

PRONÓSTICO DEL PROMEDIO INDUSTRIAL DOW JONES, APLICANDO REDES NEURONALES ARTIFICIALES

(Recibido: 04 Marzo/2011-aprobado: 04 Abril/2011)

Elsy Lisbeth Gómez Ramos*
Francisco Venegas Martínez**
Héctor Allier Campuzano**

59

Resumen

En este trabajo se compara el pronóstico del Promedio Industrial Dow Jones (DJIA) a través de dos modelos. Uno paramétrico de tipo GARCH y otro no paramétrico como la Red Neuronal Artificial (RNA) perceptrón multicapa. Para ello, se tomó una muestra de 150 observaciones en forma diaria y se pronosticaron 10 periodos. Donde se observa que la RNA logra captar el comportamiento de la serie de tiempo, pero el modelo tipo GARCH presenta un mejor ajuste dentro y fuera de la muestra.

Abstract

This paper compares the forecasting of Dow Jones Industrial Average (DJIA) through two models. A parametric model: GARCH and a non parametric one: Artificial Neural Network (ANN) multilayer perceptron. It was taken into account a sample of 150 observations on a daily basis and it was predicted 10 periods in both models. It was found that the ANN can detect the behavior of the time series but the GARCH model present more accurate forecasting.

Palabras clave: Pronóstico, Red Neuronal Artificial y perceptrón multicapa.

* Alumna del Doctorado en Ciencias Económicas, IPN.

** Profesores de la sección de Estudios de Posgrado e Investigación, ESE-IPN.

1. Introducción

Los modelos paramétricos han sido tradicionalmente aplicados a mercados bursátiles y han probado en diversos estudios ser efectivos (por ejemplo, el modelo GARCH¹). Estos modelos parten de una función de distribución, y reducen el problema a estimar los parámetros que mejor se ajustan a las observaciones de la muestra. Además, de que resultan ser muy potentes cuando el proceso generador de datos sigue una distribución propuesta, en la teoría financiera los resultados más conocidos se han obtenido a través de la distribución normal o gaussiana.

Sin embargo, la aplicación de modelos no paramétricos, en especial las Redes Neuronales Artificiales (RNA), han permitido en años recientes dar soluciones más precisas a problemas complejos, por lo que se ha expandido su aplicación debido a las limitaciones que presentan los modelos paramétricos (como la violación de supuestos del modelo de regresión lineal, por ejemplo: que las observaciones no estén normalmente distribuidas) al ser aplicados a fenómenos financieros, ya que éstas en sí son complejas para su estudio, en parte porque presentan volatilidad y por lo tanto son sensibles a las condiciones del mercado.

Por su parte, las RNA no se encuentran sujetas a ninguna forma funcional, por lo que presentan pocas restricciones y permiten *reconstruir* la función de clasificación, es decir emplean formas funcionales flexibles que se aproximen a la función objetivo, por lo que el problema consiste en calcular los parámetros de una función. La aplicación de RNA, para efectos de pronóstico se puede ver como un problema de aproximación de una función en la cual los valores de la función son representados por la serie de tiempo. (Mehrotra *et al.*, (2000), pág. 31.)

A pesar de las bondades que presentan las RNA también tienen sus limitaciones (Brooks (2002), pág. 649) como:

- a) Los coeficientes estimados resultantes de la red no tiene ninguna interpretación real,
- b) No existen pruebas específicas disponibles para considerar que un modelo sea el adecuado, y
- c) Ofrecen resultados excelentes dentro de la muestra pero fuera de ésta los pronósticos carecen de exactitud.

Sin embargo, hay que considerar que en el inciso (a) no es posible que los coeficientes estimados (pesos) tengan una interpretación debido a que éstos solo buscan aproximarse a la función generadora de los datos y no dar información sobre ajuste del modelo ya que no es un modelo paramétrico. Aunque, el inciso (b) y (c) son ciertos cuando aplicamos métodos de optimización manuales (prueba y error) como el que se presenta en esta investigación, no así

¹ Por sus siglas en inglés: Generalized autoregressive conditional heteroscedasticity.

para métodos de optimización evolutiva.² donde estos localizan un óptimo global cercano al elemento a evolucionar y posteriormente se aplica el algoritmo tradicional de la RNA.

Sin embargo, los métodos manuales ofrecen flexibilidad en su aplicación, debido a que el diseñador puede basarse en experiencias anteriores o en proponer diversos criterios para la selección de la red (análisis del error, criterio de información Schwarz, número de predicciones correctas, etc.) lo que permite de cierto modo, entender el funcionamiento de la red de una forma sencilla, aunque efectivamente carece de ajuste. El hecho de establecer un método universal no es una tarea sencilla, debido a que en la literatura existe una inmensa variedad no solo de criterios de selección de redes, sino también, de utilizar variantes (término momento) de la red como una alternativa o combinaciones de estas dos últimas o métodos evolutivos o simplemente utilizar distintos tipos de redes (modulares, recurrentes, perceptrón multicapa, etc.). Por lo que, se hace prescindible establecer acercamientos con este tipo de modelos a través de diferentes enfoques e ir estableciendo la dirección que mejores resultados arroje.

Las investigaciones que aplican métodos manuales, no siempre consideran la posibilidad de comparar dicho modelo con técnicas parecidas o de diferente naturaleza, en algunos casos se enfocan en comprobar si tales modelos pueden ser aplicados a detectar problemas, como quiebras corporativas (Bo y Zheng, 2006) donde los autores a través de estados financieros logran obtener buenos resultados; o comprobar hipótesis en los mercados financieros (White, 1996) donde a través de los rendimientos de una acción se logra tener evidencia del rechazo de la hipótesis de mercados eficientes.

Por otro lado, existe literatura que se basa en comparar la RNA con modelos adaptativos y no adaptativos lineales para series económicas (Swanson y White, 1997), donde se muestra la superioridad de la red sobre dichos modelos, aunque, es deficiente al predecir el cambio en la dirección de la variable. Pero cuando el comparativo se realiza con modelos de algoritmos genéticos para índices bursátiles asiáticos (Parisi *et al.*, 2006) estos últimos obtuvieron mejor desempeño en términos de *riesgo-retorno*, no obstante la red lo superó en términos de capacidad de predicción del signo de la variable. También, se han realizado aplicaciones con el modelo Black-Scholes (Malliaris y Salchenberger, 1996) donde el modelo basado en volatilidad histórica (red) muestra un pronóstico más preciso.

Recientemente, se han orientado las investigaciones a hacer evolucionar la red. Un caso especial, es la aplicación en el mercado de derivados (Karathanasopoulos *et al.*, 2010) en Grecia, donde los autores hacen evolucionar las entradas a la red. Y para índices bursátiles esta el caso del TEPIX (Tehran Stock Exchange Prices Indexes) (Hadavandi *et al.*, 2010) donde se hace evolucionar los pesos de la red y se comparan los pronósticos con un modelo

² Cuando hablamos de algoritmos evolutivos, generalmente nos referimos a tres grandes grupos: (a) La programación evolutiva, (b) Estrategias evolutivas y (c) Algoritmos genéticos. Aunque, en esencia existen grandes similitudes entre estas categorías lo que las diferencia es el grado de utilización del operador de mutación y en la reproducción y selección de estrategias (Mehrotra *et al.*, 2000, pág. 289).

de genética difusa, una red constructiva (manual) y la red evolutiva, los autores llegan a la conclusión de que la red evolutiva realiza el pronóstico más preciso.

La ventaja que proporciona nuestra investigación, es que presenta de manera accesible el manejo de RNA para efectos de pronóstico, es decir, se presenta paso a paso la construcción de la red y a su vez incluimos parte de la programación empleada para ejemplificar la metodología utilizada. Por lo que, nuestro objetivo es aplicar una RNA y un modelo tipo GARCH a la serie de tiempo del DJIA y analizar el desempeño de la red, que en nuestro caso será el perceptrón multicapa por ser la red más ampliamente utilizada.

2. Modelos de Volatilidad

62 Los modelos autorregresivos integrados de media móvil (ARIMA)³ o Box-Jenkins suponen que la volatilidad de cualquier variable modelada es constante en el tiempo. En muchos casos esto no es verdad. Engle (1982) introduce una nueva clase de procesos estocásticos llamados modelos de heteroscedasticidad condicional autorregresiva (ARCH), en los cuales la varianza condicionada a la información pasada no es constante y depende del cuadrado de las innovaciones pasadas. El modelo (ARCH) fue desarrollado con el fin de capturar esta propiedad en las series de tiempo financieras. Bajo este modelo, la ecuación de la media condicional (la cual describe como la variable dependiente, Y_t , varía en el tiempo) puede tomar casi cualquier forma que se desee.

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \varepsilon_t$$

$$\varepsilon_t \sim N(0, h_t)$$
(1)

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$
(2)

Donde el modelo anterior, puede ser extendido al caso más general donde el error de la varianza depende de q rezagos del error al cuadrado, el cual se le conoce como ARCH (q):

$$\text{Var}(\varepsilon_t) = \sigma_t^2 = \alpha_0 + \alpha \varepsilon_{t-1}^2 + \alpha \varepsilon_{t-2}^2 + \dots + X_p \varepsilon_{t-p}^2$$
(3)

Una generalización del modelo ARCH es el modelo GARCH propuesto por Bollerslev (1986) y Taylor (1986), en la cual la varianza condicional de, ε_t es dependiente no solamente de las perturbaciones o errores al cuadrado, sino también sobre las varianzas condicionales pasadas, en lo que subyace algún tipo de mecanismo de aprendizaje adaptativo. De modo que una estructura GARCH (p, q) queda de la forma siguiente:

$$\text{Var}(\varepsilon_t) = \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 + \varphi_1 \sigma_t^2 \dots + \varphi_q \sigma_{t-q}^2$$
(4)

³ Conocidos también como modelos homoscedásticos.

Las características principales del GARCH⁴ son la no linealidad; es decir, los modelos lineales son incapaces de explicar efectos relevantes de las series financieras, por lo que se hace necesario el uso de modelos paramétricos no lineales que permitan incluir por ejemplo la tendencia de los rendimientos de los activos financieros que se caracterizan por una distribución leptocúrtica (colas gordas y exceso de apuntamiento en la media).

Si es un modelo GARCH (1,1), σ_t^2 , es conocida como la varianza condicional desde que se estima el primer periodo hacia delante para la varianza calculada basada en cualquier información pasada relevante. Para poder estimar un GARCH (1,1) primero se debe especificar las ecuaciones apropiadas para la media y la varianza ecuaciones (5 y 6) y especificar la función de máxima verosimilitud ecuación (7) que maximiza bajo el supuesto de normalidad de los errores.

$$Y_t = \mu + \delta Y_{t-1} + \varepsilon_t \quad (5)$$

$$\varepsilon_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \varphi_1 \sigma_{t-1}^2 \quad (6)$$

$$L = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^T \frac{(y_t - \mu - \delta Y_{t-1})^2}{\sigma_t^2} \quad (7)$$

Desde el desarrollo del modelo GARCH un gran número de extensiones y variantes han sido propuestas, esto debido a los problemas previstos con el modelo GARCH estándar; por ejemplo: (a) que las condiciones de no negatividad pueden ser violadas por el modelo estimado, (b) no se consideran los efectos de apalancamiento, y (c) el modelo no permite reacción directa alguna entre la varianza condicional y la media condicional dado que se impone una respuesta simétrica de la volatilidad positiva o negativa ante los shocks (Brooks (2002), pág.468).

3. Red Neuronal Artificial tipo Perceptrón multicapa

El cerebro humano se caracteriza por su gran capacidad de procesar información y su rapidez de respuesta han hecho que se intenten desarrollar sistemas que imiten, al menos en parte sus características. Su capacidad para realizar tareas como el reconocimiento, el recuerdo o la clasificación se han intentado plasmar en diferentes tipos de sistemas.

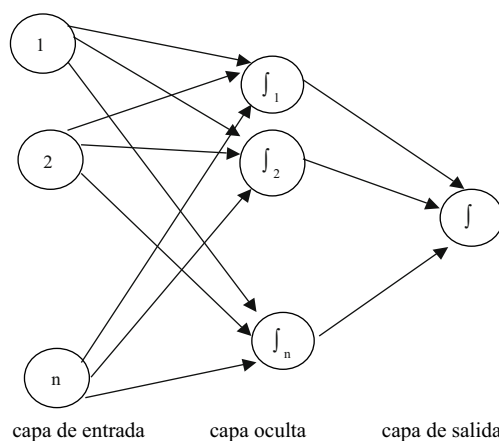
⁴ Se supone una media cero y una varianza condicional: $\sigma_t^2 = \delta + \alpha \varepsilon_{t-1}^2 + \varphi \sigma_{t-1}^2$ donde: $\varepsilon_t = \sqrt{\sigma_t^2} v_t$ y $v_t \sim iid$; además de que $\delta > 0$; $\alpha \geq 0$; $\varphi \geq 0$ -1 $\alpha + \varphi < 1$. Así mismo, se tiene que v_t es un proceso de ruido blanco con varianza uno y ε_t, v_t son independientes.

Las RNA son sistemas distribuidos,⁵ adaptativos⁶ y generalmente con mecanismos de aprendizaje no lineal, contruidos por diversos procesadores elementales o neuronas. A diferencia de otras técnicas, que consideran al cerebro humano como una caja negra, las RNA intentan modelar su estructura fisiológica básica: la neurona, así como la agrupación de neuronas en sistemas que pueden mostrar un comportamiento que se puede considerar, en alguna medida, inteligente.

La red que utilizaremos es de tipo perceptrón multicapa, que generalmente incluye una capa de entrada, otra de salida y al menos una capa oculta. Aunque no existen límites teóricos en el número de capas ocultas, normalmente hay una o dos, habiéndose comprobado experimentalmente que se pueden resolver problemas de cualquier complejidad utilizando un máximo de 3 capas ocultas. Las capas de entrada y salida indican el flujo de información durante la fase de recuerdo u operación normal de la red. El número de capas y de neuronas por capa son decisiones importantes que se deben tomar para cada aplicación particular, aunque se pueden considerar algunas reglas generales, basadas en experiencias anteriores (Gráfica 1).

En este tipo de red se puede determinar el error cometido por las unidades de salida. Sin embargo, para las neuronas de las capas ocultas no se puede determinar tal valor. Por ello, se ha extendido la regla Delta a redes con capas ocultas (regla delta generalizada) con conexiones hacia adelante y con neuronas cuyas funciones de activación sean continuas, dando lugar al algoritmo de retropropagación. La superficie de error y el mecanismo de aprendizaje buscan el error mínimo, modificando para ello los pesos en una cantidad proporcional al gradiente decreciente de la función de error. (Pérez y Martín (2003), pág. 40).

Gráfica 1
Arquitectura de una red tipo perceptrón multicapa



⁵ Cuentan con muchas neuronas a través de las cuales distribuyen su memoria

⁶ Tienen la capacidad de adaptarse al entorno modificando sus pesos y sinapsis de manera de encontrar una solución aceptable al problema.

La red aprende un conjunto predefinido de patrones de entrenamiento empleando un ciclo de propagación y adaptación en dos fases. Inicialmente se asignan a la red pesos pequeños y aleatorios. Se presenta una entrada a la red y se realiza un procesamiento de la información, a través de las capas, obteniendo una salida. Se compara dicha salida con la esperada y se calcula una señal de error para cada neurona de salida. Tales señales de error se propagan hacia atrás, desde la capa de salida hacia las neuronas de la capa intermedia anterior que contribuyen directamente a la salida. La señal recibida por cada una de ellas será una fracción de la señal de error total, tomada en función de la aportación de la neurona correspondiente a la salida obtenida.

El proceso se repite en todas las capas, hasta que cada neurona reciba una señal de error proporcional a su contribución a la salida. Cada neurona actualizará sus pesos en función de la señal de error que reciba. Lo cual permite que, a medida que avanza el entrenamiento, las neuronas de las capas ocultas se organicen a sí mismas, de forma que cada una aprende a reconocer distintas características del espacio de entrada. Para actualizar los pesos, se calcula un término de error asociado a cada neurona. El término de error de las neuronas de salida es igual al error entre salida esperada y obtenida, multiplicado por la derivada de la función de salida.

El término de error de las neuronas ocultas es igual al producto de la derivada de la función de salida de dichas neuronas por la suma ponderada de los términos de error de las neuronas de la siguiente capa con los pesos de las conexiones correspondientes. Es decir, como no se puede calcular directamente el error cometido por cada neurona oculta, pues no se conoce su salida esperada, se considera que es proporcional a la suma de los errores asociados a las neuronas a las que está conectada su salida, multiplicados éstos por los pesos de las conexiones correspondientes.

Los pesos se actualizan empezando por la capa de salida y continuando hacia la de entrada. El incremento en el peso de una neurona será proporcional al producto del término de error de la neurona por la salida que ha generado. El término de proporcionalidad es denominado tasa de aprendizaje y juega un papel fundamental en el entrenamiento, ya que condiciona la velocidad de aprendizaje. El proceso de aprendizaje se repite hasta que la suma de los cuadrados de los términos de error de las neuronas de salida tome un valor aceptable para el conjunto de patrones de entrenamiento (Mehrotra *et al.*, (2000), pág. 70.). Una vez concluido el entrenamiento, la siguiente fase es de prueba donde los pesos quedaran fijos para poder realizar el pronóstico de nuestra serie.

4. Aplicación de los modelos

Las series financieras y en especial los índices bursátiles se caracterizan en esencia por:

- a) Volatilidad *clustering* (la volatilidad aparece en grupos, es decir, cuando tenemos rendimientos altos esperamos que sigan altos, y viceversa),

- b) Efectos de apalancamiento (la volatilidad incrementa más seguida de una caída en los precios que seguida del incremento de estos) y
- c) Leptocurtosis (colas gordas y exceso de apuntamiento en la media). *Aspectos que el modelo GARCH-simétrico es incapaz de capturar* (Brooks, 2002, pág. 468).

El modelo GARCH-simétrico es el que se aplicará en esta investigación, por lo que, se hace innecesaria realizar pruebas de volatilidad y/o justificar el periodo de estudio que implique una relación directa con las características de las series financieras. Si fuera el caso de aplicar modelos GARCH-asimétricos (por ejemplo: EGARCH, GARCH, etc.) u otra extensión, se justificarían las pruebas de volatilidad y/o de la muestra. Por lo que, el periodo que se toma como muestra en este trabajo, es solo un ejercicio para aplicar el modelo GARCH-simétrico y la RNA cada uno con sus alcances, limitaciones y técnicas seleccionadas para valorar su desempeño para el pronóstico del DJIA.

66

4.1 Muestra

Los datos que se analizan corresponden a los precios de cierre en el periodo comprendido entre el 7 de junio de 2010 al 6 de enero de 2011 en forma diaria para tener un total de 150 observaciones del DJIA. La fuente de los datos es de *Económica*, debido a su alto grado de confiabilidad.

4.2 Metodología de la RNA tipo perceptrón multicapa

El software que utilizaremos para nuestra red será el Mathematica 6.0. La variante que usaremos es con el término momento y el método de diseño es constructivo.⁷ A continuación se describe los pasos más sobresalientes para la elaboración del perceptrón multicapa.

I. Inicialización de los pesos

Sera entre -0.1 y 0.1, tanto para la capa oculta y de salida. Hay que recordar, que la capa de entrada no realiza ningún tipo de procesamiento pero si influye en el número de pesos de la red y en la distribución de los datos hacia la capa oculta.

$$\begin{aligned} \text{hidWts} &= \text{Table}[\text{Table}[\text{Random}[\text{Real}, \{-0.1, 0.1\}], \{\text{inNumber}\}], \{\text{hidNumber}\}] \\ \text{outWts} &= \text{Table}[\text{Table}[\text{Random}[\text{Real}, \{-0.1, 0.1\}], \{\text{hidNumber}\}], \{\text{outNumber}\}] \end{aligned}$$

⁷ Los métodos constructivos intentan adaptar el tamaño de la red al problema, comenzando con una red pequeña y añadiendo las capas y unidades necesarias hasta que se encuentre la solución. La ventaja principal es que no se necesita hacer una estimación a priori del tamaño de la red; sin embargo, este tipo de método se basa en la experiencia del diseñador lo cual hace susceptible a no encontrar la arquitectura óptima (Castillo et al. (2001), pág. 2.).

II. Entradas y salidas

Son variaciones logarítmicas rezagadas del índice bursátil. Las entradas corresponden al número de neuronas en la capa de entrada y las salidas al valor deseado como resultado de las entradas propuestas. Estos vectores se seleccionan en forma aleatoria y se especifican tanto las entradas como las salidas.

$$ioP=ioPairs[[Random[Integer,\{1,Length[ioPairs]\}]]][\{-0.00091,-0.00053,-0.00593,-0.0003,0.0008,0.00312,-0.00798,0.00193,0.00185,0.0009\},\{0.0019\}]$$

III. Selección de la topología de la red

La topología se refiere al número de neuronas y capas de la red. La arquitectura comprende la topología, la conectividad y la función de transferencia de cada neurona. Las dos últimas las mantendremos fijas (es decir, la conectividad es hacia adelante y la función de transferencia será la función sigmoidea) y nos concentramos únicamente en la topología de nuestra red; para esto consideramos el error medio cuadrático (MSE) y observamos su comportamiento al aumentar el número de neuronas por capa. A continuación presentamos los resultados obtenidos de las tres topologías propuestas para el perceptrón multicapa.

Cuadro 1
Topologías propuestas para el perceptrón multicapa

Red	Etapas	MSE
3-2-1	Entrenamiento	0.004834
	Primer rezago	0.003657
	Múltiples rezagos	0.009601
5-3-1	Entrenamiento	0.003071
	Primer rezago	0.002151
	Múltiples rezagos	0.002959
10-6-1	Entrenamiento	0.000528
	Primer rezago	0.000405
	Múltiples rezagos	0.001382

En el *cuadro 1* se muestra el comportamiento del MSE a través de diferentes topologías propuestas, lo importante de analizar es el hecho de que el error disminuya en forma considerable al aumentar el número de neuronas, de otro modo nos indicaría que la red tiene

un tamaño superior al requerido por el problema. Otro hecho, es el error en los múltiples rezagos⁸ de la etapa de prueba, puesto que este nos indican el desempeño de la red fuera de la muestra (Chakraborty *et al.*, (1992)).

IV. Tasa de aprendizaje y término momento

La tasa de aprendizaje la definimos como *eta* y el término momento como *alpha*. Estos valores se consideraron fijos para todas las topologías propuestas (cuadro 1). Estas reglas de aprendizaje forman parte crucial en la actualización de los pesos, tanto de la capa oculta como la de salida.

68

$$\begin{aligned} \text{eta} &= 0.9, \text{ alpha} = 0.6 \\ \text{hidLastDelta} &= \text{Table}[\text{Table}[0, \{\text{inNumber}\}], \{\text{hidNumber}\}]; \\ \text{hidLastDelta} &= \text{etaOuter}[\text{Times}, \text{hidDelta}, \text{inputs}] + \text{alpha hidLastDelta} \end{aligned}$$

$$\begin{aligned} \text{outLastDelta} &= \text{Table}[\text{Table}[0, \{\text{hidNumber}\}], \{\text{outNumber}\}] \\ \text{outLastDelta} &= \text{eta Outer}[\text{Times}, \text{outDelta}, \text{hidOuts}] + \text{alpha outLastDelta} \end{aligned}$$

V. Función de activación

La función que aplicaremos será la sigmoidea, para todas las topologías propuestas. Recordemos que únicamente la capa oculta y de salida realiza procesamiento de datos y que la capa de entrada solo los distribuye.

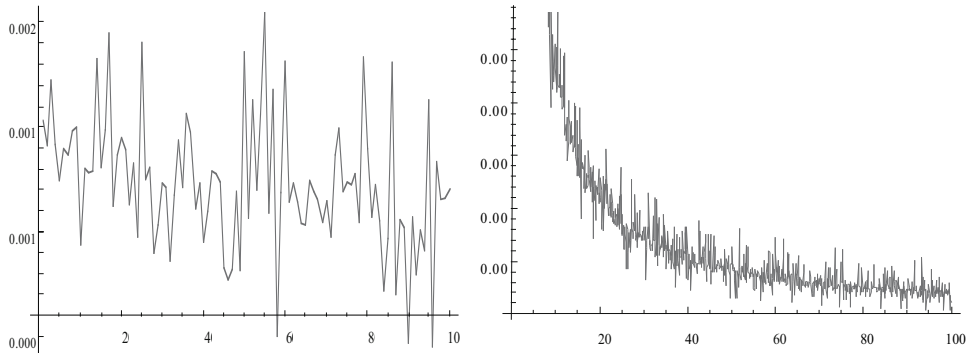
$$\text{sigmoid}[x_]=1/(1+e^{-x})$$

VI. Etapa de entrenamiento

Esta etapa comienza con la selección de datos destinados al entrenamiento de la red y concluye con la minimización de error (Gráfico 2) a través de las iteraciones para encontrar los pesos óptimos. En el eje de las ordenadas indica el error cometido, mientras que en las abscisas son el número de las iteraciones realizadas.

⁸ En la etapa de prueba se divide en la predicción con un rezago donde el valor está basado solo en los valores pasados y en la predicción de múltiples rezagos donde se consideran valores pronosticados para predecir valores futuros. Por ejemplo, si tenemos 5 entradas para la red, el primer valor pronosticado será n_6 como resultado de las entradas i_1, \dots, i_5 . Por lo que, el próximo valor a predecir será n_7 el cual se realizara utilizando las entradas i_2, \dots, i_5, n_6 , y para la siguiente predicción n_8 las entradas a la red serán i_3, i_4, i_5, n_6, n_7 . Pero si nuestra predicción es solo de un rezago, para obtener n_8 únicamente usaremos la información actual i_3, i_4, i_5, i_6, i_7 (Mehrotra *et al.*, (2000), pág. 32.)

Gráfico 2
Comportamiento del error de la RNA 100 y 1000 iteraciones



VII. Etapa de prueba

Aquí los pesos óptimos (capa oculta y de salida) recabados de la fase anterior, serán valores fijos para esta fase lo que nos permitirá realizar el pronóstico de nuestra serie.

```
hidWts+= {-0.0990461,0.0628403,-0.0752713,-0.0461095
,0.0986236,0.0469558,0.00911645 ,0.0696739,0.0938065,-
0.0584878}, {-0.0260153,0.020944,-0.0217435,0.0981858,0.038
896,-0.068399,0.0555709,0.0183582,0.0712103,-0.0571961}, {-
0.0323587,0.0236022,0.0945302,-0.049989,-0.0334737,0.0604
865,0.0695797,0.0957328,-0.0321182,0.0073245}, {-0.039643,-
0.0738314,-0.026193,-0.0341463,0.0862136,0.00498088,0.0952333,
-0.032619,-0.0528586,-0.0268982}, {-0.0606153,0.0489094,-0.0243469,-
0.0697712,0.0715713,-0.0749605,-0.0191944 ,0.0798971,0.00489365,-
0.0357111}, {0.0109588,0.0840674,-0.0633005,0.0569097,-0.0495432,
0.0576519,0.0626739,-0.0092941,-0.0358029,-0.0474587}}

outWts+= {-1.22977,-1.37388,-1.38002,-1.21776,-1.26934,-1.32275}}
```

4.3 Metodología del modelo tipo GARCH

Los pasos más sobresalientes que se siguieron para la elaboración del GARCH en EView-5 fueron los siguientes:

- a) Se utilizaron variaciones logarítmicas y se comprobó si tal transformación haría a la serie del DJIA estacionaria, por lo que, se realiza la prueba de raíz unitaria (cuadro 2) donde se muestra que la serie es estacionaria al tener una probabilidad menor a 0.05.

Cuadro 2

Prueba de contraste de raíz unitaria

Augmented Dickey-Fuller Unit Root Test on IDJ			
Null Hypothesis: IDJ has a unit root			
Exogenous: Constant			
Lag Length: 0 (Automatic based on SIC, MAXLAG=13)			
	<i>t</i> -Statistic	<i>Prob.</i> *	
Augmented Dickey-Fuller test statistic	-12.10725	0.0000	
Test critical values:	1% level	-3.474874	
	5% level	-2.880987	
	10% level	-2.577219	

- b) Se procedió a modelar primero con la metodología ARMA con la finalidad de tener un mejor ajuste en el modelo. Después de revisar el correlograma se pueden identificar la combinación posible, en nuestro caso fue ARMA (7,7) que se identifican, a través de la función de autocorrelación (FA) y de la función de autocorrelación parcial (FAP) quedando aquellas que sean significativas, es decir que la probabilidad sea menor de 0.05.
- c) Posteriormente, se estimo el GARCH (p, q) (cuadro 3). Donde el número máximo de rezagos para el modelo, se identifica en el correlograma de los residuales al cuadrado, que deben cumplir con: i) No negatividad en la ecuación de la varianza, ii) Que la sumatoria de los coeficientes de tales rezagos sea menor a 1, y iii) Verificar la significatividad individual de los coeficientes a través del Estadístico z (Perez, 2006, pág. 436)

Cuadro 3

Modelo GARCH (1,1)

	<i>Coefficient</i>	<i>Std. Error</i>	<i>z</i> -Statistic	<i>Prob.</i>
C	-0.000534	0.000290	-1.842270	0.0654
AR(7)	-0.341942	0.042782	-7.992668	0.0000
AR(1)	0.552062	0.041107	13.33241	0.0000
MA(7)	0.465193	0.015348	30.30905	0.0000
MA(1)	-0.016110	0.016110	-39.07684	0.0000
<i>Variance Equatiob</i>				
C	3.36E-07	5.35E-07	0.626868	0.5307
RESID(-7)^2	0.079030	0.045595	1.733324	0.0830
GARCH(-1)	0.891975	0.066400	13.43333	0.0000
R-squared	0.136666	Mean dependent var		-0.000357
Adjusted R-squared	0.091566	S.D. dependent var		0.003823
S.E. of regression	0.003644	Akaike info criterion		-8.369468
Sum squared resid	0.001779	Schwarz criterion		-8.202942
Log likelihood	602.2322	F-statistic		3.030311
Durbin-Watson stat	2.022370	Prob (F-statistic)		0.005457

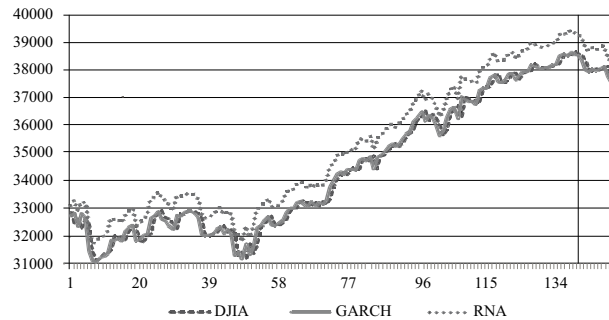
4.4 Comparativo de los pronósticos

Para realizar el comparativo se procedió a invertir las transformaciones para obtener el dato a nivel.⁹ En el gráfico 3, se ilustra el pronóstico del modelo GARCH el cual muy potente y

⁹ La muestra original es de 150 observaciones, pero al realizar las variaciones logarítmicas se pierde un dato menos los diez rezagos de la red solo contaremos con 139 observaciones. Las cuales serán analizadas en este apartado.

logra un buen ajuste, Por el lado de la RNA, se aprecia un pronóstico que sobrevalua prácticamente a lo largo de toda la serie; por lo que, se puede explicar el comportamiento de los estadísticos tanto fuera como dentro de la muestra (cuadro 4).

Gráfica 3
Comparativo de los pronósticos fuera y dentro de la muestra a nivel



Gráfica 4
Estadísticos descriptivos dentro y fuera de la muestra

DJIA	RNA	GARCH		DJIA	RNA	GARCH
10794.74	10998.84	10771.52	Media	11744.21	11958.25	11727.95
10812.04	11002.85	10755.64	Mediana	11743.67	11955.58	11721.13
11722.89	11923.94	11664.52	Maximo	11837.93	12062.85	11831.61
9686.48	9879.06	9669.59	Minimo	11637.45	11858.56	11631.22
518.59	524.8061	507.84	Dev. Std.	72.51	68.33	67.14
-0.11	-0.1	-0.11	Simetria	-0.03	0.19	0.23
1.92	1.9	1.93	Curtosis	1.55	1.87	1.87
7.01	7.20	6.82	Jarque-Bera	0.87	0.58	0.61
0.02	0.02	0.03	Probabilidad	0.64	0.74	0.73
1500468	1528839	1497241	Suma	117442.1	119582.5	117279.5
37113942	38008159	35591199	Suma Dev. C.	47327.33	42020.96	40573.28
139	139	139	Obs.	10	10	10

En cuanto al análisis del error utilizaremos el MAPE.¹⁰ La interpretación del estadístico esta en porcentajes, por lo tanto, el GARCH obtuvo en promedio un error del 0.32% para dentro de la muestra y un 0.29% para fuera de la muestra, mientras que la RNA obtuvo en promedio un error 1.92% para dentro de la muestra y un 1.82% para fuera de la muestra.

5. Conclusiones

Los resultados muestran que el modelo GARCH presentó mejor ajuste tanto fuera como dentro de la muestra, mientras que la RNA sobrevaló la serie del DJIA lo que arrojó errores más altos. Por lo que, se concluye que el método de optimización para la RNA en esta investigación, no

¹⁰ Para el cálculo se tiene que: $MAPE = \sum_{t=1}^n \left(\frac{e_t}{Y_t} \right) * 100$

es el más adecuado debido a que la búsqueda a la solución del problema sólo se basó en las topologías propuestas, quedando combinaciones que pudieron haber sido mejores.

Bibliografía

- Bo, Tian y Quin Zheng (2006). *Empirical Study of Financial Affairs Early Warning Model on Companies Based on Artificial Neural Network*, Springer, pp, 313-320.
- Brooks, C. (2002). *Introductory Econometrics for Finance*, Cambridge University press.
- Bollerslev, Tim (1986), “Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, vol. 31 (3), pp. 307-327.
- Castillo, P.A., J.G. Castellano, J.J. Merelo y A. Prieto (2001). “Diseño de Redes Neuronales Artificiales Mediante Algoritmos Evolutivos”, *Revista Iberoamericana de Inteligencia Artificial*, 14, pp. 2-32.
- Chakraborty, K., K. Mehrotra, C. Mohan, and S. Ranka (1992). *Forecasting the behavior of multivariate time series using neural networks*, Elsevier Science, 5, pp.961-970.
- Engle, Robert (1982), “Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation”, *Econometrica*, vol. 50 (4), pp. 987-1007.
- Freeman, J. A. (1993). *Simulating Neural Networks with Mathematica*, Addison-Wesley Professional.
- Hadavandi, E., A. Ghanbari, and S. Abbasian-Nagheh (2010). *Developing an Evolutionary Neural Network Model for Stock Index Forecasting*, Springer, pp.407-415.
- Karathanasopoulos, Andreas S.; Konstantinos A. Theofilatos; Panagiotis M. Leloudas y Spiridon D. Likothanassis (2010). “Modeling the Ase 20 Greek Index Using Artificial Neural Networks Combined with Genetic Algorithms, *Artificial Neural Networks*”, *Lecture note in Computer Science*, vol 6352.
- Malliaris, Mary y Linda Salchenberger (1996). *Neural Networks in Finance and Investing, Using Artificial Intelligence to Improve Real- World Performance*, En R. Trippi y E. Turban (Eds.), *Neural Networks for Predicting Options Volatility* (pp.613-622). E.U: Irwin Professional Publishing.
- Mehrotra, K., C. Mohan, and S. Ranka (2000). *Artificial Neural Networks*, 2nd ed, The MIT Press, Cambridge, Massachusetts.
- Parisi, A., F. Parisi y D. Diaz (2006). “Modelos de Algoritmos Genéticos y Redes Neuronales en la Predicción de Índices Bursátiles Asiáticos”, *Cuadernos de Economía*, 43, pp.251-284.
- Pérez, César (2006). *Econometría de las Series Temporales*, Pearson Prentice Hall, Madrid.
- Pérez, M. y Q. Martín (2003). *Aplicación de Redes Neuronales Artificiales a la Estadística*, Cuadernos de estadística, Editorial la Muralla, Madrid.
- Swanson, Norman y Halbert White (1997). “A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks”, *The Review of Economics and Statistics*, vol.79, no.4, pp540-550.
- Taylor, S. (1986), *Modelling Financial Time series*, John Wiley and Sons, New York.
- White, H. Del L. (1996). “*Neural Networks in Finance and Investing, Using Artificial Intelligence to Improve Real- World Performance*”. En R. Trippi y E. Turban (Eds.), *Economic Prediction Using Neural Networks: the case of IBM Daily Stock Prices* (pp.469-480). E.U: Irwin Professional Publishing.