

# Aplicación de modelos unifactoriales y aprendizaje automático en el mercado colombiano de deuda pública

**Application of unifactorial models and machine learning for the colombian fixed income market**

Paula Catalina Alejo\*  
David Gómez Aldana\*\*  
Bernardo León\*\*\*

---

\* Magíster en Finanzas, Universidad Externado de Colombia. Jefe de Producto, Banco de Bogotá, Bogotá (Colombia). [paulacataap@hotmail.com]

\*\* Magíster en Finanzas, Universidad Externado de Colombia. Trader de Liquidez, MiBanco, Bogotá (Colombia). [davidgomez24.dga@gmail.com]

\*\*\* Ph.D. en Economía, Universidad de Salamanca, Salamanca (España). Profesor Adjunto Universidad Javeriana. [uniandinos2007@gmail.com]; [https://orcid.org/0009-0002-7054-2019].

Artículo recibido: 20 de julio de 2024.

Aceptado: 18 de octubre de 2024.

Para citar este artículo:

Alejo, P. C. *et al.* (2024). Aplicación de modelos unifactoriales y aprendizaje de automático en el mercado colombiano de deuda pública. *Odeon*, 27, 255-298.

DOI: <https://doi.org/10.18601/17941113.n27.08>

## Resumen

En este trabajo se implementa el modelo propuesto por Vasicek y el de Cox *et al.*, en ventanas móviles de tiempo, y observaremos su desempeño de pronóstico en el lapso de un año para el mercado colombiano de deuda pública.

Los resultados se van a comparar con modelos de aprendizaje de máquinas, específicamente con modelos de árboles de regresión, redes neuronales y bosques aleatorios (*random forest*); a pesar de no tener un sustento económico, estos modelos se usan ampliamente para manejar problemas de pronóstico en diferentes áreas del conocimiento y, al igual que los modelos de tasas unifactoriales, se pueden utilizar para determinar los niveles futuros de tasas de interés, abordándolos como un problema de regresión.

**Palabras clave:** renta fija; deuda pública colombiana; pronóstico; modelos unifactoriales; tasas de corto plazo; aprendizaje de máquinas; curva de rendimientos.

**Códigos JEL:** G10, G17, G19

## Abstract

We will implement the model proposed by Vasicek (1977) and Cox *et al.* (1985), in moving time windows and we will observe its forecasting performance in the tenor of 1 year for the Colombian public debt market.

The results will be compared against machine learning models, specifically against Regression Trees, Neural Networks and Random Forest models, these models despite not having an economic support are widely used to handle forecasting problems in different areas of knowledge and like the unifactorial rate models, can be used to determine the future levels of interest rates, approaching it as a regression problem.

**Key words:** Fixed income; sovereign colombian debt; forecast; spot rates; unifactorial models; machine learning; yield curve.

**JEL classification:** G10, G17, G19

## Introducción

El mercado de renta fija colombiano es de los más robustos y con mayor operatividad. De acuerdo con la Bolsa de Valores de Colombia, para el año 2022 el 85 % de las operaciones realizadas fueron transacciones del mercado de renta fija (Bolsa de Valores de Colombia, 2023); adicionalmente, el programa de creadores de mercado que fomenta el Gobierno colombiano y administra el

Banco de la República, garantiza el correcto funcionamiento del mercado de deuda pública realizando subastas para el funcionamiento del gobierno, y la correcta formación de precios.

El pronóstico de tasas y de curvas de rendimientos cobra vital importancia porque los tenedores de deuda son distintos actores del mercado, como fondos de pensión, bancos comerciales, fiduciarias, entre otros.

En la literatura académica colombiana no se ha ahondado en el pronóstico de las tasas de interés. Si bien Diebold y Li (2006) propusieron un modelo de vectores autorregresivos para el pronóstico de los factores latentes de la curva de rendimientos de Estados Unidos, en Colombia, según Melo y Castro (2010), el Banco de la República aplica este modelo, pero para determinar el impacto de las variables macroeconómicas en la formación de la curva de rendimientos, por lo tanto, no se enfoca en el pronóstico de las tasas, ni tampoco en la curva de rendimientos, sino en los determinantes que afectan los parámetros latentes de estas (nivel, pendiente y curvatura). El trabajo concluye que existe una relación bidireccional entre la formación de las curvas y las variables macroeconómicas estudiadas.

La gran generalidad de estudios realizados en los últimos años se ha enfocado en seguir el camino propuesto por Diebold y Li (2006), el cual consiste en tener una estructura funcional de la curva de rendimientos –en la mayoría de los casos la propuesta por Nelson y Siegel (1987)– y se enfocan en pronosticar los factores latentes y después obtener una curva completa de rendimientos, para poder comparar entre la curva pronosticada y la curva realizada.

Por ejemplo, en el trabajo de Andreasen *et al.* (2019), amplían la base de datos utilizada, en este caso usan bonos transados en el mercado canadiense, para estimar la curva cero cupón, pero siguen utilizando la metodología de ajuste de Nelson y Siegel y su objetivo es estimar la curva desde los factores latentes como lo hicieron Diebold y Li (2006). La conclusión es que estimar la estructura a término directamente extraída de los precios de títulos observables en el mercado (bonos) y no con un enfoque de tasas teóricas cero cupón sintéticas, no conlleva un mejor ajuste en el pronóstico de las tasas de la curva de rendimientos.

En su trabajo, Luo *et al.* (2021) utilizan el mismo marco metodológico de Diebold y Li (2006), pero la forma en que pronostican el factor de la pendiente incluye saltos paralelos. Un trabajo mucho más similar a este artículo fue el propuesto por Matsumura *et al.* (2011), ellos no solamente trabajan con la curva de rendimientos del tesoro norteamericano, sino que además lo hacen con la

curva de rendimientos de los bonos emitidos por el Gobierno central de Brasil; el estudio resalta el bajo rendimiento que tienen los modelos actuales cuando se prueban fuera de muestra, sin embargo, ellos utilizan modelos lineales para modelar y pronosticar la curva de rendimientos. La conclusión presentada en el trabajo resalta que los modelos tradicionales tienen un bajo poder predictivo y que, por lo general, no superan de manera consistente un modelo de caminata aleatoria; adicionalmente, indican que incluir variables macroeconómicas y las condiciones de no arbitraje tampoco mejora el desempeño de los modelos fuera de muestra, pero que adicionar variables de mercado permite capturar mejor las dinámicas de la curva de rendimientos. Finalmente, al comparar los pronósticos de la curva de rendimientos de Brasil con respecto a los Estados Unidos, los errores de pronóstico de la curva brasileña son mucho mayores a los de la curva norteamericana.

A nivel local, un trabajo similar lo realizan Herrera Cardona *et al.* (2011), en el cual buscan aplicar el modelo propuesto por Vasicek (1977) en un título de referencia específica: los títulos de tasa fija con vencimiento el 24 de julio de 2020 (Nemo: TFIT15240720) con una base de datos del 5 de agosto de 2005 al 5 de noviembre de 2010, con él estiman las propiedades estadísticas del comportamiento de la tasa de interés, realizan un proceso econométrico para el cálculo de parámetros y modelan la volatilidad. Una vez realizado este proceso, proceden a calcular la tasa corta y a modelar la curva de rendimientos. El resultado final de la investigación confirma que existe un pobre desempeño del modelo con respecto a la curva de negociación al momento de comparar el resultado. Ahora bien, como un primer punto de mejora del ejercicio es tomar un título de referencia específica en vez de un tenor puntal dentro de la curva –es decir, al momento del inicio del análisis a los TES de 2020 les quedaban alrededor de 15 años para el vencimiento, y al finalizar el análisis faltaban alrededor de 10 años–, la dinámica de negociación de un título de 10 años difiere ampliamente con la de un título de 15 años, lo que lleva al siguiente punto débil del análisis: la muestra tomada de 5 años contiene diferentes ciclos económicos y, si dicha muestra se toma sin realizar segmentaciones, puede empobrecer el desempeño de los pronósticos fuera de muestra.

Un trabajo que vale la pena mencionar es el realizado por Restrepo Tobón y Botero Ramírez (2007), en el cual se proponen modelos unifactoriales, el de Hull y White (1990) y el de Black y Karasinski (1991), para la valoración de derivados de tasa de interés y se presenta una metodología estándar a pesar de la falta de información del mercado. Para la calibración de los modelos utilizan

métodos econométricos para la estimación de la varianza a través de modelos Egarch sobre la tasa interbancaria. En el trabajo se presenta la implementación de los modelos unifactoriales, el cálculo de la curva de rendimientos y la posterior comparación de dicha curva calculada por la BVC, para llegar a la valoración de derivados sobre productos de tasas de interés.

La conclusión es que estos modelos replican con exactitud la estructura a término presentada por la BVC, lo que abre la puerta para aplicar estos modelos a la valoración de derivados que no se negocian de manera activa en el mercado colombiano. Sin embargo, hace énfasis que la aplicación de modelos unifactoriales con especificaciones más simples, como la velocidad de reversión a la media y la volatilidad de tasas de interés, que se calculan de forma inequívoca y señalan que se necesitaba información adicional que, debido al desarrollo que presentaba el mercado en ese entonces, no tenían disponible. El trabajo es una excelente respuesta al problema de valoración de derivados de tasas de interés y un paso en la dirección correcta para la implementación y profundización de este, sin embargo, no es acorde con el problema que se aborda en este artículo, que se enfoca en el pronóstico de las tasas y no en la propuesta de modelos de valoración y su implementación.

En su artículo, Castello y Resta (2022) compararon dos métodos alternativos: el modelo paramétrico dinámico de cinco factores (5F-DRF) de Rezende y Ferreira (2013) y la red neuronal Feed-Forward (FFNN), para modelar la curva de rendimiento de los países BRICS (Brasil, Rusia, India, China, Sudáfrica), la evidencia demostró la capacidad superior de FFNN para interpolar el comportamiento del rendimiento de la curva, ya que esta se adaptó perfectamente a todas sus formas típicas, incluso a las más retorcidas, con múltiples puntos de inflexión como en el caso de India y Rusia. Esto supera las limitaciones comunes de los modelos paramétricos en presencia de auges y caída. Por otro lado, la mayor capacidad de predicción dentro de la muestra fue confirmada por el MSE y RMSE asociados al ajuste, y el proceso de estimación tomó menos tiempo que en el 5F-DRF, dada la menor cantidad de parámetros requeridos durante la calibración de la red. Sin embargo, una limitación de este enfoque es el hecho de que el ajuste fue dentro de la muestra, por lo que los autores señalan que una posible extensión de su trabajo podría ser si la comparación del desempeño de los modelos fuera de muestra probaron si el modelado basado en el enfoque paramétrico o de aprendizaje automático puede garantizar un ajuste simplificado en la muestra también para estos nuevos actores mundiales. Para ello, después de los estudios presentados, incluso los más recientes exponen

modelos estadísticos clásicos para el pronóstico y se enfocan en la proyección de toda la curva de rendimientos, lo cual abre el espacio para la aplicación de modelos de aprendizaje de máquinas y proyección de tenores puntuales de la curva de rendimientos.

Dado lo anterior, este trabajo presenta una propuesta para analizar la capacidad de pronóstico de modelos de tasas unifactoriales, Vasicek (1977) y Cox, Ingersoll y Ross (CIR) (1985), y los compara con los modelos de aprendizaje de máquina, que detallaremos a continuación.

## 1. Base teórica y modelo propuesto

Implementaremos los modelos propuestos por Vasicek (1977) y Cox *et al.* (1985), en ventanas móviles de tiempo, y observaremos su desempeño en el pronóstico dentro y fuera de muestra para el mercado de Deuda Pública colombiana para el lapso de un año.

Los resultados se compararán contra modelos de aprendizaje de máquinas, específicamente contra modelos de árboles de regresión, redes neuronales y bosques aleatorios (Random Forest). Estos modelos, a pesar de no tener un sustento económico, se usan ampliamente para manejar problemas de pronóstico en diferentes áreas del conocimiento y, al igual que los modelos de tasas unifactoriales, se pueden utilizar para determinar los niveles futuros de tasas de interés, abordándolos como un problema de regresión.

A continuación, presentamos las hipótesis por demostrar:

1. A pesar de que los modelos de tasas unifactoriales de Vasicek (1977) y CIR (1985), tienen diferentes supuestos, consideramos que su aplicación y capacidad de pronóstico dentro y fuera de muestra no debe ser significativamente diferente.
2. Los modelos de aprendizaje de máquina tienen un mejor desempeño del pronóstico, medido a través de estadísticos como el MSE, RMSE y el MAPE, tanto dentro como fuera de muestra versus los modelos de Vasicek (1977) y CIR (1985), lo que los convierte en una herramienta útil para el pronóstico de tasas.
3. Dentro de los modelos de aprendizaje de máquinas consideramos que el de bosques aleatorios es el que mejor capacidad de pronóstico fuera de muestra tiene, esto debido a que es un modelo de ensamble y por sus características tiene mayor capacidad de entrenamiento.

Se va a implementar parte del trabajo desarrollado en 2019 por Giuseppe Orlando, Rosa María Mininni y Michele Bufalo, presentado en sus artículos titulados: “Forecasting interest rates through Vasicek and CIR models: A partitioning approach” (Orlando *et al.*, 2019) ; “A new approach to forecast market interest rates through the CIR model” (Orlando *et al.*, 2019) e “Interest rates calibration with a CIR model” (Orlando *et al.*, 2019) donde los autores sugieren un modelo para pronosticar las tasas de interés a partir del modelo CIR mediante la partición de la muestra de datos, esto con el objetivo de capturar los cambios de la varianza estadísticamente significativos en las tasas de interés y posterior a ello, calibrar los parámetros reemplazando los residuos estandarizados normalmente distribuidos del modelo ARIMA “óptimo” elegido para cada subgrupo, en la parte aleatoria del proceso de movimiento browniano estándar (Orlando *et al.*, 2019).

En esta investigación se utiliza la base de datos del Banco de la República, donde se publica la curva cero cupón de TES tasa fija en pesos. Acá se presentan observaciones diarias de la curva, ajustadas tras las negociaciones realizadas para ese día, y con esto se obtiene la tasa *spot* para un determinado plazo, también conocida como un tenor de la curva.

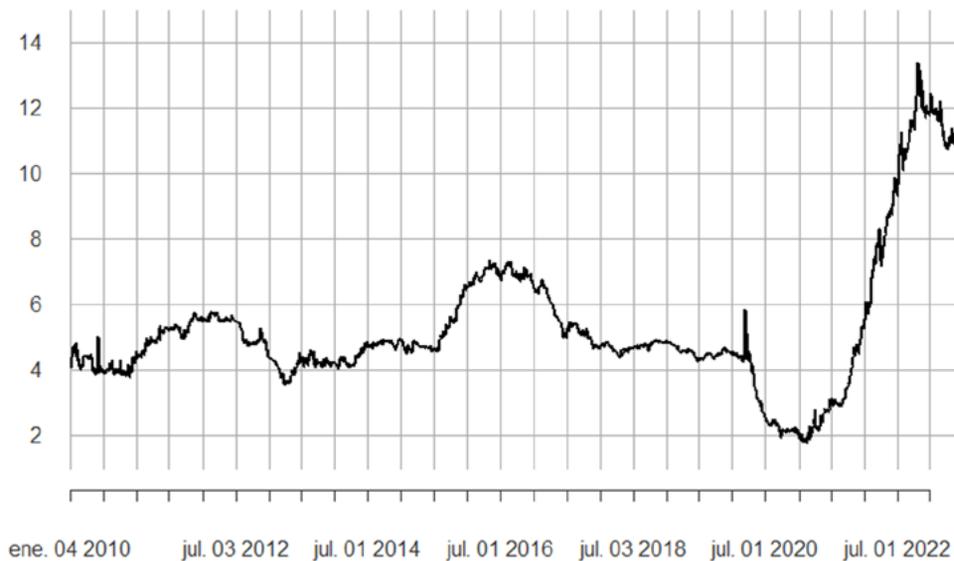
Ahora bien, para este trabajo se utilizará la tasa de TES del tenor de un año para el periodo comprendido entre el 4 enero de 2010 al 31 de diciembre de 2022, lo que da como resultado una base de datos con 3.171 observaciones en frecuencia diaria (solo días hábiles de negociación). Con esto se pretende analizar la capacidad de pronóstico que tienen los modelos tradicionales (Vasicek y CIR) y no tradicionales (árboles, bosques y redes), ajustando los modelos dentro de muestra y probando sus pronósticos fuera de esta, a través de ventanas móviles que avanzan de forma trimestral, esto dado que se realizará la partición de la base de datos en trimestres, lo que da como resultado un total de 53 particiones.

Con el objetivo de que la proyección de la tasa *spot* se ajuste a la realidad de las emisiones disponibles en el mercado, se escogió el nodo de 1 año, como se presenta en la figura 1; adicionalmente, en la figura 2 se muestra la dinámica de toda la curva de TES Tasa fija en el periodo mencionado, esto con el fin de distinguir que las tasas en cada uno de los nodos presentados por el Banco de la República pueden tener diferentes movimientos.

Para entender mejor los modelos que se utilizaron en el artículo, a continuación realizamos una revisión de la literatura relevante.

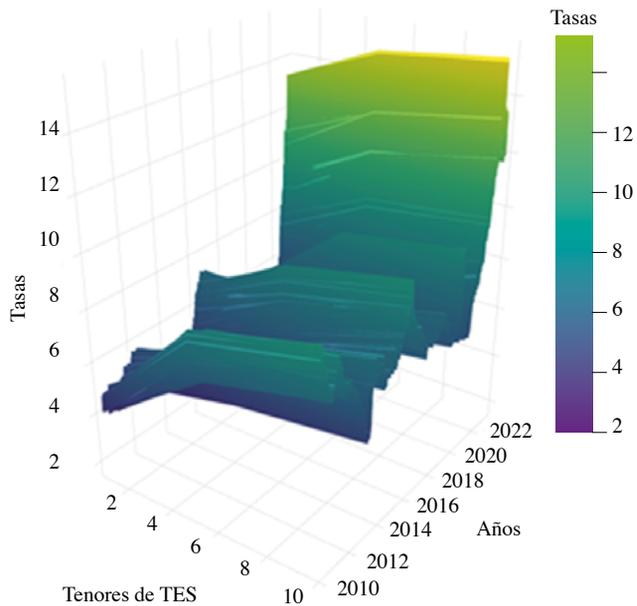
Figura 1. Tasas de TES nodo de 1 año

2010-01-04 / 2023-05-12



Fuente: Banco de la República de Colombia.

Figura 2. Dinámica de la curva de TES tasa fija (2010-2022)



Fuente: Banco de la República de Colombia.

### 1.1. Clasificación de modelos de tasa corta de interés

La literatura clasifica los modelos de tasa corta como de equilibrio y de no arbitraje. Los primeros tienen una serie de supuestos respecto a la economía en la cual operan, allí la estructura a plazos y las volatilidades son determinadas de forma endógena, los segundos suponen que el mercado es perfecto, la curva de rendimientos solo se mueve hacia arriba o hacia abajo, la volatilidad de un periodo es la misma en todos los estados del modelo y, como su nombre lo indica, no existen oportunidades de arbitraje (Josheski y Apostolov, 2021), estos últimos son los preferidos para determinar el valor de los derivados, ya que los precios de las opciones se obtienen replicando los pagos proporcionados por la opción, utilizando el activo subyacente y los préstamos libres de riesgo. Un ejemplo de esto es la paridad *put-call* que permite hallar los precios exactos de mercado para los bonos, en cambio, los modelos de equilibrio no fijarán un precio exacto.

Ahora bien, ambos modelos se pueden distinguir según el número de factores aleatorios (variables de estado que contienen toda la información disponible) o fuentes de riesgo o incertidumbre que se contemplan para determinar la estructura a plazo; por un lado están los modelos unifactoriales, los cuales asumen que la curva se desprende de un único factor (tasa de interés de corto plazo, normalmente), allí se incluyen los de Vasicek (1977), Cox *et al.* (1985), Ho y Lee (1986), Hull y White (1990), Black y Karasinski (1991), entre otros; y por otro están los modelos multifactoriales como el modelo de Brennan y Schwartz (1979), con la tasa de corto plazo y la de largo plazo como los dos factores, y el modelo de tres factores: nivel de la tasa de interés, pendiente de la curva de rendimiento y curvatura (asociada a la volatilidad) desarrollado por Litterman y Scheinkman (1991); el modelo de Longstaff y Schwartz (1992), quienes interpretan los dos factores como la tasa de interés a corto plazo y volatilidad de la tasa, entre otros. En los modelos multifactoriales, la variación de las tasas de largo plazo se explica por factores que experimentan una reversión media muy lenta.

Los modelos también se pueden clasificar según la naturaleza del proceso; reversión a la media vs. caminata aleatoria o procesos normales, log-normales vs. de raíz cuadrada (Venegas y Vasicek, 2021).

A continuación, se presentan algunos modelos que nacen (en su mayoría) a partir de la propuesta de Vasicek.

Tabla 1. Resumen de Modelos de Tasa de Interés

No. factores	Modelo	Ecuación	Distribución	Principal característica
Un factor	Vasicek (1977)	$\Delta r_t = \alpha(\beta - r_t)\Delta t + \sigma\Delta Wt$	Normal	Supone parámetros constantes $\alpha$ , $\beta$ , $\sigma$ , los cuales son estimados en función de las observaciones históricas, lo que limita su capacidad de predicción.
	CIR (1985)	$\Delta r_t = \alpha(\beta - r_t)\Delta t + \sqrt{r_t}\sigma\Delta Wt$	Chi cuadrada	La volatilidad aumenta con la raíz cuadrada de la tasa de interés, lo que ayuda a evitar que el modelo genere tasas de interés negativas.
	Ho-Lee (1986),	$\Delta r_t = \theta_t \Delta t + \sigma\Delta Wt$	Normal	Supone una volatilidad constante en el tiempo $\sigma$ , lo que simplifica su modelado.
	Hull y White (1990)	$\Delta r_t = (\theta_t - \alpha r_t)\Delta t + \sigma\Delta Wt$	Normal	Supone parámetros de volatilidad dependientes del tiempo $\alpha_t$ y $\sigma_t$ lo que permite ajustarse a los precios observados del mercado.
	Black-Derman-Toy (1990)	$\Delta \ln \ln r_t = (\theta_t + \ln \ln r_t) \Delta t + \sigma\Delta Wt$	Lognormal	Supone que las tasas de interés siguen una distribución lognormal, lo que implica que no pueden ser negativas.
	Black-Karasinski (1991)	$\Delta r_t = (\theta_t - \alpha r_t)\Delta t + \sigma r_t \Delta Wt$	Normal	Supone también un nivel de reversión a la media variable con el tiempo $\theta_t$ , pero a diferencia del modelo BDT, este permite especificar la forma y el comportamiento de la reversión, a través de estructuras más complejas y personalizables.

Fuente: elaboración propia.

## 2. Aplicación y resultados

Se pretende abordar y comparar los modelos de Vasicek (1977) y CIR (1985) realizando la división de la muestra de datos en subgrupos y, finalmente, analizar los resultados obtenidos de estos modelos tradicionalmente utilizados, frente a los de *machine learning* tales como: árboles de regresión, redes neuronales y bosques aleatorios.

Por lo tanto, se busca realizar ajustes dentro y fuera de muestra en diferentes momentos del tiempo, se propone realizar un bucle en donde se ajustan los parámetros para el caso de los modelos unifactoriales, y en el caso de los

modelos de *machine learning* se busca aprovechar los datos de manera tal que los datos de entrenamiento aumentan conforme el bucle avanza, como se presenta en la figura 3, y en las figuras 4 y 5 se muestra cómo en cada bucle de ajuste se cambian los datos de entrenamiento en el caso de los modelos tradicionales y los modelos de aprendizaje de máquinas.

Figura 3. Ejecución general del código para los modelos unifactoriales y de *machine learning*

Ajuster del modelo para Modelos de la Tasa Unifactoriales (Vasicek y CIR)

Primera ejecución del bucle:

Datos: 1Q-2010  
Se ajustan los parámetros  
-  $\beta$  (Tasa de equilibrio)  
-  $\alpha$  (Reversión a la media)

Datos: 2Q-2010  
Con los parámetros calculados ( $\beta$  y  $\alpha$ ), se realiza el pronóstico fuera de muestra y se calculan estadísticas de desempeño (MSE, RSME, MAE)

Segunda ejecución del bucle:

Datos: 2Q-2010  
Se rebalancean los datos y se ajustan los parámetros  
-  $\beta$  (Tasa de equilibrio) y  $\alpha$  (Reversión a la media)

Datos: 3Q-2010  
Con los parámetros calibrados ( $\beta$  y  $\alpha$ ), se realiza el pronóstico fuera de muestra y se calculan estadísticas de desempeño (MSE, RSME, MAE)

Ajuste del modelo para Modelos *Machine Learning*

Primera ejecución del bucle:

Datos: 1Q-2010  
Se ajustan los hiperparámetros del modelo a utilizar: Árboles binomiales, bosques aleatorios y redes neuronales

Datos: 3Q-2010  
Con la optimización de hiperparámetros con los datos de entrenamiento se proyecta y se calculan estadísticas de desempeño (MSE, RSME, MAE)

Segunda ejecución del bucle:

Datos: De 1Q-2010 hasta 2Q-2010  
Se reajustan los hiperparámetros del modelo a utilizar: Árboles binomiales, bosques aleatorios, y redes neuronales

Datos: 3Q-2010  
Una vez recalibrados los hiperparámetros con los datos de entrenamiento se proyecta y se calculan estadísticas de desempeño (MSE, RSME, MAE)

Fuente: elaboración propia.

Figura 4. Ejecución del bucle y descripción de los datos utilizados en los modelos de Vasicek y CIR



Fuente: elaboración propia.

Figura 5. Ejecución del bucle y descripción de los datos utilizados en los modelos de aprendizaje de máquina (ventana expandible)



Fuente: elaboración propia.

## 2.1 Ajuste para el modelo de Vasicek

En el trabajo "An equilibrium characterization of the term structure" (Vasicek, 1977) se establece que la dinámica de las tasas *spot*, aparte de tener un componente estocástico, presenta un componente importante de reversión a la media. Como se presentó en el marco teórico, el comportamiento de esta tasa tiene reversión a la media a través de un proceso de Ornstein-Uhlenbeck. El modelo de Vasicek describe la tasa *spot* como una ecuación diferencial estocástica, como se describe a continuación:

$$dr_t = a(b - r_t)dt + \sigma * dWt \quad (1)$$

Ahora bien, la ecuación anterior se presenta en forma continua, si se reescribe la expresión en términos discretos se obtiene la siguiente forma de la ecuación (Venegas, 2002), para poder realizar la aplicación en una regresión lineal asumiendo que los errores son no correlacionados y siguen una distribución normal estándar  $\varepsilon \sim (0, \sigma^2)$ :

$$\Delta r_t = \alpha(\beta - r_{t-1})\Delta t + \varepsilon_t(0, \sigma^2) \quad (2)$$

Para obtener los parámetros  $\alpha$  (velocidad de reversión a la media) y  $\beta$  (tasa de equilibrio de largo plazo), se puede reescribir la expresión en términos discretos ecuación (2) y, a través de una regresión lineal con la optimización de mínimos cuadrados ordinarios (MCO), obtendríamos los coeficientes; adicionalmente, el cambio en la tasa  $\Delta r_t$  al pasar a ser un proceso discreto se reescribe  $r_t - r_{t-1}$ :

$$r_t - r_{t-1} = \alpha \beta - \alpha r_{t-1} + \varepsilon_t(0, \sigma^2) \quad (3)$$

Al escribir la ecuación de la siguiente manera obtenemos que para los parámetros debemos realizar los siguientes despejes una vez se realiza la regresión lineal:

$$\alpha = -\alpha \quad \beta = \frac{\alpha\beta}{\alpha}$$

Basado en las ecuaciones obtenidas y teniendo en cuenta que este es un proceso que revierte a su media, sabemos que el valor esperado del proceso se describe de la siguiente manera (Venegas Martinez, 2002):

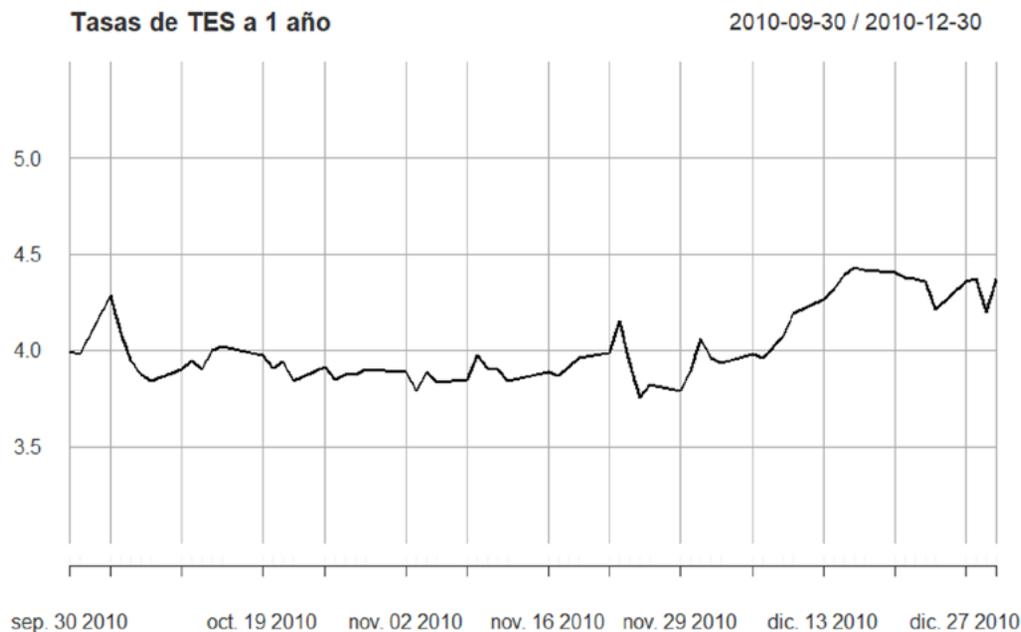
$$E[r_{t+n}|r_t] = r_t e^{-\alpha*n} + \beta * (1 - e^{-\alpha*n}) \quad (4)$$

### 2.1.1. Implementación del modelo de Vasicek (1977) dentro del bucle

Para la implementación del Vasicek dentro del Bucle, se realiza el cálculo de los coeficientes por cada vez que se realiza un recorrido; a continuación, se presenta un ejemplo suponiendo que esta es la cuarta vez que se recorre el bucle:

- i. Se tienen las siguientes tasas de TES a 1 año comprendidas entre el 30 de septiembre de 2010 al 30 de diciembre de 2010, es decir, el cuarto trimestre de 2010. Estos, al ser la cuarta vez que recorren el bucle, serían los nuevos datos de muestra, reemplazando los datos de muestra anterior. En la figura 6 se presentan las tasas de TES a 1 año, del periodo en cuestión.

Figura 6. Tasas de TES a 1 año y datos de entrenamiento—Modelos unifactoriales



Fuente: Banco de la República.

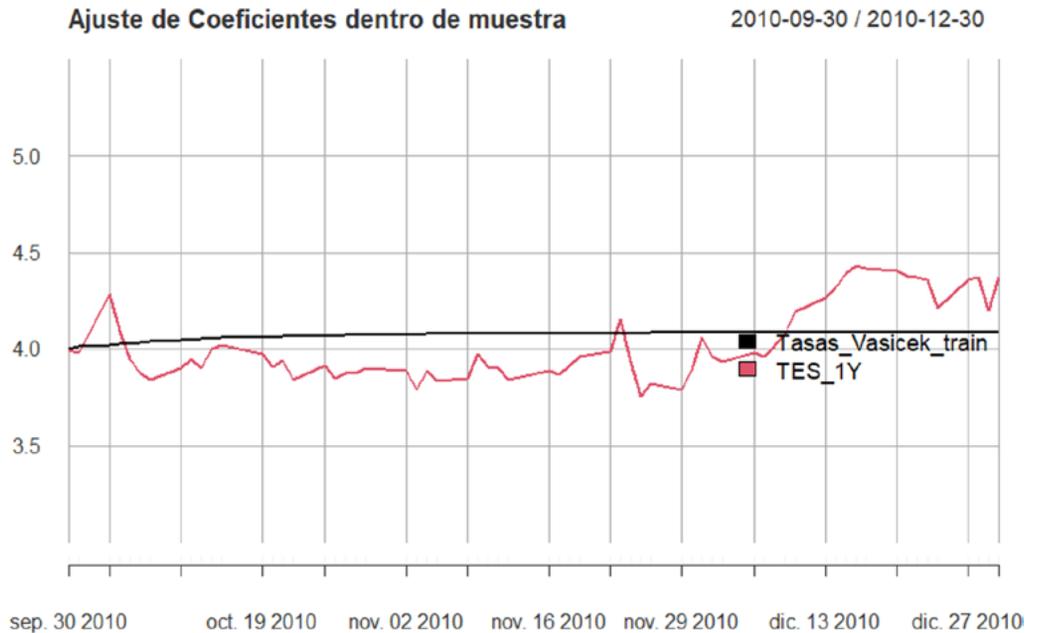
- ii. Ahora, con estos datos que se consideran de entrenamiento, se procede a calcular los coeficientes  $\alpha$  (velocidad de convergencia),  $\beta$  (tasa de equilibrio) y  $\sigma$  (volatilidad del modelo) (tabla 2).

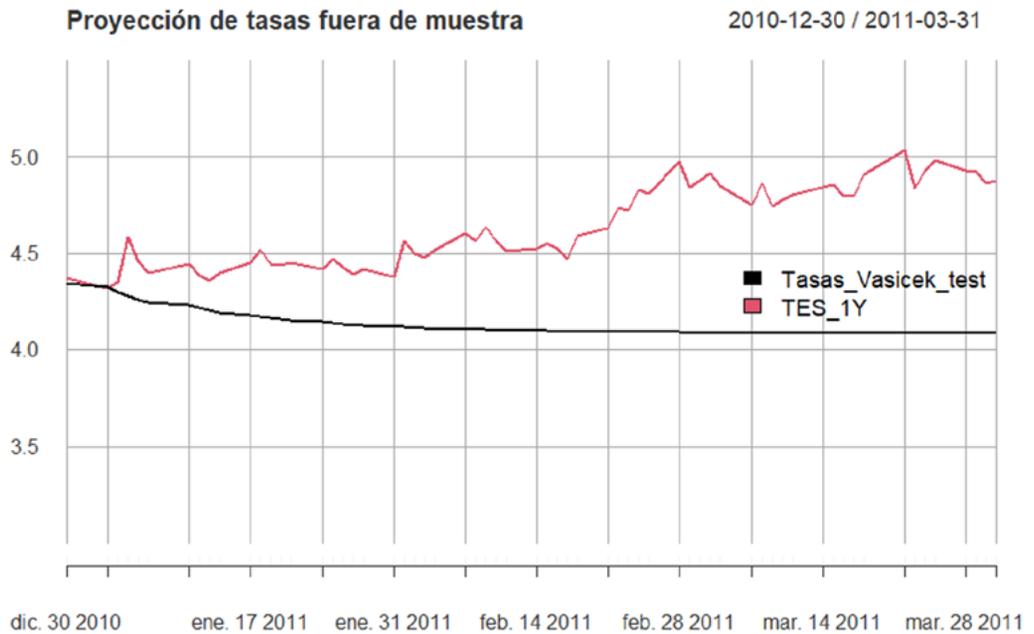
Tabla 2. Parámetros del modelo de Vasicek con datos de entrenamiento

Convergencia	Tasa_equilibrio	Sigma
0.0990396	0.0408648	0.0009661

iii. Con los coeficientes ya calculados dentro de la muestra suministrada, para el cuarto trimestre de 2010 se utiliza la última tasa suministrada de la muestra, para este caso 4,37 %, y se procede a proyectar un trimestre, es decir, el primer trimestre de 2011. En la figura 7 se presenta el ajuste dentro y fuera de muestra.

Figura 7. Ajuste del modelo de Vasicek dentro y fuera de muestra





Fuente: cálculos propios.

iv. Ahora procedemos a calcular las siguientes medidas de precisión del pronóstico: *mean square error* (MSE), *root mean square error* (RMSE), y el *mean absolute percentage error* (MAPE):

Tabla 3. Medidas de precisión dentro de muestra de Vasicek 4Q-2010

MSE	RSME	MAPE
3.9e-06	0.0019858	0.0440681

Tabla 4. Medidas de precisión fuera de muestra de Vasicek 1Q-2011

MSE	RSME	MAPE
3.21e-05	0.0056687	0.1074378

Lo que se presentó anteriormente sería el recorrido de una vez dentro del bucle. Este proceso se repite conforme el bucle avanza y se obtienen una serie de datos con las medidas en cuestión y con las tasas proyectadas para cada periodo.

## 2.2. Ajuste para el modelo CIR

El modelo CIR en una primera presentación sigue una ecuación diferencial estocástica y su comportamiento es de reversión a la media. Sin embargo, este modelo, a diferencia del modelo de Vasicek, no puede generar tasas de interés negativas y tiene un  $\sqrt{r_t}$ , en el término estocástico, por lo tanto, el proceso deja de tener distribución normal. La forma discreta de la ecuación de CIR y su forma en tiempo continuo no difieren, sin embargo, para realizar el ajuste de los coeficientes a través de una regresión lineal, se tiene que reescribir la ecuación dividiendo ambos parámetros en  $\sqrt{r_t}$ , por lo tanto, la forma discreta de la ecuación presenta la siguiente función:

$$\Delta r_t = \alpha(\beta - r_{t-1})\Delta t + \sigma\sqrt{r_{t-1}}\varepsilon_t(0, \sigma^2) \quad (5)$$

Para implementar el cálculo de los parámetros a través de la regresión lineal, la ecuación debe reescribirse

$$\frac{r_t - r_{t-1}}{\sqrt{r_{t-1}}} = \frac{\alpha\beta}{\sqrt{r_{t-1}}} - \alpha\sqrt{r_{t-1}} + \varepsilon_t(0, \sigma^2) \quad (6)$$

Al transformar la ecuación como se presentó, se pueden calcular los parámetros a través de una regresión lineal y después despejar los coeficientes  $\alpha$  (velocidad de reversión a la media) y  $\beta$  (tasa de equilibrio de largo plazo). Al igual que en la sección anterior, para obtener los parámetros se debe realizar un despeje:

$$\alpha = -\alpha \quad \beta = \frac{\alpha\beta}{\alpha}$$

Finalmente, para la proyección fuera de muestra, al ser un proceso que converge a la media, utilizamos los coeficientes calculados y los aplicamos de la siguiente manera

$$E(r_{t+n}|r_0) = r_0 e^{-\alpha n} + \beta(1 - e^{-\alpha n}) \quad (7)$$

### 2.2.1. Implementación del modelo CIR (1985) dentro del bucle

La implementación del CIR dentro del bucle genera el cálculo de los coeficientes por cada vez que se recorre el bucle con la información dentro de la muestra. Una vez se calculan los coeficientes con la información se proyectan las tasas adelante fuera de muestra. Para el desarrollo del ejemplo se utilizará la información de la sección de implementación del modelo de Vasicek (1977) dentro del bucle.

Por lo tanto, se utilizan las tasas de TES a 1 año comprendidas entre el 30 de septiembre de 2010 al 30 de diciembre de 2010, es decir, el cuarto trimestre de 2010. Esto sería equivalente al cuarto recorrido del bucle y serían los nuevos datos de muestra, reemplazando los datos de muestra anteriores.

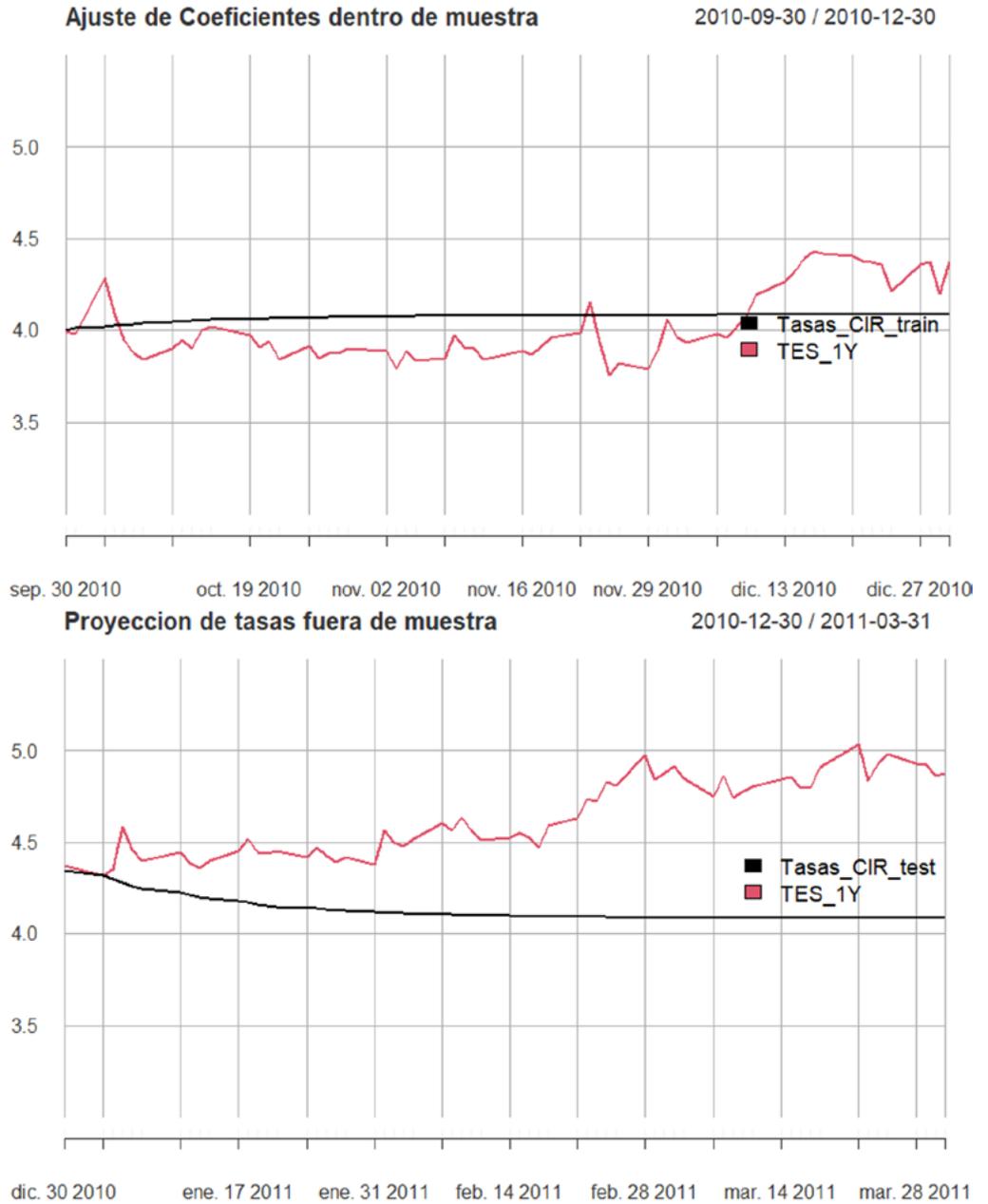
1. Utilizando la recursividad de separación de los datos se calculan los coeficientes:  $\alpha$  (velocidad de convergencia),  $\beta$  (tasa de equilibrio) y  $\sigma$  (volatilidad del modelo).

Tabla 5. Parámetros modelo CIR con datos de entrenamiento

Convergencia	Tasas_equilibrio	Sigma
0.1006482	0.0408548	0.0047963

2. Con los coeficientes calculados del modelo CIR para el cuarto trimestre de 2010, se utiliza la última tasa suministrada de la muestra, equivalente al 4,37 % y se proyecta el siguiente trimestre (enero-marzo 2011) (figura 8).

Figura 8. Ajuste del modelo CIR dentro y fuera de muestra



Fuente: cálculos propios.

3. Ahora, se calculan los indicadores de precisión de pronóstico (MSE, RMSE y MAPE) para el modelo dentro y fuera de muestra.

Tabla 6. Medidas de precisión dentro de muestra CIR 4Q-2010

MSE	RSME	MAPE
3.9e-06	0.0019856	0.0440291

Tabla 7. Medidas de precisión fuera de muestra CIR 1Q-2011

MSE	RSME	MAPE
3.23e-05	0.0056816	0.107777

### 2.3 Ajuste para el modelo de árbol de regresión

Los árboles de regresión son un mecanismo de aprendizaje supervisado, cuya finalidad para este trabajo consiste en realizar una proyección de las tasas de TES a 1 año, y observar los pronósticos realizados conforme avanza el bucle y el modelo aumenta su base de datos de entrenamiento. De acuerdo con el léxico utilizado dentro del mundo de aprendizaje de máquinas, este problema se clasifica como de pronóstico o de regresión ya que la variable salida es de tipo continuo y no cualitativo. A diferencia de los modelos tradicionales, estos van a tener memoria, ya que conforme el bucle avanza, los datos de entrenamiento son cada vez mayores.

La selección de este modelo radica en su fácil interpretabilidad, su robustez frente a los datos atípicos (*outliers*), y que al ser un modelo no paramétrico no tiene ninguna asunción respecto a los datos de ingreso. A continuación, se presenta la ecuación con la que se va a entrenar el modelo:

$$TES1Y_t = f(TES1Y_{t-1}) + \epsilon_t \quad (8)$$

Siendo  $f(\cdot)$  el algoritmo de bosques aplicado al dato anterior de la serie de tiempo, similar a un proceso de orden AR(1). Dicho de otra manera, el algoritmo utiliza el dato inmediatamente anterior y se entrena para obtener el siguiente dato.

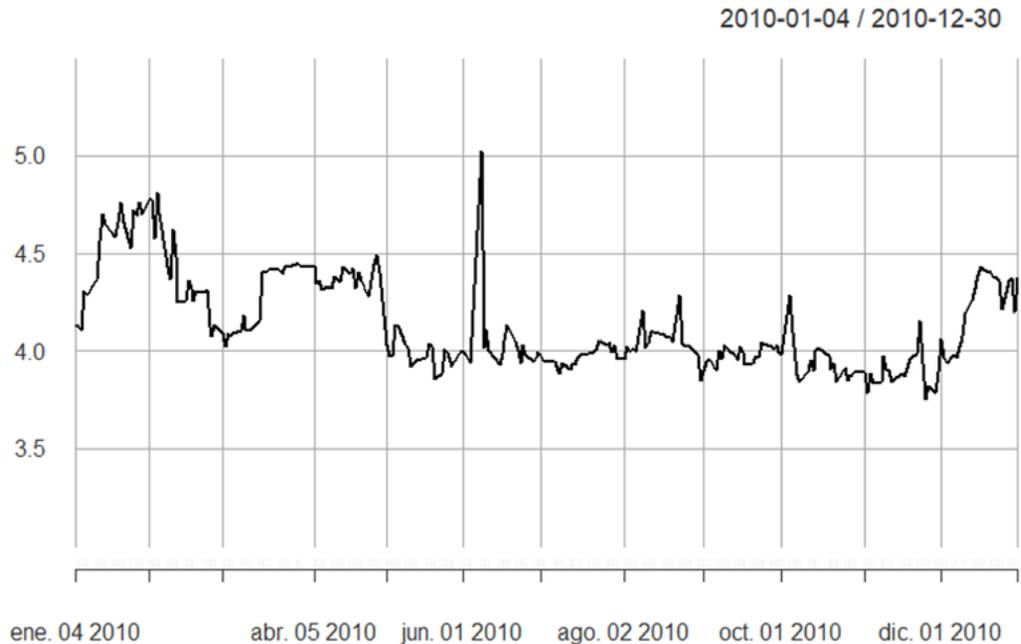
Los árboles se construyen a partir de un algoritmo conocido como segmentación recursiva, sin embargo, su explicación al detalle no hace parte de la extensión de este trabajo. Por lo tanto, se implementa este algoritmo de segmentación, para el pronóstico de las próximas tasas de interés de los TES a 1 año.

### 2.3.1. Implementación de árboles de regresión dentro del bucle

La implementación de este modelo consiste en un proceso que no solamente recalibra los datos de entrenamiento, sino que además recalibra el modelo y lo optimiza conforme los datos cambian. A continuación, se ilustra el proceso como si fuera la cuarta vez que se ingresa en el bucle, esto para que se puedan comparar los resultados obtenidos en las secciones anteriores.

1. Se utilizan las tasas de TES a 1 año comprendidas entre el 04 de enero de 2010 y el 30 de diciembre de 2010 (figura 9).

Figura 9. Tasas de TES a 1 año. Datos de entrenamiento modelos aprendizaje de máquina



Fuente: cálculos propios.

2. Ahora, con base en la función presentada en la ecuación (8) se procede a entrenar el modelo y así el mejor árbol posible dado un conjunto de hiperparámetros que se van a optimizar, los cuales se presentan a continuación.

Profundidad del árbol (`max_depth`): sugiere qué tan profundo debe ser el árbol de regresión. Entre más profundo es el árbol mayor complejidad adquiere el modelo, sin embargo, puede producir un sobreajuste con los datos de prueba y reducir su desempeño en los datos fuera de muestra.

Número mínimo de muestras (`min_sample_split`): dentro del árbol de regresión se presentan múltiples nodos conforme el modelo se ajusta. Este hiperparámetro son el número mínimo de muestras necesarias para dividir un nodo interno dentro del árbol. Esto previene el sobreajuste del modelo y propende a una generalización del modelo.

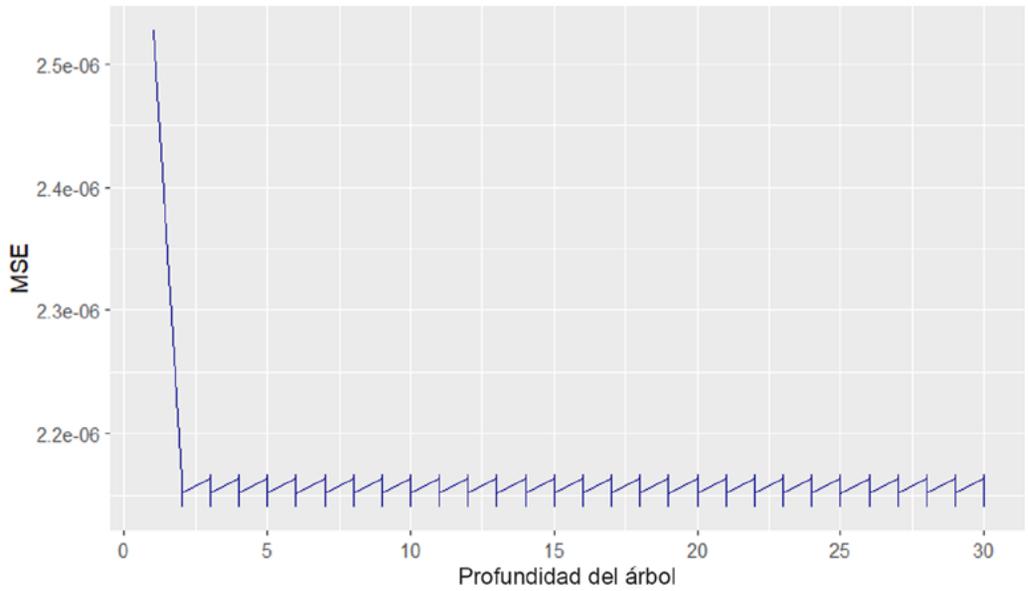
Número mínimo de muestras en el nodo terminal (`min_bucket`): el hiperparámetro refiere al número mínimo de observaciones que debe tener un nodo terminal del árbol. Al igual que el hiperparámetro anterior este propende a realizar un modelo más generalizado.

Ahora bien, utilizaremos un algoritmo de optimización aleatorio, también conocido como *random search*, que, si bien no es el más eficiente, para los propósitos de este trabajo se ajusta bastante bien. El objetivo de esta optimización es hallar el mínimo MSE de los parámetros seleccionados (ver resultados en figuras 10, 11 y 12).

Tabla 8. Hiperparámetros optimizados de árboles de regresión

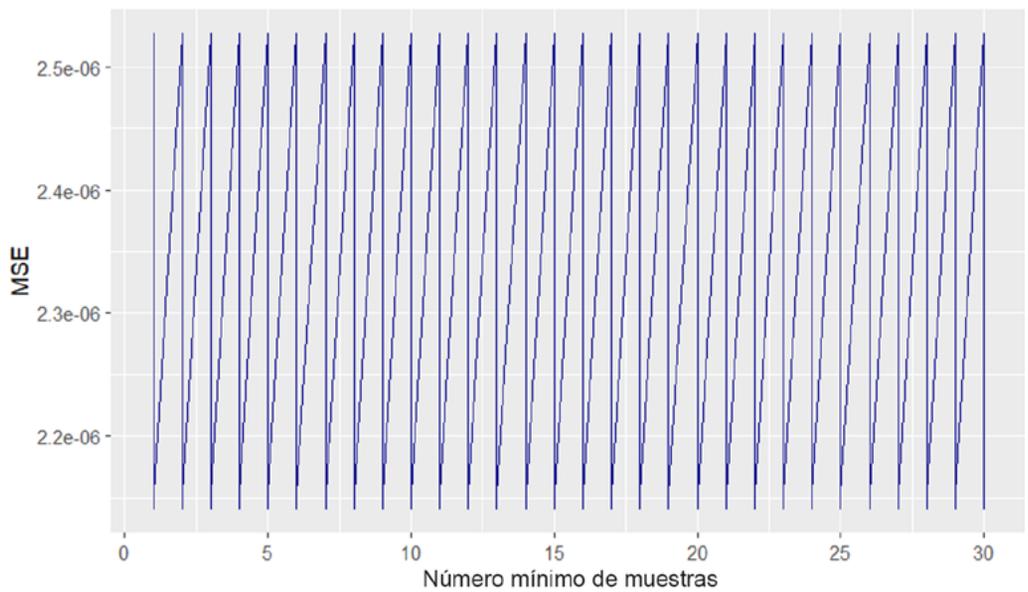
<code>maxdepth</code>	<code>minsplit</code>	<code>minbucket</code>
13	21	23

Figura 10. MSE vs. Profundidad



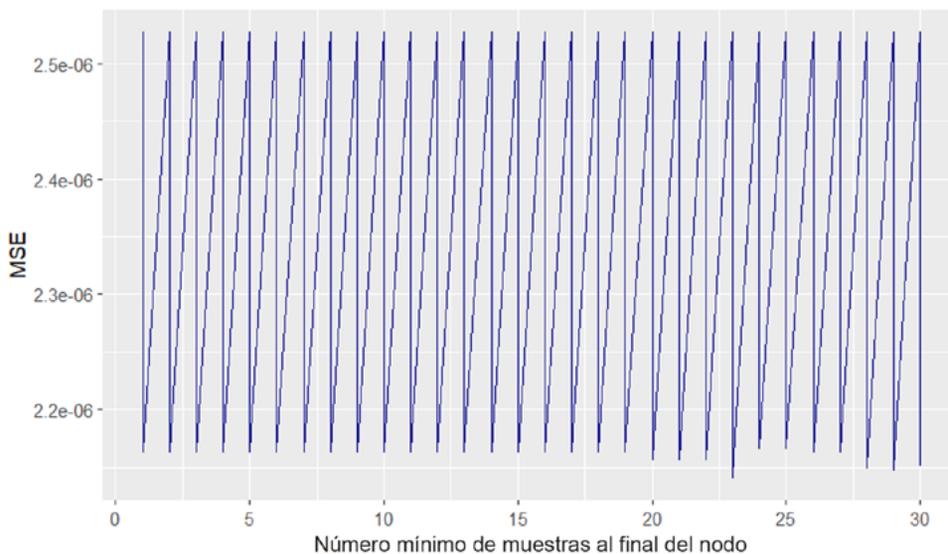
Fuente: cálculos propios.

Figura 11. MSE vs. número mínimo de muestras



Fuente: cálculos propios.

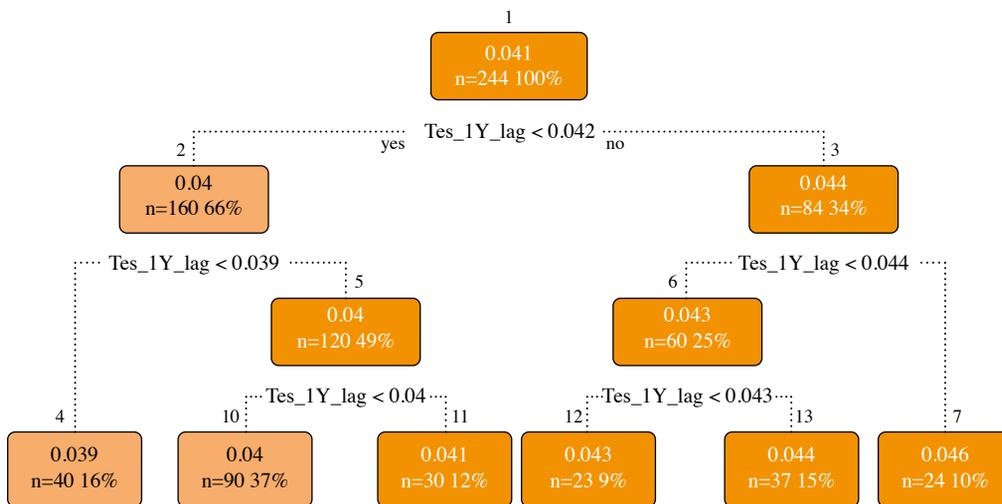
Figura 12. MSE vs. número mínimo de muestras en el nodo terminal



Fuente: cálculos propios.

3. Una vez se obtienen los mejores hiperparámetros para el modelo, se entrena el modelo de árbol de regresión. A continuación, se presenta el modelo seleccionado y entrenado:

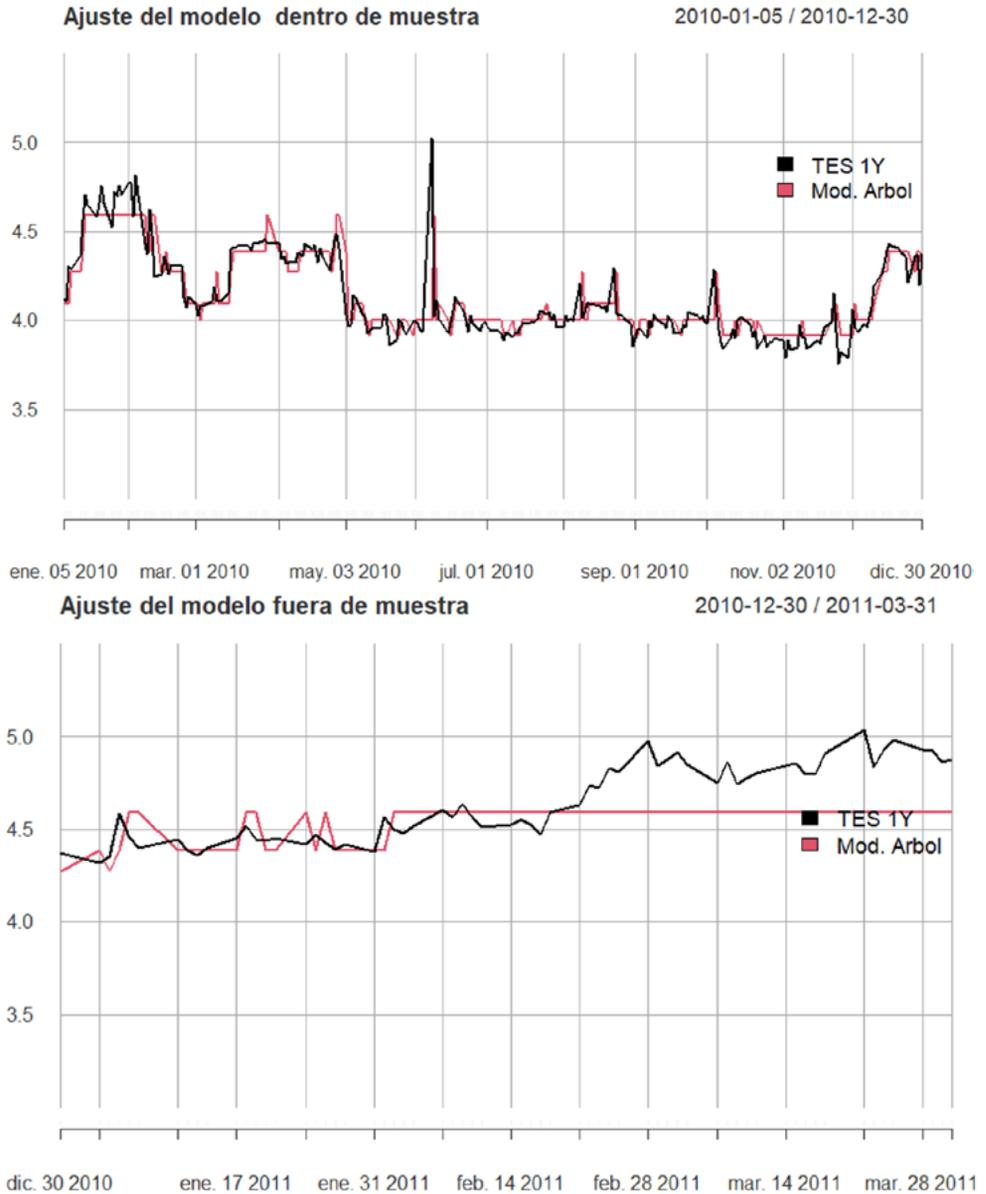
Figura 13. Modelo de árboles TES 1Q 2010–4Q 2010



Fuente: cálculos propios.

4. Con el modelo entrenado, procedemos a realizar el pronóstico de las tasas fuera y dentro de la muestra para, posteriormente, calcular los estadísticos de precisión.

Figura 14. Ajuste de modelos de árboles de regresión



Fuente: cálculos propios.

5. Finalmente, en la ejecución del bucle se calculan los estadísticos de precisión fuera y dentro de la muestra, para determinar qué tan bien se ajusta el modelo fuera de muestra.

Tabla 9. Medidas de precisión dentro de muestra, árboles 1Q 2010- 4Q 2010

MSE	RSME	MAPE
1.4e-06	0.0011667	0.0168068

Tabla 10. Medidas de precisión fuera de muestra, árboles 1Q 2011

MSE	RSME	MAPE
3.7e-06	0.0019363	0.0329775

## 2.4. Ajuste para el modelo de bosques aleatorios

El modelo de bosques aleatorios es un modelo de ensamble de los árboles de regresión, en este caso el objetivo es generar distintos árboles, que con los hiperparámetros seleccionados anteriormente lleguen en conjunto al mejor modelo y resultado posible, buscando maximizar la precisión del modelo, medida a través del MSE.

El generar distintos árboles de regresión reduce el sesgo y la varianza del modelo. Al igual que los árboles de regresión, este modelo no asume ninguna distribución de los datos. Adicionalmente, la ecuación que describe el proceso es similar a la presentada en los árboles de regresión, con la diferencia de que  $h(\cdot)$  denota la implementación del algoritmo de bosques aleatorios.

$$TES1Y_t = f(TES1Y_{t-1}) + \epsilon_t \quad (9)$$

En el caso del algoritmo implementado, este sigue un proceso denominado *bagging*, del cual se producen árboles aleatorios, que no dependen de los resultados previamente obtenidos por el código; al final del proceso se realiza una votación en la cual el dato que se escoge es aquel que fue votado por la mayoría de los árboles

### 2.4.1 Implementación de bosques aleatorios dentro del bucle

La implementación del algoritmo de bosques aleatorios dentro del bucle se explica en la figura 17, donde se seleccionan datos dentro de muestra, se calibran los hiperparámetros y se realiza una proyección fuera de muestra. A continuación, se presenta el proceso como si fuera la cuarta vez que se entra en el ciclo, esto para poder comparar los resultados anteriores.

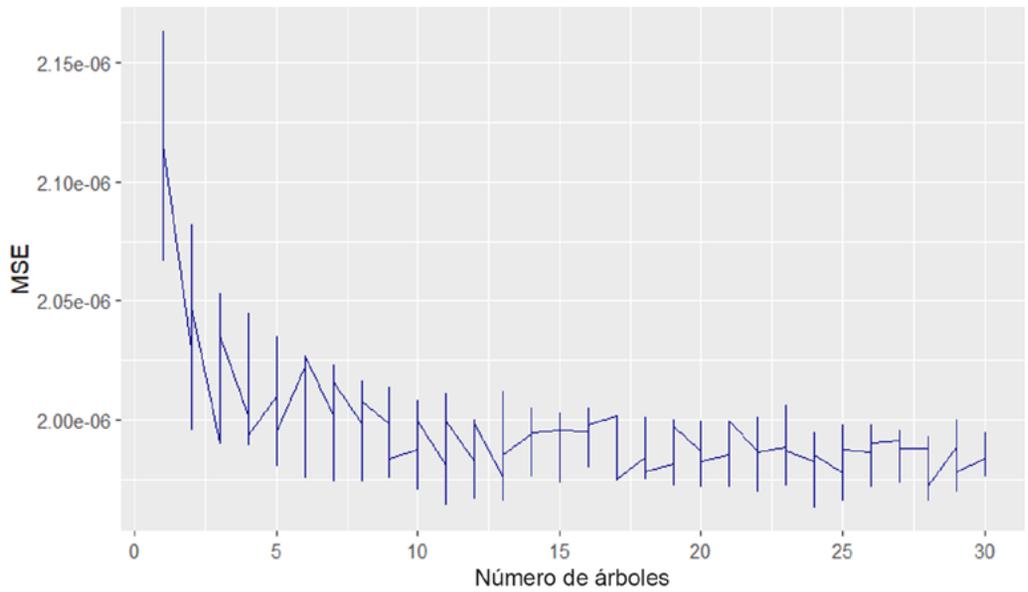
- i. Para este modelo se utilizan, al igual que en el modelo de árboles de regresión, las tasas de TES a 1 año comprendidas desde el inicio de los datos hasta el momento que se calibra el modelo dentro de muestra, esto en fecha sería desde el 04 de enero de 2010 al 30 de diciembre de 2010 (figura 17).
- ii. Al igual que en el proceso anterior, se utiliza un algoritmo de búsqueda aleatoria para la optimización de los hiperparámetros, sin embargo, ya que los bosques aleatorios cuentan con varios hiperparámetros similares al algoritmo de los árboles aleatorios, se utilizan estos hiperparámetros; al momento de optimizar solo se iteran los siguientes:
  - Número de árboles (ntree): hace referencia al número de árboles aleatorios que va a generar el modelo.
  - mtry: hace referencia al número de variables seleccionadas de manera aleatoria como candidatos para realizar un nodo dentro de un árbol.

El objetivo de seleccionar las variables es minimizar el MSE, por lo tanto, en el cuarto recorrido del bucle se utilizan los hiperparámetros de la tabla 11 y en las figuras 15 y 16 se puede observar el comportamiento del MSE en la optimización de cada hiperparámetro.

Tabla 11. Hiperparámetros bosques aleatorios

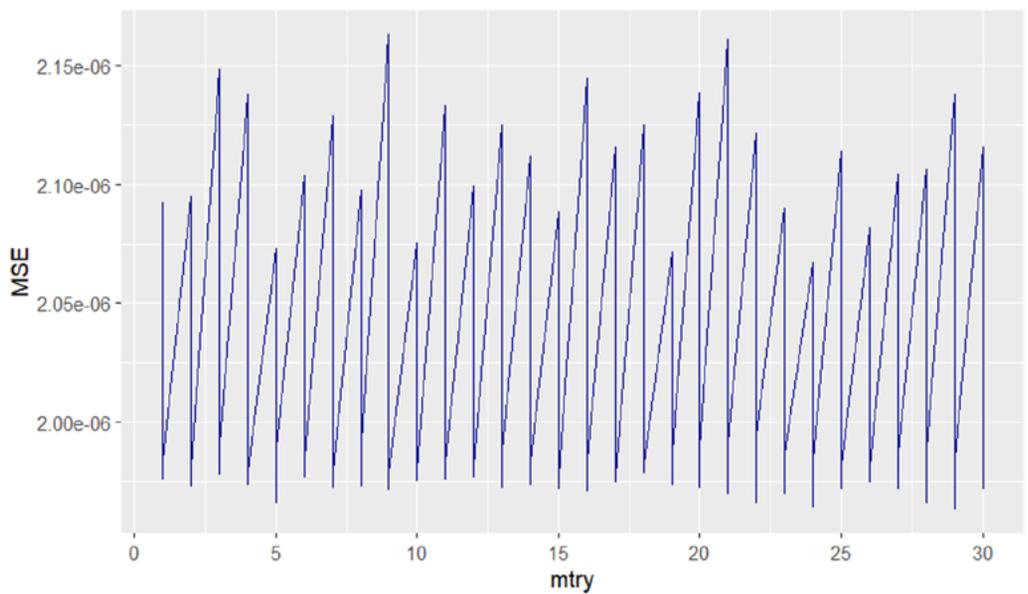
NTREE	MTRY
24	29

Figura 15. MSE vs. número de árboles



Fuente: cálculos propios.

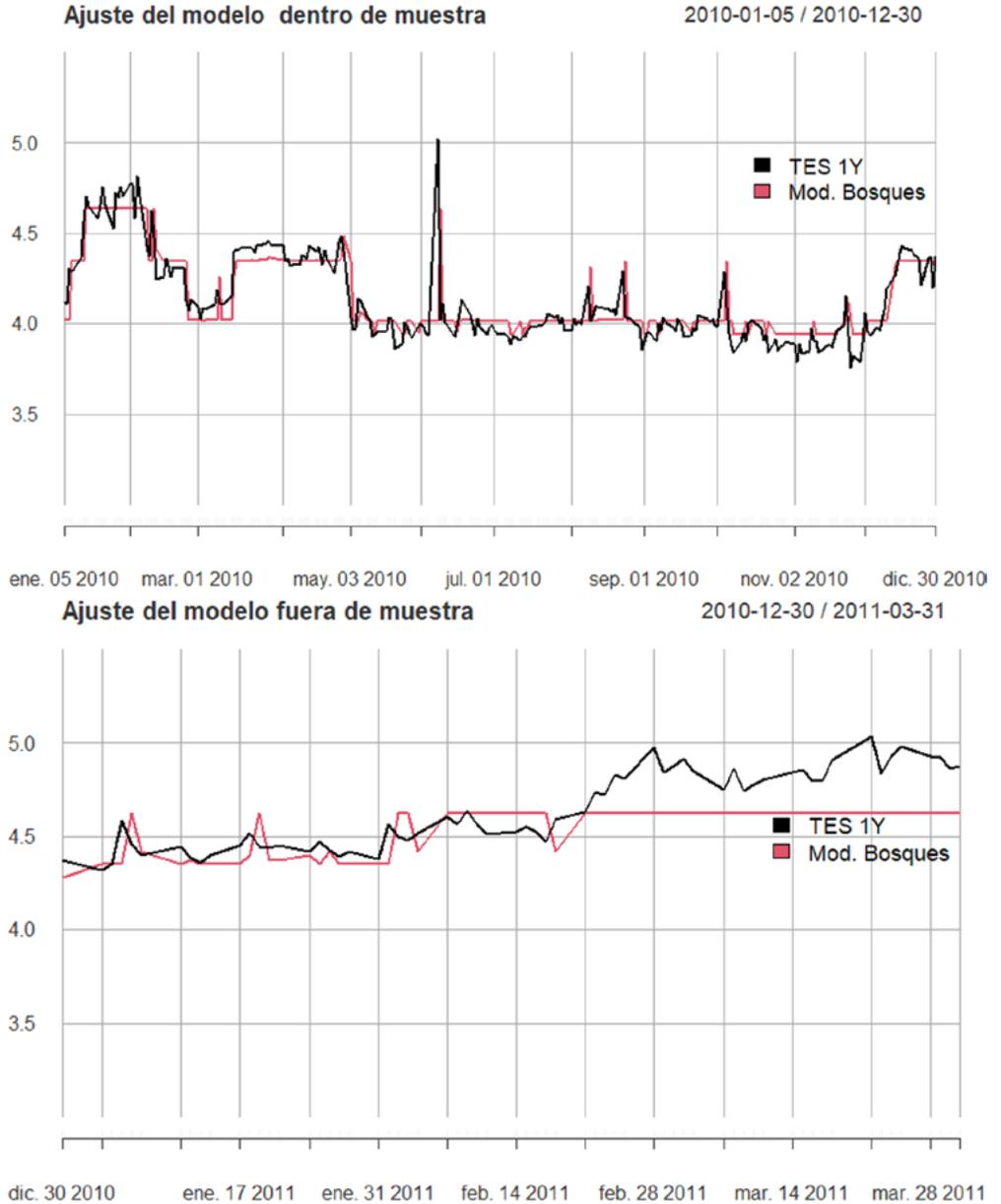
Figura 16. MSE vs. MTRY



Fuente: cálculos propios.

iii. Con el modelo ajustado procedemos a realizar el cálculo dentro y fuera de muestra de las tasas de los TES a 1 año.

Figura 17. Ajuste de modelo de bosques aleatorios



Fuente: cálculos propios.

- iv. Finalmente, procedemos a calcular los estadísticos de precisión del modelo, como los presentamos a continuación.

Tabla 12. Medidas de precisión dentro de muestra bosques aleatorios 1Q 2010 -4Q 2010

MSE	RSME	MAPE
1.4e-06	0.0011951	0.0180111

Tabla 13. Medidas de precisión dentro de muestra bosques aleatorios 1Q 2011

MSE	RSME	MAPE
3e-06	0.001733	0.0306252

## 2.5. Ajuste del modelo para redes neuronales

Los modelos de redes neuronales son una gran herramienta para predecir valores, en este caso –al igual que en los casos anteriores– el modelo funciona como un modelo de aprendizaje supervisado que, con la variable de TES con rezago de un periodo atrás, es capaz de predecir el siguiente valor; en la siguiente ecuación, la función  $g(\cdot)$ , no es más que la representación del algoritmo para predecir el siguiente valor de las tasas.

$$TES1Y_t = h(TES1Y_{t-1}) + \epsilon_t \quad (10)$$

En este caso, los datos ya se encuentran preparados y al funcionar como una serie temporal, no se realizan procesos adicionales como la separación aleatoria de los datos; resaltamos que al igual que en los anteriores algoritmos de aprendizaje de máquinas, este también aumenta los datos de entrenamiento conforme el bucle avanza.

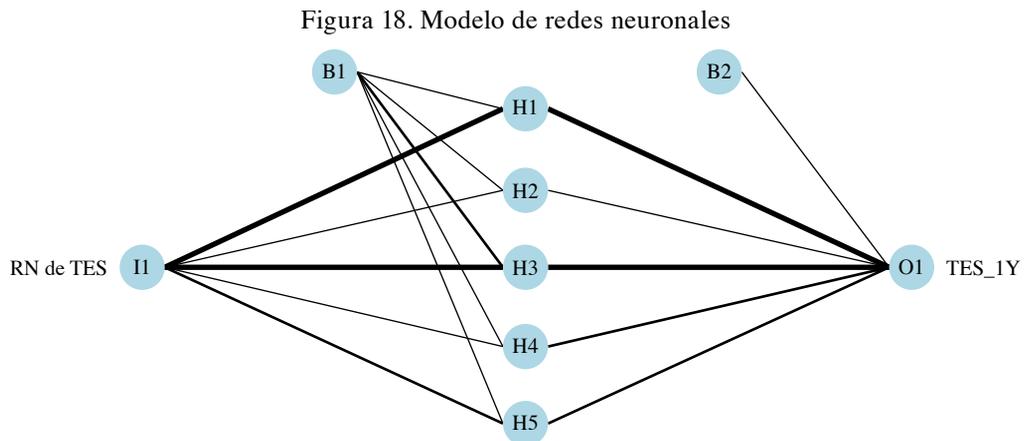
La arquitectura de la red neuronal constará de una única capa. Si bien es cierto que se puede ingresar un número grande de capas ocultas antes del resultado, como se muestra en la ecuación, la única variable independiente es el dato rezagado. En la red utilizada en el código se utiliza una arquitectura de propagación hacia atrás –también conocida como *back propagation*– que,

de acuerdo con el resultado, genera la interacción entre las distintas neuronas y optimiza la función de pérdida, que para este caso el MSE.

### 2.5.1 Implementación de redes neuronales dentro del bucle

Para el caso de las redes neuronales, la implementación es la misma que en las metodologías de árboles de regresión y bosques aleatorios, incluso se utilizan los mismos datos. Por lo tanto, a continuación se presenta la implementación desde la ejecución del modelo.

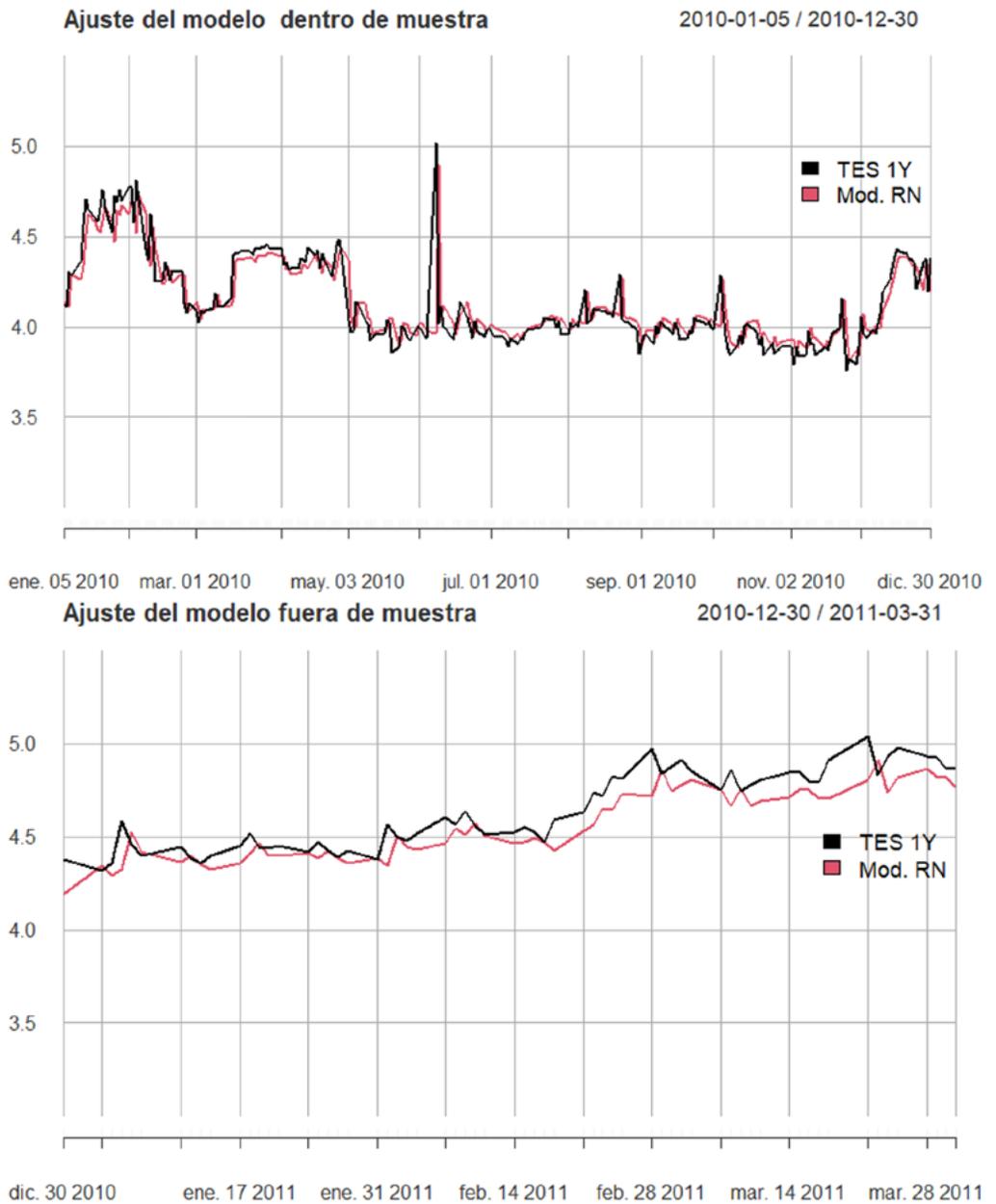
1. En el modelo implementado se utiliza una red neuronal que se entrena con una arquitectura de propagación hacia atrás, se adicionan cinco neuronas y el algoritmo busca minimizar el MSE. De acuerdo con las características del algoritmo desarrollado por Ripley y Venables (2002), el número de neuronas en la capa oculta de la red se determina de manera heurística de acuerdo con la cantidad de variables que ingresan a la red, y los datos de entrenamiento de la misma. A continuación, se presenta el resultado del modelo con los datos de entrenamiento.



Fuente: cálculos propios.

2. Una vez se entrena el modelo, se procede a calcular las tasas pronosticadas por el modelo dentro y fuera de la muestra.

Figura 19. Ajuste de modelo de redes neuronales



Fuente: cálculos propios.

3. Finalmente, se calculan los estadísticos de precisión al igual que en los modelos presentados anteriormente.

Tabla 14. Medidas de precisión dentro de muestra de redes neuronales 1Q 2010- 4Q 2010

MSE	RSME	MAPE
1.6e-06	0.0012529	0.017092

Tabla 15. Medidas de precisión fuera de muestra de redes neuronales 1Q 2011

MSE	RSME	MAPE
1.2e-06	0.00106	0.0183588

## Conclusiones

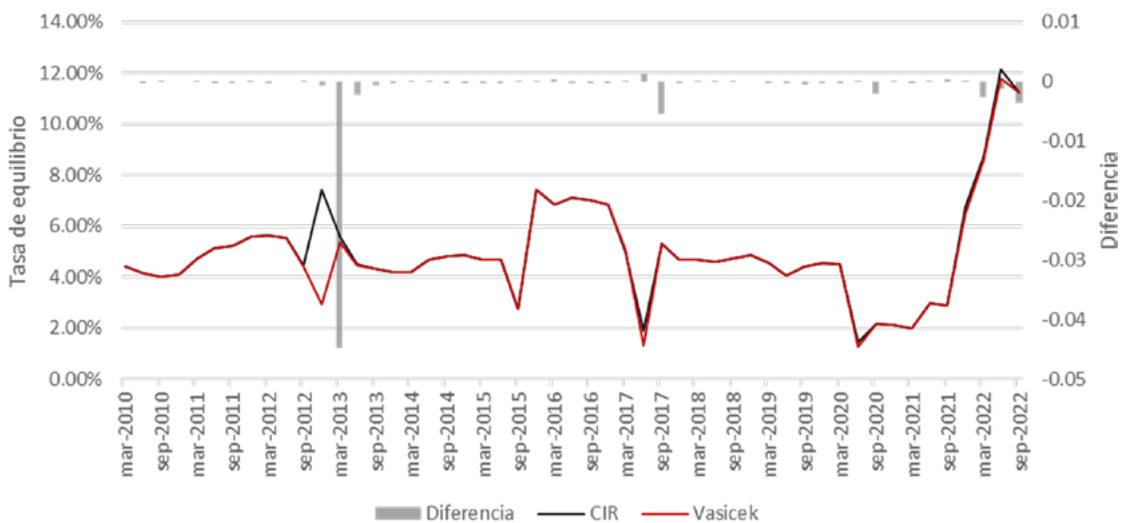
De acuerdo con las hipótesis planteadas en los objetivos, a continuación se presentan los resultados de la investigación. En la tabla 16 se muestra la diferencia entre los parámetros que se obtienen después de correr los bucles con los datos de entrenamiento para los modelos de Vasicek y CIR.

En la figura 20, que muestra la evolución de las tasas de equilibrio de largo plazo, se observa que el resultado del parámetro no suele presentar diferencias significativas entre un modelo y otro, a excepción de septiembre de 2022, cuando el modelo CIR obtuvo una tasa de equilibrio del 7,41 %. Como se presenta en la tabla 16, la diferencia promedio de los parámetros obtenidos es de 0,1 puntos básicos (pbs), pero la mediana es de 0 pbs. Por lo tanto, podemos observar que para ambos modelos la forma en la cual se obtiene la tasa de equilibrio no difiere significativamente, lo que indica que sería indiferente aplicar uno u otro modelo para el pronóstico de las tasas de interés en el nodo de 1 año.

En el parámetro en que se observa una mayor diferencia es en el de la velocidad de convergencia (figura 21); los parámetros calculados por ambos modelos muestran una correlación lineal del 99,96 %, de manera que, con los datos de entrenamiento, el parámetro tampoco muestra diferencias importantes, en promedio la diferencia de un parámetro del modelo de Vasicek, y uno del modelo CIR es de 4 pbs, y una mediana de 5 pbs.

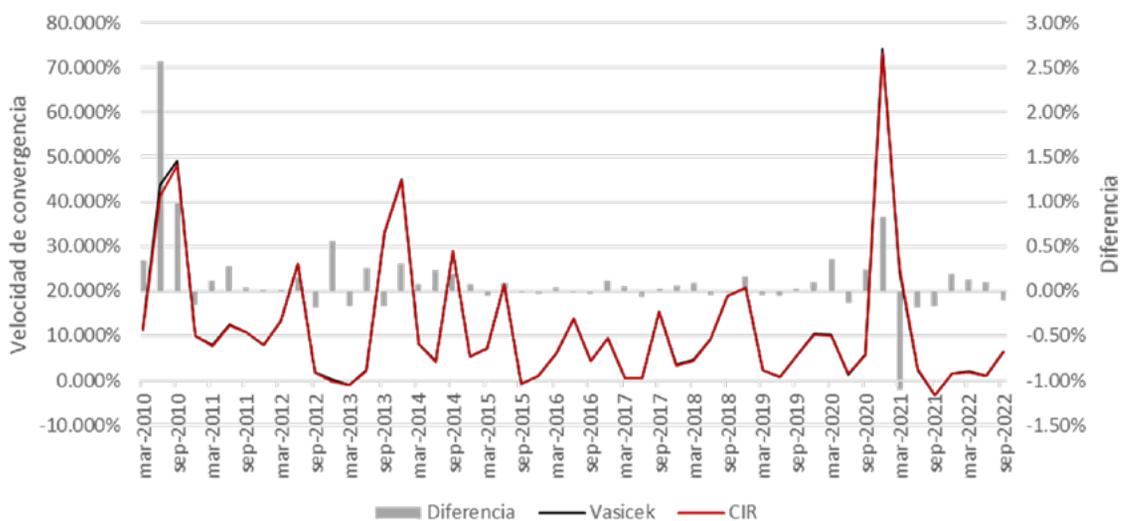
Respecto a la primera hipótesis se puede concluir que de acuerdo con la metodología utilizada para el cálculo de los parámetros de los modelos CIR, y el modelo Vasicek, no existe una diferencia significativa en la aplicación de uno u otro modelo.

Figura 20. Tasa de equilibrio ( $\beta$ )



Fuente: cálculos propios.

Figura 21. Velocidad de reversión a la media ( $\alpha$ )



Fuente: cálculos propios.

Tabla 16. Diferencias entre la estimación de parámetros Vasicek- CIR

	Tasa de Convergencia	Tasa de equilibrio
Min	-1.10%	-4.4750%
1Q	-0.04%	-0.0014%
Mediana	0.05%	0.0000%
Promedio	-0.04%	-0.0014%
3Q	1.77%	0.0006%
4Q	2.56%	0.1195%
Max	2.56%	0.1195%

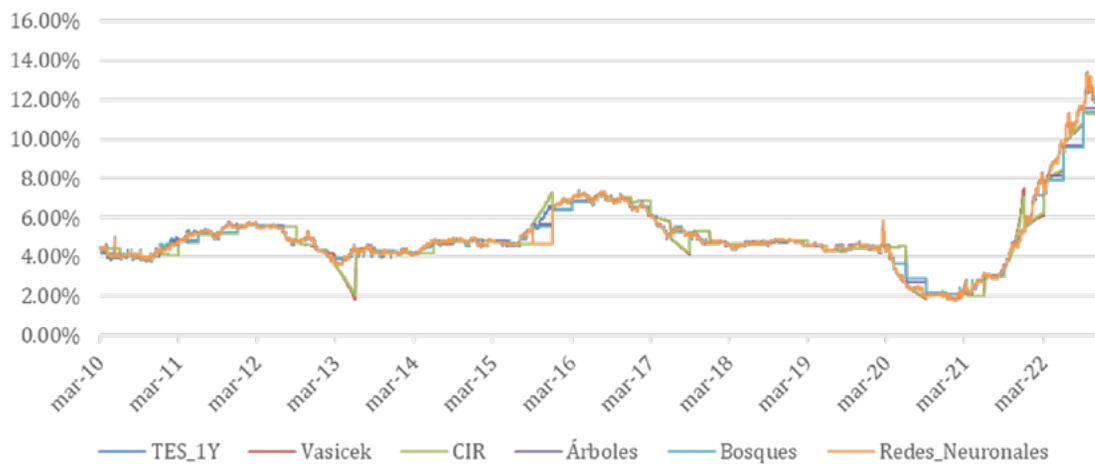
Fuente: cálculos propios.

El segundo objetivo es evaluar el desempeño de los modelos propuestos fuera de muestra, para dictaminar si aquellos definidos como tradicionales tienen mejor capacidad de pronóstico que los algoritmos implementados con aprendizaje de máquina, para esto utilizamos estadísticos de desempeño previamente descritos en el trabajo y su interpretación. En las figuras 23, 24 y 25 mostramos los gráficos de cajas y bigotes (*box plots*) de la distribución de los estadísticos de precisión de los modelos implementados en el trabajo.

En el RMSE, valores más bajos implican un mayor ajuste y se pueden interpretar como la desviación estándar del pronóstico contra la variable observada. Los modelos de Vasicek y CIR, no solamente tienen valores promedio más altos en RMSE, en los datos de testing, sino que además tienen RMSE más dispersos, mientras que los modelos de aprendizaje de máquinas presentan valores promedios más bajos y una menor dispersión. En el caso del MAPE, donde medimos el error porcentual absoluto, el promedio de los modelos tradicionales es de 6,06 % para Vasicek y de 5,89 % para CIR, mientras que para los otros modelos el promedio más alto es el que se obtiene de los árboles de regresión con un MAPE de 1,55 %; adicionalmente, se puede observar que el estadístico de precisión también es más consistente ya que no presenta variaciones significativas. Por lo tanto, podemos concluir que la aplicación de modelos de aprendizaje de máquina fuera de muestra tiene una mayor capacidad de pronóstico de la tasa corta para los TES a 1 año.

La figura 22 presenta el resultado del pronóstico de cada uno de los modelos fuera de muestra, allí se empalman las series de tiempo y los pronósticos fuera de muestra en la medida que avanzaba el bucle.

Figura 22. Resultados de proyección de tasas TES a 1 año fuera de muestra



Fuente: cálculos propios.

Finalmente, para determinar cuál es el mejor modelo nos basamos en el MAPE, en este caso vemos que, en promedio, el modelo más consistente fuera de muestra es el de bosques aleatorios, no solamente porque tiene un MAPE menor a los demás, sino porque no presenta mucha distorsión, seguido de las redes neuronales y, finalmente, de los bosques aleatorios

Tabla 17. Resumen del MAPE en datos fuera de muestra

	Vasicek	CIR	Árboles	Bosques	Redes Neuronales
Min	0.82%	0.82%	0.67%	0.03%	0.37%
1Q	2.25%	2.08%	1.25%	0.06%	0.54%
Mediana	3.80%	3.72%	1.90%	0.10%	0.86%
Promedio	6.06%	5.89%	3.21%	0.21%	1.55%
3Q	7.32%	7.28%	4.16%	0.21%	1.55%
4Q	33.91%	33.86%	17.26%	1.40%	21.45%
Max	33.91%	33.86%	17.26%	1.40%	21.45%
Desviacion Estandar	0.82%	0.82%	0.67%	0.03%	0.37%

Fuente: cálculos propios.

Tabla 18. Resumen del RMSE en datos fuera de muestra

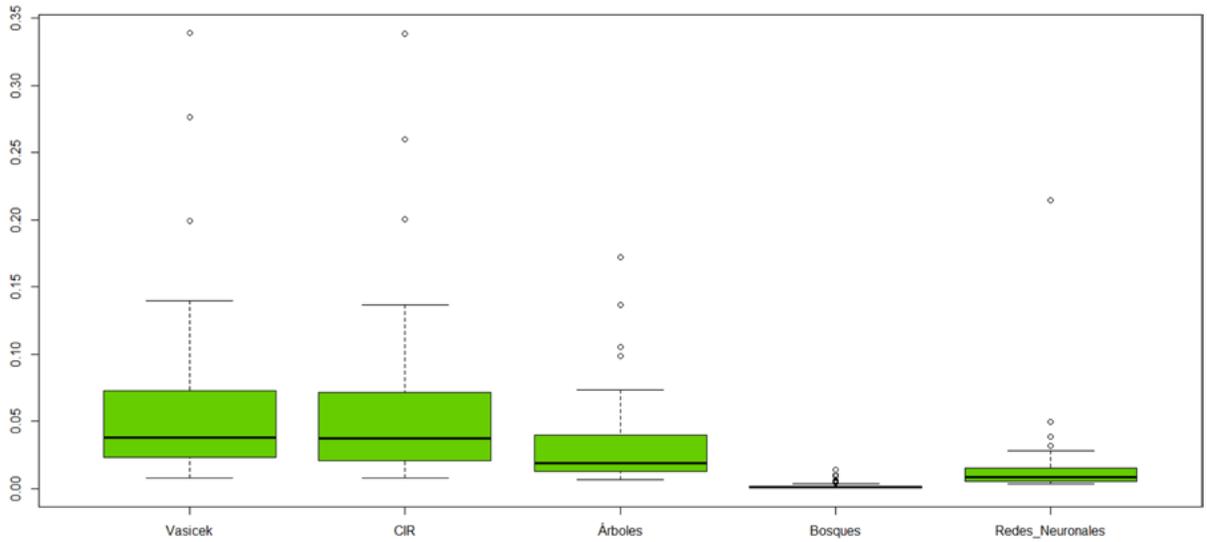
	Vasicek	CIR	Árboles	Bosques	Redes Neuronales
Min	0.04%	0.04%	0.04%	0.03%	0.02%
1Q	0.13%	0.13%	0.07%	0.06%	0.04%
Mediana	0.20%	0.20%	0.11%	0.10%	0.06%
Promedio	0.35%	0.34%	0.20%	0.21%	0.10%
3Q	0.46%	0.46%	0.22%	0.21%	0.09%
4Q	1.37%	1.29%	1.28%	1.40%	1.35%
Max	1.37%	1.29%	1.28%	1.40%	1.35%
Desviacion Estandar	0.04%	0.04%	0.04%	0.03%	0.02%

Fuente: cálculos propios.

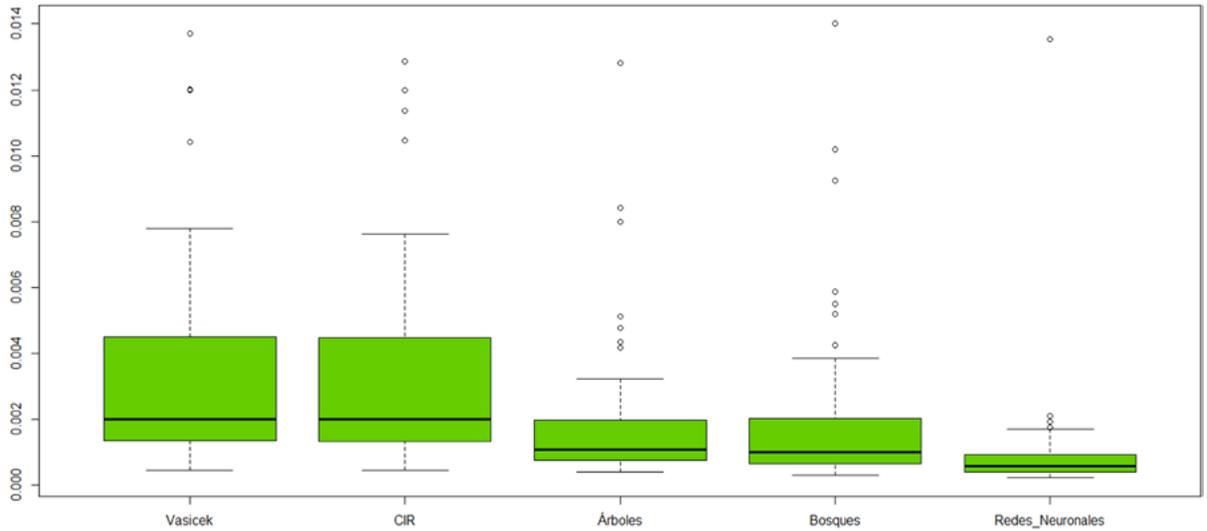
Tabla 19. Resumen del MSE en datos fuera de muestra

	Vasicek	CIR	Árboles	Bosques	Redes Neuronales
Min	0.0000001963	0.0000001935	0.0000001624	0.0000000881	0.0000000461
1Q	0.0000017956	0.0000016819	0.0000005457	0.0000004038	0.0000001547
Mediana	0.0000039580	0.0000039580	0.0000011768	0.0000009921	0.0000003407
Promedio	0.0000222487	0.0000207522	0.0000092389	0.0000115071	0.0000042944
3Q	0.0000215964	0.0000209251	0.0000046441	0.0000045998	0.0000008877
4Q	0.0001882804	0.0001657350	0.0001643097	0.0001961328	0.0001834324
Max	0.0001882804	0.0001657350	0.0001643097	0.0001961328	0.0001834324
Desviacion Estandar	0.0000001963	0.0000001935	0.0000001624	0.0000000881	0.0000000461

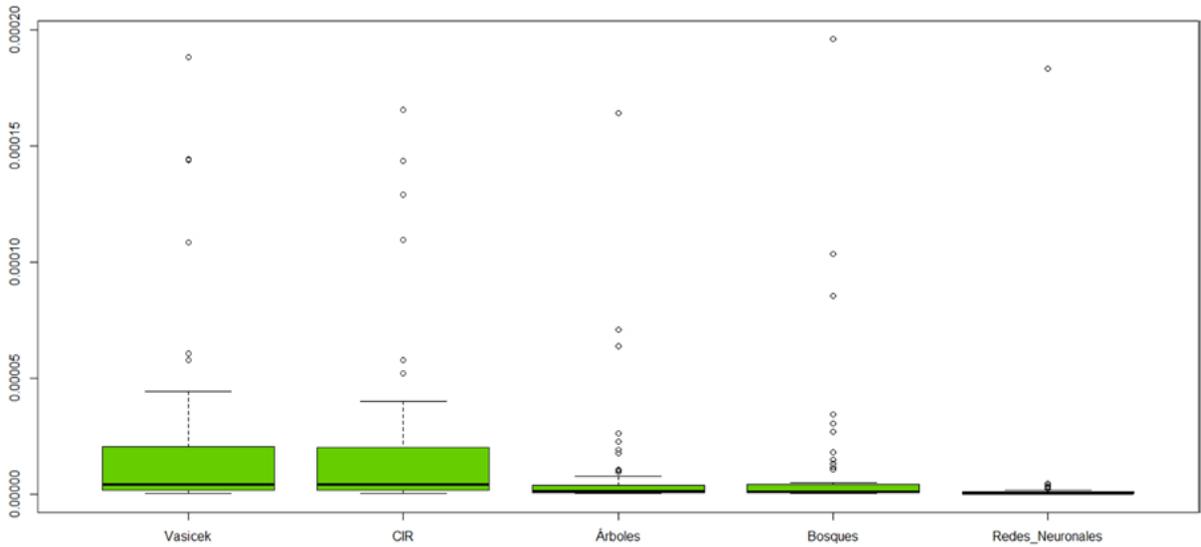
Fuente: cálculos propios.

Figura 23. *Box plot* MAPE fuera de muestra

Fuente: cálculos propios.

Figura 24. *Box plot* RMSE fuera de muestra

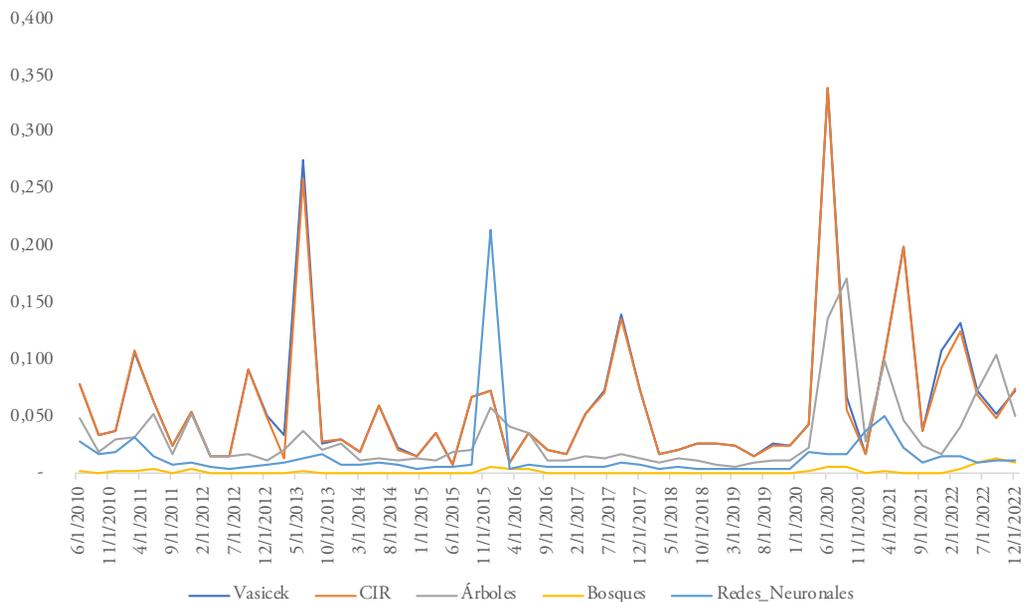
Fuente: cálculos propios.

Figura 25. *Box plot* MSE fuera de muestra

Fuente: cálculos propios.

Dentro de los resultados también presentamos el comportamiento del MAPE y RSME a través del tiempo. Se esperaría que en la medida que la base de entrenamiento aumenta para los modelos de aprendizaje de máquina, los estadísticos de error del modelo disminuyen. Sin embargo, solo observamos este comportamiento en los modelos de redes neuronales y bosques aleatorios.

Figura 26. Serie temporal MAPE fuera de muestra



Fuente: cálculos propios.

Los modelos presentados, que se denominaron tradicionales –como el Vasicek y el CIR–, tuvieron un bajo desempeño al realizar el ajuste (datos dentro de muestra), y uno aún más bajo fuera de muestra en comparación con los modelos aplicados de aprendizaje de máquinas.

Por la forma en que se obtuvieron cada uno de los parámetros de velocidad de reversion a la media y tasas de equilibrio, en los cuales se ajustó a través de una regresión lineal y un método de ajuste de mínimos cuadrados ordinarios, los modelos de Vasicek y CIR tuvieron medidas de desempeño muy similares al momento de pronosticar las tasas de interés futuras al utilizar el valor esperado como medida de pronóstico. Dentro de los modelos de aprendizaje de máquinas, los que mejor desempeño presentaron fuera de muestra fueron los bosques aleatorios, seguidos por las redes neuronales y, finalmente, los árboles de regresión. Resaltamos que estos modelos, si se calibran de manera correcta, pueden ayudar a los diferentes agentes del mercado, no solo a formular estrategias que les permitan gestionar mejor su riesgo de tasa de interés, sino que además con los resultados de los algoritmos pueden formular estrategias especulativas haciendo uso de estos modelos que tienen mejor pronóstico que los tradicionales.

Finalmente, la conclusión es una aproximación práctica a los mercados de deuda colombiana, caracterizando el comportamiento de la tasa cero cupón de TES a 1 año. Como se mencionó, el aporte principal del trabajo es aplicar y presentar los resultados de la aplicación tanto de modelos tradicionales como de modelos de aprendizaje de máquinas, así como brindar un marco metodológico para la implementación de estos.

## Referencias

- Andreasen, M. M., Christensen, J. H. y Rudebusch, G. D. (2019). Term Structure analysis with big data: One-step estimation using bond prices. *Journal of Econometrics*, 212(1), 26-46.
- Bolsa de Valores de Colombia (2023). *Bolsa de Valores de Colombia*. <https://www.bvc.com.co/renta-fija/descripcion-general>
- Brennan, E. y Schwartz, S. (1979). A continuous time approach to the pricing of bonds. *Journal of Banking & Finance*, 3(2), 133-155. [https://doi.org/10.1016/0378-4266\(79\)90011-6](https://doi.org/10.1016/0378-4266(79)90011-6).
- Black, F. y Karasinski, P. (1991). Bond and option pricing when short rates are lognormal. *Financial Analysts Journal*, 47(4), 52-59. <https://doi.org/10.2469/faj.v47.n4.52>
- Carmona, R. y Ludkovski, M. (2004). Spot convenience yield models for the energy markets. *Contemporary Mathematics*, 351, 65-80.
- Castello, O. y Resta, M. (2022). *Modeling the yield curve of BRICS countries: Parametric vs. Machine Learning techniques*. School of Social Sciences, Department of Economics and Business Studies, University of Genova.
- Cox, J. C., Ingersoll, J. E. y Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2), 385-408.
- Diebold, F. X. y Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130(2), 337-364.
- Gibson, R. y Schwartz, E. S. (1990). Stochastic convenience yield and the pricing of oil contingent claims. *The Journal of Finance*, 45(3), 959-976.

- Girsanov, I. V. (1960). On transforming a certain class of stochastic processes by absolutely continuous substitution of measures. *Theory of Probability & Its Applications*, 5(3), 285-301.
- Herrera Cardona, L. G., Cárdenas Giraldo, D. y Salcedo García, J. P. (2011). Estimación de la estructura a plazos para un título de renta fija del tesoro colombiano por el método unifactorial de Vasicek. *Gestion y Desarrollo*, 8(1), 95-114.
- Ho, T. S. Y. y Lee, S. B. (1986). Term structure movements and pricing interest rate contingent claims. *The Journal of Finance*, 41, 1011-1029. <http://dx.doi.org/10.1111/j.1540-6261.1986.tb02528.x>
- Hull, J. y White, A. (1990). Pricing interest-rate-derivative securities. *The Review of Financial Studies*, 3(4), 573-592.
- Josheski, D. y Apostolov, M. (2021). Equilibrium short-rate models vs no-arbitrage models: Literature review and computational examples. *Econometrics. Ekonomia. Advances in Applied Data Analytics*, 25(3), 42-71.
- Kellerhals, B. P. (2001). *Financial pricing models in continuous time and kalman filtering*. Springer-Verlag Berlin Heidelberg.
- Key, D. E., Azamhuzjaev, M. y Mackey, J. (2001). Economic value added®: A critical. *The Journal of Corporate Accounting & Finance*, 12(2), 65-71.
- La República (2020, mayo 12). *Momento decisivo en el Fondo Monetario Internacional sobre ayuda a emergentes*. *La República*. <https://www.larepublica.co/globoeconomia/momento-decisivo-en-el-fondo-monetario-internacional-sobre-ayuda-a-emergentes-2991149>
- Law, H. S. (2009). *On the modelling, design and valuation of commodity derivatives* (Doctoral dissertation). The University of Manchester.
- Litterman, R. y J. Scheinkman (1991), Factores comunes que afectan el rendimiento de los bonos. *Journal of Fixed Income* 1(1), 54-61.
- Longstaff, F. y Schwartz, E. (1992). Interest rate volatility and the term structure: A two-factors general equilibrium model. *Journal of Finance*, 47( 4), 1259-82.

- Luo, D., Pang, T. y Xu, J. (2021). Forecasting U.S. yield curve using the dynamic Nelson–Siegel model with random level shift parameters. *Economic Modelling*, 94, 340-350.
- Matsumura, M., Moreira, A. y Vicente, J. (2011). Forecasting the yield curve with linear factor models. *International Review of Financial Analysis*, 20(5), 237-243.
- Mejía, C. A. (2015). *Una introducción general a los mercados de commodities a nivel internacional*. SSRN. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3062233](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3062233)
- Melo Velandia, L. F. y Castro Lancheros, G. A. (2010). Relación entre las variables macro y la curva de rendimientos. *Borradores de Economía*, 605, 2-28.
- Miltersen, K. R. (2003). *Commodity price modelling that matches current observables: A new approach*. *Quantitative Finance*, 3(1), 51-58.
- Modigliani, F. (1944). Liquidity preference and the theory of interest and money. *Econometrica*, 12(1), 45-88.
- Nelson, C. R. y Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, 60(4), 473-489.
- Orlando, G., Minnini, R. M. y Bufalo, M. (2019). Forecasting interest rates through Vasicek and CIR models: A partitioning approach. *Journal of Forecasting*, 39(4), 569-579.
- Platen, E. y Bruti-Liberati, N. (2007). *Numerical Solution of Stochastic Differential Equations with Jumps in Finance*. Springer.
- Rezende, R. B. y Ferreira, M. S. (2011). Modeling and forecasting the yield curve by an extended Nelson-Siegel class of models: A quantile autoregression approach. *Journal of Forecasting*, 30(8), 339-350.
- Restrepo Tobón, D. A. y Botero Ramírez, J. C. (2007). *Modelos unifactoriales de tipos de interés: Aplicación al mercado colombiano*. Universidad EAFIT.
- Ribeiro, D. R. y Hodges, S. D. (2004b). *A two-factor model for commodity prices and futures valuation*. Semantic Scholar. <https://pdfs.semanticscholar.org/e395/228d9a0e51a302772cd50ef76108204e8ff8.pdf>

- Ribeiro, D. y Hodges, S. (2004a). *Equilibrium model for commodity prices: Competitive and monopolistic markets*. Papers Series. [https://warwick.ac.uk/fac/soc/wbs/subjects/finance/research/wpaperseries/pp\\_04.130.pdf](https://warwick.ac.uk/fac/soc/wbs/subjects/finance/research/wpaperseries/pp_04.130.pdf)
- Ripley, B. D. y Venables, W. N. (2002). *Modern Applied Statistics with S*. Springer.
- Schwartz, E. S. (1997). The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of Finance*, 52(3), 923-973.
- Thornton, D. L. y Valente, G. (2012). Out-of-sample predictions of bond excess returns and forward rates: An Asset allocation perspective. *The Review of Financial Studies*, 25(10), 3141-3168.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2), 177-188.
- Venegas Martínez, F. (2002). *Riesgos financieros y económicos. Productos derivados y decisiones económicas bajo incertidumbre*. Cengage Learning.
- Venegas, F. y Vasicek, O. (2021). Modelos de la estructura de plazos de las tasas de interés: Revisión, tendencias y perspectivas. *Revista Mexicana de Economía y Finanzas Nueva Época*, 16(2), 1-28.