



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE ECONOMÍA

**Análisis de Modelos de Volatilidad Estocástica y
Redes Neuronales para el Pronóstico de Datos
Financieros**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Licenciado en Economía

P R E S E N T A:

Gloribella Priego García



**DIRECTOR DE TESIS:
Mtro. José Alberto Reyes de la Rosa**

Ciudad Universitaria, CD. MX., 2017



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Índice

Capítulo 1. Introducción	3
1.1 Justificación de la investigación	3
1.2 Objetivo general.....	4
1.3 Objetivos particulares	4
1.4 Hipótesis	4
Capítulo 2. Los Métodos de pronóstico para series financieras	5
2.1 Métodos tradicionales	5
2.2 Métodos no tradicionales	5
Capítulo 3. Bases de econometría.....	8
3.1 Normalidad	8
3.2 Multicolinealidad	9
3.3 Homoscedasticidad	10
3.4 Autocorrelación	11
Capítulo 4. Bases de series de tiempo	12
4.1 Procesos estocásticos	12
4.1.1 Proceso estocástico estacionario.....	12
4.1.2 Proceso estocástico no estacionario (caminata aleatoria).....	12
4.1.3 Proceso estocástico de raíz unitaria	14
4.1.4 Proceso estocástico estacionario en tendencia y estacionario en diferencias	14
4.1.5 Procesos estocásticos integrados	16
Capítulo 5. Pronóstico con econometría de series de tiempo	17
5.1 Modelo ARMA	17
5.1.1 Modelo autorregresivo (AR)	17
5.1.2 Modelo de medias móviles	18
5.2 Modelo ARIMA.....	18
5.3 Modelos de volatilidad estocástica	18
5.3.1 Modelo ARCH.....	19
5.3.2 Modelo GARCH.....	23
Capítulo 6. Pronóstico con redes neuronales.....	27
6.1 La neurona artificial.....	27
6.2 Red neuronal artificial (RNA)	33
6.2.1 Arquitectura de las redes neuronales artificiales	33
6.2.2 Mecanismos de aprendizaje y recuerdo.....	36

6.2.3	Validación de la Red Neuronal.....	42
6.3	Modelos de RNA's	42
6.3.1	Perceptrón simple	42
6.3.2	Red neuronal con perceptrones.....	43
6.3.3	Adaline	46
6.3.4	Red de Hopfield.....	46
Capítulo 7.	Aplicaciones	48
7.1	Modelo ARCH para predicción	50
7.1.1	Adecuación del modelo ARCH.....	51
7.1.2	Validación del modelo ARCH.....	57
7.1.3	Pronóstico con modelo ARCH	61
7.2	Modelo GARCH para predicción	62
7.2.1	Adecuación del modelo GARCH	62
7.2.2	Validación del modelo GARCH.....	64
7.2.3	Pronóstico con modelo GARCH	68
7.3	Modelo de RNA para predicción	70
7.4.1	Adecuación de la arquitectura de RNA	70
7.4.2	Entrenamiento del modelo de RNA	75
7.4.3	Validación del modelo de RNA	79
7.4.4	Pronóstico con modelo de RNA	79
Capítulo 8.	Conclusiones.....	84
	Bibliografía.....	85
	Anexos.....	87

Capítulo 1. Introducción

Como es bien sabido, no podemos controlar los sucesos futuros sin embargo podemos realizar una aproximación a ellos o al menos simular un comportamiento que permita intuir cualitativa y cuantitativamente el desenlace de un determinado hecho. El presente trabajo está diseñado para estudiar los modelos de pronóstico para las series de tiempo financieras, las cuales debido a su alta volatilidad no han sido del todo desentrañadas en sus patrones de comportamiento que son necesarios para su proyección a futuro. Las técnicas utilizadas corresponden a métodos econométricos comunes y un método no tradicional que en específico aparenta un proceso adaptativo de los factores que conforman el comportamiento de los sucesos: las redes neuronales. Dentro de contenido general, en el capítulo 2 se hablará de estos métodos tradicionales y no tradicionales de pronóstico de datos; en los capítulos 3 y 4 se estipulan las bases de econometría y de series de tiempo necesarias para el tratamiento de los mismos; en el capítulo 5 se describen los modelos ARCH y GARCH y, en el 6 se cita el modelo de red neuronal artificial o RNA. El último apartado está dedicado a la ejecución de tales técnicas para el pronóstico de la serie de retornos intradía de la acción AAPL, correspondiente a la empresa estadounidense Apple Inc. Se utilizó el programa SPSS para la construcción RNA y el programa Eviews para los modelos econométricos. En las conclusiones se deduce finalmente la eficiencia los métodos estudiados.

1.1 Justificación de la investigación

Esta tesis está encaminada a hacer uso de técnicas innovadoras en el proceso de predicción de datos financieros. Su principal motivación ha sido alcanzar a estudiar y dominar una de las etapas más avanzadas de los cursos econométricos y estadísticos correspondientes a la modelización aplicada al pronóstico. Empero, tal dominio no sólo se limita a la implementación de la enseñanza temática que la Facultad de Economía ofrece en su plan de estudios, si no que explora las innovaciones en el ámbito económico-financiero para complementar y fortalecer los conocimientos ya adquiridos. En este sentido, la presente investigación hace análisis de los modelos de volatilidad estocástica *ARCH* y *GARCH* (procesos evolucionados de la econometría básica), así como de las redes neuronales artificiales (método pilar de la inteligencia artificial con aplicaciones en varios ámbitos, incluido el económico-financiero).

1.2 Objetivo general

Comparar la eficiencia en el pronóstico de series de datos financieros mediante dos vías: los métodos econométricos tradicionales para series de tiempo volátiles (en específico modelos *ARCH* y *GARCH*) y los métodos innovadores correspondientes a la inteligencia artificial (en este caso el modelo de red neuronal artificial).

1.3 Objetivos particulares

Conocer a detalle los métodos econométricos que adaptan la volatilidad en su desarrollo teórico y práctico para el pronóstico de series de tiempo fluctuantes.

Comprender el mecanismo bajo el cual trabajan las redes neuronales con el fin de poderlo aplicar en otras áreas correspondientes al pronóstico económico-financiero.

1.4 Hipótesis

Las redes neuronales artificiales tienen igual o mayor eficacia en el pronóstico de series de tiempo financieras que los modelos estadísticos comunes.

Capítulo 2. Los Métodos de pronóstico para series financieras

Existen varias técnicas de pronóstico en el ámbito económico-financiero, la mayor parte de ellas se basan en conocimientos estadísticos y econométricos, sin embargo el desarrollo de la inteligencia artificial en conjunto con el progreso de la tecnología computacional, ha permitido la aplicación de otros procesos predictivos eficaces como los algoritmos genéticos o las redes neuronales artificiales. Una vez adquiridos los conocimientos para su ejecución, el uso fácil de software permite obtener, procesar, y actualizar con gran rapidez la carga específica de datos que se requiera tratar.

2.1 Métodos tradicionales

Los métodos de pronóstico tradicionales consisten en técnicas de descomposición, promedios móviles, suavización exponencial y análisis de regresión o de correlación. Dichos métodos se erigen bajo el supuesto de que la varianza de una serie cronológica es constante (homoscedástica), condición para alcanzar la estacionariedad. Sin embargo, en ocasiones la varianza tiene cambios sistemáticos en el tiempo, sobre todo tratándose de series financieras, en donde el fenómeno es mucho más complejo.

En este sentido, para el tratamiento y predicción de series de tiempo volátiles, los modelos *ARCH* (Heteroscedasticidad Condicional Autorregresivo) y *GARCH* (generalización de los modelos *ARCH*) son los más comunes entre los administradores de riesgo y financieros ya que capturan considerablemente la heteroscedasticidad. Estos modelos se expondrán más adelante.

2.2 Métodos no tradicionales

Los modelos de pronóstico anteriores extrapolan el valor de la variable de interés a través de mecanismos que utilizan datos históricos, consecuentemente trabajan bajo el supuesto de que el futuro está en cierta razón predefinido por el pasado. Así mismo, la estructuración de estos modelos incluye ceñirse a reglas o suposiciones acerca de la distribución de los datos, su correlación, estabilidad, varianza, etc.; es decir, adaptan los sucesos a un modelo teóricamente correcto. Contrariamente, los métodos no convencionales adaptan un modelo a los hechos pasados o información en general para

estimar un resultado en lo posterior. Uno de estos métodos son las redes neuronales artificiales o RNA.

El programa de la red neuronal asimila un rango completo de relaciones pasadas entre un conjunto de variables “he intenta desarrollar las relaciones subyacentes a través del aprendizaje conforme lo asimila” (Hanke, 2010:489). Una de las ventajas del uso de redes neuronales es que su estructura y comportamiento se funda sobre entrenamientos suministrados y no necesita de una relación especificada como en los métodos tradicionales; tampoco requiere de datos provenientes de una población con distribución de probabilidad determinada y el desempeño de la red puede ser corregido para obtener mayor exactitud en los resultados.

Las redes neuronales son poderosos mecanismos para pronosticar datos financieros, debido a que son eficaces en la modelación de sistemas claramente no lineales o sucesos altamente correlacionados.

A los puntos anteriores se añaden las siguientes ventajas:

- Poseen un aprendizaje adaptativo, es decir los elementos procesales (neuronas) pueden ser reformados por el supervisor o autoajustarse por la misma red mediante la generación de su propia distribución de pesos en atención al algoritmo de aprendizaje adecuado.
- Son mecanismos auto-organizados, ya que con el fin de obtener los resultados especificados, la red posee la capacidad de reestructurarse en su totalidad.
- Toleran fallos en sus distintas fases de procesamiento:
 - a) Durante el entrenamiento la red es capaz de aprender aún con patrones distorsionados, incompletos o con ruido; así mismo almacena información no localizada para generar un patrón de salida.
 - b) Durante el funcionamiento la actividad de la red no se ve afectada ante daños o destrucción de su estructura.
- Operan en tiempo real, realizando modificaciones en los pesos de sus conexiones durante el aprendizaje.
- Recurren al uso de la tecnología en su construcción y entrenamiento, dado que una red puede ser trasladada fácilmente a un hardware de bajo coste.

Por consiguiente, la diferencia más radical entre los métodos tradicionales y no tradicionales es la definición y restricción del algoritmo que procesará la información y será ejecutado por el ordenador computacional.

Capítulo 3. Bases de econometría

La econometría es una disciplina que hace uso de las matemáticas y la estadística para el análisis cuantitativo de modelos económicos. Un modelo es aquella versión simplificada de la realidad que incluye variables dependientes e influyentes, y puede ser representado mediante expresiones matemáticas. En este capítulo se definirán algunos de los conceptos fundamentales correspondientes a las características estadísticas (incluyendo su detección, causas y corrección) que presentan específicamente los modelos econométricos.

3.1 Normalidad

El supuesto de normalidad para el término de error tiene como finalidad representar la influencia estable combinada de un gran número de variables independientes (no incluidas en el modelo) sobre la variable dependiente. Es decir, se desea que los errores posean una dispersión más menos comportada, lo que indicaría que la estimación del modelo matemático se acopla casi correctamente a la realidad a pesar de no incluir todas las variables de influencia posibles. El teorema del límite central proporciona la justificación teórica para el supuesto de normalidad del término de error (a una mayor cantidad de datos y experimentos realizados los errores tenderán a comportarse más uniformemente, lo que se puede plasmar en una distribución normal). Bajo este teorema se enuncia también, que cuando el número de variables no sea muy grande y éstas no sean estrictamente independientes, la distribución de su probabilidad tiende a seguir una distribución normal; lo mismo sucede con la probabilidad de los parámetros estimados.

a) Causas de no normalidad:

- Presencia de datos atípicos.
- Uso de variables con asimetría.
- Incorrecta especificación del modelo.

b) Pruebas para detectar:

- Jarque Bera.
- Shapiro-Wilk.
- Kolmogorov-Smirnov.
- Anderson-Darling.

- c) Consecuencias: La inexistencia de normalidad provoca que los estimadores ya no posean varianza mínima y por tanto que no sean eficientes.
- d) Corrección:
 - Modificación de la forma funcional del modelo.
 - Elección de nuevas variables.

3.2 Multicolinealidad

La multicolinealidad es la existencia de una relación lineal perfecta o exacta entre algunas o todas las variables explicativas de un modelo de regresión. En un modelo econométrico es inadmisibles la presencia de este efecto dado que, cuando se desea explicar de manera “pura” la alteración de la variable independiente a través de la influencia de una variable independiente, esta debe medirse individualmente sin tomar en consideración los efectos o alteraciones que se produzcan ante las relaciones con otras variables independientes.

- a) Causas:
 - Método de recolección de información deficiente.
 - Restricciones en el modelo o en la población objeto de muestra.
 - Incorrecta especificación del modelo.
 - Modelo sobredeterminado.
- b) Síntomas:
 - La razón t de uno o más coeficientes es estadísticamente no significativa.
 - R^2 , la medida global de bondad de ajuste puede ser muy alta.
 - Los estimadores presentan varianzas y covarianzas grandes.
 - Los estimadores mínimos cuadrados ordinarios y sus errores estándar son sensibles a pequeños cambios en los datos.
- c) Pruebas para detectar:
 - Un R^2 elevado con pocas razones t significativas para los parámetros.
 - Altas correlaciones entre parejas de regresores.
 - Examen de correlaciones parciales.
 - Ejecución de regresiones auxiliares entre las mismas variables independientes para observar su nivel de correlación.
 - Observación del factor de inflación de la varianza.
 - Diagrama de dispersión.
- d) Consecuencias: Los estimadores dejan de ser insesgados.

e) Corrección:

- Información a priori que justifique multicolinealidad.
- Combinación de información de corte transversal y de series de tiempo.
- Eliminación de una o varias variables.
- Transformación de variables.
- Datos nuevos o adicionales.

3.3 Homoscedasticidad

La homoscedasticidad ocurre cuando la varianza de las perturbaciones no es constante a lo largo de las observaciones. En la modelación econométrica se desea por todas partes minimizar esta característica, ya que es indicativa de que el ajuste matemático del modelo es altamente inconstante.

a) Causas:

- Aprendizaje de errores.
- Adaptación de una variable que depende de la decisión humana (por ejemplo, el consumo se adapta al ingreso en función de las expectativas de su aumento).
- Mejora en las técnicas de recolección de datos.
- Presencia de datos atípicos o aberrantes.
- Modelo incorrectamente especificado.
- Asimetría en la distribución de una o más regresoras incluidas en el modelo.
- Incorrecta transformación de los datos.

b) Pruebas para detectar:

- Método gráfico: existencia de patrones en los residuos.
- Prueba de Park.
- Prueba de Glejser.
- Prueba de correlación de orden de Spearman.
- Prueba de Goldfeld Quandt.
- Prueba Breush Pagan Godfrey.
- Prueba general de heteroscedasticidad de White.

c) Consecuencias: Los estimadores aún son lineales e insesgados pero dejan de tener varianza mínima, motivo por el cual ya no son eficientes.

d) Corrección:

- Mínimos cuadrados ponderados

3.4 Autocorrelación

Este concepto refiere la correlación existente entre los miembros de una serie de observaciones ordenadas en el tiempo o en el espacio.

a) Causas:

- Inercia o pasividad en la propia serie.
- Sesgo de especificación (datos excluidos de la serie).
- Sesgo de especificación por la forma funcional incorrecta.
- Fenómeno de la telaraña.
- Rezagos.
- No estacionariedad.
- Transformación incorrecta de datos.

b) Pruebas para detectar:

- Método gráfico o revisión del correlograma de residuos al cuadrado.
- Prueba de las rachas.
- Prueba d de Durbin Watson.
- Prueba de Breush Godfrey.

c) Consecuencias: Los estimadores dejan de ser eficientes

d) Corrección:

- Uso del método de Newey Way
- Método generalizado de mínimos cuadrados
- Modificación de la forma funcional

Capítulo 4. Bases de series de tiempo

Un modelo de series temporales describe la senda de una variable y_t en términos de factores x_t contemporáneos (incluso retardos), errores ϵ_t , y su propio pasado, y_{t-1} , por ejemplo,

$$y_t = \beta_1 + \beta_2 x_t + \beta_3 y_{t-1} + \epsilon_t \quad 4.1$$

Las series temporales son sucesivos hechos de un proceso aleatorio o estocástico (Greene, 1999: 710).

4.1 Procesos estocásticos

Un proceso estocástico es aquel que se desarrolla al azar en el tiempo, es decir, consiste en secuencias de variables aleatorias $x_0, x_1, x_2 \dots$ en las que x_n representa alguna magnitud aleatoria a tiempo n . Por lo general, el valor x_n a tiempo n puede depender del valor x_{n-1} a tiempo $n - 1$, o también de los valores x_m a otros tiempos $m < n$. (Evans, 2004: 649)

4.1.1 Proceso estocástico estacionario

Es aquella sucesión o secuencia de variables aleatorias que devienen de una población con media cero, varianza constante y cuyo valor de la covarianza entre dos periodos depende sólo de la distancia o rezago entre dichos periodos y no del tiempo en el cual se calculó la covarianza; regularmente se suponen independientes o normalmente distribuidas. Este proceso estocástico está ordenado según el tiempo t :

$$\{\epsilon_t\}, t = -\infty, +\infty$$

Donde cada elemento de la sucesión tiene $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma_\epsilon^2$, y $Cov(\epsilon_t, \epsilon_s) = 0$ para todo $s \neq t$. Este proceso también es conocido como ruido blanco y es el cimiento de los modelos de series de tiempo (Gujarati, 2009: 740).

4.1.2 Proceso estocástico no estacionario (caminata aleatoria)

Como su nombre lo indica es una sucesión de variables aleatorias que pertenecen a una población que no cumple con alguno de los supuestos antes mencionados como media cero, varianza constante o covarianza independiente del tiempo. El proceso más común es el de la caminata aleatoria. Suponiendo que ϵ_t es un término de error de ruido blanco, con media 0 y varianza σ^2 . Entonces decimos que la serie y_t es una caminata aleatoria si:

$$y_t = y_{t-1} + \epsilon_t \quad 4.2$$

El valor de y en el tiempo t es igual a su valor en el tiempo $(t - 1)$ más un choque aleatorio ϵ_t . Existen dos tipos de caminata aleatoria:

1. Caminata aleatoria sin deriva o sin desvío (es decir, sin término constante o de intercepto)
2. Caminata aleatoria con deriva o con desvío (es decir, con término constante)

Para determinar una caminata aleatoria sin deriva, de la ecuación 4.2 podemos escribir:

$$y_1 = y_0 + \epsilon_1$$

$$y_2 = y_1 + \epsilon_2 = y_0 + \epsilon_1 + \epsilon_2$$

$$y_3 = y_2 + \epsilon_3 = y_0 + \epsilon_1 + \epsilon_2 + \epsilon_3$$

En general si el proceso comenzó en el tiempo 0 con un valor de y_0 , tenemos:

$$y_t = y_0 + \sum \epsilon_t \quad 4.3$$

un modelo de caminata aleatoria sin deriva donde y_t es la suma de y_0 inicial más la suma de los choques aleatorios $\sum \epsilon_t$ (tendencia estocástica). El impacto de tales choques perdura en el tiempo, por ejemplo, si $\epsilon_2 = 2$, en vez de $\epsilon_2 = 0$, todas las y_t de y_2 en adelante serán dos unidades mayores, por lo que nunca cesa el efecto. La caminata aleatoria tiene memoria infinita. Consiguientemente:

$$E(y_t) = E\left(y_0 + \sum \epsilon_t\right) = y_0$$

Esto indica que la media de y es igual a su valor inicial (constante). De igual forma se demuestra que

$$Var(y_t) = t\sigma^2$$

Conforme se incrementa t , su varianza aumenta de manera indefinida, lo que viola una condición de la estacionariedad (Gujarati, 2009: 743). Así, el modelo de caminata aleatoria (MCA) sin deriva es un proceso estocástico no estacionario.

Para definir la caminata aleatoria con deriva modificamos (4.2) como sigue:

$$y_t = \delta + y_{t-1} + \epsilon_t \quad 4.4$$

Donde δ se conoce como el parámetro de la deriva. El término deriva proviene del hecho de que si escribimos la ecuación anterior en primeras diferencias

$$y_t - y_{t-1} = \Delta y_t = \delta + \epsilon_t$$

se demuestra que y_t se deriva o se desvía hacia arriba o hacia abajo, según δ sea positiva o negativa. Es posible demostrar que para el modelo de caminata aleatoria con deriva (4.4):

$$E(y_t) = y_0 + t \cdot \delta$$

$$Var(y_t) = t\sigma^2$$

la media, al igual que la varianza se incrementa con el tiempo, lo que viola de nuevo las condiciones de estacionariedad. Por ende, el MCA, con o sin deriva, es un proceso estocástico no estacionario.

4.1.3 Proceso estocástico de raíz unitaria

Un proceso estocástico no estacionario posee una raíz unitaria si al obtener sus primeras diferencias dicho proceso se vuelve estacionario. Esto se explica de la siguiente manera. Expresando el modelo de caminata aleatoria como:

$$y_t = \rho y_{t-1} + \epsilon_t$$

$$-1 \leq \rho \leq 1$$

Si $\rho = 1$, las primeras diferencias del modelo son:

$$(y_t - y_{t-1}) = \Delta y_t = \epsilon_t$$

Es decir, la serie de primeras diferencias resultante ϵ_t es estacionaria. En resumen, si ρ es en efecto 1, tenemos el problema de raíz unitaria, que equivale a un caso de no estacionariedad. El proceso descrito hasta ahora corresponde a una caminata aleatoria sin deriva, cuya varianza de y_t es no estacionaria. “El nombre de raíz unitaria se debe a que $\rho = 1$. Por tanto los términos no estacionariedad, caminata aleatoria, raíz unitaria y tendencia estocástica se consideran sinónimos” (Gujarati, 2009: 744).

4.1.4 Proceso estocástico estacionario en tendencia y estacionario en diferencias

Un proceso estocástico es estacionario en tendencia cuando una serie de tiempo posee una tendencia determinista, es decir, predecible y no variable. Así mismo, un proceso estocástico es estacionario en diferencias cuando la tendencia de una serie de tiempo no es

predecible, si no estocástica (y debe realizarse un proceso de primeras diferencias para ser identificada). Formalizando la definición consideremos un modelo de serie de tiempo y_t :

$$y_t = \beta_1 + \beta_2 t + \beta_3 y_{t-1} + \epsilon_t \quad 4.5$$

Donde ϵ_t es un término de error de ruido blanco y donde t es el tiempo medio cronológicamente. Tenemos las siguientes clasificaciones (Gujarati, 2009: 744):

- Caminata aleatoria pura: si en (4.5) $\beta_1 = 0, \beta_2 = 0, \beta_3 = 1$, obtenemos

$$y_t = y_{t-1} + \epsilon_t$$

un modelo de caminata aleatoria sin deriva y por tanto no estacionario, pero cuyas primeras diferencias

$$\Delta y_t = (y_t - y_{t-1}) = \epsilon_t$$

si son estacionarias. Por tanto, un MCA sin deriva es un proceso estacionario en diferencias.

- Caminata aleatoria con deriva: si en (4.5) $\beta_1 \neq 0, \beta_2 = 0, \beta_3 = 1$ obtenemos

$$y_t = \beta_1 + y_{t-1} + \epsilon_t$$

que es una caminata aleatoria con deriva y en consecuencia es no estacionaria. Si la expresamos como

$$(y_t - y_{t-1}) = \Delta y_t = \beta_1 + \epsilon_t \quad 4.5a$$

Esto significa que y_t mostrará una tendencia positiva ($\beta_1 > 0$) o negativa ($\beta_1 < 0$), es decir, una tendencia estocástica. La ecuación (4.5a) es también un proceso estacionario en diferencias, ya que estas vuelven el proceso estacionario.

- Tendencia determinista: Si en (4.5), $\beta_1 \neq 0, \beta_2 \neq 0, \beta_3 = 0$, obtenemos

$$y_t = \beta_1 + \beta_2 t + \epsilon_t \quad (3.5)$$

lo cual se llama proceso estacionario en tendencia. Aunque la media de y_t es $\beta_1 + \beta_2 t$ —no constante—, su varianza ($= \sigma^2$) sí lo es. Una vez que conocemos los valores de β_1 y β_2 , podemos pronosticar la media; si restamos la media de y_t de y_t , la serie resultante será estacionaria.

- Caminata aleatoria con deriva y tendencia determinista: Si en (4.5) $\beta_1 \neq 0, \beta_2 \neq 0, \beta_3 = 1$, obtenemos

$$y_t = \beta_1 + \beta_2 t + y_{t-1} + \epsilon_t$$

en cuyo caso tenemos una caminata aleatoria con deriva y tendencia determinista, lo cual se observa al expresarse esta ecuación como

$$\Delta y_t = \beta_1 + \beta_2 t + \epsilon_t$$

que significa que y_t es no estacionaria

- Tendencia determinista con componente estacionario: $\beta_1 \neq 0$, $\beta_2 = 0$, $\beta_3 < 1$, obtenemos:

$$y_t = \beta_1 + \beta_2 t + \beta_3 y_{t-1} + \epsilon_t$$

que es estacionaria alrededor de la tendencia determinista.

4.1.5 Procesos estocásticos integrados

“Si una serie de tiempo (no estacionaria) debe diferenciarse d veces para hacerla estacionaria, decimos que la serie es integrada de orden d ” (Gujarati, 2009: 746), por consiguiente si una serie de tiempo es diferenciada una vez se denomina proceso integrado de orden 1 y se denota como $I(1)$. De manera similar, si la serie es diferenciada dos veces (tomando las primeras diferencias de la serie ya diferenciada), entonces se denomina integrada de orden 2. Una serie de tiempo y_t integrada de orden d se denota como $y_t \sim d$.

Si una serie es de naturaleza estacionaria entonces es integrada de orden cero ($y_t \sim 0$), ya que no requiere ninguna diferenciación.

Capítulo 5. Pronóstico con econometría de series de tiempo

En este capítulo se describirán los métodos econométricos de pronóstico elementales, que hacen uso de datos históricos para su estimación futura.

5.1 Modelo ARMA

Un modelo *ARMA*, como su nombre lo indica, es aquel que incluye tanto términos autorregresivos (*AR*) como de medias móviles (*MA*). El proceso más simple se escribe como sigue:

$$y_t = \theta + \varphi_1 y_{t-1} + \epsilon_t + \beta_1 \epsilon_{t-1}$$

Donde θ es un término constante. Por ende, en un proceso *ARMA*(p, q), habrá p términos autorregresivos y q términos de promedios móviles. Estas estructuras se describen a continuación.

5.1.1 Modelo autorregresivo (AR)

Un modelo autorregresivo (*AR*) describe una clase particular de proceso en el que las observaciones en un momento dado son predecibles a partir de las observaciones previas del proceso más un término de error. El caso más simple es el *AR*(1) o de primer orden, cuya expresión matemática es:

$$y_t = \varphi_1 y_{t-1} + \epsilon_t$$

Esta expresión indica que el valor del pronóstico de y en el periodo t es una proporción φ_1 de su valor en el periodo $t - 1$, más un choque o perturbación aleatoria en el tiempo t . Consiguientemente, el proceso autorregresivo de orden p , representado por *AR*(p) toma la forma:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \epsilon_t$$

Los modelos sólo consideran los valores anteriores como regresores.

5.1.2 Modelo de medias móviles

Un modelo de medias móviles describe una serie temporal estacionaria. En este modelo el valor actual puede predecirse a partir de un promedio móvil de los términos de error presente y pasado. El modelo denotado por $MA(1)$ se expresa como sigue:

$$y_t = \mu + \epsilon_t + \beta_1 \epsilon_{t-1}$$

Donde μ es un término constante. El proceso de medias móviles de orden q , representado por $MA(q)$, viene dado por la expresión:

$$y_t = \mu + \epsilon_t + \beta_1 \epsilon_{t-1} + \beta_2 \epsilon_{t-2} + \dots + \beta_q \epsilon_{t-q}$$

Así, un proceso de promedios móviles es una combinación lineal de términos de error de ruido blanco.

5.2 Modelo ARIMA

Un modelo $ARIMA(p, d, q)$, denominado proceso autorregresivo integrado de medias móviles de orden p, d, q , es una serie temporal que se convierte en un ruido blanco (proceso puramente aleatorio) después de ser diferenciada d veces. El modelo $ARIMA(0, d, 0)$ se expresa mediante

$$(1 - L)^d y_t = \Delta^d y_t = \epsilon_t$$

Donde L es una constante, en tanto que d indica el número de diferencias que son necesarias para convertir la serie original en estacionaria.

5.3 Modelos de volatilidad estocástica

Las series de tiempo financieras se comportan como una caminata aleatoria, por lo que no son estacionarias; no obstante, sus primeras diferencias sí lo son, aunque presentan alta volatilidad o varianza desigual en el tiempo, lo que finalmente supone un problema para su tratamiento y pronóstico. Para esta clase de series se desarrolló el modelo de heteroscedasticidad condicional autorregresivo (ARCH) y sus variantes, como el modelo de heteroscedasticidad condicional autorregresivo generalizado (GARCH), los cuales son una solución a los problemas de varianza desigual y correlación entre observaciones.

5.3.1 Modelo ARCH

Como se ha venido indicando, una serie de tiempo solo podrá ser modelada matemáticamente si es estacionaria. Para que los datos de tiempo financieros se vuelvan estacionarios es necesario obtener sus primeras diferencias o simplemente calcular sus retornos, ya que estos fungen como una diferenciación al incluir en su cálculo la siguiente formulación:

$$y_t = x_t - x_{t-1}$$

No obstante, los retornos de datos financieros se caracterizan por presentar aglutinamientos de volatilidad. Este hecho, perteneciente a la variable de interés denota una clase especial de heteroscedasticidad ya que se manifiesta en un patrón determinado por agrupamientos (clusters) de periodos de variabilidad alta, seguidos de agrupamientos de periodos de variabilidad baja.

“Este patrón nos sugiere una ‘memoria’ de la volatilidad. Es decir, a una perturbación grande en el periodo actual, le corresponde una perturbación importante en los instantes inmediatamente futuros, por lo menos en promedio” (Fabris, 2009: 189). Esta peculiaridad es útil para la estipulación de un patrón y posible pronóstico en los periodos consecutivos. Siguiendo esta lógica, si un inversionista toma decisiones en un contexto financiero de baja variabilidad, entonces deberá tomar en cuenta en su pronóstico de variabilidad para los instantes prontamente subsecuentes que el presente es un momento de mercado tranquilo.

Este tipo de pronóstico se denomina condicional, ya que depende de la información disponible hasta el momento, y es más acertado que la proyección simple de la media de la variabilidad (o pronóstico incondicionado).

Los inversionistas basan su elección sobre variables financieras en dos aspectos principales: un valor promedio o esperado (expectativa) y las fluctuaciones observadas en periodos inmediatamente precedentes. En este sentido, resulta conveniente atender a un modelo de heteroscedasticidad condicionada autorregresivo, diseñado para determinar un patrón de comportamiento estadístico para la varianza (Arce, 1988:3).

“Puede existir un proceso definido a partir de un ‘ruido blanco’, en el que la media y la varianza marginales sean constantes; y, al mismo tiempo, la media condicional puede ser constante y la varianza condicional no fija” (Arce, 1998:3)

Con el objeto de poder realizar inferencia ante el fenómeno de heteroscedasticidad, el modelo ARCH se consigue multiplicando una variable aleatoria denominada ϵ_t (ruido blanco normal) por un factor σ_t , conocido como volatilidad condicional que amplificará o reducirá las oscilaciones del proceso estocástico ϵ_t .

El modelo ARCH está estipulado bajo la siguiente expresión:

$$y_t = \epsilon_t \sigma_t$$

$$\epsilon_t \sim iidN(0,1)$$

donde:

y_t : Variabilidad

ϵ_t : Término de error de ruido blanco homoscedástico, idénticamente distribuido con media cero y desviación típica igual a uno (varianza constante, no hay correlación con su pasado, luego tampoco la hay con el pasado de y_t)

σ_t : Volatilidad condicional (siempre no negativa)

Así mismo, la volatilidad está conformada de la siguiente manera:

$$\sigma_t = (\alpha_0 + \alpha_i y_{t-i}^2)^{1/2}$$

donde:

α_0 : valor medio alrededor del cual giran los demás valores (término constante positivo)

α_i : factor de amplificación que representa la influencia que los rezagos tienen sobre la volatilidad σ_t

y_{t-i}^2 : el valor de $t - i$ es una realización concreta conocida no aleatoria.

En el modelo de Engle, la denominación autorregresivo es inapropiada ya que el término contemporáneo (valuado en t) del lado izquierdo del modelo, no es el mismo que aparece rezagado (valuado en $t - 1$) en el lado derecho, pero esta es la denominación que ha permanecido.

En una notación más desarrollada el modelo ARCH se especifica como sigue:

$$y_t = \epsilon_t (\alpha_0 + \alpha_i y_{t-i}^2)^{1/2}$$

Ó lo que es lo mismo:

$$y_t^2 = \epsilon_t^2(\alpha_0 + \alpha_i y_{t-i}^2)$$

$$i = 1 \dots q$$

Con $\sigma_t = +\sqrt{\sigma_t^2}$ (raíz cuadrada positiva)

La variabilidad de las series financieras se expresa bajo un modelo denotado por y_t , que depende de un choque aleatorio ϵ_t (ruido blanco) potenciado por la volatilidad contextual σ_t que no es otra cosa que valores conocidos pasados que giran alrededor de una media o expectativa financiera.

Esta volatilidad está conformada por un primer término constante (positivo) α_0 y un segundo término que depende del valor de la variable de interés en el periodo anterior σ_t . El coeficiente α_1 (entre cero y uno) representa la influencia de σ_t en el valor actual y_t ; así mismo, los cuadrados aparecen para establecer relaciones entre las magnitudes absolutas de las variables sin considerar el signo de las mismas, el que queda determinado aleatoriamente por el signo de ϵ_t .

Una característica de las series financieras es que el cuadrado de los retornos está autocorrelacionado, mientras que los retornos no. El proceso ARCH está diseñado para capturar este hecho empírico.

En el modelo:

1. La dispersión del error respecto a su media está cimentado sobre la información pasada.
2. Su construcción se justifica ante la hipótesis de permanencia estructural en los fenómenos financieros y económicos.

Para cumplirse la condición de estacionariedad de la media de y_t , la suma de todos los parámetros debe ser menor que la unidad:

$$\alpha_0 > 0$$

$$\alpha_i \geq 0$$

$$i = 1 \dots q$$

Dado que ϵ_t se distribuye normalmente, entonces y_t es condicionalmente normal y su varianza es σ_t^2 . Así mismo, el proceso y_t generado también es estacionario, por lo que el

proceso ARCH es en sí mismo diferenciado, su media marginal es cero y no posee autocorrelación (todas las autocovarianzas son cero).

La distribución marginal de y_t no tiene forma conocida, no obstante, dado que todos sus momentos impares son cero, su distribución es simétrica. Los cuadrados de las realizaciones pasadas permiten relacionar las magnitudes absolutas de las variables sin considerar el signo de las mismas, el que queda determinado aleatoriamente por el signo de ϵ_t .

El orden 1 del ARCH se refiere al alcance de la memoria del proceso. En $ARCH(p)$, p indicaría el máximo de rezagos considerados en el modelo. Por ejemplo, en un modelo $ARCH(3)$ la volatilidad σ_t^2 se modelaría como:

$$\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 y_{t-2}^2 + \alpha_3 y_{t-3}^2$$

Los coeficientes se estiman por regresiones realizadas por software, obteniéndose las estimaciones α_0 y α_i .

Los momentos del modelo $ARCH$ se detallan de la siguiente manera:

Esperanza:

$$E(y_t) = E(\epsilon_t(\alpha_0 + \alpha_i y_{t-i}^2)^{1/2})$$

$$E(y_t) = E(\epsilon_t)E((\alpha_0 + \alpha_i y_{t-i}^2)^{1/2})$$

$$\rightarrow E(\epsilon_t) = 0$$

$$E(y_t) = 0$$

Varianza:

$$E(y_t^2) = E(\epsilon_t^2(\alpha_0 + \alpha_i y_{t-i}^2))$$

$$E(y_t^2) = E(\epsilon_t^2)E(\alpha_0 + \alpha_i y_{t-i}^2)$$

$$\rightarrow E(\epsilon_t^2) = \sigma_t^2$$

$$E(y_t^2) = \sigma_t^2(\alpha_0 + \alpha_i E(y_{t-i}^2))$$

$$\rightarrow E(y_{t-i}^2) = \sigma_{t-i}$$

$$E(y_t^2) = \sigma_t^2(\alpha_0 + \alpha_i \sigma_{t-i})$$

$$E(y_t^2) = \frac{\alpha_0}{1 - \alpha_i} \sigma_t^2$$

Mientras que los momentos condicionales se expresan como sigue:

Esperanza condicional:

$$E_{t-1}(y_t) = E_{t-1}(\epsilon_t(\alpha_0 + \alpha_i y_{t-i}^2)^{1/2})$$

$$E_{t-1}(y_t) = E_{t-1}(\epsilon_t) E_{t-1}((\alpha_0 + \alpha_i y_{t-i}^2)^{1/2})$$

$$E_{t-1}(y_t) = 0$$

Varianza condicional:

$$E_{t-1}(y_t^2) = E_{t-1}(\epsilon_t^2(\alpha_0 + \alpha_i y_{t-i}^2))$$

$$E_{t-1}(y_t^2) = E_{t-1}(\epsilon_t^2) E_{t-1}(\alpha_0 + \alpha_i y_{t-i}^2)$$

$$E_{t-1}(y_t^2) = \sigma_t^2(\alpha_0 + \alpha_i y_{t-i}^2)$$

5.3.2 Modelo GARCH

El modelo Garch se especifica mediante:

$$y_t = \epsilon_t \sigma_t$$

$$\epsilon_t \sim iidN(0,1)$$

Donde:

$$\sigma_t = (\alpha_0 + \alpha_i y_{t-i}^2 + \beta_j y_{t-j}^2)^{1/2}$$

Por lo que el modelo queda construido como se muestra a continuación:

$$y_t = \epsilon_t (\alpha_0 + \alpha_i y_{t-i}^2 + \beta_j y_{t-j}^2)^{1/2}$$

Los parámetros se estipulan bajos los siguientes rangos:

$$\alpha_0 > 0$$

$$\alpha_i, \beta_i \geq 0$$

$$i = 1 \dots q$$

$$j = 1 \dots p$$

La suma de todos los parámetros es menor que la unidad. La función de distribución marginal de y_t tampoco es conocida, pero se pueden calcular los primeros momentos y definir el proceso respecto a su media y su varianza.

Esperanza:

$$E(y_t) = E\left(\epsilon_t(\alpha_0 + \alpha_i y_{t-i}^2 + \beta_j \sigma_{t-j}^2)^{1/2}\right)$$

$$E(y_t) = E(\epsilon_t)E\left((\alpha_0 + \alpha_i y_{t-i}^2 + \beta_j \sigma_{t-j}^2)^{1/2}\right)$$

$$\rightarrow E(\epsilon_t) = 0$$

$$E(y_t) = 0$$

Varianza:

$$E(y_t^2) = E\left(\epsilon_t^2(\alpha_0 + \alpha_i y_{t-i}^2 + \beta_j \sigma_{t-j}^2)\right)$$

$$E(y_t^2) = E(\epsilon_t^2)E(\alpha_0 + \alpha_i y_{t-i}^2 + \beta_j \sigma_{t-j}^2)$$

$$\rightarrow E(\epsilon_t^2) = \sigma_t^2$$

$$E(y_t^2) = \sigma_t^2(\alpha_0 + \alpha_i E(y_{t-i}^2) + \beta_j E(\sigma_{t-j}^2))$$

$$\rightarrow E(y_{t-i}^2) = \sigma_{t-i}$$

$$E(y_t^2) = \sigma_t^2(\alpha_0 + \alpha_i \sigma_{t-i} + \beta_j E(\sigma_{t-j}^2))$$

$$E(y_t^2) = \frac{\alpha_0}{1 - \alpha_i - \beta_j}$$

Los momentos condicionales del modelo se detallan consecutivamente.

Esperanza condicional:

$$E_{t-1}(y_t) = E_{t-1}\left(\epsilon_t(\alpha_0 + \alpha_i y_{t-i}^2 + \beta_j \sigma_{t-j}^2)^{1/2}\right)$$

$$E_{t-1}(y_t) = 0$$

Varianza condicional:

$$E_{t-1}(y_t^2) = E_{t-1}(\epsilon_t^2(\alpha_0 + \alpha_i y_{t-i}^2 + \beta_j \sigma_{t-j}^2))$$

$$E_{t-1}(y_t^2) = \sigma_t^2$$

Como se estipuló, el proceso $GARCH(p, q)$ está conformado por dos partes

1. ϵ_t que es un término aleatorio comportado como ruido blanco y
2. σ_t volatilidad condicionada o varianza en el periodo actual t , que está constituida a su vez por:
 - a) α_0 valor medio constante de largo plazo, sobre el cual se genera expectativa.
 - b) y_{t-i}^2 variaciones observadas previamente amplificadas por un factor α_i .
 - c) σ_{t-j}^2 predicción de la varianza en el último periodo histórico conocido y amplificado por un factor β_j .

En específico, el modelo $GARCH(p, q)$, como su nombre lo indica, generaliza el uso de los datos a partir de la inclusión de sus valores pasados y de su propia varianza. Con esto se pretende modelar a un más las fluctuaciones inherentes a la serie financiera.

Como se enunció anteriormente, el hecho de que no exista autocorrelación entre observaciones del ruido blanco desplazadas en el tiempo, no significa necesariamente que no haya dependencia entre estas de un modo no lineal. Es a partir de la diferenciación o el uso de los retorno de la serie cuando se descubren relaciones de causalidad que responden como proceso estocástico estacionario para el contraste lineal. Es necesario recordar que en la especificación dicho proceso estacionario se debe cumplir con:

$$E(\epsilon_t) = 0$$

$$Var(\epsilon_t) = E(\epsilon_t - 0)^2 = \sigma_t^2$$

$$Cov(\epsilon_t | \epsilon_{t-1}) = 0$$

El valor esperado de los errores se iguala a cero, mientras que la varianza de los mismos con respecto a la media se mantiene constante. La covarianza de los errores pasados con los actuales también debe ser cero.

En la tabla siguiente (5.1) se muestran las pruebas que se llevarán a cabo para validar la correcta especificación de los modelos. Es necesario lograr que la media sea estacionaria

para normalizar el proceso. Una vez realizado el modelo se debe comprobar que la serie de residuos no contiene autocorrelación (no relación en entre el residuo y el periodo precedente); que el contrastaste de homocedasticidad con el test LM ARCH no muestre significativa la existencia de varianzas inconsistentes; que la bondad de la regresión haya sido representativa (aunque en estos modelos por lo general se presenta un coeficiente R^2 mínimo debido a la dificultad de ajuste de las variables explicativas); y que la suma de los coeficientes de la regresión sea menor que uno (en el caso que se modelen retornos financieros).

Tabla 5.1 Contraste econométrico y método estadístico de examen.

Contraste	Método
Raíz unitaria en el proceso aleatorio de la serie	Test Dickey-Fuller
Especificación (homocedasticidad)	Test del Multiplicador de Lagrange ARCH(q) o LM ARCH Test así como la observación del correlograma de los residuos al cuadrado.
Autocorrelación	Prueba Durbin Watson u observación del correlograma de probabilidad estadística de los coeficientes (Q-statistics)
Bondad	R^2 de la regresión calculada
Normalidad	Contraste Jarque-Bera
Inssegadez Pegan y Swert(1990)	Suma de los coeficientes de la regresión $y_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \varepsilon_t$ menor a uno.

Fuente: Elaboración propia con base las pruebas sugeridas por Arce, (1998), p.3

Una vez concluidos los pasos anteriores se podrá realizar la predicción de los datos subsecuentes (Forecast).

Capítulo 6. Pronóstico con redes neuronales

Las redes neuronales artificiales son usadas por los ‘traders’ para pronosticar el comportamiento de los mercados financieros, ya que se perfilan como un instrumento matemático adaptativo que simulan las relaciones o factores que de manera conjugada determinan aproximadamente hechos futuros de interés. Así, los traders pueden tener un panorama no sólo anticipativo, si no retrospectivo del comportamiento de datos financieros en el tiempo. “Desde el punto de vista de la toma de decisiones en condiciones de incertidumbre estos métodos de pronóstico se han convertido en una herramienta imprescindible” (Tharp, 2006:131).

6.1 La neurona artificial

Una neurona biológica humana está conformada, como toda célula, por tres partes: membrana, citoplasma y núcleo, las cuales se especifican como sigue (Bustamante, 2007:53):

1. Membrana exterior: órgano de intercambio con el medio.
2. Citoplasma: cuerpo de la célula y prolongaciones, las cuales son:
 - a) Cilindro eje o axón: prolongación que se divide en,
 - I. Colaterales del axón: órganos efectores (salida) o motores de comunicación con otras neuronas, que a su vez constan de:
 - i. Telodendrias o ramificaciones finales
 - ii. Botón terminal, que es un corpúsculo con relevador químico (neurotransmisor), actuador e inhibidor.
 - b) Dendritas: órganos de recepción divididos en:
 - I. Espinas dendríticas o conductores de señal (el número de estas determina el número de información a recibir).
3. Núcleo: material genético de la célula.

En general, las señales de entrada son captadas por las dendritas y el cuerpo de la neurona, este último las integra y suministra la señal al axón, que está conectado a otro conjunto de dendritas. En esta última conexión se produce la sinapsis (transmisión de la excitación) hacia otras unidades. Por consiguiente las etapas de funcionamiento de la neurona biológica son:

1. Recepción
2. Procesamiento
3. Envío

Tales etapas de funcionamiento fueron adaptadas matemáticamente para una neurona artificial y se organizaron de la siguiente manera:

1. Recepción
2. Suma
3. Activación
4. Salida

En resumen “una neurona es un dispositivo muy simple que transforma (en su soma o cuerpo celular) varias señales de entrada (dendritas) en una única salida (axón)” (De Abajo, 2001:29). Por tanto las neuronas artificiales no son más que dispositivos de cálculo que a partir del procesamiento de un vector de entrada provee una respuesta o salida.

Matemáticamente cada neurona artificial se caracteriza por los siguientes elementos básicos:

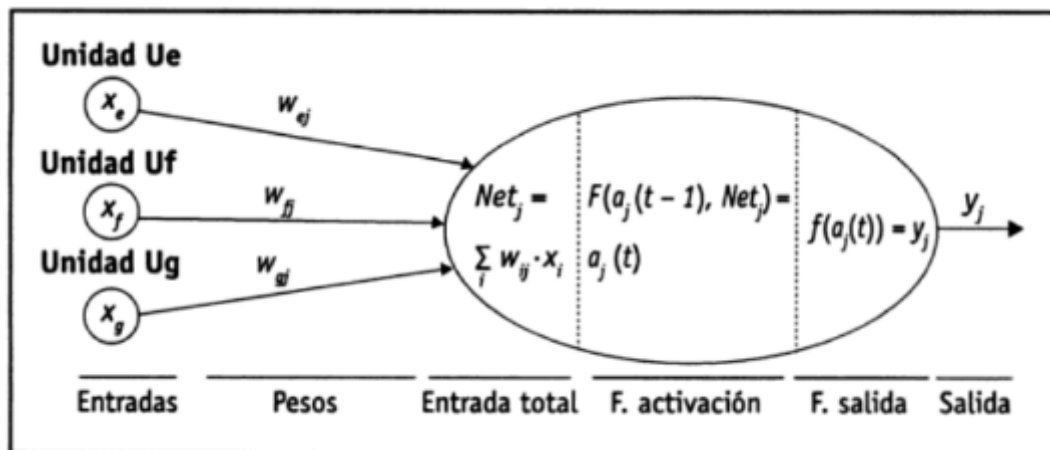
- Un valor o estado de activación inicial (α_{t-1}), anterior a la recepción de los estímulos.
- Unos estímulos o entradas a la neurona (x_i), con unos pesos asociados (w_{ij}).
- Una función de propagación o entrada, que determina la entrada total de la neurona (Net_j).
- Una función de activación o transferencia (f), que combina las entradas a la neurona con el estado de activación inicial para producir un nuevo valor de activación.
- Una función de salida (F), que transforma el estado final de activación en la señal de salida.
- Una señal de salida que se transmite, en su caso, a otras neuronas artificiales (y_j).
- Una regla de aprendizaje, que determina la formación de actualización de los pesos de la red (aprendizaje).

Las funciones de activación o transferencia se estudian en los siguientes párrafos.

A. Función de entrada (input function):

Para iniciar la explicación, es necesario definir el vector de entrada de la red, el cual corresponde al conjunto de x variables $x_i = (x_{i1}, x_{i2}, \dots, x_{in})$ que serán procesadas por las capas subsecuentes ya que influyen sobre la variable dependiente o el dato que se desee obtener.

Figura 6.1 Elementos de una neurona artificial



Fuente: Lévy, Jean-Pierre (2008), p.21

Dos tipos de variables (x) pueden ser codificadas:

- 1) Variables continuas (atributos numéricos): son aquellas que poseen una ordenación intrínseca y pueden tomar cualquier valor en un intervalo dado, es decir entre dos valores cualesquiera puede existir un número infinito de otros valores.
- 2) Variables discretas (atributos simbólicos): son aquellas que sólo pueden tomar valores en específico y no es posible ningún valor entre dos consecutivos.
- 3) Variables pseudodiscretas, refieren a las que poseen ordenación y asumen valores concretos.

La función de entrada es aquella que, precisamente, utiliza como argumento las entradas simples asociadas a su peso w_{ij} correspondiente y bajo la aplicación de operadores específicos emite una única entrada global Net_i , que será procesada por el resto de la neurona. De esta manera, por ejemplo:

- Las variables continuas pueden ser codificadas mediante una función lineal.
- Mientras que las variables discretas pueden ser codificadas mediante neuronas binarias.

El tratamiento que realiza la función se denota de la siguiente manera:

$$Net_i = (x_{i1} * w_{i1}) * (x_{i2} * w_{i2}) \dots (x_{in} * w_{in})$$

Donde * representa al operador apropiado (ejemplo: máximo, sumatoria, productoria, etc.), n al número de entradas a la neurona N_i y w_i al peso.

Así mismo: x_{i1} es la entrada número 1 a la neurona N_i ; w_{i1} es el peso correspondiente a x_{i1} ; mientras que x_{i2} es la entrada número 2 a la N_i , y por tanto w_{i2} es el peso correspondiente a x_{i2} , y así sucesivamente hasta x_{in} , y y_i la salida de la neurona N_i . Algunas de las funciones de entrada son:

a) Sumatoria de las entradas pesadas:

$$\sum_j (x_{ij} w_{ij})$$

Con $j = 1, 2, \dots, n$

b) Productoria de las entradas pesadas:

$$\prod_j (x_{ij} w_{ij})$$

Con $j = 1, 2, \dots, n$

c) Máximo de las entradas pesadas:

$$Max(x_{ij} w_{ij})$$

Con $j = 1, 2, \dots, n$

B. Función de activación

La función de activación calcula el estado de activación de la neurona, el cual se divide en dos, activo (1) e inactivo (0 o -1). La función transforma la entrada global (Net_i menos el umbral Θ_i) en un valor cuyo rango va generalmente de (0 a 1) o de (-1 a 1). En este caso, Θ_i corresponde a un umbral de activación, es decir, una cantidad que refuerza la salida del sumador y que se calcula según la necesidad de la función.

$$f(z) = (Net_i - \Theta_i)$$

A continuación se expresarán las funciones más comunes:

a) Función lineal

$$f(z) = \begin{cases} -1 & z \leq -1/a \\ a * z & -1/a < z < 1/a \\ 1 & z \geq 1/a \end{cases}$$

Con $z = Net_i - \Theta_i$, $a =$ estado de activación inicial, $a > 0$

Los valores obtenidos son:

$a * (Net_i - \Theta_i)$, cuando el argumento de $(Net_i - \Theta_i)$ esté comprendido en el rango $(-1/a, 1/a)$. Cuando $a = 1$, la salida es igual a la entrada.

b) Función sigmoidea:

$$f(z) = \frac{1}{1+e^{-gz}}$$

Con $z = Net_i - \Theta_i$

Los valores de salida de esta función van de 0 a 1. La modificación de g afecta la pendiente de la función de activación.

c) Función tangente hiperbólica

$$f(x) = \frac{e^{gz} - e^{-gz}}{e^{gz} + e^{-gz}}$$

Con $z = gin_i - \Theta_i$

Los valores de salida van de -1 a 1. La modificación de g afecta la pendiente de la función de activación.

Si en el esquema de neurona estándar consideramos que las entradas puedan ser tanto discretas como continuas (analógicas), y las salidas exclusivamente continuas, puede emplearse como función de activación una sigmoide (cuya gráfica tiene forma de letra 'S' inclinada y aplastada) que es una función continua y diferenciable en cierto intervalo, por ejemplo, en el $(-1, +1)$ o en el $(0, +1)$.

Esta función de activación, como se verá más adelante, es utilizada en el perceptrón multicapa. El requisito de trabajar con funciones diferenciables lo puede imponer la regla de aprendizaje, como sucede con la regla de propagación hacia atrás o backpropagation (BP), que también se expondrá en como parte de los capítulos posteriores.

C. Función de salida (output function)

La función de salida F es aquella que obtiene un valor final (y_j) que se transferirá a neuronas vinculadas. Si el resultado de la función de activación se encuentra debajo del

umbral determinado, entonces ninguna salida podrá pasar a la neurona subsecuente. Los valores de salida están comprendidos en el rango:

$$[0,1] \text{ o } [-1,1]$$

También pueden ser binarios:

$$\{0,1\} \text{ o } \{-1,1\}$$

Dos de las funciones de salida más comunes son:

a) Función binaria:

$$\begin{cases} 1 & \text{si } a_i \geq \xi_i \\ 0 & \text{de lo contrario} \end{cases}$$

Donde ξ_i es el umbral de activación

b) Función identidad: la salida es la misma que la entrada.

La función de activación se suele considerar determinista, y en la mayor parte de los casos es monótona creciente y continua, como se observa habitualmente en las neuronas biológicas. La más simple de todas es la función identidad (que se puede generalizar al caso de una función lineal cualquiera), empleada por ejemplo en la Adalina. Otro caso muy simple es la función escalón, empleada en el Perceptrón Simple y la red de Hopfield discreta (redes que se examinarán más adelante). Por otro lado, también es permitido el uso de funciones lineales o funciones lineales saturadas que son de gran sencillez computacional y resultan más adaptables desde el punto de vista biológico pues, como se ha explicado, las neuronas se activan más a mayor excitación, hasta saturarse a la máxima respuesta que se puede proporcionar.

En ocasiones los algoritmos de aprendizaje requieren que la función de activación sea derivable. La más empleada en este sentido es la función de tipo sigmoideo, como la del BP. Otra función clásica es la gaussiana, que es utilizada junto con reglas de activación que involucran el cálculo de cuadrados de distancias, (por ejemplo la euclídea), entre los vectores de entradas y pesos. Estas últimas funciones no serán descritas ya que superan el alcance de estudio del presente trabajo.

Por otro lado, existen tres tipos de neuronas artificiales: unidades de entrada, de salida y unidades ocultas (Lévy, 2008:21).

- Las neuronas de entrada reciben señales desde el entorno, provenientes de sensores o de otros sectores del sistema (como archivos de almacenamiento de patrones de aprendizaje).
- Las neuronas de salida envían su señal directamente fuera del sistema una vez finalizado el tratamiento de la información (salidas de la red).
- Las neuronas ocultas reciben estímulos y emiten salidas dentro del sistema, sin mantener contacto alguno con el exterior. En ellas se lleva a cabo el procesamiento básico de la información, estableciendo la representación interna de ésta.

Otra forma de clasificar las neuronas artificiales en el modelaje de la red es según los valores que pueden tomar.

- a) Neuronas binarias: pueden tomar valores dentro del intervalo $\{0,1\}$ o $\{-1,1\}$.
- b) Neuronas reales: pueden tomar valores dentro del intervalo $[0,1]$ o $[-1,1]$.

Es importante destacar que cada nodo o neurona entabla contacto con neuronas superiores o inferiores, más no con aquellas que pertenecen a su mismo nivel.

6.2 Red neuronal artificial (RNA)

Una red neuronal artificial es una estructura organizativa, compuesta por neuronas (de entrada, de salida u ocultas), que trata de imitar el aprendizaje cerebral caracterizado por la experiencia y la extracción de conocimiento genérico a partir de un conjunto de datos (Lévy, 2008:21). En los siguientes apartados se estudiará su composición, funcionamiento e implementación.

6.2.1 Arquitectura de las redes neuronales artificiales

“La topología o arquitectura de una RNA consiste en la organización o disposición de las neuronas, formando capas o agrupaciones de neuronas más o menos alejadas de la entrada o salida de la red” (Matich, 2001:27); depende de cuatro parámetros principales (Lévy, 2008:29):

- A. El número de capas del sistema.
- B. El número de neuronas por capa.
- C. El grado de conectividad entre las neuronas (transmisión).
- D. El tipo de conexiones neuronales (conexiones hacia adelante, conexiones hacia atrás, conexiones laterales, conexiones autorrecurrentes).

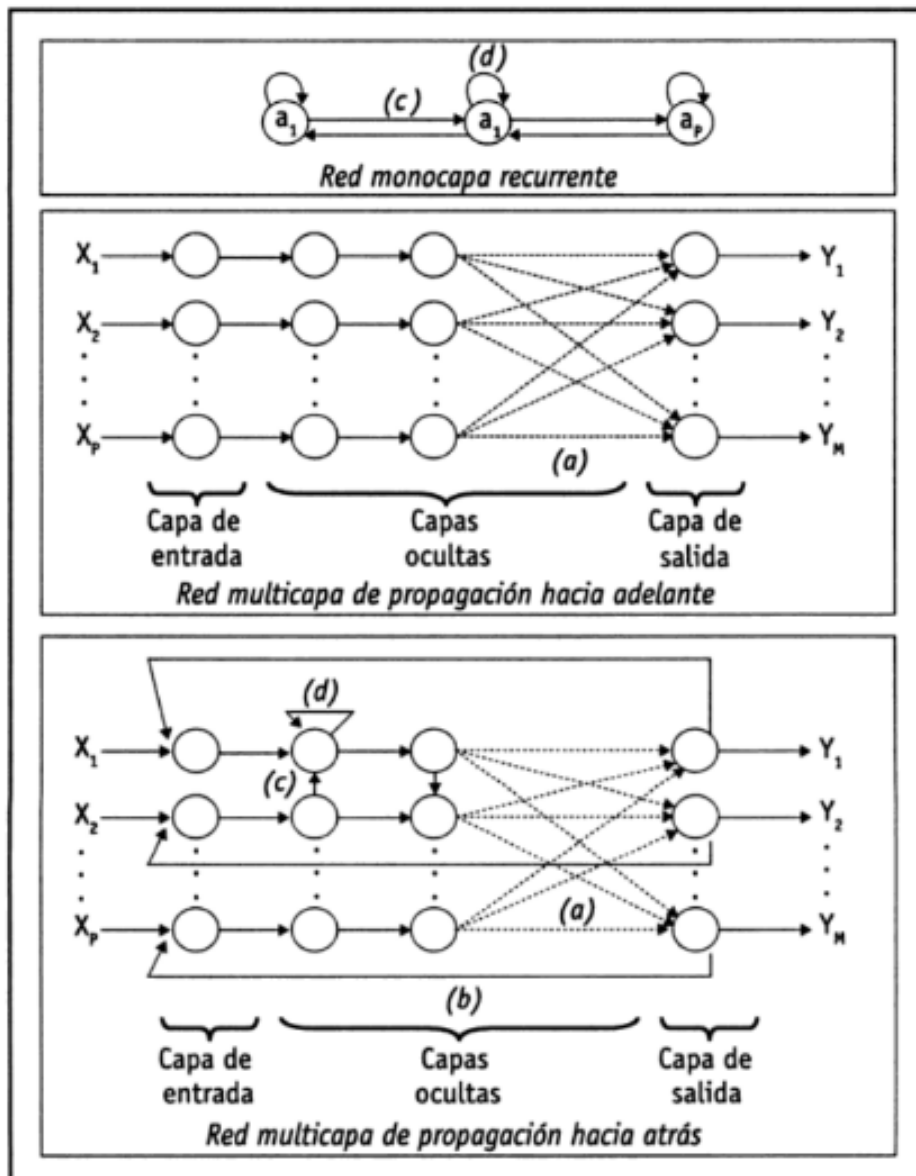
Por su parte la conectividad entre los nodos de la red está relacionada con la forma de transmisión de las salidas de las distintas unidades y su transformación en inputs de otros procesadores.

En este contexto, las arquitecturas neuronales pueden clasificarse atendiendo a distintos criterios:

- I. Según su estructura en capas.
 - 1) Redes monocapa, compuestas por una única capa de neuronas, entre las que se establecen conexiones laterales y, en ocasiones, autorrecurrentes. Este tipo de redes suele utilizarse para la resolución de problemas de autoasociación y clusterización.
 - 2) Redes multicapa (layered networks), cuyas neuronas se organizan en varios niveles o capas (de entrada, oculta(s) y de salida). Las neuronas de una capa, reciben señales de entrada de una capa anterior; así, la capa a la que pertenece la neurona puede distinguirse mediante la observación del origen de las señales que recibe y el destino de la señal que genera.
- II. Según el flujo de datos en la red
 - 1) Redes unidireccionales (feedforward) o de propagación hacia adelante, en las que ninguna salida neuronal es entrada de unidades de la misma capa o de capas precedentes. La información circula en único sentido, desde las neuronas de entrada hacia las neuronas de salida de la red.
 - 2) Redes de propagación hacia atrás (feedback), en las que las salidas de las neuronas pueden servir de entradas a unidades del mismo nivel (conexiones laterales) o de niveles previos. Las redes neuronales de propagación hacia atrás que presentan lazos cerrados se denominan sistemas recurrentes.
 - 3) Redes de propagación autorrecurrente, refiere aquellas en donde la señal que es emitida por la neurona se recibe por la misma. La figura 6.2 ilustra resumidamente algunos de los tipos más comunes de arquitecturas neuronales de acuerdo con las clasificaciones previas.
- III. Según el tipo de respuesta de la red, cabe distinguir entre:
 - 1) Redes heteroasociativas, entrenadas para que, ante la presentación de un determinado patrón A, el sistema responda con otro diferente B. Estas redes precisan al menos de dos capas, una para captar y retener la información de entrada y otra para mantener la salida de la información asociada. Las redes

heteroasociativas pueden clasificarse a su vez según el objetivo pretendido con su utilización, distinguiéndose entre las RNAs destinadas a computar una función matemática a partir de las entradas que reciben, redes utilizadas para tareas de clasificación, redes empleadas para la asociación de patrones, etc.

Figura 6.2 Arquitecturas típicas de RNA's



Nota: (a) conexiones hacia adelante, (b) conexiones hacia atrás; (c) conexiones laterales; (d) conexiones autorrecurrentes.

Fuente: Lévy, Jean-Pierre (2008), p.30.

- 2) Redes autoasociativas, entrenadas para que asocien un patrón consigo mismo, residiendo su interés en que, ante la presentación de un patrón A' afectado por su ruido, su respuesta sea el patrón original A, actuando como un filtro. Estas redes pueden implementarse con una única capa de neuronas, que comenzará reteniendo la información autoasociada; si se desea mantener la información de entrada y salida, deberán añadirse capas adicionales. Estos modelos suelen emplearse en tareas de filtrado de información, para analizar las relaciones de vecindad entre los datos considerados (clustering) y para resolver problemas de optimización.

La tabla resume los modelos más conocidos de redes heteroasociativas y autoasociativas.

Tabla 6.1 Redes neuronales heteroasociativas vs asociativas.

Redes Heteroasociativas		Redes Autoasociativas
Perceptrón	Memoria Asociativa Lineal	Caja de estado cerebral
Adaline	Memoria Asociativa Lineal Óptima	Hopfield
Retropropagación	De manejo reforzado	Memoria Asociativa Lineal Óptima
Crítica Heurística Adaptativa	Retrocontabilización	Grosberg Aditivo
Máquina de Boltzmann	De Memoria Activa Bidireccional (BAM)	Grosberg Cerrado
Máquina Cauchy	BAM Adaptativo	Mapa de características autorganizado
Matriz de Aprendizaje	Cognitiva y Neocognitiva	
Memoria Asociativa Temporal	Teoría de la Resonancia Adaptativa	

Fuente: Lévy, Jean-Pierre (2008), p.31

6.2.2 Mecanismos de aprendizaje y recuerdo

El aprendizaje de una red sucede cuando esta ha sufrido cambios totales o parciales en sus pesos o estructura para proporcionar la respuesta u objetivo deseado.

- 1) Adaptación de pesos: radica en la corrección de los pesos de las interconexiones para obtener resultados específicos.
- 2) Adaptación de la estructura de la red: consiste en la renovación de la topología de la red o de sus diferentes funciones de cada capa o neurona.

Una red aprende cuando “modifica, crea o destruye conexiones entre sus neuronas. La creación implica una nueva conexión con un peso distinto de cero y la destrucción, se produce cuando el peso de la misma pasa a ser cero” (Matich, 2001:18).

En el entorno neuronal, dependiendo la forma en que se produce el ajuste de los parámetros, el modo de aprendizaje puede clasificarse en (Torra, 2013:24):

1. Supervisado: consiste en estructurar un modelo neuronal que permita estimar relaciones entre los inputs y los outputs sin la necesidad de una forma funcional a priori. Si el output no coincide con el deseado se genera un error de salida que se usará para mejorar el proceso de aprendizaje.
2. No supervisado: o auto organizado, es aquel que no dispone de output y supone descubrir patrones y agruparlos según su similitud (clúster) a través de una función de densidad de probabilidad $p(x)$, que describirá los patrones x del espacio R^n de entrada.
3. Reforzado: es una combinación de las características de los aprendizajes anteriores ya que emplea la información del error cometido, pero no cuenta con un output deseado sino con objetivos y señales críticas. El proceso sigue un camino de premio-castigo, reforzándose la acción que mejore el modelo.
4. Híbrido: usa de manera completa el aprendizaje supervisado y no supervisado en distintas capas de neuronas.

Cada aprendizaje se consigue con estrategias específicas:

- Supervisado:
 - a. Por corrección de error: consiste en ajustar los pesos sinápticos de las conexiones de la red en función de la diferencia entre los valores deseados y los obtenidos en la salida de la red, es decir, en función del error cometido. El ajuste de los pesos se realiza con distintos algoritmos dependiendo el caso, entre los cuales se encuentran:
 - ✓ La regla del aprendizaje Delta o regla del mínimo error cuadrado (LMS Error o Least Mean Squared Error) que además de utilizar la

desviación con respecto a la salida objetivo, cuantifica la desviación de neuronas predecesoras, con lo que se adquiere el error global. Este último es igualmente repartido entre las conexiones neuronales para efectuar su corrección.

- ✓ La regla de aprendizaje de propagación hacia atrás (LMS multicapa o backpropagation), la cual es una generalización de la regla anterior y se explicará más adelante.
- ✓ La regla de aprendizaje del Perceptrón que se denota como sigue:

$$\Delta w_{ij} = \eta * y_j * (t_j - y_i)$$

donde: η denota el aprendizaje, t_j es la salida deseada/objetivo de la neurona de salida N_j , y_i es la salida de la neurona N_i y por tanto $\delta_i = (t_j - y_i)$ es la desviación objetivo de la neurona N_i . La salida de la neurona $N_j(y_j)$, se utiliza dado que este valor formará parte de la entrada global y por ende de la activación de la neurona N_i . Las correcciones se completan mediante este “efecto cadena”.

- b. Por instrucción estocástica: gravita en realizar cambios aleatorios en los valores de los pesos de la red y evaluar el efecto a partir del objetivo deseado mediante distribuciones de probabilidad.
- c. Aprendizaje por gradiente: se denomina gradiente en un punto de una función real de varias variables reales al conjunto ordenado de las derivadas parciales de esa función en ese punto. Es decir por cada variación en unidades de las variables independientes, el gradiente informa en un punto determinado cuánto se modifica la función. Así, el gradiente de $f(x, y, z)$ de dimensión tres podría ascender a (a, b, c) donde a, b, c son constantes. Si f es función de una sola variable, su gradiente en cada punto sólo tiene una componente, que es la derivada de f en ese punto o su pendiente. En este caso las neuronas se activan cuando alcanzan el gradiente de la función que se esté utilizando. Otro método reconocido es el gradiente conjugado, que minimiza las funciones cuadráticas convexas; su nombre proviene dado que utiliza vectores de entradas conjugados para la optimización de los resultados (Marsden, 2015:180).

- No supervisado:
 - a. Aprendizaje hebbiano: se consume cuando dos neuronas N_i y N_j alcanzan un mismo estado simultáneamente (ambas activas o ambas inactivas), por lo que el peso de la conexión entre ambas se incrementa. Por tanto, el ajuste de los pesos de las conexiones se debe formular de acuerdo con la correlación de los valores de las dos neuronas conectadas. La entrada o salida se categoriza en: $\{-1,1\}$ o $\{0,1\}$.
 - b. Aprendizaje competitivo: las neuronas compiten por activarse, quedando una por grupo ganador. El objetivo es formar un clúster o clasificación de los datos de entrada en patrones reconocidos anteriormente para proceder a reformar la estructura o los pesos de conexión. El número de categorías es decidido por la propia red.
- Reforzado:
 - a. Aprendizaje por refuerzo: los pesos se ajustan probabilísticamente mediante señales de refuerzo (éxito=+1 o fracaso=-1) indicadas por el supervisor. De esta manera se memorizan las conductas reforzadas positivamente y aquellas que son reforzadas de manera negativa.

Estos modos de aprendizaje suelen ejecutarse bajo un criterio de actividad neuronal determinada:

- a) Aprendizaje on line: la red aprende durante su funcionamiento habitual,
- b) Aprendizaje off line: el proceso de aprendizaje supone la desconexión de la red. Se distinguen dos fases, la fase de aprendizaje o entrenamiento y la fase de operación o funcionamiento; cada una se reforma según el contraste obtenido de un conjunto de datos de entrenamiento y un conjunto de datos de test o prueba. Los pesos de las conexiones permanecen fijos después de que termina el entrenamiento.

Un tipo de aprendizaje supervisado más utilizado es el aprendizaje por retropropagación del error, el cual consiste en el uso de un algoritmo que aplica el error (generado de la diferencia entre la salida real y la deseada) para ajustar los pesos de varias capas de interconexiones. Para explicar su funcionamiento es posible utilizar una neurona de tres capas (capa de entrada, capa intermedia u oculta y capa de salida). Cada una de las neuronas de la capa intermedia está conectada con cada una de las pertenecientes a la capa

de entrada y la capa de salida; no obstante, no se admiten conexiones entre neuronas de la misma capa.

El algoritmo está conformado por dos fases:

- a) Fase de retropropagación hacia adelante: el vector de entrada x_i es transmitido por la capa de entrada hacia cada neurona j de la capa intermedia, las cuales proceden a realizar una suma ponderada S disponiendo de los pesos de sus interconexiones. Consiguientemente, el resultado de la suma es transformada por una función sigmoide $f(S)$, que se caracteriza por ser un elemento activador no lineal. La salida de esta función se constituye como la entrada de cada neurona k en la capa de salida, y es consecuentemente transformada por un activador no lineal que constituye la salida final de la red.
- b) Fase de retropropagación hacia atrás: utiliza el cálculo de error de salida en la capa de salida, para corregir los pesos sinápticos. El resultado se obtiene mediante la siguiente formulación:

$$\delta_j = (t_k - y_k)f'(s_k)$$

Donde:

t_k = la salida deseada de la neurona j

y_k = la salida real de la neurona j

$f'(x)$ = la derivada de la función activadora sigmoideal, $f(x)$ en el punto x

s_k = la suma ponderada de las entradas a la neurona k

Consiguientemente se realiza lo mismo para la capa intermedia mediante el producto resultado de la multiplicación de la suma ponderada de los pesos de interconexión w_{jk} entre la neurona intermedia j y la neurona de salida k , por la derivada f' de la función activadora sigmoideal $f(S_j)$ de la capa intermedia.

$$\delta_j = \left[\sum_k \delta_k w_{jk} \right] f'(S_j)$$

El resultado para modificar los pesos sinápticos se usa cuando este queda cerca de la zona de pendiente máxima de la curva sigmoideal. Esto se logra a partir del cálculo de los gradientes, método descrito anteriormente.

Por consiguiente se deduce el incremento Δw_{ji} a los pesos, de acuerdo con la fórmula (regla delta generalizada):

$$\Delta w_{ji} = \eta \delta_j q_i$$

Donde:

δ_j = el valor de error de la neurona posterior

q_i = el valor de salida real de la neurona anterior

η = la tasa de aprendizaje

Finalmente se repite el procedimiento anterior para la capa de salida, pero usando la función activadora correspondiente a dicha capa.

Todo proceso de corrección de pesos puede llevarse a cabo de dos formas, una de ellas se denomina corrección por lotes (batch), que consiste en una fase de ejecución para todas y cada una de las observaciones del conjunto de entrenamiento; consiguientemente se calcula la variación en los pesos debido a cada error acumulado y sólo entonces se procede a la actualización de los pesos. En otras palabras, si una red neuronal es entrenada d veces, la reasignación de pesos sólo se llevará a cabo en cuánto el número de entrenamientos haya sido cumplido en su totalidad, lo que implica una mayor cantidad de cálculos y almacenamiento. La segunda forma de corrección consiste en actualizar los pesos sinápticos tras la presentación de las observaciones por cada entrenamiento (en vez de presentarlos todos y luego actualizar), esquema denominado aprendizaje en serie (on line).

El aprendizaje en serie es habitualmente empleado en aquellos problemas en los que se dispone de un muy numeroso conjunto de observaciones de entrenamiento (compuesto por cientos o miles de observaciones), en el que habrá mucha redundancia en los datos. Si se emplease en este caso el modo por lotes, el tener que procesar todos los patrones antes de actualizar los pesos demoraría considerablemente el entrenamiento (además de precisar el almacenamiento de numerosos resultados parciales).

En el aprendizaje en serie se debe tener presente que el orden en la presentación de las observaciones debe ser aleatorio, puesto que si siempre se sigue un mismo orden el entrenamiento estaría viciado en favor del último patrón del conjunto de entrenamiento, cuya actualización, por ser la última, siempre predominaría sobre las anteriores.

Además, esta aleatoriedad presenta una importante ventaja, puesto que puede permitir escapar de mínimos locales en determinadas ocasiones, por lo que al final del proceso puede alcanzarse un mínimo más profundo.

El proceso de aprendizaje generalmente se detiene cuando el cálculo de error cuadrado ha alcanzado un mínimo o se encuentra debajo de determinado umbral, así mismo se consigue al haberse cubierto un cierto número de ciclos correctivos, cuando los pesos de

entrenamiento hayan sido totalmente corridos o cuando la *función f* de transformación de los datos de entrada haya sido resuelta. En varias ocasiones el proceso de aprendizaje simplemente termina cuando los valores de la red permanecen estables:

$$dy_{ij}/dt_{ij} \approx 0$$

El cociente entre la diferencia de salidas reales y objetivo deberá aproximarse a cero.

6.2.3 Validación de la Red Neuronal

Reside en comprobar si la red puede resolver eficientemente un determinado y nuevo problema, para lo cual fue entrenada. Para ello se necesita un conjunto de validación o testeo de datos, que se compara con la solución calculada por la red.

6.3 Modelos de RNA's

En lo subsecuente se expondrá el perceptrón simple, la red neuronal multicapa, la red Adaline y la red de Hopfield, que son los modelos más comunes en el mundo de la inteligencia artificial.

6.3.1 Perceptrón simple

Reside en una sola neurona procesadora de información, la cual conserva sus elementos básicos: valores de entrada, función de introducción, función de activación y salida. La función de entrada consiste en la suma ponderada de los pesos con sus respectivas variables de ingreso $X_i = (x_1, x_2, \dots, x_p)$ (que son de carácter binario), y se estipula como sigue:

$$\sum_i W_i X_i + b$$

Donde b es un término independiente. La función activadora radica en una función de escalón umbral, ésta suministrará una única salida y_i dependiendo si la suma ponderada es mayor o igual a un umbral U .

$$f(x) = \begin{cases} 1 & \text{si } \sum_i W_i X_i + b > U \\ 0 & \text{si } \sum_i W_i X_i + b < U \end{cases}$$

La regla de aprendizaje del Perceptrón (en donde el error se genera como diferencia entre una señal tutora y una señal de salida, lineal o no-lineal de la neurona), como se apuntó anteriormente, es el proceso con el cual se entrena este modelo.

En la clasificación de patrones, este modelo es el más utilizado. En este caso se debe disponer de una función de activación tipo signo que denota como (CYTED, 1999:33):

$$y_i = \varphi(V_j) = \text{sgn}\left(\sum_{i=1}^p W_i X_i - \theta_j\right)$$

El vector de patrones sinápticos y_i es igual a una función de activación tipo signo (sgn) de la diferencia entre la suma producto del vector de pesos sinápticos W_i y un valor umbral de disparo θ . Donde V_j se define como

$$V_j = \sum_{\substack{k=i \\ k=j}}^q \cos(a_k, a_j) b_k$$

El coseno del producto de vectores de entrada a_k , a_j y el vector de salida b_k . El propósito del perceptrón es clasificar el conjunto de patrones x_1, x_2, \dots, x_p en dos clases determinadas por las salidas $+1$ y -1 .

La arquitectura de redes neuronales más comunes se compone de varias capas de perceptrones simples.

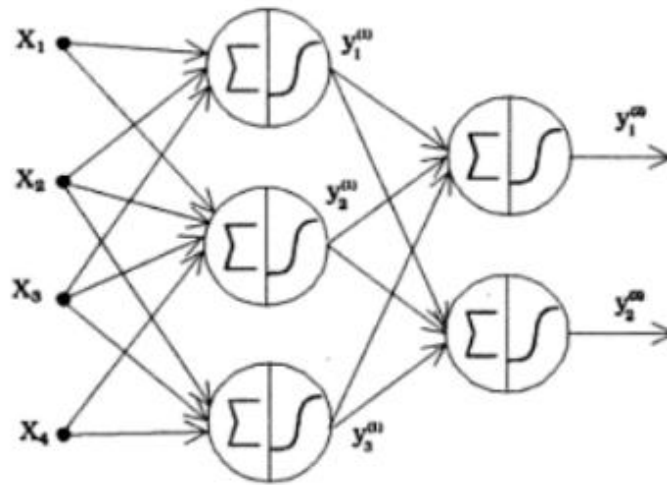
6.3.2 Red neuronal con perceptrones

Los perceptrones multicapa son una extensión del perceptrón simple, al que se le ha agregado una o más etapas similares en cascada. La función que denota su estructura es

$$f(x, v) = \theta\left(\sum_i X_i \cdot \theta\left(\sum_j V_j \cdot (\dots)\right)\right)$$

Donde $X_i = (x_1, x_2, \dots, x_p)$ es el vector de parámetros de entrada. La estructura de la red se muestra en la siguiente figura:

Figura 6.3 Perceptrón multicapa



Fuente: CYTED, 1999, p.34

A pesar de la teoría no existe un procedimiento para determinar el número de variables importantes de la red, tales como número de capas, número de unidades por capa, cantidad de muestras necesarias para entrenamiento, etc.

La red neuronal con preceptrones es el modelo más sencillo de red multicapa (MLP), ya que está conformada por dos capas, una receptora de entrada y una de salida ensamblada a la primera bajo pesos ajustables. La operación de un MLP con una capa oculta y neuronas de salida lineal (estructura que constituye un aproximador universal de funciones) se expresa matemáticamente de la siguiente manera.

$$z_k = \sum_j w'_{kj} y_j - \theta'_i = \sum_j w'_{kj} f \left(\sum_i w_{ji} x_i - \theta_j \right) - \theta'_i$$

Se denomina a x_i como las entradas de la red, y_j como las salidas de la capa oculta y z_k las de la capa final (y globales de la red); t_k a las salidas objetivo (target). Por otro lado, w_{ij} son los pesos de la capa oculta y θ_j sus umbrales, siendo $f(.)$ de tipo sigmoideo, como por ejemplo, las siguientes funciones

$$y = f(x) = \frac{1}{1+e^{-x}} \text{ con } y \in [0,1]$$

$$y = f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} = tgh(x), \text{ con } y \in [-1,1]$$

Esta red es entrenada con iteración de patrones, los cuales son formados por un vector de entrada x y su vector de salida y deseable, siendo la dimensión de cada vector igual al número de neuronas correspondiente a su capa. El reajuste de los pesos w de las interconexiones se lleva a cabo mediante el cálculo de la diferencia entre la salida de la red y el vector de salida ideal, es decir, el aprendizaje de este modelo se lleva a cabo mediante la corrección supervisada del de error. En específico, el reajuste se consigue mediante la Regla Delta de Widrow y Hoff (1960):

$$w_1 = w_0 + \eta(t_i - y_i)$$

El nuevo peso w_1 es igual al peso anterior más una cantidad proporcional a la diferencia entre la salida deseada t_i y la salida real y_i . Donde η es una constante de proporcionalidad menor que la unidad que corresponde a la razón de aprendizaje.

El proceso de aprendizaje también puede llevarse a cabo mediante el algoritmo de retropropagación de errores. La adaptación de los coeficientes de peso de cada una de las capas de una red multicapa debe cubrir los siguientes aspectos: relacionar las variaciones de estos coeficientes con una reducción del error e implementar un algoritmo capaz de recorrer el espacio de estados de la superficie del error, hasta encontrar un vector de pesos óptimo en cada capa de la red (CYTED, 1999:34). La solución, es el algoritmo de retropropagación de errores. Los conceptos para su construcción son similares a los parámetros del perceptrón simple. La ecuación que lo denota es la siguiente:

$$W_{K+1} = W_K + 2\mu\delta_k X_k$$

Donde $\delta_k = \varepsilon_t \cdot sgm'(Y_k)$ refiere la derivada de la función sigmoide (sgm), que nos brinda información de la sensibilidad del error cuadrático a una variación de la salida lineal Y_k .

En general si añadimos capas intermedias a un perceptrón simple, obtendremos un perceptrón multicapa o MLP. Esta arquitectura suele entrenarse mediante el algoritmo de retropropagación de errores o BP, o bien haciendo uso de alguna de sus variantes o derivados, motivo por el que en muchas ocasiones el conjunto arquitectura MLP y aprendizaje BP suele denominarse red de retropropagación, o simplemente BP.

6.3.3 Adaline

Consiste en una neurona lineal adaptativa que calcula la suma ponderada de las entradas y toma en cuenta el error global para su aprendizaje. En seguida se enlistan algunas de sus características.

- 1) Adicional a sus valores de entrada, hace uso de un valor de entrada constante que comúnmente es igual a +1.
- 2) Usa una función de transformación tipo escalón.
- 3) La respuesta de salida es +1 o -1.
- 4) Para el reajuste, el error utilizado consiste en la diferencia entre la salida s_j de la suma ponderada de las entradas y la salida deseada, puesto que el tratamiento de aquella es más sencilla.
- 5) Para el aprendizaje, se vale de la Regla Delta:

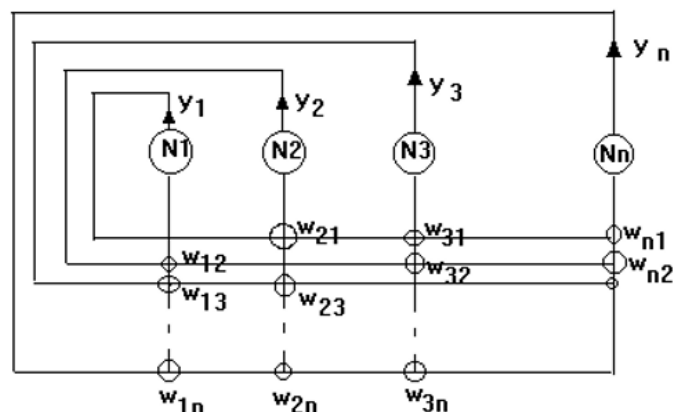
$$\Delta w_{ij} = y_i \eta (t_j - s_j)$$

Donde Δw_{ij} es el incremento en el peso entre la neurona i y la neurona j , y_i es la salida de la neurona i , η es la tasa de aprendizaje, t_j es la salida deseada de la neurona j y s_j la salida del sumador de la neurona j .

6.3.4 Red de Hopfield

Es una red neuronal de carácter recursiva o recurrente, ya que se halla estructurada por una capa con un número simétrico de neuronas conectadas entre sí por arcos ponderados que tienen un sentido bidireccional, lo que permite su retroalimentación.

Figura 6.4 Red de Hopfield



Fuente: Lévy, Jean-Pierre (2008), p.47

La naturaleza de la red es binaria, esto es porque el conjunto de valores de entrada y de salida permitidos son $\{1,0\}$ o en pocas ocasiones $\{1,-1\}$. En este sentido, las salidas finales o estados de activación de la red son activa (1) cuando la suma ponderada de las salidas es mayor que el umbral T , o inactiva (-1 o 0) cuando la suma ponderada es menor que tal umbral.

Posee una memoria asociativa no lineal, dado que almacena patrones incompletos o con ruido y emite patrones totalmente integrados; así mismo es capaz de reproducir toda una señal a partir de una fracción de ella. Es por esta característica que la red de Hopfield es considerada una herramienta de optimización.

Para cumplir como memoria asociativa, la red debe asumir una matriz de pesos formulada de la siguiente manera:

$$W = \sum_j y_i^T y_i$$

Donde y_i es el i ésimo factor a reconocer (memoria), expresado como vector renglón, mientras que $y_i^T y_i$ es la matriz de pesos correspondientes al i ésimo vector a reconocer y la sumatoria abarca todas las matrices de pesos correspondientes a todos los vectores (memorias a reconocer).

El aprendizaje de la red se da de forma secuencial, de una neurona a la vez en forma aleatoria. Como se mencionó anteriormente, si la suma ponderada de entradas al nodo (s_j) es menor o igual que el umbral T , entonces el estado-salida y_j de la neurona j se hace igual a cero:

$$s_j = \sum_i x_i w_{ij}$$

Si la suma es mayor que el umbral T , el estado salida de la neurona j se hace igual a uno. En este sentido, los pesos permanecen fijos y únicamente se cambia el estado de la neurona.

Capítulo 7. Aplicaciones

En este capítulo se pondrá en práctica toda la teoría antes especificada para lograr realizar modelos de pronóstico aplicados a la serie de tiempo financiera correspondiente a la acción AAPL INC propiedad de la empresa multinacional estadounidense Apple Inc.

Las acciones de Apple han inundado el mercado financiero internacional en los últimos años. El alza de su valor ha sido un fuerte impulso en la recuperación de índices tan importantes como el Nasdaq y el Dow; por el contrario, la baja en su cotización ha traído graves consecuencias. Las principales razones por las que las cotizaciones de Apple han declinado, se centran en una caída de las ventas o ingresos de la empresa. En este sentido, en el tercer trimestre fiscal del año 2016 la acción AAPL sufrió una caída del 24%, dadas las condiciones inciertas en que se presentaría el iPhone 7 en su conferencia de desarrolladores, hecho que se ha repetido en años anteriores como en septiembre de 2012, cuando perdió el 40% de su valor ante el lanzamiento del iPhone 5 y el cambio en la directiva de la empresa. Para 2014, las cotizaciones se recuperaron y se posicionaron en los primeros rankings de acciones más valiosas; lo mismo sucedió para el segundo trimestre de 2015. Para este año las acciones han mostrado una tendencia al alza incrementándose hasta cerca del 30% con respecto a 2016. También, para el segundo trimestre fiscal de este año la compañía reportó un crecimiento de sus ingresos en 5%, mientras que su margen bruto se posicionó entre 37.5% con respecto al mismo periodo en 2016, cifras que reflejan un desempeño favorable.

Dadas las razones anteriores, la serie accionaria se puede describir en pos de los sucesos fortuitos que acaparan el estado de la empresa. Por otra parte, las acciones de una empresa dependen en gran medida de la incertidumbre económica y social que se ve reflejada en el consumo, las tasas de interés, el tipo de cambio, la inflación y otras variables macroeconómicas.

Para el año 2016 varios analistas estimaron que los pronósticos para las acciones de AAPL no eran alentadores ya que las ventas de la empresa disminuyeron; no obstante se observó un repunte de las mismas, tendencia que se ha mantenido hasta hoy. De ahí el interés por presentar en el siguiente documento distintos métodos para pronosticar su precio intradía.

Algunas de las características correspondientes a la naturaleza de la empresa son las siguientes: los componentes pertenecientes a los productos de Apple son manufacturados

en EU e Irlanda, así como en Asia por medio de outsourcing. Entre los factores de riesgo se encuentra que la producción puede ser afectada si los proveedores de outsourcing no cumplen su entrega de componentes. Así mismo, todos los productos tratan de obtener propiedad intelectual (patentes), por lo que se requiere un gran desarrollo científico y tecnológico en todos los ámbitos de la producción. El consumidor final es el encargado de probar las innovaciones y estas deberán convencerlo para que pueda continuar su demanda.

Las condiciones operativas y financieras actuales de la compañía varían directa o indirectamente conforme a los hechos materializados en el pasado, así como las expectativas sobre sus resultados operativos y financieros.

Algunos factores que influyen negativamente sobre la empresa son una menor accesibilidad al crédito, alto desempleo, volatilidad en los mercados financieros, programas de austeridad gubernamentales, noticias financieras negativas, decline en el ingreso o el valor de los activos. Estas variables macroeconómicas y de mercado podrían afectar adversamente la demanda de los productos y servicios de la compañía.

Otros factores que podrían influenciar la demanda regional o internacional incluyen cambios en el precio de los combustibles y otros costos energéticos; condiciones variables en el mercado hipotecario; desempleo; costos en laborales; difícil acceso a créditos; baja confianza de los consumidores; cambios en el valor justo de los instrumentos derivados y otros factores que afecten el comportamiento de los consumidores. La compañía invierte en equipos de manufactura, muchos de los cuales han sido respaldados por cierta clase de socios subcontratados y se han hecho prepagos a cierta clase de proveedores asociados con acuerdos de provisión de largo plazo; de invalidarse o fallar estos acuerdos existirían inconvenientes productivos.

El precio de las acciones de la compañía ha experimentado una substancial volatilidad, el ejercicio financiero de la misma es sujeto de los riesgos asociados con cambios en el valor del dólar frente a otras monedas locales. Así mismo, la compañía está expuesta a riesgos de crédito y fluctuaciones en el mercado de valores de inversiones en portafolios. Dada la naturaleza global de su negocio, la compañía tiene ambas, inversión nacional e internacional. Consecuentemente el rango de los créditos y precios para la inversión en la compañía en cada país se pueden ver negativamente afectados por la liquidez, el deterioro del crédito, resultados financieros negativos, riesgo económico, riesgo político, riesgo soberano y otros factores.

Comparado con el retorno del Índice en Información y Tecnología S&P 500 y el Índice del Supersector Tecnológico de EU Dow Jones, el retorno de Apple Inc. ha sido superior en varios años consecutivos, no obstante, estos índices pueden ser tomados en cuenta como sus principales agentes competidores en el mercado accionario.

Mientras que la compañía está expuesta a las fluctuaciones en la tasa de interés globales, el ingreso y gasto de la compañía es más sensible a fluctuaciones en la tasa de interés de los Estados Unidos. Este último país en conjunto con China fueron los únicos que contabilizaron más del 10% en las ventas netas de la compañía en 2016, 2015 y 2014. No hubo un solo cliente que contabilizara más allá del porcentaje mencionado en las ventas netas durante estos años. En periodos anteriores, el mercado Japonés era el de mayor crecimiento de Apple, aunque la reducción del precio del producto para tal mercado en pos de la competencia con otras varias marcas nacionales provocó también el descenso de los ingresos hacia la compañía; no obstante, sigue siendo uno de los principales consumidores hasta el día de hoy.

Con base en estos rasgos, es posible delimitar cuáles son las variables que influyen en las fluctuaciones del precio de la acción y la ponderación intuitiva que recibe cada una de ellas. Los programas estadísticos para llevar a cabo los cálculos correspondientes al modelado fueron dos: el programa Eviews se destinó al estimado ARCH y GARCH, mientras que el programa SPSS se designó a la construcción de la red neuronal artificial.

7.1 Modelo ARCH para predicción

Como ya se mencionó, la serie que se analizó en el presente documento corresponde a los retornos de los precios de la acción de la compañía Apple en Estados Unidos. Los datos de dicha serie se conformaron a partir de las cotizaciones de días hábiles, obtenidos de la plataforma Bloomberg, desde el 14 de enero de 2015 hasta el 15 de febrero del 2017, con lo cual se analizaron aproximadamente 2 años de cotizaciones diarias (500 observaciones) al cierre de la operación.

En consecuencia, se pronosticó el precio de la acción para el siguiente día hábil, 16 de febrero de 2017, y para efectos comparativos sobre la eficiencia del pronóstico se analizó la suma de los errores cuadráticos en cada método utilizado.

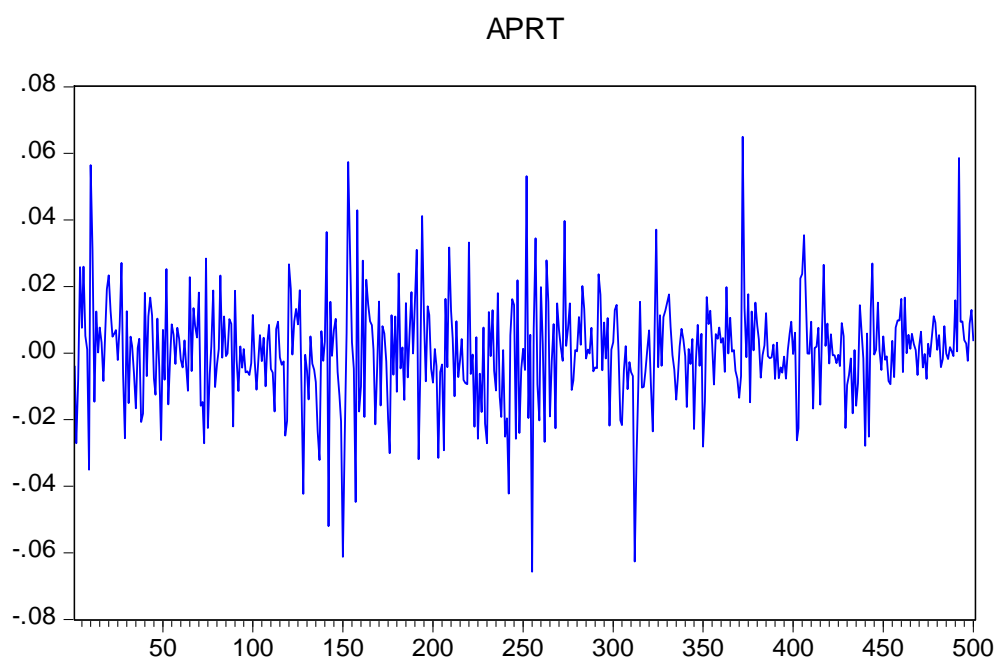
Según la teoría, el retorno de los precios de las acciones posee una volatilidad intrínseca y su distribución de probabilidad se asemeja a una de tipo normal, con excepción de que

presenta colas más anchas dado que los valores atípicos tienen asociadas significativas probabilidades de ocurrencia. Bajo este supuesto es necesario destacar que la prueba de normalidad para los siguientes modelos no se verá cumplida del todo, pero ante pruebas experimentales iterativas es posible que posea un mayor ajuste, según el teorema del límite central. No obstante, en lo que a este trabajo concierne, la tarea iterativa se encomendará para estudios futuros.

7.1.1 Adecuación del modelo ARCH

Antes de llevar a cabo cualquier tipo de cálculo, es necesario realizar un examen estadístico y gráfico de los datos a utilizar. En este caso pudo ser comprobada la forma de distribución de los retornos a partir de una gráfica de líneas que se muestra a continuación.

Gráfica 7.1 Retornos de la acción AAPL (APRT)



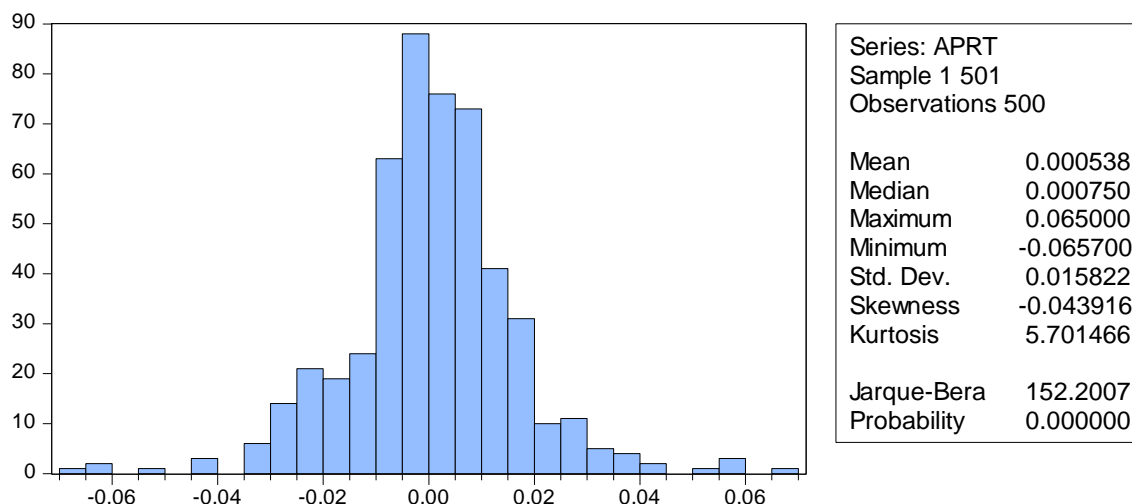
Fuente: Elaboración propia con datos de Bloomberg.

En la ilustración se observa claramente la volatilidad que presenta la serie mediante las bruscas y recurrentes oscilaciones de los datos. La normalidad de los mismos se examinó a partir de la elaboración de un histograma de frecuencias y del cálculo de la curtosis, sesgo y el estadístico Jarque-Bera como prueba de normalidad (Gráfica 7.2).

El histograma de frecuencias muestra una distribución leptocúrtica y sesgada mínimamente hacia la izquierda. El coeficiente Jarque Bera, al ser mayor que 7, acabó por precisar que la

distribución de los datos no es normal, si no que presenta un ajuste más estrecho con respecto a la media.

Gráfica 7.2 Pruebas de normalidad



Fuente: Elaboración propia con datos de Bloomberg.

Como se mencionó capítulos atrás, toda serie que desee ser modelada debe ser estacionaria en sus observaciones. Para confirmar tal requisito se investigó la ausencia de raíces unitarias mediante el test Dickey-Fuller (Tabla 7.1)

Tabla 7.1 Test de Raíz Unitaria Dickey-Fuller

Null Hypothesis: APRT has a unit root

Exogenous: Constant

Lag Length: 0 (Automatic - based on SIC, maxlag=17)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-22.03444	0.0000
Test critical values: 1% level	-3.443228	
5% level	-2.867112	
10% level	-2.569800	

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

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(APRT)

Method: Least Squares

Date: 09/24/17 Time: 15:57

Sample (adjusted): 2 500

Included observations: 499 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
APRT(-1)	-0.988274	0.044851	-22.03444	0.0000
C	0.000540	0.000710	0.761212	0.4469
R-squared	0.494156	Mean dependent var	1.48E-05	
Adjusted R-squared	0.493138	S.D. dependent var	0.022265	
S.E. of regression	0.015851	Akaike info criterion	-5.447151	
Sum squared resid	0.124876	Schwarz criterion	-5.430267	
Log likelihood	1361.064	Hannan-Quinn criter.	-5.440525	
F-statistic	485.5164	Durbin-Watson stat	1.995258	
Prob(F-statistic)	0.000000			

Fuente: Elaboración propia con datos de Bloomberg.

Es necesario recordar que los parámetros pueden ser o no estadísticamente significativos, es decir, existen dos posibles hipótesis:

Hipótesis nula: el parámetro no es estadísticamente significativo.

Hipótesis alterna: el parámetro es estadísticamente significativo.

En términos matemáticos:

$$H_0: \beta_0 = 0$$

$$H_1: \beta_0 \neq 0$$

En este sentido, el p-value permite tomar una decisión sobre la hipótesis nula y observar la fuerza de dicha decisión. Si el p-value es menor que 0.05 hay una evidencia fuerte de que la hipótesis nula es verdadera (el parámetro no es estadísticamente significativo), por el contrario, si p-value es mayor que 0.05 se termina por rechazar la hipótesis nula (el parámetro es estadísticamente significativo).

Regla de decisión:

- La hipótesis nula se acepta si $p - value > 0.05$
- La hipótesis nula se rechaza si $p - value < 0.05$

Para el presente modelo no se confirmó la existencia de una raíz unitaria como parámetro de la serie, ya que el p-value asociado no rebasa el 5%. Además el coeficiente t-calculado (1%, 5% o 10%) es mayor que el t-crítico de la tabla de MacKinnon. Por otro lado, el intercepto tampoco resultó estadísticamente significativo, esto es indicativo de que la aleatoriedad de la serie no se mueve alrededor de una constante proporcional en el tiempo.

Tabla 7.2 Correlograma de los retornos.

Date: 09/24/17 Time: 18:51
Sample: 1 501

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.052	0.052	1.3763	0.241
		2	0.010	0.007	1.4269	0.490
		3	0.003	0.002	1.4312	0.698
		4	-0.014	-0.014	1.5292	0.821
		5	0.003	0.004	1.5333	0.909
		6	0.054	0.054	2.9905	0.810
		7	0.098	0.093	7.8849	0.343
		8	0.028	0.018	8.2981	0.405
		9	0.089	0.086	12.367	0.193
		10	0.021	0.014	12.582	0.248
		11	-0.006	-0.006	12.599	0.320
		12	0.025	0.023	12.916	0.375
		13	0.102	0.095	18.329	0.145
		14	0.009	-0.011	18.366	0.191
		15	-0.028	-0.042	18.762	0.225
		16	-0.022	-0.037	19.009	0.268
		17	-0.002	-0.002	19.010	0.328
		18	-0.051	-0.063	20.384	0.312
		19	-0.000	-0.014	20.384	0.372
		20	0.059	0.042	22.232	0.328
		21	-0.016	-0.026	22.369	0.379
		22	-0.046	-0.059	23.460	0.376
		23	0.012	0.022	23.535	0.430
		24	-0.006	0.009	23.557	0.487
		25	-0.049	-0.039	24.807	0.473
		26	0.043	0.037	25.784	0.475
		27	-0.012	-0.008	25.856	0.527
		28	-0.029	-0.013	26.311	0.556
		29	-0.020	-0.016	26.528	0.597
		30	-0.001	0.007	26.528	0.648
		31	-0.058	-0.037	28.311	0.605
		32	0.119	0.129	35.906	0.290
		33	0.004	-0.026	35.915	0.333
		34	0.013	0.028	36.006	0.375
		35	-0.062	-0.056	38.068	0.332
		36	-0.004	0.008	38.076	0.375

Fuente: Elaboración propia con datos de Bloomberg.

Una serie es estacionaria si todos los momentos de corte transversal de su distribución, el promedio, la varianza y otros, no dependen del tiempo, y que además las correlaciones a

través del tiempo no cambian. Dados los resultados anteriores la serie puede ser calificada como estacional y no será necesario aplicarle ningún método de ajuste como la diferenciación.

El correlograma de probabilidades Q-statistics comprueba los argumentos anteriores, dado que la probabilidad de que los datos sean netamente aleatorios y no estén correlacionados con el tiempo es alta en todos los casos. El 'p-value' de las pruebas rebasa el 0.5% (Tabla 7.2).

Una vez comprobado que la serie es estacionaria, se puede modelar mediante una regresión que la considere como variable dependiente y como variables independientes aquellas que sean de impacto o influencia sobre esta. Concluido lo anterior, se verificará que la heteroscedasticidad haya sido correctamente ajustada mediante un test LM ARCH o correlograma de los residuos al cuadrado; así mismo que la suma de los coeficientes en la regresión sea menor que uno y que el modelo esté correctamente especificado.

Las variables que se utilizaron para llevar a cabo la regresión fueron las siguientes

- El Índice Dow Jones de Rendimiento Total de Tecnología Estadounidense (US Technology Total Return Index) cuyo símbolo es DJUSTCT, uno de los instrumentos que compiten con la acción AAPL en el mercado.
- La paridad USDCNY, dólar estadounidense vs yuan chino (por sus siglas en inglés United State Dollar vs Chinese Yuan), dado que China representó, junto con los EU la mayor participación porcentual en ventas.
- El dólar estadounidense vs yen Japonés (United State Dollar vs Japanese Yen) o USDJPY, por figurar como una de las principales paridades con las que se comercializaron los productos de Apple y que sigue siendo predominante en la actualidad.
- El Índice de Volatilidad VIX (Chicago Board Options Exchange Market Volatility Index), puesto que refleja las altas fluctuaciones del mercado que afectan la estabilidad de la acción AAPL.
- El Índice de Tecnología Dow Jones, W1TEC (Dow Jones Global Technology Index); otro de los principales instrumentos competidores de la acción-
- Y finalmente la el Índice de oro y plata (Philadelphia Gold and Silver Index) o XAU, dado que los componentes más costosos incluidos en los productos devienen de tales materiales.

La regresión que se obtuvo fue la siguiente:

$$\begin{aligned} \text{ARCH} &= 0.000176211099513 + 0.325200523607*(-)\text{RESID}^2 \\ \text{APRT} &= 0.000243668706171 * \text{DJUSTCT} - 0.0219205667189 * \text{USDCNY} + \\ &0.000594904946491 * \text{USDJPY} - 0.000729461119113 * \text{VIX} - \\ &0.000476527393931 * \text{W1TEC} + 7.15951614063e - 05 * \text{XAU} \end{aligned}$$

En seguida se detalla la salida de estimación del modelo.

Tabla 7.3 Estimación ARCH

Dependent Variable: APRT

Method: ML - ARCH

Sample (adjusted): 1 500

Included observations: 500 after adjustments

Convergence achieved after 28 iterations

Presample variance: backcast (parameter = 0.7)

GARCH = C(7) + C(8)*RESID(-1)^2

Variable	Coefficient	Std. Error	z-Statistic	Prob.
DJUSTCT	0.000244	5.10E-05	4.778797	0.0000
USDCNY	-0.021921	0.005297	-4.138410	0.0000
USDJPY	0.000595	0.000121	4.930808	0.0000
VIX	-0.000729	0.000211	-3.453796	0.0006
W1TEC	-0.000477	0.000100	-4.753487	0.0000
XAU	7.16E-05	1.43E-05	5.011652	0.0000
C	0.000176	1.24E-05	14.17234	0.0000
RESID(-1)^2	0.325201	0.068011	4.781593	0.0000
R-squared	0.034029	Mean dependent var		0.000538
Adjusted R-squared	0.024252	S.D. dependent var		0.015822
S.E. of regression	0.015629	Akaike info criterion		-5.508072
Sum squared resid	0.120661	Schwarz criterion		-5.440638
Log likelihood	1385.018	Hannan-Quinn criter.		-5.481611
Durbin-Watson stat	1.895230			

Fuente: Elaboración propia con datos de Bloomberg.

Las variables significativas fueron el Índice Dow Jones de Rendimiento Total de Tecnología Estadounidense cuyo símbolo es DJUSTCT y cuyo coeficiente fue de 0.000244, lo que implica que un aumento de una unidad en el índice modificaría el precio de AAPL en 0.000244 dólares; la paridad USDCNY, dólar estadounidense vs yuan chino que obtuvo un coeficiente de -0.02192 lo que indica que por cada unidad que se aprecie la paridad, la cotización de AAPL disminuirá en 0.02192 dólares; así mismo la apreciación en una unidad de la paridad USDJPY, dólar estadounidense vs yen Japonés, provocaría un incremento de 0.000595 dólares en el precio de la acción; el Índice de Volatilidad VIX disminuye el precio de AAPL en 0.000729 dólares al aumentar una unidad; una disminución de -0.000477 también podría suscitarse ante el aumento del Índice de Tecnología Dow Jones, W1TEC; y finalmente la apreciación del Índice de oro y plata (Philadelphia Gold and Silver Index) o XAU, en una unidad aumentaría el precio de la acción en la mínima cantidad de 0.00000716 dólares.

El coeficiente R cuadrado ascendió a 0.0344 unidades, lo que indica que el modelo no es estadísticamente significativo; no obstante es común que en este tipo de modelos se presenten coeficientes de tal magnitud, dado que es expresamente difícil establecer las variables cuya alteración pueda provocar movimientos significativos en la variable dependiente.

7.1.2 Validación del modelo ARCH

Como ya se indicó, el coeficiente de correlación (R-squared) ascendió a 0.34%, lo que determina que únicamente el 0.34% de las variaciones en los retornos se explican por modificaciones en las variables independientes. En otras palabras únicamente el 0.34% de los datos están ajustados a la línea de regresión. Así mismo, el error estándar de la regresión (S.E. squared resid) indica que la distancia que existe entre los valores reales y la línea de regresión estimada es de 0.15 unidades porcentuales, lo que revela que la bondad de ajuste no es elevada.

El estadístico Durbin-Watson ascendió a 1.89, lo que indica que el modelo no presenta autocorrelación. Esto se verificó mediante un correlograma de probabilidades Q-statistics:

Tabla 7.4 Correlograma Q-statistics

Date: 09/24/17 Time: 18:51
 Sample: 1 501
 Included observations: 500

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*	
		1	0.021	0.021	0.2164	0.642
		2	0.013	0.012	0.2992	0.861
		3	0.008	0.007	0.3308	0.954
		4	-0.015	-0.016	0.4513	0.978
		5	-0.021	-0.021	0.6754	0.984
		6	0.041	0.042	1.5243	0.958
		7	0.081	0.080	4.8752	0.675
		8	0.020	0.016	5.0744	0.750
		9	0.072	0.068	7.7047	0.564
		10	0.007	0.004	7.7304	0.655
		11	0.003	0.005	7.7353	0.737
		12	0.004	0.005	7.7424	0.805
		13	0.094	0.091	12.261	0.506
		14	-0.001	-0.009	12.261	0.585
		15	-0.021	-0.031	12.481	0.642
		16	-0.007	-0.019	12.504	0.709
		17	-0.005	-0.004	12.516	0.768
		18	-0.032	-0.034	13.049	0.789
		19	0.002	-0.007	13.051	0.836
		20	0.059	0.044	14.861	0.784
		21	0.005	0.001	14.872	0.829
		22	-0.041	-0.053	15.763	0.827
		23	0.012	0.015	15.840	0.862
		24	0.003	0.013	15.845	0.894
		25	-0.060	-0.052	17.721	0.854
		26	0.058	0.050	19.510	0.814
		27	0.000	-0.003	19.510	0.851
		28	-0.013	-0.007	19.594	0.879
		29	-0.008	-0.010	19.628	0.904
		30	-0.003	-0.003	19.633	0.926
		31	-0.073	-0.058	22.490	0.867
		32	0.116	0.125	29.714	0.583
		33	-0.023	-0.047	30.005	0.617
		34	-0.004	0.002	30.013	0.663
		35	-0.056	-0.057	31.699	0.628
		36	-0.012	-0.007	31.781	0.670

*Probabilities may not be valid for this equation specification.

Fuente: Elaboración propia con datos de Bloomberg.

La serie no presenta correlación serial dado que las probabilidades asociadas a los coeficientes de correlación son mayores al 5%. La prueba de heteroscedasticidad también se llevó a cabo mediante la observación del correlograma de residuos al cuadrado (Tabla 7.5).

Tabla 7.5 Correlograma de residuos al cuadrado

Date: 09/24/17 Time: 18:51

Sample: 1 501

Included observations: 500

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*	
		1	0.005	0.005	0.0113	0.915
		2	-0.034	-0.034	0.5841	0.747
		3	0.036	0.037	1.2475	0.742
		4	0.019	0.017	1.4299	0.839
		5	-0.026	-0.024	1.7814	0.878
		6	-0.042	-0.042	2.6987	0.846
		7	0.024	0.022	3.0011	0.885
		8	0.013	0.012	3.0919	0.928
		9	0.050	0.056	4.3941	0.884
		10	0.028	0.028	4.8024	0.904
		11	0.027	0.027	5.1765	0.922
		12	0.041	0.037	6.0204	0.915
		13	0.015	0.015	6.1312	0.941
		14	0.038	0.041	6.8725	0.940
		15	0.048	0.051	8.0522	0.922
		16	-0.004	-0.002	8.0604	0.947
		17	-0.032	-0.031	8.6054	0.952
		18	0.046	0.042	9.7291	0.940
		19	0.008	0.002	9.7614	0.959
		20	-0.008	-0.002	9.7985	0.972
		21	0.006	0.000	9.8153	0.981
		22	0.028	0.017	10.242	0.984
		23	-0.048	-0.057	11.455	0.978
		24	-0.015	-0.018	11.576	0.984
		25	-0.019	-0.032	11.766	0.988
		26	-0.061	-0.064	13.706	0.977
		27	0.013	0.007	13.794	0.983
		28	-0.003	-0.011	13.799	0.989
		29	-0.002	-0.007	13.802	0.992
		30	0.050	0.046	15.148	0.989
		31	0.029	0.025	15.613	0.990
		32	-0.023	-0.019	15.897	0.992
		33	-0.031	-0.029	16.412	0.993
		34	0.055	0.056	18.032	0.989
		35	0.020	0.038	18.253	0.991
		36	-0.003	0.014	18.258	0.994

*Probabilities may not be valid for this equation specification.

Fuente: Elaboración propia con datos de Bloomberg.

La serie es homoscedástica dado que los coeficientes presentan un p-value mayor al 5%.

Esta prueba también se verificó mediante el Test LM ARCH (Tabla 7.6).

Tabla 7.6 Test LM ARCH

Heteroskedasticity Test: ARCH

F-statistic	0.424814	Prob. F(1,497)	0.5148
Obs*R-squared	0.426159	Prob. Chi-Square(1)	0.5139

Test Equation:

Dependent Variable: WGT_RESID^2

Method: Least Squares

Date: 09/24/17 Time: 18:18

Sample (adjusted): 2 500

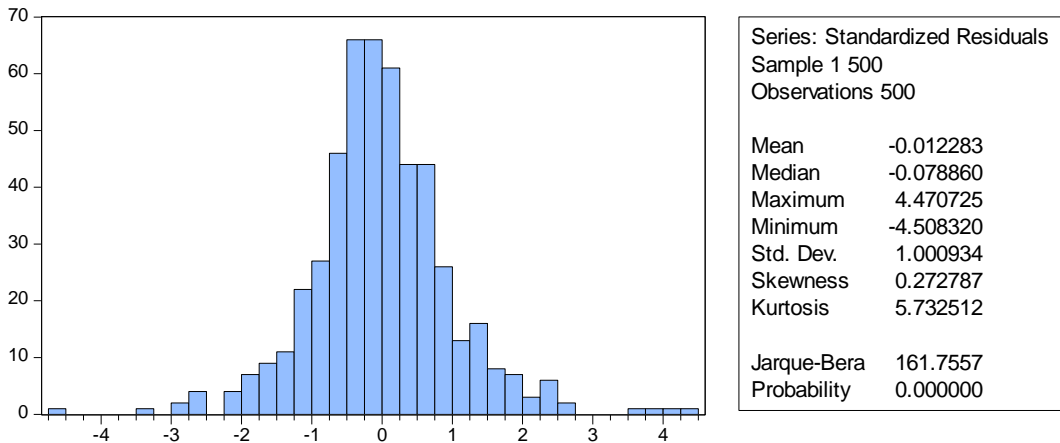
Included observations: 499 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.031284	0.107318	9.609613	0.0000
WGT_RESID^2(-1)	-0.029221	0.044833	-0.651778	0.5148
R-squared	0.000854	Mean dependent var	1.002016	
Adjusted R-squared	-0.001156	S.D. dependent var	2.176096	
S.E. of regression	2.177354	Akaike info criterion	4.398098	
Sum squared resid	2356.212	Schwarz criterion	4.414982	
Log likelihood	-1095.325	Hannan-Quinn criter.	4.404724	
F-statistic	0.424814	Durbin-Watson stat	1.995642	
Prob(F-statistic)	0.514846			

Fuente: Elaboración propia con datos de Bloomberg.

Dado que el p-value de la prueba es mayor que 0.05 se asume que el modelo no posee heteroscedasticidad. El fenómeno de normalidad se observó mediante siguiente gráfica 7.3. Dado que el coeficiente de asimetría es de 0.2727 los residuos se comportan con una mínima desviación positiva en su distribución; así mismo el coeficiente de kurtosis de 5.73 indica que se comportan de manera leptocúrtica. El coeficiente Jarque Bera rebasó los 7 puntos, lo que revela que la distribución no es normal.

Gráfica 7.3 Normalidad de los residuos del modelo

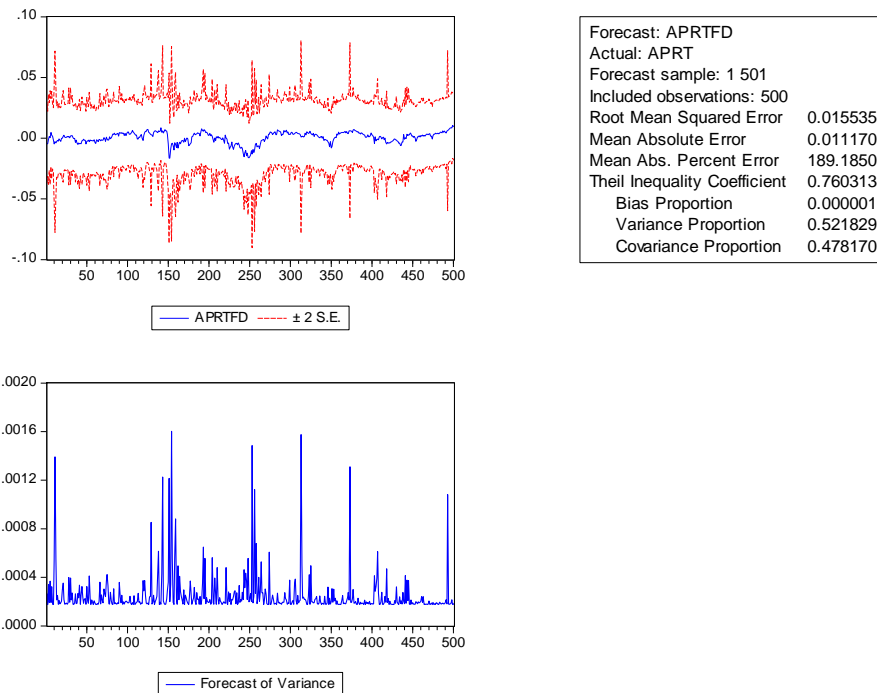


Fuente: Elaboración propia con datos de Bloomberg

7.1.3 Pronóstico con modelo ARCH

En la determinación del modelo se erogaron los siguientes errores:

Gráficas 7.4 Residuos al cuadrado y varianza.

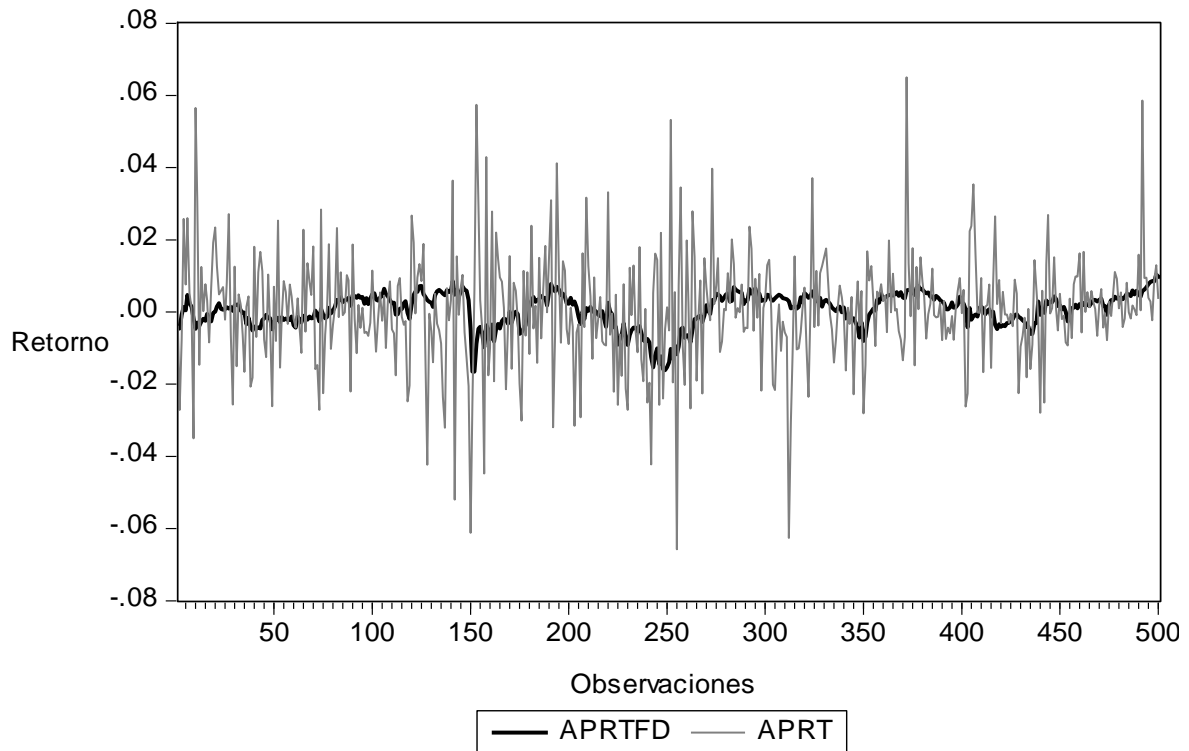


Fuente: Elaboración propia con datos de Bloomberg

La raíz de los errores cuadráticos medios ascendió a 0.015535 unidades. La media del error absoluto fue de 0.01117 unidades y la suma de los errores cuadráticos fue de 0.12066. Con base a este último criterio se comparará la eficiencia de los modelos estimados, ya que es el único que también se estima automáticamente en el modelaje SPSS de redes neuronales.

En la gráfica 7.4 es posible observar el comportamiento de los residuos al cuadrado y de la varianza; mientras tanto, en la gráfica 7.5 se expone la conducta de los datos reales (línea gris delgada) y los datos ajustados (línea negra gruesa al fondo), los cuales manifiestan una importante brecha de diferencia.

Gráfica 7.5 Datos pronosticados vs datos reales



Fuente: Elaboración propia con datos de Bloomberg.

La estimación del retorno para el día 16 de febrero de 2017 fue de .0094 o lo que es lo mismo 0.94%, mientras que el retorno real se posicionó en -0.0012 o -0.12%.

7.2 Modelo GARCH para predicción

En los siguientes apartados se ejecutará y verificará el modelo GARCH siguiendo los lineamientos y pruebas antes mencionados.

7.2.1 Adecuación del modelo GARCH

El modelo GARCH desarrollado en Eviews obtuvo la siguiente regresión:

$$\text{GARCH} = 2.58159037151e - 05 + 0.148717629407 * \text{RESID}(-1)^2 + 0.751777879959 * \text{GARCH}(-1)$$

$$\begin{aligned} \text{APRT} = & 0.000230635625471 * \text{DJUSTCT} - 0.0195424646871 * \text{USDCNY} + \\ & 0.000535272082226 * \text{USDJPY} - 0.000794419809064 * \text{VIX} - \\ & 0.000456099599825 * \text{W1TEC} + 6.81071293303e - 05 * \text{XAU} \end{aligned}$$

La tabla de resultados fue la siguiente:

Tabla 7.7 Estimación GARCH

Dependent Variable: APRT

Method: ML - ARCH

Sample (adjusted): 1 500

Included observations: 500 after adjustments

Convergence achieved after 21 iterations

Presample variance: backcast (parameter = 0.7)

GARCH = C(7) + C(8)*RESID(-1)^2 + C(9)*GARCH(-1)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
DJUSTCT	0.000231	5.55E-05	4.154694	0.0000
USDCNY	-0.019542	0.006861	-2.848489	0.0044
USDJPY	0.000535	0.000126	4.232185	0.0000
VIX	-0.000794	0.000249	-3.194990	0.0014
W1TEC	-0.000456	0.000102	-4.489737	0.0000
XAU	6.81E-05	1.64E-05	4.153331	0.0000
Variance Equation				
C	2.58E-05	7.50E-06	3.442822	0.0006
RESID(-1)^2	0.148718	0.039545	3.760691	0.0002
GARCH(-1)	0.751778	0.056183	13.38085	0.0000
R-squared	0.032725	Mean dependent var	0.000538	
Adjusted R-squared	0.022935	S.D. dependent var	0.015822	
S.E. of regression	0.015639	Akaike info criterion	-5.542173	
Sum squared resid	0.120824	Schwarz criterion	-5.466310	
Log likelihood	1394.543	Hannan-Quinn criter.	-5.512405	
Durbin-Watson stat	1.894955			

Fuente: Elaboración propia con datos de Bloomberg.

Las variables significativas fueron el Índice Dow Jones de Rendimiento Total de Tecnología Estadounidense o DJUSTCT cuyo coeficiente ascendió a 0.000231, lo que implica que un aumento de una unidad en el índice modificaría el precio de AAPL en 0.000231 dólares; la paridad USDCNY, dólar estadounidense vs yuan chino que obtuvo un coeficiente de -0.01954 lo que indica que por cada unidad que se aprecie la paridad, la cotización de AAPL disminuirá en 0.01954 dólares; así mismo la apreciación en una unidad de la paridad USDJPY, dólar estadounidense vs yen Japonés provocaría un incremento de 0.000535 dólares en el precio de la acción; el Índice de Volatilidad VIX disminuye el precio de AAPL en 0.000794 dólares al aumentar una unidad; una disminución de -0.000456 también podría suscitarse ante el aumento del Índice de Tecnología Dow Jones, W1TEC; y finalmente la apreciación del Índice de oro y plata (Philadelphia Gold and Silver Index) en una unidad aumentaría el precio de la acción en la mínima cantidad de 0.00000681 unidades. Como se puede observar, los resultados son muy parecidos a los obtenidos por el modelo ARCH.

7.2.2 Validación del modelo GARCH

Como en el caso preliminar se verificará que el modelo no posea autocorrelación, sea homoscedástico y presente normalidad. Para esto se observaron una vez más los correlogramas y sus pruebas estadísticas complementarias.

Tabla 7.8 Correlograma Q-statistics

Date: 09/24/17 Time: 18:51
 Sample: 1 501
 Included observations: 500

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*	
		1	0.041	0.041	0.8630	0.353
		2	0.008	0.006	0.8969	0.639
		3	-0.018	-0.019	1.0675	0.785
		4	-0.006	-0.005	1.0860	0.896
		5	-0.017	-0.016	1.2257	0.942
		6	0.041	0.043	2.1007	0.910
		7	0.072	0.069	4.7339	0.692
		8	0.032	0.025	5.2510	0.730
		9	0.072	0.070	7.8656	0.548
		10	0.014	0.011	7.9636	0.632
		11	0.007	0.009	7.9912	0.714
		12	-0.002	0.001	7.9925	0.786
		13	0.096	0.093	12.699	0.471
		14	0.018	0.008	12.867	0.537
		15	-0.031	-0.042	13.364	0.574
		16	0.005	-0.000	13.376	0.645
		17	-0.005	-0.009	13.388	0.710
		18	-0.036	-0.041	14.054	0.726
		19	0.008	0.001	14.085	0.779
		20	0.051	0.035	15.438	0.751
		21	0.008	-0.001	15.474	0.798
		22	-0.039	-0.051	16.271	0.802
		23	0.021	0.024	16.505	0.833
		24	0.001	0.008	16.505	0.869
		25	-0.044	-0.040	17.514	0.862
		26	0.059	0.056	19.381	0.820
		27	-0.006	-0.013	19.398	0.855
		28	-0.021	-0.016	19.634	0.877
		29	-0.023	-0.022	19.926	0.895
		30	0.005	0.004	19.940	0.918
		31	-0.067	-0.054	22.352	0.872
		32	0.128	0.134	31.121	0.511
		33	0.007	-0.019	31.148	0.560
		34	0.014	0.013	31.259	0.603
		35	-0.053	-0.043	32.751	0.577
		36	-0.021	-0.018	33.001	0.612

*Probabilities may not be valid for this equation specification.

Fuente: Elaboración propia con datos de Bloomberg.

El correlograma Q-Statistics (Tabla 7.8) demostró que el modelo no posee autocorrelación, dado que sus coeficientes se posicionan más allá del 5%.

Tabla 7.9 Correlograma de residuos al cuadrado.

Date: 09/24/17 Time: 18:51
 Sample: 1 501
 Included observations: 500

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob*	
		1	-0.009	-0.009	0.0442	0.833
		2	-0.024	-0.024	0.3235	0.851
		3	0.006	0.006	0.3447	0.951
		4	-0.038	-0.038	1.0640	0.900
		5	-0.016	-0.017	1.1949	0.945
		6	-0.044	-0.046	2.1803	0.902
		7	-0.030	-0.031	2.6239	0.917
		8	0.055	0.051	4.1872	0.840
		9	-0.002	-0.003	4.1889	0.899
		10	0.040	0.040	5.0250	0.890
		11	0.026	0.023	5.3769	0.912
		12	-0.012	-0.009	5.4528	0.941
		13	-0.009	-0.010	5.4986	0.963
		14	0.021	0.027	5.7331	0.973
		15	-0.034	-0.028	6.3427	0.973
		16	0.009	0.010	6.3804	0.983
		17	-0.014	-0.012	6.4847	0.989
		18	-0.009	-0.011	6.5265	0.994
		19	-0.028	-0.035	6.9380	0.995
		20	-0.022	-0.022	7.1891	0.996
		21	0.011	0.006	7.2507	0.998
		22	0.030	0.025	7.7100	0.998
		23	-0.009	-0.006	7.7510	0.999
		24	-0.033	-0.039	8.3208	0.999
		25	-0.003	-0.006	8.3271	0.999
		26	-0.037	-0.038	9.0350	0.999
		27	-0.008	-0.005	9.0669	1.000
		28	-0.050	-0.050	10.383	0.999
		29	-0.006	-0.004	10.403	0.999
		30	0.030	0.020	10.897	0.999
		31	-0.001	-0.003	10.898	1.000
		32	0.014	0.008	11.010	1.000
		33	0.014	0.010	11.118	1.000
		34	-0.014	-0.009	11.221	1.000
		35	0.000	-0.000	11.221	1.000
		36	-0.050	-0.043	12.572	1.000

*Probabilities may not be valid for this equation specification.

Fuente: Elaboración propia con datos de Bloomberg

Así mismo, los coeficientes correspondientes al correlograma de residuos al cuadrado se ubicaron por encima del 5%, lo que indica que el modelo es homoscedástico.

Tabla 7.10 Test LM-ARCH

Heteroskedasticity Test: ARCH

F-statistic	0.043710	Prob. F(1,497)	0.8345
Obs*R-squared	0.043882	Prob. Chi-Square(1)	0.8341

Test Equation:

Dependent Variable: WGT_RESID^2

Method: Least Squares

Date: 09/24/17 Time: 19:03

Sample (adjusted): 2 500

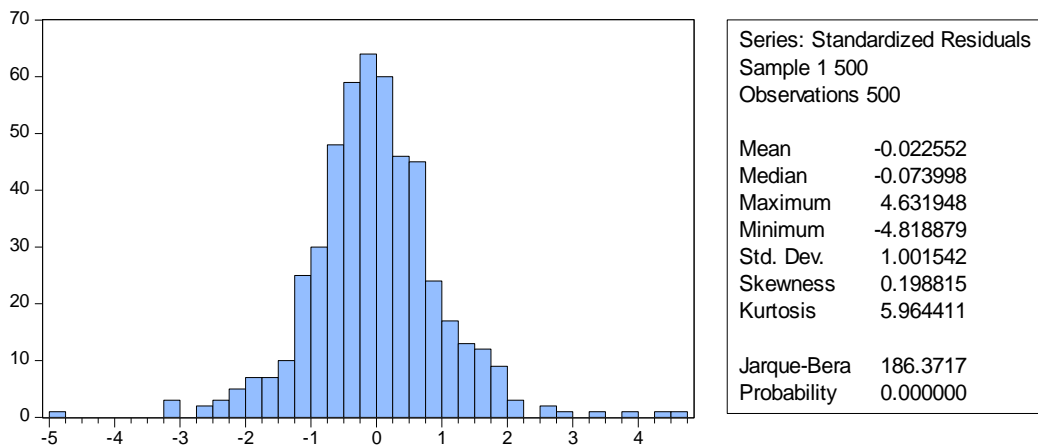
Included observations: 499 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.012987	0.109626	9.240352	0.0000
WGT_RESID^2(-1)	-0.009377	0.044851	-0.209070	0.8345
R-squared	0.000088	Mean dependent var	1.003580	
Adjusted R-squared	-0.001924	S.D. dependent var	2.230952	
S.E. of regression	2.233097	Akaike info criterion	4.448656	
Sum squared resid	2478.401	Schwarz criterion	4.465540	
Log likelihood	-1107.940	Hannan-Quinn criter.	4.455282	
F-statistic	0.043710	Durbin-Watson stat	1.998925	
Prob(F-statistic)	0.834479			

Fuente: Elaboración propia con datos de Bloomberg

La probabilidad del estadístico F en la prueba LM ARCH fue de 83.45%, por lo que se rechaza la hipótesis nula de heteroscedasticidad. El modelo posee varianza constante en el tiempo.

Gráfica 7.6 Normalidad de los residuos de la regresión



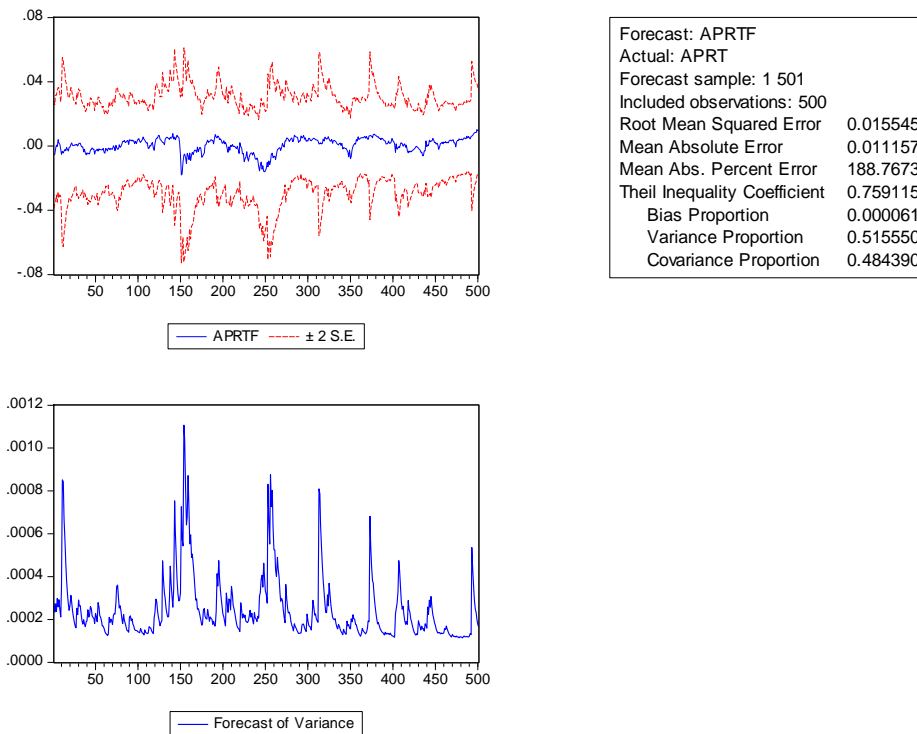
Fuente: Elaboración propia con datos de Bloomberg

Los residuos al cuadrado poseen una distribución asimétrica positiva, leptocúrtica y no se comportan de manera normal.

7.2.3 Pronóstico con modelo GARCH

El programa Eviews calculó los siguientes errores ante el modelado GARCH:

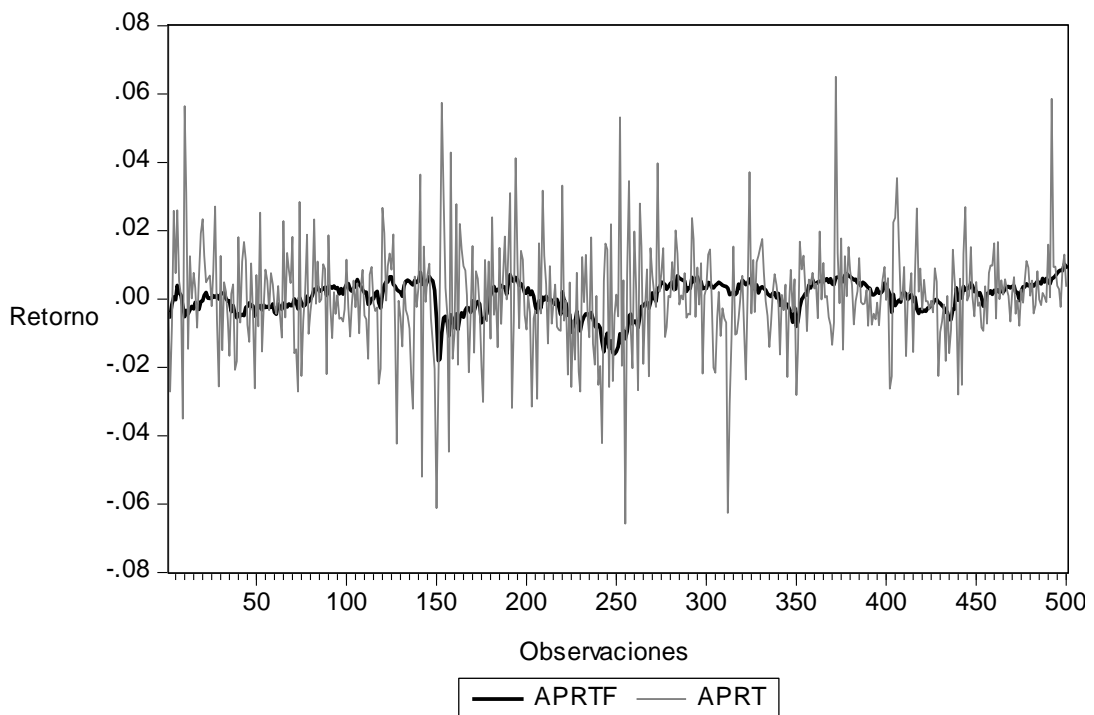
Gráficas 7.7 Raíz cuadrada de los residuos y varianza.



Fuente: Elaboración propia con datos de Bloomberg

La raíz de los errores medios al cuadrado alcanzó las 0.015545 unidades, la media del error absoluto fue de 0.011157 unidades, mientras que la suma de los errores cuadráticos alcanzó las 0.12082 unidades, cifra superior en 0.00018 unidades a la registrada en el cálculo del modelo ARCH. La expresión visual de los errores y la varianza se revela en la gráfica 7.7. Por otro lado, la gráfica 7.8 muestra los datos reales (línea gris delgada) en contraste con los datos ajustados mediante el modelo (línea negra gruesa).

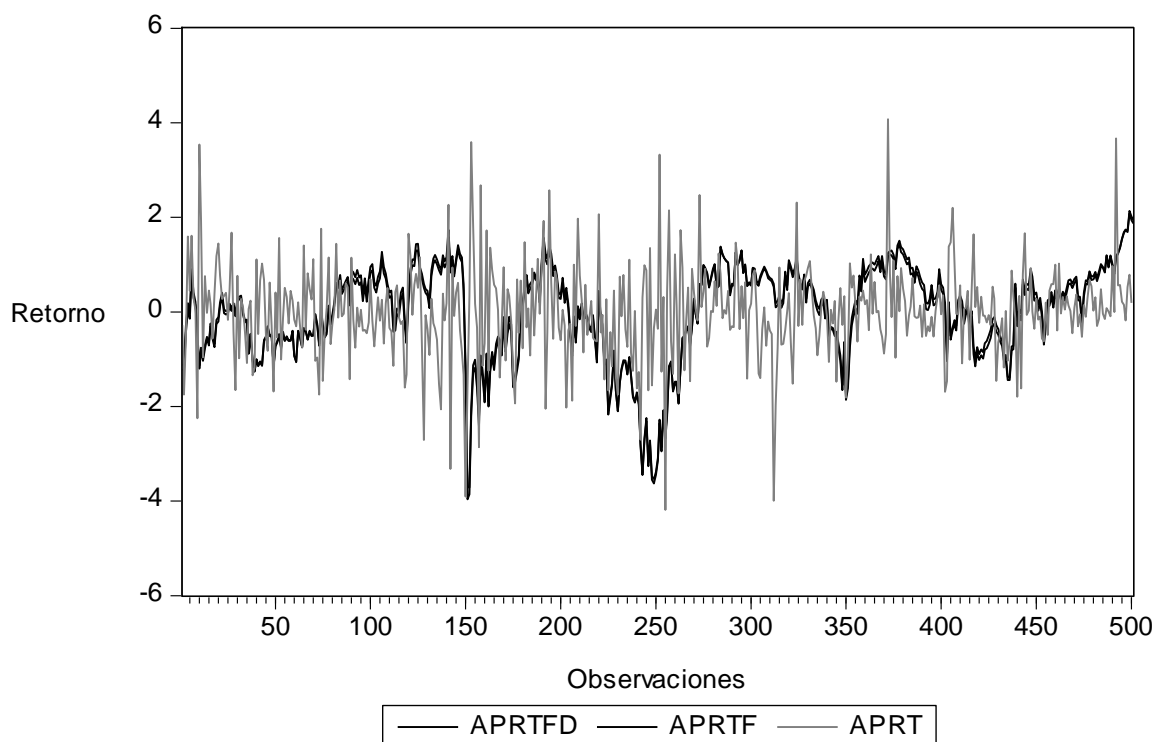
Gráfica 7.8 Datos pronosticados vs datos reales



Fuente: Elaboración propia con datos de Bloomberg.

Bajo el modelo GARCH la estimación de retorno para el día 16 de febrero de 2017 fue de 0.009146 o lo que es lo mismo 0.9146% mientras que, como ya se dijo anteriormente, el retorno real descendió a 0.0012, o 0.12% con respecto al día anterior. A continuación se muestra la gráfica que resume los datos pronosticados por los modelos ARCH y GARCH vs los datos originales.

Gráfico 7.9 Datos reales vs pronosticados los métodos anteriores.



Fuente: Elaboración propia con datos de Bloomberg.

La escala de la gráfica fue normalizada para la mejor apreciación de los resultados, no obstante la línea de datos pronosticados por el modelo GARCH es bastante similar a la conseguida por el modelo ARCH, razón por la cual se superponen en una sola línea obscura, siendo la línea más tenue la perteneciente a los datos reales.

7.3 Modelo de RNA para predicción

En este apartado se definirá el modelo RNA más adecuado, según la serie de tiempo financiera, para realizar su pronóstico; para ello se acotará la estructura, aprendizaje, y validación de la red propuesta.

7.4.1 Adecuación de la arquitectura de RNA

Las características que se eligieron para la red construida en SPSS fueron las siguientes:

- Estructura en capas: red de perceptrón multicapa
- Flujo de datos: de propagación hacia adelante.
- Tipo de respuesta: heteroasociativa.
- Mecanismo de aprendizaje: supervisado, por gradiente.
- Funciones de activación: Sigmoide.

La red de perceptrones multicapa, aunque no posee supuestos a priori también funciona minimizando el error de predicción de las variables de destino. La estructura utilizada será de propagación hacia adelante o feedforward, con la cual, la información de los nodos se encamina de la capa de entrada a la de salida sin la capacidad de retroalimentación durante el flujo. Se eligió este tipo de estructura dada la simplicidad de su aplicación. La respuesta es heteroasociativa puesto que se corresponde con un patrón o cifra distinta a la introducida y el aprendizaje es de tipo supervisado ya que se contó con datos de salida deseados que permitieron realizar el entrenamiento y el cálculo del error.

Para la creación del modelo se usó como variable dependiente el retorno de las acciones de Apple y como variables independientes la paridad USDCNY (dólar estadounidense vs yuan chino) y el Índice de Volatilidad (VIX), las cuales obtuvieron los mayores coeficientes de influencia en las estimaciones realizadas con los métodos econométricos. La variable dependiente así como las independientes son de tipo continuo ya que sus valores son categorías ordenadas entre las cuales puede existir un número infinito de otros valores que mantienen el mismo orden. El entrenamiento será realizado en línea y no en lotes dado el tamaño de la muestra.

Figura 7.1 Introducción del conjunto de datos

	APRT	VIX	USDCNY	var	var	var	var	var	var	var	var	var	var	var	var	var
1	.00	21.48	6.20													
2	-.03	22.39	6.19													
3	-.01	20.95	6.21													
4	.03	19.89	6.21													
5	.01	18.85	6.21													
6	.03	16.40	6.21													
7	.01	16.66	6.23													
8	.00	15.52	6.26													
9	-.04	17.22	6.24													
10	.06	20.44	6.25													
11	.03	18.76	6.25													
12	-.01	20.97	6.25													
13	.01	19.43	6.26													
14	.00	17.33	6.26													
15	.01	18.33	6.25													
16	.00	16.85	6.25													
17	-.01	17.29	6.24													
18	.01	18.55	6.25													
19	.02	17.23	6.24													
20	.02	16.96	6.24													
21	.01	15.34	6.25													
22	.00	14.69	6.24													
23	.01	15.80	6.25													

Fuente: Elaboración propia con datos de Bloomberg.

En el menú de SPSS se da por default la opción de analizar los datos por medio de redes neuronales. En este caso se eligió el modelo de Perceptrón Multicapa. En seguida se introdujeron los datos correspondientes a los retornos, la paridad USDCNY y el índice VIX.

Consecuentemente se especificó que los datos son de tipo numérico y de escala o continuos.

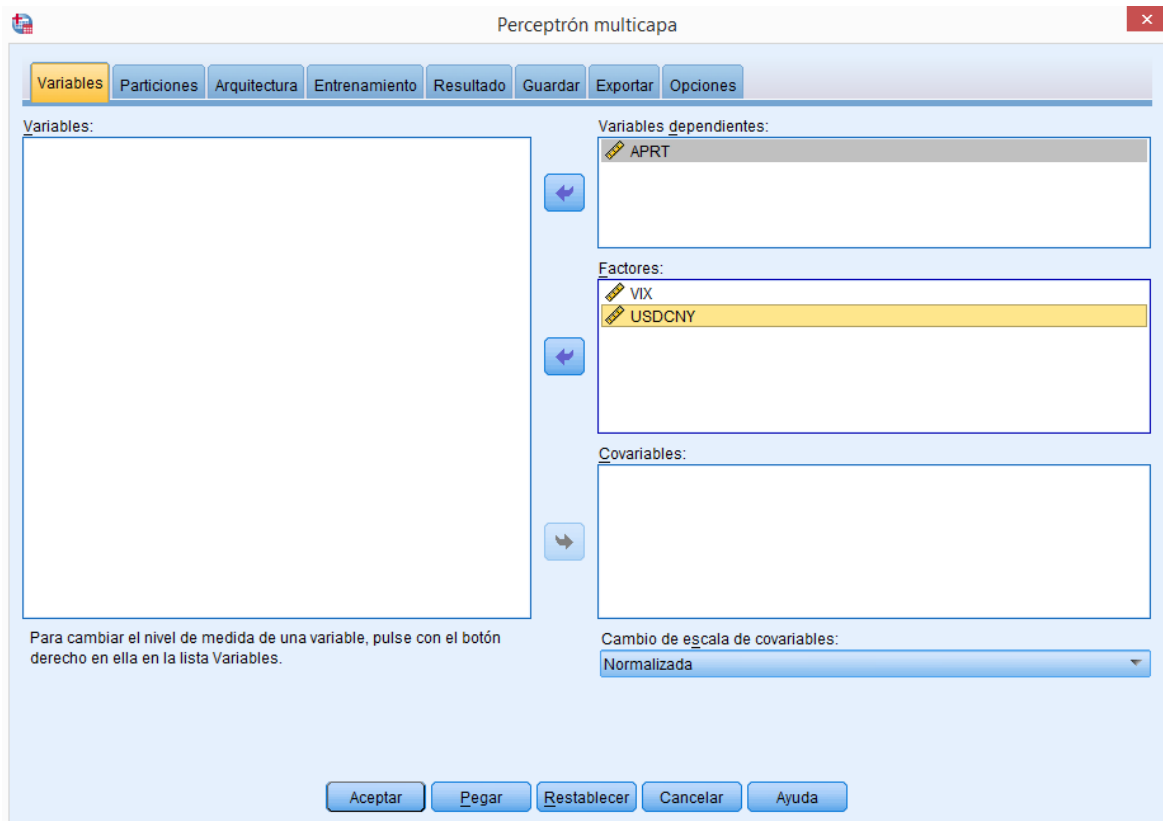
Figura 7.2 Especificación del tipo de variable

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas	Alineación	Medida	Rol
1	APRT	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
2	VIX	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
3	USDCNY	Numérico	8	2		Ninguna	Ninguna	8	Derecha	Escala	Entrada
4											
5											
6											
7											
8											
9											
10											
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
~											

Fuente: Elaboración propia con datos de Bloomberg.

Una vez finalizado el paso anterior se seleccionó al menos una variable dependiente y un factor o variable independiente. Si se requiere, en la pestaña “Cambio de escala de covariables” es posible cambiar la escala de las mismas, no obstante en este caso no se trabajó con tales elementos.

Figura 7.3 Selección de variables

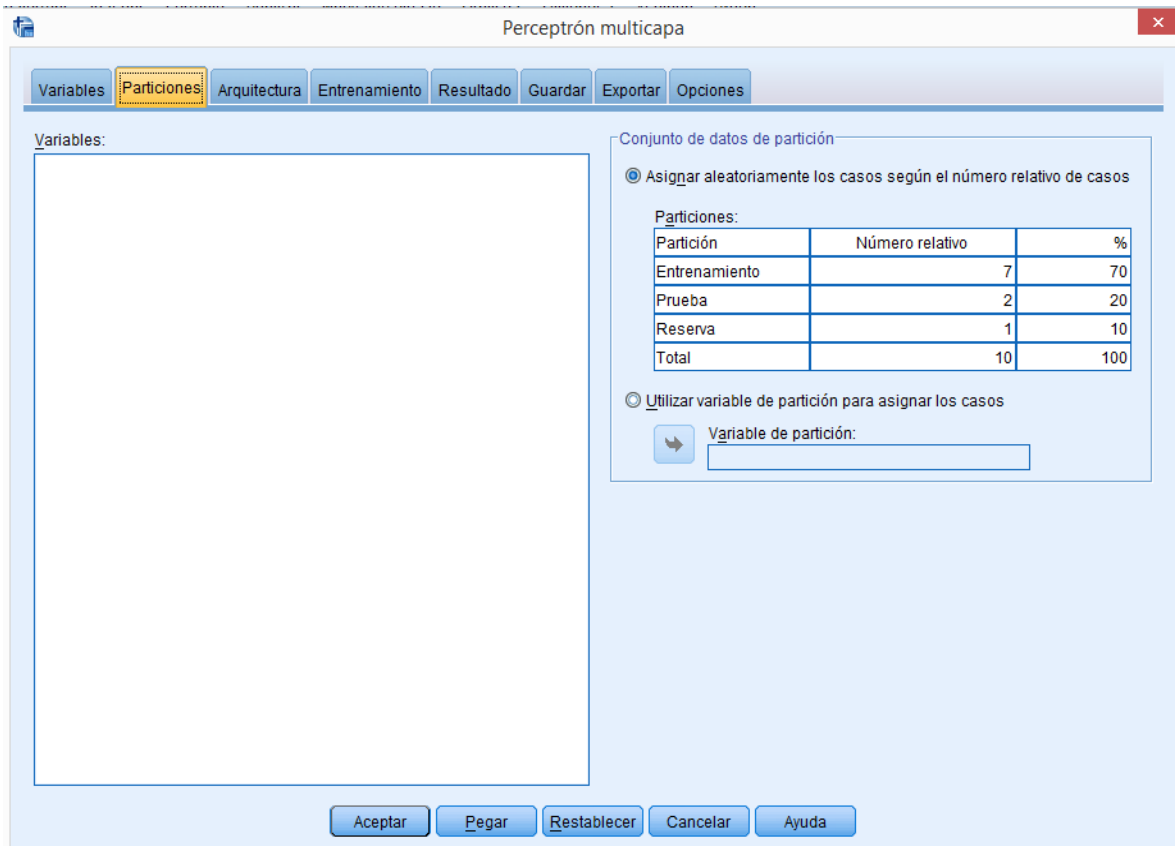


Fuente: Elaboración propia con datos de Bloomberg.

Después de esto se inició el registro y la exploración de los datos para continuar con las etapas de partición, arquitectura, entrenamiento y generación de resultados. La muestra de datos introducida se “partió” o dividió en un conjunto de datos de entrenamiento, de prueba y de reserva. Los primeros corresponden a aquellos que son procesados por la red, los siguientes se toman como un conjunto independiente de registros para desarrollar previamente un proceso de entrenamiento que es aplicado finalmente a los datos iniciales, esto se realiza con el propósito de evitar un exceso de entrenamiento en la red definitiva que pueda ralentizar su ejercicio. La muestra de reserva se utilizó para evaluar los resultados de la red ya construida en base a los errores generados.

El número de datos que fue utilizado por división se debió expresar porcentualmente para el programa. Generalmente se destina un 70% del total de los datos al entrenamiento, un 20% a un conjunto de prueba y el resto a la partida de reserva.

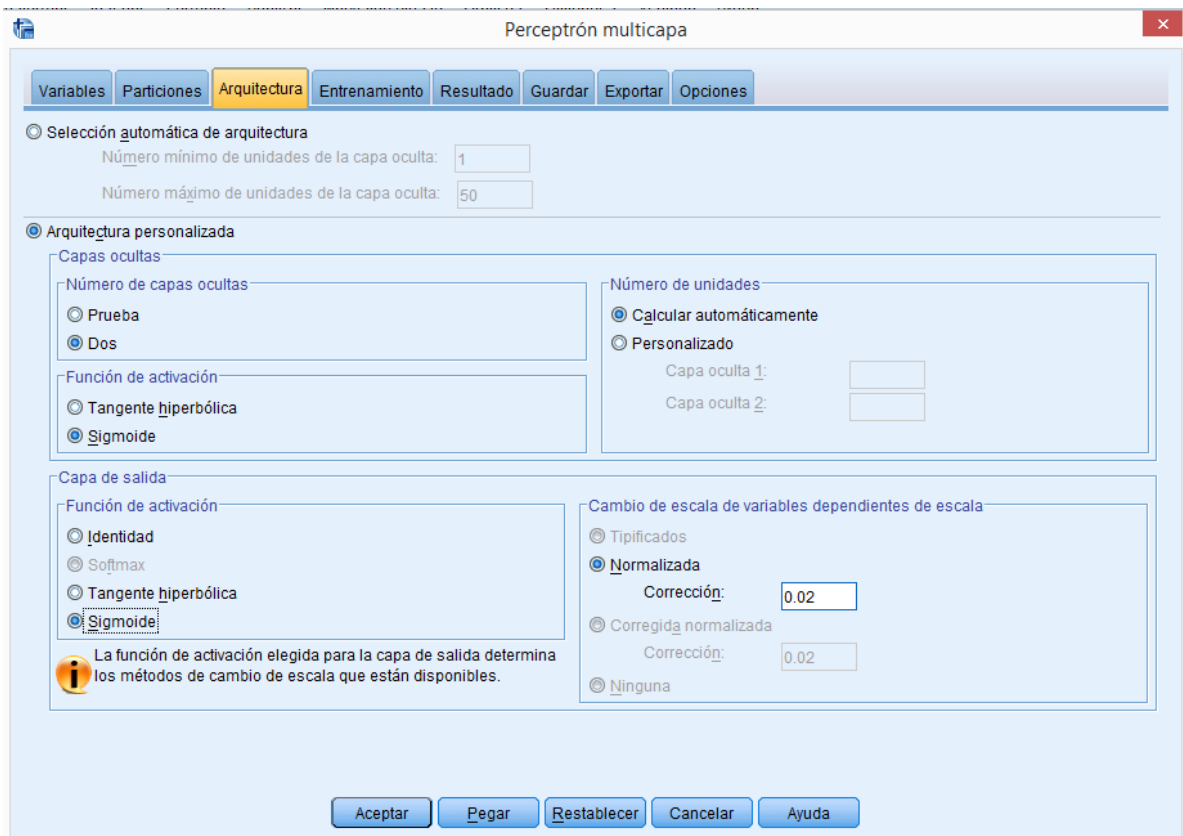
Figura 7.4 Conjunto de datos de partición



Fuente: Elaboración propia con datos de Bloomberg.

En la construcción de la arquitectura de la red el programa ofrece dos opciones, la automática y la personalizada. La primera genera una red con una sola capa oculta y calcula el número más eficiente de nodos o neuronas en ella. Así mismo hace uso de las funciones de activación que aparecen por defecto, la tangente hiperbólica e identidad para las capas ocultas y de salida respectivamente. La construcción personalizada permite al usuario elegir el número de capas ocultas así como la función de activación para estas y para la capa de salida, ofreciendo también el mejor número de nodos por capa. No obstante ninguna de las dos opciones mencionadas anteriormente admite la manipulación de los pesos sinápticos, ya que estos son determinados en un principio aleatoriamente y consecuentemente modificados por el algoritmo de estimación que se estipula en la pestaña de entrenamiento.

Figura 7.5 Arquitectura de la red.



Fuente: Elaboración propia con datos de Bloomberg.

Como es de esperar se optó por la construcción personalizada de la arquitectura, dado que la teoría expuesta anteriormente brinda las herramientas necesarias para concretar el tipo de características de una red de perceptrones multicapa destinada al pronóstico de datos. En este sentido, se eligieron dos capas ocultas con función de activación sigmoide. La escala de la variable dependiente se normaliza automáticamente con la selección de la función sigmoide; esta corrección de escala garantiza que los valores de todas las variables dependientes queden comprendidos dentro del rango $[0,1]$ que corresponden a los límites de la función sigmoide.

7.4.2 Entrenamiento del modelo de RNA

El objetivo del entrenamiento es ajustar los pesos de la red con el fin de que las entradas introducidas produzcan las salidas deseadas. Para este caso, dado que se dispone del vector de salida objetivo, el aprendizaje es de tipo supervisado.

Antes de comenzar el proceso de entrenamiento, todos los pesos deben ser inicializados a pequeños valores aleatorios. Esto asegura que la red no sea saturada por grandes pesos y

que no se comience en un punto circunstancial de la superficie de error. El entrenamiento se conforma por una serie de pasos:

1. Elegir el vector de entrenamiento del conjunto total de datos e introducirlo como entrada a la red.
2. Obtener el vector de salida de la red.
3. Calcular el error entre el vector de salida de la red y el vector de salida esperado.
4. Modificar los pesos sinápticos de la red con el fin de minimizar el error.
5. Modificar los pasos 1 al 4 para cada vector del conjunto de entrenamiento hasta disminuir convenientemente el error.

Los primeros pasos (1 y 2) constituyen la fase de aprendizaje hacia adelante, mientras que los pasos 3 y 4 pueden fungir como la fase hacia atrás en una red backpropagation. Como ya se indicó, usualmente, de todo el conjunto de entrenamiento se emplean aproximadamente un 70% de los patrones para entrenar, reservándose un 30% como conjunto de test y reserva.

Figura 7.6 Entrenamiento

The screenshot shows the 'Perceptrón multicapa' software interface with the 'Entrenamiento' tab selected. The interface includes several sections for configuring training parameters:

- Tipo de entrenamiento:** Radio buttons for 'Lote', 'En línea' (selected), and 'Mini lote'. Under 'Mini lote', there are options for 'Número de registros en cada mini lote': 'Calcular automáticamente' (selected) and 'Personalizado' with a text input field containing '2'.
- Algoritmo de optimización:** Radio buttons for 'Gradiente conjugado escalado' and 'Pendiente de gradiente' (selected).
- Opciones de entrenamiento:** A table with the following data:

Opción	Valor
Tasa de aprendizaje inicial	0.4
Límite inferior de la tasa de aprendizaje	0.001
Reducción de la tasa de aprendizaje, en Épocas	10
Impulso	0.9
Centro de intervalo	0
Desplazamiento de intervalo	±0.5

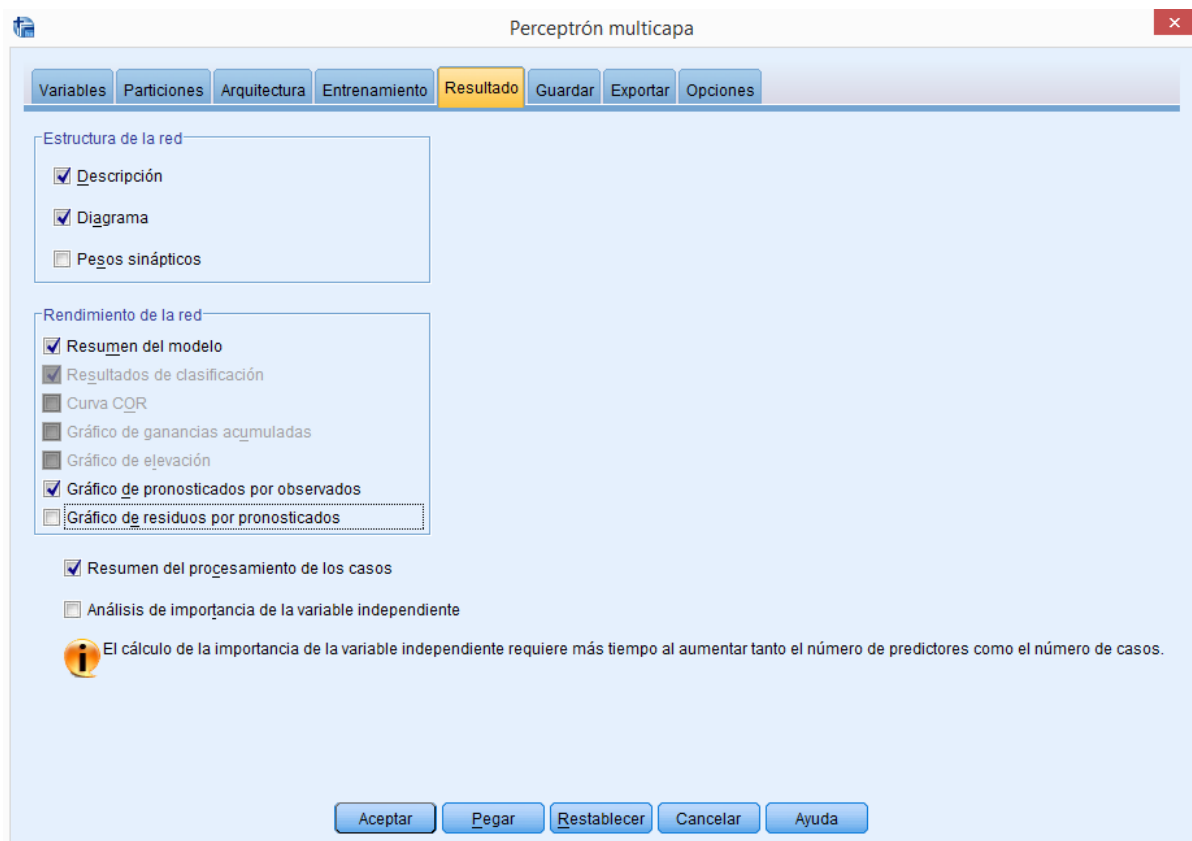
At the bottom of the window, there are buttons for 'Aceptar', 'Pegar', 'Restablecer', 'Cancelar', and 'Ayuda'.

Fuente: Elaboración propia con datos de Bloomberg.

En la pestaña siguiente se establece el entrenamiento en línea (figura 7.6) tras ocupar entradas que rebasan las 100 unidades y el algoritmo de optimización corresponde a la pendiente de gradiente para alcanzar máximos y mínimos globales de optimización.

Enseguida se opta por los instrumentos gráficos y estadísticos que se desean observar. En este caso fue menester analizar el diagrama de la red final, su descripción y resumen cuantitativo, así como el gráfico de datos observados por pronosticados.

Figura 7.7 Resultado



Fuente: Elaboración propia con datos de Bloomberg.

Finalmente se ordena al programa el registro de los valores pronosticados en la hoja de cálculo donde se insertaron los datos iniciales.

Figura 7.8 Opciones de guardado

Perceptrón multicapa

Variables Particiones Arquitectura Entrenamiento Resultado **Guardar** Exportar Opciones

Guardar categoría o valor pronosticado para cada variable dependiente
 Guardar pseudoprobabilidad pronosticada para cada variable dependiente

Variables:

Variable dependiente	Categoría o valor pronosticado	Pseudoprobabilidad pronosticada	
	Nombre de la variable guardada	Nombre raíz de las variables guardadas	Categorías para guardar
APRT	MLP_PredictedValue		25

Nombres de las variables guardadas

Generar automáticamente nombres únicos
 Seleccione esta opción si desea añadir un nuevo conjunto de variables guardadas al conjunto de datos cada vez que ejecute un modelo.

Nombres personalizados
 Especificar nombres de las variables. Si selecciona esta opción, se reemplazarán todas las variables existentes con el mismo nombre o nombre raíz cada vez que se ejecute un modelo.

Aceptar Pegar Restablecer Cancelar Ayuda

Fuente: Elaboración propia con datos de Bloomberg.

Figura 7.7 Otras opciones

Perceptrón multicapa

Variables Particiones Arquitectura Entrenamiento Resultado Guardar Exportar **Opciones**

Valores definidos como perdidos por el usuario
 Especifique cómo desea tratar los casos con valores definidos como perdidos por el usuario en los factores y las variables dependientes categóricas.
 Excluir Incluir
 Los casos con valores definidos como perdidos por el usuario en las covariables o en las variables dependientes de escala siempre se excluyen.

Reglas de parada
 Las reglas de parada se prueban en el orden que se indica a continuación.

Número máximo de pasos sin una disminución del error: 1

Datos para usar para calcular el error de predicción:
 Elegir automáticamente
 Datos de entrenamiento y de prueba

Tiempo máximo de entrenamiento Minutos: 15

Número máximo de épocas de entrenamiento
 Calcular automáticamente
 Especificar valor personalizado Número máximo de épocas:

Cambio mínimo relativo del error de entrenamiento: 0.0001

Cambio mínimo relativo de la tasa de errores de entrenamiento: 0.001

Número máximo de casos que se almacenarán en memoria: 1000

Aceptar Pegar Restablecer Cancelar Ayuda

Fuente: Elaboración propia con datos de Bloomberg.

En otras opciones se puede estipular el tiempo de entrenamiento así como el número de pasos o ciclos de aprendizaje que se desean seguir antes de detener la preparación, no obstante la elección de estos factores serán delegados al programa.

7.4.3 Validación del modelo de RNA

En este apartado se verificará la funcionalidad de la red construida y se vislumbrará su modelado gráfico.

7.4.4 Pronóstico con modelo de RNA

El número de datos destinados al entrenamiento fueron 338 lo que representó el 88.3% del total; como prueba se utilizaron 29 datos y de reserva 16, lo que corresponde al 7.6% y 4.2% respectivamente. Por otro lado el programa excluyó 118 datos.

Tabla 7.11 Resumen del procesamiento de casos

		N	Porcentaje
Muestra	Entrenamiento	338	88.3%
	Prueba	29	7.6%
	Reserva	16	4.2%
Válidos		383	100.0%
Excluidos		118	
Total		501	

Fuente: Elaboración propia con datos de Bloomberg.

El número de datos utilizados fueron 350. Las capas ocultas de la red fueron dos, la primera cuenta con 20 neuronas y la segunda con 15 neuronas. Como ya se especificó anteriormente la función de activación fue de tipo sigmoide y se entrenó a partir de la suma de los errores cuadrados.

Tabla 7.12 Información general de la red

Capa de entrada	Factores	1	VIX
		2	USDCNY
	Número de unidades ^a		350
Capas ocultas	Número de capas ocultas		2
	Número de unidades de la capa oculta 1 ^a		20
	Número de unidades de la capa oculta 2 ^a		15
	Función de activación		Sigmoide

Capa de salida	VARIABLES DEPENDIENTES	1	APRT
	Número de unidades	1	
	Método de cambio de escala para dependientes de escala	Normalizada	
	Función de activación	Sigmoide	
	Función de error	Suma de cuadrados	
a. Sin incluir la unidad de sesgo			

Fuente: Elaboración propia con datos de Bloomberg.

La suma de los errores cuadráticos del conjunto de entrenamiento ascendió a 1.462 unidades, en tanto que la suma de los errores cuadráticos de la muestra de prueba fue de 0.069. Únicamente fueron utilizados 0.98 segundos para el procesamiento de la red y se aplicó un solo ciclo o paso de entrenamiento.

Tabla 7.13 Resumen del modelo

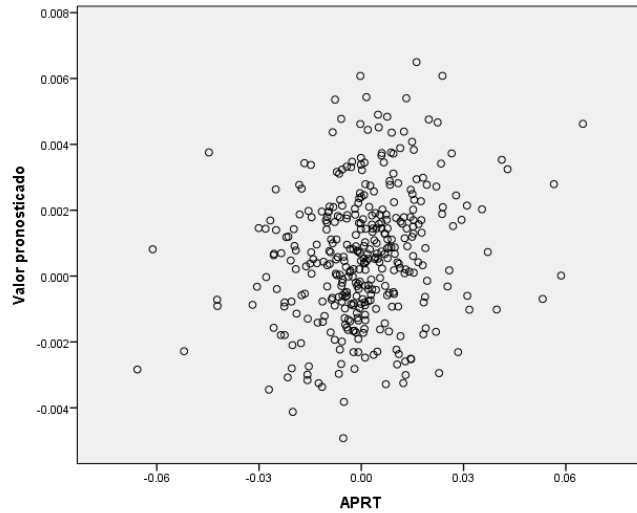
Valoraciones		
Entrenamiento	Suma de errores cuadráticos	1.462
	Error relativo	.959
	Regla de parada utilizada	1 pasos consecutivos sin disminución del error ^a
	Tiempo de entrenamiento	0:00:00.98
Prueba	Suma de errores cuadráticos	.069
	Error relativo	.947
Reserva	Error relativo	1.256
Variable dependiente: APRT		
a. Los cálculos del error se basan en la muestra de prueba.		

Fuente: Elaboración propia con datos de Bloomberg.

La gráfica de dispersión de valores pronosticados por observados (7.10) se interpreta de la siguiente manera: mientras la dispersión de los datos esté más concentrada entonces el modelo de pronóstico se ajusta de manera más eficiente al comportamiento real del

fenómeno; el caso contrario sucede cuando existe poca concentración. Para este modelo el ajuste fue regular, ya que se observa que los valores no se sobrepone del todo.

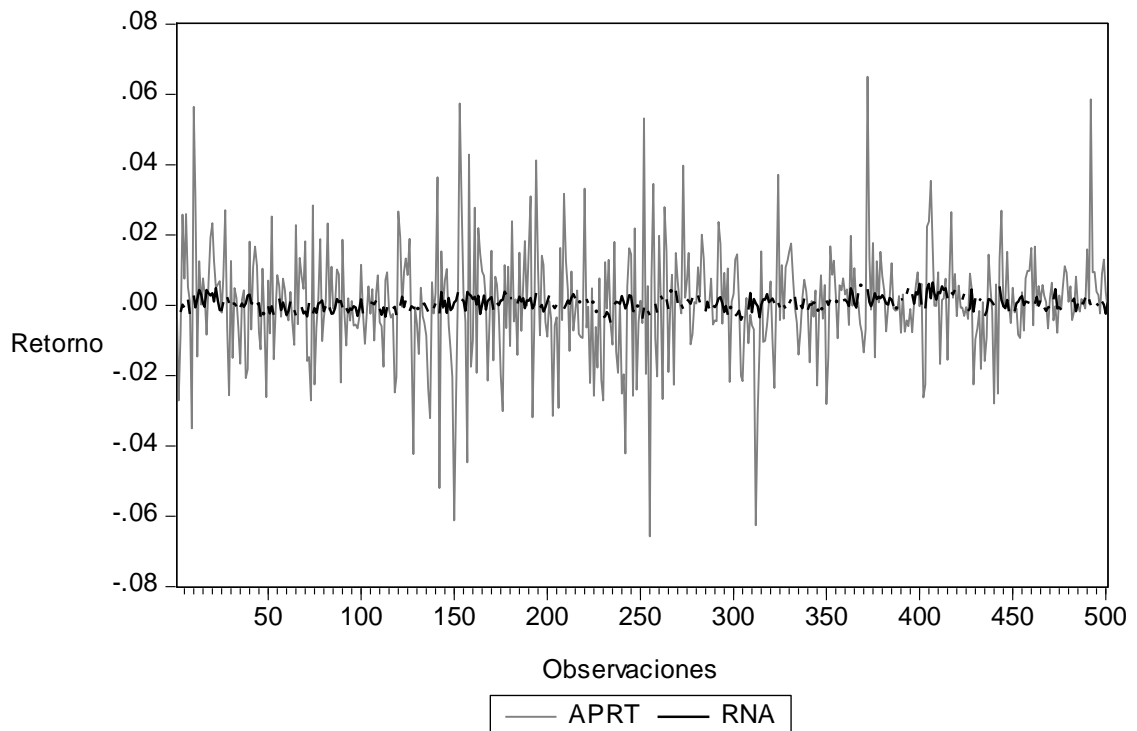
Grafica 7.10 Valores pronosticados por observados



Fuente: Elaboración propia con datos de Bloomberg.

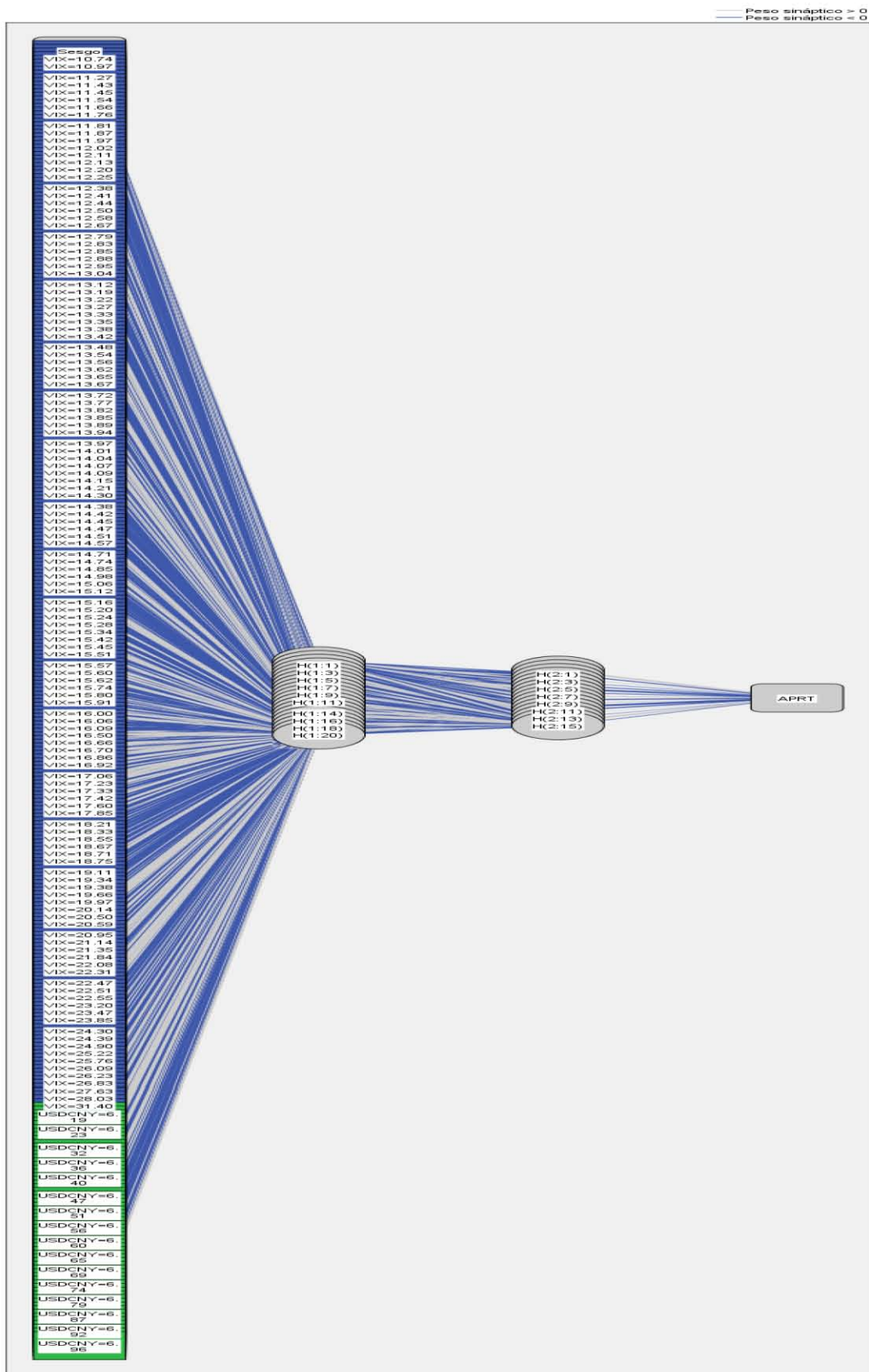
A continuación se muestra una gráfica de los valores estimados en contraposición con los valores reales. El ajuste parece más débil que es logrado por los modelos econométricos.

Gráfica 7.10 Datos pronosticados vs datos reales



Fuente: Elaboración propia con datos de Bloomberg.

Figura 7.8 Red neuronal multicapa



Fuente: Elaboración propia con datos de Bloomberg.

Bajo el modelo RNA la estimación de retorno para el día 16 de febrero de 2017 fue de 0.008 o lo que es lo mismo 0.8%, en tanto que el retorno real descendió a 0.0012, o 0.12% con respecto al día anterior.

Se observa la red neuronal de manera gráfica en la figura anterior (7.8).

Capítulo 8. Conclusiones

Cada uno de los métodos de pronóstico tienen marcadas cualidades y diferencias. Mientras que los modelos de heteroscedasticidad condicional pretenden adaptar los patrones de volatilidad, las redes neuronales son en sí un proceso dinámico cuya adaptación está intrínseca en su construcción. No obstante, este último método resulta un tanto versátil, ya que no existen enfoques cuantitativos que dicten a priori su construcción.

La ventaja de las redes neuronales es que se puede aplicar a una multitud de problemas de diversa índole, proporcionando con frecuencia buenas soluciones con no demasiado tiempo de construcción. No obstante, si se requiere una solución realmente excelente, habrá que dedicar más empeño al desarrollo del sistema neuronal, teniendo en cuenta diferentes cuestiones adicionales como algoritmos eficientes, arquitecturas óptimas, selección de los pesos iniciales, estilo de aprendizaje, procesamiento de los datos de entrada, conjunto de aprendizaje empleado, utilización de técnicas que eviten el sobreajuste, etc. Por otra parte debe tenerse en cuenta que los algoritmos utilizados por el programa SPSS no garantizan alcanzar el mínimo global de la función error, sino tan sólo un mínimo local en algunas ocasiones, por lo que el proceso de aprendizaje puede quedarse estancado, resultando en cifras menos consistentes.

Realizando la comparación de los pronósticos el modelo ARCH estimó un retorno de 0.0094 o lo que es lo mismo 0.94% para el día 16 de febrero de 2017, mientras que el modelo GARCH lo estimó en 0.009146 o 0.9146%, en tanto que el retorno real se posicionó en -0.0012 o -0.12%. La red neuronal por su parte estimó un retorno de 0.008 o 0.8%, cifra que no difiere en demasía de las obtenidas por los modelos econométricos. No obstante la suma de los errores cuadráticos se posicionó en 0.120661 y 0.120824 unidades para los modelos ARCH y GARCH respectivamente, y en 1.462 unidades para el modelo artificial. Esto evidencia la mayor eficacia de los métodos econométricos en el cálculo de un modelo más aproximado a los datos reales. Con esta sencilla comparación podría manifestarse inicialmente que la hipótesis del presente trabajo se rechaza, empero todos los modelos son perfectibles, por lo que se podría maximizar la estimación de cada uno realizando los ajustes correspondientes.

Bibliografía

1. Bowerman, B (2010) Forecasting, Time series and Regression, EU, Cengage Learning.
2. Bonifacio, Martín (2002) Redes Neuronales y Sistemas Difusos, España, Alfa Omega.
3. Caridad, JM, (2005) Econometría, modelos econométricos y series de tiempo, España, Reverte.
4. Carlin and David Soskice (1990) Macroeconomics and the Wage Bargain: A Modern Approach to Employment, Inflation, and the Exchange Rate, U.U.EE. Oxford University Press.
5. CYTED (1999) Aplicaciones de las Redes de Neuronas en Supervisión, Diagnóstico y Control de Procesos, Universidad Simón Bolívar, Venezuela
6. De Abajo, Nicolás (2001) Introducción a la inteligencia Artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva, Universidad de Oviedo, España.
7. De Lara, Alfonso (2005) Medición y control de riesgos financieros, Limusa, México.
8. Evans, Michael J. y Jeffrey S. Rosenthal (2004) Probabilidad y Estadística, Reverté, México.
9. Fabris, Julio (2009) Econometría Financiera: modelos y pronósticos utilizando QMS EViews, Buenos Aires, Omicron System.
10. Greene, William H. (1999), Análisis Económico, tercera edición, Pearson Educación, España
11. Gujarati, Damodar N. (2009) Econometría, McGraw-Hill, quinta edición, México
12. Hanke, John E. (2010) Pronósticos en los negocios, Pearson Educación, novena edición, México.
13. <https://www.wikipedia.org/>
14. Lara, Felipe () Fundamentos de Redes Neuronales Artificiales, UNAM, México. (artículo)
15. Lévy, Jean-Pierre (2008) Las Redes Neuronales Artificiales: Fundamentos teóricos y aplicaciones prácticas, Netbiblo, España.

16. Makridakis, Spyros (1983) *Forecasting: Methods and Applications*, segunda edición, Nueva York.
17. Matich, Damián Jorge (2001) *Redes Neuronales: Conceptos Básicos y Aplicaciones*, Universidad Tecnológica Nacional – Facultad Regional Rosario, Argentina.
18. Morales, Efraín (2001) *Introducción a la Econometría*, Ministerio de Educación, Ecuador.
19. Osorio, Edgar (2006) *Gestión Financiera Empresarial*, Pontificia Universidad Javeriana, Bogotá.
20. Pérez, Cesar (2006) *Econometría de las series temporales*, Pearson Educación, Madrid.
21. Pérez, López (2006), *Problemas resueltos de econometría*, Thomson, España.
22. Tharp, Van K. (2006) *Tener Éxito en Trading*, Netbiblo, Estados Unidos
23. Torra, Salvador (2013) *Modelos Neuronales Aplicados en Economía*, Addlink, España.
24. Universidad de Alcalá (1993) *Redes Neuronales Artificiales: Fundamentos y Aplicaciones*, Servicios de Publicaciones de la Universidad de Alcalá. España.
25. Weatherall, James (2013) *Cuando los Físicos asaltaron los mercados*, Ariel, México.

Anexos

Date	APRT	USDJPY	XAU	VIX	WITEC	DJUSTCT	USDCNY	RNA	ARCH	GARCH
14/01/2015	-0.0038	117.33	1228.72	21.48	656.43	1239.42	6.20		-0.0046	-0.0053
15/01/2015	-0.0271	116.17	1262.75	22.39	648.76	1219.34	6.19		-0.0045	-0.0052
16/01/2015	-0.0078	117.51	1280.45	20.95	652.43	1230.97	6.21	-0.002	-0.0008	-0.0016
20/01/2015	0.0258	118.82	1295.32	19.89	658.57	1241.31	6.21	0.0002	0.0014	0.0006
21/01/2015	0.0076	117.97	1293.09	18.85	662.21	1244.05	6.21	-6E-04	0.0005	-0.0002
22/01/2015	0.0260	118.49	1302.25	16.40	672.03	1269.44	6.21		0.0047	0.0040
23/01/2015	0.0052	117.77	1294.08	16.66	674.86	1272.42	6.23	0.0016	0.0025	0.0019
26/01/2015	0.0011	118.46	1281.40	15.52	674.19	1268.63	6.26	-0.003	0.0015	0.0011
27/01/2015	-0.0350	117.87	1292.34	17.22	659.02	1227.26	6.24		-0.0017	-0.0020
28/01/2015	0.0565	117.54	1284.49	20.44	657.52	1225.79	6.25	0.0028	-0.0047	-0.0052
29/01/2015	0.0311	118.29	1257.27	18.76	660.03	1239.01	6.25	-6E-04	-0.0029	-0.0034
30/01/2015	-0.0146	117.49	1283.79	20.97	651.54	1219.45	6.25	0.0018	-0.0038	-0.0044
02/02/2015	0.0125	117.57	1274.41	19.43	655.63	1230.98	6.26	0.0044	-0.0027	-0.0032
03/02/2015	0.0002	117.57	1260.52	17.33	661.83	1244.31	6.26	0.0032	-0.0019	-0.0022
04/02/2015	0.0077	117.28	1269.35	18.33	663.54	1245.45	6.25	0.0005	-0.0024	-0.0029
05/02/2015	0.0032	117.53	1264.82	16.85	667.32	1256.90	6.25	-4E-04	-0.0005	-0.0010
06/02/2015	-0.0084	119.12	1234.04	17.29	664.59	1252.14	6.24	0.0044	-0.0018	-0.0022
09/02/2015	0.0066	118.64	1239.03	18.55	662.73	1249.63	6.25	0.0034	-0.0026	-0.0030
10/02/2015	0.0192	119.43	1233.55	17.23	671.11	1270.53	6.24	0.0028	-0.0002	-0.0008
11/02/2015	0.0234	120.46	1218.83	16.96	672.96	1276.22	6.24	0.0034	0.0001	-0.0005
12/02/2015	0.0127	119.11	1221.98	15.34	681.28	1296.89	6.25	0.0011	0.0015	0.0010
13/02/2015	0.0049	118.75	1229.76	14.69	687.22	1308.00	6.24	0.0049	0.0024	0.0019
17/02/2015	0.0059	119.25	1209.89	15.80	686.31	1308.51	6.25	0.0014	0.0008	0.0003
18/02/2015	0.0069	118.79	1212.42	15.45	687.88	1311.03	6.26	-0.001	0.0006	0.0002
19/02/2015	-0.0021	118.95	1206.84	15.29	690.39	1316.39	6.25	-0.001	0.0008	0.0003
20/02/2015	0.0081	119.03	1201.94	14.30	693.44	1323.96	6.26	0.0009	0.0014	0.0010
23/02/2015	0.0271	118.81	1201.83	14.56	694.23	1325.09	6.26		0.0009	0.0005
24/02/2015	-0.0062	118.97	1200.53	13.69	696.07	1328.05	6.26	0.0023	0.0014	0.0011
25/02/2015	-0.0256	118.86	1205.00	13.84	693.87	1318.59	6.26	0.0007	0.0003	0.0000
26/02/2015	0.0126	119.41	1209.49	13.91	697.11	1328.63	6.26		0.0018	0.0014
27/02/2015	-0.0150	119.63	1213.18	13.34	693.97	1321.02	6.27	0.0004	0.0020	0.0017
02/03/2015	0.0049	120.13	1206.83	13.04	700.12	1333.18	6.27	0.0005	0.0021	0.0018
03/03/2015	0.0021	119.73	1203.74	13.86	695.95	1322.62	6.27	-3E-04	0.0005	0.0002
04/03/2015	-0.0063	119.68	1200.34	14.23	693.93	1317.78	6.27		-0.0003	-0.0006
05/03/2015	-0.0166	120.13	1198.40	14.04	693.43	1316.04	6.27	-5E-04	-0.0002	-0.0005
06/03/2015	0.0015	120.83	1167.29	15.20	687.70	1301.65	6.26	0.0005	-0.0034	-0.0037
09/03/2015	0.0043	121.15	1167.21	15.06	687.61	1307.82	6.26	-4E-04	-0.0016	-0.0019
10/03/2015	-0.0207	121.13	1161.84	16.69	675.38	1280.20	6.26	0.0005	-0.0040	-0.0044
11/03/2015	-0.0182	121.45	1155.31	16.87	673.01	1272.78	6.26	0.0028	-0.0051	-0.0055
12/03/2015	0.0181	121.29	1153.73	15.42	675.17	1278.00	6.26	0.003	-0.0040	-0.0043
13/03/2015	-0.0069	121.40	1158.48	16.00	673.54	1272.38	6.26	0.0006	-0.0047	-0.0049