
Una aplicación de redes neuronales artificiales para el pronóstico de los rendimientos de la serie de Ecopetrol

An application of artificial neural networks to forecast the performance of the Ecopetrol series

Autor : Duván Alejandro García Ríos^a
duvan.garcia@usantotomas.edu.co

Director: Edna Carolina Moreno López^b
ednamoreno@usantotomas.edu.co

Director: Gil Robert Romero^c
gilromero@usantotomas.edu.co

Resumen

Este trabajo propone un modelo basado en redes neuronales recurrentes (RNN) para pronosticar el rendimiento la acción de Ecopetrol, la cual es una de las acciones más tranzadas en el país. Se estudian varias configuraciones de RNN, incluyendo como variables de entrada la tasa representativa de mercado (TRM), euro (EUR), precio de cierre de la respectiva acción a estimar y los precios del petróleo (Brent y WTI). Una parte de los datos es tomada como entrenamiento y la restante para la predicción, finalmente para contrastar los resultados obtenidos, se ajusta un modelo EGARCH.

Palabras clave: redes neuronales artificiales, rendimientos, EGARCH, volatilidad condicional.

Abstract

This work proposes a model based on recurrent neural networks (RNN) to forecast the performance of Ecopetrol, which is one of the most successful actions in the country. Various configurations of RNN are studied, including as input variables the representative market rate (TRM), euro (EUR), closing price of the respective share to estimate and oil prices (Brent and WTI). A part of the data is taken as training and the rest for prediction, finally to contrast the obtained results, an EGARCH model is fitted.

Keywords: artificial neural networks, returns, EGARCH, conditional volatility

^aEstudiante pregrado Estadística, Universidad Santo Tomás, sede Bogotá

^bDocente Facultad de Estadística, Universidad Santo Tomás, sede Bogotá

^cDocente Facultad de Estadística, Universidad Santo Tomás, sede Bogotá

1. Introducción

El rendimiento de una acción incluye los componentes de ganancia o pérdida de capital, los cuales son el resultado del movimiento en los precios de la acción; una ganancia es lo que resulta de un aumento de precio, mientras que una pérdida lo que resulta de una disminución de este, por lo tanto, generar un modelo para predecir este comportamiento que indique las señales de compra o venta de títulos para los inversionistas, se ha convertido en una tarea de vital interés, pero encontrar el modelo adecuado no es sencillo, ya que el precio de la acción puede verse afectado por diversas variables externas.

Es por ello que se han venido desarrollado técnicas para la estimación de la volatilidad, gran parte de los métodos de hoy en día están basados en la metodología Box-Jenkins donde se propone un método de modelación con modelos ARMA (Box,G;Jenkins,G,1976). Sin embargo, los modelos de series de tiempo tradicionales que suponen varianza homocedástica, no son adecuados para modelar series de tiempo financieras, ya que estas presentan períodos de volatilidad turbulenta, seguidos de períodos de calma con apenas fluctuaciones. Para solucionar este problema, Robert Engle introduce en 1982 el modelo ARCH (autorregresivos condicionalmente heterocedásticos), donde la varianza condicionada a la información del pasado no es constante, sino que se relaciona con los cuadrados de las innovaciones anteriores.

Los modelos autorregresivos de heterocedasticidad condicional generalizados (GARCH) fueron desarrollado por Tim Bollerslev en 1986, en este tipo de modelo la varianza condicional depende no solo del cuadrado de los errores retrasados q períodos como en el modelo ARCH(q), sino además, de las varianzas condicionales retrasadas p períodos, la principal ventaja de este modelo es que consiguiendo resultados similares a los de un ARCH, éste logra mejoras en la parsimonia y robustez. (Monsegny,M ; Cuervo,E.,2008,pp.3-4)

Sin embargo, este también tiene sus desventajas a la hora de ser implementado, pues como lo asegura Abascal *“La desventaja principal de estos modelos GARCH, es su ineficiencia a la hora de identificar las asimetrías existentes en las series temporales financieras, provocadas por los desequilibrios entre las reacciones de los activos ante buenas y malas noticias.”* (Abascal, M.,2016, p.4).

Una solución a este problema es el modelo EGARCH, también conocido como GARCH exponencial, el cual fue presentado por Nelson, quien formula un modelo para la varianza condicional, que no se comporta de una manera simétrica en las perturbaciones positivas y negativas, ya que en general la volatilidad tiende a ser mayor en los choques negativos (Nelson, D .,1991,pp.347-370).

En este trabajo se realiza un procedimiento alternativo a los métodos de modelado clásicos (GARCH, EGARCH) utilizados para el pronóstico de series de tiempo financieras, el procedimiento alternativo empleado será el de algoritmos de redes neuronales recurrentes, debido a que son capaces de modelar y predecir series de tiempo lineales y no lineales con un alto grado de precisión, de la misma manera calculan cualquier tipo de interacción entre los datos y no se requiere de conocimiento previo del problema que se está modelando (Escobar L.,Valdés H., Zapata C.,2010, pp.9-21.), en ese sentido tienen una buena precisión, y permiten dar solución a diversos problemas que presentan las metodologías existentes.

Las redes neuronales actualmente tienen un papel muy importante dentro de la inteligencia artificial y se han inspirado en el comportamiento del cerebro humano, en las neuronas y las conexiones que estas generan, de ahí su nombre. En 1888 Santiago Ramón y Cajal demuestra que el sistema nervioso está compuesto por una red de células individuales, las células se conectan a otras neuronas estimulando su actividad para formar circuitos que pueden procesar información entrante y a su vez dar una respuesta. Los circuitos neuronales pueden ser muy simples estando compuestos de solo unas pocas neuronas, o pueden componerse de redes neuronales más complejas.

Basado en lo anterior McCulloch y Pitts desarrollaron el primer modelo computacional en 1943, el cual era un modelo binario, donde cada neurona tenía un escalón o umbral prefijado, y sirvió de base para los modelos posteriores. (Marín, J., 2012, p.5).

2. Marco teórico

2.1. Modelo EGARCH (m, s)

El modelo autorregresivo de heterocedasticidad condicional generalizado, propuesto por (Nelson, D., 1991, pp.347-370). modela el efecto de asimetría, considerando la función g de las innovaciones ϵ_t , que son variables independientes e idénticamente distribuidas (i.i.d). Estas variables tienen como media cero. El modelo EGARCH involucra el valor de la innovación ϵ_t , el cual tiene la siguiente forma:

$$a_t = \sigma_t \epsilon_t \quad (1)$$

$$\ln \sigma_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i \ln(\sigma_{t-i}^2) + \sum_{j=1}^s \beta_j g(\epsilon_{t-j}) \quad (2)$$

$$g(\epsilon_t) = \delta \epsilon_t + \alpha(|\epsilon_t| - E|\epsilon_t|) \quad (3)$$

El efecto de asimetría puede verse claramente al expresar la función g por casos así:

$$g(\epsilon_t) = \begin{cases} (\delta + \alpha)\epsilon_t - \alpha E(|\epsilon_t|), & \text{si } \epsilon_t \geq 0 \\ (\delta - \alpha)\epsilon_t - \alpha E(|\epsilon_t|), & \text{si } \epsilon_t < 0 \end{cases} \quad (4)$$

Si ϵ_t es normal, entonces

$$E(|\epsilon_t|) = \sqrt{\frac{2}{\pi}} \quad (5)$$

Si ϵ_t es *Student - t*, entonces

$$E(|\epsilon_t|) = \frac{2\sqrt{v-2} \Gamma(\frac{v+1}{2})}{(v-1) \Gamma(\frac{v}{2})\sqrt{\pi}} \quad (6)$$

Ya que el verdadero proceso de volatilidad es inobservable, se opta por usar un proxy de volatilidad $(r_t - \bar{r})^2$

$$ECMProxy(n) = \frac{1}{N} \sum_{t=1}^N [(r_{t+n} - \bar{r})^2 - \hat{h}_t(n)]^2 \quad (7)$$

En esta ecuación, (r_{t+n}) es el retorno en el tiempo $t+n$, $(\bar{r})$ la media de todo el periodo de observación del retorno y $\hat{h}_t(n)$ es el pronóstico de la varianza con origen t n pasos adelante. (Kosapattarapim, C.; Lin, Y.; MCCare, M., 2011, pp.9-21).

Para el calculo del ECM Proxy en la red, $\hat{h}_t(n)$ es el resultado obtenido después de simular la red con menor ECM 30 veces y calcular la varianza estimada para cada punto del tiempo, mientras que en el modelo EGARCH $\hat{h}_t(n)$ es la varianza pronosticada n pasos adelante, calculada mediante.

$$S^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1} \quad (8)$$

2.2. Redes neuronales artificiales

Las Redes Neuronales Artificiales (RNA o ANN por sus siglas en inglés, Artificial Neural Networks), son sistemas de aprendizaje automático, cuya estructura y funcionamiento están inspirados en las dinámicas de la corteza cerebral, aunque poseen otras funcionalidades y estructuras de conexión distintas a las vistas desde la perspectiva biológica. Una red neuronal artificial es una unidad de proceso masivo que tiene una propensión natural para almacenar conocimiento y posteriormente usarlo para hacer predicciones o clasificaciones.

Las redes neuronales son principalmente usadas de cuatro formas: modelos de sistemas nerviosos biológicos, inteligencia artificial, adaptador en tiempo real de procesos simples o implemento de control en hardware para aplicaciones tales como: robótica, análisis de datos y reconocimiento de patrones (Rodríguez, E;Pérez, E; Ovalle, Á;Domínguez, S.,2002,pp.31- 37).

Ya en la práctica, cada neurona recibe una serie de entradas, las cuales viajan por la red neuronal interconectada, para finalmente, emitir una salida, como lo indica (Sepúlveda,P,2011), una neurona artificial puede ser modelada con tres elementos básicos

1. Una serie de sinapsis o entradas, cada una está caracterizada por un peso que puede tomar tanto valores negativos como positivos.
- 2.Función de transferencia: consiste en el sumatorio de cada entrada multiplicada por su peso sináptico correspondiente. Si el peso es positivo, la conexión se denomina excitatoria; si es negativo, se denomina inhibitoria.

$$\varphi_j = \sum_{i=1}^n W_{ij} * X_i + \theta \quad (9)$$

Donde φ_j es la señal de entrada a la neurona j, W_{ij} los pesos sinápticos que llegan a la neurona j desde cada entrada X_i y n el número de entradas.

3. Función de activación: se aplica al valor devuelto por la función de transferencia y se utiliza para acotar la salida de la neurona, generalmente el conjunto de valores de salida está en un rango determinado a un intervalo cerrado $[0,1]$ o alternativamente $[-1,1]$. Algunas de las más utilizadas son: la función escalón, logística, tangente hiperbólica ,gaussiana, polinomial, Gompertz entre otras.

1. Función logística

El uso de esta función sería asignar todas las entradas en el dominio de número real en el intervalo de 0 a 1, donde los valores altos tienden de manera asintótica a 1 y los valores muy bajos tienden de manera asintótica a 0.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (10)$$

2. Tangente hiperbólica - Tanh

La función tangente hiperbólica transforma los valores introducidos a una escala entre -1 y 1, donde los valores altos tienden de manera asintótica a 1 y los valores muy bajos tienden de manera asintótica a -1.

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (11)$$

3. Gompertz

La función Gompertz transforma los valores introducidos a una escala de 0, 1, donde los valores altos tienen de manera asintótica a 1 y los valores muy bajos tienden de manera asintótica a 0, es una función que describe el crecimiento como más lento al comienzo y al final de un período de tiempo dado.

$$f(x) = e^{-e^{-x}} \quad (12)$$

Una red neuronal artificial tiene entradas y salidas que se conectan a otras neuronas, en este caso llamadas nodos que se catalogan de la siguiente manera: **nodos de entrada** aquellos que reciben información del exterior y **nodos de salida** aquellos que transmiten la información al exterior, además están aquellos que no tienen ningún contacto con el exterior y solamente intercambian información con otros, a estos se les llaman **nodos ocultos**, que están organizados por capas y se encuentran fuertemente interconectados entre sí. Estos nodos adquieren valores de entrada que se van modificando en un proceso llamado “aprendizaje”, existen diferentes algoritmos de aprendizaje para identificar qué conexiones serán más o menos significativas. De esta manera se toma una decisión sobre qué entrada es más importante con base al modelo que se está aplicando, una vez obtenida la estructura de la red, se pasa a la fase de aprendizaje y entrenamiento donde por medio de patrones objetivos, los valores en los nodos se ajustan de forma interactiva hasta que se muestren respuestas satisfactorias, es decir, basándose en ello la red sabrá cuándo crear, destruir o modificar los nodos; para este proceso podemos distinguir tres tipos de aprendizaje.

2.3. Tipos de aprendizaje

2.3.1. Aprendizaje supervisado

La red dispone de patrones de entrada y de salida que se quieren obtener para una determinada entrada de información y en función de ello se modifican los nodos ocultos para ajustar la entrada a la salida, este será capaz de predecir el valor correspondiente a cualquier objeto de entrada después de haber visto una serie de ejemplos; el objetivo de la red es que la salida de la red esté lo más cercana posible a la salida deseada.

2.3.2. Aprendizaje no supervisado

Consiste en no proporcionar a la red los patrones de salida, sino solo los de entrada y dejar que la red los clasifique en función de características comunes que encuentran entre ellos.

2.3.3. Aprendizaje híbrido

Donde no se proporcionan los patrones objetivos, sino que solo se le dice si la respuesta acierta o falla ante un patrón de entrada.

De acuerdo Matich, D.,2001,pp.8-10) algunas de las ventajas que presentan las redes neuronales artificiales son:

- **Aprendizaje Adaptativo:** tiene la capacidad de aprender a realizar tareas basadas en un entrenamiento o en una experiencia inicial, esta puede aprender a diferenciar patrones mediante ejemplos y entrenamientos, por lo cual no es necesario elaborar modelos a priori ni se necesita especificar funciones de distribución de probabilidad, además tampoco requiere de un algoritmo para resolver un problema, ya que ella puede generar su propia distribución de pesos en los enlaces mediante el aprendizaje.
- **Auto organización:** una RNA puede crear su propia representación de la información recibida mediante una etapa de aprendizaje adaptativo, la autoorganización consiste en la modificación de la red neuronal completa para llevar a cabo un objetivo específico.
- **Tolerancia a fallos:** la destrucción parcial de una red conduce a una degradación de su estructura; sin embargo, algunas capacidades de la red se pueden retener, por lo cual esta puede seguir respondiendo de manera aceptable aun si esta se daña parcialmente.
- **Operación en tiempo real:** los cómputos neuronales pueden ser realizados en paralelo, con el fin de obtener una respuesta en tiempo real, para esto se diseñan y fabrican máquinas con hardware especial que permita obtener esta capacidad.

Por otro lado, (Andrade, E.,2013,p.46). expone las siguientes desventajas para las RNA:

- **Complejidad de aprendizaje para grandes tareas:** cuantas más cosas se necesiten que aprenda una red, más complicado será enseñarle.
- **Tiempo de aprendizaje elevado:** el cual depende de si se incrementa la cantidad de patrones a identificar o clasificar y si se requiere mayor flexibilidad o capacidad de adaptación de la red neuronal para reconocer patrones muy similares, se tiene que invertir más tiempo en lograr unos pesos adecuados que permitan representar lo que se quiere enseñar a la red neuronal.
- **No permite interpretaciones:** la salida de la red es un número, que no puede ser interpretado por ella misma, sino que se requiere de la intervención del programador y de la aplicación para darle un significado a dicho valor.
- **Elevada cantidad de datos para el entrenamiento:** para que una red aprenda de forma correcta a identificar patrones, se requerirá de una mayor información, lo que adicional implica más tiempo del CPU.

2.4. Topología de las redes neuronales

2.4.1. Red neuronal Monocapa – Perceptrón simple

Una red monocapa es correspondiente “con la red neuronal más sencilla, ya que se tiene una capa de neuronas que proyectan las entradas a una capa de neuronas de salida donde se realizan los diferentes cálculos” (Nacelle, A .,2009,pp.5-7). Esta red neuronal normalmente es usada en tareas relacionadas con la autoasociación, por ejemplo, para regenerar información de entrada que se presenta como incompleta o distorsionada.

2.4.2. Red neuronal Multicapa – Perceptrón multicapa

Es una generalización de la red neuronal monocapa, la diferencia está en el conjunto de capas intermedias, también llamadas capas ocultas, que están entre la capa de entrada y la de salida, estas capas adicionales le permiten establecer regiones de decisión más complejas.

2.4.3. Redes de base radial (RBF)

Se caracterizan principalmente por ser una red multicapa de propagación hacia adelante, esta calcula la salida en función de la distancia a un punto denominado centro, en este tipo de red solo se presenta una capa oculta. Aquí las salidas son las que se encargan de realizar una combinación lineal en los resultados obtenidos de las activaciones de las neuronas de la capa oculta. (Buitrago, V.,2012,pp.11-12).

En ese sentido y de acuerdo con el artículo redes de base radial (González,M, 2004,p.13) publicado en el 2004 por la revista Varpa de la universidad de La Coruña; este tipo de red tiene la característica de aprendizaje más rápido, ya que el cambio de peso solo afecta a la neurona oculta, además este tipo de red es menos sensible al orden de presentación de patrones.

2.4.4. Red neuronal recurrente (RNN)

Se denominan recurrentes, ya que realizan la misma tarea para cada elemento de una secuencia, este tipo de red, permiten conexiones de las neuronas consigo mismas y con neuronas en capas inferiores, no tiene una estructura de capas definida, sino que permiten conexiones arbitrarias entre las neuronas, consiguiendo crear la temporalidad, lo cual permite que la red tenga memoria, haciéndola una herramienta apropiada para modelar series temporales.

Estas redes son un intento de establecer una correspondencia entre secuencias de entrada y de salida que no son más que patrones temporales. “Se puede clasificar en parcial y/o totalmente recurrente. Las totalmente recurrentes son aquellas que cada neurona puede estar conectada a cualquier otra y sus conexiones recurrentes son variables. Las redes parcialmente recurrentes son aquellas que sus conexiones recurrentes son fijas”. (Rodríguez, E;Pérez, E; Ovalle, Á;Domínguez, S.,2002,pp.48-57)

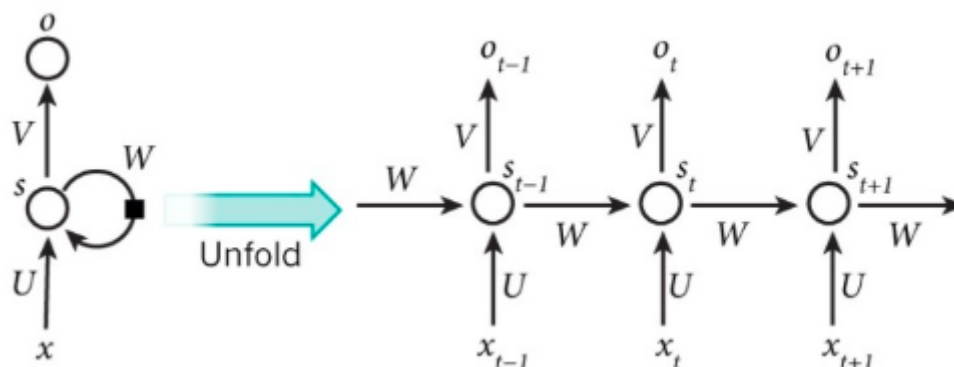


Figura 1: (IA-LATAM 2019)

Debido a las características expuestas de las redes neuronales recurrentes, este tipo de red es seleccionado para aplicar al pronóstico de los rendimientos de la serie de Ecopetrol, ya que es la más adecuada para evaluar series temporales.

3. Objetivos

3.1. Objetivo General

Realizar el pronóstico de los rendimientos en la acción de Ecopetrol con el método de Red Neuronal Artificial.

3.2. Objetivos Específicos

- Determinar si la utilización de variables explicativas permiten mejorar el pronóstico de la serie de interés.
- Comparar los resultados obtenidos mediante los algoritmos de redes neuronales con el método clásico (EGARCH).

4. Metodología

El proceso metodológico usado para el desarrollo del trabajo es el siguiente:

1. La serie de datos histórica del precio de cierre de la acción de Ecopetrol, fue obtenidas de la página del Grupo AVAL (<https://www.grupoaval.com/wps/portal/grupo-aval/aval/portal-financiero/renta-variable/acciones-bolsa-colombia/resumen-dia>); los datos fueron tomados desde el 4 enero del 2010 hasta el 30 de agosto del 2019, es una serie diaria.
2. Estimación de datos faltantes: las series descargadas, presentaban datos faltantes, los cuales fueron estimados a través de TSW, software de gran potencia, que permite la realización de una gran cantidad de procedimientos de forma automática, entre estos la estimación de datos faltantes. TSW está basado en los programas TRAMO y SEATS; este software fue desarrollado en el Banco de España por Gianluca Caporello y Agustín Maravall.
3. Se toma una muestra de los datos como entrenamiento, desde el 04 de enero del 2010, hasta el 9 agosto del 2019; los 15 días restantes serán utilizado para la predicción, donde se evaluará el desempeño de la red, del de 12 de agosto, al 30 de agosto.
4. La estructura empleada fue la de red neuronal totalmente recurrente.
5. Se tomo como variables explicativas la TRM, EURO, el precio el crudo Brent y el crudo WTI.

5. Resultados

A continuación se presenta el gráfico de los precios de la acción.



Figura 2: Precio cierre de la acción de Ecopetrol

Los rendimientos logarítmicos se expresan como:

$$R_i = \ln\left(\frac{P_i}{P_j}\right) \quad (13)$$

Siendo P la serie de precios, y cumpliendo que $i > j$

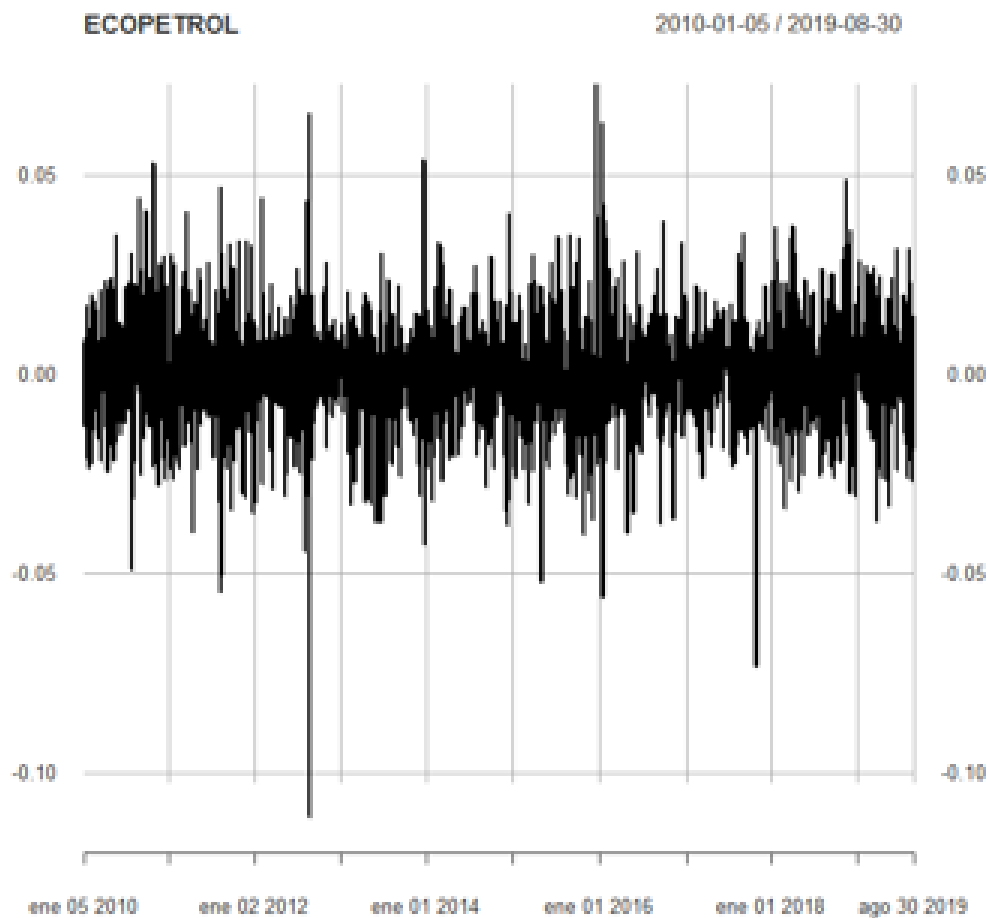


Figura 3: Rendimientos diarios de la acción de Ecopetrol

En la figura tres se puede observar los rendimientos logarítmicos diarios de la acción de Ecopetrol desde el 5 de enero del 2010 hasta el 30 de agosto del 2019, donde se puede apreciar que la media de los rendimientos gira entorno a cero, esta cuentan con la presencia de valores atípicos y algunos conglomerados de volatilidad.

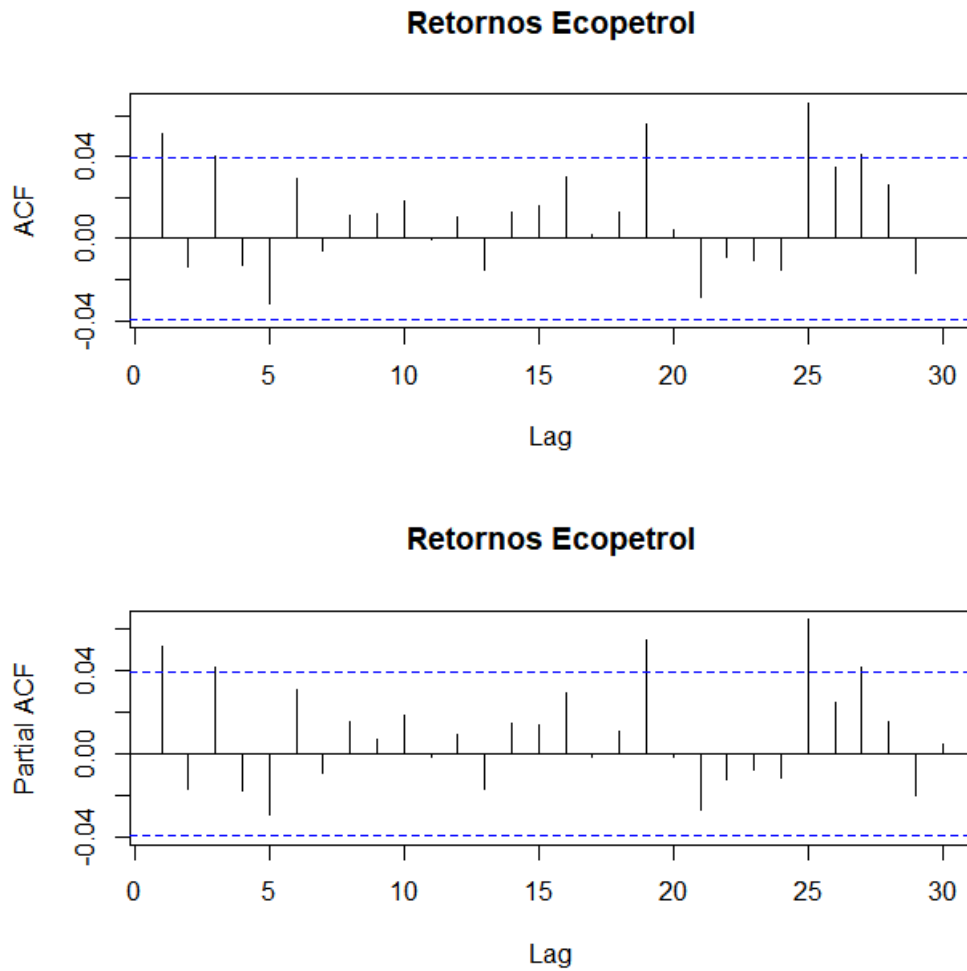


Figura 4: Autocorrelaciones de los retornos

De los gráficos de autocorrelación simple y parcial de los retornos, parece ser que hay dos correlaciones significativas en los primeros rezagos, al rezago uno y tres, para corroborar esta información se realiza la estadística Q Ljung-Box la cual verifica si las primeras (m) autocorrelaciones, son simultáneamente cero.

Hipótesis de la Prueba

$H_0 : \rho_1 = \rho_2 = \dots = \rho_m = 0$ (Las primeras m autocorrelaciones son simultáneamente cero)

$H_1 : \rho_i \neq 0$ (Al menos una autocorrelación es diferente de cero)

$$Q_{LB} = T(T+2) \sum_{k=1}^m \frac{1}{T-k} \hat{\rho}^2(k) \sim X_m^2 \quad (14)$$

Con un $m = 15$ el p-value es 0.1787 por lo tanto a un nivel de significancia del 5 %, no hay evidencia suficiente para rechazar la hipótesis nula de que en los primeros quince rezagos hay al menos una correlación significativa.

Con un $m = 8$ el se evidencia un p-value menor al 5 % ($p - value = 0.03344$), por lo cual se rechaza la hipótesis nula indicando que hay al menos una correlación significativa en los primeros 8 rezagos, debido

a esto se justifica hacer un modelo para la media de los datos, por lo tanto se explorara un posible modelo para la media que capture la dependencia de los rezagos uno y tres que se observan en los gráficos de autocorrelación simple y parcial de los retornos, inicialmente se probara con un modelo $MA(1)$. Después de una breve exploración de los residuales, el modelo $MA(1)$ genera buenos resultados.

Evaluando si existe heterocedasticidad condicional

En esta sección se determina si existe o no heterocedasticidad condicional mediante el análisis de los gráficos de autocorrelación simple y parcial de los residuales al cuadrado y con el uso de la prueba de McLeod-Li

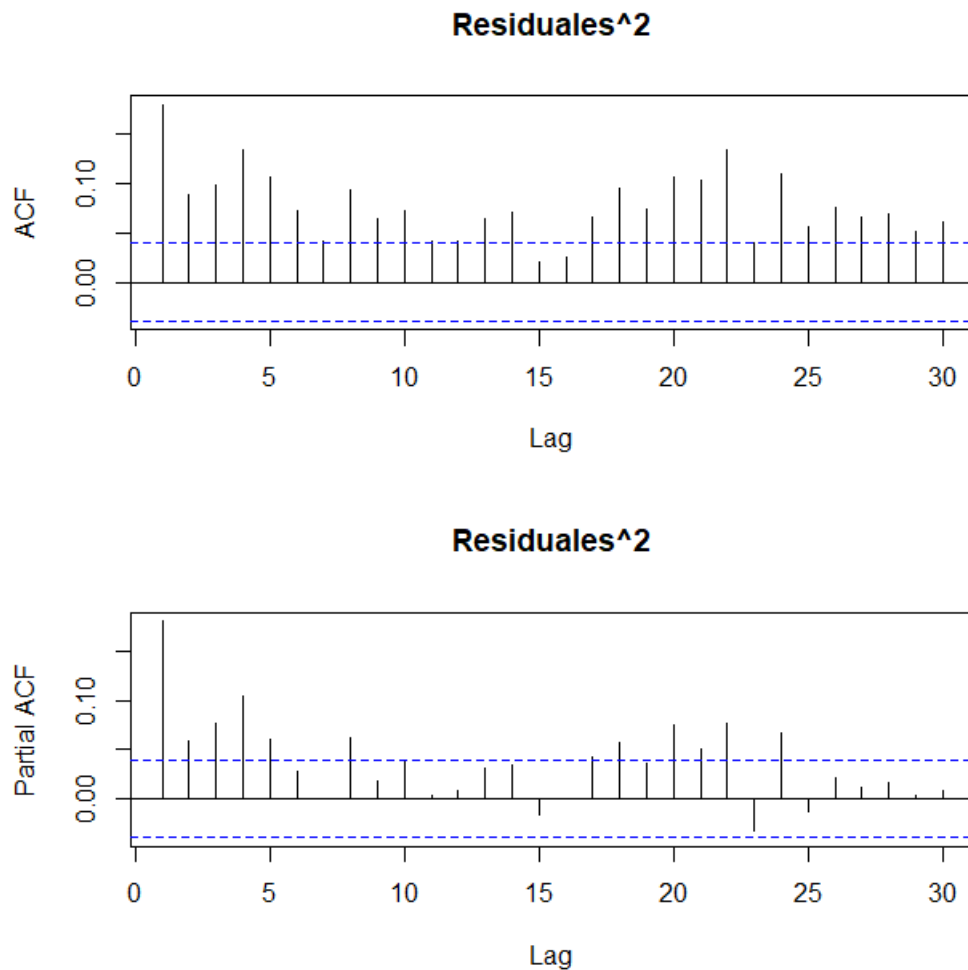


Figura 5: Autocorrelación de los residuales al cuadrado

El gráfico muestra una correlación alta en los cuadrados de los residuales, evidencia de heterocedasticidad condicional, para corroborar esta información, se realiza una prueba de hipótesis de McLeod-Li la cual es estadística Q Ljung-Box para los *residuales*²

Hipótesis de la Prueba

$H0 : \rho_1 = \rho_2 = \dots = \rho_m = 0$ (Las primeras m autocorrelaciones de los *residuales*² son cero)

$H1 : \rho_i \neq 0$ (Al menos una autocorrelación de los *residuales*² es diferente de cero)

test McLeod-Li

$X - squared = 294.59, df = 15, p - value < 2.2e - 16$

$X - squared = 238.32, df = 8, p - value < 2.2e - 16$

Dado que el p-value es menor al 5 % el se rechaza la hipótesis nula, esto prueba que hay heterocedasticidad condicional en los datos, justificando un modelo de volatilidad.

El modelo seleccionado fue un MA(1) EGARCH(1,1), un modelo sencillo para la media, ya que la correlación de los retornos era pequeña.

MODELO EGARCH PARA LA SERIE DE LOS RETORNOS

En esta sección se presenta el modelo matemático y los resultados generados por este modelo.

Tabla 1: Coeficientes del modelo

mu	0.0002630922
ma1	0.0547803001
omega	-0.3629787568
alpha1	-0.0368274184
beta1	0.9560694256
gamma1	0.2838826112
shape	4.4393474572

Se escribe el modelo

$$\begin{cases} r_t = 0.000262 + 0.05478 a_{t-1} + a_t \\ a_t = \sigma_t * \epsilon_t \\ \epsilon_t \sim iid t4(0, 1) \\ Ln(\sigma_t^2) = -0.36297 + 0.95606 \ln(\sigma_{t-1}^2) - 0.03682(\epsilon_{t-1}) + 0.28388(|\epsilon_{t-1}|) - E(|\epsilon_{t-1}|) \end{cases} \quad (15)$$

Dado que $\epsilon_t \sim t_4$ se calcula $E|\epsilon_t|$. De acuerdo a la fórmula 6 de la sección 2.1.

$$E(|\epsilon_t|) = \frac{2\sqrt{4.439 - 2}\Gamma(\frac{4.439+1}{2})}{(4.439 - 1)\Gamma(\frac{4.439}{2})\sqrt{\pi}} = 0.721966 \quad (16)$$

Se reescribe el modelo de la siguiente manera después de agrupar algunas constantes y abriendo el valor absoluto ($|\epsilon_{t-1}|$). La ventaja de esta nueva expresión, es que se puede observar como se calcula la asimétrica de la volatilidad ante choques positivos y ante choques negativos.

$$Ln(\sigma_t^2) = -0.5679324 + 0.95606 \ln(\sigma_{t-1}^2) + \begin{cases} (-0.03682 + 0.28388)\epsilon_{t-1} & si \ \epsilon_{t-1} \geq 0 \\ (-0.03682 - 0.28388)\epsilon_{t-1} & si \ \epsilon_{t-1} < 0 \end{cases} \quad (17)$$

$$Ln(\sigma_t^2) = -0.5679324 + 0.95606 \ln(\sigma_{t-1}^2) + \begin{cases} (0.2470552)\epsilon_{t-1} & si \ \epsilon_{t-1} \geq 0 \\ (-0.32071)\epsilon_{t-1} & si \ \epsilon_{t-1} < 0 \end{cases} \quad (18)$$

Para considerar una expresión de σ_t^2 podemos comparar la volatilidad estimada en el tiempo t ante un choque negativo $\epsilon_{t-1} = -1$ y la volatilidad estimada ante un choque positivo de la misma magnitud, de esta forma calculamos el porcentaje:

$$\frac{\sigma_t^2 \text{ si } \epsilon_{t-1} = -1}{\sigma_t^2 \text{ si } \epsilon_{t-1} = 1} \quad (19)$$

Tomando exponencial a ambos lados

$$\frac{e^{-0.32071*(-1)}}{e^{0.2470552*(1)}} = 1.076435 \quad (20)$$

Por consiguiente el impacto sobre la volatilidad de un choque negativo, es de alrededor del 7.64 % más alto que un choque positivo del mismo tamaño

Verificación de residuales

Como veremos a continuación este modelo logra capturar la poca dependencia lineal que había en los datos a través del modelo de la media y la heterocedasticidad condicional a través del modelo para la varianza.

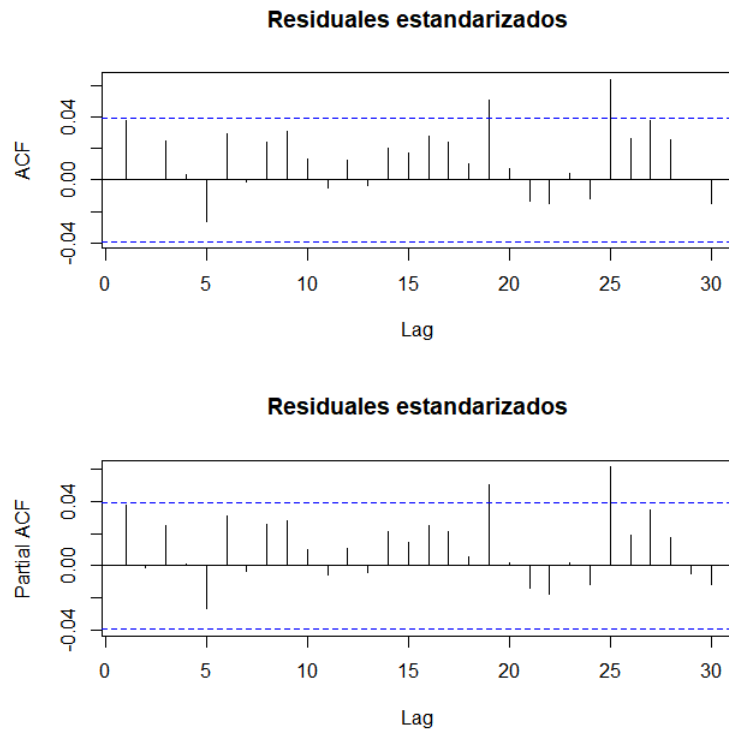


Figura 6: Residuales estandarizados

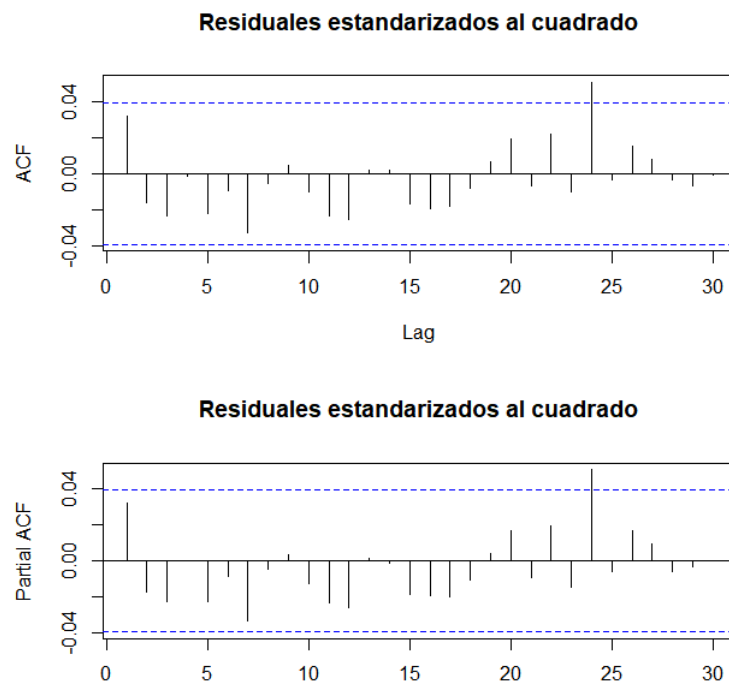


Figura 7: Residuales estandarizados al cuadrados

Por medio del modelo se capturó la heterocedasticidad condicional de los datos, la correlación que está siendo significativa, al rezago de veinticuatro días no se tendrá en cuenta, a pasear de ser significativa su correlación es pequeña y su rezago es muy lejano. Se realiza una prueba t para evaluar si la media de los Residuales estandarizados es cero o no.

$H_0 : M = 0$ (La media verdadera es igual a 0)

$H_1 : M \neq 0$ (La media verdadera no es igual a 0)

$p - value = 0.4572$, no hay evidencia para rechazar H_0 lo cual indica que la media de los residuales estandarizados es igual a cero.

Prueba de Autocorrelación

test Ljung-Box:

$df = 8, p - value = 0.2282$

No se rechaza la hipótesis nula, las primeras 8 autocorrelaciones de los Residuales estandarizados son cero.

$df = 15, p - value = 0.4046$

No se rechaza la hipótesis nula, las primeras 15 autocorrelaciones de los Residuales estandarizados son cero.

Con lo anterior se valida la ecuación de la media.

Prueba de heterocedasticidad

test McLeod-Li $X - squared = 8.7758, df = 8, p - value = 0.3616$

No se rechaza la hipótesis nula, la ecuación de la varianza captura de forma adecuada la heterocedasticidad condicional.

Face de Pronósticos

	Series	Sigma
T+1	0.0004480	0.01093
T+2	0.0002631	0.01112
T+3	0.0002631	0.01130
T+4	0.0002631	0.01147
T+5	0.0002631	0.01164
T+6	0.0002631	0.01181
T+7	0.0002631	0.01197
T+8	0.0002631	0.01213
T+9	0.0002631	0.01228
T+10	0.0002631	0.01242
T+11	0.0002631	0.01256
T+12	0.0002631	0.01270
T+13	0.0002631	0.01283
T+14	0.0002631	0.01296
T+15	0.0002631	0.01308

Figura 8: Pronósticos

$ECMPROXY = 0,0000001698$

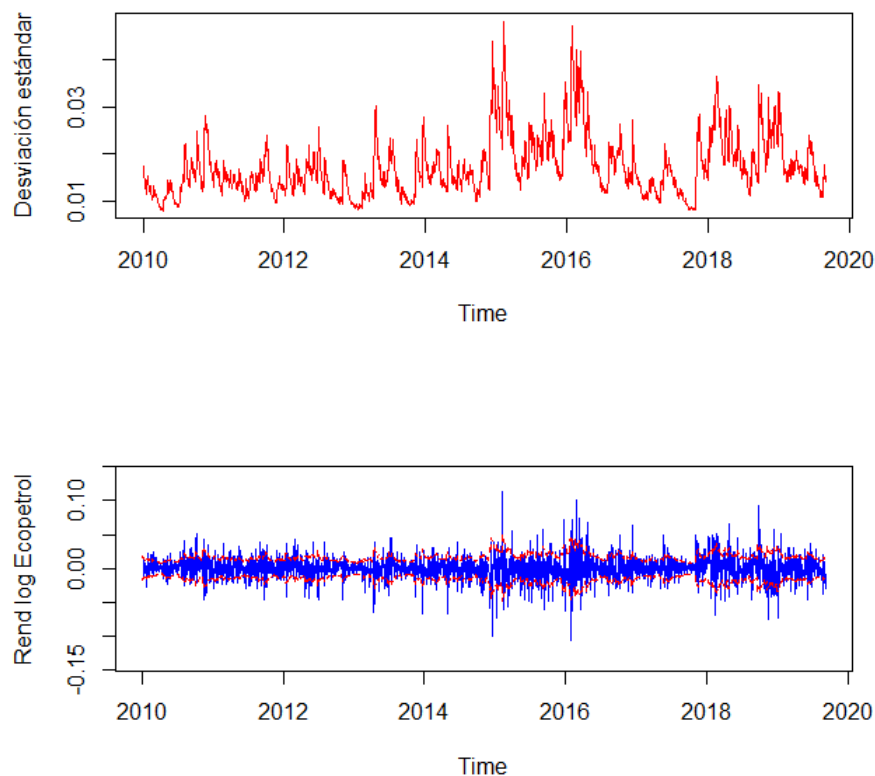


Figura 9: Desviación / Rendimientos

MODELOS DE RED RECURRENTE PARA LA SERIE DE LOS RETORNOS

En esta sección se presentaran los resultados obtenidos mediante las redes neuronales recurrentes aplicadas a la serie de los retornos de Ecopetrol.

MODELOS DE RED RECURRENTE PARA LA SERIE DE LOS RETORNOS CON VARIABLES AUXILIARES.

Para este modelo, además de utilizar los retornos rezagados hasta 3 días atrás, también se utilizaron las siguientes variables explicativas.

- PrecioR1, PrecioR2, PrecioR3: Precio de cierre de la acción con uno, dos y tres rezagos.
- F.P.BrentR, F.P.BrentR2, F.P.BrentR3: Precio de cierre del crudo Brent con uno, dos y tres rezagos.
- F.P.C.WTIR, F.P.C.WTIR2, F.P.C.WTIR3: Precio de cierre del crudo WTI con uno, dos y tres rezagos. .P.Brent1: Rendimiento del crudo Brent con un rezago. .P.C.WTI1: Rendimiento del crudo WTI con un rezago.
- TRMR1, TRMR2, TRMR3: Precio de cierre de la TRM con uno, dos y tres rezagos.
- euroR1, euroR2, euroR3: Precio de cierre de la EUR/COP con uno, dos y tres rezagos.
- TRM1: Rendimiento de la TRM con un rezago. 1: Rendimiento del EUR/COP con un rezago .

TRM “La tasa de cambio representativa del mercado (TRM) es la cantidad de pesos colombianos por un dólar de los Estados Unidos. La TRM se calcula con base en las operaciones de compra y venta de divisas entre intermediarios financieros que transan en el mercado cambiario colombiano, con cumplimiento el mismo día cuando se realiza la negociación de las divisas.” (Banco de la República 2019, p.1) Euro Corresponde al cambio entre el par de divisas peso colombiano y euro

Según el portal de noticias (dailyfx) el cual proporciona información y material sobre los mercados financieros, indica que las mezclas más negociadas son el crudo Brent y el crudo WTI. el primero se extrae de campos petroleros en el Mar del Norte y el segundo se extrae de campos petroleros en los Estados Unidos, principalmente en Texas, Louisiana y Dakota del Norte.(dailyfx,2018).

Calibración del modelo

Selección de la tasa de aprendizaje

Para la calibración de la red se escogen dos capas con 50 neuronas cada una y varia la tasa de aprendizaje desde 0.01 a 0.05, el número de iteraciones será de 100 para todas las redes aplicadas de aquí en adelante, para todos los casos se calcula el error cuadrático medio.

Error cuadrático medio:

$$ECM = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2 \quad (21)$$

Donde $\hat{Y}_i$ es un vector de predicciones, Y_i es el vector de los valores verdaderos y n es el número de predicciones realizadas.

El ECM que se tiene en cuenta para la calibración de la red es el Error Cuadrático Medio Train, ya que este indica que tan eficiente fue la red en los datos tomados como entrenamiento, evidenciando cual sería el mejor modelo para emplear en datos futuros.

Selección de la Tasa de aprendizaje 1

Tabla 2: Tasa de aprendizaje

Tasa	0,01	0,02	0,03	0,04	0,05
Error Cuadrático Medio Train	0,01165284	0,01328577	0,01328577	0,01328577	0,01165283
Error Cuadrático Medio Test	0,01109806	0,01386668	0,01386676	0,01386675	0,01109806

La tasa seleccionada fue la de 0.05, ya que es la que presenta un menor ECM.

Selección de la función de activación

Para este se corren la red neuronal con mejor comportamiento en el pasado anterior, variando la función de activación

Tabla 3: Función de activación

Función de activación	logistica	Gompertz,
Error Cuadrático Medio Train	0,01165263	0,01165284
Error Cuadrático Medio Test	0,01109788	0,01109806

La función de activación seleccionada es la logisticaen, el caso de la tangente hiperbólica, no arroja resultados para esta red.

Selección del número de capas

Para esto se escogen una tasa de aprendizaje de 0.01 y función de activación logistica y se varia el número de capas.

Tabla 4: Tasa de aprendizaje

RED	5 capas	10 capas	15capas	20 capas
Error Cuadrático Medio Train	0,01328577	0,01165284	0,01165284	0,01165284
Error Cuadrático Medio Test	0,01386676	0,01109806	0,01109806	0,01109806

El tipo de red Recurrente seleccionada fue una con 0.05 de tasa de aprendizaje, función logistica de 10 capas cada una con 50 neuronas. A continuación se presentan las estimaciones de este tipo de red.

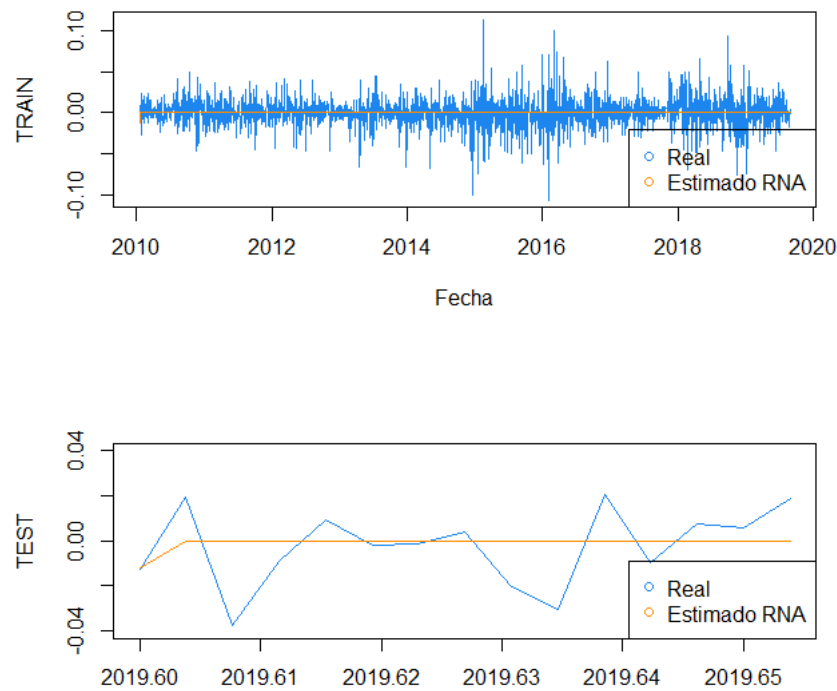


Figura 10: Real/ Estimado

A pesar de ser esta la mejor red obtenida del entrenamiento, sus predicciones no fueron buenas como se muestra en el gráfico anterior, por lo cual se decide explorar la función de activación tangente hiperbólica, se probaron otras combinaciones de red con diferente número de capas para esta función, es de aclarar que no todas funcionaban, ya que con demasiadas capas o neuronas esta no converge, algunas de las que convergieron se presentan continuación .

Tabla 5: Función de activación tangente hiperbólica

tangente hiperbólica	1 capa con 1 neurona	2 capa con 1 neurona
Error Cuadrático Medio Train	0,00058964	0,00101103
Error Cuadrático Medio Test	0,0008492	0,00081732

Otras combinaciones empleadas fueron : (50-50) (10-10) (5-5) (3-3) (2-2) (2-1) (1-2) (4-0) (3-0) (2-0).

El resultado generado por la Función de activación tangente con una tasa de aprendizaje de 0.01, una capa oculta y una neurona se presenta a continuación.

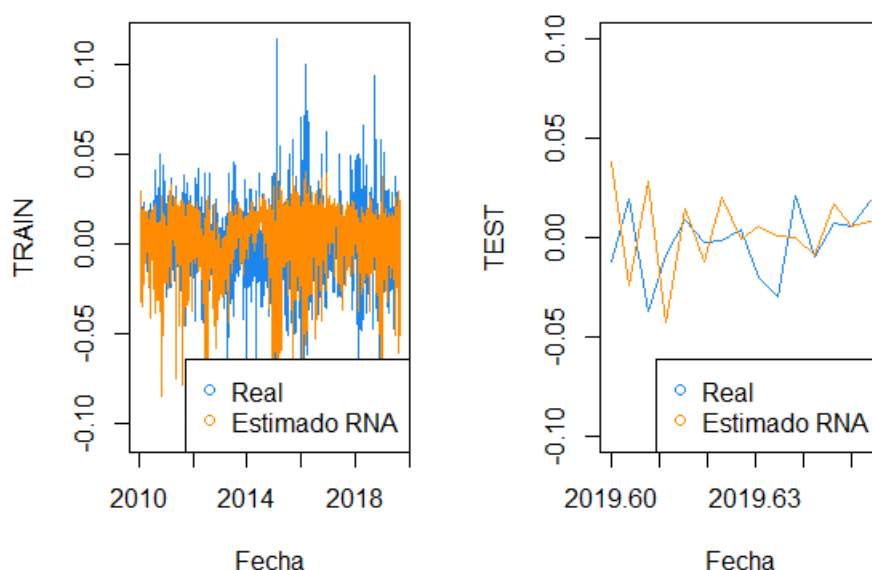


Figura 11: Real / Estimado

Debido a la mejora en los resultados, este tipo de red es seleccionado para compararlo contra el modelo EGARCH, en esta comparación se necesita calcular la varianza y la desviación estándar, por lo cual se ejecuta 30 veces esta RNN y con las 30 estimaciones obtenidas, se calcula la varianza y la desviación estándar para cada día.

Comparación entre los resultados obtenidos del EGARCH y el modelo con variables auxiliares de red recurrente

Tiempo	valor Real	Modelo de RRN con variables auxiliares				EGARCH	
		Mejores resultado	Aciertos	series	Sigma	series	Sigma
T+1	-0,01242	0,04239	0	0,00107	0,032779	0,00045	0,01093
T+2	0,01945	-0,01172	0	0,00180	0,04237	0,00026	0,01112
T+3	-0,03747	0,02338	0	0,00119	0,03443	0,00026	0,01130
T+4	-0,00913	0,00288	0	0,00164	0,040546	0,00026	0,01147
T+5	0,00913	0,02457	1	0,00156	0,039458	0,00026	0,01164
T+6	-0,00216	-0,02623	1	0,00148	0,03845	0,00026	0,01181
T+7	-0,00148	0,02667	0	0,00121	0,03483	0,00026	0,01197
T+8	0,00364	-0,00273	0	0,00134	0,03923	0,00026	0,01213
T+9	-0,02020	0,00413	0	0,00109	0,03299	0,00026	0,01228
T+10	-0,03130	0,00956	0	0,00140	0,03743	0,00026	0,01242
T+11	0,02081	0,01661	1	0,00115	0,03395	0,00026	0,01256
T+12	-0,00941	-0,00564	1	0,00157	0,03965	0,00026	0,01270
T+13	0,00753	0,02310	1	0,00112	0,03350	0,00026	0,01283
T+14	0,00561	0,00122	1	0,00165	0,04066	0,00026	0,01296
T+15	0,01848	-0,01059	0	0,00106	0,03261	0,00026	0,01308
Total		40,00%	6				
Ecuadrático medio Proxy		0,000001412815				0,000000169770	
Ecuadrático medio Proxy*1000000		1,412815159000				0,169770015000	

Figura 12: Modelo RNN 1 - EGARCH

Las estimaciones del modelo que presento un menor ECM al recrear las 30 veces el tipo de red seleccionado, se muestra en la figura 12, en la columna de mejores resultados, donde se puede apreciar que este

modelo, logro estimar bien 6 de las 15 veces la dirección del rendimiento, lo que representa el 40 %. Si se compara el ECM-Proxy de ambos modelos, se puede concluir que el modelo con variables auxiliares de red recurrente no obtuvo unos resultados tan eficientes, ya que este presenta un ECM-Proxy mayor, esto indica que no logra captar de una buena manera la volatilidad de la serie de los retornos. Como se puede ver en la figura 11, el modelo no logra captar los valores lejanos a la media.

MODELOS DE RED RECURRENTE PARA LA SERIE DE LOS RETORNOS SIN LAS VARIABLES AUXILIARES

Para este modelo se utilizó los retornos rezagados hasta 3 días atrás.

Con el fin de determinar que función de Activación es apropiada usar, se emplea los errores cuadráticos medios.

Tabla 6: Función de Activación

Función de Activación	logística	Gompertz	tangente hiperbólica
Error Cuadrático Medio Train	0,000299	0,000303	0,000427
Error Cuadrático Medio Test	0,000294	0,00028	0,000542

Gompertz fue la función activación que presento un menor error cuadrático, para comparar la eficiencia de la red, a continuación se presenta las estimaciones dentro y fuera de muestra con los tres tipos diferentes de función Activación .

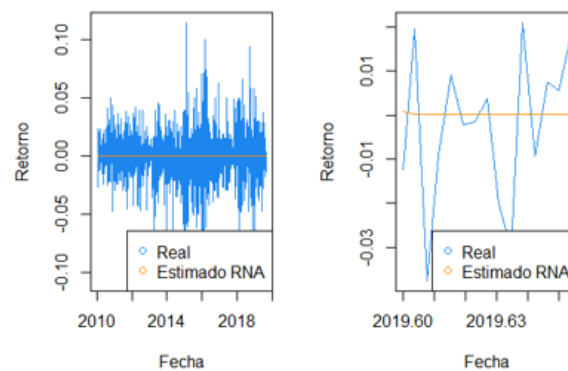


Figura 13: función de activación logística

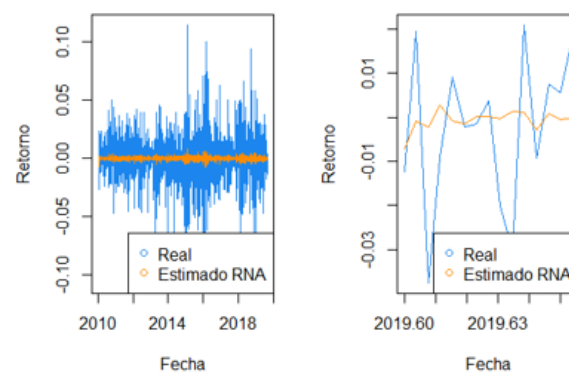


Figura 14: función de activación Gompertz

A pesar de que las Funciones de activación logística y Gompertz fueron las que presentaron los errores cuadráticos más bajos, se puede apreciar en las gráficas de estimaciones, que con estas características, la red no logra captar la volatilidad de la serie de Ecopetrol.

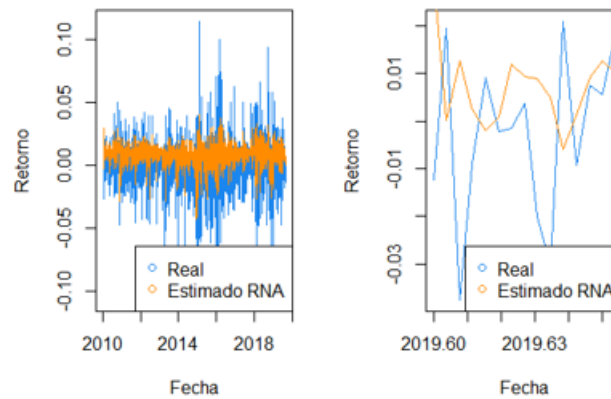


Figura 15: función de activación tangente hiperbólica

La función de activación tangente hiperbólica logra captar de una mejor manera la volatilidad de la serie, para poder compararla con el modelo EGARCH, se repite este proceso 30 veces con las mismas características, los resultados obtenidos, se presentan a continuación.

Comparación entre los resultados obtenidos del EGARCH y los modelos de redes recurrentes con variables auxiliares y sin ellas

Tiempo	valor Real	Modelo de RRN con variables auxiliares				Modelo de RRN sin variables auxiliares				EGARCH	
		Mejores resultados	Aciertos	series	Sigma	Mejores resultados	Adertos	series	Sigma	series	Sigma
T+1	-0,01242	0,04239	0	0,00107	0,032779	0,00594	0	0,00048	0,02195	0,00045	0,01093
T+2	0,01945	-0,01172	0	0,00180	0,04237	-0,00082	0	0,00059	0,02415	0,00026	0,01112
T+3	-0,03747	0,02338	0	0,00119	0,03443	0,00327	0	0,00046	0,02150	0,00026	0,01130
T+4	-0,00913	0,00288	0	0,00164	0,040546	-0,00484	1	0,00065	0,02556	0,00026	0,01147
T+5	0,00913	0,02457	1	0,00156	0,039458	-0,00472	0	0,00062	0,02497	0,00026	0,01164
T+6	-0,00216	-0,02623	1	0,00148	0,03845	-0,01029	1	0,00062	0,02493	0,00026	0,01181
T+7	-0,00148	0,02667	0	0,00121	0,03483	0,00143	0	0,00048	0,02187	0,00026	0,01197
T+8	0,00364	-0,00273	0	0,00154	0,03923	0,00323	1	0,00048	0,02196	0,00026	0,01213
T+9	-0,02020	0,00413	0	0,00109	0,03299	0,00115	0	0,00048	0,02190	0,00026	0,01228
T+10	-0,03130	0,00956	0	0,00140	0,03743	-0,00214	1	0,00055	0,02335	0,00026	0,01242
T+11	0,02081	0,01661	1	0,00115	0,03395	-0,00949	0	0,00065	0,02542	0,00026	0,01256
T+12	-0,00941	-0,00564	1	0,00157	0,03965	-0,00810	1	0,00059	0,02419	0,00026	0,01270
T+13	0,00753	0,02310	1	0,00112	0,03350	-0,00319	0	0,00057	0,02378	0,00026	0,01283
T+14	0,00561	0,00122	1	0,00165	0,04066	0,00658	1	0,00047	0,02166	0,00026	0,01296
T+15	0,01848	-0,01059	0	0,00106	0,03261	0,00171	1	0,00048	0,02201	0,00026	0,01308
Total		40,00%	6					46,67%	7		
Ecuadratico medio Proxy		0,000001412815				0,000000226439				0,000000169770	
Ecuadratico medio Proxy*1000000		1,412815159000				0,226438521000				0,169770015000	

Figura 16: Modelos RNN y EGARCH

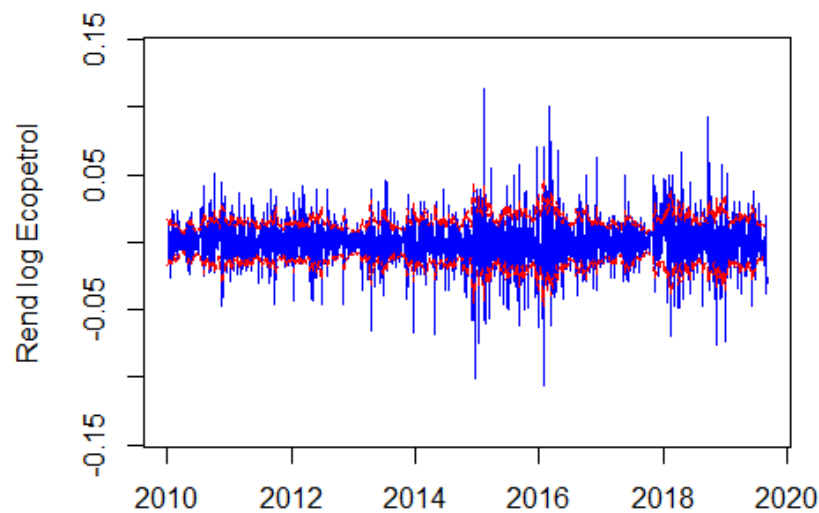


Figura 17: EGARCH

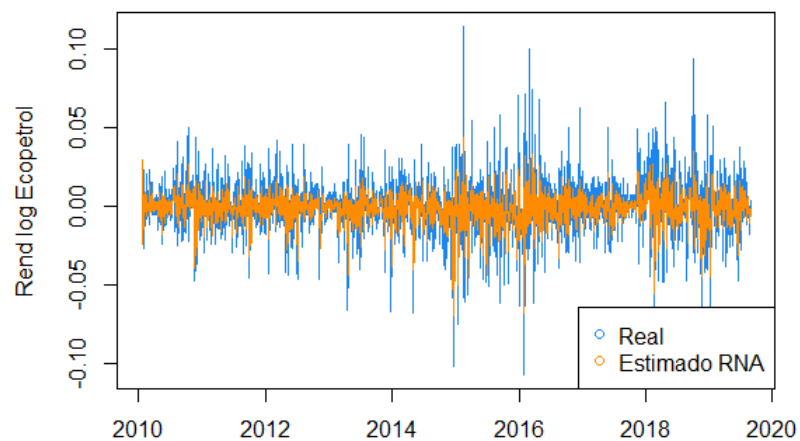


Figura 18: Modelo de RNN sin variables auxiliares

Como se puede observar en la figura 19, el modelo de RNN sin variables auxiliares logra captar cierta volatilidad de la serie diaria de los retornos de Ecopetrol y también algunos datos negativos lejanos de la media, la red logra estimar la dirección correcta del retorno en el test, el 46,67% de las veces, siendo un modelo más sencillo que el modelo con variables auxiliares de RNN, este obtiene mejores resultados, si se compara el Error cuadrático medio Proxy de las RNN y el del modelo Egarch, se puede observar que el generado por el modelo EGARCH sigue siendo menor.

6. Conclusiones

- Si bien las estimaciones generadas por las RRN empleadas no fueron exactas, este demuestra que si es posible aproximarse a los resultados, además se puede inferir que, con otro tipo de variables o configuraciones en la arquitectura de la red, se podrían obtener resultados más exactos.
- La implementación de variables explicativas para este caso, no permitió mejorar el pronóstico de la serie, ya que el modelo que únicamente dependía del retorno y el valor rezagado 1,2 y 3 pasos atrás, arrojó un menor ECM PROXY.
- El mejor modelo dado el ECM PROXY para los rendimientos diarios de la serie de Ecopetrol fue el modelo EGARCH.
- Se propone que en futuras investigaciones se considere implementar otros tipos de redes con diferentes algoritmos de aprendizaje para identificar si hay alguna mejora la precisión de las estimaciones a través de nuevos diseño en los modelos.

7. Trabajo Futuro

- Se propone que en futuras investigaciones se considere implementar otros tipos de redes con diferentes algoritmos de aprendizaje para identificar si hay alguna mejora la precisión de las estimaciones a través de nuevos diseño en los modelos.

Referencias

- Abascal, M. (2016). *"Análisis de series temporales financieras"*, Universidad de Cantabria, España, (p.4).
- Andrade, E.(2013). *"Estudio de los principales tipos de redes neuronales y las herramientas para su aplicación."*,Universidad politecnica salesiana, Quito, (p.46).
- Banco de la República.(2019). *"Tasa Representativa del Mercado (TRM - Peso por dólar)."*,Banco de la Republica de Colombia, Borradores de Economía, (p.1).
- Rodríguez, E ;Pérez, E; Ovalle, Á;Domínguez, S.(2002). *"Aplicación de redes neuronales artificiales y técnicas sig para la predicción de coberturas forestales."*,Revista Chapingo. Serie Ciencias Forestales y del Ambiente, vol. 8, núm. 1 (Universidad Autónoma), México, (pp.31- 37,pp.48-57).
- Box, G;Jenkins,G .(1976). *"Times Series Análisis: Forecasting and Control."*
- Buitrago, V.(2012). *"Simulador de las redes de base radial."*,Universidad Carlos III, Escuela politécnica superior,Madrid, (pp.11-12).
- dailyfx .(2018). *"WTI vs Brent: Las 5 principales diferencias entre la mezcla de crudo WTI y Brent."*,<https://www.dailyfx.com/espanol/precio-del-petroleo/wti-vs-brent.html> (último acceso: 7 de 10 de 2019).
- Escobar,L;Valdes,H;Zapata,C.(2010). *"Redes neuronales artificiales en predicción de series de tiempo. Una aplicación a la industria. Chile"*,Universidad Tecnológica Metropolitana , (pp.9-21).
- González,M.(2004). *"Redes de base radial"*,Revista Varpa de la universidad de La Coruña , (p.13).
- Marín,J.(2012). *"Introducción a las Redes Neuronales Aplicadas."*,Universidad Carlos III,Madrid, (p.5).
- Matich, D.(2001). *"Redes Neuronales: Conceptos Básicos y Aplicaciones."*,Universidad Tecnológica Nacional – Facultad Regional Rosario,Argentina, (pp.8-10).
- Monsegny,M ; Cuervo,E.(2008). *"Modelos arch, garch y egarch: aplicaciones a series financieras."*,Cuadernos de Economía (scielo) 27, n° 48, (pp.3-4).
- Nacelle, A .(2009). *"Las redes neuronales: de la biología a los algoritmos de clasificación."*,Universidad de la Republica,Uruguay, (pp.5-7).
- Nelson, D .(1991). *"Conditional Heteroskedasticity in Asset Returns: A New Approach.» Econometrica (The Econometric Society)."*,vol 59 n° 2, (pp.347-370).
- Kosapattarapim,C.; Lin,Y.; MCCare,M.(2011). *"Evaluating the volatility forecasting performance of best fitting GARCH models in emerging asian stock markets."*,University of Wollongong Australia (Faculty of engineering and information sciences),Australia, (pp.9-21).
- Sepúlveda ,P.,(2011). *"Redes neuronales recurrentes con algoritmo de optimización por enjambre de partículas para pronóstico multivariado de captura de anchovetas en el norte de Chile."*,Pontificia universidad católica de Valparaíso (Facultad de ingeniería),Chile.