



UNIVERSIDAD DEL BÍO-BÍO



MAGISTER EN GESTIÓN DE EMPRESAS

“Modelo predictivo ARIMA, aplicado en las variaciones del precio del cobre y aluminio.”

Programa Ejecutivo

Autor : Paulo Venegas Guajardo.

Profesor Guía : Dr. Luis Améstica Rivas

Profesor Co-Guía : Antonino Parisi Fernández

Chillán, 18 de febrero de 2020.

Resumen

La predicción de la variación de los precios de commodities se ha convertido en el foco de atención de muchos investigadores, inversionistas, entre otros, debido a la oportunidad de generar beneficio económico de esta. Este documento tiene como objetivo, dar a conocer los modelos predictivos que existen en la actualidad junto con sus características, y específicamente desarrollar modelos predictivos ARIMA, optimizado con fuerza bruta operacional para la predicción del signo de la variación del precio del cobre y aluminio. Se utilizaron los precios de cierre semanales de estos commodities durante 9 años (2011-2019), observando las variaciones de los precios y comparando los datos reales con los de la predicción hecha por el modelo. Para cada modelo se utilizaron ocho variables, generando 3.000.000 de iteraciones con fuerza bruta, con el propósito de encontrar el mejor modelo y ayude a mejorar la toma de decisiones para los inversionistas. De acuerdo a los resultados obtenidos se estableció una capacidad predictiva en ambos activos superior al 50%, específicamente para el cobre 56% y para el aluminio 51%.

Palabras Clave: ARIMA, modelos predictivos, cobre, aluminio, fuerza bruta.

Abstract

The prediction of the variation in commodity prices has become focus of the attention of many researchers, investor and others, due to the opportunity to generate economic benefits from this. The present document aims to publicize the predictive models that currently exist with their characteristics, and specifically develop a predictive model ARIMA, optimized with operational brute force to predicting de sign of the price variation of copper and aluminum. It used the weekly closing prices of these commodities for nine years (2011-2019), observing the prices variation and comparing the real data with the prediction made by the model. For each model eight variables were used, generating 3.000.000 of iteration with brute force, with the purpose to find the best model and help investors to improve the decision making. According to the results obtained, a predictive ability was established in both assets higher than 50%, specifically 56% for the case of copper and 51% for the aluminum.

Keywords: ARIMA, predictive models, cooper, aluminum, brute force.

Introducción

En el mundo de los negocios e inversiones en activos, existe una constante incertidumbre, es por eso que siempre los participantes se encuentran en busca de gran cantidad de información para la toma de decisiones. Es debido a lo anterior que ha hecho necesaria la creación y utilización de modelos predictivos que tengan una gran capacidad de procesamiento de datos y sean de utilidad para los tomadores de decisiones a la hora de enfrentarse a la incertidumbre y riesgo que presentan los mercados, además de lograr predecir los cambios en los precios y obtener beneficios económicos de esto.

En consecuencia de lo anterior, es que se han creado distintos modelos predictivos con el pasar del tiempo y que han ido evolucionando. Estudios anteriores y recientes han sido aplicados especialmente en commodities, dentro de estas investigaciones se destacan autores como Torbat, Khashei y Bijari (2018) con su artículo “A hybrid probabilistic fuzzy ARIMA model for consumption forecasting in commodity”, también a Pincheira, & Hardy (2019) con su artículo “Forecasting base metal prices with the Chilean exchange rate”, y a Parisi, Améstica, & Chileno (2019) con su artículo “Predicción de variaciones en el precio del petróleo con el modelo de optimización ARIMA, innovando con fuerza bruta operacional”. Es en esta línea en que se desenvuelve esta investigación, utilizando un modelo de regresión lineal (ARIMA) para la predicción del signo de la variación del precio del cobre y aluminio.

Al utilizar la palabra modelos predictivos, hacemos referencia a una representación de la realidad basada en un intento de relacionar un conjunto de variables con otra. En este aspecto, se utilizan las tecnologías de información, trabajando con grandes cantidades de datos y con el objetivo de apoyar la toma de decisiones en personas interesadas. Estos modelos son útiles para poder, mediante un esfuerzo analítico, revelar oportunidades de inversión o venta, conocer la previsión de ventas o la cuota de mercado, conocer los mercados de destino más rentables e incluso los precios futuros de activos.

Además, se realizará una revisión de los modelos predictivos más importantes. Para desarrollar el modelo se utilizó información disponible en la web para conocer los precios de ambos commodities comprendidos en el periodo 2011-2019, donde se observará las variaciones de precios, de esta manera comparar las variaciones reales con la pronosticadas y poder encontrar el modelo ARIMA que obtiene el mejor porcentaje de predicción del signo.

Revisión de literatura

En la actualidad existen muchos estudios que respaldan con evidencia significativa la predictibilidad de los precios de un título accionario lo que confirma que estos no son aleatorios, y, por ende, sí pueden llegar a ser predecibles en algún grado. Estudios como el de Lo y Mackinlay (1988), evidenciaron que a través de la utilización de datos semanales obtenidos de mercados desarrollados tales como Estados Unidos, Japón y Europa Occidental, entre 1962 y 1985, donde estos autores encontraron una correlación serial positiva entre los retornos semanales, es decir que los precios accionarios no siguen un camino aleatorio y los retornos son predecibles en algún grado. Así mismo, Conrad y Kaul (1988), mediante un proceso autorregresivo de primer orden, encontraron evidencia de poder predecir los retornos en el corto plazo usando datos semanales y para el mismo período estudiado por Lo y MacKinlay (1988). En relación a lo anterior, otros estudios como Bondt & Thaler (1985), Fama & French (1988), Poterba & Summers (1988) y Chopra (1992), encontraron una correlación serial negativa en los retornos de los activos individuales y varios portafolios sobre intervalos de tres a diez años.

Así mismo, la búsqueda de variables que permitan predecir los retornos se remonta a los años veinte, pero si se piensa de forma más general, es intuitivo querer contar con herramientas que permitan disminuir la aleatoriedad en los resultados esperados. Dentro de las herramientas se pueden encontrar diferentes modelos predictivos, como autómatas celulares Neumann y Ulam (1940), algoritmos genéticos Holland (1970), redes neuronales McCulloch y Pitts (1943) y ARIMA Box y Jenkins (1976).

Mientras tanto, los autómatas celulares son un modelo matemático para un sistema dinámico, compuesto por un conjunto de celdas o células que adquieren distintos estados o valores. Estos estados son alterados de un instante a otro en unidades de tiempo discreto, es decir, que se puede cuantificar con valores enteros a intervalos regulares. De esta manera este conjunto de células logra una evolución según una determinada expresión matemática, que es sensible a los estados de las células vecinas, la cual se le conoce como regla de transición local.

Por otro lado, los algoritmos genéticos son métodos adaptativos que pueden usarse para resolver problemas de búsqueda y optimización. Están basados en el proceso genético de los organismos vivos. A lo largo de las generaciones, las poblaciones evolucionan en la naturaleza de acorde con los principios de la selección natural y la supervivencia de los más fuertes, postulados por Darwin (1859). Por imitación de este proceso, los Algoritmos Genéticos son capaces de ir creando soluciones para

problemas del mundo real. La evolución de dichas soluciones hacia valores óptimos del problema depende en buena medida de una adecuada codificación de las mismas.

Igualmente, en la búsqueda de predicciones precisas, se encuentra el modelo de redes neuronales como concepto, puede ser explicada como un poderoso sistema computacional paralelo, capaz de resolver problemas altamente complejos. Está organizada de manera de que su funcionamiento es una representación de cómo funcionan las neuronas reales en el cerebro humano (Ayala, Letelier y Zagal, 2009), donde lo que se busca es resolver problemas mediante el empleo del conocimiento obtenido de casos similares resueltos en el pasado.

Por otra parte, dentro de los modelos de predicción lineal más utilizados en la literatura, se encuentra el Modelo Autorregresivo Integrado de Media Móvil (ARIMA) fue desarrollado por Box y Jenkins (1976), y tienen como objetivo proyectar una variable en función del comportamiento pasado de series de tiempo o, mejor dicho, variables temporales. Este modelo utiliza información más reciente como valor de partida y luego analiza los errores de proyección más reciente para así ajustarlo apropiadamente.

De manera análoga, existen diferentes estudios que han utilizado el modelo ARIMA en sus predicciones como lo es la “Estimación del precio internacional del arroz *Oryza sativa* L. bajo el modelo ARIMA” (Ceballos y Pire 2015) el cual plantea una manera confiable para pronosticar el precio de la tonelada de arroz con el fin de planificar racionalmente la toma de decisiones basada en estos pronósticos, dado el estudio el precio de la tonelada del arroz se ubicó en US \$665 bajo condiciones normales del mercado, en caso de ocurrir cambios en las políticas agrarias de los países productores y/o se presenten eventos o fenómenos naturales que afecten a la producción, el precio de la tonelada sufriría variaciones.

En el contexto de series de tiempo, pronosticar significa predecir valores futuros de una variable aleatoria basándose en el estudio de la estructura definida por las observaciones pasadas de variables que explican su variación, suponiendo que la estructura del pasado se conserva en el futuro (Pajouh y Danech, 2002). En consecuencia, se supone que las regularidades que puedan contener las series históricas de variaciones de precios e índices bursátiles son percibidas por los modelos de series de tiempo, de redes neuronales y de algoritmos genéticos. No obstante, el grado de predictibilidad de los rendimientos es por lo general considerado económicamente no significativo, por lo que Hodgson y

Nicholls (1991) sugieren evaluar la significancia económica de predecir la dirección de los cambios en los precios de los activos.

Los modelos ARIMA, en algunos casos se hacen llamar híbridos, debido a que se pueden optimizar a través de diferentes métodos para mejorar la predicción del signo, es así tal que en estudios como el de Aamir, Shabari y Ishaq (2018), fue utilizado un modelo ARIMA híbrido basado en la descomposición de modo empírico completo con ruido adaptativo (CEEMDAN) para lograr predecir los precios mundiales del petróleo crudo, además, estos antes de directamente predecir utilizan un método para filtrar los datos y así obtener una mayor precisión, el filtro que se utilizó en esta investigación es el Filtro Kalman. También otro método que se encuentra es el de fuerza bruta operacional, esta técnica hace referencia a lo expuesto por Parisi (2015), quien plantea que se utiliza la capacidad de los computadores para así poder encontrar la solución óptima. Esta técnica aplicada a los modelos ARIMA simula la inteligencia humana, debido a que genera escenarios distintos donde cada uno de ellos entrega una solución única al problema. La función del modelo ARIMA optimizado con fuerza bruta, es contrastar los escenarios generados con los anteriores y seleccionar el mejor.

De igual manera, Parisi, Améstica y Chileno (2019) en su investigación evalúan la eficacia del modelo ARIMA multivariable optimizado con fuerza bruta para el caso del precio del petróleo y empresas petroleras, prediciendo el comportamiento de las acciones a la semana siguiente de una última fecha analizada, a través de la técnica de fuerza bruta se pudo establecer una capacidad de predicción superior al 60% para el caso del precio del petróleo y las acciones de empresas petroleras.

Por otro lado, Leung, Daouk y Chen (2000) , realizan un estudio comparando la capacidad predictiva de los modelos de clasificación (linear discriminant analysis, logit model, probit model y probabilistic neural network) con los de estimación de nivel (adaptative exponential smoothing, vector autoregression with Kalman filter, multivariate transfer function y multilayered feedforward neural network), concluyendo que los primeros se desempeñan mejor a la hora de utilizarlos para predecir el signo o la dirección de retornos de los índices bursátiles. Lo anteriormente planteado sobre centrar la atención en la predicción es muy relevante para los traders (persona que compra y vende activos financieros en cualquier mercado financiero), ya que apunta a que estos deben centrar sus esfuerzos en predecir con precisión la dirección de los movimientos en vez de minimizar la desviación estándar de las estimaciones de los valores observados, debido a que estas estrategias pueden generar beneficios más altos (Wu y Zhang, 1997), esto también se encuentra apoyado según lo expuesto por

O' Connor, Remus y Griggs (1997) quienes afirman que la utilidad será mayor al proyectar la dirección del cambio en el nivel de precios, en vez que el nivel de precios en sí.

Así mismo, Leung, Dauk y Chen a mediados del siglo XXI concluyen que la predicción de los retornos accionarios es un hecho real, haciendo que la pregunta ya no sea si se puede o no predecir los retornos, si no que como se debe usar la información de forma óptima para poder predecirlos.

Datos y metodología

Modelo ARIMA

ARIMA (p, d, q)

P = AR () autorregresivo como variable explicativa.

D= Integrado

Q= Error como variable explicativa (media móvil de los errores).

Por otra parte, la metodología de Box y Jenkins supone que no se conocen o que no están disponibles aquellas variables que determinan el valor proyectado, siendo en esta situación donde los Modelos ARIMA tienen una gran ventaja sobre los modelos de regresión típicos, y a principios de los 80 fueron ampliamente utilizados para modelar la media de un activo (Alonso & García, 2009) y según (Pai & Lin, 2004) el modelo ARIMA, ha sido uno de los enfoques más utilizados en predicción.

Proceso Autorregresivo (AR)

Sea Y_t el Precio de la acción en el período t . Si se modela Y_t como

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + u_t \quad (1)$$

Donde δ es la media de Y y donde u_t es un término de error aleatorio no correlacionado con media cero y varianza constante σ^2 (ruido blanco), entonces se dice que Y_t sigue un proceso estocástico autorregresivo de primer orden, o AR(1). Aquí el valor de Y en el tipo de t depende de su valor en el período de tiempo anterior y de un término aleatorio; los valores de Y están expresados como desviaciones de su valor medio. En otras palabras, este modelo dice que el valor de pronóstico de Y en el período t es simplemente alguna proporción ($= \alpha_1$) de su valor en el período $(t - 1)$ más una

innovación aleatorio o perturbación en el tiempo t ; nuevamente los valores de Y están expresados alrededor del valor de su media.

Pero, si se considera este modelo

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + u_t \quad (2)$$

Entonces, se dice que Y_t sigue un proceso autorregresivo de segundo orden, o AR(2). Es decir, el valor de Y en el tiempo t depende de sus valores en los dos períodos anteriores, los valores de Y expresados alrededor del valor de su media.

En general, se tiene

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \dots + \alpha_p(Y_{t-p} - \delta)u_t \quad (3)$$

In which case, Y_t is an autoregressive process of order p , or AR (p).

Note that in all the previous models only the current and previous values of Y ; are being considered; there are no other regressors.

Moving average process (MA)

The recently exposed AR process is not the only mechanism that Y may have generated. Suppose Y is modeled as follows:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} \quad (4)$$

Where μ is a constant and u , as before, is the stochastic error term white noise. Here, Y in the period equals a constant plus a moving average of the present and past error terms. Thus, in the present case, Y is said to follow a first-order mobile media process, or MA (1).

But if Y follows the expression

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} \quad (5)$$

So it is a MA (2) process. More generally,

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} \quad (6)$$

Es un proceso MA(q). En resumen, un proceso de medio móvil es sencillamente una combinación lineal de términos de error ruido blanco.

Proceso autorregresivo y de media móvil (ARMA)

Por supuesto, es muy probable que Y tenga características de AR y de MA a la vez y, por consiguiente, sea ARMA. Así, Y_t sigue un proceso ARMA si éste puede describirse como:

$$Y_t = \theta + \alpha_1 Y_{t-1} + \beta_0 u_t + \beta_1 u_{t-1} \quad (7)$$

Porque hay un término autorregresivo y uno de media móvil.

En general, en un proceso ARMA (p, q) , habrá p términos autorregresivos y q términos de medias móviles.

Técnica Fuerza bruta

La técnica fuerza bruta (Parisi, 2015; Parisi, Améstica-Rivas y Chileno, 2019), utiliza la capacidad de las computadoras para poder encontrar la mejor solución a un problema de optimización. Esta técnica aplicada a los modelos multivariados dinámicos simula la inteligencia humana, puesto que genera diferentes escenarios en los cuales cada uno de ellos brinda una solución única al problema. La función de este modelo multivariado dinámico con fuerza bruta es comparar los nuevos escenarios generados con los anteriores y elegir el mejor. Dicho de otra manera, la técnica de la fuerza bruta recuerda al igual que un humano, el propio comportamiento para ofrecer una mejor solución a determinado problema; si ese comportamiento solucionó el problema, cada vez que suceda un escenario parecido se utilizará el mismo comportamiento. De la misma manera, los modelos multivariados dinámicos con fuerza bruta utilizan el mejor modelo.

Como afirma Durán (2006), la fuerza bruta consiste en enlistar todos los casos y para cada uno calcular la solución, identificando de este modo el caso que ofrezca la mejor solución. Asimismo, Riveros (2015), en un estudio para encontrar la solución óptima al problema del camino más corto para una empresa de logística, comenta que la solución más directa es con fuerza bruta, es decir evaluar todas las posibles combinaciones (de recorridos) y quedarse con el trayecto que utiliza una menor distancia.

En el caso de los modelos ARIMA, que son modelos de regresión, la aplicación de la técnica fuerza bruta permite generar infinitos coeficientes de un universo establecido, para darle un peso a cada variable establecida y evaluada en el modelo. Incluso se puede afirmar que usando fuerza bruta se puede contemplar todos los escenarios del universo establecido, siendo así, una mejora a los algoritmos genéticos, los cuales sólo buscan alrededor de un punto en el universo que ofrece una solución de primera instancia óptima (Parisi; Améstica-Rivas y Chileno, 2019).

Los métodos mencionados anteriormente sirven para resolver los mismos problemas de optimización simulando la inteligencia humana. Muchos científicos y expertos en ciencias sociales a lo largo de los años han estado tratando de desarrollar cada vez mejores métodos para la solución de problemas. Lo curioso es que el primer método empleado para resolver problemas es el conocido como “fuerza bruta”. Desde tiempos antiguos el hombre utiliza la fuerza bruta para resolver los problemas. Pero ¿por qué se crearon nuevos métodos si con este se podía? Sencillo, los problemas cada vez alcanzaron dimensiones mayores y complejas, lo cual hacía prominente crear métodos que demoraran menos en resolver un problema.

Como se ha dicho, la fuerza bruta prueba una a una las diferentes condiciones y características de un sistema para resolver el problema. Una vez que se encuentra la solución, se queda con dicho sistema. Se dejó de usar fuerza bruta puesto que no existía la capacidad para resolver problemas que requerían la evaluación de una cantidad de variables considerada demasiado grande. Sin embargo, hoy en día la tecnología ha vuelto a superar los problemas, y cualquier persona puede contar con un computador con recursos altamente efectivos capaces de procesar información mucho más rápido que aquellas computadoras de hace 10 años atrás. Por lo tanto, ahora es prudente e incluso más eficaz volver a utilizar la técnica fuerza bruta utilizando la capacidad de una computadora de alta tecnología (Parisi; Améstica-Rivas y Chileno, 2019).

Tipo de investigación y diseño

El método de investigación que se utilizó es el cuantitativo el cual se basa en la recolección de información a través de diferentes bases de datos, en este estudio se recopilan los precios históricos del cobre y aluminio que serán utilizados en el modelo ARIMA. Así mismo tiene un carácter exploratorio ya que otorga una visión general, de tipo aproximativo, que se realiza especialmente cuando el tema elegido ha sido poco explorado como lo es la construcción de un modelo optimizado con fuerza bruta, para el cual fue necesario la elaboración de una macro, en Microsoft Visual Basic, para obtener la predicción de los precios, y así calcular el porcentaje de predicción de signo en el precio del cobre y aluminio. Por otra parte, se utilizó una técnica de recolección basada en el registro del precio del cobre y aluminio con un periodo desde el año 2011 al 2019, por lo tanto, la investigación es de un corte longitudinal ya que se realiza en un tiempo determinado, es a través de la observación de estos datos se buscó analizar el comportamiento del metal en el mercado.

Diseño de la investigación

Se construyó un modelo ARIMA con fuerza bruta, para obtener el porcentaje de predicción de signo en el precio del cobre, utilizando 3.000.000 de iteraciones en el modelo.

Población y muestra

La población para esta investigación son los precios históricos del cobre y aluminio publicados en la Comisión Chilena del Cobre (Cochilco), el cual es un organismo técnico y especializado del Estado, que actúa como asesor y depende del Ministerio de minería.

Recolección de datos

Como se ha mencionado los precios históricos del cobre y aluminio fueron obtenidos de la página de internet Cochilco. En la plataforma se encuentran las cotizaciones históricas de los commodities que se tranzan a nivel mundial, así como también sus precios de cierres. Los datos pueden ser recolectados con frecuencia diaria, mensual y anual. Para este estudio se utilizó una frecuencia semanal de los precios de cierre del cobre y aluminio, que comprende entre el 5 de septiembre del 2011 hasta el 19 de agosto de 2019, con un total de 416 observaciones, estas se obtuvieron desde http://boletin.cochilco.cl/estadisticas/grafico.asp?tipo_metal=1

Análisis estadístico

Se construyó un modelo ARIMA optimizado con fuerza bruta computacional para poder evaluar su capacidad predictiva para frecuencias semanales. De esta manera se buscó determinar el comportamiento de este modelo en su función predictiva para conocer la variación del signo del precio del cobre y el aluminio para el valor en estudio.

El planteamiento de la técnica fuerza bruta de Parisi (2015) y utilizado en Parisi, Améstica y Chileno (2019), tiene como objetivo encontrar los coeficientes óptimos que maximicen el porcentaje de predicción de signo de las variaciones semanales de las cotizaciones para el valor en estudio.

El modelo autorregresivo integrado de promedio móvil de orden p, d, q - ARIMA (p, d, q): Al igual que un modelo ARMA, corresponde a la combinación de los modelos autorregresivo y al de medias móviles, con la característica de incluir un proceso de restablecimiento (el cual se designa como integración) de inestabilidad original presente en una serie de tiempo.

La forma general de un modelo ARIMA es semejante al de un modelo ARMA:

$$Y'_t = \alpha_1 Y'_{t-1} + \alpha_2 Y'_{t-2} + \alpha_3 Y'_{t-3} + \dots + \alpha_p Y'_{t-p} + \beta_1 \cdot \varepsilon_{t-1} + \beta_2 \cdot \varepsilon_{t-2} + \beta_3 \cdot \varepsilon_{t-3} + \beta_p \cdot \varepsilon_{t-p} + \varepsilon_t \quad (8)$$

Donde:

Y'_t : Es la serie inducida a la estabilidad.

Evaluación de la predicción

La siguiente etapa consistió en evaluar la calidad de cada modelo en función a su porcentaje de predicción de signo alcanzado (PPS). La evaluación se realizará sobre la base de un conjunto extramuestral de 416 datos semanales para cada modelo. Se utilizará la muestra total para estimar los coeficientes α , β y θ de cada modelo respectivamente, por medio de la minimización de la suma del cuadrado de los residuos del modelo, como para evaluar la capacidad predictiva de los modelos.

Además, para realizar esto se comparó el signo de la proyección con el signo de la variación observada en cada i -ésimo período, en el que $i = 1, 2, \dots, m$. Si los signos entre la proyección y el observado coinciden, entonces aumentan la efectividad del modelo analizado y, en caso contrario, disminuye su capacidad predictiva.

Una vez proyectado el signo de la variación del precio para el período $n+1$, la variación observada correspondiente se incluye en la muestra de tamaño de n con objeto de reestimar los coeficientes del modelo, contando ahora con una observación más. Así, el mismo modelo, pero con sus coeficientes nuevamente calculados, este se utilizará para realizar la proyección correspondiente al período $n+2$. Este procedimiento recursivo se efectuará una y otra vez hasta concluir con las observaciones del conjunto extramuestral. Finalmente, el PPS de cada modelo se calcula de la siguiente forma:

$$\text{Porcentaje} = \frac{\sum_{t=n+1}^{n+m} p_i}{m} \quad (9)$$

$$p_i = \left\{ \begin{array}{l} 1, \\ 0, \end{array} \right. \text{ si } \left. \begin{array}{l} \Delta \text{Precio}_i * \hat{\Delta} \text{Precio}_i > 0 \\ \text{en otro caso} \end{array} \right\}, i = [n + 1, \dots, n + m] \quad (10)$$

En la que Δ en la ecuación (10), representa la variación observada, $\hat{\Delta}$ la variación estimada, $n = 0$ y $m = \dots, p_i$ se refiere al precio observado en el período i que es representado por t que también se refiere al período o tiempo en el que se observa el precio. De esta manera, los modelos multivariados dinámicos construidos optimizados con técnica fuerza bruta serán evaluados en función de su capacidad para predecir el signo de los movimientos del precio del valor en estudio.

Evaluación estadística

En esta etapa se aplicó la prueba de acierto direccional de Pesaran y Timmermann (1992), que proponen Parisi y Parisi (2010), en su libro Teoría de Inversiones, para medir la capacidad de predicción de los modelos de ARIMA, el cual permite medir la significancia estadística de los modelos y su capacidad de predicción.

El test de Pesaran y Timmermann testea las siguientes hipótesis:

H0: la proyección basada en el modelo y el valor observado son independientes

H1: la proyección basada en el modelo y el valor observado no son independientes

La fórmula para el cálculo del test DA de Pesaran y Timmermann, es la siguiente:

$$DA = \frac{(SR - SRI)}{\sqrt{Var(SR) - Var(SRI)}} \quad (11)$$

Donde:

$$SR = PPS = \frac{\text{Cantidad de signos acertados}}{\text{Número de predicciones}} = \frac{\sum_{i=1}^n I_i(Y * \hat{Y} > 0)}{m} \quad (12)$$

Luego:

$$SRI = P * \hat{P} + (1 - P) * (1 - \hat{P}) \quad (13)$$

Donde:

$$P = \frac{\text{Signos } > 0 \text{ Observados}}{\text{Datos totales}} = \frac{\sum_{i=1}^n I_i(Y > 0)}{m} \quad (14)$$

$$\hat{P} = \frac{\text{Signos } > 0 \text{ proyectados}}{\text{Datos totales}} = \frac{\sum_{i=1}^n I_i(\hat{Y} > 0)}{m} \quad (15)$$

Finalmente:

$$\mathit{Var}(SRI) = \frac{[m(2\hat{P}-1)^2 * P(1-P) + m(2P-1)^2 * \hat{P}(1-\hat{P}) + 4P\hat{P}(1-P)*(1-\hat{P})]}{m^2} \quad (16)$$

$$\mathit{Var}(SR) = \frac{SRI*(1-SRI)}{m} \quad (17)$$

El Test DA tiene como supuesto que los valores obtenidos de la aplicación del test se distribuyen según una función normal, por lo tanto, los puntos críticos con un 95% de confianza son [-1,96; +1,96]. Este intervalo representa la zona de no rechazo de la hipótesis nula “H0”; si se rechaza la hipótesis nula, entonces se probará que los resultados obtenidos no son producto del azar.

Resultados

El análisis de la serie de valores de precios de cierre del aluminio y el cobre que se pueden observar en la tabla 1, en el caso del precio del aluminio entregó como resultado un coeficiente de asimetría positivo, esto quiere decir que es una distribución asimétrica donde existe mayor concentración de valores a la derecha de la media que a su izquierda, esto quiere decir que hay más valores separados de la media a la derecha. Por otra parte, para los valores de cierre del cobre se observó un coeficiente de asimetría negativo, esto significa que la distribución presenta valores que tienden agruparse hacia la izquierda de la curva. También, otro estadístico a destacar es el valor de la curtosis, el cual resultó ser negativo para ambos commodities, esto quiere decir que la distribución de los valores observados adoptó una forma platicúrtica, dicho de otra manera, hay una menor concentración de datos en torno a la media, esto quiere decir que la distribución toma una forma más achatada.

Tabla 1: *Estadísticos descriptivos*

Estadístico	Aluminio	Cobre
Media	84,895	293,769
Error estandar de la media	0,45701	2,30855
Varianza	86,885	2217,035
Desviación estándar	9,32121	47,0854
Asimetría	0,045	-0,097
Curtosis	-0,479	-0,684
Mediana	84,45	297,045
Moda	82,89	307,67
Mínimo	64,59	197,77
Máximo	111,22	405,28
Rango	46,63	207,51
Cuenta	416	416
Suma	35316,34	122207,91

Fuente: elaboración propia

A continuación, se presenta el mejor modelo multivariado dinámico para cada caso de estudio de acuerdo con el PPS:

$$\Delta Cobre_t = \alpha_1 \cdot \Delta Cobre_{t-1} + \alpha_2 \cdot \Delta Cobre_{t-2} + \alpha_3 \cdot \Delta Cobre_{t-3} + \alpha_4 \cdot \Delta Cobre + \beta_1 \cdot \varepsilon_{t-1} + \beta_2 \cdot \varepsilon_{t-2} + \beta_3 \cdot \varepsilon_{t-3} + \beta_4 \cdot \varepsilon_{t-4} + \varepsilon_t \quad (18)$$

$$\Delta\text{Aluminio}_t = \alpha_1 \cdot \Delta\text{Aluminio}_{t-1} + \alpha_2 \cdot \Delta\text{Aluminio}_{t-2} + \alpha_3 \cdot \Delta\text{Aluminio}_{t-3} + \alpha_4 \cdot \Delta\text{Aluminio}_t + \beta_1 \cdot \varepsilon_{t-1} + \beta_2 \cdot \varepsilon_{t-2} + \beta_3 \cdot \varepsilon_{t-3} + \beta_4 \cdot \varepsilon_{t-4} + \varepsilon_t \quad (19)$$

En la tabla 2 se muestran los coeficientes α_i y β_i de los mejores modelos, en términos del porcentaje de predicción de signo (PPS) que se obtuvieron con los modelos, que maximicen para cada caso el PPS, estos datos se obtuvieron luego de haber corrido el modelo utilizando fuerza bruta, alcanzando 3 millones de iteraciones, logrando con esto los mejores coeficientes para maximizar el PPS.

Tabla 2: *Resumen de parámetros*

Resumen de los parámetros estimados para el Aluminio.		Resumen de los parámetros estimados para el cobre.	
Variables	Coficiente	Variables	Coficiente
$\Delta\text{Aluminio}_{t-1}$	0,320754759	ΔCobre_{t-1}	-0,30233611
$\Delta\text{Aluminio}_{t-2}$	0,018146841	ΔCobre_{t-2}	0,211214778
$\Delta\text{Aluminio}_{t-3}$	0,083322945	ΔCobre_{t-3}	-0,55516873
$\Delta\text{Aluminio}_{t-4}$	0,879231098	ΔCobre_{t-4}	-0,651335
$\Delta\varepsilon_{t-1}$	-0,077513919	$\Delta\varepsilon_{t-1}$	0,233184483
$\Delta\varepsilon_{t-2}$	0,224445389	$\Delta\varepsilon_{t-2}$	0,493187011
$\Delta\varepsilon_{t-3}$	-0,565987222	$\Delta\varepsilon_{t-3}$	-0,83134293
$\Delta\varepsilon_{t-4}$	-0,792090942	$\Delta\varepsilon_{t-4}$	-0,71514648

Fuente: elaboración propia

En cuanto a la Tabla 3 se exponen los mejores modelos producidos por ARIMA optimizado con fuerza bruta para ambos casos, para el aluminio y el cobre, de los cuales se obtuvo el porcentaje de predicción de signo para cada uno de los casos. Esta capacidad predictiva fue estimada en un conjunto extramuestral de 416 datos semanales, para el caso del aluminio no resultó estadísticamente significativa en cada uno de los valores, de acuerdo con la prueba de acierto direccional (DA), rechazándose así la hipótesis de que si existe capacidad predictiva en los modelos ARIMA con fuerza bruta, por el contrario en el caso del cobre, según el resultado que entregó la prueba de acierto direccional se concluye que esta capacidad predictiva es estadísticamente significativa en cada uno de sus valores, comprobándose con esto la hipótesis de que si existe capacidad predictiva del modelo optimizado con fuerza bruta operacional aplicado al cobre.

Tabla 3: *Resumen de resultados*

Modelo	PPS	Prueba DA^a
Aluminio	51%	0,36359
Cobre	56%	2,33320

^aEl valor de z crítico es de 1.96 y 1.64, para una significación de 5% y 10%, respectivamente.

*Significativo al 5% y al 10%.

Fuente: elaboración propia

Finalmente, el valor del test DA se comparó con el valor crítico de una tabla de distribución normal (z^* , el cual es del 1,96, para un nivel de significancia de 5%). Debido a que el valor absoluto del test DA es menos que Z^* para el caso del aluminio, no se rechazó la hipótesis nula que plantea que la proyección basada en el modelo ARIMA optimizado con fuerza bruta aplicado en el aluminio y el valor observado 0,17 son independientes y, en consecuencia, se concluyó que el modelo no tiene capacidad para predecir el signo de las variaciones del precio del cobre observado.

Por lo expuesto en la Tabla 6 y lo mencionado anteriormente se pudo confirmar que el modelo ARIMA construido con fuerza bruta obtuvo un porcentaje de predicción de signo menor al 60% por lo tanto no se logró rechazar la hipótesis nula, confirmando con esto que el modelo no tiene una capacidad predictiva para el caso del aluminio.

Mientras tanto, se puede afirmar que el modelo ARIMA optimizado con fuerza bruta operacional alcanzó un PPS del 56% y una significancia estadística que permitió rechazar la hipótesis nula, debido a que el valor absoluto del test DA es mayor que Z^* , se rechazó la hipótesis nula y, en consecuencia, se concluyó que el modelo tiene capacidad para predecir el signo de las variaciones del precio del cobre.

Conclusión

La presente investigación determinó que es factible construir un modelo ARIMA optimizado con fuerza bruta operacional, con una capacidad de predicción del 51% para la variación del precio del aluminio, el cual no presenta significancia estadística según el resultado obtenido en el test DA. Por otro lado, este modelo presenta una capacidad predictiva del 56% para la variación del precio del cobre siendo este un mejor resultado que el del aluminio, que a su vez presenta significancia estadística de acuerdo el resultado del test DA.

Además, como se dijo anteriormente los resultados de la prueba de acierto direccional de Pesaran y Timmermann (1992), el cual es “útil cuando se está evaluando la capacidad predictiva de un modelo de signo” (Parisi, F., Parisi A., 2010, p.533), indicaron que el modelo ARIMA, para el caso del aluminio no presentó una capacidad predictiva estadísticamente significativa. Pero, este test si presentó para el cobre una capacidad predictiva estadísticamente significativa, dado los resultados obtenidos.

El modelo ARIMA aplicado al precio de estos commodities, contaba con la misma cantidad de datos analizados e iteraciones de fuerza bruta. Ahora bien, los dos resultados no presentan una gran diferencia, lo que permitió, por el lado del cobre entregar información valiosa para la toma de decisiones de traders, inversionistas, entre otras personas pertenecientes a este mercado, no así en el caso del aluminio.

Para finalizar, la investigación entregó bases de que los modelos ARIMA optimizados con fuerza bruta operacional pueden ser utilizados para la investigación de predicciones de series de tiempos, basadas en la variación de precios. Y de acuerdo a los resultados obtenidos, queda abierta la línea de investigación para que se sigan desarrollando estudios para la PPS del aluminio especialmente e intentar seguir mejorando la del cobre, así llegar a construir modelos predictivos con una capacidad predicción superior al 60% y estadísticamente significativa.

Referencias

Aamir, M., Shabari A., Ishaq M. (2018). Crude oil price forecasting by CEEMDAN based hybrid model of ARIMA and Kalman Filter. *Jurnal Teknologi (Sciences & Engineering, Vol 80 (4)*, 67-79.

Alonso, J., & García, J. (2009). ¿Qué tan buenos son los patrones del IGBC para predecir su comportamiento? Una aplicación con datos de alta frecuencia. *Estudios Gerenciales*, 13-36.

Ayala, L., Letelier, S. Y Zagal, P. (2009). *Modelo de Redes Neuronales para la Predicción de la Variación del Valor de la Acción de First Solar*. (Seminario de título Ingeniero comercial). Facultad de economía y negocios, Universidad de Chile, Santiago

Bondt, W. & Thaler, R (1985). Does The Stock Market Overreact? *The Journal of finance. Vol. 40*, No. 3, Papers and proceedings of the forty-third annual meeting American finance association, Dallas, Texas, (Jul., 1985), 793-805.

Box, G., & Jenkins, G. (1976). *Time series analysis: Forecasting and Control*. San Francisco: Holden Day.

Ceballos, S. y Pire, R. (2015). Estimación del precio internacional del arroz (*Oryza sativa* L.) bajo el modelo ARIMA. *Revista Mexicana de ciencias agrícolas*, (11), 2083-2089.

Chopra, N., Lakonishok, J., & Ritter, J. R. (1992). Measuring abnormal performance: do stocks overreact? *Journal of financial economics*, 235- 268.

Conrad, J., Kaul, G. (1988). Time-variation in expected returns. *The journal of business*, Vol 61 (4), 409-425.

Duran, G. (2016). Introducción a los algoritmos genéticos. Obtenido de: http://bioinformatica.uab.cat/base/documents/genetica_gen/portfolio/Introducci%C3%B3n%20a%20los%20algoritmos%20gen%C3%A9ticos%202016_5_11P20_38_44.pdf

Fama, E. & French K. R. (1988). Permanent and temporary components of stock prices. *Journal of political economy*, Vol 96(2), 247-273.

Hodgson, A., & Nicholls, D. (1991). The impact of index futures markets on australian sharemarket volatility. *Journal of business finance & accounting*, Vol 18 (2), 267-280.

Holland, H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. Oxford, Inglaterra: The University of Michigan Press.

Leung, Mark T., Hazem Daouk y An- Sing Chen (2000), Forecasting stock indices: a comparison of classification and level estimation models. *International journal of forecasting*, Vol 16, 173-190.

Lo, A., & Mackinlay, A. (1988). Stock market prices do not follow random walks: evidence from a simple specification test. *The review of financial studies*, Vol 1 (1), 41-66.

McCulloch, W. S., and Pitts, W. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity. *The Bulletin of Mathematical Biophysics*, Vol 5, 115-133.

O' Connor, M., Remus y K.Griggs (1997), Going up-going down: how good are people at forecasting trends and changes in trends? *Journal of forecasting*, Vol 16(3), 165-176.

Pai, P., & Lin, C. (2004). A hybridarima and support vector machines model in stock price forecasting. *Omega*, 497-505.

Pajouh, Danech. Methodology for traffic forecasting. The French National Institute for Transport and Safety Research (INRETS). Arcuel. 2002.

Parisi, A. (2015). *Modelo predictivo accionario adoptando fuerza bruta. Documento de trabajo Programa Magister Dirección de Empresas (MDE), Universidad del Bío Bío, Chile. 7Pág.*

Parisi, A., Améstica, L., & Chileno, O. (2019). *Predicción de variaciones en el precio del petróleo con el modelo de optimización ARIMA, innovando con fuerza bruta operacional. Vol 13 (1), 53-70.*

Parisi, F., and Parisi, A. (2010). *Teoría de Inversiones*. Chile: Copygraph.

Pesaran, M., Timmermann, A. (1992). *A Simple Nonparametric Test of Predictive Performance. Journal of Business & Economic Statistics*, 1992, vol. 10, issue 4, 561-65.

Pincheira, P., & Hardy, N. (2019). Forecasting base metal prices with the Chilean exchange rate. *Resources Policy. Vol 62, 256-281.*

Poterba, J. & Summers L. (1988). Mean reversion in stock prices: Evidence and Implications. *Journal of Financial Economics*, Vol 22 (1), 27-59.

Riveros, D. (2015). Aplicación de la investigación de operaciones al problema de la distribución a una empresa logística. Universidad Nacional Mayor de San Marcos, Perú. Obtenido de: <https://core.ac.uk/download/pdf/54216940.pdf>

Torbat, S., Khashei, M., Bijari, M. (2018). A hybrid probabilistic fuzzy ARIMA model for consumption forecasting in commodity. *markets Economic Analysis and Policy* vol 58, 22-31.

Wu, Y. & Zhang, H. (1997). Forward premiums as unbiased predictors of future currency depreciation: A non-parametric analysis. *Journal of International Money and Finance*, Vol 16 (4),609-623.