

CONTRASTE METODOLÓGICO EN LA PREDICCIÓN DE LA VOLATILIDAD DEL ÍNDICE DE PRECIOS Y COTIZACIONES DE LA BOLSA MEXICANA DE VALORES: EWMA, GARCH Y REDES NEURONALES

(Recibido: 29 de junio de 2012 – Aceptado: 27 de julio de 2012)

45

Lilia Alejandra Flores Castillo*
Miguel Flores Ortega**

Resumen

En este trabajo se presenta el contraste metodológico en la predicción de la volatilidad del Índice de Precios y Cotizaciones de la bolsa mexicana de valores (IPyC), al utilizar los modelos de: media móvil con ponderación exponencial (EWMA), el modelo generalizado con heteroscedasticidad condicional autoregresivo (GARCH) y un modelo de redes neuronales artificiales (RNA) con estructura multicapa.

El análisis se presenta a partir de la capacidad de reproducción de las características de la serie histórica y la velocidad de ajuste ante condiciones cambiantes para incorporar la característica de heteroscedasticidad de la muestra histórica en cada modelo. Para el modelo EWMA se determinó el valor óptimo del coeficiente de decaimiento para establecer el peso de los registros históricos y minimizar el error del pronóstico, de forma similar se ajustó la estructura y los parámetros del modelo GARCH, a partir de la revisión de sus características. En el caso de la RNA se utiliza un perceptrón con tres capas y se seleccionó el número de neuronas de la capa interna que mejor identificó los patrones de comportamiento de la serie histórica, para el aprendizaje se utiliza el método tradicional de propagación hacia atrás en el modelo multicapa.

Palabras Clave: Pronóstico financiero, modelos de series de tiempo, Redes neuronales, Modelos de pronóstico EWMA, GARCH y RNA

Clasificación JEL: G17, C22, C45, C53

* Egresada de la Maestría en Ciencias Económicas, ESE-IPN.

** Profesor de la Sección de Estudios de Posgrado e Investigación. ESE-IPN.

Introducción

En el estudio de los mercados financieros el análisis de la volatilidad de los activos financieros es indispensable. Un referente obligado al estudiar el riesgo de múltiples activos es el trabajo de Harry Markowitz (1952), su tesis considera que la variación de los rendimientos presenta reversión a la media, por lo que asume que la varianza permanece constante en el periodo de estudio, con ésta consideración se simplifica la formulación matemática para determinar la volatilidad.

46 Cuando la varianza permanece constante el pronóstico corresponde al cálculo del valor histórico, pero conforme la varianza varía en el tiempo y los cambios son importantes para el pronóstico el tema se vuelve relevante; en un mundo global la dinámica de los mercados financieros hace que aumente la volatilidad y el pronóstico se dificulta, con esta motivación, los investigadores han buscado desarrollar modelos que permitan calcular y pronosticar la volatilidad con precisión, con esta idea se desarrolló este trabajo en donde se prueban tres métodos para determinar la volatilidad del IPyC y realizar su pronóstico.

La revisión de la literatura sobre el tema muestra que existen resultados empíricos de la evaluación de modelos de pronóstico de la volatilidad, el trabajo de Akgiray (1989), presenta el efecto de la variación de la varianza en el tiempo que corresponde a la heteroscedasticidad y analiza el efecto en el pronóstico, trabajos posteriores como el de Andersen, Bollerslev, y Lange (1999), concluyen que el modelo GARCH captura adecuadamente las dependencias dinámicas de la serie histórica que afectan a la varianza.

En el trabajo de McMillan, Speight y A P Gwilym (2000), se presenta la evaluación de los modelos GARCH y EWMA, los resultados indican que el desempeño de los modelos es similar, sin embargo para los autores el modelo EWMA es mejor dado que el valor de la raíz cuadrática media del error es menor, por esto, se considera un mejor modelo de pronóstico para los datos no ajustados, ésta afirmación se corrobora con los resultados del trabajo que presentaron Bluhm y Yu (2001), quienes concluyeron que ningún método es claramente superior a pesar de que el modelo EWMA muestra ser un modelo adecuado para el pronóstico de la volatilidad. En el trabajo de Shyan, C., Andy C., Chang, C., en conjunto con los trabajos de Chien W. (2010) y Chong (2005), se encontró que coinciden en afirmar que en un horizonte de 10 a 20 días el modelo EWMA proporciona el mejor pronóstico.

El trabajo de Swales y Young (1992) presenta que para los datos que utilizó, el pronóstico del rendimiento de las acciones con un modelo de múltiple discriminante el nivel de confianza del pronóstico corresponde al 74%, mientras que los resultados obtenidos a partir de una RNA llegaron a un 91% de nivel de confianza, en su trabajo aseveran que el modelo de RNA es significativamente superior. En el trabajo de Villamil y Delgado (2007), quién aplicó un modelo de redes neuronales para el pronóstico y la negociación del tipo de cambio EUR/USD, afirma que es posible obtener rentabilidades positivas de manera sostenida en el

mercado Forex con la estrategia adecuada. En otro campo el trabajo de Villada, Cadavier, y Molina (2008), presenta que el pronóstico del precio de la energía mediante el uso de una RNA tiene ventajas sobre los resultados obtenidos con un modelo GARCH. De forma similar el trabajo de Jhonson y Padilla (2005) presenta que las RNA son mejores para predecir que los modelos lineales tradicionales.

Desde el punto de vista de Fadlalla y Lin (2001), la aplicación de RNA en finanzas para el pronóstico de la varianza en contraste con los modelos econométricos y estadísticos, muestra una ventaja significativa en la práctica financiera cuando se utiliza para el pronósticos del precio de un bien, de una acción, bonos, la determinación del riesgo de fraude o bancarrota, y esta afirmación motiva la presente investigación. El uso de RNA es una alternativa que ha mostrado un buen desempeño en el análisis de los rendimientos de los activos financieros especialmente cuando se habla de distribuciones no lineales, su implementación puede contribuir a obtener resultados adecuados para la estimación de los parámetros necesarios para obtener portafolios con rendimientos óptimos.

El objetivo de la investigación es evaluar el pronóstico de la volatilidad del IPyC a partir de un análisis empírico de los resultados obtenidos con los modelos EWMA, GARCH y una RNA multicapa. Para la prueba se utilizó una muestra histórica compuesta de 5119 observaciones del rendimiento diario del IPyC, que comprende el periodo del 2 de Enero de 1992 al 18 de Mayo del 2012, el horizonte de pronóstico corresponde a $t+15$ del 21 de Mayo de 2012 al 8 de Junio del 2012, con esta muestra se establece qué modelo permite obtener el mejor pronóstico de la volatilidad del rendimiento del IPyC.

El artículo se organiza en cinco apartados, a los que precede la introducción al tema, el primer apartado se dedica al análisis del modelo EWMA, el segundo al modelo GARH y el tercero al modelo RNA multicapa con entrenamiento con propagación hacia atrás, el cuarto se dedica a presentar la evidencia empírica y el contraste metodológico en la predicción de la volatilidad del IPyC, finalmente en el apartado cinco se presentan las conclusiones de la investigación.

1. Modelo de media móvil con ponderación exponencial

Para dar respuesta a las necesidades de los analistas financieros J. P. Morgan (1996) propuso una metodología denominada RiskMetrics que fue adoptada a nivel mundial para predecir la varianza y covarianza de los rendimientos de los activos financieros, mediante el modelo EWMA (por su nombre en inglés Exponential Weighted Moving Average), el modelo pondera de manera diferente cada observación, de forma que asigna mayor peso a las observaciones recientes y menor peso a las observaciones más alejadas en el tiempo; de acuerdo con Botha, Van, y Styger (2001), éste enfoque tiene dos ventajas, en primer lugar, la volatilidad reacciona con mayor rapidez en condiciones de crisis en el mercado, porque utiliza los datos

más recientes a los que da mayor peso. En segundo lugar, después de un cambio brusco en los rendimientos, la volatilidad disminuye exponencialmente a medida que el peso de la observación reduce el efecto del cambio, la velocidad corresponde al factor λ^{T-1} .

La ecuación 1, representa la varianza de los rendimientos:

$$\sigma^2 = (1 - \lambda) \sum_{t=1}^T \lambda^{t-1} (r_t - r)^2 \quad (1)$$

Cuando la ecuación se escribe de forma recursiva es posible realizar el pronóstico de la volatilidad para $t+1$, como se presenta en la ecuación 5:

$$\sigma_{1,t+1}^2 = (1 - \lambda) r_{1,t}^2 + \lambda \sigma_{1,t}^2 \quad (2)$$

$$\sigma_{1,t+1}^2 = (1 - \lambda)(r_{1,t}^2 + \lambda r_{1,t-1}^2 + \lambda^2 r_{1,t-2}^2 + \dots) \quad (3)$$

$$\sigma_{1,t+1}^2 = (1 - \lambda)(r_{1,t}^2 + \lambda(1 - \lambda)r_{1,t-1}^2 + \lambda^2 r_{1,t-2}^2 + r_{1,t-3}^2) \quad (4)$$

$$\sigma_{1,t+1}^2 = \lambda \sigma_{1,t}^2 + (1 - \lambda)r_{1,t}^2 \quad (5)$$

Donde:

λ : es el factor de decaimiento, va a determinar el peso de las observaciones, toma valores entre 0 y 1

σ^2 : simboliza la varianza

r^2 : representa el rendimiento en el tiempo t , al cuadrado

En la ecuación 5, se establece que la varianza futura es igual al factor lambda multiplicado por la volatilidad del día anterior, más el cuadrado de la rentabilidad del día por uno menos lambda, es decir, si hoy la rentabilidad es alta, ésta induce a un incremento en la volatilidad estimada.

1.1 Factor de decaimiento óptimo

Con el propósito de realizar un adecuado pronóstico de la varianza es necesario determinar el valor del factor de decaimiento óptimo, uno de los métodos para estimarlo, es el criterio de la raíz del error cuadrado medio (RMSE) para el cálculo de la varianza $\sigma_{1,t}^2$ que está en función del factor λ que se expresa en la ecuación 6, que presentó RiskMetric.

$$RMSE_v = \sqrt{\frac{1}{t} \sum_{t=1}^T r_{1,t}^2 - \left(\frac{1}{t} \sum_{t=1}^T r_{1,t} \right)^2} \quad (6)$$

De forma empírica se determina a partir de una muestra de la series de tiempo el factor de decaimiento, mediante la búsqueda del menor RMSE para un conjunto de valores de λ , se determina el factor de decaimiento óptimo que es el que minimiza el error cuadrado del pronóstico de la varianza. Cuando hay varios activos financieros se determina un factor de decaimiento óptimo para todas las series a partir del siguiente procedimiento:

Se encuentra Π , la suma del producto de todo N que minimiza su RMSE y su τ_i :

$$\Pi = \sum_{i=1}^N \tau_i \tag{7}$$

Se define el error medio relativo:

$$\theta_i = \tau_i / \left(\sum_{i=1}^N \tau_i \right) \tag{8}$$

Se definen los pesos ϕ :

$$\phi_i = \theta_i^{-1} / \sum_{i=1}^N \theta_i^{-1} \text{ donde } \sum_{i=1}^N \theta_i^{-1} = 1 \tag{9}$$

El factor de decaimiento λ se define por:

$$\tilde{\lambda} = \sum_{i=1}^N \theta_i \hat{\lambda}_i \tag{10}$$

que se expresa en la ecuación 6, que presentó RiskMetric. Se denota como factor de decaimiento óptimo a $\hat{\lambda}_i$ y N ($i = 1, 2, \dots, N$) que corresponde al número de series de tiempo. Por otro lado τ_i denota la i th soluciones del RMSE asociado con $\hat{\lambda}_i$, ie., τ_i es el mínimo RMSE para la i th series de tiempo.

El factor de decaimiento óptimo que aplica RiskMetrics, es un promedio pesado de los factores de decaimiento óptimos individuales donde cada peso es la media del pronóstico individual.

2. Modelo generalizado de heteroscedasticidad condicional autoregresiva

El modelo generalizado de heteroscedasticidad condicional autoregresiva (GARCH) propuesto por Bollerslev(1986), presenta un enfoque que asume que la varianza de los rendimientos sigue un proceso predecible y depende del último evento, así como también de la varianza condicional previa, establece que la varianza condicional h_t , está dada por la expresión 11.

$$h_t = \alpha_0 + \sum_{i=1}^q \alpha_i r_{t-1}^2 + \sum_{i=1}^p \beta_i h_{t-1} \tag{11}$$

Donde

$$\begin{aligned} \alpha_0 &> 0, & \alpha_i &\geq 0, & i &= 1, \dots, q, \\ \beta_1 &\geq 0, & & & i &= 1, \dots, p. \\ \alpha_1 + \beta_1 &< 1 \end{aligned}$$

El modelo más simple de estos tipos de procesos es el GARCH(1,1):

$$\sigma_n^2 = \gamma V + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \quad (12)$$

El γ es el peso asignado para V , α es el peso asignado para u_{n-1}^2 , y β es el peso asignado para σ_{n-1}^2 , para que este modelo sea estacionario la suma de los parámetros debe ser menor a la unidad o igual 1: $\gamma + \alpha + \beta \leq 1$

50

Si se supone que $\omega = \gamma V$, el modelo GARCH (1,1) se expresa de acuerdo a la expresión 13.

$$\sigma_n^2 = \omega + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \quad (13)$$

Se utiliza la ecuación 10, para la estimación de los parámetros, el parámetro γ se calcula como $1 - \alpha - \beta$, la varianza promedio condicional se calcula como $\bar{\omega} / \gamma$, o se deriva de la expresión 14.

$$V = \frac{\alpha_0}{1 - \alpha_1 - \beta} \quad (14)$$

El pronóstico de la varianza mediante la metodología GARCH (1,1) se expresa en la ecuación 15:

$$E[\sigma_{n+k}^2] = V + (\alpha + \beta)^k (\sigma_n^2 - V) \quad (15)$$

La ventaja del enfoque GARCH es que permite reaccionar más rápido al conferir una ponderación mayor a la información reciente lo que permite tratar series de tiempo que presentan heteroscedasticidad y proporcionar estimaciones de la volatilidad de manera eficiente sobre todo en horizontes de corto tiempo. Una limitación importante del modelo es que la varianza condicional responde de la misma manera a los residuos positivos que a los negativos, característica que contradice el comportamiento observado en las series temporales de datos financieros.

3. Redes Neuronales Artificiales

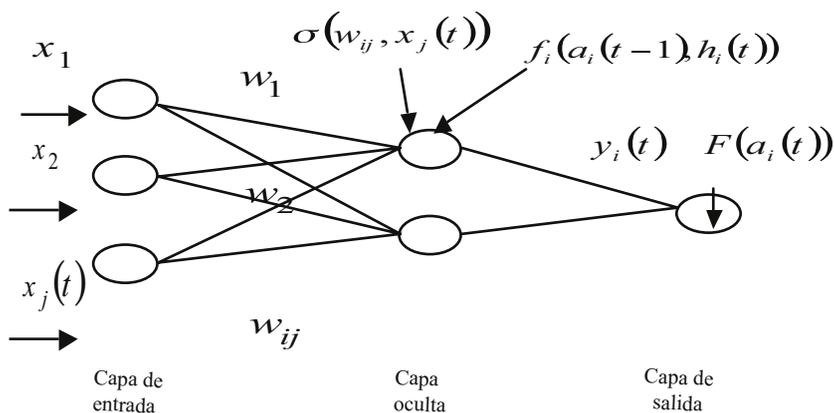
Las redes neuronales artificiales (RNA), son un modelo de procesamiento de información, que pueden expresarse mediante métodos matemáticos que pretenden emular el funciona-

miento del cerebro humano, son modelos estadísticos no lineales, no paramétricos utilizados principalmente para la clasificación, predicción de datos y variables.

3.1 Estructura de una red neuronal

Dentro de ésta investigación se considera la estructura de una red perceptrón multicapa, generalmente se compone de tres capas de neuronas con diferentes funciones; la primera capa es la capa de entrada, mediante la cual se introducen los datos que van alimentar la red. La segunda capa es denominada capa oculta la cual tiene asociada una regla de propagación y una función de activación. En la tercera capa mediante la función de salida se obtiene la salida de la red.

Figura 1
Elementos básicos de una RNA



Fuente: Elaboración propia.

De acuerdo con Martin y Sanz (2007), los elementos básicos de una RNA son los siguientes:

- Conjunto de entradas, $x_j(t)$.
- Pesos sinápticos de la neurona i , w_{ij} que representan la intensidad de interacción entre cada neurona pre sináptica j y la neurona pos sináptica i .
- Regla de propagación $\sigma(w_{ij}, x_j(t))$, que proporciona el valor del potencial pos sináptico $h_i(t) = \sigma(w_{ij}, x_j(t))$ de la neurona i en función de sus pesos y entradas.
- Función de activación de la neurona i en función de su estado anterior $a_i(t-1)$ y de su potencial pos sináptico actual, representa simultáneamente la salida de la neurona y su estado de activación.

- Función de salida $f_i(a_i(t))$, que proporciona la salida actual de la neurona i en su estado de activación.

La operación de una neurona i puede expresarse como:

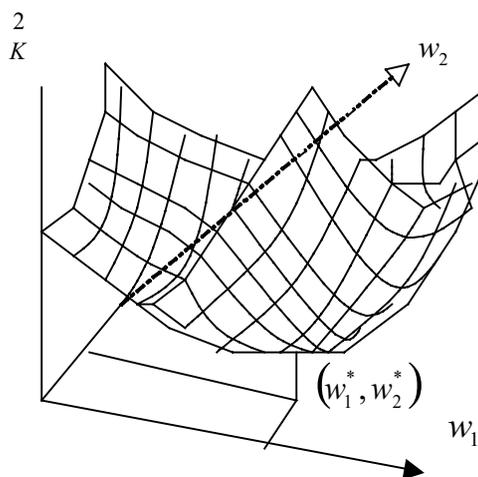
$$y_i(t) = F_i(f_i[a_i(t-1), \sigma(w_{ij}, x_j(t))]) \quad (16)$$

3.2 Fase de entrenamiento

Durante la fase de entrenamiento se modifican los pesos de las conexiones, se sigue la regla de aprendizaje que trata de optimizar su respuesta mediante la minimización del error de ajuste. El modo de aprendizaje más sencillo consiste en la presentación de un conjunto patrones de entrada y salida deseados para cada patrón de entrada. El problema para caracterizar al modelo de la red neuronal es determinar el vector de pesos que asocia el vector de entrada con el valor de salida para obtener con precisión el valor deseado, el proceso de aprendizaje corresponde a minimizar las diferencias entre las salidas deseadas y la salida real para todos los vectores de entrada para determinar los pesos que minimizan el conjunto de los errores, de forma gráfica se representa por la superficie paraboloides que se muestra en la figura 2.

52

Figura 2. Pesos que minimizan el error



Fuente: Feeman y skapura (2003).

En el desarrollo de la investigación se utilizó la arquitectura del perceptrón multicapa y se entrenó con el algoritmo de propagación hacia atrás en donde los valores de entrada se propagan a través de la neurona hasta llegar a la salida y los compara con los valores esperados o valores objetivo, cuando se cumple con las restricciones propuestas se asume que

la red neuronal está entrenada, de lo contrario, la diferencia del valor de salida y el objetivo se propaga hacia atrás para ajustar los pesos de las neurona, el proceso se repite hasta lograr el ajuste. La importancia de este proceso consiste en que a medida que se entrena la red, las neuronas de las capas intermedias se organizan de tal forma que aprenden a reconocer distintas características del conjunto de las entradas.

De acuerdo con Hagan, Demuth, y Beale (2002), en la fase de entrenamiento además de propagar el error hacia atrás se implementa el algoritmo Levenberg Marquardt el cuál optimiza el proceso de entrenamiento.

$$x_{k+1} = x_k - [J^T(x_k) J(x_k) + \mu_k I]^{-1} J^T(x_k) v(x_k) \tag{17}$$

$$\Delta x_k = - [J^T(x_k) J(x_k) + \mu_k I]^{-1} J^T(x_k) v(x_k) \tag{18}$$

En donde: μ determina la tendencia de ajuste.

Se presentan las entradas a la red y se calculan las salidas de la red, el error se expresa por $e_q = t_q - a_q^M$, se calcula la suma del cuadrado de los errores sobre todas las entradas, $F(x)$.

$$F(x) = \sum_{q=1}^Q (t_q - a_q)(t_q - a_q) \tag{19}$$

$$F(x) = \sum_{q=1}^Q e_q e_q = \sum_q \sum_{j=1}^{S^M} (e_{j,q})^2 = \sum_{i=1}^N (v_i)^2 \tag{20}$$

Donde: $e_{j,q}$ es el elemento $j_{i,mo}$ del error para $q_{mo} q_l$ par de entrada-objetivo.

Se calcula la matriz Jacobiana formada por las derivadas parciales de primer orden del error respecto a los pesos.

$$\begin{matrix} \frac{\partial e_{1,1}}{\partial w_{1,1}^1} & \frac{\partial e_{1,1}}{\partial w_{1,2}^1} & \dots & \frac{\partial e_{1,1}}{\partial w_{s^1,2}^1} & \frac{\partial e_{1,1}}{\partial b_1^1} & \dots \\ \frac{\partial e_{2,1}}{\partial w_{1,1}^1} & \frac{\partial e_{2,1}}{\partial w_{1,2}^1} & \dots & \frac{\partial e_{2,1}}{\partial w_{s^1,2}^1} & \frac{\partial e_{2,1}}{\partial b_1^1} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial e_{S^M,1}}{\partial w_{1,1}^1} & \frac{\partial e_{S^M,1}}{\partial w_{1,2}^1} & \dots & \frac{\partial e_{S^M,1}}{\partial w_{s^1,R}^1} & \frac{\partial e_{S^M,1}}{\partial b_1^1} & \dots \\ \frac{\partial e_{1,2}}{\partial w_{1,1}^1} & \frac{\partial e_{1,2}}{\partial w_{1,2}^1} & \dots & \frac{\partial e_{1,2}}{\partial w_{s^1,R}^1} & \frac{\partial e_{1,2}}{\partial b_1^1} & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \end{matrix} \tag{21}$$

Con esto se estima el factor de sensibilidad de la red neuronal

$$\tilde{S}_q^M = -F^M(n_q^M) \quad (22)$$

$$\tilde{S}_q^m = -F^m(n_q^m)(W^{m+1})^T \tilde{S}_q^{m+1} \quad (23)$$

Donde $F^m(n_j^m)$

$$F^m(n^m) = \begin{bmatrix} F^m(n_1^m) & 0 & \dots & .0 \\ 0 & F^m(n_2^m) & \dots & .0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & F^m(n_{s^m}^m) \end{bmatrix} \quad (24)$$

Se aumentan las matrices individuales en las sensibilidades Marquardt

54

$$\tilde{S}_q^m = \tilde{S}_1^m | \tilde{S}_2^m | \dots | \tilde{S}_Q^m \quad (25)$$

Después se calcula los elementos de la matriz Jacobiana con:

$$|J|_{h,l} = \frac{\partial v_h}{\partial x_l} = \frac{\partial e_{k,q}}{\partial w_{i,j}^m} = \frac{\partial e_{k,j}}{\partial n_{i,q}^m} = \tilde{S}_{i,h}^m \frac{\partial n_{i,q}^m}{\partial w_{i,j}^m} \tilde{S}_{j,q}^m a_{j,q}^{m-1} \quad (26)$$

Al solucionar la siguiente ecuación se obtiene Δx_k

$$x_{k+1} = x_k - [J^T(x_k) J(x_k) + \mu_k I]^{-1} J^T(x_k) v(x_k) \quad (27)$$

Se calcula la suma del error al cuadrado utilizando $x_k + \Delta x_k$; si el valor de la suma es menor que la calculada anteriormente, entonces dividimos μ por ϑ , se establece $x_{k+1} = x_k + \Delta x_k$ y se vuelven a calcular las salidas de la red. El algoritmo converge cuando la norma del gradiente $\Delta F(x) = 2J^T(x) v(x)$ es menor que un determinado valor, o cuando la suma de los cuadrados tiene que ser reducido a un valor objetivo.

3.3 Evaluación y ventajas de la red

La evaluación del funcionamiento de la red neural se realiza una vez que ha culminado la fase de entrenamiento, es usual disponer de un conjunto de datos distintos a los usados para el entrenamiento para los cuales se conoce la respuesta correcta y se usan como examen para evaluar el aprendizaje de la red, se observa si responde adecuadamente al conjunto de entradas y salidas conocido.

De acuerdo con Ponce (2010), las ventajas de las RNA con respecto a otros modelos es que:

- Las RNA sintetizan algoritmos a través de un proceso de aprendizaje.
- Para utilizar la tecnología neuronal no es necesario conocer los detalles matemáticos. Solo se requiere estar familiarizado con los datos del trabajo.

- La solución de los problemas no lineales es uno de los puntos fuertes de las RNA.
- Las RNA son robustas, pero pueden fallar elementos en el procesamiento y la red continúa trabajando, esto es contrario a lo que sucede en programación tradicional.

Las desventajas de las redes neuronales son:

- Las RNA se deben entrenar para cada problema. Además, es necesario realizar múltiples pruebas para determinar la arquitectura adecuada. El entrenamiento es largo y puede consumir varias horas de la computadora.
- Debido a que las redes se entrenan en lugar de programarlas, estas necesitan muchos datos.
- Las RNA representan un aspecto complejo para un observador externo que desee realizar cambios. Para añadir nuevo conocimiento es necesario cambiar las iteraciones entre muchas unidades para que su efecto unificado sintetice este conocimiento.

4. Contraste metodológico en la predicción de la volatilidad del IPyC, evidencia empírica

Al considerar los aspectos metodológicos planteados en esta investigación, la evidencia empírica se observa al realizar el pronóstico de la varianza del IPyC con cada uno de los modelos expuestos, EWMA, GARCH y RNA.

4.1 Descripción de los datos

El análisis y pronóstico de la varianza para cada una de los modelos, parte de una muestra de 5119 observaciones de los rendimientos diarios históricos del IPyC, se considera el precio de cierre desde el 2 de Enero de 1992 hasta el 18 de Mayo del 2012 y los rendimientos se obtienen de forma logarítmica:

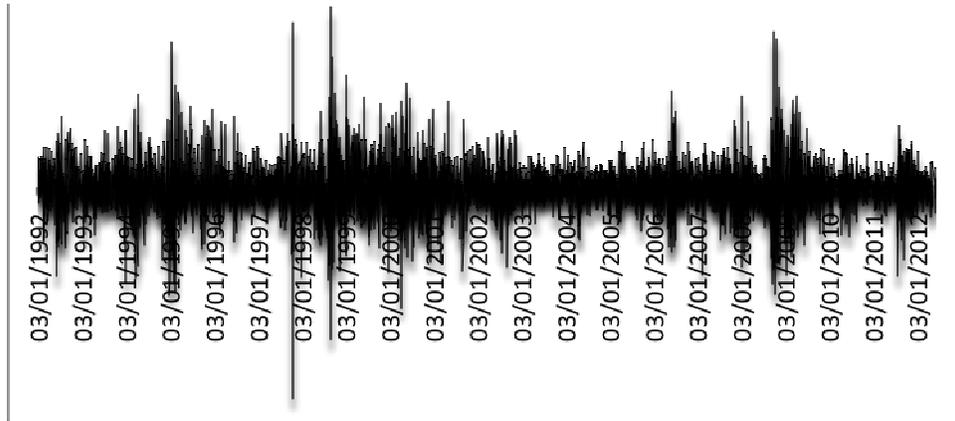
$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (28)$$

El horizonte para realizar el pronóstico de la varianza es de $t+15$ corresponde al 21 de Mayo de 2012 al 8 de Junio del 2012.

4.2 Pronóstico de la varianza mediante la metodología EWMA

Con el objetivo de obtener un pronóstico adecuado de la varianza se calculó el factor de decaimiento óptimo mediante el criterio de optimización de la raíz del error cuadrado medio; se establecieron 21 muestras con 200 observaciones, cada una lleva asociada un factor de decaimiento λ que minimiza el error cuadrado del pronóstico de la varianza.

Gráfica 1
Rendimientos logarítmicos del IPyC



Fuente: Elaboración propia con datos del IPyC

Cuadro 1
Cálculo del factor de decaimiento óptimo

No. de muestra	Factor óptimo	RMSE	Θ	θ	ϕ	$\phi\lambda$
1	0,95	0,0008	0,053406	18,7246	0,0214	0,0203
2	0,99	0,0012	0,080214	12,4666	0,0142	0,0141
3	0,92	0,0012	0,082215	12,1632	0,0139	0,0128
4	0,97	0,0006	0,042271	23,657	0,027	0,0262
5	0,97	0,0003	0,016795	59,5403	0,068	0,066
6	0,9	0,0017	0,112564	8,8839	0,0102	0,0091
7	0,9	0,0016	0,108816	9,1898	0,0105	0,0095
8	0,96	0,0005	0,035986	27,7887	0,0318	0,0305
9	0,95	0,0009	0,057089	17,5165	0,02	0,019
10	0,98	0,0008	0,053218	18,7906	0,0215	0,021
11	0,95	0,0004	0,024931	40,111	0,0458	0,0435
12	0,97	0,0003	0,019043	52,5127	0,06	0,0582
13	0,97	0,0001	0,007973	125,4291	0,1433	0,139
14	0,9	0,0002	0,015937	62,7481	0,0717	0,0645
15	0,9	0,0002	0,015734	63,5564	0,0726	0,0654
16	0,9	0,0005	0,034272	29,1782	0,0333	0,03
17	0,98	0,0003	0,021672	46,1433	0,0527	0,0517
18	0,9	0,0021	0,137191	7,2891	0,0083	0,0075
19	0,99	0,0009	0,05891	16,9749	0,0194	0,0192
20	0,9	0,0002	0,015421	64,848	0,0741	0,0667
21	0,97	0,0001	0,006343	157,6425	0,1801	0,1747
		$\Pi = 0,015156$	N	Factor de decaimiento óptimo 0,949		
			$\sum_{i=1}^N \theta_i^{-1}$	$= 875,1547$		

Fuente: Elaboración propia con datos del IPyC, en base a la metodología Riskmetrics.

Se considera que el factor de decaimiento óptimo es adecuado para estimar el pronóstico de la varianza para el periodo referido. En el cuadro 2, se presenta el pronóstico diario de la volatilidad sobre los rendimientos de la acción IPyC, en base a la ecuación 29:

$$\sigma_{1,t+1|t}^2 = \lambda \sigma_{1,t+1|t}^2 + (1 - \lambda) r_{1,t}^2 \quad (29)$$

Cuadro 2
Cálculo del pronóstico de la varianza mediante el modelo EWMA

<i>Fecha</i>	<i>Rendimientos IPyC</i>	<i>Varianza con modelo EWMA</i>
21/05/2012	0.01712905	7.57E-05
22/05/2012	-0.0008326	8.88E-05
23/05/2012	-0.00157029	8.35E-05
24/05/2012	0.00348994	7.86E-05
25/05/2012	-0.00178466	7.47E-05
28/05/2012	0.0041675	7.04E-05
29/05/2012	0.01277706	6.72E-05
30/05/2012	-0.00147459	7.30E-05
31/05/2012	-0.00520703	6.87E-05
01/06/2012	-0.01840241	6.62E-05
04/06/2012	-0.00330244	8.26E-05
05/06/2012	0.00079785	7.83E-05
06/06/2012	0.00498709	7.36E-05
07/06/2012	-0.00073669	7.07E-05
08/06/2012	0.00205361	6.65E-05

Fuente: Elaboración propia con datos del IPyC

4.3 Pronóstico de la varianza mediante la metodología GARCH (1,1).

El primer paso que se realiza para obtener el pronóstico en cualquier modelo que involucra series temporales es verificar si la serie es estacionaria. Un criterio formal, para detectar la estacionariedad, son los contrastes de raíces unitarias (Augmented Dickey-Fuller, Dickey Fuller GLS, Phillips Perron, Kwiatkowski-Phillips-Schmidt-Shin, Elliot-Rothenberg-Stock Point-Optimal, Ng-Perron), para considerar la hipótesis de estacionariedad se utiliza el contraste de la prueba de raíces unitarias y la prueba de Dickey Fuller aumentada, se obtiene un p-valor menor a 0.05, se acepta la hipótesis de estacionariedad como se muestra en el cuadro 3.

Cuadro 3
Prueba de estacionariedad Dickey Fuller

<i>Prueba</i>	<i>t-Statistic</i>	<i>Prob. *</i>
<i>Augmented Dickey-Fuller test statistic</i>	-64.66353	0.0001
<i>Test critical values:</i>		
1% level	-3.431442	
5% level	-2.861907	
10% level	-2.567008	

Fuente: Elaboración propia con datos del IPyC en Eviews

Con el objetivo de obtener el pronóstico de la varianza se estima el modelo GARCH (1,1), obteniendo los parámetros: ω , α y β .

Cuadro 4**Estimación de los parámetros mediante el modelo GARCH(1,1)**

Dependent Variable: IPyC				
Method: ML – ARCH				
Sample: 1 5136				
Included observations: 5136				
Convergence achieved after 14 iterations				
Variance backcast: ON				
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)				
	<i>Coefficient</i>	<i>Std. Error</i>	<i>z-Statistic</i>	<i>Prob.</i>
C	0.001104	0.000178	6.203479	0.0000
	<i>Variance Equation</i>			
C	4.04E-06	5.08E-07	7.957067	0.0000
RESID(-1)^2	0.109585	0.005043	21.73040	0.0000
GARCH(-1)	0.878796	0.005523	159.1017	0.0000

58

La varianza en n mediante el modelo GARCH (1,1), se obtiene de la siguiente manera:

$$\sigma_n^2 = \lambda V + \alpha u_{n-1}^2 + \beta \sigma_{n-1}^2 \quad (30)$$

donde:

$$\omega = \gamma V$$

$$\gamma = 1 - \alpha - \beta$$

$$V = \frac{\omega}{\gamma}$$

Especificando la ecuación de la varianza a partir del cuadro (4) y la ecuación (30), tenemos:

$$\sigma_n^2 = 4.04E-6 + -0.109585u_{n-1}^2 + 0.78796\sigma_{n-1}^2 \quad (31)$$

Se realiza el pronóstico de la varianza:

$$E[\sigma_{n+k}^2] = V + (\alpha + \beta)^k (\sigma_n^2 - V) \quad (32)$$

Cuadro 5**Cálculo del pronóstico de la varianza mediante el modelo GARCH (1,1)**

<i>Fecha</i>	<i>Rendimientos IPyC</i>	<i>Varianza con modelo GARCH (1,1)</i>
21/05/2012	0.01712905	0.00012886
22/05/2012	-0.0008326	0.00014542
23/05/2012	-0.00157029	0.00013225
24/05/2012	0.00348994	0.00012104
25/05/2012	-0.00178466	0.00011103
28/05/2012	0.0041675	0.00010253
29/05/2012	0.01277706	0.00009517
30/05/2012	-0.00147459	0.00010261
31/05/2012	-0.00520703	0.00009494
01/06/2012	-0.01840241	0.00009184
04/06/2012	-0.00330244	0.00012644
05/06/2012	0.00079785	0.00011728
06/06/2012	0.00498709	0.00010712
07/06/2012	-0.00073669	0.00009983
08/06/2012	0.00205361	0.00009214

Fuente: Elaboración propia con datos del IPyC.

4.4 Pronóstico de la varianza mediante RNA

El pronóstico de la varianza con RNA se realizó con el software Matlab y se utiliza el Neural Network Toolbox . El objetivo es predecir los valores futuros de la varianza del IPyC de forma autoregresiva no lineal, se predicen los valores de $\sigma(t)$ dado n valores pasados $\sigma(t)$:

$$\sigma(t) = f(\sigma(t-1), \sigma(t-2), \sigma(t-2), \dots, y(t-n)) \tag{33}$$

targetSeries = tonndata(varianza,false,false)

Los datos que representan la entrada de la red corresponden a los rendimientos del IPyC de donde se estima la varianza, los datos se dividen aleatoriamente de acuerdo a la proporción 70% se utilizan en la fase de entrenamiento, 15% validación y medir la generalización del comportamiento de la red neuronal y 15% para efectos de prueba de la determinación de la red neuronal entrenada.

```
net.divideFcn = 'dividerand'; % Divide data randomly
net.divideMode = 'time'; % Divide up every value
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
```

Se seleccionó la arquitectura de la red con una capa de entrada, una capa oculta y una capa de salida, la red se entrenó y se seleccionó el número de neuronas de la capa oculta que minimizó el error medio cuadrático.

Cuadro 6

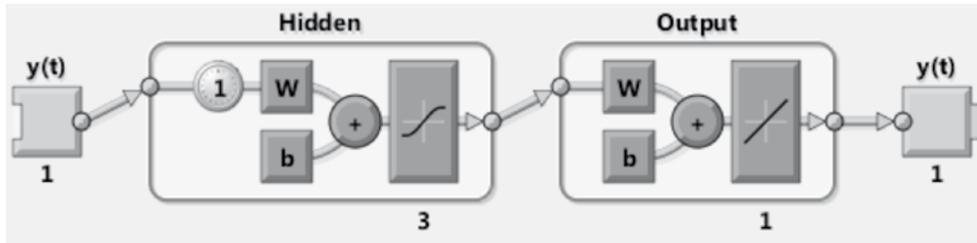
Selección del número de neuronas que conforman la capa oculta

<i>Nº de capas ocultas</i>	<i>Fase de prueba</i>
	<i>EMC</i>
1-1-1	3.94E-06
1-2-1	3.73E-06
1-3-1	2.76E-06
1-4-1	3.40E-06
1-5-1	3.54E-06
1-6-1	2.81E-06
1-7-1	4.75E-06
1-8-1	4.13E-06
1-9-1	3.89E-06
1-10-1	3.82E-06
1-11-1	3.52E-06
1-12-1	3.68E-06
1-13-1	5.09E-06
1-14-1	2.93E-06
1-15-1	3.29E-06

Fuente: Elaboración propia con datos del IPyC.

De acuerdo al cuadro 6, se puede observar que es la arquitectura que minimiza el error medio cuadrático medio es la estructura que está conformada por una capa de entrada una capa oculta que contiene 3 neuronas y una de salida con una neurona, con una función de transferencia sigmoide en la capa oculta y una función de transferencia lineal en la capa de salida.

Figura 3
Estructura de la red



Fuente: Elaboración propia con datos del IPyC, utilizando el software Matlab

```
% Choose a Training Function
60 % For a list of all training functions type: help ntrain
net.trainFcn = 'trainlm' % Levenberg-Marquardt
```

Una vez entrenada la red se realiza el pronóstico:

```
netc = closeloop(net);
[xc,xic,aic,tc] = preparets(netc, {}, {}, targetSeries);
yc = netc(xc,xic,aic);
perfc = perform(net,tc,yc);
nets = removedelay(net);
[xs,xis,ais,ts] = preparets(nets, {}, {}, targetSeries);
ys = nets(xs,xis,ais);
closedLoopPerformance = perform(net,tc,yc)
```

Cuadro 7
Pronóstico de la varianza mediante una RNA

Fecha	Rendimientos IPyC	Varianza con RNA
21/05/2012	0.01712905	0.00105107
22/05/2012	-0.0008326	0.00104279
23/05/2012	-0.00157029	0.00103554
24/05/2012	0.00348994	0.00102983
25/05/2012	-0.00178466	0.0010241
28/05/2012	0.0041675	0.00102009
29/05/2012	0.01277706	0.00102738
30/05/2012	-0.00147459	0.00102186
31/05/2012	-0.00520703	0.00101884
01/06/2012	-0.01840241	0.00103912
04/06/2012	-0.00330244	0.0010329
05/06/2012	0.00079785	0.00102664
06/06/2012	0.00498709	0.00102289
07/06/2012	-0.00073669	0.00101777
08/06/2012	0.00205361	0.00101352

Fuente: Elaboración propia con datos del IPyC.

4.5 Evaluación de la capacidad predictiva

La capacidad de pronóstico de cada uno de los modelos se llevó a cabo a partir del error cuadrático medio y del coeficiente de desigualdad de Theil.

$$ECM = \frac{1}{N} \sum_{n=1}^N (f_n - O_n)^2 \tag{34}$$

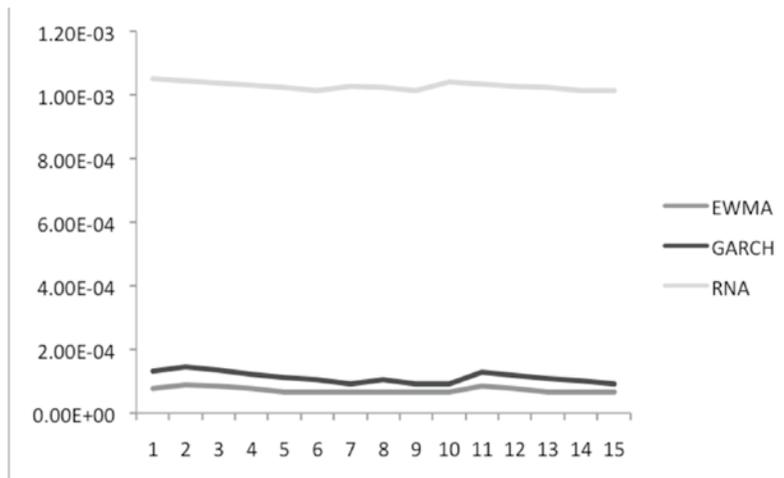
$$CDT = \frac{\sqrt{\frac{\sum_{i=1}^n (\hat{Y}_i - Y_i)^2}{n}}}{\sqrt{\frac{\sum_{i=1}^n \hat{Y}_i^2}{n} + \frac{\sum_{i=1}^n Y_i^2}{n}}} \tag{35}$$

El coeficiente de Theil varía entre el cero y el uno, indicando una mejor capacidad predictiva del modelo cuando más se acerca a cero.

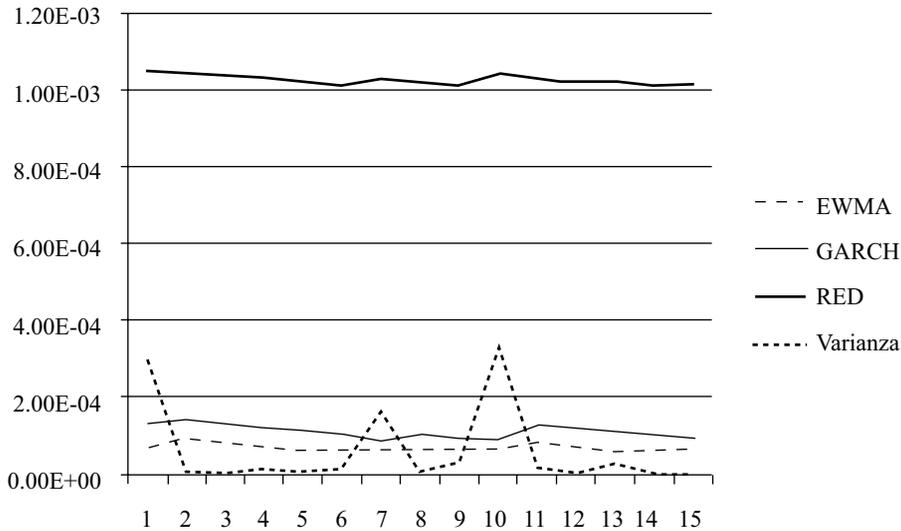
Cuadro 8
Error cuadrático medio y coeficiente de Theil

	<i>ECM</i>	<i>CDT</i>
EWMA	0.00011152	0.672556026
GARCH	0.00012306	1.265104058
RNA	0.00097358	5.417609788

Gráfica 2
Pronóstico de la varianza del IPyC



Fuente: Elaboración propia con datos del IPC.

Gráfica 3**Comprobación del pronóstico de la varianza del IPyC con la varianza observada**

Fuente: Elaboración propia con datos del IPyC.

Los valores que representan el error cuadrático medio y coeficiente de desigualdad de Theil para los 15 valores pronosticados se muestran en el cuadro 8, para esta prueba empírica el modelo que minimiza el error para ambos indicadores es el modelo EWMA, mientras que la evaluación del rendimiento de pronóstico mediante redes neuronales muestra que la red logra pronosticar la tendencia de la serie aunque se observa que el error es mayor. En la gráfica 3, se observa la dinámica del pronóstico con el modelo EWMA, es semejante a la producida por el modelo GARCH(1,1).

5. Conclusiones

En esta investigación se realizó el contraste metodológico en la predicción del pronóstico de la volatilidad a partir los modelos EWMA, GARCH y RNA, la revisión teórica muestra que todos tienen ventajas y desventajas y su utilidad depende de la aplicación que se les quiera dar y el esfuerzo que se quiera realizar al obtener el pronóstico, desde el punto de vista empírico se encuentra relevante el tamaño de la muestra y la selección del periodo de estudio para la determinación de los parámetros de los modelos.

En el caso del pronóstico de la volatilidad mediante la metodología EWMA, se encontró que el factor de decaimiento óptimo mediante el criterio de optimización de la raíz del error cuadrado medio se ubicó en un valor de 0.949, el resultado es semejante al parámetro recomendado por RiskMetrics, resultado que establece el comportamiento de la volatilidad en el

mercado financiero mexicano es semejante al comportamiento de mercados financieros de países desarrollados.

Para la determinación del modelo GARCH se efectuó la prueba Dickey Fuller y se encontró un p-valor menor a 0.05, por lo que se aceptó la hipótesis de estacionariedad para el modelo GARCH (1,1), que se utilizó para el pronóstico de la volatilidad.

La estructura de redes neuronales perceptrón multicapa confirma ser una herramienta robusta que cuenta con un mecanismo de entrenamiento de forma tal que minimizará los errores de ajuste antes y después de observar desviaciones fuertes, una de las ventajas de haber utilizado esta metodología fue que en todos los análisis se logró una convergencia adecuada y una reducción sistemática de los errores de ajuste durante los procesos de entrenamiento.

En la investigación documental se encontró una importante cantidad de métodos para entrenar la red multicapa en donde la eficiencia consiste en la reducción del tiempo de entrenamiento y dado que no es un problema crítico para esta investigación se optó por utilizar herramientas convencionales que facilitan el estudio.

Para el modelo de red neuronal multicapa se realizaron múltiples pruebas para determinar el número de neuronas en la capa interna y el valor que se seleccionó fue de 3 porque muestra el menor error de ajuste, La red se configuró con una capa de entrada, una capa oculta que contiene 3 neuronas y una de salida con una neurona, el procedimiento de entrenamiento que se utilizó fue el algoritmo de propagación hacia atrás de Levenberg-Marquardt, en el estudio se encontró que la convergencia es muy rápida y adecuada.

Los resultados empíricos de la prueba para realizar el contraste de la capacidad predictiva de cada uno de los modelos permite analizar los errores definidos por el ECM y el CDT, en el modelo EWMA se encuentra el valor más pequeño del error, le sigue el modelo GARCH y en último lugar el modelo de RNA seleccionado, en principio los resultados confirman lo encontrado en la literatura sobre el poder predictivo del modelo EWMA que en el corto plazo cuando no hay cambios importantes previos. Se encuentra un tema a desarrollar para diferenciar el comportamiento del modelo GARCH y la RNA cuando hay grandes choques y crece el horizonte de predicción.

La aportación de la investigación se centra en el establecimiento de la evidencia empírica de la capacidad de ajuste y predicción de los modelos seleccionados a partir de la evaluación de los resultados con información real del IPyC en el periodo de estudio. La investigación sobre el modelado de la RNA es incipiente y existen múltiples alternativas, el avance que se presenta muestra la bondad de utilización de nuevas herramientas para pronosticar la volatilidad en condiciones de heteroscedasticidad.

Una conclusión adicional dentro del desarrollo del pronóstico de la varianza mediante cada una de las metodologías es considerar la importancia de la longitud de la memoria de la serie de tiempo con relación a la calidad y el nivel de confianza de la estimación de la volatilidad, a mayor nivel de confianza mayor debe ser el tamaño de la muestra histórica. Además de forma

empírica se corroboró que la metodología que sugiere RiskMetrics es adecuada en el entorno de la Bolsa Mexicana de Valores aunque se presentan desviaciones cuando ocurren cambios bruscos en el rendimiento de las variables, de ahí la importancia de determinar el coeficiente óptimo de decaimiento para representar de forma adecuada los efectos de persistencia.

Bibliografía

- 64 Akgiray, Vedat. (1989). “Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts”, *The Journal of Business*, Vol. 62, Num.1, pp. 56–79.
- Andersen, Torben G., Bollerslev, Tim., y Lange, Steve. (1999), “Forecasting Financial Market Volatility: Sample Frequency vis-à-vis Forecast Horizon”, *Journal Empirical Finance* 6:5, pp. 457–77.
- Bluhm, H., y Yu, Jun. (2001). “Forecasting Volatility: Evidence from the German Stock Market”. Work paper, U. Auckland.
- Bollerslev Tim. (1986). “Generalized Autoregressive Conditional Heteroscedasticity”. *Journal of Econometrics*, Vol.31, pp. 307-327.
- Botha, Marius, Van, Vuuren G., y Styger Pau. (2001). “The effect of dynamic exponential decay factor on volatility and VaR”. *Journal of Applied Sciences*, Vol.1, pp.24-32.
- Chong James. (2005). “The forecasting abilities of implied and econometric variance-covariance models across financial measures”, *Journal of economics and business*, Vol. 57, pp. 463-490.
- Fadlalla, Adam y Lin, Chien Hua. (2001). “An Analysis of the Applications of Neural Networks in Finance”. *Interfaces*, Vol. 31, Núm. 4, Jul. – Aug, 2001, pp. 112-122.
- Freeman, James A. y Skapura, David M. (1993). *Redes Neuronales, algoritmos, aplicaciones y técnicas de programación*. Addison-wesley/ Diaz de Santos.
- Hagan, Martin, T., Demuth, Howard B., y Beale, Mark H. (2002). *Neural network design*. USA, PWS Publishing Company.
- Jhonson, Christian A., y Padilla, Miguel A. (2005), “Regularidades no lineales en índices accionarios. Una aproximación con redes neuronales”, *El Trimestre Económico*, México, Fondo de cultura económica, octubre-diciembre, Vol. LXXII (4), Num. 288, pp. 765-821.
- J.P Morgan. (1996). *RiskMetrics*. Technical Document, 4a ed, New York.
- Markowitz Harry M. (1952), “Portfolio Selection”, *The journal of finance*, Vol II, Num.1, pp. 77-91.
- Martin, del Rio Bonifacio y Sanz, Molina Alfredo. (2007). *Redes Neuronales y Sistemas Borrosos*. México, 3ª ed, Alfaomega.
- McMillan, David, Speight, Alan H. y A. P. Gwilym, Owain. (2000). “Forecasting U.K. Stock Market Volatility”, *Applied Financial Economics*, Vol.10, pp. 435–48.
- Ponce Cruz Pedro. (2010). *Inteligencia Artificial con aplicaciones a la ingeniería*. México, Alfaomega.
- Shyan, C., Andy C., Chang, C., y Chien W. (2010). “A comparison of the forecasting volatility performance of EWMA family models”. *International Research Journal of finance and economics*, Vol.54, pp. 19-28.
- Swales, G, y Young, Y. (1992). “Applying Artificial Neural Networks to Investment Analysis”, *Financial Analysts Journal*, Vol. 48, Núm. 5, pp. 78-81.
- Villada, Fernando., Cadavie, Diego R., y Molina Juan D. (2008). “Pronóstico del precio de la energía eléctrica usando redes neuronales artificiales”. *Revista Facultad de ingeniería*, julio, Núm.44, pp.111-118.
- Villamil Torres, Jaime A. y Delgado Rivera, Jesús A. (2007). “Entrenamiento de una red neuronal multicapa para la tasa de cambio euro - dólar (EUR/USD)”. *Ingeniería e investigación*, Vol.27, Núm.3, pp. 106-117.