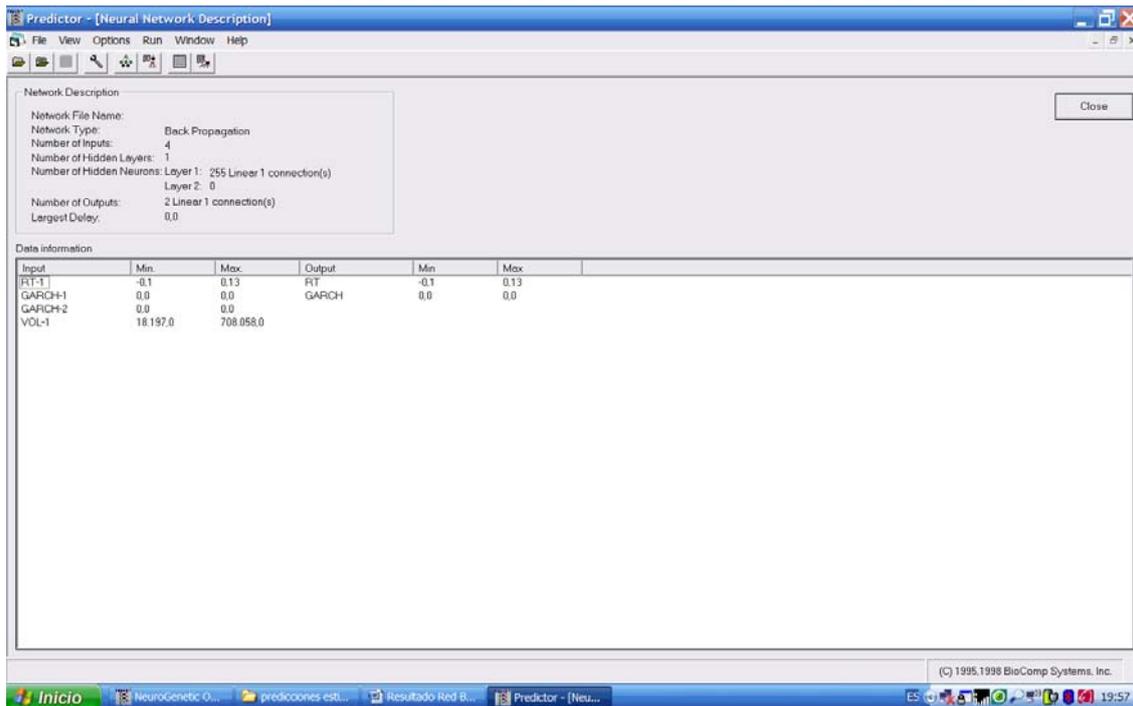
**Figura 4-11: Listado Red Backpropagation: criterio Mean Squared Error**

Las neuronas de salida utilizaron funciones de transferencia de tipo lineal.

Los residuos obtenidos para ambas variables han sido:

**Tabla 4-33: Resultados Red Neuronal Backpropagation: criterio Mean Squared Error**

	RT	GARCH
<b>MAPE</b>	<b>1,071136</b>	<b>0,148348765</b>
<b>MAE</b>	<b>0,010802</b>	<b>0,000037</b>
<b>MSE</b>	<b>0,000236805</b>	<b>0,0000000196</b>
<b>MPE</b>	<b>-0,99804707</b>	<b>-0,006030978</b>
<b>RMSE</b>	<b>0,015388459</b>	<b>0,000140036</b>



**Figura 4-12: Datos Red Backpropagation seleccionada: criterio Mean Squared Error**

### Aprendizaje de la red Backpropagation mediante criterio de selección de red Root Mean Squared Error

En este caso la red elegida ha sido generada en 25 segundos. Los inputs que la red ha seleccionado han sido la rentabilidad con un retardo, la varianza condicional con uno y dos retardos y el volumen con un retardo.

Como en los dos casos anteriores la red se ha construido con una sola capa oculta. En este caso también con 255 neuronas, siendo 85 neuronas con función de transferencia de tipo logístico sigmoideal, 85 neuronas de tipo tangente hiperbólica y 85 neuronas con función de transferencia lineal. Para las neuronas de salida de la red las funciones de transferencia utilizadas han sido de tipo lineal.

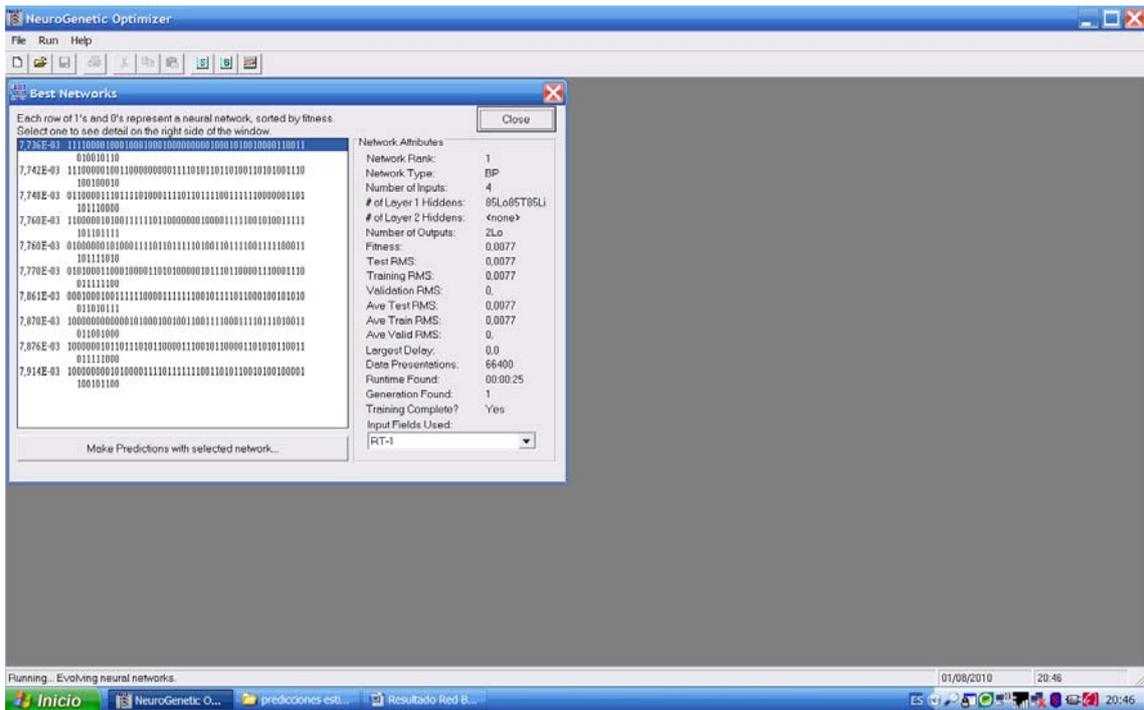


Figura 4-13: Listado Red Backpropagation: criterio Root Mean Squared Error

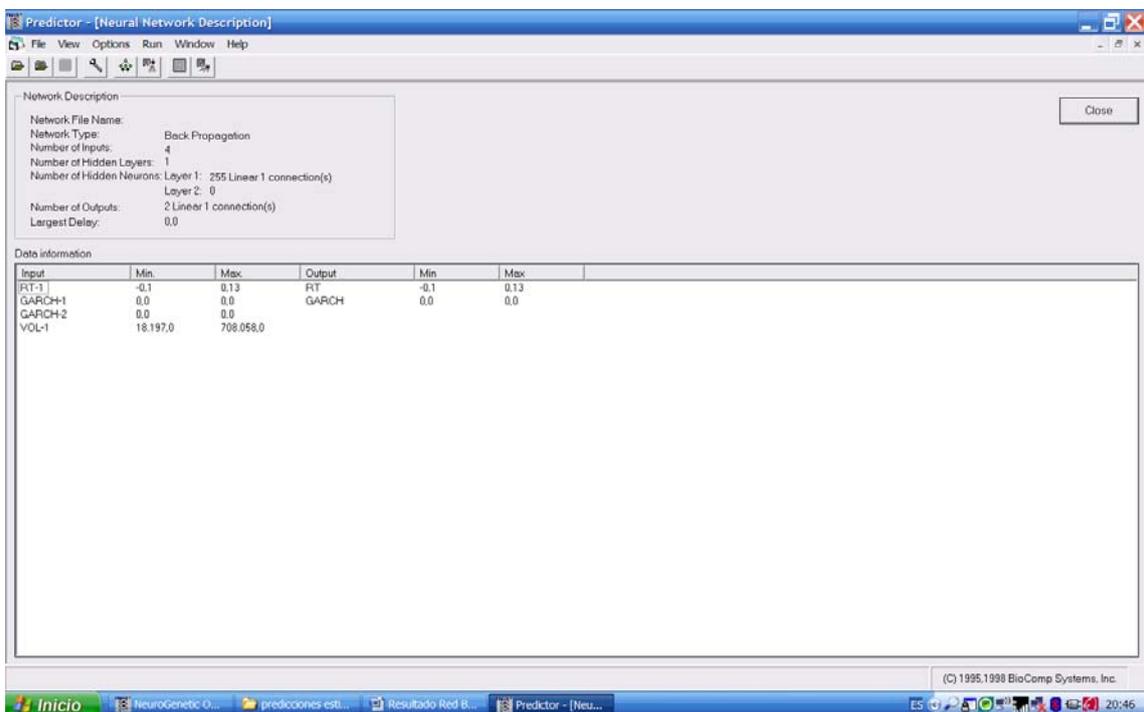


Figura 4-14: Datos Red Backpropagation seleccionada: criterio Root Mean Squared Error

Los errores cometidos en la predicción de ambas variables con esta red neuronal han sido:

**Tabla 4-34: Resultados Red Neuronal Backpropagation: criterio Root Mean Squared Error**

	<b>RT</b>	<b>GARCH</b>
<b>MAPE</b>	<b>1,148959</b>	<b>0,175423745</b>
<b>MAE</b>	<b>0,010844</b>	<b>0,000038</b>
<b>MSE</b>	<b>0,000237048</b>	<b>0,0000000121</b>
<b>MPE</b>	<b>-1,024735411</b>	<b>0,089807449</b>
<b>RMSE</b>	<b>0,015396354</b>	<b>0,000110112</b>

## 5. Conclusiones

Vamos a incorporar un cuadro resumen de todos los resultados obtenidos tanto con el modelo econométrico como con las redes neuronales a efectos de realizar una comparativa y extraer las principales conclusiones.

**Tabla 5-1: Cuadro resumen resultados para la rentabilidad del índice Ibex-35**

	<b>ARMA(1,1)- GARCH- M(4,2)</b>	<b>RED NEURONAL CRIT. AVERAGE ABSOLUTE ERROR</b>	<b>RED NEURONAL CRIT. MEAN SQUARED ERROR</b>	<b>RED NEURONAL CRIT. ROOT MEAN SQUARED ERROR</b>
<b>MAPE</b>	<b>1,098356625</b>	<b>1,142042409</b>	<b>1,071136</b>	<b>1,148959</b>
<b>MAE</b>	<b>0,010810</b>	<b>0,010792093</b>	<b>0,010802</b>	<b>0,010844</b>
<b>MSE</b>	<b>0,000238249</b>	<b>0,000237482</b>	<b>0,000236805</b>	<b>0,000237048</b>
<b>MPE</b>	<b>-0,971461761</b>	<b>-0,99147652</b>	<b>-0,99804707</b>	<b>-1,024735411</b>
<b>RMSE</b>	<b>0,015435317</b>	<b>0,015410446</b>	<b>0,015388459</b>	<b>0,015396354</b>

Una de las primeras conclusiones que podemos obtener es el hecho que para el caso de la estimación y predicción de la rentabilidad, las redes neuronales obtienen de forma aproximada los mismos errores con cualquiera de los tres criterios de selección.

Comparando las redes neuronales con el modelo econométrico podemos afirmar, en general, que las redes neuronales mejoran los resultados. Ahora bien, se trata de mejoras muy reducidas. Esto puede ser debido a varias cuestiones. La primera está relacionada con las variables o inputs que se han incluido para la estimación de la rentabilidad que parece que no recoge suficiente información, con lo que debería incluirse otro tipo de variables que pudieran afectar a la rentabilidad del índice y no sólo sus retardos y volatilidad como en el modelo. El segundo aspecto a tener en cuenta es el tiempo de entrenamiento de las redes. Como se indicó anteriormente, el entrenamiento fue suspendido a la media hora de inicio eligiendo la mejor red encontrada hasta ese momento (a excepción de la primera red que se suspendió a los treinta y cinco minutos). Quizás con un periodo de tiempo de entrenamiento más elevado podría obtenerse redes que mejorasen, con la información disponible, aún más los resultados obtenidos y obteniendo una diferenciación más clara sobre el modelo econométrico.

**Tabla 5-2: Cuadro resumen resultados para la volatilidad condicional del índice Ibex-35**

	<b>ARMA(1,1)- GARCH- M(4,2)</b>	<b>RED NEURONAL CRIT. AVERAGE ABSOLUTE ERROR</b>	<b>RED NEURONAL CRIT. MEAN SQUARED ERROR</b>	<b>RED NEURONAL CRIT. ROOT MEAN SQUARED ERROR</b>
<b>MAPE</b>	<b>2,046341064</b>	<b>0,188839984</b>	<b>0,148348765</b>	<b>0,175423745</b>
<b>MAE</b>	<b>0,000234</b>	<b>0,0000379</b>	<b>0,000037</b>	<b>0,000038</b>
<b>MSE</b>	<b>0,000000104</b>	<b>0,000000006</b>	<b>0,0000000196</b>	<b>0,0000000121</b>
<b>MPE</b>	<b>1,890341047</b>	<b>0,104597277</b>	<b>-0,006030978</b>	<b>0,089807449</b>
<b>RMSE</b>	<b>0,000323262</b>	<b>0,0000800898</b>	<b>0,000140036</b>	<b>0,000110112</b>

En este caso, en los resultados obtenidos con las redes neuronales sí existen algunas pequeñas diferencias en los errores obtenidos en la estimación y predicción para la volatilidad condicional en función del criterio elegido de selección. No obstante, con cualquiera de los tres criterios de elección de la red neuronal, se obtiene una clara mejoría de los resultados respecto el modelo econométrico.

A modo de resumen, podemos afirmar que, al menos, para esta muestra la red neuronal mejora los resultados de los modelos econométricos para estimar tanto la rentabilidad como la varianza condicional dada su mayor flexibilidad y capacidad de obtener relaciones más complejas entre las variables que las que se obtienen mediante un modelo econométrico. No obstante, para dar mayor fiabilidad de los resultados sería conveniente realizar más estudios similares con otros horizontes temporales, otros índices bursátiles, títulos, etc...

Tampoco se han contemplado modelos GARCH más complejos al utilizado, como el asimétrico, que podría explicar y captar mejor el comportamiento pasado y las relaciones existentes entre las variables e, incluso en el periodo de predicción. En la fase de construcción del modelo ARMA-GARCH-M se observó cierta asimetría negativa en los residuos del modelo seleccionado, que, aunque ésta pueda ser mínima, podría recoger mejor el comportamiento y predicción de las variables rentabilidad y varianza condicional del índice Ibex-35.

## 6. Bibliografía

ARAGÓ MANZANA, V.; FERNÁNDEZ IZQUIERDO, M<sup>a</sup>.A. (2004) Modelos GARCH Asimétricos y Volumen de Negociación: Aplicación para el Índice Ibex-35. Revista Española de Financiación y Contabilidad, vol nº 121, pp 443-463

BOLLERSLEV, T. (1986) Generalized Autorregressive Conditional Heteroskedasticity. Journal of Econometrics, pp 307-327

BOLLERSLEV, T.; OLE MIKKELSEN, H. (1996) Modeling and Pricing long memory in Stock Market Volatility. Journal of Econometrics, pp 151-184

BONILLA, M.; OLMEDA, I. (2002) Redes Neuronales Artificiales: Predicción de la Volatilidad del Tipo de Cambio de la Peseta. Editor: Instituto Valenciano de Investigaciones Económicas. WP-EC 2002-08

CHEN, A-S. (1997) Forecasting the S&P 500 Index Volatility. International Review of Economics and Finance, pp 391-404

DALY, K. (2008) Financial Volatility : Issues and Measuring Techniques. Physica A, pp 2377-2393

DEO, M.; SRINIVASAN, K.; DEVANADHEN, K. (2008) The empirical relationship between Stock Returns, Trading Volume and Volatility. European Journal of Economics

ENGLE, R.F. (1982) Autorregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica

ENGLE, R.F.; LILIEN, D.M.; ROBINS, R.P. (1987) Estimating Time-Varying Risk Premia in the Term Structure: the ARCH-M Model. Econometrica, pp 391-407

FRENCH, K.R.; SCHWERT, G.W.; STAMBAUGH, R.F. (1987) Expected Stock Returns Volatility. Journal of Financial Economics, pp 3-29

GARCÍA ESTEVEZ, P. (2002) Aplicaciones de las Redes Neuronales en las Finanzas. Madrid: Universidad Complutense de Madrid.

GARCÍA MONTALVO, J. (1998) Volumen y Volatilidad en mercados financieros. Revista Española de Financiación y Contabilidad nº95, pp 367-393

GATELY, E. (1996) Neural Networks for Financial Forecasting. USA: Editorial John Wiley & Sons Inc.

GHAHRAMANI, M.; THAVANESWARAN, A. (2008) A note on GARCH model Identification. An International Journal computers & Mathematics, pp 2469-2475

GUZMÁN PLATA, M<sup>a</sup>.P. (1998) Los modelos CAPM y ARCH-M. Economía Teoría y Práctica, nº 9.

HENSCHL, L. (1995) All in the family Nesting symmetric and asymmetric GARCH

models. *Journal of Financial Economics*, pp 71-104

HILERA, J.R; MARTÍNEZ, V.J. (1995) *Redes Neuronales Artificiales*. Madrid: Editorial Ra-ma

JOCHUM, C. (1999) Volatility spillovers and the price of risk: Evidence from Swiss stock market. *Empirical Economics*, pp 303-322

KARANASOS, M. (1999) Prediction in ARMA models with GARCH in MEAN Effects. Editor: Universidad de York.

KHO, B-C. (1996) Time-varying risk premia, volatility, and technical trading rule profits: Evidence from foreign currency futures markets. *Journal of Financial Economics*, pp 249-290

LEE, B-S.; RUI, O.M. (2002) The dynamic relationship between stock returns and trading volume. *Journal of Banking & Finance* n° 26, pp 51-78

LEUNG LAI, T.; XING, H. (2008) *Statistical Models and Methods for Financial Markets*. USA: Editorial Springer

LUNDBERGH, S.; TERÄSVIRTA, T. (2002) Evaluating GARCH models. *Journal of Econometrics*, pp 417-435

MONTAÑO MORENO, J.J. (2002) *Redes Neuronales Artificiales aplicadas al Análisis de Datos*. Editorial Universidad de las Islas Baleares.

MUBARIK, F.; JAVID, A. (2009) Relationship between stock return, trading volume and volatility: evidence from Pakistan stock market. *Asia Pacific Journal of Financial and Banking Research* vol 3 n° 3

NELSON, D.B. (1991) Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, pp 347-370

OLMEDA, I.; OTROS. (1996) *Avances Recientes en Predicción Bursátil*. Madrid: Ediciones de la Universidad de Alcalá.

PAGAN, A. (1996) The econometrics of financial markets. *Journal of Empirical Finance*, pp 15-102

PEIRÓ, A. (1992) *La volatilidad del mercado de acciones español*. Valencia: Instituto Valenciano Investigaciones Económicas. WP-EC 92-12

PÉREZ LÓPEZ, C. (2008) *Econometría avanzada. Técnicas y herramientas*. Madrid: Editorial Pearson Prentice Hall

SCHEPPER, A.D.; GOOVAERTS, M.J. (1999) The GARCH(1,1)-M model: results of densities of the variance and the mean. *Insurance Mathematics & Economics*, pp 83-94

TANCO, F. *Introducción a las Redes Neuronales Artificiales*. Editorial Universidad

Tecnológica Nacional de Buenos Aires.

URIEL JIMENEZ, E.; PEIRÓ GIMENEZ, A. (2000) Introducción al análisis de series temporales. Madrid: Editorial AC

URIEL, E.; CONTRERAS, D.; MOLTÓ, M<sup>a</sup>.L.; PEIRÓ, A. (1990) Econometría. Madrid: Editorial AC