



Universidad CENFOTEC

Maestría en Tecnología de Bases de Datos

Documento final Proyecto de Investigación Aplicada 2

Tema

Desarrollo de un modelo avanzado para predecir las fluctuaciones del tipo de cambio entre el colón costarricense y el dólar estadounidense mediante técnicas de aprendizaje automático y análisis de series temporales

Autor

Felipe Corrales Karpukhina

Fecha

Marzo, 2024

TRIBUNAL EXAMINADOR

Este proyecto fue aprobado por el Tribunal Examinador de la carrera: **Maestría en Tecnología de Bases de Datos**, requisito para optar por el título de grado de **Maestría**, para el estudiante: **Corrales Karpukhina Felipe**.

TOMAS DE CAMINO BECK (FIRMA)
Firmado digitalmente por TOMAS DE CAMINO BECK (FIRMA)
Fecha: 2024.02.29 10:58:56 -06'00'

Tomás de Camino Beck, Ph.D.
Tutor

ELIECER VARGAS OCONITRILLO (FIRMA)
Firmado digitalmente por ELIECER VARGAS OCONITRILLO (FIRMA)
Fecha: 2024.03.01 09:12:30 -06'00'

Máster Carlos Vargas Oconitrillo
Lector 1

ALONSO JOSE GUEVARA FERNANDEZ (FIRMA)
Digitally signed by ALONSO JOSE GUEVARA FERNANDEZ (FIRMA)
Date: 2024.03.11 14:08:57 -06'00'

M.Sc. Alonso Guevara Fernández
Lector 2



San José, Costa Rica, 28 de febrero de 2024

Tema:

Desarrollo de un modelo avanzado para predecir las fluctuaciones del tipo de cambio entre el colón costarricense y el dólar estadounidense mediante técnicas de aprendizaje automático y análisis de series temporales

I. Introducción	2
II. Materiales y métodos	3
Evolución histórica del tipo de cambio	3
Métodos de inteligencia artificial	4
Árbol de regresión	4
Bosques Aleatorios	6
Series de tiempo	7
Modelo ARIMA	7
Modelo ARCH	8
Modelo GARCH	9
III. Resultados	10
Construcción de métodos de aprendizaje automático	11
Modelo árbol de regresión	12
Modelo bosques aleatorios	13
Construcción de series de tiempo	14
ARIMA	14
Componente I	14
Componente AR	14
Componente MA	16
ARCH	17
GARCH	17
Resultados de las series de tiempo generadas	17
IV. Discusión	18
V. Conclusiones	24
V. Agradecimientos	25
VI. Referencias	25
Anexos	26

Resumen

El tipo de cambio impacta la vida de las personas y las empresas en una economía no dolarizada. El riesgo cambiario afecta las decisiones financieras de los agentes económicos a corto, mediano y largo plazo. Se pueden utilizar técnicas de minería de datos y métodos de aprendizaje automático para pronosticar la clase de cambio. En este trabajo se compara el desempeño de los modelos de pronóstico: árboles de regresión, bosques aleatorios, ARIMA, GARCH y ARCH. Se utiliza la técnica del *backtesting* y se calcula el error porcentual absoluto medio para determinar el mejor modelo de pronóstico del tipo de cambio entre el colón costarricense y el dólar estadounidense.

Palabras clave

Tipo de cambio USD/CRC, ARIMA, ARCH, GARCH, pronóstico, series de tiempo, aprendizaje automático y minería de datos.

Abstract

The exchange rate is a matter that impacts the lives of individuals and companies. In an economy that is not dollarized, exchange rate risk affects the financial decisions that economic agents must make in the short, medium, and long term. By utilizing data mining techniques through machine learning methods and time series, estimates can be made to forecast the exchange rate. This paper compares the performance of forecasting models: regression trees, random forests, ARIMA, GARCH, and ARCH; using backtesting technique and by calculating the mean absolute percentage error, it aims to determine the best forecasting model for the exchange rate between the Costa Rican colon and the US dollar.

Keywords:

Exchange rate USD/CRC, ARIMA, ARCH, GARCH, Forecast, time series, machine learning, data mining

I. Introducción

En los mercados cambiarios es muy importante pronosticar el tipo de cambio. Esto permite realizar una estimación de este y facilitar una mejor toma de decisiones por parte de los agentes económicos. Por esta razón, se realizan mediciones de cuánto se desvía el pronóstico de la clase de cambio esperado en un horizonte de tiempo t y los métodos de pronóstico que se utilizan.

El pronóstico del tipo de cambio se ha abordado en varios trabajos en la región latinoamericana. Destacan los trabajos de Guzmán (2006), donde se aplica un modelo de pronóstico utilizando series de tiempo GARCH, ARCH y ARIMA para la clase de cambio de *spot* entre el peso mexicano y el dólar estadounidense entre 2005 y 2007. Por otra parte, Viales (2011) utiliza las series de tiempo ARIMA, ARCH y GARCH para realizar pronósticos sobre la volatilidad del tipo de cambio entre el colón y el dólar estadounidense entre 2009 y 2010, usando *backtesting* para determinar la mejor metodología de predicción entre diversos modelos. Finalmente, Dupouy (2014) usa diferentes técnicas de minería de datos, como árboles de decisión, redes neuronales y series de tiempo para estimar la tasa de interés en Chile entre 2008 y 2011 para distintas estrategias de inversión.

No se ha tratado de forma comparativa el tipo de cambio para Costa Rica utilizando técnicas de aprendizaje automático, como los árboles de regresión y los bosques aleatorios, contra las series de tiempo para estimar el valor de la clase de cambio en un horizonte de tiempo. En el presente trabajo se revisan y aplican métodos avanzados de pronóstico estadístico y de inteligencia artificial. Se utilizan técnicas como los árboles de regresión, bosques aleatorios y técnicas de series temporales como ARIMA, ARCH y GARCH para estimar el tipo de cambio en Costa Rica.

El presente trabajo busca de manera exhaustiva aplicar en los modelos de aprendizaje automático diferentes escenarios y para las series de tiempo órdenes que tengan valores significativos estadísticamente. Con el objetivo de determinar el mejor modelo de predicción del tipo de cambio, se realiza una prueba retrospectiva (*backcasting*) para cada modelo,

utilizando valores conocidos de la clase de cambio en un tiempo específico. Los modelos no se entrenarán para ese periodo específico, con el fin de asegurar la objetividad de los modelos de predicción. Se utiliza el error absoluto medio porcentual (MAPE) como métrica de desempeño para definir el modelo que se acerque más a los valores conocidos del tiempo especificado.

Este documento se divide en dos partes principales: la primera se enfoca en la selección y entrenamiento de modelos de inteligencia artificial y análisis de series temporales. La segunda se centra en mostrar los resultados que se derivan del *backcasting* y en elegir el mejor modelo de predicción por medio del error absoluto medio porcentual durante este periodo.

En los mercados de divisas se realizan investigaciones que modelan la volatilidad de las variaciones del tipo de cambio o el impacto de variables que se relacionan con el tipo de cambio. Sin embargo, no se llegó a evaluar un pronóstico para el tipo de cambio, a pesar de los avances en los campos de la inteligencia artificial y la minería de datos en los últimos años.

En el presente trabajo se realiza una revisión de métodos de pronóstico estadístico y de inteligencia artificial que estiman, con base en una serie de parámetros, el pronóstico de una variable cuantitativa. Se utiliza *backtesting* para evaluar el mejor método para pronosticar el tipo de cambio. Se emplean métodos de inteligencia artificial como los árboles de regresión y los bosques aleatorios y métodos estadísticos como las series de tiempo: ARIMA, ARCH y GARCH. Estos modelos pueden utilizarse para pronosticar el tipo de cambio mediante estimación de variaciones, la identificación de patrones a través del tiempo o la relación entre distintas variables macroeconómicas, con el fin de que se determine un posible valor.

El presente documento estructura una primera parte en la cual se recapitulan las variables cualitativas y cuantitativas que permiten entrenar los modelos de inteligencia artificial, así como los análisis que permiten la elección de los modelos ARIMA, ARCH y GARCH que se ajustan al mejor pronóstico del tipo de cambio. Posteriormente, se muestran los resultados de utilizar el *backcasting* con cada uno de estos modelos, con el fin de especificar el mejor pronóstico.

II. Materiales y métodos

Evolución histórica del tipo de cambio

El mercado cambiario costarricense ha tenido cuatro etapas en su política cambiaria. La primera etapa es la política cambiaria fija que estuvo vigente desde la fundación del BCCR en la década de 1950 hasta la década de los 80. Esta se caracterizó por un BCCR que establecía el tipo de cambio entre el colón y el dólar. Según Saborío (1997), en este periodo: “El manejo del comercio exterior era muy conservador, con el propósito de mantener un tipo de cambio estable y un nivel de reservas monetarias internacionales sano” (p. 100). De vez en cuando había alguna devaluación del colón, pero no era la norma.

En 1980, la economía costarricense entró en crisis, como consecuencia del aumento en el precio internacional del petróleo y de la caída de los precios del café, principal producto de exportación del país en ese momento. La crisis se caracterizó, especialmente, por una altísima inflación y por la disminución del poder adquisitivo

de la moneda, ambas violentas respecto del dólar, y por un aumento del desempleo (Chacón, 2019, p. 126).

Esta crisis inició la segunda etapa del mercado cambiario, conocida como el esquema de minidevaluaciones, que estuvo vigente desde 1984 hasta 2006. En esta política monetaria, el colón se devaluaba diariamente, lo que provocaba un aumento casi lineal en el tipo de cambio. “Este esquema le permitió al BCCR tener control y estabilidad del tipo de cambio teniendo como costo de oportunidad un limitado control sobre la inflación” (Viales, 2011, s. p.).

Posteriormente, se entra a la tercera etapa en la cual el BCCR tiene como objetivo reducir el impacto de la inflación en la economía mediante una mayor liberalización del mercado cambiario y buscando formas de enfrentar movimientos de alta volatilidad a través de intervenciones. Se propusieron las bandas cambiarias, en las cuales el BCCR evitaría que el tipo de cambio de compra bajara de un nivel o piso y el tipo de cambio de venta subiera más allá de un nivel o techo.

La última etapa corresponde a la liberalización parcial del tipo de cambio por medio de una flotación administrada, la cual entró en vigor el 2 de febrero de 2015. Rodríguez (2015) argumenta que:

En el esquema de flotación administrada el precio del dólar lo fijan las fuerzas de la oferta y la demanda, sin que existan límites a su cotización, como sucede en el modelo de banda cambiaria. Para controlar fluctuaciones violentas el Banco Central fija reglas que marcan su intervención en el mercado (s. p.).

Con base en el esquema de flotación administrada se pretende evaluar la posibilidad de pronosticar el tipo de cambio utilizando métodos estadísticos y algoritmos de inteligencia artificial. Principalmente, los actores económicos tienen incentivos para modificar sus expectativas a corto, mediano y largo plazo en cuanto a la oferta y demanda de dólares. En virtud de esto, el BCCR solo intervendría en casos extremos o cuando lo exija la ley.

La convertibilidad de la moneda es un asunto que atañe a los bancos centrales, el caso costarricense no es una excepción. En la Ley Orgánica del Banco Central de Costa Rica (BCCR) se indica que es la institución encargada de dirigir la política cambiaria, cosa que ha realizado desde su fundación en la década de los 50.

Desde ese momento hasta la actualidad, han existido varias formas de coordinar la política cambiaria en el país. En una primera instancia el BCCR decretaba que el tipo de cambio es fijo a través del tiempo, esto sin importar los cambios que sucedan en el mercado cambiario. El Banco Central y el gobierno hacían todo lo posible por mantener el tipo de cambio en una tasa que no cambiara. Esta política cambiaria estuvo vigente desde la fundación del BCCR hasta 1983. Durante este periodo el tipo de cambio fluctuó entre ₡6.62 a ₡40.00 por dólar, este periodo se caracterizó por prosperidad económica, particularmente entre 1960 y 1970. Este tipo de política monetaria se mostró ineficaz en el instante de manejar *shocks* externos como los que ocurrieron al final de la década de los 70 y principios de los 80 con la crisis del petróleo. Lo anterior fue particularmente crítico entre 1980 y 1982 donde, por ejemplo, el tipo de cambio se movió de ₡8.54 a ₡36.00 para el tipo de cambio de compra, representando un aumento de aproximadamente 300 % en 2 años, algo inédito para el país. De esta forma, se desató una crisis muy profunda que hoy en día se recuerda de forma muy negativa entre las personas que vivieron esa época.

Ante la situación descrita, el BCCR decidió perseguir una nueva política cambiaria: la de las minidevaluaciones. El régimen cambiario costarricense se caracterizó principalmente por:

Ajustes pequeños y periódicos del tipo de cambio nominal, de acuerdo al comportamiento de los diferenciales de inflación y de los tipos de cambio nominales de los principales socios comerciales. Este sistema fue conocido como un régimen de minidevaluaciones y tuvo su origen en la tesis jurídica, de que todas o casi todas las transacciones de divisas se llevarían a cabo en el mercado libre interbancario y el Banco Central en ejercicio de su virtual monopolio en la compra de divisas, pondría el precio en ese mercado libre (Guardia, 1993, citado por Saborío, 2014, p. 103).

Esto permitió no solo traer estabilidad a la moneda, sino que también permitió la entrada de capital y buscar equilibrio en la balanza de pagos del comercio exterior generada por la crisis. El esquema de minidevaluaciones funcionó desde el 1 de enero de 1983 hasta el 7 de octubre de 2006. En dicho periodo, el tipo de cambio de compra pasó de ₡40.00 a ₡520.04 en los 23 años que estuvo en vigencia dicha política.

A pesar de que la economía pudo recuperarse de la grave crisis provocada en los años 80, el BCCR se vio incapaz de cumplir las metas de inflación que se plantearon, ya que la moneda sufrió una devaluación muy acelerada. Por lo tanto, se decide cambiar hacia una política cambiaria de bandas.

Este esquema, de carácter más flexible que los anteriores, propició una mayor autonomía al Banco Central, adquiriendo un mayor control y efectividad de la política monetaria, con el objetivo principal de reducir las tasas de inflación a niveles de los principales socios comerciales y de esta forma promover la estabilidad macroeconómica. En este sistema, el Banco Central se comprometía a mantener el tipo de cambio entre ciertos límites (Techo y Piso), dentro de los cuales la cotización de la moneda extranjera puede fluctuar libremente (Alvarado, 2020, p. 7).

Esta política cambiaria estuvo vigente hasta el 2 de febrero de 2015, donde el tipo de cambio de compra pasó de ₡520.04 a ₡530.85. A partir de esta fecha el BCCR decide perseguir un esquema cambiario de flotación administrada donde:

Hay dos situaciones que favorecen la eliminación de la banda: la baja en el precio del petróleo y que viene en camino una nueva emisión de eurobonos, con lo cual va a haber una menor demanda de divisas y más bien se podría observar una presión a la baja del tipo de cambio (Morales y Delgado, 2015, s. p.).

Alvarado (2020) indica que:

Bajo el esquema de flotación administrada el tipo de cambio es determinado por el mercado, pero el Banco Central se reserva la posibilidad de realizar operaciones de intervención en el mercado de divisas para moderar fluctuaciones importantes en el tipo de cambio y prevenir desvíos de éste con respecto al que sería congruente con el comportamiento de las variables que explican su tendencia de mediano y largo plazo, esto como parte de la transición hacia un esquema monetario de metas de inflación que requiere de una mayor flexibilidad cambiaria (p. 11).

Con base en esa flexibilidad cambiaria, se pretende evaluar la posibilidad de poder pronosticar el tipo de cambio entre el colón costarricense y el dólar estadounidense utilizando métodos estadísticos y algoritmos de inteligencia artificial. Esto se debe principalmente a que los actores económicos tienen ciertos incentivos ante determinados parámetros para modificar sus expectativas a corto, mediano y largo plazo en cuanto a la oferta y demanda de dólares, en virtud de la cual el BCCR solo intervendría en casos extremos o los que le demande la ley. De acuerdo con lo anterior, para la siguiente

investigación se utilizan los métodos de pronóstico fundamentados en inteligencia artificial y series de tiempo.

Métodos de inteligencia artificial

Árbol de regresión

Los árboles de regresión permiten hacer agrupaciones sobre datos tanto cuantitativos como cualitativos. A grandes rasgos, un árbol de regresión categoriza los datos con base en ciertos parámetros y luego realiza una función de agrupamiento, como promedio o mediana, para aproximar el valor de una variable de predicción. Esto se logra utilizando una serie de valores de entrada de la misma clase que los parámetros elegidos.

Según Martínez (2023):

Los árboles de clasificación y regresión (CART=Classification and Regression Trees) son una alternativa al análisis tradicional de clasificación/discriminación o a la predicción tradicional (regresión). Entre las ventajas de estos árboles CART podemos destacar su robustez a outliers, la invarianza en la estructura de sus árboles de clasificación o de regresión a transformaciones monótonas de las variables independientes, y sobre todo, su interpretabilidad (p. 1).

Los árboles de regresión consisten en tres partes:

- Nodo raíz: este es el que contiene los datos generales del conjunto de datos.
- Nodos secundarios o internos: estas son las subdivisiones que categorizan a los datos.
- Hojas o nodos terminales: son los valores finales que el algoritmo pronostica para el valor, para el caso de los árboles de regresión este suele ser el resultado de una función de agrupación.

En el momento de construir un árbol de regresión, es necesario tomar en cuenta que existe un coeficiente de pureza y uno de entropía. La pureza de los datos existe cuando un conjunto de datos entra en el algoritmo y se clasifica a lo largo de los nodos internos hasta llegar a uno terminal. Si los datos corresponden a un mismo valor o clase en ese nodo, se puede afirmar que tienen una pureza alta (o un nivel de impureza baja). Si los datos muestran clases diferentes al seguir los nodos internos, se puede afirmar un índice de pureza bajo (o de impureza alto). Matemáticamente, se representa de la siguiente forma:

$$1 - \sum_{i=1}^k (p_i^2)$$

Donde:

p: es la probabilidad de que todos los valores pertenezcan a la misma clase.

i: la clase elegida.

k: el número de clase.

La fórmula anterior se denomina un índice de coeficiente de Gini. Se busca que el índice sea el menor posible (0) o el máximo posible (1), cualquier número real entre estos valores indicaría un nivel bajo de pureza.

Los árboles de regresión también deben tener en cuenta la entropía al construir árboles de regresión, ya que esta se relaciona con la pureza. La entropía representa el grado de distancia entre las hojas en términos de los valores que clasifican. Si la probabilidad de que los nodos terminales sean similares es alta, significa que el árbol no está categorizado correctamente en sus nodos internos, lo que resultará en una alta variación entre el valor actual y el valor pronosticado de algunos datos. Matemáticamente, se representa de la siguiente manera:

$$-\sum_{i=1}^N p_i \log_2 p_i$$

Donde

i: es la clase elegida.

p es la probabilidad de elegir aleatoriamente un elemento de la clase i.

N: es el número total de clases.

Si todos los elementos de la clase son de 100 %, la entropía como consecuencia es 0. Esto representa una pureza muy alta e indica que no es necesario separar el conjunto de datos en más nodos.

La entropía es importante porque permite evaluar la ganancia de información (GI) en el árbol. En el momento de segregar los conjuntos de datos en nodos internos desde un nodo raíz es necesario valorar cuál clasificación logra la mayor ganancia de información al reducir la cantidad de entropía de un nodo raíz a un nodo interno. Matemáticamente, se representa de la siguiente forma:

$$GI = E_{nodo\ raíz} - E_{nodos\ internos}$$

Generalmente, los nodos internos se pueden dividir en varios nodos. Se utiliza el promedio simple o promedio ponderado de cada uno, donde el peso del anterior puede ser la distribución de frecuencias de los datos en cada nodo. Se busca que la variable que se utilice para particionar el nodo interno genere una ganancia de información. Este proceso es iterativo y, generalmente, lo realizan las computadoras.

Bosques aleatorios

Los árboles, a pesar de ser algoritmos de aprendizaje automático muy efectivos, tienen dos desventajas: sesgo y varianza. El sesgo se representa como el grado de pureza que tienen las hojas en el árbol. Desde que este tiene una decisión binaria en cada nodo, existe una gran probabilidad de que no clasifique correctamente los datos. Esto último lleva a que las predicciones de las hojas generen discrepancias entre los datos propuestos por el modelo y los datos actuales, es decir, la varianza.

Para minimizar esto, se desarrollaron los bosques aleatorios, los cuales son una forma recursiva y aleatoria de llevar a cabo conjuntos de árboles de regresión. Cada uno de estos árboles produce un valor determinado por la lógica en la que se construyeron los nodos y se obtiene un resultado global.

La construcción de un bosque aleatorio implica una serie de pasos adicionales. Sin embargo, un algoritmo de bosque aleatorio es esencialmente un conjunto de árboles de

regresión. En el desarrollo del algoritmo de bosque aleatorio, se particiona aleatoriamente el conjunto de datos de entrenamiento. Esto implica excluir ciertas filas y repetir otras. Una vez hecho esto, el algoritmo construye árboles de regresión. En cada iteración, se seleccionan de forma aleatoria los campos o categorías, excluyendo algunos de ellos.

Dado un conjunto de datos universo C , $D = \{(X_1, Y_1), \dots, (X_N, Y_N)\}$ los datos de entrenamiento, con $X_i = (X_1, \dots, X_p)$ como el conjunto de datos predictores o campos, con valores posibles de $J = 1$ hasta K , donde K es el número de árboles que se decide hacer del bosque. Se toma un conjunto D_j de tamaño N con $D_j \subseteq C$, usando el conjunto D_j , se construye un árbol de regresión seleccionando al azar m predictores del conjunto X_i de p predictores disponibles, con $i \leq p$. Este proceso se repite K veces.

La forma en la que los bosques aleatorios toman decisiones sobre los valores para realizar pronósticos consiste en que cada uno de los K árboles creados determina un valor para un conjunto de datos de entrada con base en el algoritmo del árbol a_l con $l = \{1, \dots, K\}$. Esto da un grupo de K resultados que se evaluaron por votación en caso de los bosques aleatorios de clasificación o una función de agregación (promedio, suma, etc.).

Series de tiempo

Según Hernández (2015): “Las series de tiempo son colecciones de observaciones sobre un determinado fenómeno efectuadas en sucesivos momentos del tiempo, usualmente equiespaciados. Corresponde a una realización de un proceso generador de datos” (s. p.). Las series de tiempo se pueden dividir en dos clases: las estocásticas que son series temporales o componentes de ellas que tienen una parte que es fija o que tiene muchas probabilidades de poder predecirse y una parte aleatoria que no tiene ninguna correlación de ninguna clase y las determinísticas, estas son series de tiempo o componente que no tienen la variable de cambio aleatorio. Por lo tanto, los pronósticos son muy fiables.

Según Minitab (2023) estos son componentes:

Tendencia (Te): La tendencia a largo plazo de una serie de aumentar o disminuir (tendencia creciente o tendencia decreciente).

Estacionalidad (Es): La fluctuación periódica en las series de tiempo dentro de un período determinado. Estas fluctuaciones forman un patrón que tiende a repetirse de un período estacional al siguiente.

Ciclos (Ci): Largas desviaciones de la tendencia debido a factores diferentes de la estacionalidad. Los ciclos por lo general se producen durante un intervalo de tiempo extenso, y los tiempos que transcurren entre los picos o valles sucesivos de un ciclo no necesariamente son iguales.

Movimiento Irregular (MI): El movimiento que queda después de explicar los movimientos de tendencia, estacionales y cíclicos; ruido aleatorio o error en una serie de tiempo (s. p.).

Matemáticamente, los componentes de la serie temporal X_t se pueden relacionar de las siguientes formas:

- $X_t = Te + Es + Ci + MI$

$$X_t = Te * Es * Ci * MI$$

Pueden usarse otros tipos de relación, sin embargo, esos son los más comunes y los que se utilizan en esta investigación.

Modelo ARIMA

Los modelos ARIMA consideran los factores autorregresivos (AR) y de media móvil (MA) de las series de tiempo estacionarias para revelar las correlaciones de los valores a través del tiempo. Las series de tiempo estacionarias implican un ciclo aislado de los movimientos aleatorios o el ruido blanco.

Generalmente, las series temporales no son estacionarias en la vida real. Para lograr eso, es necesario aislar el componente lineal que impacta a la serie de tiempo. Con el fin de lograrlo, la serie de tiempo se deriva (I). Para determinar si una serie temporal es estacionaria, se pueden hacer pruebas como la Dicky-Fuller o la prueba Phillip-Perron. Explicar estas pruebas sobrepasa los objetivos de esta investigación. Sin embargo, se puede resaltar que, para el caso de ambas pruebas, la hipótesis nula H_0 implica que la serie temporal es no estacionaria. En caso contrario, de descartarse la hipótesis nula H_1 implica una serie de tiempo estacionaria.

El componente AR(p) o componente autorregresivo de orden p de la serie de tiempo mide el impacto de los valores pasados con un retraso p en los valores presentes de forma independiente. Este toma en cuenta los valores que se encuentran en el medio. Matemáticamente, se expresa mediante la siguiente fórmula:

$$y(t) = \phi_0 + \phi_1 * y(t - 1) + \phi_2 * y(t - 2) + \dots + \phi_p * y(t - p) + \varepsilon(t)$$

Donde:

$y(t)$: es el valor presente

$y(t-p)$: es el valor p unidades en el pasado

ϕ_p : es el coeficiente de correlación que tiene un retraso p

$\varepsilon(t)$: el factor del ruido blanco en la serie temporal

ϕ_0 : la tendencia más básica de serie temporal.

El método ARIMA también cuenta con el componente MA (q) o componente del promedio móvil, es muy importante no confundirlo con un promedio móvil que busca el promedio de un valor tomando en cuenta q retrasos, más bien intuye un modelo de predicción de datos pasados atrasados q unidades para la serie de tiempo, se compara el estimado por el modelo de predicción contra el valor real atrasado (esto se llama error o *shock*) y le asigna un peso para evaluar cuánto impacta al valor actual una desviación significativa. Es relevante indicar que el componente MA no evalúa los elementos independientemente como el componente AR, sino que toma en cuenta los valores en medio del valor actual y el retraso q. Matemáticamente, se representa de la siguiente forma:

$$y(t) = c + \varepsilon(t) + \theta_1 * \varepsilon(t - 1) + \theta_2 * \varepsilon(t - 2) + \dots + \theta_p * \varepsilon(t - q)$$

Donde:

$y(t)$: es el valor presente

$y(t-p)$: es el valor p unidades en el pasado

$\varepsilon(t-q)$: es el error entre el valor actual menos el pronosticado

θ_p : es el peso que tiene cada error pasado en el valor actual

c: es la constante básica del error

Entre algunos métodos que se pueden usar para realizar el pronóstico del MA se tiene la posibilidad de utilizar, por ejemplo, el método de mínimos cuadrados, también el de máxima verosimilitud. Sin embargo, existen otros métodos para hacer este pronóstico describirlos no forma parte de la presente investigación.

Modelo ARCH

El modelo ARIMA tiene una serie de limitaciones al estudiar ciertos fenómenos debido al supuesto inicial de una varianza constante. Por ejemplo, en las series temporales de las finanzas, en la mayoría de los casos, la volatilidad del periodo anterior en los datos impacta la volatilidad del periodo actual y es distinta según el momento en el que se encuentre el activo. En estos casos, el modelo ARIMA no presenta resultados tan buenos.

En virtud de lo anterior, se crearon los modelos heterocedásticos e introdujeron los mismos. Estos modelos buscan medir el grado de dispersión que presentan los datos con un p retraso en los datos. Es decir, se busca analizar el elemento heterocedástico de las series de tiempo, ya que las varianzas o $\sigma^2(t)$ varían a medida que se avanza en la serie temporal.

Para el caso del ARCH(p) o modelo autorregresivo con heterocedasticidad condicional con p retrasos, no se trabajan con valores, sino con los retornos (o cambios porcentuales) de estos. Lo anterior tiene el fin de que se determine cuánto varían los datos. El modelo ARCH(p) es equivalente en funcionamiento al componente AR(p) del modelo ARIMA, ya que los retornos anteriores impactan en los retornos actuales. Matemáticamente, se representan de la siguiente forma:

$$\sigma^2(t) = \alpha_0 + \alpha_1 * \varepsilon^2(t-1) + \alpha_2 * \varepsilon^2(t-2) + \dots + \alpha_p * \varepsilon^2(t-p)$$
$$\varepsilon(t) = \sigma(t) * \epsilon(t)$$

Donde:

- $\varepsilon(t)$ es el error del periodo actual
- $\epsilon(t)$ es el factor del ruido blanco en el periodo actual
- $\sigma(t)$ es la desviación estándar del periodo t (la cual generalmente se estima)
- $\sigma^2(t)$ es la varianza del periodo t
- α_0 representa el parámetro base o la intersección
- α_p sería el valor del peso del parámetro p siendo p el retraso elegido
- $\varepsilon^2(t-p)$ sería el error al cuadrado, es decir, el valor del retorno esperado menos el actual en el periodo p , o en el pasado.

Modelo GARCH

Existe una extensión de los modelos ARCH. Estos se conocen como modelos generalizados autorregresivos condicionalmente heterocedásticos (GARCH). En estos modelos, además del componente autorregresivo del retorno, se agrega también el componente de promedio móvil (MA). Este componente intenta medir el impacto que tienen las varianzas extremas en el pasado. El funcionamiento de los modelos GARCH es similar en cuanto a conceptos al modelo ARIMA. Sin embargo, es importante tener en cuenta que estos modelos solo toman en cuenta la variabilidad de la serie de tiempos.

Matemáticamente, un modelo GARCH(p,q), con p como los periodos de atraso sobre la

varianza (componente AR) y q como los periodos de atraso de los errores (con base en un pronóstico, el componente MA) de la varianza, se expresa de la siguiente forma:

$$h(t) = a_0 + \sum_{i=1}^p \alpha_i \varepsilon_{2t-i} + \sum_{j=1}^q \beta_j h_{t-j}$$

Donde:

a_0 : es mayor a cero y representa a una varianza base, generalmente es una volatilidad a largo plazo, como una media.

$\sum_{i=1}^p \alpha_i \varepsilon_{2t-i}$: este es el componente AR que mide el impacto de las pasadas varianzas sobre la varianza actual a través de un parámetro α_i

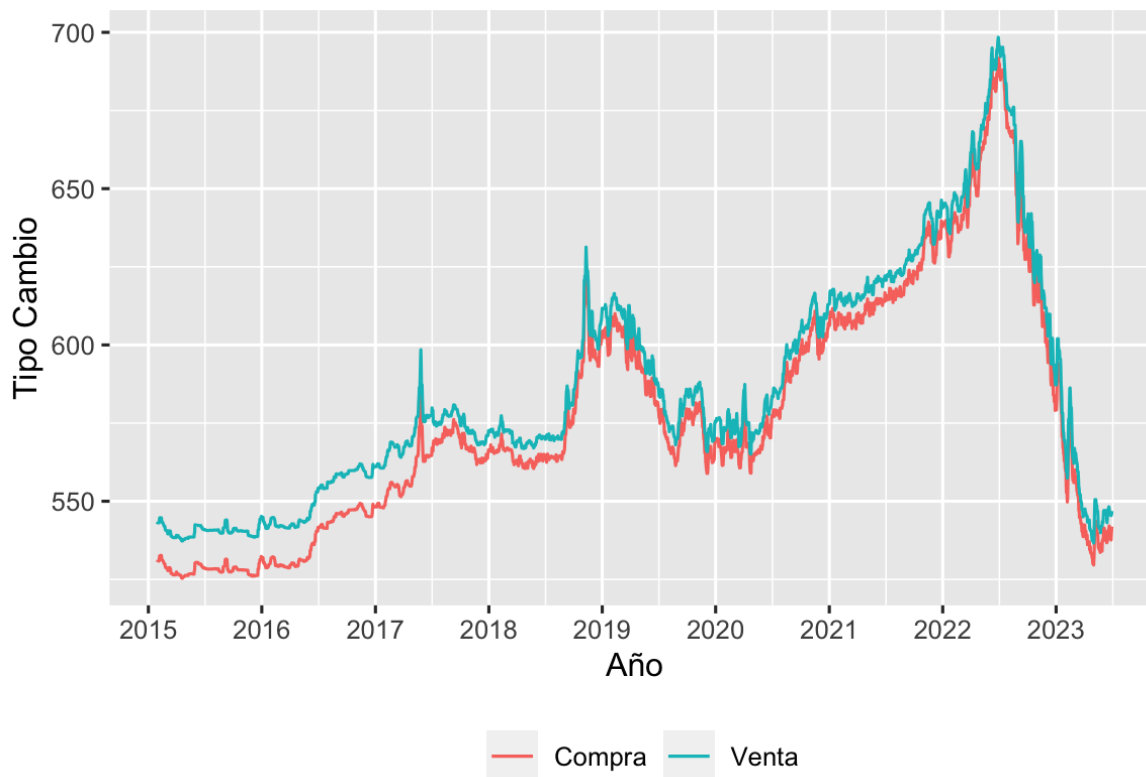
$\sum_{j=1}^q \beta_j h_{t-j}$: este es el componente MA donde se hace un pronóstico de las varianzas y se estima el impacto que tiene sobre el valor de la varianza actual pesando cada uno de los parámetros con un atraso q teniendo el peso β_j .

Finalmente, en el caso de las series de tiempo, se deben utilizar combinaciones entre los modelos ARIMA y los modelos GARCH. Se puede observar que un modelo ARCH(p) es un modelo GARCH(p, q) con un $q=0$. El algoritmo que se utiliza para pronosticar debe usar tanto los componentes del ARIMA como las variaciones históricas con el GARCH, con el objetivo de lograr un pronóstico más preciso.

III. Resultados

Con el fin de realizar el estudio comparativo es necesario crear un conjunto de datos para entrenar los modelos de aprendizaje supervisado. Por otra parte, la serie de tiempo no requiere otra entrada de datos aparte de la serie histórica del tipo de cambio. El presente estudio se enfoca en los datos de la política cambiaria vigente a partir del 2 de febrero de 2015, que corresponde a la flotación administrada. Esto se hace para mantener una homogeneidad en la forma en la que el BCCR toma las decisiones sobre el tipo de cambio.

Figura 1. Tipo de cambio del esquema de flotación administrada entre 2015 y 2023



Fuente: elaboración propia con datos del BCCR.

Se puede evidenciar que la compra y la venta se mueven en la misma dirección según la Figura 1. Por lo tanto, se evalúa la capacidad del modelo de pronóstico para predecir el tipo de cambio de compra, que representa el suelo. Tanto el conjunto de datos para los modelos de aprendizaje automático como las series de tiempo utilizan los datos disponibles desde el 2 de febrero de 2015 hasta el 31 de mayo de 2023. Se excluyen los datos del 1 de junio de 2023 hasta el 30 de junio de 2023, con el fin de realizar el *backtesting*.

Construcción de métodos de aprendizaje automático

Para construir el árbol de regresión y los bosques aleatorios es necesario realizar una serie de transformaciones al conjunto de datos. Esto permite categorizar los datos, de manera que el algoritmo pueda asignar los coeficientes de Gini y segmentarlos. Los datos que no se repiten, que son llaves primarias como la fecha, se segmentan de forma tal que puedan mapearse datos que permitan calificarlos, tales como:

- Número de mes
- Día de la semana
- Número de trimestre
- Número de semana Iso
- Días para la siguiente elección

Estos datos permiten categorizarlos y agruparlos. Por lo tanto, son candidatos para ser datos cualitativos. Asimismo, se remueven todos los datos desconocidos o nulos en los modelos cuantitativos. Esto incrementa la pureza del conjunto de datos y permite hacer agrupaciones más adecuadas. Los datos cuantitativos del conjunto de datos por utilizar incluyen:

- Promedio móvil 30 días del tipo de cambio.
- Tasas TRI, cada uno en una columna distinta de:
 - 1 semana
 - 1 mes
 - 3 meses
 - 6 meses
 - 9 meses
 - 12 meses
 - 24 meses
 - 36 meses
 - 60 meses
- Balanza comercial trimestral
- Exportaciones trimestrales
- Importaciones trimestrales
- Reservas del BCCR
- Total de intervenciones de compra hechas por el BCCR.
- Total de intervenciones de venta hechas por el BCCR.

El conjunto de datos se divide en un 70 % de datos para entrenar los modelos de aprendizaje y un 30 % de la muestra para la prueba del modelo entrenado. Esta separación se realiza de forma aleatoria. Como medida de referencia, el modelo ingenuo para este caso asume que el tipo de cambio tiene un valor igual al promedio de todos los tipos de cambio registrados hasta el año 2023 para la muestra.

Tabla 1. Correlación de las variables con el tipo de cambio sobre los datos del conjunto de entrenamiento

Variable	Correlación con el tipo de cambio de compra
Promedio.Movil.30D.Compra	0.98552707
Tasa.TRI.6meses	-0.75627408
Tasa.TRI.12meses	-0.72703223
Tasa.TRI.9meses	-0.67507173
Precio.WTI.USD	0.64909834
Tasa. TRI.3meses	-0.63532278
Tasa. TRI.60meses	-0.62617195
Tasa.TRI.24meses	-0.60247824
Importaciones.Trimestral.MillonesUSD	0.55155404
Exportaciones.Trimestral.MillonesUSD	0.54980548
Reservas.BCCR.MillonesUSD	-0.51702571
Tasa.TRI.1semana	-0.50636824

Tasa.TRI.36meses	-0.48912659
Tasa.TRI.1mes	-0.32618636
Balanza.Comercial.Trimestral.MillonesUSD	0.16310586
Intervenciones.Total.Ventas	0.08708587
Intervenciones.Total.Compras	0.00416874

Modelo árbol de regresión

Para la creación de los modelos de árbol de regresión se entrenan varios modelos en los cuales se combina la variable predictiva de cambio compra con diferentes circunstancias, tanto variables cuantitativas como cualitativas. Lo anterior tiene el fin de encontrar el modelo más apropiado utilizando el paquete rpart de R. Para el presente trabajo se proponen las siguientes pruebas:

- Modelo árbol-Todas las variables.
- Modelo árbol-Solo variables correlación mayor que 0.6 y -0.6 sin promedio móvil.
- Modelo árbol-Solo variables correlación mayor que 0.6 y -0.6 con promedio móvil.
- Modelo árbol-Solo con variables cualitativas.

Al entrenar los modelos sobre el conjunto de entrenamiento, se obtienen los siguientes resultados de las métricas.

Tabla 2. Métricas desempeño para los modelos de árbol sobre datos de entrenamiento y de prueba

Nombre del modelo	RSME en entrenamiento	RSME en prueba	R cuadrado en prueba
Modelo árbol-Todas las variables	5.89	7.04	98.2
Modelo árbol-Solo variables correlación mayor que 0.6 y -0.6 sin promedio móvil	8.49	18.89	86.99
Modelo árbol-Solo variables correlación mayor que 0.6 y -0.6	5.89	7.14	98.14
Modelo árbol-Solo con variables cualitativas	32.7	32.23	62.13

Modelo bosques aleatorios

Para la creación de los modelos de bosques aleatorios, se entrenan varios modelos en los cuales se combina la variable predictiva de cambio compra con las variables cuantitativas y cualitativas en diferentes circunstancias. Lo anterior tiene el fin de encontrar el modelo más apropiado utilizando el paquete randomForest de R. Para el presente trabajo se proponen las siguientes pruebas:

- Modelo bosque-Todas las variables.
- Modelo bosque-Solo variables correlación mayor que 0.6 y -0.6 sin promedio móvil.
- Modelo bosque-Solo variables correlación mayor que 0.6 y -0.6 con promedio móvil.
- Modelo bosque-Solo con variables cualitativas.

Ahora, para los árboles aleatorios, la librería randomForest tiene la función randomForest que construye el modelo de aprendizaje supervisado. Además, permite calcular la cantidad de árboles necesarios para alcanzar el mínimo error cuadrado (MSE por sus siglas en inglés). Por lo tanto, en estos experimentos se construye el modelo, se consulta la cantidad óptima de árboles para reducir el MSE y, finalmente, se aplica dicho parámetro nuevamente al modelo para optimizarlo. Se obtienen los siguientes resultados:

Tabla 3. Métricas de los modelos de bosque sobre datos de entrenamiento y de prueba

Nombre del modelo	RSME en entrenamiento	RSME en prueba	R cuadrado en prueba
Modelo bosque-Todas las variables	1.91	1.62	99.9
Modelo bosque-Solo con variables cualitativas	5.79	32.65	61.14
Modelo bosque-Solo variables correlación mayor que 0.6 y -0.6 con promedio móvil	3.71	2.74	99.73
Modelo bosque-Solo variables correlación mayor que 0.6 y -0.6 sin promedio móvil	2.28	1.96	99.86

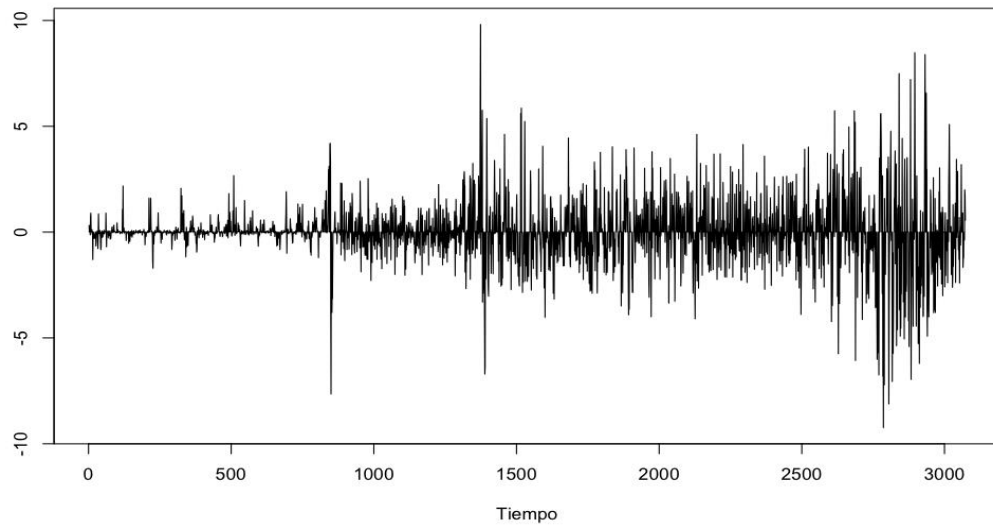
Construcción de series de tiempo

ARIMA

Componente I

Al realizar la prueba de Dickey-Fuller aumentada se obtiene un valor p poco significativo de 0.99. Por lo tanto, es necesario aplicar una derivada sobre los datos de la serie de tiempo. Al aplicar la misma prueba sobre la serie de tiempo derivada una vez, se obtiene que la serie de tiempo resultante tiene un valor p menor que 0.01, lo que hace el valor muy significativo y rechaza la hipótesis nula.

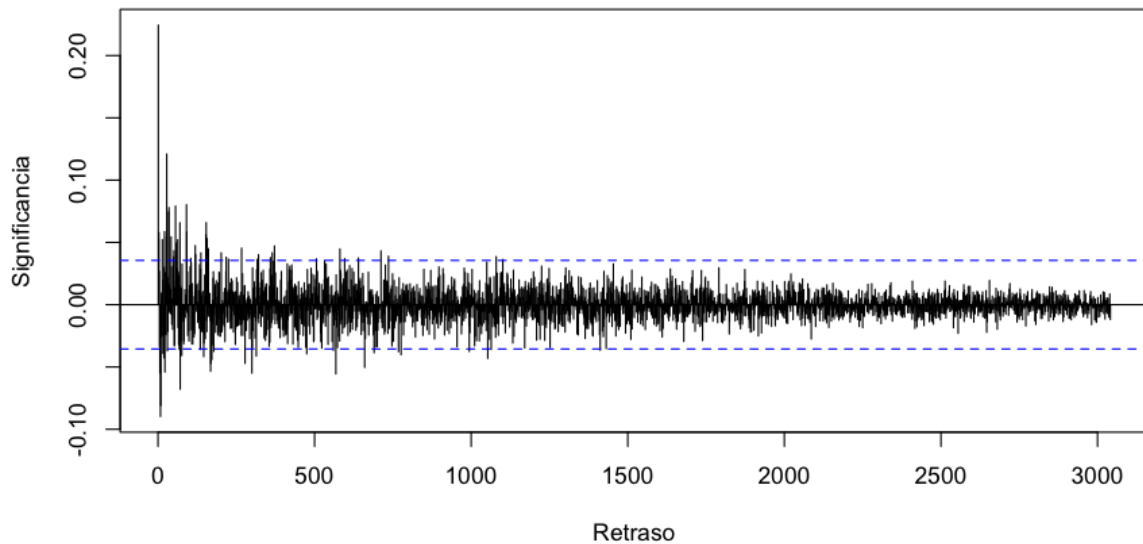
Figura 2. Serie de tiempo del tipo de cambio de compra con una diferenciación



Componente AR

En la serie de tiempo diferenciada, se calcularán los coeficientes del componente autorregresivo (AR). Para aplicar la autorregresión, es necesario calcular la correlación de los valores con un retraso (lag). Esto permite extraer los valores más significativos que tienen más peso en la serie de tiempo.

Figura 3. Función PACF en la serie de tiempo de compra diferenciada



Todos los valores que estén por encima del valor $\rho = 0.029$ o que sean menores $\rho = -0.029$, representados por la línea azul (con 10 % de significancia), son los retrasos (en este caso días) que influyen sobre el valor del tipo de cambio presente.

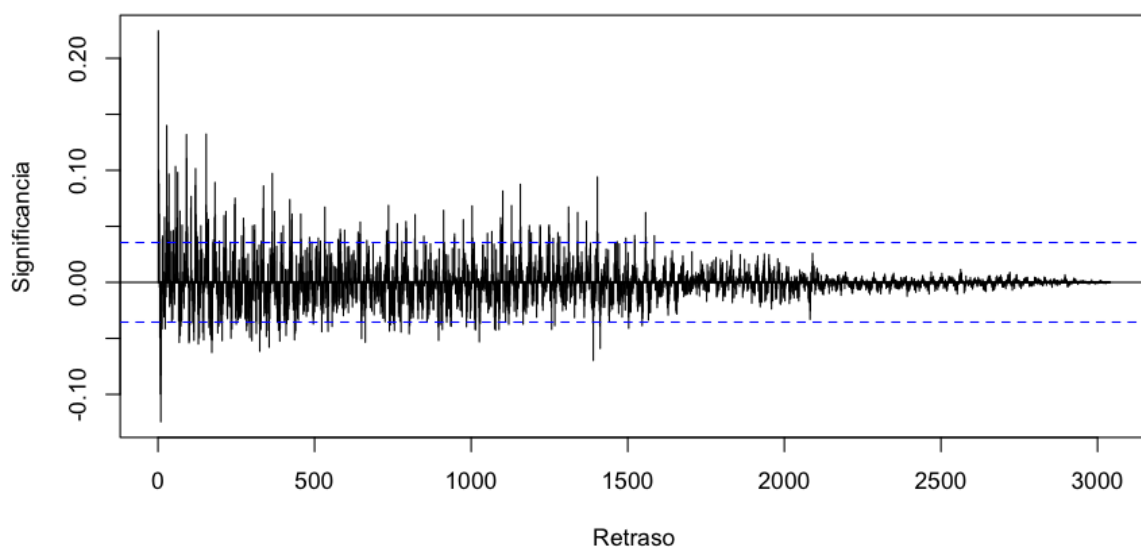
Tabla 4. Los 10 valores más relevantes de la función autorregresiva pacf sobre la serie de tiempo de cambio diferenciada

Valores p	Retraso	Valores absolutos de p
0.2275445	1	0.2275445
0.12299287	28	0.12299287
-0.08931883	8	0.08931883
-0.08189771	9	0.08189771
0.07973326	91	0.07973326
0.07912618	35	0.07912618
0.07888825	56	0.07888825
0.07357835	34	0.07357835
0.06733022	154	0.06733022
-0.06591573	71	0.06591573

Componente MA

El componente MA también debe ser considerado, ya que permite calcular el peso de los errores de pronóstico en la serie de tiempo. Esto da la posibilidad de identificar cuáles días son más relevantes para el cálculo del error actual. Para esto, R tiene una función llamada `acf` o correlograma total, que mide la significancia de cada uno de los retrasos.

Figura 4. Función ACF en la serie de tiempo de compra diferenciada



Todos los valores que estén por encima del valor $p= 0.029$ o que sean menores $p= -0.029$ representados por la línea azul (con 10 % de significancia), son los retrasos (en este caso días) que influyen sobre el valor del tipo de cambio presente.

Tabla 5. Los 10 valores más relevantes de la función parcial autorregresiva acf sobre la serie de tiempo de cambio diferenciada

Correlación	Retraso	Valores absolutos
0.2275445	1	0.2275445
0.14395335	28	0.14395335
0.13255882	154	0.13255882
0.13112695	91	0.13112695
-0.1252373	9	0.1252373
0.10821868	92	0.10821868
0.10401933	120	0.10401933
0.10261727	56	0.10261727
0.10071689	63	0.10071689
0.09907301	35	0.09907301

ARCH

Los modelos ARCH son el componente análogo al AR en los modelos ARIMA. Es importante destacar que la serie de tiempo sobre la cual se aplica un proceso ARCH no necesariamente requiere una serie de tiempo estacionaria. Sin embargo, en este caso, para homologarlo con el ARIMA, se decide aplicar una diferenciación. Asimismo, los coeficientes de correlación para los retrasos en este procedimiento son los mismos que los mostrados en la Tabla 5 del modelo ARIMA.

GARCH

El GARCH agrega el componente MA (promedio móvil) a los modelos heterocedásticos, es decir, al error de pronóstico. En este modelo heterocedástico, donde la varianza no es constante, se puede aplicar sobre curvas no estacionarias. Sin embargo, con el mismo argumento que para el componente ARCH, se utilizan como coeficientes de correlación relevantes que se generan por la curva ACF en la Tabla 6.

Resultados de las series de tiempo que se generan

Se aplican los coeficientes de correlación de las curvas ACF y PACF correspondientemente para los modelos ARIMA, ARCH y GARCH. Lo anterior tiene el fin de que se determine la curva con menor AIC como marca distintiva del mejor modelo de cada caso.

Es importante indicar que el AIC no se puede comparar entre modelos, debido principalmente a la forma en la que la función *fgarch* y *Arima* calculan el AIC en R. Una particularidad de los modelos calculados por el *fgarch* es que pueden generar AIC negativos. A pesar de eso, el criterio de elección no cambia, el menor AIC representa el modelo con mejor perspectiva de pronóstico para las series de tiempo.

Tabla 6. Los cinco resultados con menor AIC de los modelos ARIMA probados

Modelo	AIC
ARIMA(92,1,9)	10216.58
ARIMA(28,1,56)	10217.6
ARIMA(92,1,8)	10218.24
ARIMA(91,1,56)	10219.47
ARIMA(56,1,9)	10224.59

Tabla 7. Los cinco resultados con menor AIC de los modelos ARIMA-ARCH probados

Modelo	AIC
ARIMA(28,1,28)~ARCH(91)	-10.9846
ARIMA(1,1,1)~ARCH(9)	-10.0116
ARIMA(28,1,28)~ARCH(1)	-9.7478
ARIMA(1,1,1)~ARCH(2)	-9.59284
ARIMA(1,1,1)~ARCH(1)	-9.57639

Tabla 8. Los cinco resultados con menor AIC de los modelos ARIMA-GARCH probados

Modelo	AIC
ARIMA(1,1,1)~GARCH(91,1)	-11.5055

ARIMA(1,1,1)~GARCH(91,56)	-11.3976
ARIMA(1,1,1)~GARCH(91,8)	-11.2639
ARIMA(28,1,28)~GARCH(28,28)	-11.1942
ARIMA(1,1,1)~GARCH(9,9)	-11.1812

IV. Discusión

Ahora bien, hacer un análisis sobre cada uno de esos modelos de las series de tiempo solo es necesario para los modelos que tengan el mejor AIC. Como se indicó, se utilizan técnicas de *backtesting* sobre los datos de junio de 2023. Esto quiere decir que todos los modelos presentados a continuación se entrenaron desde el 2 de febrero de 2015 hasta el 31 de mayo de 2023. Estos son los modelos elegidos:

- Predicción modelo árbol-Todas las variables, con datos exactos de los parámetros.
- Predicción modelo árbol-Todas las variables, sin datos de los parámetros.
- Predicción modelo bosque-Todas las variables, con datos exactos de los parámetros.
- Predicción modelo bosque-Todas las variables, sin datos de los parámetros.
- Predicción modelo ARIMA(28,1,56).
- Predicción modelo ARIMA(1,1,1)~ARCH(9).
- Predicción modelo ARIMA(1,1,1)~GARCH(91,1).

Una vez entrenados los modelos, se comparan con los datos de junio de 2023. Como métrica de comparación se utiliza el error porcentual absoluto medio (MAPE en inglés) como la métrica de desempeño del pronóstico, que se define de la siguiente manera.

$$\frac{\sum_{t=1}^n \frac{|A_t - F_t|}{|A_t|}}{n}$$

Donde:

A_t : es el valor real de la observación t

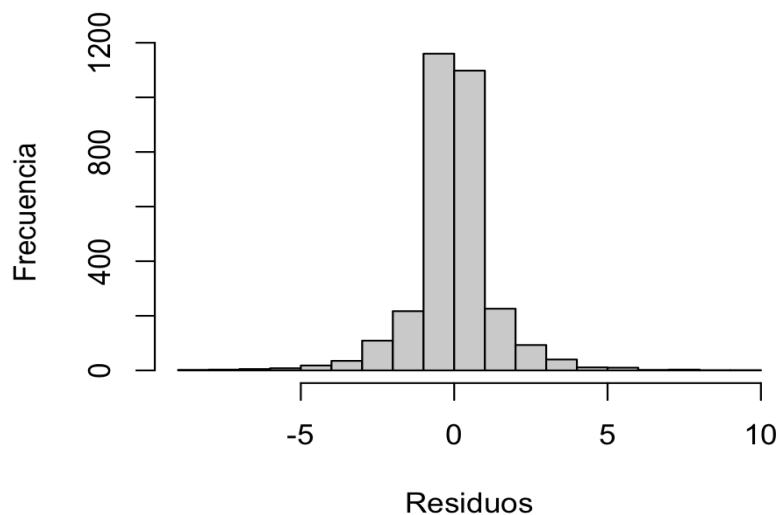
F_t : es el valor pronosticado para la observación t

n: es el número de observaciones.

Cuanto menor sea el MAPE, mayor es la capacidad de pronóstico del modelo en cuestión, ya que el valor real está más cerca del pronosticado, estos resultados son importantes. Los datos de la balanza trimestral del BCCR no estaban publicados, por lo tanto, se supone que permanecen constantes y se toman los del trimestre anterior. Una vez aclarado esto, cuando el pronóstico indica que utiliza datos, es exclusivo para los modelos de aprendizaje supervisado. Después de pasar el entrenamiento, se quiere observar si pueden pronosticar nuevos datos con base en el agrupamiento realizado en el entrenamiento.

Los modelos supervisados en R pueden producir un pronóstico sin necesidad de nuevos datos utilizando la función predict del paquete base de R. Esto se refiere a los pronósticos de los modelos de aprendizaje supervisado sin datos. En el caso de los modelos ARIMA, se puede hacer un pronóstico de los datos usando la función predict para las series de tiempo. Sin embargo, los modelos ARIMA~GARCH y ARIMA~ARCH no generan valores, sino valores de varianzas. Al analizar la distribución de los residuos, se obtiene una distribución normal sesgada, como muestra la siguiente figura.

Figura 5. Distribución de residuos del modelo ARIMA(1,1,1)~ARCH(9)



Se realiza una especificación sobre la serie de tiempo univariada. Esta especificación utiliza una distribución normal sesgada para generar pronósticos de las volatilidades. Además, se lleva a cabo una simulación Montecarlo para producir desviaciones estándares futuras. Estas desviaciones se multiplican con el valor del último tipo de cambio de compra de mayo de 2023. El mismo proceso se realiza para los modelos ARIMA~GARCH, por lo que no es necesario repetirlo. Para garantizar la reproducibilidad, se usa la fórmula set.seed. Lo anterior tiene el fin de asegurar una aleatoriedad fija para esta simulación. Sin embargo, es importante destacar que los modelos GARCH tienen la particularidad de pronosticar valores en un rango aproximado. A continuación, se presentan los resultados de los modelos.

Figura 6. Pronósticos para junio de 2023 utilizando los modelos de árbol

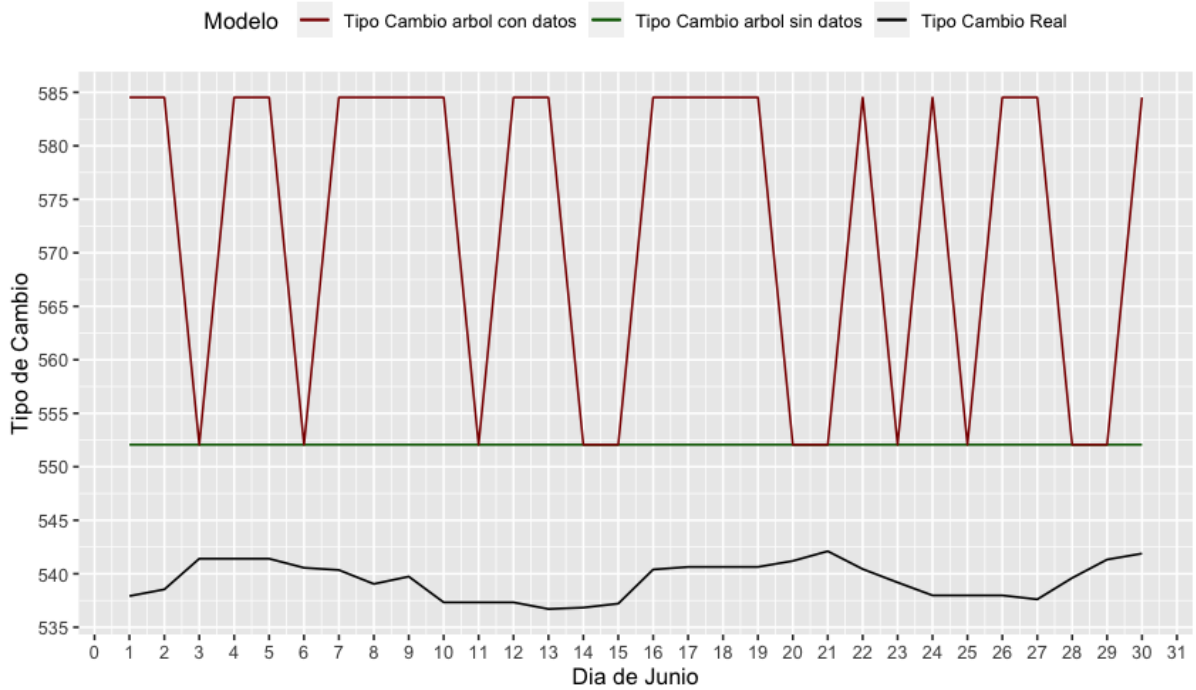


Figura 7. Pronósticos para junio de 2023 utilizando los modelos de bosque

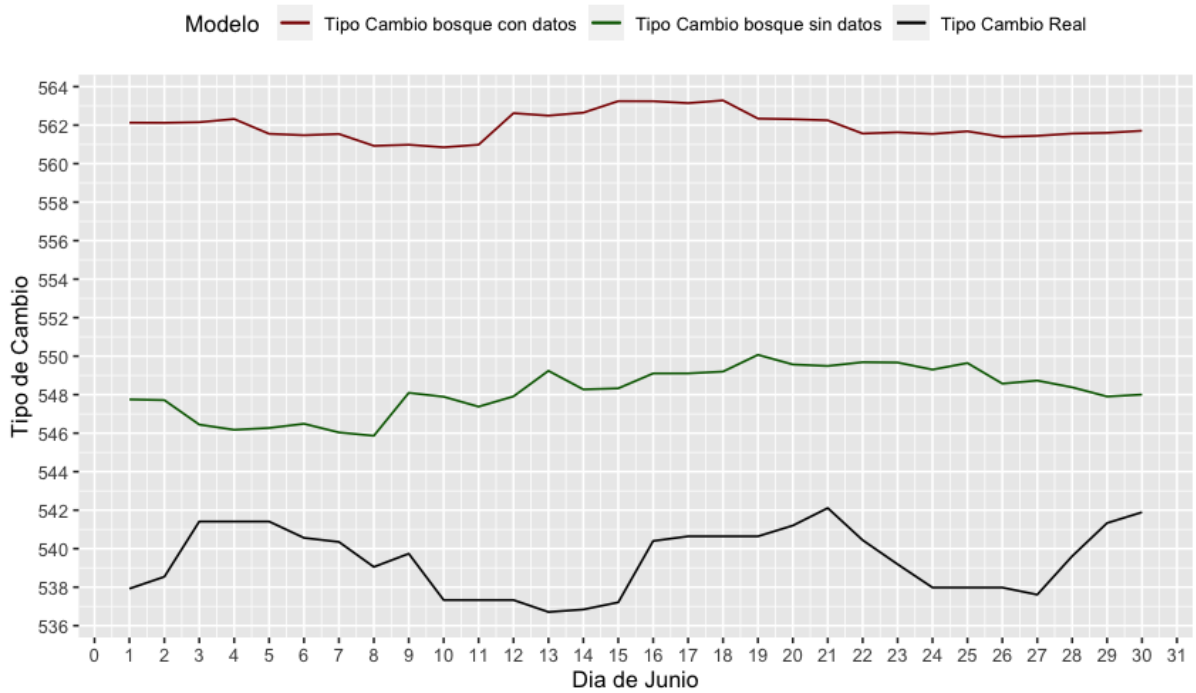


Figura 8. Pronósticos para junio de 2023 utilizando los modelos ARIMA

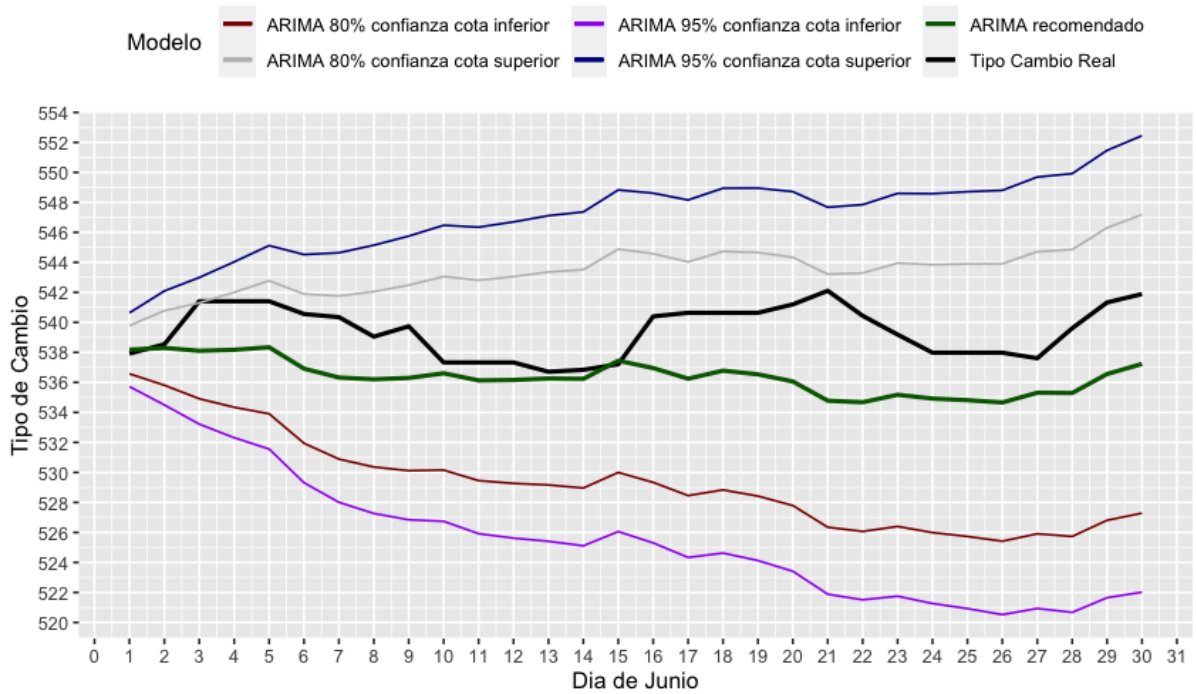


Figura 9. Pronósticos para junio de 2023 utilizando los modelos ARIMA-ARCH y ARIMA-GARCH

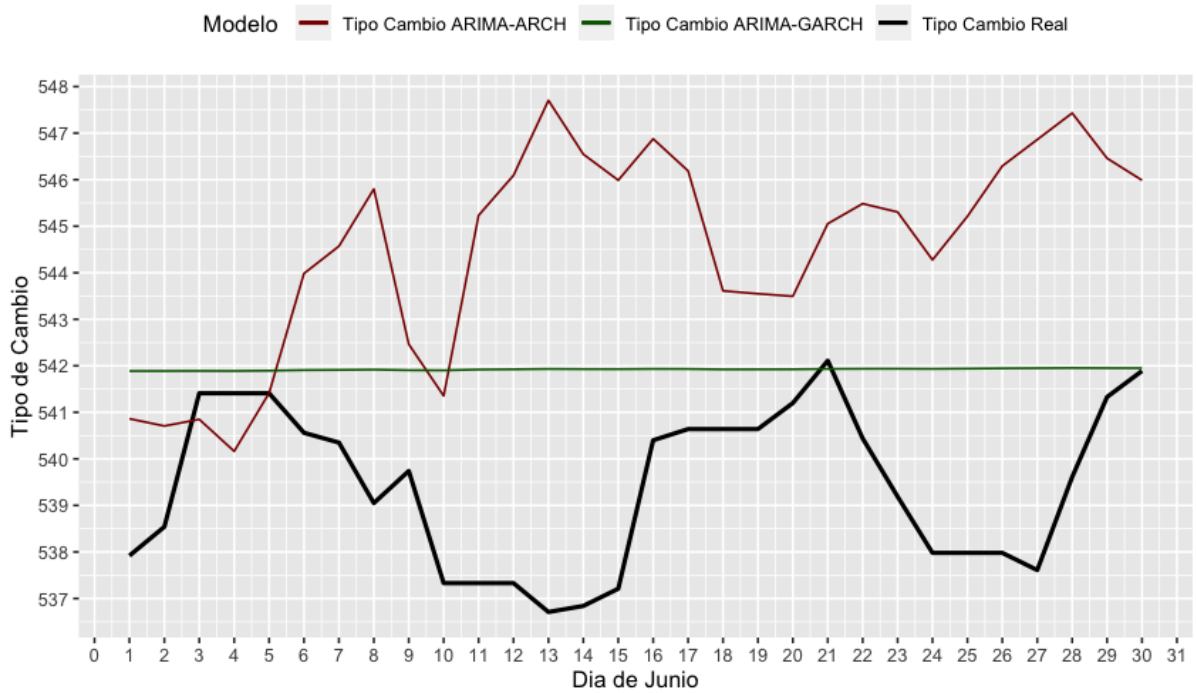


Tabla 9. Error porcentual absoluto medio (MAPE) de los modelos propuestos para el mes de junio de 2023

Nombre modelo	MAPE(%)
Tipo cambio ARIMA(1,1,1)~GARCH(91,1)*	0.47
ARIMA(92,1,9) recomendado	0.57
ARIMA(92,1,9) 80 % confianza cota superior	0.75
Tipo cambio ARIMA(1,1,1)~ARCH(9)*	0.97
ARIMA(92,1,9) 95 % confianza cota superior	1.44
Tipo cambio bosque sin datos	1.63
ARIMA(92,1,9) 80 % confianza cota inferior	1.87
Tipo cambio árbol sin datos	2.34
ARIMA(92,1,9) 95 % confianza cota inferior	2.56
Tipo cambio bosque con datos	4.18
Tipo cambio árbol con datos	6.16

* Utilizando la simulación Montecarlo con aleatoriedad establecida.

La tabla de los datos se encuentra en los Anexos.

V. Conclusiones

Durante la obtención de los datos del MAPE, se puede evidenciar que las series de tiempo son los mejores modelos para pronosticar a corto plazo. Es decir, un plazo de aproximadamente 30 días, con errores porcentuales absolutos rondando números menores de 1 %.

Para el caso del modelo ARIMA(1,1,1)~GARCH(91,1), el MAPE es de 0.47 %, lo que indica que el modelo está cercano al tipo de cambio actual. Es importante tener en cuenta que los modelos GARCH pueden tener una mayor variabilidad y este número puede cambiar considerablemente.

Los modelos de series de tiempo no son infalibles y parecería que pierden fuerza conforme pasa el tiempo. Esto se puede observar en los gráficos de los residuos de las series temporales, donde a partir de los 1000 días se observa mayor variación en los datos de los residuos.

Los métodos de aprendizaje supervisado, aunque muestran un buen rendimiento dentro de las métricas de los conjuntos de entrenamiento y de prueba, no tienen un desempeño tan bueno como las series de tiempo al tomar conjuntos de datos desconocidos. Es decir, no

están en ninguno de los conjuntos mencionados. El principal problema de estos modelos es que, a pesar de ser computacionalmente menos complejos, son más complejos en sí mismos debido a la agrupación de parámetros que requieren.

En la economía suelen ocurrir eventos extremos cada cierto tiempo. Estos eventos son generalmente raros, pero pueden provocar cambios significativos en poco tiempo.

Los modelos de aprendizaje supervisado tienen la ventaja en cuanto al planteamiento de escenarios para las empresas. Con los modelos propuestos en este trabajo, se puede evaluar la posibilidad de que el petróleo tenga cierto valor o que las tasas TRI suban a cierto nivel. Lo anterior tiene el fin de minimizar el riesgo de cambio.

V. Agradecimientos

VI. Referencias

Alvarado, R. (2020). *Efecto multiplicador de la devaluación en el tipo de cambio sobre la estructura de costos de Costa Rica, con un modelo de insumo producto* (Tesis de grado). <https://repositorio.una.ac.cr/bitstream/handle/11056/21145/Tesis%20de%20Licenciatura.pdf?sequence=1&isAllowed=y>

Ayuya, C. (2021). *Entropy and Information Gain to Build Decision Trees in Machine Learning*. <https://www.section.io/engineering-education/entropy-information-gain-machine-learning/#entropy>

Chacón, M. (2019). *Del papel moneda a los billetes de banco en Costa Rica*. <https://museosdelbancocentral.org/wp-content/uploads/2021/03/billetes-del-papel-moneda-a-los-billetes-de-banco.pdf>

Cruz, R.; Rivera, F.; Sánchez, M. y Uribe, A. (2013). *Modelos ARCH y GARCH*. http://personal.cimat.mx:8181/~jortega/MaterialDidactico/ST2013/Modelo_ARCH.pdf

Cutler, A.; Stevens, J. y Cutler, D. (2011). *Random Forests*. https://www.researchgate.net/profile/Adele-Cutler/publication/236952762_Random_Forests/links/0a85e52f509178323a000000/Random-Forests.pdf

Dupouy, C. (2014). *Aplicación de árboles de decisión para la estimación del escenario económico y la estimación de movimiento de la tasa de interés en Chile*. <https://repositorio.uchile.cl/bitstream/handle/2250/117556/Dupouy%20Berrios%20Carlos.pdf;sequence=1>

El Sol. (2017). *Política Cambiaría en Costa Rica y sus Regímenes Cambiarios*. <https://elsoldeoccidente.com/destacadas/politica-cambiaría-en-costa-rica-y-sus-regimenes-cambiarios/>

Fernández, S. (2013). *Series temporales*. <https://www.doccity.com/es/series-temporales/3061108/>

Guzmán, M. (2006). *Un modelo de predicción del tipo de cambio spot para la economía mexicana*. <https://www.redalyc.org/pdf/413/41304707.pdf>

Hernández, S.(2015). *Análisis de Series de Tiempo*.
https://www.cepal.org/sites/default/files/courses/files/01_1_conociendo_una_serie_de_tiempo.pdf

Kevin, A. (2018). *Modelos ARCH Y GARCH: Aplicación a series financieras*.
<https://diposit.ub.edu/dspace/bitstream/2445/125023/2/memoria.pdf>

Kumar, R. (2020). *Time Series Model(s) — ARCH and GARCH*.
<https://medium.com/@ranjithkumar.rocking/time-series-model-s-arch-and-garch-2781a982b448>

León, J. (2012). *Nueva Historia Monetaria De Costa Rica: De La Colonia a la Década de 1930*. <https://revistas.ucr.ac.cr/index.php/dialogos/article/view/13638/13450>

Lizano, E. y López, G(2006). *Régimen Cambiario En Costa Rica*.
https://www.academiaca.or.cr/wp-content/uploads/2017/05/Libroanual_06.pdf

Martínez, I. (2023). *Árboles de Clasificación y Regresión*.
<https://www.uv.es/mlejarza/actuariales/tam/arbolesdecision.pdf>

Morales, S. y Delgado, E. (2015). *Banco Central eliminó la banda cambiaria y adopta flotación administrada del dólar*. <https://www.elfinancierocr.com/finanzas/banco-central-elimino-la-banda-cambiaria-y-adopta-flotacion-administrada-del-dolar/RAMTCQAJ6VBHXLERBSH2EFKUKI/story/>

Rodríguez, A.(2015). *Lo que usted debe saber sobre la flotación administrada del dólar*.
<https://www.elfinancierocr.com/finanzas/lo-que-usted-debe-saber-sobre-la-flotacion-administrada-del-dolar/J2LSOTOZ7FCSPOOYBGFOIGV3KI/story/>

Saborío, G. (1997). *Los Regímenes Cambiarios en Costa Rica: Evolución Histórica y evaluación de algunas alternativas teóricas*.
<https://www.revistas.una.ac.cr/index.php/economia/article/download/1641/1556/>

Anexos

Para el conjunto de datos que entrenará los modelos de árbol de regresión y bosque aleatorio se construye el siguiente conjunto de datos.

Tabla 1. Descripción de los campos del conjunto de datos

Columna	Tipo de dato	Explicación	Formato
Fecha	Date	Esto representa la fecha en la que suceden los eventos, cada fecha es única y se puede considerar la llave primaria para unir los diversos reportes financieros	MMDDYYYY
Trimestre	chr	Esto representa el trimestre al cual la fecha pertenece, en algunos casos los reportes se lanzan de forma trimestral por lo que se usa este elemento como campo en común	T#[1-4]-YYYY
Numero.Mes	int	Este es un número que representa el número de mes al cual pertenece la fecha	#
Numero.Ano	int	Este es un número que representa el número de año al cual pertenece la fecha	#
Fecha.Excel	int	Este es un número entero que representa la fecha en formato legible para Microsoft Excel, con el fin de integrarlo.	#

Tipo.Cambio.Comp ra	num	Este es el valor del tipo de cambio de la compra de dólares establecido por el BCCR en esa fecha	###
Tipo.Cambio.Venta	num	Este es el valor del tipo de cambio de la venta de dólares establecido por el BCCR en esa fecha	###
Dias.Para.Siguien te.Eleccion	int	Este número representa el número de días que faltan para las siguientes elecciones	#
Promedio.Movil.30 D.Compra	num	Este número representa el promedio móvil de 30 días del tipo de cambio de la compra	###
Promedio.Movil.30 DPromVenta	num	Este número representa el promedio móvil de 30 días del tipo de cambio de la venta	###
Tasa.TRI.1semana	num	Esta es la tasa TRI publicada por la cámara de bancos diariamente a una semana	###%
Tasa.TRI.1mes	num	Esta es la tasa TRI publicada por la cámara de bancos diariamente a un mes	###%
Tasa.TRI.3meses	num	Esta es la tasa TRI publicada por la cámara de bancos diariamente a 3 meses	###%
Tasa.TRI.6meses	num	Esta es la tasa TRI publicada por la cámara de bancos diariamente a 6 meses	###%
Tasa.TRI.9meses	num	Esta es la tasa TRI publicada por la cámara de bancos diariamente a 9 meses	###%
Tasa.TRI.12meses	num	Esta es la tasa TRI publicada por la cámara de bancos diariamente a 12 meses	###%
Tasa.TRI.24meses	num	Esta es la tasa TRI publicada por la cámara de bancos diariamente a 24 meses	###%
Tasa.TRI.36meses	chr	Esta es la tasa TRI publicada por la cámara de bancos diariamente a una 36 meses	###%
Tasa.TRI.60meses	chr	Esta es la tasa TRI publicada por la cámara de bancos diariamente a 60 meses	###%
Balanza.Comercial. Trimestral.Millones USD	num	Este el resultado de las exportaciones menos las importaciones, publicados de manera trimestral por el BCCR	###
Exportaciones.Trim estral.MillonesUSD	num	Este el resultado de las exportaciones publicadas de manera trimestral por el BCCR	###
Importaciones.Trim estral.MillonesUSD	num	Este el resultado de las importaciones publicadas de manera trimestral por el BCCR	###
Intervencion.limite.i nferior.MilesUSD	int	Este es el resultado de intervenciones para conservar el límite inferior del tipo de cambio publicado diariamente por el BCCR	###
Intervencion.limite. superior.MilesUSD	int	Este es el resultado de intervenciones para conservar el límite superior del tipo de cambio publicado diariamente por el BCCR	###
Operaciones.estabi lizacion.Compras.M ilesUSD	int	Este es el resultado de intervenciones de compra de divisas para evitar que el tipo de cambio se desestabilice, publicado diariamente por el BCCR	###
Operaciones.estabi lizacion.Ventas.Mil esUSD	int	Este es el resultado de intervenciones de venta de divisas para evitar que el tipo de cambio se desestabilice, publicado diariamente por el BCCR	###
Operaciones.propia s.Compras.MilesU SD	num	Este es el resultado de intervenciones de compra de divisas para necesidades internas del BCCR, publicado diariamente por el BCCR	###

Operaciones.propias.Ventas.MilesUSD	int	Este es el resultado de intervenciones de venta de divisas para necesidades internas del BCCR, publicado diariamente por el BCCR	###
Operaciones.requerimientos.SPNB.BCCR.Compras.MilesUSD	num	Este es el resultado de intervenciones de compras de divisas directas para necesidades internas del SNPB, publicado diariamente por el BCCR	###
Operaciones.requerimientos.SPNB.BCCR.Ventas.MilesUSD	num	Este es el resultado de intervenciones de ventas de divisas directas para necesidades internas del SNPB, publicado diariamente por el BCCR	###
Operaciones.requerimientos.SPNB.MONEX.Compras.MilesUSD	int	Este es el resultado de intervenciones de compras de divisas por medio de Monex para necesidades internas del SNPB, publicado diariamente por el BCCR	###
Operaciones.requerimientos.SPNB.MONEX.Ventas.MilesUSD	int	Este es el resultado de intervenciones de ventas de divisas por medio de Monex para necesidades internas del SNPB, publicado diariamente por el BCCR	###
Reservas.BCCR.MillonesUSD	num	Este es el monto de reservas que tiene disponibles el BCCR, publicado diariamente por el BCCR	###
Precio.WTI	num	Este el precio de cierre del barril de petróleo WTI publicado diariamente por investing.com	###

Tabla 2. Resultados de tipos de cambio pronóstico.

DiaJunio	TipoCambio.Actual	TipoCambio.arbol.sin.datos	TipoCambio.arbol.con.datos	TipoCambio.bosque.sin.datos	TipoCambio.bosque.con.datos	ARIMA.recomendado	ARIMA.80.bajo	ARIMA.80.alto	ARIMA.95.bajo	ARIMA.95.alto	TipoCambio.ARI.MA.ARCH	TipoCambio.ARI.MA.GARCH
1	537.92	552.07	584.52	547.75	562.13	538.18	536.57	539.79	535.72	540.64	540.86	541.89
2	538.54	552.07	584.52	547.71	562.12	538.30	535.81	540.78	534.50	542.09	540.71	541.89
3	541.41	552.07	552.07	546.44	562.15	538.11	534.91	541.30	533.22	542.99	540.85	541.89
4	541.41	552.07	584.52	546.17	562.32	538.17	534.34	542.01	532.31	544.04	540.16	541.89
5	541.41	552.07	584.52	546.27	561.55	538.34	533.91	542.78	531.56	545.12	541.41	541.90
6	540.56	552.07	552.07	546.49	561.48	536.92	531.94	541.89	529.31	544.52	543.98	541.91
7	540.35	552.07	584.52	546.04	561.54	536.32	530.89	541.76	528.01	544.64	544.57	541.91
8	539.05	552.07	584.52	545.86	560.93	536.21	530.36	542.05	527.27	545.15	545.80	541.92
9	539.74	552.07	584.52	548.09	560.98	536.30	530.11	542.48	526.84	545.75	542.46	541.91
10	537.33	552.07	584.52	547.89	560.85	536.61	530.16	543.06	526.74	546.48	541.35	541.90
11	537.33	552.07	552.07	547.38	560.98	536.13	529.45	542.81	525.92	546.35	545.23	541.92
12	537.33	552.07	584.52	547.91	562.62	536.16	529.27	543.05	525.62	546.70	546.09	541.92
13	536.71	552.07	584.52	549.24	562.49	536.26	529.16	543.36	525.41	547.12	547.71	541.93
14	536.84	552.07	552.07	548.27	562.65	536.24	528.96	543.51	525.11	547.36	546.55	541.93
15	537.21	552.07	552.07	548.33	563.25	537.44	530.00	544.89	526.06	548.83	545.99	541.93
16	540.40	552.07	584.52	549.10	563.24	536.96	529.34	544.58	525.30	548.61	546.88	541.93
17	540.64	552.07	584.52	549.10	563.14	536.25	528.46	544.04	524.33	548.16	546.19	541.93
18	540.64	552.07	584.52	549.20	563.28	536.78	528.83	544.73	524.62	548.94	543.61	541.92
19	540.64	552.07	584.52	550.07	562.34	536.54	528.42	544.65	524.13	548.95	543.55	541.92

20	541.20	552.07	552.07	549.57	562.31	536.06	527.79	544.34	523.41	548.72	543.50	541.93
21	542.11	552.07	552.07	549.49	562.25	534.78	526.34	543.21	521.88	547.68	545.06	541.93
22	540.44	552.07	584.52	549.69	561.56	534.68	526.07	543.29	521.51	547.85	545.49	541.94
23	539.19	552.07	552.07	549.67	561.63	535.18	526.40	543.95	521.75	548.60	545.31	541.94
24	537.98	552.07	584.52	549.30	561.54	534.92	525.99	543.85	521.26	548.58	544.28	541.93
25	537.98	552.07	552.07	549.64	561.68	534.82	525.73	543.90	520.92	548.71	545.21	541.94
26	537.98	552.07	584.52	548.57	561.39	534.66	525.42	543.91	520.52	548.80	546.30	541.95
27	537.61	552.07	584.52	548.73	561.45	535.31	525.91	544.71	520.93	549.69	546.86	541.95
28	539.61	552.07	552.07	548.38	561.56	535.29	525.73	544.86	520.67	549.92	547.43	541.95
29	541.33	552.07	552.07	547.90	561.60	536.56	526.81	546.31	521.65	551.47	546.46	541.95
30	541.89	552.07	584.52	548.00	561.71	537.23	527.28	547.18	522.02	552.45	545.99	541.95

Tabla 3. Resultados MAPE simulación I sin función set.seed

Nombre modelo	MAPE(%)
ARIMA recomendado	0.57
ARIMA 80 % confianza cota superior	0.75
ARIMA 95 % confianza cota superior	1.44
Tipo cambio bosque sin datos	1.63
ARIMA 80 % confianza cota inferior	1.87
Tipo cambio árbol sin datos	2.34
ARIMA 95 % confianza cota inferior	2.56
Tipo cambio ARIMA GARCH	2.8
Tipo cambio ARIMA ARCH	3.26
Tipo cambio bosque con datos	4.18
Tipo cambio árbol con datos	6.16

Tabla 4. Resultados MAPE simulación II sin función set.seed

Nombre modelo	MAPE(%)
Tipo cambio ARIMA ARCH	0.46
Tipo cambio ARIMA GARCH	0.47
ARIMA recomendado	0.57
ARIMA 80 % confianza cota superior	0.75
ARIMA 95 % confianza cota superior	1.44
Tipo cambio bosque sin datos	1.63
ARIMA 80 % confianza cota inferior	1.87
Tipo cambio árbol sin datos	2.34
ARIMA 95 % confianza cota inferior	2.56
Tipo cambio bosque con datos	4.18
Tipo cambio árbol con datos	6.16

Carta de aprobación del filólogo

Cartago, 02 de febrero de 2024

Los suscritos, Elena Redondo Camacho, mayor, casada, filóloga, incorporada a la Asociación Costarricense de Filólogos con el número de carné 0247, portadora de la cédula de identidad número 3-0447-0799 y, Daniel González Monge, mayor, casado, filólogo, incorporado a la Asociación Costarricense de Filólogos con el número de carné 0245, portador de la cédula de identidad número 1-1345-0416, ambos vecinos de Quebradilla de Cartago, revisamos el artículo que se titula: *Desarrollo de un modelo avanzado para predecir las fluctuaciones del tipo de cambio entre el colón costarricense y el dólar estadounidense mediante técnicas de aprendizaje automático y análisis de series temporales*, sustentado por Felipe Corrales Karpukhina.

Hacemos constar que se corrigieron aspectos de ortografía, redacción, estilo y otros vicios del lenguaje que se pudieron trasladar al texto. A pesar de esto, la originalidad y la validez del contenido son responsabilidad directa de la persona autora.

Esperamos que la participación de Filólogos Bórea Costa Rica satisfaga los requerimientos de la universidad.

X

Elena Redondo Camacho
Filóloga - Carné ACFIL n.º 0247

X

Daniel González Monge
Filólogo - Carné ACFIL n.º 0245