

Predictibilidad de la rentabilidad observada en el mercado accionario colombiano, basado en
análisis de series de tiempo.

Tatiana Andrea Ruiz Cañas

Trabajo de Grado para Optar al Título de Magister en Ingeniería Industrial

Director

Carlos Enrique Vecino Arenas,

Ph.D.

Codirector

Henry Lamos Diaz,

Ph.D.

Universidad Industrial de Santander

Facultad de Ingenierías Fisicomecánicas

Escuela de Estudios Industriales y Empresariales

Maestría en Ingeniería Industrial

Bucaramanga

2025

Tabla de Contenido

| | Pág. |
|--|-------------|
| Introducción | 11 |
| 1. Revisión de literatura | 14 |
| 1.1 Hipótesis de mercado eficiente | 14 |
| 1.2 Predictibilidad | 16 |
| 1.3 Modelos de pronósticos | 19 |
| 1.3.1 Modelo autoregresivo de media móvil (ARIMA) | 22 |
| 1.3.2 Modelo redes neuronales artificiales (RNA) | 24 |
| 2. Planteamiento del problema y justificación | 29 |
| 3. Objetivo..... | 32 |
| 3.1 Objetivo General..... | 32 |
| 3.2 Objetivos Específicos..... | 32 |
| 4. Hipótesis | 33 |
| 5. Metodología | 33 |
| 5.1. Determinación del enfoque..... | 33 |
| 5.2. Tipo de diseño..... | 34 |
| 5.3. Metodología KDD | 35 |
| 6. Análisis de series de tiempo, modelos y resultados..... | 39 |
| 6.1. Criterio de Información de Akaike | 51 |
| 6.2. Modelos para la serie del COLCAP..... | 52 |
| 6.3. IBC..... | 55 |
| 6.4. Serie junta | 78 |

| | |
|---|-----|
| 6.5 Redes neuronales artificiales..... | 82 |
| 6.5.1 Métodos estadísticos | 83 |
| 6.5.2 Predicción numérica..... | 84 |
| 6.5.3 Estadísticas del ajuste de redes neuronales a las series..... | 86 |
| Para el COLCAP..... | 88 |
| 6.6 Conclusiones para RNA..... | 95 |
| 6.6.1 Modelo híbrido..... | 96 |
| 7. Análisis de resultados | 100 |
| 8. Conclusiones..... | 103 |
| 9. Recomendaciones | 105 |
| Referencias Bibliográficas | 106 |

Lista de Tablas

| | Pág. |
|--|-------------|
| Tabla 1. Modelos para series de tiempo..... | 21 |
| Tabla 2. Canasta COLCAP vigente para el primer trimestre del 2019..... | 31 |
| Tabla 3. Estadísticas descriptivas de los índices..... | 42 |
| Tabla 4. Estadísticas descriptivas de las rentabilidades de los índices | 45 |
| Tabla 5. Prueba de rachas de los rendimientos de la serie con el índice IGBC del 2001 al 2013 | 48 |
| Tabla 6. Prueba de Kolmogorov-Smirnov de los rendimientos de la serie con el índice IGBC del 2001 al 2013..... | 49 |
| Tabla 7. Prueba de rachas de la rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018..... | 50 |
| Tabla 8. Prueba de Kolmogorov- Smirnov de la rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018 | 51 |
| Tabla 9. Modelos de ajuste | 52 |
| Tabla 10. Criterio AIC para la serie del COLCAP..... | 53 |
| Tabla 11. Modelos ajustados serie del COLCAP | 54 |
| Tabla 12. Criterio AIC para la serie del IBC | 55 |
| Tabla 13. Modelos ajustados serie del IBC | 77 |
| Tabla 14. Criterio AIC para la serie junta..... | 78 |
| Tabla 15. Modelos ajustados de la serie conjunta..... | 79 |
| Tabla 16. Prueba de aleatoriedad | 88 |
| Tabla 17. Rango Intercuartil | 89 |

| | |
|---|-----|
| Tabla 18. Prueba chi cuadrado..... | 90 |
| Tabla 19. Autocorrelación | 91 |
| Tabla 20. Comparación métrica de predictibilidad..... | 100 |

Lista de Figuras

| | Pág. |
|---|-------------|
| Figura 1. Representación de una RNA | 25 |
| Figura 2. IGBC desde el 3 de julio del 2001 hasta 31 de octubre del 2013..... | 40 |
| Figura 3. COLCAP desde noviembre del 2013 a diciembre del 2018..... | 41 |
| Figura 4. Rendimiento de IGBC desde el 3 de julio del 2001 hasta octubre del 2013 | 42 |
| Figura 5. Diagrama de caja de rendimientos de la serie con el índice IGBC del 2001 al 2013. .. | 43 |
| Figura 6. Rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018..... | 44 |
| Figura 7. Diagrama de caja de rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018..... | 44 |
| Figura 8. Diagrama de ACF de rendimientos de la serie con el índice IGBC del 2001 al 2013 .. | 46 |
| Figura 9. Diagrama de ACF de rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018..... | 46 |
| Figura 10. Diagrama de PACF de rendimientos de la serie con el índice IGBC del 2001 al 2013 | 47 |
| Figura 11. Diagrama de PACF de rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018..... | 47 |
| Figura 12. Red neuronal IGBC | 87 |
| Figura 13. Gráfica del error | 88 |
| Figura 14. Prueba Chi cuadrado..... | 90 |
| Figura 15. Topología de la RNA para el COLCAP | 93 |
| Figura 16. Diagrama del error para el COLCAP | 94 |

Figura 17. Gráfica de datos para el modelo híbrido..... 96

Figura 18. Parámetros 97

Figura 19. Predicción modelo ARIMA..... 97

Figura 20. Nuevo dataset 98

Figura 21. Modelo final 98

Lista de Apéndices

Se anexa carpeta con los siguientes apéndices:

Apéndice A. Revisión de la literatura

Apéndice B. Gráficas de los modelos con mayor AIC

Apéndice C. Resultados del ajuste del IBC mediante Python

Apéndice D. Ajuste a las RNA para las series

Apéndice E. Código red neuronal

Resumen

Título: Predictibilidad de la rentabilidad observada en el mercado accionario colombiano, basado en análisis de series de tiempo *

Autor: Tatiana Andrea Ruiz Cañas **

Palabras Clave: Pronóstico, índice COLCAP, modelo ARIMA, redes neuronales artificiales, modelos combinados.

Descripción: En este trabajo de investigación, se presenta el estudio del comportamiento estadístico de las series de tiempo de los índices IGBC (Índice General de la Bolsa de Valores de Colombia) y COLCAP (Índice de capitalización que refleja las variaciones de los precios de las acciones más líquidas de la Bolsa de Valores de Colombia) del 2001 al 2018 mediante el análisis de series de tiempo, con el fin de ajustarlas al modelo ARIMA (Modelo autorregresivo integrado de media móvil), redes neuronales artificiales y un modelo híbrido exploratorio basado en los anteriores para identificar cuál de ellos pronostica mejor los rendimientos.

Se observó que las series históricas del índice bursátil son asimétricas negativas, leptocúrticas y no son estacionarias, con un mayor apuntamiento en la distribución de frecuencias de la serie del IGBC. Además, se realizó la Prueba de Rachas a las series de tiempo estudiadas, encontrándose que estas son dependientes. Por otra parte, el IGBC, el COLCAP y la unión de las anteriormente mencionadas se ajustaron a modelos sencillos como MA(1), AR(1), MA(2), AR(2), ARMA(1,1), ARCH(1,1), GARCH(1,1).

Además, se determinó que las series del IGBC y COLCAP presentan un mayor ajuste a los modelos de redes neuronales con un solo rezago. Para el IGBC, la red neuronal con menor error es una Red multinivel hacia adelante (MLF), con dos capas ocultas y un nodo en la primera capa, dos nodos en la segunda capa y cuatro rezagos. Para el COLCAP, el mayor ajuste se da con una capa oculta, diez nodos y un solo rezago. Para el ajuste del modelo híbrido ARIMA-Red neuronal artificial se obtuvo un RMSE (Error de raíz cuadrada media) igual a 0,02.

Finalmente, se validan las hipótesis demostrando que las redes neuronales artificiales son más precisas en pronosticar los rendimientos, capturando el comportamiento y las relaciones que existen entre los mismos datos de las series. Adicionalmente, se incluye una métrica de predictibilidad instantánea en las series, lo que constituye una estrategia práctica para evaluar la idoneidad de métodos de pronóstico.

* Trabajo de Grado

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Estudios Industriales y Empresariales.
Director: Carlos Enrique Vecino Arenas. Ph.D. Codirector: Henry Lamos Diaz. Ph.D.

Abstract

Title: Predictability of the observed profitability in the colombian stock market, based on time series analysis.*

Author(s): Tatiana Andrea Ruiz Cañas**

Key Words: Forecast, COLCAP index, ARIMA model, artificial neural networks, combined models.

Description: This research work presents a study of the statistical behavior of the time series of the IGBC and COLCAP indices from 2001 to 2018 through time series analysis, to adjust them to the ARIMA model, artificial neural networks, and an exploratory hybrid model based on the previous ones for choose which of them have the best predicts.

It was observed that the historical series of the stock index are non-stationary, negative asymmetric, and leptokurtic. Likewise, a greater pointing was found in the frequency distribution of the IGBC series. Besides, the Runs test was performed on the time series, finding that these are dependent and, therefore, have a level of predictability. On the other hand, the IGBC, COLCAP, and the joint series were adjusted to simple models such as MA (1), AR (1), MA (2), AR (2), ARMA (1,1), ARCH (1, 1), GARCH (1,1).

On the other hand, it was determined that the IGBC and COLCAP series show a greater fit to the neural network models with a single lag. For the IGBC, the neural network with the lowest error is a Multilevel Forward Network (MLF), with two hidden layers and one node in the first layer, two nodes in the second layer, and four lags. The best fit for COLCAP index was obtained using a hidden layer and ten nodes for the historical data series and a single lag. For the adjustment of the ARIMA-Artificial Neural Network hybrid model, an RMSE equal to 0.02 was obtained.

Finally, the hypotheses are validated that artificial neural networks are more accurate in predicting yields, capturing the behavior and relationships that exist between the same data in the series. Additionally, is included an instantaneous predictability metric in the series, which constitutes a practical strategy to evaluate the suitability of forecasting methods.

* Degree Work

**Faculty of Physical-Mechanical Engineering. School of Industrial and Business Studies. Director: Carlos Enrique Vecino Arenas. Ph.D. Co-director: Henry Lamos Diaz. Ph.D.

Introducción

La literatura financiera asume la impredecibilidad de los retornos financieros por la formulación de la hipótesis del mercado eficiente (Fama, 1963, 1965, 1970), según la cual los precios se reflejan de manera prácticamente instantánea, donde la información llega de manera aleatoria en forma de noticia (Malkiel, 2003). En el mercado existen tres tipos de eficiencia; la eficiencia débil, en la cual la información pasada no tienen ningún poder de predicción sobre el futuro precio de los valores, porque los precios son independientes de un período a otro, la eficiencia semifuerte; donde los valores se ajustan rápidamente cuando la información se hace pública y la eficiencia fuerte; la cual incorpora las dos anteriores y la información interna, de esta forma los precios no sólo reflejan el histórico y la información pública, sino también toda la información que pueda obtenerse mediante el análisis de la empresa y de la economía (Roberts, 1967).

El análisis técnico ha sido usado en los mercados por décadas como una herramienta para predecir las tendencias de los precios de los activos. En la actualidad, es el modo dominante de análisis para transar divisas y mercancías (Lo & Hasanhodzic, 2010), siendo una técnica respaldada y utilizada por inversionistas y quienes piensan que el mercado si es eficiente (cf. Brealey, 1986).

Con el análisis técnico como herramienta de análisis, es posible determinar si el precio de un activo sube o baja con base en información histórica, especialmente con los datos de los precios del activo y los volúmenes de transacción pasados (cf. Achelis, 2000).

Algunos investigadores como Han, Yan y Zhou (2013) establecen por primera vez, que una estrategia de media móvil simple conduce a retornos anormales, un resultado empírico

inexplicable a partir de la eficiencia de los mercados y de los modelos de valoración de activos tradicionales. El resultado de Han et al. (2013), es importante porque demuestra que una de las estrategias del análisis técnico es rentable, incluso si se ajusta la rentabilidad por el riesgo utilizando los principales modelos de valoración de activos. Tal rentabilidad contradice la hipótesis de los mercados eficientes.

Por su parte, Ospina (2017) evaluó la asociación estadística entre la predictibilidad y la rentabilidad ajustada al riesgo en la anomalía de Han et al. (2013), por medio de la medición de la predictibilidad y de la rentabilidad como variables dinámicas cuyos valores se establecen en ventanas disjuntas de longitud que progresan en el tiempo de manera consecutiva, el autor estableció que la ventana en que se medía la predictibilidad estaría rezagada respecto a la ventana en que se medía la rentabilidad un número de días igual a la longitud con que se calculaba la media móvil, ya que el primer retorno de la estrategia de la media móvil solo sucede después de utilizar días de información de entrada, para ello utilizó a las acciones de los mercados de la Bolsa de Nueva York (NYSE).

En Colombia, se ha estudiado muy poco sobre el tema, por lo que vale la pena explorar si estos resultados son válidos para un mercado emergente, motivo por el cual en esta tesis se realizó la evaluación de la predictibilidad de la rentabilidad observada con el histórico del COLCAP, el índice bursátil más representativo de Colombia, definido por el

Banco de la República de Colombia como el índice de capitalización que refleja las variaciones de los precios de las acciones más líquidas de la Bolsa de Valores de Colombia (BVC), donde la participación de cada acción en el índice, está determinada por el correspondiente valor de la capitalización bursátil ajustada (flotante de la compañía multiplicado por el último precio). La canasta del índice COLCAP está compuesta por mínimo 20 acciones de 20 emisores diferentes.

Las acciones sobre empresas son los títulos más transados comúnmente en un mercado de valores, el cual es un componente fundamental de los mercados financieros. El pronóstico del precio de las acciones es de alto interés para los inversionistas ya que indicará las señales de compra o venta de estos títulos, con el fin de maximizar sus beneficios. Sin embargo, esta no es una tarea fácil dada la cantidad de variables macroeconómicas y microeconómicas que determinan su valor (Hadavandi, 2010).

En el contexto colombiano, el mercado de valores ha ganado gran importancia debido a su crecimiento, por esto en este estudio se analizaron los modelos de ajuste más representativos a las series de tiempo, utilizando técnicas de computación flexible caso particular de modelo autorregresivo ARIMA (modelo autorregresivo integrado de promedio móvil), redes neuronales y un modelo combinado de ambos para la ventana de tiempo del 2001 al 2018.

El presente trabajo de investigación está organizado de la siguiente manera: revisión de literatura enfocada en predictibilidad, medidas de predictibilidad, hipótesis de mercado eficiente, modelos matemáticos de pronóstico utilizados en series de tiempo, modelo ARIMA y redes neuronales artificiales. Posteriormente, se presenta el planteamiento del problema y justificación, las hipótesis, objetivos de la investigación, metodología, análisis de series de tiempo, modelos y resultados, análisis de resultados, conclusiones y finalmente, referencias bibliográficas.

1. Revisión de literatura

1.1 Hipótesis de mercado eficiente

Las acciones juegan un papel determinante en la cadena de valor del mercado bursátil y en las decisiones que ejecutan los agentes del mercado de valores. La hipótesis de mercado eficiente (HME) ha sido ampliamente analizada por diversos exponentes como: Samuelson (1965), Jensen (1978), Malkiel (1992), Bachelier (1900), entre otros, quienes confluyen en términos generales en que: un mercado es eficiente cuando en él se cuenta con la suficiente liquidez y racionalidad económica por parte de los agentes como para que cualquier tipo de información relevante sea absorbida por los precios de forma instantánea, generando un comportamiento aleatorio entre ellos, lo que hace imposible su pronóstico sistemático.

Además, la hipótesis del mercado eficiente supone que los inversionistas, motivados por las oportunidades de lucro, utilizan hasta la más insignificante de las informaciones disponibles para intentar sacar ventaja al transar con el activo. Al hacerlo, esta información se incorpora en el precio del activo. Así, en un mercado perfectamente eficiente, sin costos de transacción, ni fricciones, la información se incorporará en el precio de manera prácticamente instantánea (Lo & MacKinlay, 2002, p. 5) sugiriendo que no hay relación entre los precios pasados y futuros de los títulos, por lo que se supone que son independientes en el tiempo. Por lo tanto, el mercado de valores se caracteriza por la linealidad, la continuidad, la estática y la independencia; a través de las cuales, el riesgo de la inversión puede estimarse y controlarse. (Liu, G., Yu, C.-P., Shiu, S.-N., y Shih, I.-T, 2022).

La hipótesis del mercado eficiente predomina en la literatura, sin embargo, ha recibido críticas especialmente fuertes de parte de la economía comportamental, según las cuales, los

actores del mercado no son siempre racionales y exhiben comportamientos predecibles (Kay, 2011; Lo, 2008). En el mercado pueden existir tres tipos de eficiencia: la eficiencia débil, en la cual los títulos reflejan plenamente la información pasada, contenida en las series históricas de los precios, bajo la hipótesis de eficiencia débil ningún inversor podría obtener rendimientos extraordinarios (superiores al promedio del mercado), analizando la información contenida en las series históricas de precios; la eficiencia semifuerte es aquella cuyos precios reflejen toda la información histórica y pública disponible, lo que supone que ningún inversor podrá conseguir un rendimiento superior al promedio del mercado en base a análisis de series históricas de precios, balances, proyecciones de variables macroeconómicas y demás información de dominio público y la eficiencia fuerte se da cuando los precios de los títulos reflejan totalmente la información pasada, pública y privada, lo cual supone que ningún inversor puede obtener rendimientos extraordinarios, excepto por medio del azar, y que no existe información privilegiada (Roberts, 1967).

Según los economistas comportamentales, los modelos cuantitativos de los mercados eficientes, que están basados en la elección racional, probablemente están equivocados (Lo, 2008), mientras que las finanzas conductuales brindan orientación y asistencia cuando se intenta comprender las decisiones financieras que toman los inversores, al considerar la influencia de los rasgos psicológicos de esos inversores. También sostiene que los fenómenos financieros se pueden comprender mejor cuando los inversores no son completamente racionales. (Noreen U, Shafique A, Ayub U, Saeed SK, 2022)

La literatura sobre la predictibilidad de los retornos del mercado de valores es muy amplia, gracias a desarrollos en los algoritmos de extracción de conocimiento de aprendizaje automático, o la disponibilidad de datos cuantitativos y cualitativos como los que están disponibles en las redes sociales, las actitudes hacia el nivel de eficiencia del mercado han cambiado. En consecuencia,

existe una rica literatura dedicada al área de finanzas que utiliza las matemáticas y las estadísticas con el objetivo de modelar y predecir series de tiempo financieras. Estos métodos varían desde los primeros métodos estadísticos y econométricos hasta las últimas técnicas de optimización basadas en el aprendizaje automático (ML). (Aminimehr, A., Raoofi, A., Aminimehr, A. et al. 2022)

Dicho lo anterior diferentes artículos concluyen en que según la eficiencia del mercado y la ventana de tiempo utilizada de series de tiempo financieras es posible establecer predictibilidad utilizando técnicas de modelado.

1.2 Predictibilidad

El término predictibilidad es usado en diferentes contextos y es definida como la cualidad de lo que es predecible, una predicción es un pronóstico que puede expresarse con un grado de probabilidad, en el caso de series financieras el término es utilizado para hacer referencia a la capacidad de generar pronósticos a partir de datos históricos, de manera tal que para generar dicha predicción es necesario que el mercado no sea eficiente. Así, el estudio de la predictibilidad del mercado de valores tiene implicaciones importantes, tanto teóricas como prácticas: teóricamente, es interesante vincular la predictibilidad a otros factores económicos y financieros; empíricamente, la búsqueda de una mayor predictibilidad conforma la ocupación normal de los que comercian con acciones, especialmente las entidades institucionales (Ospina, 2015).

De acuerdo con lo anterior, el estudio sobre el tema ha llevado a que diversos autores establezcan la predictibilidad de distintas formas. Por ejemplo, Goyal y Welch (2008, p. 1455) escriben: “la literatura es difícil de absorber, diferentes artículos usan técnicas, variables y períodos de tiempo distintos”. Asimismo, algunos trabajos utilizan diversas metodologías para medir directamente una predictibilidad en un solo punto siendo esta una medida “instantánea” que no varía en el tiempo y que permite además establecer posibles determinantes económicos y

financieros. Por otro lado, otros autores miden la predictibilidad dinámica (variable en el tiempo) la cual posibilita describir mejor el fenómeno de la predictibilidad, y construir algoritmos que predican mejor y que pueden superar la estrategia de comprar y mantenerse incluso teniendo en cuenta los costos de transacción (Ospina, 2015). Además, se puede medir la predictibilidad mediante pruebas de caminata aleatoria o métricas de capacidad de ajuste de modelos de pronóstico.

Estas medidas de predicción, constituyen una estrategia práctica para evaluar la idoneidad de métodos de pronóstico para determinadas series de tiempo. De esta forma, si el pronóstico producido por una serie que contiene una cantidad significativa de estructura predictiva, la medida puede deducir que ese método es inadecuado para la serie de datos y por ende, se debería buscar otro método que se adapte mejor al fenómeno. (Garland, J, James, R & Bradley, E, 2014).

Articulando lo anteriormente mencionado, desde la literatura se han planteado diversas medidas de predictibilidad, asociándola con la forma como se generan los datos y otros respecto a la capacidad para predecir datos futuros mediante el ajuste a modelos de pronóstico.

Una manera de medir la predictibilidad es mediante la comprobación de la teoría de la caminata aleatoria o random walk, el cual asume que el mercado de valores se mueve de una manera completamente impredecible; la hipótesis sugiere que el precio futuro de cada acción es independiente de su propio movimiento histórico y del precio de otros valores (Lo & MacKinlay, 1988; Urrutia, 1994).

Dentro de las medidas de predictibilidad usadas con frecuencia está el criterio de información de Akaike (AIC) el cual fue propuesto por Akaike (1974) como un estimador insesgado asintótico de la información de Kullback-Leibler esperada entre un modelo candidato ajustado y el verdadero modelo. Igualmente, el criterio de información Bayesiana BIC derivado

por Schwarz en 1978, es uno de los métodos más populares usado para la selección de modelos, dado que tiene la capacidad de medir la eficacia del modelo de medidas con el fin de predecir los datos y penaliza la complejidad del modelo, mediante la aproximación a una transformación de la probabilidad posterior de un modelo candidato (Montesinos López, 2011).

Otros trabajos precursores propusieron medidas de predictibilidad de una serie de tiempo como las que la determinan considerando la aleatoriedad de los datos, las cuales miden la capacidad predictiva de un modelo mediante la comparación del error que tiene la serie original y la que presenta el modelo de ajuste con las metodologías, caso del presente proyecto, en el cual se utiliza esta métrica adicional con el fin de comparar las familias de los modelos (ARIMA, redes neuronales, modelo híbrido) y el test de la razón de varianzas el cual realiza una comparación sistemática con el camino aleatorio que no es predecible.

Por su parte, Granger y Newbold (1986) usan una medida de la predictibilidad de una serie de covarianza estacionaria similar al R^2 de una regresión lineal:

$$G=1 - \frac{Va(e_{t+j,t})}{Var(r_{t+j})}$$

Donde r_t es el retorno en el tiempo t , \hat{r}_{t+j} es la predicción óptima (la media condicional) y $e_{t+j} = r_{t+j} - \hat{r}_{t+j}$. De forma similar, se halla en Galbraith (2003) la función de contenido de la predicción para generalizar y compararla con diferentes modelos como se presenta a continuación:

$$C(s)=1 - \frac{MSE_{\tilde{r}(s)}}{MSE_{\hat{r}(s)}}$$

Donde $\tilde{r}(s)$ es la media muestral, $\hat{r}(s)$ es la predicción s -pasos hacia adelante y MSE denota el error cuadrático medio.

Esta última función descrita es la utilizada en la presente tesis donde el error cuadrático medio de los datos se contempla mediante la fórmula:

$$MSE_{\hat{r}(s)} = \frac{1}{N} \sum (r_t - \hat{r}_t)^2$$

Donde \hat{r}_t es la media muestral y r_t es el dato actual.

El error cuadrático medio de los modelos se calcula mediante la siguiente ecuación:

$$MSE_{\hat{r}(s)} = \frac{1}{N} \sum (y_t - \hat{y}_t)^2$$

Donde: y_t es el dato actual y \hat{y}_t es el dato predicho bajo la metodología seleccionada.

En cuanto al análisis de la medida de predictibilidad se busca que al introducir los datos y calcular el valor predictivo de la serie, esta sea cercana a uno describiendo que la metodología seleccionada contiene un porcentaje de error interior, comparado con la dispersión de los datos originales.

Mediante la revisión de literatura no se encuentra un consenso sobre la mejor alternativa de medición de predictibilidad (Diebold & Kilian, 2001; Garland, James & Bradley, 2014) y el debate se divide en estudios que buscan los mejores predictores de la rentabilidad de las acciones y estudios que confirman su imprevisibilidad (Lopes Moreira Da Veiga, Mao & Casas, 2020). Por esta razón, la medida global que finalmente se seleccionó es la de Galbraith (2003), la cual trabaja el término de predictibilidad bajo la definición de relación directa con la capacidad de ajuste del modelo usado en el pronóstico dado que la métrica apoya las suposiciones del análisis técnico y permite comparar el modelo más óptimo.

1.3 Modelos de pronósticos

Para abordar las series de tiempo financieras se utilizan modelos para datos los cuales pueden tener muchas formas y representar diferentes procesos estocásticos, entre los que se pueden

encontrar los modelos autorregresivos (AR), los modelos de media móvil (MA) y los modelos integrados (ARIMA). Estas clases de modelos dependen linealmente de los puntos de datos anteriores.

Adicionalmente, se han estudiado modelos de series de tiempo, para representar los cambios de varianza en el tiempo (heteroscedasticidad). Estos modelos representan heteroscedasticidad condicional autorregresiva (ARCH) y la colección comprende una amplia variedad de representaciones (GARCH, TARARCH, EGARCH, FIGARCH, CGARCH) introducidos por Engle (1982) y Bollerslev (1986).

Otros trabajos, emplean la combinación de los métodos GARCH, EGARCH, paseo aleatorio y promedios móviles exponenciales para predecir el comportamiento de la volatilidad en los mercados de valores (Jing-rong, 2007). Su aplicación al índice de la bolsa de Shenzhen en la China arrojó que el modelo propuesto entregaba errores inferiores al obtenido con los pronósticos de modelos individuales.

Sin embargo, dependiendo de las características de las series y las ventanas de tiempo y con el auge de la inteligencia artificial se han ampliado los modelos de solución y se han combinado métodos con el fin de mejorar el pronóstico y por ende la toma de decisiones.

Algunos enfoques recientes de aprendizaje profundo utilizan Redes Neuronales Artificiales (ANN), estos han recibido un interés particular como una herramienta para minería de datos, puesto que ofrece los medios para modelar de manera efectiva y eficiente problemas grandes y complejos, son capaces de encontrar relaciones (patrones) de forma inductiva por medio de los algoritmos de aprendizaje basado en los datos existentes, más que requerir la ayuda de un modelador para especificar la forma funcional y sus interacciones.

Las redes neuronales artificiales son un modelo inspirado en la estructura del sistema nervioso de los seres humanos, el cual consiste en un conjunto de unidades conectadas entre sí para transmitirse señales. La información de entrada atraviesa la red neuronal (donde se somete a diversas operaciones) produciendo unos valores de salida.

Dado que los pronósticos se realizan con el fin de orientar las decisiones, se busca que los modelos de series de tiempo sean más precisos en sus proyecciones, algunos estudios han probado la superioridad de modelos combinados. **En la Tabla 1.**

Modelos para series de tiempo se presenta un resumen de modelos de solución para series de tiempo financieras utilizadas en trabajos recientes.

Tabla 1.

Modelos para series de tiempo

| Autor y Año | Método |
|---|--|
| Xiao, et. al (2019) | Propusieron un nuevo modelo combinado que integra el análisis de espectro singular (SSA), y máquina de vectores de soporte (SVM). Además, compararon el sistema de inferencia difusa basado en red adaptativa (ANFIS), la descomposición empírica en modo de conjunto-ANFIS (EEMD-ANFIS), EEMD-SVM y SSA-ANFIS, para concluir que SSA-SVM demuestra el mejor rendimiento de predicción según cuatro criterios. |
| Garcia, Diego; Kristjanpoller, Werner (2019) | Proponen un marco compuesto por un conjunto de modelos de series de tiempo tales como ARIMA y GARCH, modelos no paramétricos como ANN y sistemas de inferencia difusa (FIS), y especificaciones híbridas de ambos, probando la superioridad del modelo adaptativo GARCH-FIS. |
| Gao , Tingwei ; Chai, Yueting (2018) | Predicen los precios de cierre de las acciones mediante el uso de redes neuronales recurrentes (RNN). Un modelo de memoria de largo a corto plazo (LSTM), un tipo de RNN junto con acciones básicas de datos comerciales y los indicadores técnicos, los cuales se presentan como un nuevo método para predecir el precio de cierre del mercado. |
| Brasileiro, Souza, Victor LF ; e Oliveira, Adriano LI (2017) | Estudian el uso de técnicas de minería de datos para predecir la dirección futura de los precios de las acciones y descubrir patrones en los datos de series de tiempo para proporcionar soporte de decisión. |

 Continuación Tabla 2. *Modelos para series de tiempo*

| | |
|--|--|
| Brasileiro, Rodrigo C ; Souza, Victor LF ; De Oliveira, Adriano LI (2017) | Proponen el enfoque PAA-MS-IDPSO-V (aproximación agregada por pieza - Multi-enjambre de optimización mejorada autoadaptativa de enjambre de partículas con validación). Los patrones descubiertos por el método se utilizan junto con las reglas de inversión para respaldar las decisiones y, por lo tanto, ayudan a los inversores a maximizar el beneficio en sus operaciones en el mercado de valores. |
| Liu, Xiaochun; Luger, Richard (2015) | Utilizan un nuevo modelo de tipo GARCH para la volatilidad condicional autorregresiva, la asimetría y la curtosis. Realizaron la estimación del modelo con métodos bayesianos utilizando un método adaptativo: Cadena de Markov del algoritmo de Monte Carlo. |
| Wei, Liang-Ying (2013) | Utilizan el modelo híbrido, un algoritmo genético de expectativa adaptativa para optimizar el sistema de inferencia difusa basado en red (ANFIS) adaptativo para predecir las tendencias del precio de las acciones. La efectividad del modelo propuesto se verifica con evaluaciones de desempeño y error cuadrático medio (RMSE). |

1.3.1 Modelo autoregresivo de media móvil (ARIMA)

Una serie de tiempo es un conjunto de datos históricos que pueden descomponerse en cuatro componentes que no son directamente observables, de los cuales únicamente se pueden obtener estimaciones. Estos cuatro componentes son:

- Tendencia (T): representa el comportamiento predominante de la serie. Esta puede ser definida vagamente como el cambio de la media a lo largo de un extenso período de tiempo.
- Ciclo (C): caracterizado por oscilaciones alrededor de la tendencia con una larga duración, y sus factores no son claros.
- Estacionalidad (E): es un movimiento periódico que se producen dentro de un período corto y conocido.
- Aleatorio (A): son movimientos erráticos que no siguen un patrón específico y que obedecen a diversas causas (Hurtado C, 2008).

Debido al comportamiento de las series de tiempo, se han desarrollado modelos para encontrar patrones futuros, entre los cuales se encuentra el modelo de promedio móvil integrado autorregresivo (ARIMA), siendo uno de los modelos lineales más populares en predicción de series de tiempo durante las últimas tres décadas (Zhang, 2001).

El modelo, también conocido como el modelo o metodología Box - Jenkins se usa comúnmente en análisis, es decir, identificación, estimación, verificación y pronóstico de diagnóstico, además es popular debido a su generalidad, ya que puede manejar cualquier serie, estacionaria o no estacionaria (Susruth & Mulukalapally, 2017).

La metodología de Box-Jenkins verifica la estacionariedad de la serie; si ésta no es estacionaria, la diferencia hasta alcanzarla, además, trata de identificar el proceso estocástico que ha generado los datos, estima los parámetros que caracterizan dicho proceso y verifica que se cumplan las hipótesis que han permitido la estimación de dichos parámetros (Hurtado, 2008).

El modelo ARIMA supone que el valor futuro de una variable es una función lineal de varias observaciones pasadas y errores aleatorios. El proceso subyacente que genera la serie temporal tiene la forma:

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} \\ + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

Donde, Y_t y ε_t son el valor real y el error aleatorio en el período de tiempo t , respectivamente; ϕ_i ($i = 1; 2; \dots p$) y θ_j ($j = 0; 1; 2; \dots q$), son parámetros del modelo. p y q son enteros y a menudo son referidos como órdenes del modelo. Se supone que los errores aleatorios, ε_t , se distribuyen de forma independiente e idéntica con una media de cero y una varianza constante de σ^2 .

La ecuación implica varios casos especiales importantes de la familia de modelos ARIMA. Si $q = 0$, entonces se convierte en un modelo AR (autorregresivo) de orden p . Cuando $p = 0$, el modelo se reduce a un modelo MA (media móvil) de orden q . Por lo tanto, la tarea central del modelo ARIMA es determinar el orden de modelo apropiado ($p; q$) (Zhang, 2001). Asimismo, el entero d que es positivo, controla el nivel de diferenciación. Si $d = 0$, entonces es equivalente a un modelo ARMA, el cual es un proceso estocástico que puede ser descrito por la suma ponderada de sus valores anteriores y la suma ponderada de un error de ruido blanco de períodos anteriores.

Siguiendo la formulación del modelo, los parámetros participantes deben estimarse con la ayuda del método del gradiente, buscando el gradiente cero de la suma cuadrática media de errores de ajuste sobre los datos históricos. La estimación de parámetros básicamente apunta a la minimización del error de pronóstico.

Los modelos estadísticos como el modelo Box - Jenkins ARIMA usan más supuestos que pueden afectar la precisión de sus predicciones cuando no se cumplen los mismos (Qumsiyeh, Maher, Derecha, R y Gannon, D, 2017).

Si bien, el modelo ARIMA es uno de los modelos lineales más populares, estudios demuestran que el modelo tradicional puede combinarse con técnicas lineales y no lineales y compararlas en el modelado y pronóstico de series de tiempo financieras verificando la mejora de la capacidad de modelado.

1.3.2 Modelo redes neuronales artificiales (RNA)

La búsqueda de modelos para predecir los precios de los mercados financieros sigue siendo un tema muy investigado, a pesar de los principales desafíos relacionados. Los precios de los activos financieros son no lineales, dinámicos y caóticos, por lo que las series de tiempo financieras son difíciles de predecir. Según Henrique, BM, Sobreiro, VA y Kimura, H, 2019 quienes realizaron

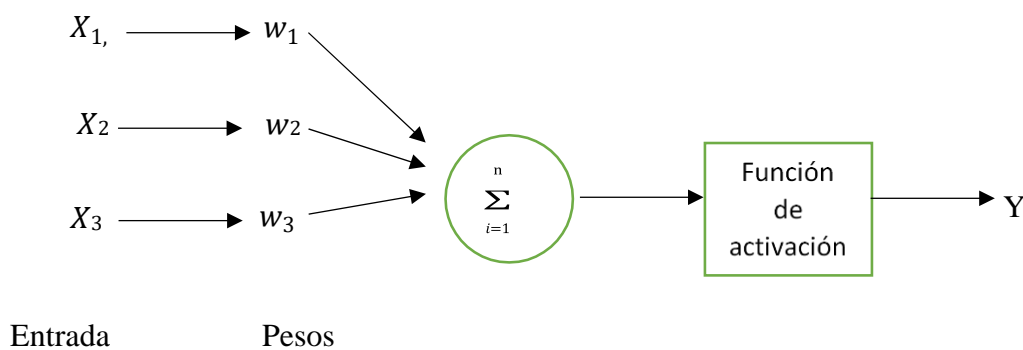
una revisión profunda de artículos, concluyeron que los modelos de aprendizaje automático son muy útiles en series de tiempo dada su capacidad para reconocer patrones complejos en diversas aplicaciones, por lo que son muy investigadas y utilizadas las metodologías de máquinas de vectores de soporte (SVM) y redes neuronales.

La Inteligencia Artificial estudia la forma de diseñar procesos que exhiban características que comúnmente se asocian con el comportamiento humano inteligente como la red neuronal, esta consiste en un conjunto de elementos de procesamiento, llamados neuronas, conectándose entre sí (Koehn, 1994), esta relaciona variables independientes con variables de salida; la organización y disposición de las neuronas dentro de una red neuronal se denomina topología, y viene dada por el número de capas, la cantidad de neuronas por capa, el grado de conectividad, y el tipo de conexión entre neuronas. Las Redes Neuronales Artificiales (RNA) pueden representarse en la

Figura 1:

Figura 1.

Representación de una RNA



Las RNA requieren de un entrenamiento que consiste en presentar un vector de entrada a la red, calcular la salida de la red, compararla con la salida deseada, y el error o diferencia resultante

se utiliza para realimentar la red y cambiar los pesos de acuerdo con un algoritmo que tiende a minimizar el error.

En la literatura se encuentran varios tipos de aprendizaje como el aprendizaje por corrección de error, que consiste en ajustar los pesos de las conexiones de la red en función de la diferencia entre el valor deseado y el obtenido en la salida de la red, el aprendizaje por refuerzo, el cual es un aprendizaje más lento, dado que no se indica la salida durante el entrenamiento y finalmente, el aprendizaje estocástico el cual consiste en cambiar aleatoriamente los valores de los pesos de las conexiones de la red para evaluar el efecto que se produce a partir del objetivo deseado y la distribución de probabilidad. Por lo tanto, la tarea del aprendizaje neuronal es construir un estimador $g^\lambda(x, w)$ de la función desconocida el cual se conoce sólo un conjunto de datos.

$$g^\lambda(x, w) = \gamma_2 \left[\sum_{j=1}^{\lambda} w_{ij} \gamma_1 \left[\sum_{i=1}^m w_{ij} x_i + w_{m+1,j} \right] + w_{\lambda+1} \right]$$

Donde, $w = (w_1, \dots, w_d)^T$ es el vector paramétrico a ser estimado y equivale a las ponderaciones de las conexiones entre las neuronas de la red, γ_1 es una función no-lineal acotada y diferenciable con forma de función sigmoide o de radio basal, γ_2 es una función que puede ser lineal o no lineal y λ es el parámetro de control que indica el número de neuronas escondidas. La función γ_1 típicamente es la función sigmoideal. Si la función γ_2 se elige no-lineal, debe ser estrictamente monótona, acotada y diferenciable.

El proceso de aprendizaje de una red neuronal *feed forward* se caracteriza por ser supervisado por los parámetros de la red, conocidos como pesos, los cuales son estimados a partir de un conjunto de patrones de entrenamiento compuesto por patrones de entrada y salida. Por lo tanto, el ajuste de la red se produce como resultado de la estimación de los parámetros basados en una muestra de tamaño n .

Los modelos de RNA dependen de la elección del valor de capas ocultas (q), el cual depende de los datos, además de determinar la dimensión del vector de entrada y el número de observaciones rezagadas (p). Este último valor es el parámetro más importante a estimar, sin embargo, en los parámetros q y p no existe un estándar del cual se base para determinar su valor.

Las redes neuronales artificiales han sido utilizadas en diferentes aplicaciones en el ámbito colombiano como en pronósticos de índices, precios de acciones, entre otros. Esto se demuestra por los trabajos realizados por Misas, López & Borrero (2002), Misas, López, Arango & Hernández (2003), Jalil & Misas (2006) y Aristizabal & Misas (2006), los cuales determinaron la tipología de las redes neuronales artificiales mediante prueba y error, con el fin de encontrar el mejor ajuste a los datos concluyendo que las RNA tienen mayor bondad de ajuste que los métodos tradicionales.

Otros estudios como los de Hornik et al. (1989) y Cybenko (1989), han demostrado que las redes neuronales son aproximadores universales y que las formadas por múltiples capas (MLP) son una de las arquitecturas más utilizadas en la solución de problemas, debido a su fácil uso y aplicabilidad. Además, el MLP con conexiones feed-forward (FNN) y algoritmo de aprendizaje back-propagation (BP) y la RNA es más común y utilizada en la identificación de sistemas no lineales (Subudhi y J. Debashisha, 2011).

Asimismo, Aminimehr, A., Raoofi, A., Aminimehr, A. et al. (2022) examinaron el proceso de prueba de la eficiencia del mercado y la viabilidad de la predicción del mercado con un enfoque detallado, coherente y estadístico, para ello utilizaron los datos del mercado S&P 500 y realizaron redes neuronales, con memoria a corto y largo plazo tipo LSTM, debido a que intrínsecamente, considera la memoria a largo y corto plazo en sus cálculos. Por lo tanto, probaron diferentes

métodos de preprocesamiento para mejorar su precisión, con datos de mercado durante septiembre de 2000 a febrero de 2021.

Pradeepkumar, D & Ravi, V. (2017) presentaron una nueva red neuronal de regresión cuántica entrenada con la optimización del enjambre de partículas (PSO) denominada PSOQRNN, para pronosticar la volatilidad de las series de tiempo financieras. Estos investigadores compararon modelos tradicionales de predicción de volatilidad, como el Heteroscedasticidad Condicional Autorregresiva Generalizada (GARCH) y tres Redes Neuronales Artificiales (RNA) que incluyen Perceptrón Multicapa (MLP), Red Neural de Regresión General (GRNN), Grupo Método de Manejo de Datos (GMDH), Bosque Aleatorio (RF) y dos híbridos basados en Regresión Cuantil (QR), incluida la Red Neuronal de Regresión Cuantil (QRNN) y el Bosque Aleatorio de Regresión Cuantil (QRRF), indicando que el PSOQRNN propuesto superó a estos modelos en términos de error cuadrático medio (MSE), corroborado por la prueba Diebold Mariano de significación estadística.

En el mundo de las Finanzas, el uso de redes neuronales profundas, en los últimos años ha venido aumentando dado que contiene una cantidad mayor de capas de neuronas que almacenan el aprendizaje y suelen mejorar el ajuste de los datos ofreciendo nuevos métodos y enfoque para abordar de forma más eficiente los problemas, siendo una herramienta que permite manejar las series de tiempo de forma más sencilla como las Redes Neuronales Recurrentes, RNR, (RNN Recurrent Neural Networks) y en particular con las redes conocidas como LSTM, (LSTM Long Short Term Memory).

Las RNA con aprendizaje profundo se diferencian de las redes tradicionales dado que estas contienen dos o tres capas ocultas, mientras que en las de este tipo, puede tener hasta 150. Estas redes neuronales profundas funcionan de manera similar a las tradicionales, por back-propagation

(BP), ajustando el error de igual forma, utilizando el método de gradiente descendente. Este proceso se repite miles de veces, ajustando los pesos del modelo, en función de los errores que produce durante el aprendizaje.

Según la revisión de la literatura, se encuentra que los modelos combinados y las redes neuronales de manera general han mostrado mejores ajustes a las series de tiempo financieras describiendo su comportamiento lineal y no lineal, obteniendo pronósticos más precisos, de esta manera, es importante continuar aplicando modelos que puedan disminuir el error del pronóstico de los rendimientos de índices financieros sobre todo en mercados emergentes.

2. Planteamiento del problema y justificación

En el capítulo anterior, se presentó la hipótesis de mercado eficiente donde se afirma la impredecibilidad de la rentabilidad, la teoría ha sido el paradigma de referencia de numerosos trabajos tanto teóricos como empíricos sobre el comportamiento del mercado de valores dado que está asume que los precios tendrían que ser impredecibles, ya que la información llega de manera aleatoria en la forma de noticias y los precios la reflejan (Malkiel, 2003) (Ospina 2017). Sin embargo, distintos trabajos afirman que la eficiencia no es una característica estática de los mercados, sino que varía en el tiempo, dando lugar a otras vertientes, las cuales se han propuesto la tarea de estudiar la relación de la predictibilidad y los rendimientos encontrando patrones lineales y no lineales en las series de tiempo, cuyos resultados depende de las ventanas móviles seleccionas.

El pronóstico del precio de las acciones es, por lo tanto, de alto interés para los inversionistas ya que indicará las señales de compra o venta de estos títulos con el fin de maximizar

sus beneficios, sin embargo, no es una tarea fácil dada la cantidad de variables macroeconómicas y microeconómicas que determinan su valor (Hadavandi, 2010).

La precisión de la previsión de rendimientos se ha convertido en una fuente adicional ayudando a los participantes del mercado a hacer mejores decisiones informadas bajo incertidumbre (Garcia, D y Kristjanpoller, W, 2019).

En este sentido, Han et al. (2013) generan evidencias a favor de la vertiente académica que apoya la predictibilidad, en el cual, se demostró que al medir la predictibilidad (rezagada) y la rentabilidad ajustada al riesgo de manera dinámica, ambas podían relacionarse mediante una regresión lineal para cada portafolio. Los resultados de su tesis sugirieron una relación causal desde la predictibilidad hacia la rentabilidad ajustada al riesgo, en este sentido, cuando no hay predictibilidad, la rentabilidad desaparece, y reaparece en presencia de la predictibilidad.

Por otra parte, analizar la serie de rentabilidad observada en mercado de valores en Colombia es interesante dado que ha tomado gran importancia por su crecimiento, realizándose estudios como el de Talero, Duarte y Garcés, 2017 en el que analizaron el comportamiento histórico de las rentabilidades latinoamericanas, con el fin de analizar la complejidad del mercado latinoamericano en la que mediante comparaciones observaron que dependiendo del período, Colombia, Chile, Argentina, México, Perú y Brasil presentaron comportamientos similares.

En el presente trabajo se pretende probar la hipótesis, con el índice COLCAP del 2001 al 2018 y ajustar modelos en la búsqueda de dichos patrones, definiéndose este como el índice de capitalización que refleja las variaciones de los precios de las acciones más líquidas de la BVC, donde la participación de cada acción en el índice está determinada por el correspondiente valor de la capitalización bursátil ajustada.

En la Tabla 2.

Canasta COLCAP vigente para el primer trimestre del 2019 se encuentra la canasta del índice la cual está compuesta siempre por 20 emisores, descritos a continuación.

Tabla 2.

Canasta COLCAP vigente para el primer trimestre del 2019

| Canasta COLCAP Vigente para el Trimestre | | | | |
|--|---|--------------|-----------|---------------|
| Nemotécnico | Descripción | Ponderador | Precio | Participación |
| ECOPETROL | ECOPETROL S.A. | 0.0751317649 | 3,415.00 | 16.160% |
| PFBCOLOM | BANCOLOMBIA S.A. | 0.0054903928 | 40,400.00 | 13.970% |
| GRUPOSURA | GRUPO INVERSIONES SURAMERICANA | 0.0035599976 | 36,640.00 | 8.215% |
| ISA | INTERCONEXION ELECTRICA S.A. E.S.P. | 0.0069334709 | 15,960.00 | 6.970% |
| BCOLOMBIA | BANCOLOMBIA S.A. | 0.0027647996 | 39,600.00 | 6.896% |
| NUTRESA | GRUPO NUTRESA S.A. | 0.0036733823 | 25,780.00 | 5.964% |
| GEB | GRUPO ENERGIA BOGOTA S.A. E.S.P. | 0.0435893278 | 2,075.00 | 5.697% |
| GRUPOARGOS | GRUPO ARGOS S.A. | 0.0046179613 | 18,100.00 | 5.264% |
| PFAVAL | GRUPO AVAL ACCIONES Y VALORES S.A. | 0.0616628512 | 1,235.00 | 4.796% |
| PFGRUPSURA | GRUPO INVERSIONES SURAMERICANA | 0.0016595056 | 34,400.00 | 3.595% |
| PFDVVNDA | BANCO DAVIVIENDA S.A | 0.0013285662 | 37,400.00 | 3.130% |
| CEMARGOS | CEMENTOS ARGOS S.A. | 0.0062411814 | 7,880.00 | 3.098% |
| PFGRUPOARG | GRUPO ARGOS S.A. | 0.0032255614 | 14,800.00 | 3.007% |
| ÉXITO | ALMACENES EXITO S.A. | 0.0029772801 | 14,520.00 | 2.723% |
| BOGOTA | BANCO DE BOGOTA S.A. | 0.0005876583 | 66,000.00 | 2.443% |
| CELSIA | CELSIA S.A E.S.P | 0.0079580417 | 4,430.00 | 2.220% |
| CORFICOLCF | CORPORACION FINANCIERA COLOMBIANA S.A. | 0.0012582558 | 26,500.00 | 2.100% |
| PFCMARGOS | CEMENTOS ARGOS S.A. | 0.0032092321 | 6,550.00 | 1.324% |
| CLH | CEMEX LATAM HOLDING | 0.0021071721 | 4,800.00 | 0.637% |
| GRUPOAVAL | GRUPO AVAL ACCIONES Y VALORES S.A. | 0.0076992197 | 1,220.00 | 0.592% |
| PFAVH | AVIANCA HOLDINGS S.A | 0.0050684476 | 1,675.00 | 0.535% |
| BVC | BOLSA DE VALORES DE COLOMBIA S.A. | 0.0003174518 | 12,060.00 | 0.241% |
| CNEC | CANACOL ENERGY LTD | 0.0003040165 | 10,800.00 | 0.207% |
| CONCRET | CONSTRUCTORA CONCRETO S.A | 0.0042499238 | 470.00 | 0.126% |
| ETB | EMPRESA DE TELECOMUNICACIONES DE BOGOTA S.A. E.S.P. | 0.0050781544 | 286.00 | 0.091% |

Con el índice descrito anteriormente se realizan pronósticos útiles para la toma de decisiones de los inversionistas de compra o venta de sus títulos, y dado que el mercado de valores es una de las fuentes de recursos más importantes para las empresas, cotizando en la bolsa, las empresas pueden captar capital adicional para su expansión, mediante la venta de acciones, y, por otra parte, los compradores también reciben un beneficio.

Es esta la razón por la cual el trabajo tiene una pertinencia teórica dado que busca evaluar la predictibilidad de la serie de la rentabilidad observada con el histórico del COLCAP, también se destaca que en el contexto colombiano se han utilizado modelos clásicos como ARIMA (modelo autorregresivo integrado de promedio móvil) y modelos ARCH (modelo auto regresivo condicionalmente heterocedástico), para lo cual en el estudio, se analizaron los modelos ARIMA, Redes Neuronales Artificiales y un modelo combinado de ambos aplicado a la serie temporal, con el fin de verificar cuál modelo presenta mayor poder de predicción y generar conclusiones que puedan generalizarse a otras bolsas con características similares.

3. Objetivo

3.1 Objetivo General

Evaluar la posible predictibilidad de la rentabilidad observada en el mercado accionario colombiano.

3.2 Objetivos Específicos

Identificar las principales características de la serie histórica del índice bursátil COLCAP mediante un análisis descriptivo del índice durante el período de estudio (2001 al 2018) para determinar puntos críticos, tendencias, estacionalidad y fluctuaciones aleatorias.

Analizar los patrones de comportamiento que permiten la predictibilidad de la serie a partir del ajuste de modelos ARIMA, con el fin de capturar la dinámica que presenta la serie de retornos observados en el COLCAP durante el período de estudio.

Analizar los patrones que permiten la predictibilidad por medio de Redes Neuronales Artificiales para la serie de retornos observados en el COLCAP durante el período de estudio.

Evaluar la capacidad predictiva y de ajuste de los modelos para la serie histórica de la rentabilidad observada y se explorará un modelo combinado de modelos ARIMA y Redes Neuronales Artificiales.

4. Hipótesis

De acuerdo con el problema de investigación se plantean las siguientes hipótesis:

- H 1: Existe predictibilidad en la serie histórica de la rentabilidad observada mediante el índice COLCAP para la ventana de tiempo.
- H 1.2: De existir predictibilidad, la capacidad predictiva y de ajuste del modelo combinado (redes neuronales artificiales-ARIMA) para la serie histórica de la rentabilidad observada es estadísticamente mejor comparado con los modelos ARIMA y Redes Neuronales Artificiales.

5. Metodología

5.1. Determinación del enfoque

La presente investigación sigue un enfoque cuantitativo, utilizando la información histórica del COLCAP para el análisis de series de tiempo y teniendo como características más notables, la de la objetividad de los instrumentos de recogida y análisis cuantitativos de los datos.

La investigación se apoya en el método hipotético deductivo, la experimentación y técnicas estadísticas, por lo que el modelo experimental aplicado, pretende probar su efectividad a través del procesamiento de los datos, mediante procedimientos matemáticos de ajuste del modelo econométrico híbrido.

Por otra parte, este tipo de enfoques busca la exactitud de las mediciones o indicadores sociales con el fin de generalizar sus resultados a poblaciones o situaciones amplias. Por lo que trabajan fundamentalmente con el número, el dato cuantificable (Galeano, 2004).

5.2. Tipo de diseño

La investigación es de tipo no experimental longitudinal dado que estas se concentran en estudiar cómo evolucionan una o más variables o las relaciones entre ellas, y/o analizar los cambios a través del tiempo de un evento, una comunidad, un fenómeno, una situación o un contexto Hernández, et al (2010).

Al introducir la dimensión temporal se abren nuevas perspectivas explicativas, puesto que se puede aprovechar la información obtenida del “sentido del cambio” que presentan los datos u observaciones a lo largo de esa dimensión. Como ha señalado Kandel (cit. en Arnau, 1995), los diseños longitudinales permiten medir los cambios para cada caso o sujeto, sin que para ello sea necesario inferirlo a partir de la diferencia entre los individuos. Esto evita algún tipo de incertidumbre y permite establecer, de forma más clara, la acción causal de una variable.

Los diseños de series temporales, dado que son una secuencia cronológicamente ordenada de valores de medición, a lo largo de cincuenta o más puntos equidistantes de tiempo, trabajan con un solo dato por punto del tiempo y por variable medida, es decir, datos de un solo sujeto en una o más variables, con el fin de buscar regularidades del comportamiento en el tiempo, así como el pronóstico o predicción de su comportamiento más probable en el futuro.

5.3. Metodología KDD

El KDD (Knowledge Discovery in Databases) es un modelo que define el descubrimiento de conocimiento en bases de datos como un “proceso” (Moine, 2013). En el año 1996, Fayyad define a KDD como el “proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia entendibles en los datos”. El término proceso, se refiere a la secuencia iterativa de etapas o fases que lo componen. Los patrones deberían ser válidos para nuevos datos, novedosos en el sentido que deberían aportar nuevo conocimiento al dominio de aplicación y potencialmente útiles para el usuario final o tomador de decisiones (Moine, 2013).

El modelo de proceso KDD se resume en las siguientes fases:

Fase 1. Selección de los datos: se eligieron el índice general de la Bolsa de Valores de Colombia (IGBC), el cual antes del 01 de noviembre del año 2013, era el principal indicador del mercado de valores, luego el COLCAP se estableció como el principal indicador de la Bolsa de Valores de Colombia (BVC), por lo que la ventana de tiempo para el proyecto se tomó desde el 2001 al 31 de diciembre del 2018, tomando los dos indicadores.

Fase 2. Pre-procesamiento de los datos: se realizó un tratamiento de los datos incorrectos y ausentes, en esta etapa se hace limpieza de ruido o datos anómalos (outliers) y tratamiento de datos faltantes (momentos en los que la bolsa de valores no operó).

Fase 3. Identificación de características: con el fin de dar cumplimiento al primer objetivo, inicialmente es importante determinar la existencia de valores atípicos, extremos y valores perdidos antes de tratar de modelar la serie. Por medio de herramientas como:

1. Gráfico de secuencias.
2. Gráfico de Caja y Sesgo.

3. Cálculo de estadísticas descriptivas: media, mediana, moda, desviación típica, mínimo y máximo, sesgo, curtosis.

De manera empírica con el fin de identificar aleatoriedad, el gráfico de secuencia, demuestra si éste tiene una tendencia hacia arriba o hacia abajo y con ello la no aleatoriedad. Por otra parte, se puede verificar dicha percepción por medio de la prueba de rachas (El procedimiento Prueba de rachas contrasta si es aleatorio el orden de aparición de dos valores de una variable. Una racha es una secuencia de observaciones similares. Una muestra con un número excesivamente grande o excesivamente pequeño de rachas sugiere que la muestra no es aleatoria). Con lo anterior, se determina la dependencia o no de las series. En caso de que no sea aleatoria, la serie se conoce como ruido blanco o caminata aleatoria y por ende no tiene las componentes de tendencia, estacionalidad ni ciclos.

Fase 4. Determinar si la hipótesis sobre la predicción de las series es aceptada o rechazada: en esta fase se realiza la prueba de dependencia con el fin de probar la existencia de predictibilidad en la serie histórica de la rentabilidad observada mediante el índice COLCAP para la ventana de tiempo.

Fase 5. Determinar y aplicar el algoritmo de minería: de acuerdo con la tarea de minería establecida, en esta etapa se define el algoritmo (o algoritmos) que se aplican para la búsqueda de patrones sobre los datos. Incluye la determinación de qué modelos y parámetros son los más adecuados según la naturaleza del problema y de los datos disponibles.

En el proceso de minería de datos, se obtienen los patrones de interés según la tarea de minería que se lleva a cabo (descriptiva o predictiva), en esta etapa se aplican los algoritmos y técnicas seleccionadas al conjunto de datos. Para abordar este problema, las técnicas de computación flexible proporcionan herramientas útiles para pronosticar los mercados de valores,

capturando su comportamiento no lineal (Atsalakis G. , Valavanis, 2009). La base de la formulación y selección del proceso estocástico es la comprensión de las características de la muestra de observaciones y el análisis de la generación de los errores. Por lo tanto, para dar cumplimiento al objetivo se plantea la aplicación de los modelos; ARIMA, redes neuronales artificiales (ANN) y el posible modelo combinado (ANN–ARIMA), para capturar comportamientos lineales y no lineales.

Un proyecto de minería de datos tiene varias fases necesarias que son, esencialmente de acuerdo a las etapas del trabajo de investigación a abordar:

1. Comprensión del problema que se quiere resolver.
2. Determinación, obtención y limpieza de los datos necesarios.
3. Creación de modelos matemáticos.
4. Validación y comunicación de los resultados obtenidos.

Para el modelo ARIMA se utilizará la metodología Box – Jenkins, cuyo objetivo es identificar y estimar un modelo estadístico que puede ser interpretado como generador de la información de la muestra.

Las etapas a seguir son:

Etapas 1. Identificación.

Esta fase consiste en detectar el tipo de proceso estocástico que ha generado los datos. Esto significa encontrar los valores adecuados de p , d y q del modelo ARIMA con base al análisis de autocorrelación y autocorrelación parcial.

Etapas 2. Estimación.

En esta etapa se estiman los coeficientes de los términos autorregresivos integrados de media móvil incluidos en el modelo, cuyo número de rezagos p y q ya han sido identificados en la

etapa anterior. Algunas veces la estimación se efectúa por mínimos cuadrados lineales, pero en otras se recurre a la estimación no lineal de los parámetros.

Etapa 3. Verificación de Diagnóstico.

En esta etapa se busca evaluar si el modelo estimado se ajusta a los datos en forma razonablemente buena, ya que es posible que exista otro modelo ARIMA que también lo haga. A esta etapa también se le conoce como validación o comprobación de diagnóstico en la cual se efectúan algunas pruebas antes de hacer uso del modelo para la predicción.

Etapa 4. Pronóstico.

Para pronosticar un período futuro a partir del modelo seleccionado; es decir aquel que es “el mejor” resultante de las etapas anteriores, es importante considerar si la variable original fue diferenciada.

El modelo de redes neuronales artificiales comprende varias fases:

- Fase 1. Definición de la red neuronal. - Se determina el número de neuronas de las capas de entrada, oculta y salida, así como también las diferentes funciones de activación que se utilizarán en las neuronas de la capa oculta y salida.
- Fase 2. Entrenamiento de la red neuronal.- En esta etapa se define el tipo de entrenamiento que se realizará, si es aprendizaje supervisado o no supervisado, también se determinan los algoritmos de entrenamiento.
- Fase 3. Utilización de la red neuronal. - La fase de utilización es propiamente la ejecución de la estructura de la red, se inicia cuando se presenta una entrada a la red y termina generando una salida en función a las entradas proporcionadas.

- Fase 4. Mantenimiento de la red neuronal.- Sobre el mantenimiento se puede decir que la mayoría de las redes se entrenan para solucionar problemas dinámicos, por lo tanto es necesaria e indispensable una validación continua para garantizar una buena utilización.
- Fase 5. Con los resultados obtenidos de la aplicación de los modelos es necesario validar los supuestos de aleatoriedad y normalidad de los residuales para determinar si el modelo encontrado es o no el adecuado para realizar pronósticos.

También, para comprobar la hipótesis de la capacidad de ajuste del modelo econométrico híbrido para la serie histórica de la rentabilidad observada y dar cumplimiento al último objetivo se comparan los modelos y se determina cuál permite una mejor estimación por medio del cálculo de la Raíz cuadrada del error cuadrático medio (RECM), los valores de criterio de información de Akaike (AIC) y el criterio de información bayesiano de Schwarz (SBC).

- Fase 6. Interpretación y evaluación del nuevo conocimiento en el dominio de aplicación (Fayyad et.al, 1996), en esta etapa se analizan los resultados obtenidos en la aplicación de modelo híbrido y de las pruebas realizadas.

Fase 6. Determinar la capacidad predictiva de los modelos: en esta fase mediante métricas de capacidad predictiva de los modelos utilizados se busca comparar y seleccionar el mayor ajuste a la serie de los datos.

6. Análisis de series de tiempo, modelos y resultados

Para dar cumplimiento a los objetivos a continuación se describen las series analizadas de los índices generales de Colombia:

- IGBC: para el análisis de este índice se consideró el periodo desde el 3 de julio del 2001 hasta el 31 de octubre del 2013.

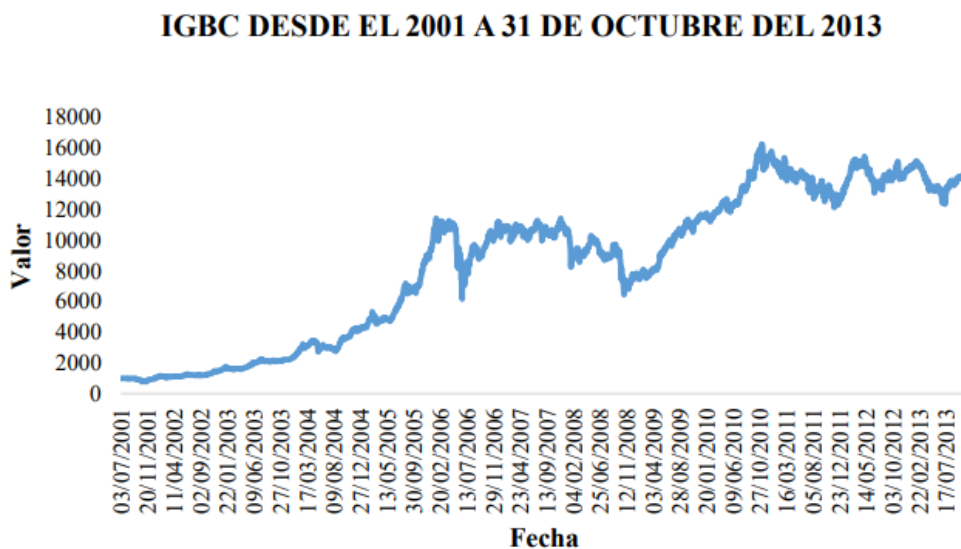
- COLCAP: en cuanto a este índice, el periodo analizado fue desde el 1 de noviembre del 2013 hasta el 28 de diciembre del 2018.

El IGBC es el resultado de ponderar las acciones más líquidas y de mayor capitalización que se negocian en la Bolsa, es decir, aquellas que tienen una mayor rotación y frecuencia. Por tanto, aquellas acciones que tengan una rotación mayor a 0,5% semestral y una frecuencia superior a 40% en el trimestre anterior, son incluidas dentro del IGBC, lo que refleja el comportamiento promedio de los precios de las acciones en el mercado ocasionado por la interacción de la oferta y la demanda, como resultado de las negociaciones de dichas acciones.

Por su parte, el índice COLCAP, que es un Índice Accionario de Capitalización bursátil ajustada, refleja las variaciones de los precios de las 20 acciones más líquidas de la Bolsa de Valores de Colombia (BVC). Por esta razón se optó por trabajar las series por separado.

Figura 2.

IGBC desde el 3 de julio del 2001 hasta 31 de octubre del 2013

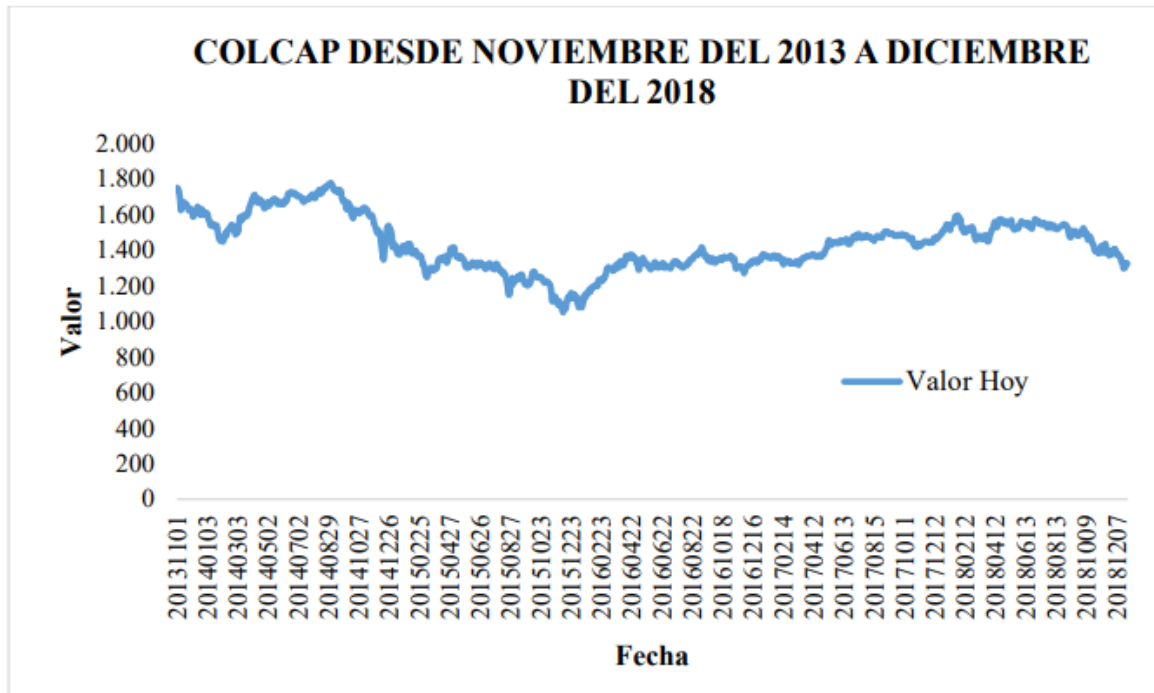


En la Figura 2 se muestran períodos con tendencia al alza o baja de manera marcada y

variabilidad, por lo cual la serie se considera como no estacionaria. Esto hace necesario realizar la transformación de la serie, con la primera diferencia logarítmica.

Figura 3.

COLCAP desde noviembre del 2013 a diciembre del 2018



En la Figura 3.

COLCAP desde noviembre del 2013 a diciembre del 2018 se muestran periodos con tendencia al alza o baja menos marcados que en el IGBC y variabilidad. Aun así, también se trata de una serie no estacionaria. Por lo tanto, se hace necesario realizar la transformación de la serie, con el mismo procedimiento de diferencia logarítmica.

En la tabla se presentan las características descriptivas de los índices:

Tabla 3.

Estadísticas descriptivas de los índices

| | IGBC | COLCAP |
|---------------------|-------------|-------------|
| 1er cuartil | 3278.1 | 1,304.62 |
| Mediana | 9598.0 | 1,307.17 |
| Media | 8504.0 | 1,329.58 |
| 3er cuartil | 12908.1 | 1,330.31 |
| Máximo | 16278.9 | 1,332.85 |
| Mínimo | 776.6 | 1,297.64 |
| Desviación estándar | 4807.533 | 150.07231 |
| Varianza | 23112372 | 22521.69874 |
| Simetría | -0.2886919 | 0.09347 |
| Curtosis | -1.33144982 | -0.43212 |
| Numero de datos | 3019 | 1255 |

El rendimiento de las series de tiempo, son calculados como:

$$r_t = \ln z_t - \ln z_{t-1}$$

Donde, \ln representa el logaritmo natural; z_t y z_{t-1} son los valores de la serie de tiempo en los periodos t y $t-1$, respectivamente. En la Figura 4 se presenta la evolución en el tiempo de los retornos de cada una de las series analizadas.

Figura 4.

Rendimiento de IGBC desde el 3 de julio del 2001 hasta octubre del 2013

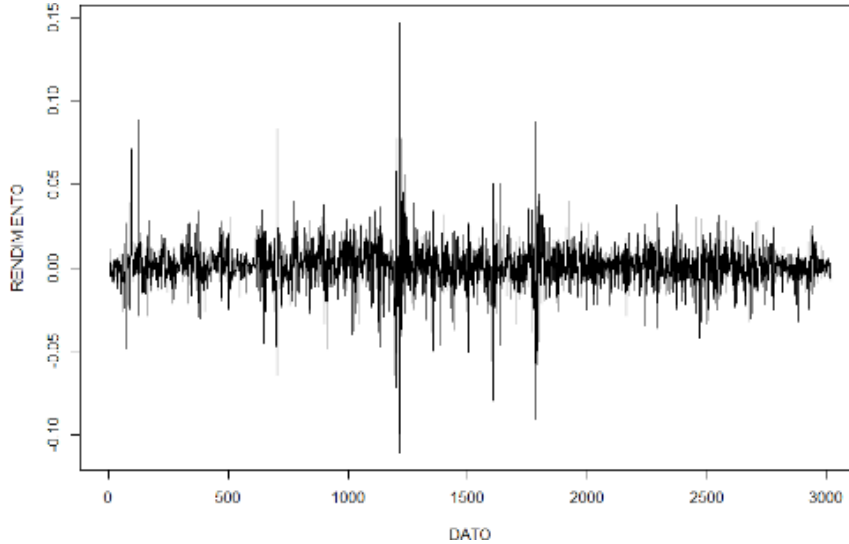
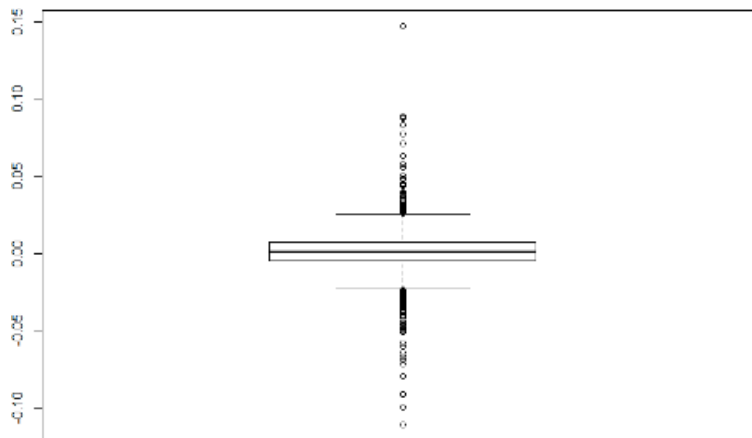


Figura 5.

Diagrama de caja de rendimientos de la serie con el índice IGBC del 2001 al 2013.



En la Figura 4 y Figura 5 se observa que los retornos logarítmicos tienen fluctuaciones, es decir, se presentan períodos de mayor o menor volatilidad. También se observa en el diagrama de caja que la mayoría de los datos se encuentran muy agrupados alrededor de la media, sin embargo, se visualizan varios puntos atípicos para esta serie.

Figura 6.

Rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018

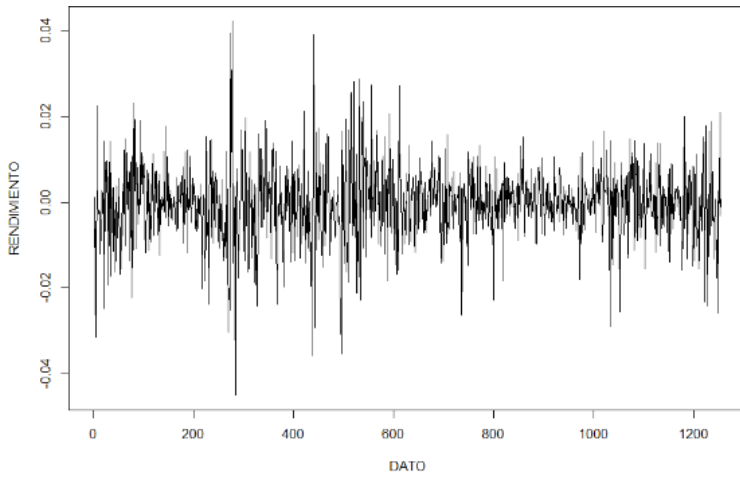
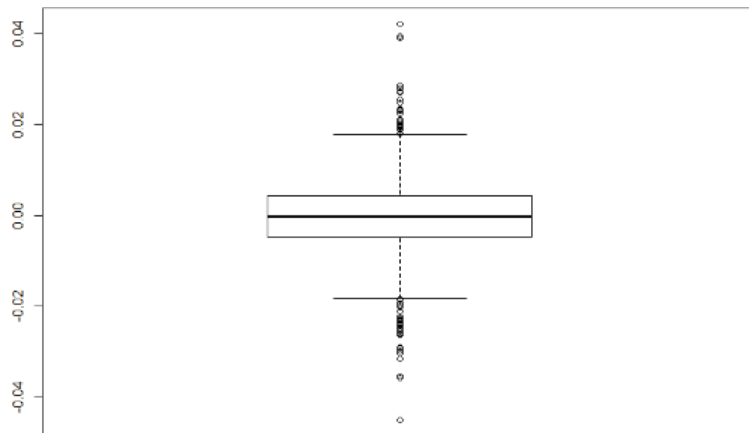


Figura 7.

Diagrama de caja de rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018.



En las Figura 6 y Figura 7 se observa que los retornos logarítmicos de la serie del COLCAP para la ventana de tiempo de noviembre del 2013 a diciembre del 2018 tienen fluctuaciones. Se observan periodos con volatilidad variada.

A continuación, se presenta en la Tabla 4 las estadísticas descriptivas de las rentabilidades de los índices.

Tabla 4.

Estadísticas descriptivas de las rentabilidades de los índices

| | IGBC | COLCAP |
|---------------------|--------------|---------------|
| 1er cuartil | -0.0049449 | -0.0046213 |
| Mediana | 0.0012090 | -0.0001596 |
| Media | 0.000877813 | -0.0002226454 |
| 3er cuartil | 0.0072795 | 0.0044487 |
| Máximo | 0.1468804 | 0.0421546 |
| Mínimo | -0.1105201 | -0.0448979 |
| Desviación estándar | 0.01355241 | 0.00873964 |
| Varianza | 0.0001836679 | 7.638132E-05 |
| Simetría | -0.207678632 | -0.15486 |
| Curtosis | 12.7164662 | 2.59407 |
| Numero de datos | 3018 | 1254 |

En las estadísticas descriptivas se puede observar que las dos series son asimétricas negativas según los criterios de la distribución en el cual,

- Si $g=0$: La distribución es simétrica.
- Si $g>0$: La curva es asimétrica positiva.
- Si $g<0$: La curva es asimétrica negativa.

En las series se presenta un g inferior a cero presentando un sesgo, donde se encuentran valores más separados de la media a la izquierda, la curtosis de las series COLCAP y el IGBC indican que son series de tipo leptocúrticas, en este caso con mayor apuntamiento en la distribución de frecuencias de la serie del IGBC. En la Figura 8, 9, 10 y 11, se presentan los diagramas de autocorrelación y de autocorrelación parcial, las cuales indican los valores de series pasadas más

útiles para predecir valores futuros. Con estos datos se determina el orden de los procesos en un modelo ARIMA.

Figura 8.

Diagrama de ACF de rendimientos de la serie con el índice IGBC del 2001 al 2013

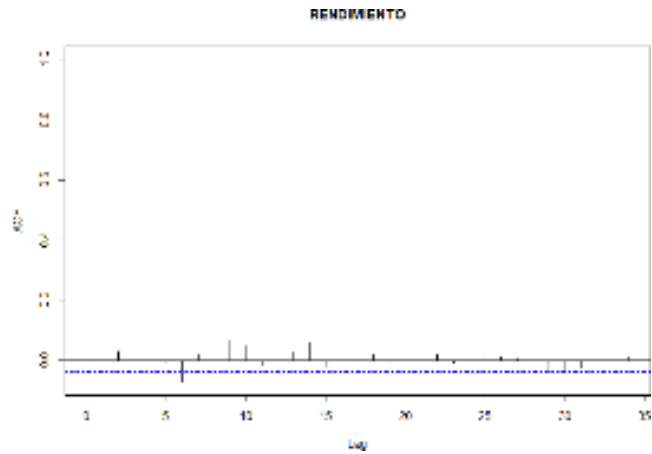


Figura 9.

Diagrama de ACF de rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018

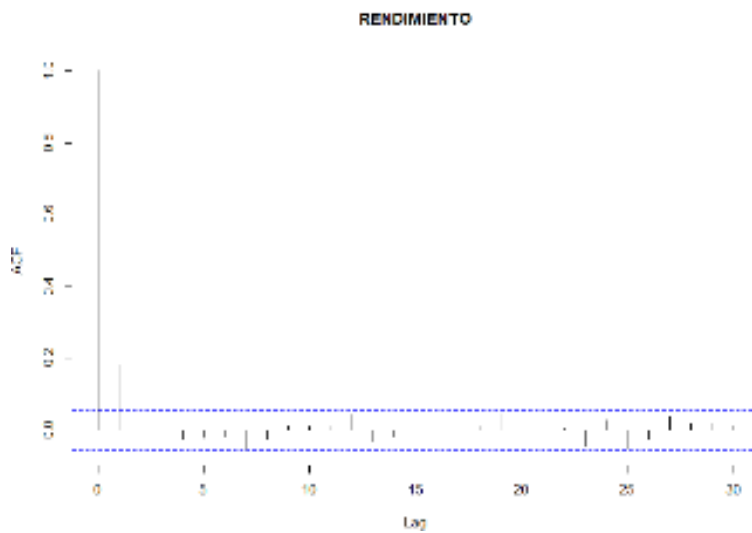


Figura 10.

Diagrama de PACF de rendimientos de la serie con el índice IGBC del 2001 al 2013

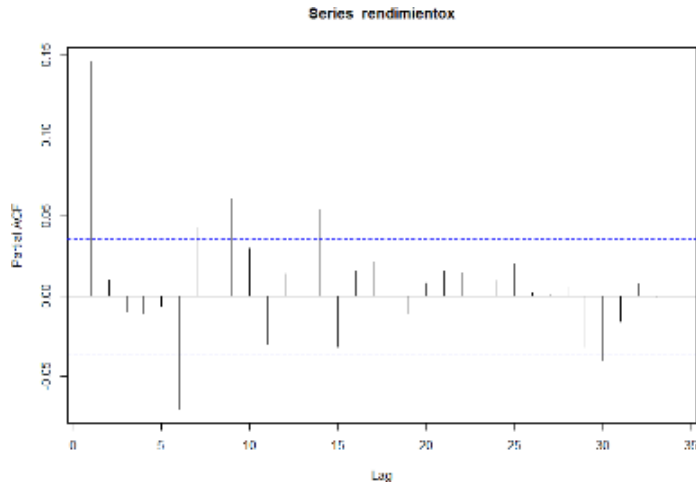
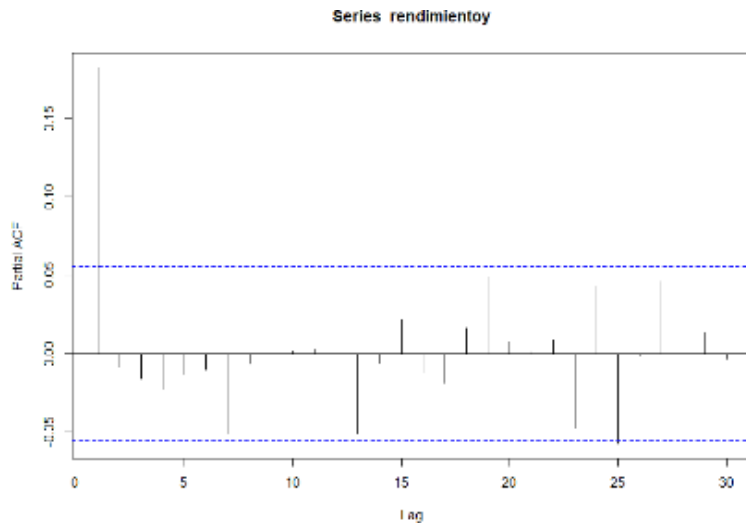


Figura 11.

Diagrama de PACF de rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018



Una racha es una secuencia de observaciones similares, una sucesión de símbolos idénticos consecutivos, en la cual los datos numéricos se transforman en signos y cada cadena máxima de signos iguales es una racha, y su longitud es el número de signos que tiene. En esta prueba se

contrasta si las rachas que se han formado están bajo la hipótesis de aleatoriedad (independencia). En series de tiempo se busca que las observaciones no sean aleatorias, para que una observación cualquiera sea dependiente de las características de la observación anterior y pueda predecirse la serie, (Berlanga, V y Rubio, M, 2012) por lo que una muestra con un número excesivamente grande o excesivamente pequeño de rachas sugiere que la muestra no es aleatoria, lo que rechaza la hipótesis nula (Ho) y acepta la hipótesis alternativa (H1)

Donde, Ho= Los retornos son independientes e idénticamente distribuidos.

H1= Los retornos son dependientes.

$$z = \frac{R - \mu}{\sigma} \text{ Donde } \mu = \left[\frac{2 * n1 * n2}{n1 + n2} \right] + 1$$

$$\sigma = \sqrt{\left[\frac{(2 * n1 * n2 * (2 * n1 * n2 - n1 - n2))}{(n1 + n2)^2 * (n1 + n2 - 1)} \right]}$$

Por lo tanto, si el p-valor es inferior al nivel de confianza estipulado se rechaza la hipótesis nula planteada.

A continuación, se presenta la prueba de rachas de los rendimientos de la serie con el índice IGBC del 2001 al 2013

Tabla 5.

Prueba de rachas de los rendimientos de la serie con el índice IGBC del 2001 al 2013

| Prueba de rachas | |
|------------------------------------|-------------------|
| | Rendimiento |
| valor de prueba^a | 0,000877813045502 |
| casos < valor de prueba | 1463 |
| casos ≥ valor de prueba | 1555 |
| casos totales | 3018 |
| número de rachas | 1278 |
| z | -8,404 |
| Sig. asintótica(bilateral) | 0,000 |
| ^a media | |

Según lo descrito anteriormente, se realizó la prueba de rachas de los rendimientos de la serie con el índice IGBC del 2001 al 2013, dando como resultado un P-valor inferior al nivel de confianza de 0,05 y 0,01, por lo cual se rechaza la hipótesis nula planteada, determinando que la serie en esa ventana de tiempo es dependiente y por lo tanto tiene un nivel de predictibilidad, dado que el mercado no es eficiente durante el periodo.

A continuación, en la Tabla 6 se presenta la prueba de Kolmogorov-Smirnov de los rendimientos de la serie con el índice IGBC del 2001 al 2013.

Tabla 6.

Prueba de Kolmogorov-Smirnov de los rendimientos de la serie con el índice IGBC del 2001 al 2013

| Prueba de Kolmogorov-Smirnov para una muestra | | |
|--|----------------------------|---------------------|
| | N | RENDIMIENTO 3018 |
| Parámetros normales ^{a,b} | Media | 0,000877813045502 |
| | Desv. Desviación | 0,013552411838997 |
| Máximas diferencias | Absoluto | 0,088 |
| extremas | Positivo | 0,077 |
| | Negativo | -0,088 |
| | Estadístico de prueba | 0,088 |
| | Sig. asintótica(bilateral) | 0,000 ^c |
| ^{a.} La distribución de prueba es normal. | | |
| ^{b.} Se calcula a partir de datos. | | |
| ^{c.} Corrección de significación de Lilliefors. | | |

La prueba Kolmogorov-Smirnov para una muestra, se puede describir como un procedimiento de "bondad de ajuste" que mide el grado de concordancia existente entre la distribución de un conjunto de datos y una distribución teórica específica. Lo anterior sirve para contrastar si las observaciones podrían proceder en forma razonable de la distribución

especificada. Con un nivel de significación de 0,05 se puede determinar que la distribución no es normal.

En la Tabla 7 se presenta la prueba de rachas de la rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018.

Tabla 7.

Prueba de rachas de la rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018

| Prueba de rachas | |
|------------------------------|-------------|
| | Rendimiento |
| Valor de prueba ^a | 0,0017253 |
| Casos < Valor de prueba | 772 |
| Casos ≥ Valor de prueba | 485 |
| Casos totales | 1257 |
| Número de rachas | 544 |
| Z | -3,140 |
| Sig. asintótica(bilateral) | 0,002 |

^a. Media

También, se realizó la prueba de rachas de los rendimientos de la serie con el índice COLCAP desde 2013 hasta 2018, dando como resultado un p-valor de 0,002 siendo inferior al nivel de confianza de 0,05 y 0,01, rechazando la hipótesis nula de independencia de la serie, determinando que esta es dependiente y demostrando un nivel de predictibilidad, dado que el mercado no es eficiente durante el periodo.

A continuación, se presenta

Tabla 8.

Prueba de Kolmogorov- Smirnov de la rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018

Tabla 8.

Prueba de Kolmogorov- Smirnov de la rentabilidad del COLCAP desde noviembre del 2013 hasta diciembre del 2018

| Prueba de Kolmogorov-Smirnov para una muestra | | |
|--|--|--------------------------|
| | | RENDIMIENTO |
| | N | 1257 |
| Parámetros normales^{a,b} | Media | 0,0017253 |
| | Desv. Desviación | 0,07382458 |
| Máximas diferencias | Absoluto | 0,395 |
| | Positivo | 0,395 |
| Extremas | Negativo | -0,365 |
| | Estadístico de prueba | 0,395 |
| | Sig. asintótica(bilateral) | 0,000^c |
| | ^{a.} La distribución de prueba es normal. | |
| | ^{b.} Se calcula a partir de datos. | |
| | ^{c.} Corrección de significación de Lilliefors. | |

Con la prueba Kolmogorov-Smirnov para una muestra, se puede concluir a partir del p-valor menor al nivel de significancia de 0,05 que no tiene una distribución normal.

6.1. Criterio de Información de Akaike

Para el análisis fue utilizado el criterio de información de Akaike (An Information Criterion, AIC) el cual proporciona un método simple y objetivo que selecciona el modelo más adecuado para caracterizar los datos experimentales.

Este criterio, se define como: $AIC = -2 \log(L(\theta)) + 2K$

Donde $\log(L(\theta))$ es el logaritmo de la máxima verosimilitud, que permite determinar los valores de los parámetros libres de un modelo estadístico, y K es el número de parámetros libres del modelo (Martínez, D; Albín, J, et. Al, 2009).

Para la selección de modelos con el criterio de información de Akaike se hace una búsqueda del modelo que presente un menor valor AIC, entre un conjunto de modelos candidatos.

Dentro de los modelos encontrados en las series, se presenta a continuación, una descripción generalizada del modelo, las reglas y el cálculo de la varianza en la Tabla 9.

Tabla 9.

Modelos de ajuste

| Modelos de ajuste | | | |
|-------------------|---|---|--|
| Modelo | Descripción | Regla | Var |
| AR1 | AR1 (μ, σ, Y_0) genera un proceso autorregresivo de primer orden (AR1) con una media μ , parámetro de volatilidad σ , coeficiente autorregresivo a_1 , y valor Y_0 en tiempo 0. Se caracteriza por una función de autocorrelación (ACF) que se reduce geométricamente y una función de autocorrelación parcial (PACF) que corta hasta 0 después de un retraso de 1 | $ a_1 < 1$ Es una condición necesaria para la estacionariedad | $\text{Var}(Y_t) = \frac{\sigma^2}{(1-a_1^2)}$ |
| MA1 | MA1 ($\mu, \sigma, b_1, \varepsilon_0$) genera un proceso de promedio móvil de primer orden con una media μ , parámetro de volatilidad σ , coeficiente de promedio móvil b_1 , y término de error inicial ε_0 . | | $\text{Var}(Y_t) = \frac{\sigma^2}{(1+b_1^2)}$ |
| ARMA11 | ARMA11 ($\mu, \sigma, a_1, b_1, Y_0, \varepsilon_0$) genera un proceso de promedio móvil autorregresivo de primer orden (ARMA11) con una media μ , parámetro de volatilidad σ , coeficiente autorregresivo de promedio móvil b_1 , valor Y_0 , en tiempo 0, y término de error inicial ε_0 . | $ a_1 < 1$ Es una condición necesaria para la estacionariedad | $\text{Var}(Y_t) = \frac{\sigma^2}{(1-a_1^2)(1+b_1^2+2a_1b_1)}$ |
| GARCH11 | GARCH11 ($\mu, \omega, b_1, a_1, Y_0, \sigma_0$) genera un proceso ARCH generalizado con media μ , parámetro de volatilidad ω , coeficiente de error b_1 , coeficiente autorregresivo a_1 , valor Y_0 en tiempo 0, y desviación estándar inicial σ_0 . | $b_1 \geq 0, a_1 \geq 0$, al menos uno de los dos, b_1 o a_1 , es positivo, $\omega > 0$ | $\sigma_t^2 = \omega + b_1(Y_{t-1} - \mu)^2 + a_1\sigma_{t-1}^2$ |

6.2. Modelos para la serie del COLCAP

En la Tabla 10 se encuentran los modelos que se ajustan a la serie de rendimientos con el índice COLCAP desde el 2013 al 2018 según el criterio AIC con el fin de identificar los parámetros de cada uno de los modelos.

Tabla 10.

Criterio AIC para la serie del COLCAP

| MODELO | AIC |
|---------------|------------|
| MA(1) | -8378,73 |
| AR(1) | -8373,12 |
| MA(2) | -8372,40 |
| AR(2) | -8371,26 |
| ARMA(1,1) | -8367,63 |
| BMMR | -8363,93 |
| ARCH(1,1) | -8323,73 |
| GARCH(1,1) | -8321,66 |

En la Tabla 11 se encuentran los parámetros de los modelos para la serie del COLCAP comprendida desde noviembre del 2013 hasta diciembre del 2018.

Tabla 11.

Modelos ajustados serie del COLCAP

| MODELO DE LA SERIE | | | | | | | | | | | |
|--------------------|--------------|---------------------|---------------------------------|---|-----------------|--------------------|----------|---------------|--------------|----------------|--------------------|
| MODELO | | μ | P. volatilidad Σ | a_1 | b_1 | a_2 | b_2 | ϵ_0 | ϵ_1 | Y ₀ | Y ₋₁ |
| I | MA1 | 0,0002179 | 0,008695 | | 0,1820 1 | | | 0,002167 1 | | | |
| II | AR1 | - 0,0002179 | 0,0085924 | 0,18164 | | | | | | 0,00070 919 | |
| III | MA2 | - 0,000217 9 | 0,00846 17 | | 0,16011 | | 0,027389 | 0,001355 4 | | 0,00627 38 | |
| IV | AR2 | - 0,0002179 | 0,008592 1 | 0,18319 | | - 0,008496 3 | | | | 0,00070 919 | - 0,00313 47 |
| V | ARM A1,1 | - 0,000755 51 | 0,008731 1 | 0,064646 | 0,125 | | | 0,002431 1 | | 0,00070 919 | |
| VI | BMM R | - 0,0002233 6 | 0,016133 | velocidad de reversión $\alpha = 1,7065$ | | | | | | 0,00070 919 | |
| VII | ARCH 1,1 | - 0,0002179 | 0,000078 128 | | 0,00012 38 | | | | | 0,00070 919 | |
| VIII | GARC H1,1 | - 0,0002179 | 0,0000746 66/ $\sigma_0 = 0$ | 0,000054 409 | 0,00004 3993 | | | | | 0,00070 919 | |

6.3. IBC

En la Tabla 12 se encuentran los modelos que se ajustan a la serie de rendimientos con el índice IBC del 2001 al 2013 según el criterio AIC.

Tabla 12.

Criterio AIC para la serie del IBC

| IBC | |
|-------------------|------------|
| Ajustar | AIC |
| AR(1) | -17457,241 |
| MA(1) | -17456,843 |
| AR(2) | -17455,575 |
| BMMR | -17448,778 |
| MA(2) | -17445,337 |
| ARMA(1,1) | -17438,422 |
| ARCH(1) | 17386,882 |
| GARCH(1,1) | 17384,404 |

A continuación, se presentan los parámetros de los modelos ajustados en la Tabla 13 para el índice IBC en el periodo de estudio del 2001 al 2013.

PREDICTIBILIDAD DE LA RENTABILIDAD

Tabla 13.

Modelos ajustados serie del IBC

| MODELO DE LA SERIE | | | | | | | | | | | |
|--------------------|-------------|------------|-------------------------------|---|------------|-----------|----------|-----------------|--------------------|------------|------------|
| MODELO | | μ | P. volatilidad σ | a_1 | b_1 | a_2 | b_2 | ε_0 | ε_{-1} | Y_0 | Y_{-1} |
| I | AR1 | 0,00087781 | 0,013407 | 0,146 | | | | | | -0,0073012 | |
| II | MA1 | 0,00087781 | 0,013727 | | 0,15757 | | | - 0,0072322 | | | |
| III | AR2 | 0,00087781 | 0,013407 | 0,14455 | | 0,0099186 | | | | -0,0073012 | - 0,005178 |
| IV | BMMR | 0,00087414 | 0,026582 | Velocidad de reversión $\alpha = 1,9244$ | | | | | | -0,0073012 | |
| V | MA2 | 0,00087781 | 0,013736 | | 0,19045 | | 0,018337 | - 0,0070492 | -0,005926 | | |
| VI | ARMA 1,1 | 0,0018661 | 0,013649 | 0,42772 | -0,30135 | | | -0,008097 | | -0,0073012 | |
| VII | ARCH1 | 0,00087781 | 0,00018714 | | 0,0001003 | | | | | -0,0073012 | |
| VIII | GARCH(1,1) | 0,00087781 | 0,00017841/ $\sigma_0 = 0$ | 0,00012995 | 0,00010512 | | | | | -0,0073012 | |

PREDICTIBILIDAD DE LA RENTABILIDAD

6.4. Serie junta

En la Tabla 14 se presentan los modelos de ajuste.

Tabla 14.

Criterio AIC para la serie junta

| SERIE TOTAL | |
|--------------------|------------|
| Ajustar | AIC |
| MA(1) | -25407,922 |
| ARMA(1,1) | -25263,342 |
| AR(2) | -24398,496 |
| AR(1) | -24031,465 |
| ARCH(1) | -23161,496 |
| GARCH(1,1) | -23161,693 |

En Tabla siguiente se encuentran los parámetros de los modelos con mayor ajuste a la serie de rendimientos del IBC y el COLCAP con modelos tradicionales.

PREDICTIBILIDAD DE LA RENTABILIDAD

Tabla 15.

Modelos ajustados de la serie conjunta

| MODELO DE LA SERIE | | | | | | | | | | | |
|--------------------|----------|---------------|----------------------------|------------|------------|---------|-------|-----------------|--------------------|-----------|----------|
| MODELO | | μ | P. volatilidad σ | a_1 | b_1 | a_2 | b_2 | ε_0 | ε_{-1} | Y_0 | Y_{-1} |
| I | MA1 | -0,0000017983 | 0,012461 | | -0,97984 | | | 0,0023336 | | | |
| II | ARMA1,1 | 0,0001604 | 0,01247 | 0,10332 | -0,93468 | | | 0,00062468 | | 0,0038439 | |
| III | AR2 | -0,0000017983 | 0,013906 | -0,55209 | | -0,2876 | | | | 0,0038439 | -0,02408 |
| IV | AR1 | -0,0000017983 | 0,01452 | -0,42878 | | | | | | 0,0038439 | |
| V | ARCH1 | -0,0000017983 | 0,00026207 | | 0,00010844 | | | | | 0,0038439 | |
| VI | GARCH1,1 | -0,0000017983 | 0,00026205/ $\sigma_0 = 0$ | 0,00019105 | 0,00015453 | | | | | 0,0038439 | |

PREDICTIBILIDAD DE LA RENTABILIDAD

A partir de los resultados para la serie histórica del índice bursátil IGBC y COLCAP del 2001 al 2018, se encontraron periodos con marcada variabilidad, convirtiéndose en una serie no estacionaria en su varianza.

Se observa que para el IGBC, el COLCAP y la serie unificada pueden ajustarse a modelos sencillos como MA(1), AR(1), MA(2), AR(2), ARMA(1,1), ARCH(1,1), GARCH(1,1).

La sencillez de estos modelos univariantes radica en que permiten explicar el valor que la variable que presenta dependencia temporal toma en el momento t , de tal manera que la información pasada y su evolución en el tiempo permite explotar el patrón de regularidad que muestran los datos; específicamente, la variable de interés o de salida depende linealmente de sus valores anteriores.

En el caso de los modelos ARCH (1,1) (AutoRegressive Conditional Heteroscedastic) y GARCH (1,1), que también se encuentran en la serie pueden ser una alternativa exitosa en el ajuste a pesar de la dirección del mercado y es el mejor describiendo la volatilidad de la serie. Estos modelos capturan las agrupaciones de volatilidad de las rentabilidades a través de la varianza condicional, encontrando la volatilidad promedio a mediano plazo mediante una autorregresión que depende de la suma de perturbaciones rezagadas y de la suma de varianzas rezagadas. Por lo tanto, la varianza futura depende de la varianza histórica y varía en función de las observaciones, además trata el exceso de curtosis.

Adicionalmente, se puede observar que en la serie conjunta no se presenta el modelo BMMR, dado que esta supone un Movimiento Browniano Geométrico que sigue el valor de la serie, también, este incluye que el cambio en los índices es estacionario; por tanto, la media y volatilidad no cambian en el tiempo. El cambio es independiente; no hay una correlación significativa con el cambio en el valor anterior y la serie debería tener una distribución normal.

PREDICTIBILIDAD DE LA RENTABILIDAD

Sin embargo, se debe tener en cuenta que, al combinar los dos índices, la transición del IGBC al COLCAP produce una mayor variabilidad en los datos. En los Apéndice B y C se encuentra mayor información sobre las gráficas de los modelos y el código de Python para ajuste de los índices (Apéndice B. Gráficas de los modelos con mayor AIC), (Apéndice C. Resultados del ajuste del IBC mediante Python).

Adicionalmente, es importante medir la predictibilidad “instantánea” con el fin de describir mejor el fenómeno y determinar el algoritmo que mejor predice. La medida de predictibilidad utilizada es la de Galbraith (2003) con la siguiente función:

$$C(s) = 1 - \frac{MSE_{\tilde{r}(s)}}{MSE_{\hat{r}(s)}}$$

Donde $\tilde{r}(s)$ es la media muestral, $\hat{r}(s)$ es la predicción s - pasos hacia adelante y MSE denota el error cuadrático medio.

Tomando en cuenta esta métrica se calcula el $MSE_{\tilde{r}(s)}$ de la serie del COLCAP y el IGBC donde el $MSE_{\tilde{r}(s)} = \frac{1}{N} \sum (r_i - \hat{r}_i)^2$ donde \hat{r}_i es la media muestral y r_i es el dato actual dando como resultado:

Para el COLCAP

$$MSE_{\tilde{r}(s)} = 22521,70$$

$$RMSE_{\tilde{r}(s)} = 150,072312$$

Para el IGBC

$$MSE_{\tilde{r}(s)} = 23104716,62$$

$$RMSE_{\tilde{r}(s)} = 4806,736587$$

En el caso de $MSE_{\hat{r}(s)}$ se utiliza cada uno de los modelos con el fin de comparar la predictibilidad, donde: $MSE_{\hat{r}(s)} = \frac{1}{N} \sum (y_i - \hat{y}_i)^2$, y_i es el dato actual y \hat{y}_i es el dato predicho

PREDICTIBILIDAD DE LA RENTABILIDAD
bajo la metodología con un paso adelante.

A continuación, se presenta los datos de la métrica para la metodología ARIMA (1,1,1)

Métrica con la serie IGBC = 0,8965053

Métrica con la serie COLCAP = 0,979972

6.5 Redes neuronales artificiales

Se puede definir como un sistema que toma valores numéricos de entrada, realiza permutaciones con esos valores, y genera uno o más valores numéricos de salida. Cuando se diseña y entrena una red neuronal para una aplicación específica, los valores de salida son aproximadamente correctos para valores de entrada determinados. Estas son útiles ya que, son capaces de aprender complejas relaciones de datos, diferenciando patrones de datos y luego extrapolar predicciones cuando se introducen nuevos datos.

Los problemas en los que se usan redes neuronales se pueden dividir en dos grupos generales:

- Problemas de clasificación: Problemas en los que se trata de determinar el tipo de categoría al que pertenece un elemento desconocido.
- Problemas numéricos: Situaciones en las que se debe predecir un resultado numérico específico. Como ejemplo están la previsión del precio de acciones o la predicción del nivel de ventas durante un periodo de tiempo futuro.

Las redes neuronales se desarrollan y utilizan en cuatro pasos:

- Preparación de datos – Los datos que se usan se definen en conjuntos de datos.
- Entrenamiento – Con el entrenamiento se genera una red neuronal a partir de un conjunto de datos compuesto de casos con valores de salida conocidos. Estos datos son frecuentemente casos históricos de los cuales se conocen los valores de salida / variable dependiente.

PREDICTIBILIDAD DE LA RENTABILIDAD

- Prueba –evalúa la predicción de los valores de salida conocidos. Los datos usados para las pruebas son normalmente un subconjunto de los datos históricos. Este subconjunto no se utiliza para el entrenamiento de la red. Después de la prueba, se mide el funcionamiento de la red mediante estadísticas como el porcentaje de respuesta conocida cuya predicción ha sido correcta.
- Predicción –Una vez entrenada y probada, la red se puede usar para predecir salidas de datos de casos nuevos.

La predicción numérica se puede realizar mediante redes MLF (Redes multinivel siempre hacia adelante), así como Redes neuronales de regresión generalizada (GRN), que están estrechamente relacionadas con las redes PN (redes neuronales probabilísticas).

Para determinar la mejor configuración de red se utilizan los datos de prueba, es por esto que se realizaron diferentes configuraciones posibles, y se identificó la de mejor funcionamiento. (Apéndice D. Ajuste a las RNA para las series)

6.5.1 Métodos estadísticos

La estructura de una red neuronal consta de unidades conectadas denominadas “nodos” o “neuronas”. Cada neurona realiza una parte del cálculo dentro de la red: la neurona toma algunos números como valores de entrada y genera un valor de salida que se convierte en uno de los valores de entrada de otra neurona, excepto en las neuronas que generan los valores de salida finales de todo el sistema.

Las neuronas están organizadas en capas o niveles, las de capas de entrada reciben los valores de entrada para los cálculos, como la longitud, diámetro y peso. Estos valores se trasladan a las neuronas de la primera capa oculta, que realizan los cálculos de sus valores de entrada y trasladan sus valores de salida a la siguiente capa. Los valores de salida de las neuronas de la última

PREDICTIBILIDAD DE LA RENTABILIDAD

capa oculta se trasladan a la neurona o neuronas que genera los resultados de salida finales de la red.

6.5.2 Predicción numérica

El entrenamiento de una red, es un proceso de ajuste fino de los parámetros de un cálculo cuyo propósito es generar valores de salida que sean aproximadamente correctos para una serie de valores de entrada. Este proceso está dirigido por los datos de entrenamiento y por el algoritmo de entrenamiento. Este último, selecciona varios grupos de parámetros de cálculo y evalúa cada grupo aplicando la red a cada caso de entrenamiento para determinar lo aproximadas que son las respuestas dadas por la red.

Cada grupo de parámetros es una “prueba”; el algoritmo de entrenamiento selecciona nuevos grupos de parámetros basándose en los resultados de pruebas anteriores.

La red multinivel siempre hacia adelante: se puede especificar si debe haber una o dos capas de neuronas ocultas, y cuántas neuronas deben contener, mientras que las redes neuronales de regresión generalizada -GRN tienen siempre dos capas ocultas de neurona, con una neurona por caso entrenado en la primera capa oculta, y un tamaño de la segunda capa determinado por ciertas características de los datos de entrenamiento.

El funcionamiento de las redes MLF se determina por:

- Su topología (el número de capas ocultas y el número de nodos de esas capas)
- Los “pesos” de las conexiones (un parámetro asignado a cada conexión) y los términos de tendencia (un parámetro asignado a cada neurona)
- La función de activación que se usa para convertir los valores de entrada de cada neurona en su valor de salida. Concretamente, se usa la función de tangente hiperbólica.

PREDICTIBILIDAD DE LA RENTABILIDAD

En cuanto al entrenamiento de esta red, consiste en encontrar una serie de pesos y términos de tendencia para las conexiones que permitan que la red genere respuestas correctas cuando se le presentan nuevos casos. Esto comienza con la asignación de una serie de pesos de conexiones seleccionados aleatoriamente. Se hace una predicción para cada caso del entrenamiento (usando valores independientes como valores de entrada para obtener un valor de salida). Si el valor de salida es diferente al conocido proporciona un valor de error y a partir de estos valores se calcula una medida de error para el conjunto de entrenamiento completo, que indica el funcionamiento de la red con los pesos iniciales utilizados.

Para el algoritmo de entrenamiento el método busca minimizar la medida de error cambiando los pesos de las conexiones, mediante el método de “Gradiente descendente conjugado”, el cual ha sido diseñado para encontrar el mínimo local de una función. Para reducir el riesgo de encontrar el mínimo local en lugar del global, combina métodos de optimización “determinísticos” y “estocásticos”. Más específicamente, se usa el método estocástico “Recocido Simulado” junto con el método Gradiente Descendente Conjugado por el cual el algoritmo decide el método que se debe usar en un punto determinado, basándose en los resultados de pruebas anteriores.

La medición del error que se utiliza cuando se entrenan redes de predicción numérica es el error cuadrático medio de todos los casos del entrenamiento; es decir, el cuadrado de la media de la diferencia entre la respuesta correcta y la respuesta dada por la red. La selección del número de capas y del número de neuronas de cada capa determina si la red es capaz de aprender la relación entre las variables independientes y las dependientes. Normalmente, una sola capa oculta es suficiente para casi cualquier problema; además, unas pocas neuronas en una sola capa ocultan normalmente son suficientes.

PREDICTIBILIDAD DE LA RENTABILIDAD

Ventajas de las redes GRN/PN:

Se entrenan rápido y no requieren una especificación de topología (número de capas ocultas y nodos).

Ventajas de las redes MLF:

Son más rápidas para hacer predicciones, son más fiables fuera del rango de los datos de entrenamiento y son capaces de generalizar a partir de conjuntos de entrenamiento muy pequeños.

6.5.3 Estadísticas del ajuste de redes neuronales a las series

Las Redes multinivel siempre hacia adelante (MLF) tienen uno o dos niveles de nodos ocultos.

Para la selección de la topología de la red neuronal se realizaron iteraciones a fin de determinar el mejor ajuste de la misma, para ello, ambas series se entrenaron con el 80% de los datos y su evaluación se realizó con el 20% restante, se ejecutaron pruebas hasta con 2 capas ocultas, de 1 a 4 rezagos, de 1 a 6 y 10 neuronas para las redes con una sola capa oculta, para los que contenían 2 capas se realizaron diversas combinaciones posibles ejemplo: (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (1,10) entre otras.

En la predicción numérica, las estadísticas y gráficas clave del informe de resumen de entrenamiento son las siguientes:

- 1) % de predicciones incorrectas – Una predicción cuenta como “incorrecta” si se encuentra fuera del margen definido alrededor del valor (“Tolerancia correcta / incorrecta (entrenamiento)”).
- 2) Error cuadrático medio – Es una medida de desviación de las predicciones con respecto al valor (calculado como la raíz cuadrada de la desviación promedio al cuadrado).
- 3) Error absoluto medio – Desviación promedio de las predicciones con respecto a los valores.

PREDICTIBILIDAD DE LA RENTABILIDAD

4) Histograma de residuales – “Residual” es la diferencia que hay entre el valor real y el de la predicción.

5) Diagramas de dispersión que muestran las relaciones entre los valores reales, los valores de predicción y los residuales.

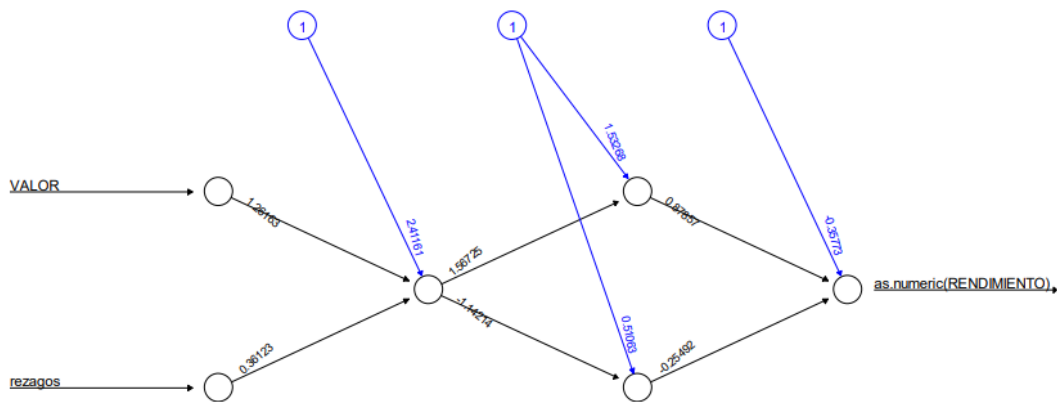
6) Error residual – la diferencia entre el valor dependiente real y el de la predicción.

Ajuste de RNA para la serie del IBC

En la Figura 12 se presenta la red neuronal con los parámetros de la misma.

Figura 12.

Red neuronal IGBC

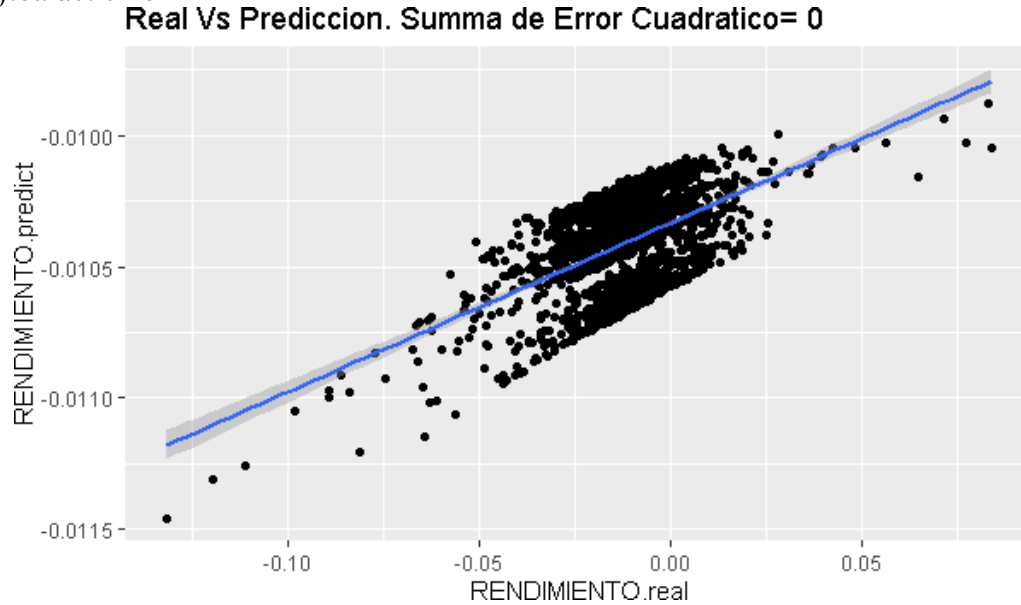


Contiene una sola capa de entrada, dos capas ocultas y una capa de salida, en la primera capa oculta se tiene una sola neurona y en la segunda dos neuronas con un rezago de 4.

PREDICTIBILIDAD DE LA RENTABILIDAD

Figura 13.

Gráfica del error



En la Figura 13 se presenta la gráfica del error para el modelo ajustado, pudiendo extraer que a pesar de tener el menor MSE es importante resaltar que en las iteraciones realizadas, la red presentada en la figura 13 con mayor ajuste, son las que tienen un solo rezago y con dos capas ocultas y pocas neuronas ((1,1), (1,2), (1,3))

Para el COLCAP

Tabla 16.

Prueba de aleatoriedad

| | Valor Hoy | RENDIMIENTO |
|--|-----------|-------------|
| <i>Prueba de ejecuciones para Aleatoriedad</i> | | |
| Observaciones | 1255 | 1255 |
| Por debajo Media | 635 | 618 |
| Por encima Media | 620 | 637 |
| Número de ejecuciones | 22 | 597 |
| Media | 1435,93 | -0,000218 |
| E(R) | 628,4104 | 628,3562 |
| Desv Est (R) | 17,7034 | 17,7019 |
| Valor Z | -34,2539 | -1,7713 |
| Valor P (dos opciones) | < 0.0001 | 0,0765 |

PREDICTIBILIDAD DE LA RENTABILIDAD

En la Tabla 16 se encuentra el resumen de las estadísticas básicas de la serie del COLCAP, como son el conjunto de datos arriba y abajo de la media, el valor de la media, la desviación estándar y el p-valor del valor histórico y del rendimiento del mismo.

Tabla 17.

Rango Intercuartil

| | | |
|--------------------|---------|-----------|
| Rango intercuartil | 208,28 | 0,009109 |
| 1,00% | 1098,32 | -0,025170 |
| 2,50% | 1129,78 | -0,019376 |
| 5,00% | 1198,51 | -0,014495 |
| 10,00% | 1255,39 | -0,010171 |
| 20,00% | 1314,38 | -0,006125 |
| 80,00% | 1557,59 | 0,005750 |
| 90,00% | 1662,73 | 0,009462 |
| 95,00% | 1700,91 | 0,013824 |
| 97,50% | 1725,13 | 0,016658 |
| 99,00% | 1749,93 | 0,022457 |

En la Tabla 17 se encuentra el resumen de las estadísticas para graficar el diagrama de caja para la serie de datos históricos y del rendimiento del COLCAP.

Por otra parte, en las siguientes tablas se presenta la prueba chi cuadrado y los errores según los retardos.

PREDICTIBILIDAD DE LA RENTABILIDAD

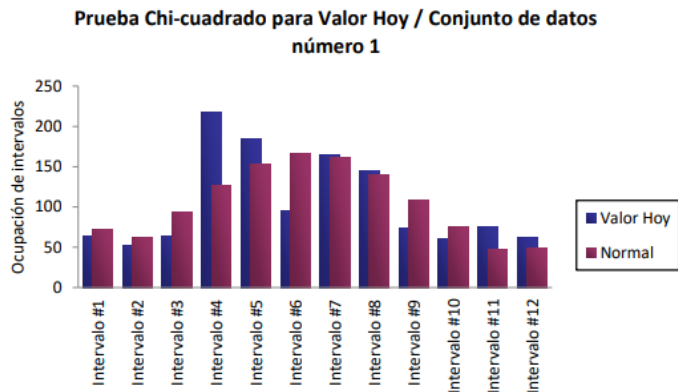
Tabla 18.

Prueba chi cuadrado

| Prueba de Normalidad | | | | | |
|--------------------------------|----------------------------|---------------|------|----------|-----------|
| Valor Hoy | | | | | |
| <i>Prueba Chi-Cuadrado</i> | Conjunto de datos número 1 | | | | |
| Media | 1435,93 | | | | |
| Desviación Estándar | 150,13 | | | | |
| Estadística Chi-Cuadrado | 147,9038 | | | | |
| Valor P | < 0.0001 | | | | |
| <i>Intervalos Chi-cuadrado</i> | Intervalo Mín | Intervalo Máx | Real | Normal | Distancia |
| Intervalo #1 | -Inf | 1198,51 | 63 | 71,3967 | 0,9875 |
| Intervalo #2 | 1198,51 | 1248,75 | 53 | 61,9306 | 1,2878 |
| Intervalo #3 | 1248,75 | 1298,99 | 63 | 93,6319 | 10,0213 |
| Intervalo #4 | 1298,99 | 1349,23 | 217 | 126,6946 | 64,3679 |
| Intervalo #5 | 1349,23 | 1399,47 | 184 | 153,4297 | 6,0910 |
| Intervalo #6 | 1399,47 | 1449,71 | 96 | 166,2949 | 29,7145 |
| Intervalo #7 | 1449,71 | 1499,95 | 164 | 161,3121 | 0,0448 |
| Intervalo #8 | 1499,95 | 1550,19 | 145 | 140,0469 | 0,1752 |
| Intervalo #9 | 1550,19 | 1600,43 | 73 | 108,8172 | 11,7892 |
| Intervalo #10 | 1600,43 | 1650,67 | 60 | 75,6724 | 3,2459 |

Figura 14.

Prueba Chi cuadrado



PREDICTIBILIDAD DE LA RENTABILIDAD

Tabla 19.

Autocorrelación

| <i>Tabla de autocorrelación</i> | Conjunto de datos número 1 |
|---------------------------------|----------------------------|
| Número de valores | 1255 |
| Error estándar | 0,0282 |
| Retraso número 1 | 0,1815 |
| Retraso número 2 | 0,0247 |
| Retraso número 3 | -0,0133 |
| Retraso número 4 | -0,0283 |
| Retraso número 5 | -0,0241 |
| Retraso número 6 | -0,0151 |
| Retraso número 7 | -0,0535 |
| Retraso número 8 | -0,0244 |
| Retraso número 9 | 0,0103 |
| Retraso número 10 | 0,0081 |
| Retraso número 11 | 0,0079 |
| Retraso número 12 | 0,0426 |
| Retraso número 13 | -0,0337 |
| Retraso número 14 | -0,0186 |
| Retraso número 15 | 0,0151 |
| Retraso número 16 | -0,0077 |
| Retraso número 17 | -0,0207 |
| Retraso número 18 | 0,0090 |
| Retraso número 19 | 0,0467 |
| Retraso número 20 | 0,0281 |
| Retraso número 21 | 0,0112 |
| Retraso número 22 | 0,0084 |
| Retraso número 23 | -0,0456 |
| Retraso número 24 | 0,0272 |
| Retraso número 25 | -0,0486 |
| Retraso número 26 | -0,0237 |
| Retraso número 27 | 0,0387 |
| Retraso número 28 | 0,0150 |
| Retraso número 29 | 0,0158 |
| Retraso número 30 | 0,0113 |

PREDICTIBILIDAD DE LA RENTABILIDAD

A continuación, también se presenta en la topología de la red la cual es la siguiente: contiene una sola capa de entrada, una capa oculta y una capa de salida, en la primera capa oculta se tiene diez neuronas y un solo rezago; se presenta en la Figura 16.

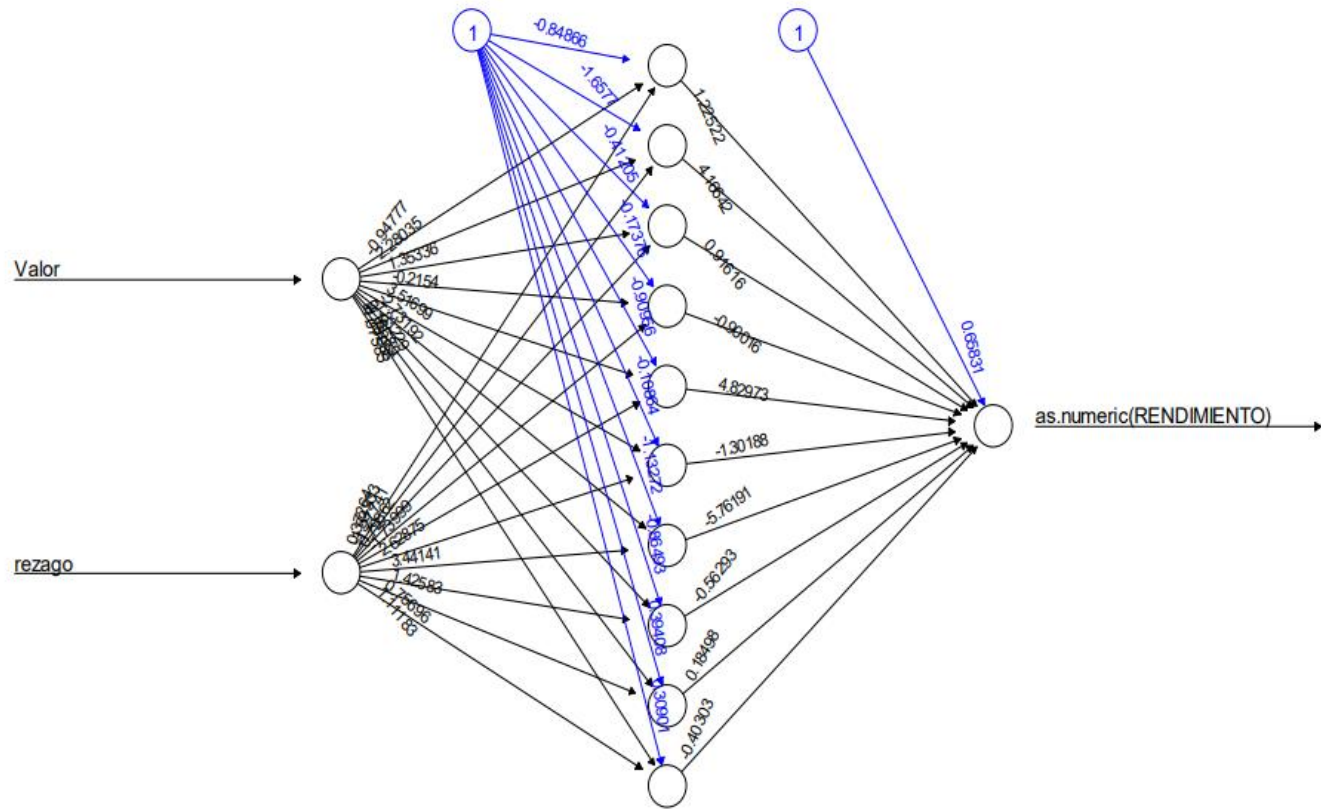
Diagrama del error para el COLCAP:

De las iteraciones realizadas el menor MSE pertenece a la red con un solo rezago y hasta 10 neuronas con un valor de $7,7609E-06$ el cual genera el mejor ajuste al comportamiento del índice.

PREDICTIBILIDAD DE LA RENTABILIDAD

Figura 15.

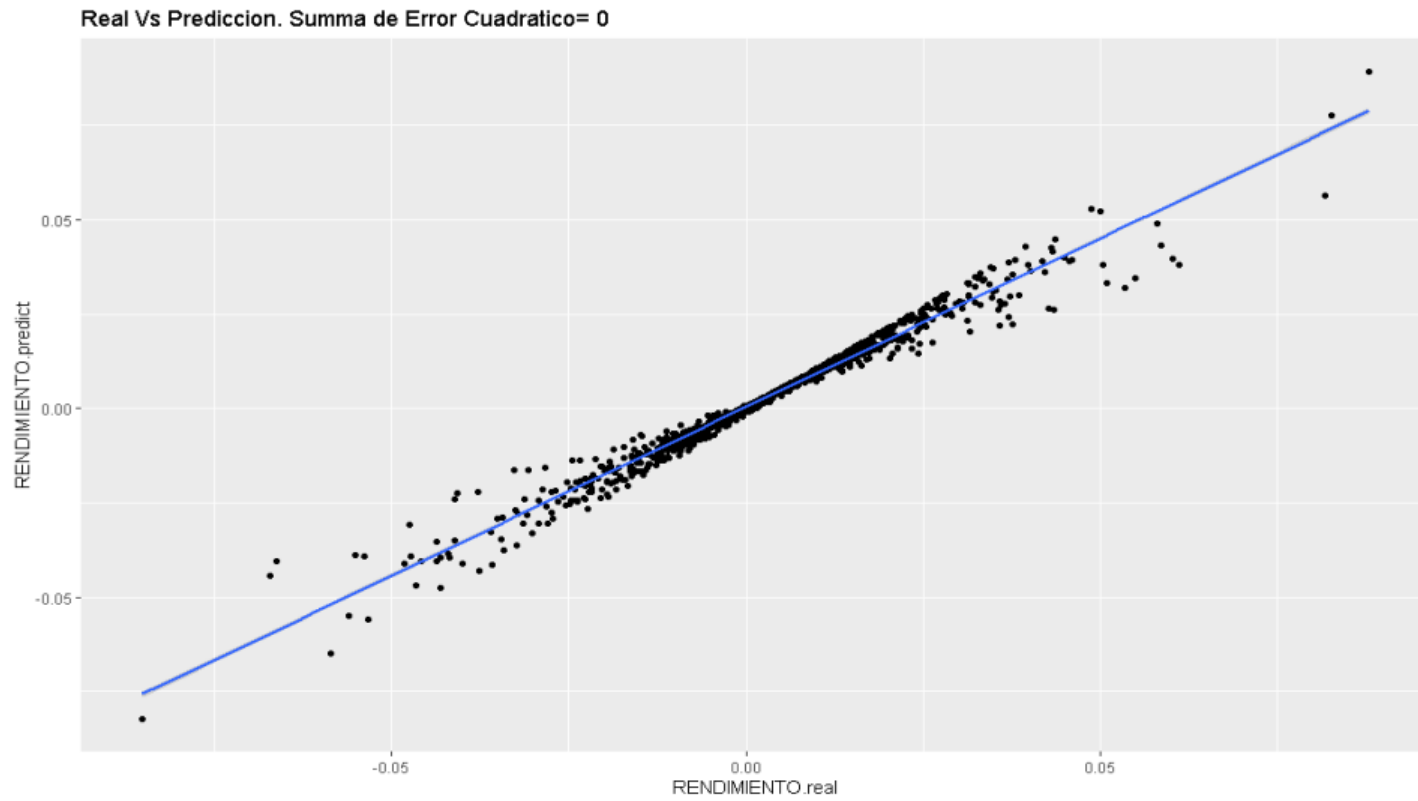
Topología de la RNA para el COLCAP



PREDICTIBILIDAD DE LA RENTABILIDAD

Figura 16.

Diagrama del error para el COLCAP



PREDICTIBILIDAD DE LA RENTABILIDAD

6.6 Conclusiones para RNA

Se observa para la serie histórica del índice bursátil IGBC y COLCAP durante el periodo de estudio (2001 al 2018), un mayor ajuste a los modelos de redes neuronales con un solo rezago.

Así mismo, se observa que para el IGBC la red neuronal con menor error es una Red multinivel hacia adelante (MLF), con el algoritmo de entrenamiento estocástico “*Resilient Backpropagation*” con dos capas ocultas, 1 nodo en la primera capa y 2 nodos en la segunda capa, para la serie de los datos históricos y cuatro rezagos; la suma cuadrática del error en promedio es de 0,031143 con las iteraciones realizadas, sin embargo, mediante el análisis de todas las iteraciones se puede concluir que al entrenar la red con rezagos mayores el MSE aumentó, para un solo rezago utilizando 2 capas ocultas, se puede ver que este, en promedio mantiene el error más bajo.

Se observa que, para el COLCAP la red neuronal con menor error es una Red multinivel hacia adelante (MLF), con el algoritmo de entrenamiento estocástico “*Resilient Backpropagation*” con una capa oculta y diez nodos, para la serie de los datos históricos y un solo rezago; la suma cuadrática del error en promedio para esta red es de 7,7609E-06 en las iteraciones realizadas.

Según la revisión de literatura, en diversos trabajos las redes neuronales presentan frente a modelos tradicionales mayor capacidad de ajuste, dado que son principalmente utilizados para estimar modelos no lineales, como es el caso del rendimiento de los índices estudiados. Mediante las iteraciones realizadas se encontraron buenos ajustes para las series, con errores pequeños. Sin embargo, es importante resaltar que este ajuste depende de una especificación correcta de la topología en aspectos como: número de rezagos, número de nodos de entrada, número de nodos en la capa oculta.

PREDICTIBILIDAD DE LA RENTABILIDAD

En el Apéndice E se presenta finalmente el código en el software estadístico R-studio para la red neuronal. (Apéndice E. Código red neuronal)

Para la métrica de predicción con la metodología de red neuronal se presenta la ecuación utilizada:

$$C(s) = 1 - \frac{MSE_{(s)}}{MSE_{r(s)}}$$

Dando como resultado:

Métrica con la serie IGBC = 0,9799999

La red neuronal tiene la topología: 4 rezagos, 2 capas ocultas, 1 nodo en la primera capa y 2 nodos en la segunda capa.

Métrica con la serie COLCAP= 0,9999999

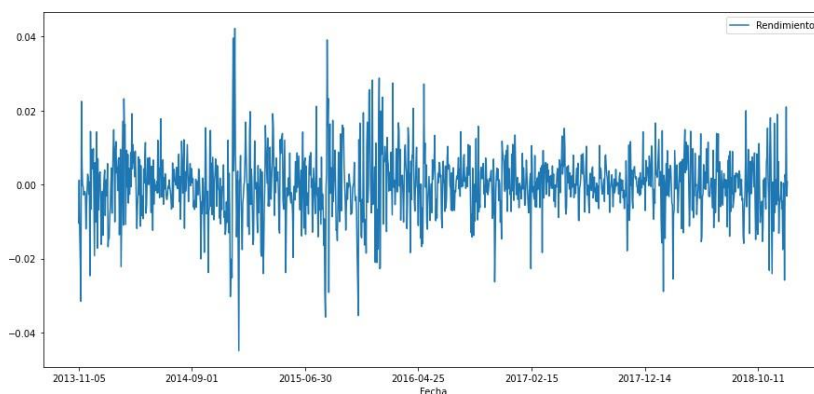
La red neuronal tiene la tipología: 1 rezago, 1 capa oculta, 10 nodos.

6.6.1 Modelo híbrido

Se exploró un modelo híbrido mediante el código que se encuentra en el Apéndice F. Código del modelo híbrido y a continuación se presenta las figura de los datos, los parámetros y predicción del modelo.

Figura 17.

Gráfica de datos para el modelo híbrido



PREDICTIBILIDAD DE LA RENTABILIDAD

Figura 18.

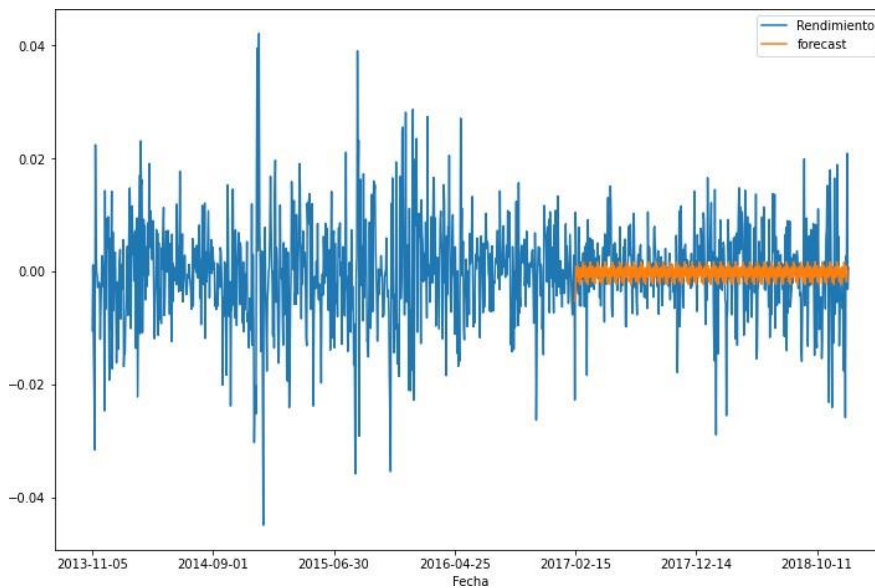
Parámetros

SARIMAX Results

| | | | | | | |
|-------------------------|----------------------------------|-------------------|-----------|-------|----------|----------|
| Dep. Variable: | y | No. Observations: | 451 | | | |
| Model: | SARIMAX(1, 0, 0)x(0, 1, [1], 12) | Log Likelihood | 1541.286 | | | |
| Date: | Tue, 30 Mar 2021 | AIC | -3076.573 | | | |
| Time: | 10:37:51 | BIC | -3064.319 | | | |
| Sample: | 0 | HQIC | -3071.738 | | | |
| | | | - 451 | | | |
| Covariance Type: | opg | | | | | |
| | coef | std err | z | P> z | [0.025 | 0.975] |
| ar.L1 | 0.0943 | 0.044 | 2.162 | 0.031 | 0.009 | 0.180 |
| ma.S.L12 | -0.8834 | 0.030 | -29.467 | 0.000 | -0.942 | -0.825 |
| sigma2 | 5.045e-05 | 2.68e-06 | 18.793 | 0.000 | 4.52e-05 | 5.57e-05 |
| Ljung-Box (Q): | 37.21 | Jarque-Bera (JB): | 40.13 | | | |
| Prob(Q): | 0.60 | Prob(JB): | 0.00 | | | |
| Heteroskedasticity (H): | 2.49 | Skew: | -0.29 | | | |
| Prob(H) (two-sided): | 0.00 | Kurtosis: | 4.36 | | | |

Figura 19.

Predicción modelo ARIMA



PREDICTIBILIDAD DE LA RENTABILIDAD

Figura 20.

Nuevo dataset

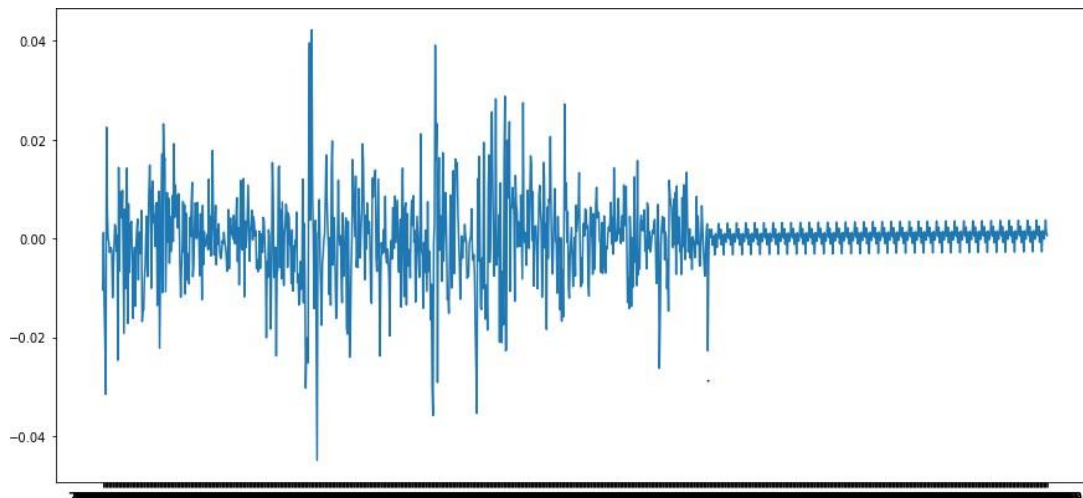
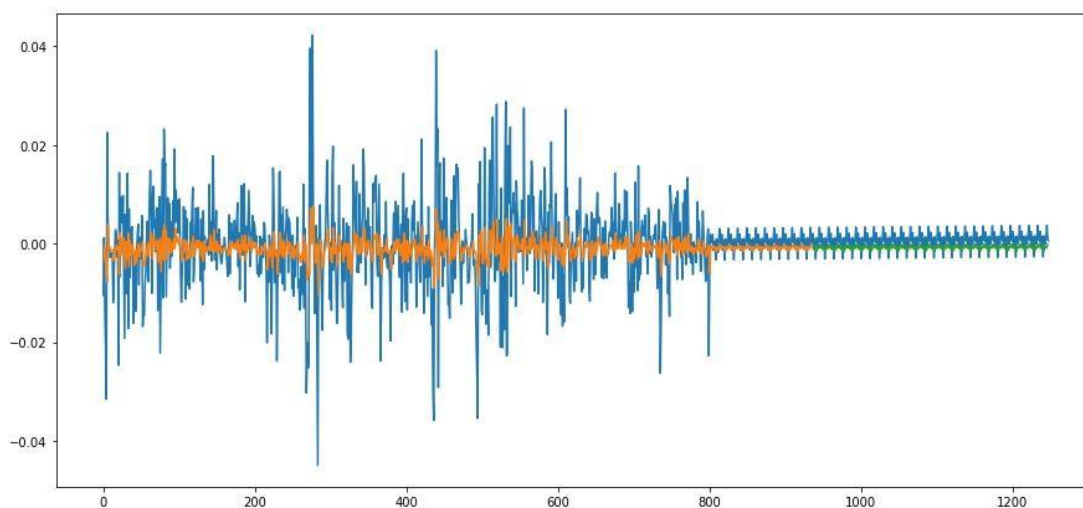


Figura 21.

Modelo final



Mediante el ajuste del modelo ARIMA-red neuronal se pudo encontrar un RMSE 0,02 y un MSE de 0,0004 siendo este un buen indicador, se destaca la utilización de un algoritmo RMSProp (Root Mean Square Propagation), este es una variación del algoritmo AdaGrad el cual introduce una variación muy interesante en el concepto de factor de entrenamiento en vez de considerar un valor uniforme para todos los pesos, se mantiene un factor de entrenamiento

PREDICTIBILIDAD DE LA RENTABILIDAD

específico para cada uno de ellos. Partiendo del factor de entrenamiento inicial se escala y adapta para cada dimensión con respecto al gradiente acumulado en cada iteración. La diferencia radica en que el RMSProp en el caso del escalado el factor de entrenamiento se realiza dividiéndolo por la media del declive exponencial del cuadrado de los gradientes.

Por otra parte, aplicando la métrica para la metodología del modelo híbrido se encontró lo siguiente:

Métrica con la serie IGBC = 0

Métrica con la serie COLCAP= 0.9999995

Se puede concluir de la medición que en el caso del índice IGBC no es posible aplicar el modelo híbrido debido a que no ejerce buenas predicciones, dado que presenta un error elevado con respecto a lo que contiene los datos originales.

En la hipótesis planteada sobre si existe predictibilidad en la serie histórica de la rentabilidad observada mediante el índice IBGC y COLCAP para la ventana de tiempo, se comprueba con la prueba de rachas la posibilidad de ajustar modelos, debido a que las sucesivas observaciones no son independientes. En cuanto a la capacidad predictiva y de ajuste en el proyecto, se pudo concluir que el del modelo combinado (redes neuronales artificiales-ARIMA) para la serie de la rentabilidad observada es estadísticamente mejor comparado con los modelos ARIMA. Sin embargo, en cuanto a las Redes Neuronales Artificiales este presenta el mejor modelo de ajuste dado que la suma cuadrática del error en promedio para el IGBC es de 0,031143 y en el COLCAP el MSE es de 7,7609E-06.

Finalmente, en la Tabla 22 se presenta los datos de la métrica comparando los tres modelos utilizados:

PREDICTIBILIDAD DE LA RENTABILIDAD

Tabla 20.

Comparación métrica de predictibilidad.

| Metodología | Modelo | Métrica con la serie | |
|--------------|-----------------|----------------------|------------------|
| | | IGBC | COLCAP |
| ARIMA | (1,1,1) | 0,8965053 | 0,979972 |
| Red neuronal | (4,2,2)-IGBC | 0,9799999 | 0,9999999 |
| | (1,1,10)-COLCAP | | |
| Híbrido | | - | 0,9999995 |

7. Análisis de resultados

Mediante la revisión de literatura financiera se presenta la formulación de la hipótesis del mercado eficiente, los tres tipos de eficiencia, algunos modelos de pronóstico y métricas de predictibilidad a fin de determinar los modelos de ajuste más representativos a las series de tiempo, en particular de modelo autorregresivo ARIMA (modelo autorregresivo integrado de promedio móvil), redes neuronales y un modelo combinado de ambos para la ventana de tiempo del 2001 al 2018.

Según el problema de investigación que se plantearon dos hipótesis las cuales fueron si existe predictibilidad en la serie histórica de la rentabilidad observada mediante el índice COLCAP para la ventana de tiempo y de existir predictibilidad, la capacidad predictiva y de ajuste del modelo combinado (redes neuronales artificiales-ARIMA) para la serie histórica de la rentabilidad observada es estadísticamente mejor comparado con los modelos ARIMA y Redes Neuronales Artificiales, para ello se identificaron las principales características de la serie histórica del índice bursátil IGBC y COLCAP mediante un análisis descriptivo del índice durante el periodo de estudio (2001 al 2018), del IGBC desde el 3 de julio del 2001 a octubre del 2013 y el COLCAP desde noviembre del 2013 a diciembre del 2018 se observaron periodos con tendencia al alza o a la baja de manera marcada y alta variabilidad, convirtiéndose en una serie no estacionaria.

PREDICTIBILIDAD DE LA RENTABILIDAD

En el análisis de las estadísticas descriptivas se observa que las dos series son asimétricas negativas según los criterios de la distribución, presentan un sesgo, donde se encuentran valores más separados de la media a la izquierda, además la curtosis de las dos series COLCAP y el IGBC indican que son de tipo leptocúrticas, en este caso con mayor apuntamiento en la distribución de frecuencias de la serie del IGBC.

Se realizó el procedimiento de Prueba de Rachas en el índice IGBC del 2001 al 2013 obteniendo un resultado de P-valor inferior al nivel de confianza de 0,05 y 0,01 rechazando la hipótesis nula planteada y determinando que la serie en esa ventana de tiempo es dependiente, ya que el mercado no es eficiente durante el periodo. También, se realizó la prueba con el índice COLCAP desde 2013 al 2018, dando como resultado un P-valor de 0,002 siendo inferior al nivel de confianza lo que determina su dependencia.

Para aceptar o rechazar la segunda hipótesis planteada en el proyecto, se realizaron los ajustes para las series del IGBC y COLCAP durante el periodo de estudio (2001 al 2018), en el cual se presenta un mayor ajuste a los modelos de redes neuronales con un solo rezago. Para el IGBC la red neuronal con menor error es una Red multinivel hacia adelante (MLF), con el algoritmo de entrenamiento estocástico "Resilient Backpropagation" con dos capas ocultas y 1 nodo en la primera capa y 2 nodos en la segunda capa, para la serie de los datos históricos y cuatro rezagos. Para el COLCAP la red neuronal con menor error es una Red multinivel hacia adelante (MLF), con el algoritmo de entrenamiento estocástico "Resilient Backpropagation" con una capa oculta y diez nodos, para la serie de los datos históricos y un solo rezago.

Para el ajuste del modelo híbrido ARIMA-Red neuronal artificial se obtuvo un RMSE 0,02 y un MSE de 0,0004 con el algoritmo RMSProp (Root Mean Square Propagation), para darle respuesta a la hipótesis planteada se comparan los indicadores demostrando que la capacidad

PREDICTIBILIDAD DE LA RENTABILIDAD

predictiva del combinado es estadísticamente mejor comparado con los modelos ARIMA, pero menos preciso que las Redes Neuronales.

En cuanto al ajuste a los modelos tradicionales se encuentra que en la serie histórica del índice bursátil IGBC y COLCAP durante el periodo de estudio (2001 al 2018), se observaron periodos con marcada variabilidad, convirtiéndose en una serie no estacionaria en su varianza. Se evidencia que las series pueden ajustarse a modelos sencillos como MA(1), AR(1), MA(2), AR(2), ARMA(1,1), ARCH(1,1), GARCH(1,1).

PREDICTIBILIDAD DE LA RENTABILIDAD

8. Conclusiones

Según la hipótesis de mercado eficiente se afirma la impredecibilidad de la rentabilidad, dado que en un mercado financiero eficiente se reflejan toda la información existente y se ajustan total y rápidamente a los nuevos datos que puedan surgir, sin embargo, en la revisión de literatura se encuentran otras vertientes que se oponen a esta hipótesis donde han estudiado la relación de la predictibilidad y los rendimientos, encontrando patrones lineales y no lineales en las series de tiempo mediante modelos matemáticos, donde se destaca que los modelos combinados y las redes neuronales han demostrado mejores ajustes a las series de tiempo financieras.

En el presente proyecto se probó la hipótesis de mercado eficiente que dependen de las ventanas móviles seleccionadas, mediante el índice COLCAP del 2001 al 2018 y se ajustaron modelos tradicionales, redes neuronales y un modelo combinado.

En la serie histórica del índice bursátil IGBC y COLCAP durante el periodo de estudio (2001 al 2018), se evidencia que las series pueden ajustarse a modelos sencillos como MA(1), AR(1), MA(2), AR(2), ARMA(1,1), ARCH(1,1), GARCH(1,1), para el caso de las

redes neuronales, se revisó la literatura a fin de identificar algoritmos frecuentes o tipologías de la red que pudieran acortar la búsqueda, sin embargo, para encontrar la estructura adecuada, se hicieron múltiples iteraciones presentando un mayor ajuste a los modelos de redes neuronales (RNA) con un solo rezago. Para el IGBC la RNA con menor error es una red multicapa, con el algoritmo de entrenamiento "Resilient Backpropagation" con dos capas ocultas y un nodo en la primera capa, dos nodos en la segunda capa y cuatro rezagos; y para el COLCAP la topología de la red tiene una capa oculta y diez nodos, con un solo rezago.

PREDICTIBILIDAD DE LA RENTABILIDAD

Para el modelo híbrido ARIMA-Red neuronal artificial se obtuvo un RMSE 0,02 que comparada con el modelo ARIMA se encuentra que es estadísticamente mejor comparado, pero con un error mayor que la Red neuronal con la topología descrita anteriormente.

Para la validación de las hipótesis, se resalta que en la serie si es posible la predicción mediante el análisis técnico con la aplicación de algoritmos, sustentado en la aplicación de la métrica de predictibilidad. En cuanto a la segunda hipótesis, se prueba que las redes neuronales artificiales son más precisas en el pronóstico de rendimientos de los índices IBGC y el COLCAP, siendo estos los índices más relevantes utilizados en Colombia capturando su comportamiento.

PREDICTIBILIDAD DE LA RENTABILIDAD

9. Recomendaciones

Se recomienda para futuros proyectos introducir diferentes métricas de medición de la predictibilidad, como el utilizado en la presente tesis variando los parámetros como los pasos hacia adelante e introducir otras medidas de predicción como los mencionados en la revisión de literatura.

Estudiar las evoluciones de los enfoques estadísticos del desarrollo de la hipótesis de mercado eficiente al mercado adaptativo y la comprensión moderna del mismo.

Desarrollar métodos de ingeniería para predicción de datos futuros en bolsas de valores o modelos híbridos distintos a los elaborados en la presente tesis con el fin de identificar un mayor ajuste de las series

Otras líneas de estudio futuras son investigar cómo las informaciones generadas por distintas noticias macroeconómicas tienen un impacto diferente en los movimientos cambiarios de corto plazo, así como el tiempo que tarda el mercado colombiano en incorporar esta nueva información.

PREDICTIBILIDAD DE LA RENTABILIDAD
Referencias Bibliográficas

Arias, S, González, M y Fuentes, H (2015) “predictibilidad de los análisis técnico y fundamental en mercados latinoamericanos: evidencia empírica y aplicación práctica” Recuperado a partir de:

<http://repositorio.uchile.cl/bitstream/handle/2250/129706/Predictibilidad%20de%20los%20an%C3%A1lisis%20t%C3%A9cnico%20y%20fundamental%20en%20mercados%20latinoam.pdf;sequence=1>

Bachelier, L. (1900). Theory of speculation. *Annales Scientifiques de l’Ecole Normale Supérieure*, III(17),21–86. Recuperado a partir de <http://www.e-m-h.org/Bachelier1900.pdf>

Berlanga, V y Rubio, M, (2012). Clasificación de pruebas no paramétricas. Como aplicarlas en SPSS, ICE Universitat de Barcelona. Recuperado a partir de <http://hdl.handle.net/2445/45283>. <http://dx.doi.org/10.1344/reire2012.5.2528>

Bertona L. F (2005). Entrenamiento De Redes Neuronales Basado En Algoritmos Evolutivos. Laboratorio de Sistemas Inteligentes

Bollerslev, T. (1986). Heteroscedasticidad condicional auto regresiva generalizada. *Diario de la econometría*, 31 (3), 307–327. [http://dx.doi.org/10.1016/0304-4076\(86\)90063-1](http://dx.doi.org/10.1016/0304-4076(86)90063-1)

Brock, W. A., Lakonishok, J., & LeBaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, 47(5), 1731–1764. <https://doi.org/10.2307/2328994>

Charisios Christodoulos, C.hristos Michalakelis, Dimitris Varoutas. Predicción con datos limitados: Combinando ARIMA y modelos de difusión. Aceptado: el 23 de enero de 2010.

PREDICTIBILIDAD DE LA RENTABILIDAD

Chen, TL (Chen, Tai-Liang); Cheng, CH (Cheng, Ching-Hsue); Liu, JW (Liu, Jing-Wei).

Un modelo de serie temporal causal basado en la regresión de perceptrón multicapa para pronosticar el índice bursátil de Taiwán. Revista internacional de tecnología de la información y toma de decisiones. Páginas: 1967-1987. DOI: 10.1142 / S0219622019500421

Economic Review. Vol. 89, Núm. 1, marzo. Office of Communications (2002). Daniel Kahneman wins Nobel Prize y Kahneman gets warm reception after winning Nobel. Princeton University, octubre 9.

Engle, R (1982) “Heteroscedasticidad condicional autoregresiva con estimaciones de la variación de la inflación del Reino Unido” Recuperado a partir de: [http://links.jstor.org/sici?sici=0012-9682%2819820 ... O% 3B2-3 & origin = repec](http://links.jstor.org/sici?sici=0012-9682%2819820...O%3B2-3&origin=repec)

Fama, E. (1965). The behavior of stock-market prices. The Journal of Business, 38(1),34–105.

Fama, E. (1970). Efficient capital markets: A review of Theory and Empirical Work. The Journal Of Finance, 25(2), 383–417.

Fama, E. (1991). Efficient capital markets: II. The Journal of Finance, 46(5),1575–1617.

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996). The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), 27-34.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). Knowledge discovery and data mining: Towards a unifying framework. Knowledge Discovery and Data Mining, 82- 88.

Fernández, A., y Grau, P (2011) “The hypothesis of efficient market, chaos and capital market” La empresa en el dominio de la complejidad / Marcial Jesús López Moreno (dir.), 2011, ISBN 978-84-96889-88-0, págs. 263-313.

PREDICTIBILIDAD DE LA RENTABILIDAD

F.-M. Tseng, H.-C. Yu, G.-H. Tzeng, Combining neural network model with seasonal time series ARIMA model, *Technol. Forecast. Soc. Change* 69 (2002) 71–87

Galarce, A y Marcet Orellana, F (2010) “Mercado cambiario chileno: — Fundamentalismo v/s chartismo” Recuperado a partir de http://repositorio.uchile.cl/tesis/uchile/2010/ec-galarce_ae/html/index-frames.html

García, D. Kristjanpoller, W. (2019) Un enfoque de pronóstico adaptativo para la volatilidad del precio del cobre a través de modelos híbridos y no híbridos. Recuperado de WOS. DOI: 10.1016/j.asoc.2018.10.007

Garland, Joshua & James, Ryan & Bradley, Elizabeth. (2014). Model-free quantification of time-series predictability. *Physical review. E, Statistical, nonlinear, and soft matter physics*. 90. 052910. 10.1103/PhysRevE.90.052910.

Goyal, A. y Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science* 49(5), 639–654.

Hadavandi, E.; Shavandi, H. y Ghanbari, A. (2010). Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting. *Knowledge-Based Systems*, 800-08.

Hamid, S. A. y Iqbal, Z. (2004). Using neural networks for forecasting volatility of S&P 500 Index futures prices. *Journal of Business Research*, 57(10):1116–1125.

Han, Y., Yang, K., & Zhou, G. (2013). A new anomaly: The cross-sectional profitability of technical analysis. *Journal of Financial and Quantitative Analysis*, 48(5), 1433–1461. <https://doi.org/10.1017/S0022109013000586>

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of Financial Economics*, 6(2–3), 95–101. [https://doi.org/10.1016/0304-405X\(78\)90025-9](https://doi.org/10.1016/0304-405X(78)90025-9)

PREDICTIBILIDAD DE LA RENTABILIDAD

Jing-rong, D., Combining stock market volatility forecasts using an EWMA technique, 4th International Conference on Wireless Communications, Networking and Mobile Computing, 5277- 5280 (2007)

Kahneman, D. y A. Tversky (1979). "Prospect Theory: An Analysis of Decision Under Risk", *Econometrica*. Vol. 47, Núm. 2, marzo.

Kay, J. (2011). The map is not the territory: An essay on the state of economics. Recuperado, a partir de <http://ineteconomics.org/sites/inet.civicaactions.net/files/kay-john-state-of-economics-v11.pdf>

Kuvulmaz, J , Usanmaz, S y Engin, SN (2005). Tiempo - serie predicción mediante modelos lineales y no lineales. Volumen: 3789. Páginas: 504-513

Lo, A. W., & MacKinlay, A. C. (2002). A non-random walk down Wall Street. Princeton, N. J.: Princeton University Press.

Lo, A. (2004). The adaptive markets hypothesis: Market efficiency from an evolutionary perspective. *The Journal of Portfolio Management*, 30(5),15–29.

Malkiel, B. (1992). Efficient Market Hypothesis. En *New Palgrave Dictionary of Money and Finance*, Macmillan. London: Newman, P., M.Milgate and J. Eatwell.

Marschinski, R., & Matassini, L. (2001). Financial markets as a complex system: A short time scale perspective (Research notes in economics & statistics No. 01-4).

Recuperado a partir de <http://hdl.handle.net/10419/40278>

Maya C. y Torres, G. (2004) "The Unification of the Colombian Stock Market: A Step Towards Efficiency. Empirical Evidence". En: *Latin American Business Review*. Vol. 5, No.4.

PREDICTIBILIDAD DE LA RENTABILIDAD

Naidu, GM; Reddy, BR; Murthy, BR. Pronósticos De La Serie De Tiempo Utilizando ARIMA Y Enfoques De Red Neural. REVISTA INTERNACIONAL DE CIENCIAS AGRÍCOLAS Y ESTADÍSTICAS. Volumen: 14, No:1, Páginas: 275-278. Junio 2018.

Ospina Holguín, Javier Humberto (2015). Medidas dinámicas de predictibilidad en el índice S&P 500 y sus determinantes. INNOVAR. Revista de Ciencias Administrativas y Sociales, 25(56),69-82.[fecha de Consulta 1 de Mayo de 2021]. ISSN: 0121-5051. Disponible en: <https://www.redalyc.org/articulo.oa?id=81835367006>

Ospina, J (2017) “Predictibilidad de la serie de tiempo y rentabilidad: una aproximación desde la anomalía de la rentabilidad de la sección transversal del análisis técnico” Tesis de Doctorado en Administración, Universidad del Valle.

Perez-Villalobos, J. y Mendoza-Gutiérrez, J. (2010). Efecto día en el mercado accionario colombiano: una aproximación no paramétrica. Borradores de economía Banco de la Republica.

Pradeepkumar, D (Pradeepkumar, Dadabada); Ravi, V (Ravi, Vadlamani) (2017). Pronosticar la volatilidad de las series de tiempo financieras utilizando la optimización de enjambre de partículas Red neuronal capacitada en regresión cuantil. Volumen 58, Páginas: 35-52. Septiembre 2017. DOI: 10.1016 / j.asoc.2017.04.014

Samuelson, P. A. (1965). Proof that properly anticipated prices fluctuate randomly. Industrial Management Review, 6(2), 41–49. Recuperado a partir de http://static.stevereads.com/papers_to_read/proof_that_properly_anticipated_prices_fluctuate_randomly.pdf

Shiller, R. (2003). From Efficient Markets Theory to behavioral finance. Journal of Economic Perspectives, 17(1), 83–104.

PREDICTIBILIDAD DE LA RENTABILIDAD

Subrahmanyam, A. (2010). The cross-section of expected stock returns: What have we learnt from the past twenty-five years of research? *European Financial Management*, 16(1), 27–42. <https://doi.org/10.1111/j.1468-036X.2009.00520.x>

Susruth, M (2017). Previsiones financieras: un estudio empírico sobre Box - Metodología Jenkins con referencia al mercado de valores indio. *RESEÑA DE NEGOCIOS DEL PACÍFICO INTERNACIONAL*. Volumen: 10, No 2, Páginas: 115-123.

Agosto 2017.

Talero, L; Duarte, J y Garces, L. (2017) The Complexity of Latin-American Stock Market using a Behavioral Cellular Automaton Model. vol. 36, n. 64, pp. 199-223. ISSN 0120-3053. <http://dx.doi.org/10.19053/01203053.v36.n64.2017.5421>.

W. Jacobs, A. M. Souza and R. R. Zanini, "Combination of Box-Jenkins and MLP/RNA models for forecasting," in *IEEE Latin America Transactions*, vol. 14, no. 4, pp. 1870-1878, April 2016. doi: 10.1109/TLA.2016.7483528

Zhang G. Peter (2001). Time series forecasting using a hybrid ARIMA and neural network model. Department of Management, J. Mack Robinson College of Business, Georgia State University, University Plaza, Atlanta, GA 30303, Aceptado 23 Noviembre 2001.