



Escuela Internacional de Ciencias Económicas y Administrativas
Economía y Finanzas Internacionales

TRABAJO DE GRADO PARA OPTAR POR EL TÍTULO DE
ECONOMISTA CON ÉNFASIS EN FINANZAS INTERNACIONALES

**Pronóstico del spread de la curva cero cupón:
Una aproximación desde las redes neuronales**

Directora: Martha Alicia Misas Arango
Codirector: Fredy Alexander Pulga Vivas

Presentado por:
Pedro Caballero Fernández de Castro
Nicolás Sarmiento Gómez

Chía, Noviembre 9 de 2021

Pronóstico del spread de la curva cero cupón

Agradecimientos

Queremos agradecer a Dios por guiar siempre nuestros pasos y por el privilegio de estar culminando nuestra carrera profesional. Queremos agradecer el esfuerzo de nuestras familias con nosotros y por la formación que nos dieron. Agradecemos también a la profesora Martha Misas por su esfuerzo y dedicación en todo nuestro proceso de formación profesional, también por su dirección en el desarrollo econométrico de esta tesis. Agradecemos al profesor Fredy Pulga por su gran disposición y excelente dirección en el área de finanzas de la tesis.

Resumen

El spread de la curva de rendimientos es una variable frecuentemente utilizada como indicador de la realidad económica de un país. Su capacidad predictiva ha sido ampliamente estudiada en la literatura con resultados que apoyan la utilización de esta variable en previsión de recesiones en el corto plazo. Debido a esto, este trabajo se enfoca en la construcción y evaluación de un modelo predictivo como lo son las redes neuronales artificiales para modelar esta variable con una frecuencia diaria. Esta investigación innova sobre el uso de modelos predictivos para el spread de la curva cero cupón en Colombia y compara sus ventajas y desventajas con el de un modelo lineal.

Palabras clave: Spread de tasas de interés, Corto-Largo Plazo, Curva Cero Cupón, Renta Fija, Redes Neuronales Artificiales, Pronóstico, Modelo de Factores.

Clasificación JEL: G17, G11, C45, C52, C53, C58, C61

Pronóstico del spread de la curva cero cupón

Pronóstico del spread de la curva cero cupón: Una aproximación desde las redes neuronales

El objetivo de investigar el spread de la curva cero cupón recae en el estudio de la rentabilidad que presentan las deudas del Gobierno Nacional Central para diferentes plazos de vencimiento. Su importancia radica en que la estructura de las tasas de interés proporciona indicios respecto a cambios futuros de la economía de un determinado país en el corto plazo (Campbell, 2014).

Asimismo, Christophers (2017) resalta la naturaleza predictiva del spread desde un análisis cualitativo. Este autor reúne investigaciones realizadas y metodologías de medición y control de inflación esperada de los miembros del comité de política monetaria del Banco de Inglaterra. Dentro de este análisis, se menciona que la curva refleja algunas incertidumbres del futuro de acuerdo a la información actual, haciendo un paralelismo a un pronóstico de un modelo econométrico sobre las condiciones futuras. Este símil, surge debido a que lo que le da mayor poder de predicción es que es una observación que recopila toda la información del mercado y no una simplificación de la realidad en conjunto con una técnica estadística. De hecho, en este artículo se menciona también un estudio realizado por Estrella y Hardouvelis (1991), donde muestra que la curva de los tesoros en Estados Unidos para el periodo de 1950-1990 presenta un spread entre el corto y largo plazo negativo entre 4 y 6 trimestres anteriores a una recesión, presentando solamente una señal errónea de recesión en 1967.

En general, diversos estudios muestran las diferentes señales que envía el spread sobre la salud de la economía de un país. Por esta razón, puede llegar a ser de gran utilidad para administradores de portafolios de renta fija y autoridades monetarias el desarrollo de modelos predictivos que permitan anticiparse a los movimientos de esta variable, y en consecuencia, a las perspectivas que se tienen sobre el país.

En consonancia con lo anterior, uno de los modelos predictivos que ha tomado relevancia en los últimos años es el de las redes neuronales artificiales. Su popularidad recae en que su capacidad de pronóstico es elevada al capturar las relaciones no lineales entre las variables de entrada o de capturar variables latentes no observadas. Así, esta investigación hace una primera

Pronóstico del spread de la curva cero cupón

aproximación a la construcción de un modelo de redes neuronales para el spread de la curva cero cupón. No obstante, si bien este modelo puede dar pronósticos sobresalientes, también requiere de una infraestructura tecnológica elevada y de una ventana de tiempo importante para su estimación. Por eso, esta investigación evalúa el costo-beneficio de este proceso de estimación y los pronósticos resultantes con el de un modelo lineal tradicional como el ARIMAX.

La relevancia de este trabajo recae en que, según la investigación de los autores, en la literatura no se encuentran metodologías de redes neuronales aplicadas al spread de tasas de interés en Colombia. También es importante aclarar que, si bien la literatura recopilada utiliza el spread de la curva de rendimientos, este trabajo se enfoca en el spread de la curva cero cupón. Esta decisión se toma debido a que las tasas de la curva cero cupón tienen una duración igual a su madurez, asegurando la consistencia de la referencia de las tasas a lo largo del tiempo. Asimismo, esta decisión es fundamental para este trabajo debido a que Colombia no tiene un esquema de emisiones “On the Run” que permita que los títulos que hacen referencia a los plazos de la curva de rendimientos sean consistentes a lo largo del tiempo.

Así, en la sección 2 se encuentra una revisión de literatura que recopila teorías y factores comúnmente utilizados para modelar el spread. La sección 3 considera una metodología sobre el modelo ARIMAX y el modelo de redes neuronales artificiales. La sección 4 menciona y justifica la elección de variables. Finalmente, las secciones 5, 6 y 7 muestran y discuten los resultados, concluyendo sobre cuál de estos dos modelos presenta mejores pronósticos a partir de indicadores de bondad de ajuste.

2. Revisión de Literatura

Al evaluar la capacidad predictiva de la curva, Álvarez, Ramírez y Rendón (2010) modelan la probabilidad de recesión dado el spread en Colombia. Estos autores concluyen para el periodo entre el 2000-2010 que, si bien el spread envía señales de recesiones futuras, también se deben tener en cuenta si los movimientos del spread corresponden a una aversión al riesgo de los agentes que esperan una recesión, o a fenómenos sobre las expectativas de inflación y de gasto público.

Pronóstico del spread de la curva cero cupón

Ahora, Christophers (2017) menciona las causas de los movimientos del corto y el largo plazo de la curva comenzando por sus implicaciones en la política monetaria cuya interacción sucede así: (1) La autoridad monetaria fija unas condiciones de liquidez, entre ellas la tasa de interés de intervención, que repercuten en el producto, la inversión, el desempleo y la inflación; (2) Las condiciones de estas variables dan como consecuencia en el mercado decisiones de inversión de los agentes de acuerdo a su búsqueda de maximización de beneficio bajo incertidumbre y de su rentabilidad real, dando como resultado movimientos en la curva de rendimientos; (3) Al observar este efecto en la curva, la autoridad monetaria considera la forma en que el mercado responde (o se anticipa) a las decisiones de ella misma o a la coyuntura, y de igual forma, toma decisiones de política monetaria.

Habiendo expuesto las causas de los movimientos, el autor ofrece también las interpretaciones de la forma de la curva para evaluar la futura coyuntura en el corto plazo de un país. Por un lado, la interpretación de una curva de rendimientos “normal”, cuya estructura consiste en que a mayor plazo se presenta mayor rentabilidad, es que invertir capital en el largo plazo incurre en un costo de oportunidad de invertir en activos más rentables y/o de menor duración. Por esta razón, la demanda de estos títulos es inferior a los de corto plazo, los cuales presentan condiciones de liquidez y de confianza alrededor de la perspectiva crediticia e inflacionaria del dinero subyacente a la deuda, lo que da como resultado un precio inferior al de corto plazo y una rentabilidad mayor. Por otro lado, la interpretación de la curva invertida recae en el interés de los inversionistas de invertir en el largo plazo, principalmente por perspectivas de riesgo de impago en el corto plazo y de políticas monetarias expansivas que disminuyen las tasas de política monetaria para amortiguar choques negativos en el crecimiento.

Profundizando desde la teoría, Hull (2015) menciona que existen tres principales teorías sobre los movimientos y la estructura de la curva de rendimientos. La primera es la teoría de las expectativas, la cual considera que los rendimientos a largo plazo están explicados por las expectativas sobre las futuras tasas de interés a corto plazo, es decir, la curva forward que se construye a partir de la curva cero cupón.

Pronóstico del spread de la curva cero cupón

La segunda es la teoría de la segmentación del mercado, en la cual se considera que los niveles de rendimiento entre el corto, mediano y largo plazo no tienen que ver una con la otra. Más bien, cada término se explica bajo la conjetura que el corto, mediano y largo plazo se determinan por la oferta y la demanda del mercado de bonos respectivamente, y que estas dinámicas de oferta y demanda entre temporalidades no se relacionan entre sí.

La tercera es la teoría de la preferencia por la liquidez, la cual asume que los inversionistas tienen la preferencia de invertir en el corto plazo para mantener unos niveles de liquidez. Por otro lado, los prestatarios prefieren pagar las deudas a tasas fijas en el largo plazo, por esta razón, el equilibrio de preferencias lleva a que la tasa a largo plazo sea mayor que la tasa cero cupón futura. Esta teoría es particularmente compartida, dado que es consistente con lo observado empíricamente sobre la forma normal de la curva, donde la tendencia es que el largo plazo presente una mayor tasa interna de retorno que el corto plazo.

Considerando la teoría de las expectativas, Campbell y Shiller (1987) ponen a prueba esta teoría con una estimación VAR en un sistema de dos ecuaciones. Por un lado, están las tasas a un mes explicadas por sus rezagos y por el spread entre el corto y largo plazo, y por otro, está el spread explicado también por sus rezagos y los de la tasa a un mes.

De esta forma, a partir de su desarrollo teórico y la estimación concluyen que: (1) los cambios en las tasas de corto plazo causan Granger al spread entre el corto y largo plazo; (2) el spread causa Granger al cambio de las tasas de corto plazo; (3) la varianza entre el spread observado y el spread estimado son estadísticamente iguales; (4) la relación entre el cambio de la tasa de largo plazo y el spread es negativa, lo que es contrario a la teoría de las expectativas. Este último resultado es consistente con el resultado de Shiller et al. (1983), el cual, pese a que los puntos anteriores indican 'cierto elemento de verdad' de la teoría, el punto (4) la rechaza.

No obstante, Carriero, Favero y Kaminska (2006) se basan en el trabajo de Campbell y Shiller (1987) para poner a prueba la teoría. Lo que diferencia su trabajo es que tienen en cuenta variables latentes para describir la estructura temporal, a la par que incluyen variables macroeconómicas para incluir los efectos de la política monetaria en el spread.

Pronóstico del spread de la curva cero cupón

Al considerar la estructura temporal, estos investigadores se basan en el modelo de Nelson y Siegel (1987), donde la curva cero cupón se genera a partir de la siguiente función que considera el nivel, la pendiente y la curvatura de los rendimientos bajo las diferentes temporalidades:

$$y_{t,k} = L_t + SL_t \left(\frac{1 - e^{-\frac{k}{\tau_1}}}{k/\tau_1} \right) + C_t \left(\frac{1 - e^{-\frac{k}{\tau_1}}}{k/\tau_1} - e^{-\frac{k}{\tau_1}} \right)$$

Donde $y_{t,k}$ es el rendimiento del título a la madurez k en el momento t . L_t es el intercepto estimado que mide el nivel de los rendimientos, se asocia como una variable de largo plazo, dado que no es función de la madurez. SL_t es el parámetro estimado que mide la pendiente de la curva en el momento t , se asocia como un factor de corto plazo al descender rápidamente cuando k aumenta. C_t es la curvatura de la estructura en el momento t , se asocia como un factor de mediano plazo, dado su comportamiento parabólico alrededor del mediano plazo en función a la madurez. Finalmente, τ_1 es un parámetro que es constante en el tiempo y habla de la velocidad en que los regresores tienden a cero.

Bajo estas condiciones, Carriero, Favero y Kaminska (2006) utilizan los rendimientos a una madurez de 1, 2, 3, 6, 9, 12, 24, 36, 60, 84, 120 meses para la serie mensual de 1975 hasta 2001, donde estimaron los parámetros L_t , SL_t y C_t para cada periodo observado en corte transversal. La decisión del parámetro τ_1 recayó en ejecutar ese procedimiento con diferentes valores y seleccionar aquel que se ajuste mejor a los datos observados según una medida de bondad de ajuste, en el caso de su investigación, $\tau_1 = 0.87$.

Además de esto, estiman diferentes modelos con distintas variables, en general, el resultado del modelo bivariado de Campbell y Shiller (1987) es el mismo. No obstante, al incluirse variables que explican las dinámicas de los cambios de política monetaria en el tiempo se llega a un resultado consistente a la teoría de las expectativas.

Para este modelo, fueron añadidas variables macroeconómicas que explican el comportamiento de la política sobre tasas de interés que se describen desde la regla de Taylor,

Pronóstico del spread de la curva cero cupón

particularmente, tomando variables de inflación y desempleo. Considerando que la teoría de las expectativas por definición busca explicar las tasas futuras con la información existente en el momento, los autores utilizan datos no revisados para mantener consistencia circunstancial del mercado. Así, los autores muestran bajo un esquema VAR y una estimación por filtro bayesiano que, incluyendo los factores de Nelson y Siegel y variables macroeconómicas, la tasa de interés estimada futura es estadísticamente igual a las tasas por fuera de muestra, concluyendo en que las expectativas sobre las tasas de interés futuras de corto plazo determinan las tasas de interés de largo plazo.

Ahora, considerando los movimientos de la curva, Shatbolt y Taylor (2002) mencionan que la variable recibe las influencias de oferta y demanda, expectativas de inflación, sentimiento de riesgo alrededor del mercado y movimientos de capitales. Adicionalmente, ellos hacen énfasis en la dificultad de obtener algunas variables, particularmente en medir el sentimiento del mercado y la demanda agregada, de esta forma, se hace la consideración de que es necesario utilizar variables 'proxy' para acercarse a los movimientos que representan algunas variables no observadas.

Por ejemplo, para el sentimiento de riesgo del mercado se puede tomar como punto de partida variables que reflejen el optimismo o pesimismo de los traders, comparando los precios de mercado con un nivel teórico apoyándose de un análisis de cointegración donde el corto plazo debe volver a la senda 'teórica' de largo plazo eventualmente.

A partir de lo anterior, dado el impacto que recibe la curva de rendimientos por la política monetaria, las condiciones macroeconómicas y el sentimiento de mercado, es recurrente encontrar en la literatura el uso de variables proxy que reflejen las condiciones de estos factores tal como lo mencionan Shatbolt y Taylor (2002). Concretamente, se busca aproximarse a factores de riesgo intemacional, factores de política monetaria y de liquidez, factores macroeconómicos y factores de riesgo crediticio del deudor.

Empezando por el factor de riesgo internacional, este factor es frecuentemente utilizado para mejorar el pronóstico de la curva de rendimientos. Holló et al. (2012) mediante una prueba

Pronóstico del spread de la curva cero cupón

de estrés que ellos proponen llamada CISS (Composite Indicator of Systemic Stress) donde, a grandes rasgos, se busca medir el riesgo sistémico ya materializado en el mercado financiero y la volatilidad que este riesgo genera en los mercados. Así, utilizando datos de la eurozona desde 1987 hasta 2011 encontraron que, a mayor nivel de riesgo internacional, mayores eran los spreads de las curvas de rendimiento de los países de la Unión Económica y Monetaria de la Unión Europea.

En adición a esto, Geyer et al. (2003), Barrios et al. (2009) y Sgherri and Zoli (2009) también concuerdan sobre la importancia de este factor para los spreads de la curva de rendimientos. Además, Haugh et al. (2009) y Barrios et al. (2009) encuentran también que estos efectos no son lineales, siendo que, a mayor sensación de riesgo en mercados internacionales, mayor spread, donde los efectos del riesgo internacional sobre el spread son más fuertes cuando las condiciones fiscales del país en cuestión se deterioran.

Siguiendo con los factores de liquidez, existen diversas opiniones alrededor de si este es un factor determinante de la curva. Por una parte, Beber et al. (2009) concuerdan sobre la importancia de este factor y encuentran que su efecto es mayor en periodos con altas tasas de interés. Contrario a esta conclusión, Jankowitsch et al. (2006), muestran que este factor no es significativo para explicar el spread. Este estudio se hizo midiendo el impacto de la liquidez en los precios de los bonos soberanos de la Unión Económica y Monetaria de la Unión Europea con una frecuencia diaria para datos de Alemania, Francia, Italia, España, Países Bajos y Austria desde enero de 1999 a marzo de 2001.

Considerando el factor de riesgo de crédito, Afonso et al. (2011) hicieron una evaluación de impacto sobre los efectos de los cambios de calificación y de perspectiva de las calificadoras Standard&Poor's, Moody's y Fitch en los spreads de las curvas de los bonos soberanos en los países de la Unión Europea usando datos diarios de junio de 1995 hasta octubre del 2010. Su metodología busca medir los efectos de valoración de un acontecimiento analizando su respuesta o efecto sobre algún parámetro o variable, esto entorno al anuncio o realización del evento. En este estudio se encontró una respuesta significativa de los spreads a cambios en las calificaciones y en las perspectivas, con ciertas diferencias dependiendo de la agencia

Pronóstico del spread de la curva cero cupón

responsable. Además, encontraron que hay un efecto más severo cuando los cambios son negativos, siendo esto un motivo de no linealidad en los determinantes de los títulos de renta fija en Europa.

3. Estrategia Empírica

Esta sección considera inicialmente una explicación sobre la metodología de un modelo ARIMAX, el cual es el modelo lineal elegido para compararse con el modelo de redes neuronales. De entrada, se tiene en cuenta el mismo conjunto de información para ambos modelos. Por un lado, para el ARIMAX se seleccionan aquellas variables y rezagos que minimicen el error de pronóstico y cumplan las propiedades de parsimonia, significancia y ruido blanco sobre los residuales.

Así, se compara el modelo redes neuronales y el modelo ARIMAX y se evalúa la calidad de pronóstico mediante el Root Mean Square Error (RMSE) y el Mean Absolute Error (MAE). El resultado de esta evaluación concluirá sobre la capacidad de pronóstico de un modelo lineal contra uno no-lineal para el spread. Después, se expone la estrategia de elección de variables para el modelo de redes neuronales artificiales, su arquitectura del modelo y cuál es su relación con las redes neuronales biológicas. Además, se realiza una breve discusión sobre la utilidad y la interpretabilidad de este modelo. Subsecuentemente, se explican los métodos de estimación.

3.1 Modelo ARIMAX

Los modelos ARIMAX son una ampliación del modelo auto-regresivo integrado de media móvil, mejor conocido como modelo ARIMA por su nombre en inglés. Como lo menciona Enders (2014), estos modelos constan de dos partes, la parte auto-regresiva del modelo que contiene un número de rezagos de la variable objetivo, y la parte de medias móviles que contiene un número de innovaciones pasadas. Asumiendo que el modelo contiene p rezagos de la variable en cuestión y q perturbaciones anteriores, a este se le conoce como un ARMA (p, q) , cuya generalización se describe como:

$$y_t = a_o + \sum_{i=1}^p a_i y_{t-i} - \sum_{i=1}^q \beta_i \varepsilon_{t-i} - \varepsilon_t .$$

Pronóstico del spread de la curva cero cupón

En caso de tener $q = 0$, es decir, que el modelo no incluya una parte de medias móviles, este se conoce como un modelo auto-regresivo puro o modelo $AR(p)$, mientras que si $p = 0$, se tendría un modelo de medias móviles puro o modelo $MA(q)$.

Complementando lo anterior, estos modelos requieren el manejo de una serie o variable estacionaria, por lo que en caso de que las raíces de la ecuación en diferencia se encuentren por fuera del círculo unitario, a la serie se le debe añadir la parte I , y se debe diferenciar d veces hasta obtener una serie estacionaria. De esta forma, se conseguiría un modelo ARIMA de la forma $ARIMA(p, d, q)$.

Además, otro requerimiento de estos modelos es que cumplan con la propiedad de ruido blanco en su término de perturbación, característica que se consigue dependiendo de los valores de p y de q . Por lo que, si se sigue el principio de parsimonia, un ARIMA apropiado sería aquel que incluya el mínimo número de parámetros en la estimación para que genere ruido blanco en sus residuales.

Por último, como lo menciona Andrew et al. (2013), los modelos ARIMAX son modelos ARIMA que tienen en cuenta variables exógenas para capturar el efecto de factores externos que pueden afectar a la variable objetivo. Siguiendo esta idea, para poder incluir una variable exógena al modelo, se requiere que esta no presente multicolinealidad con alguna otra variable exógena a incluir.

3.2 Modelo de Redes Neuronales Artificiales

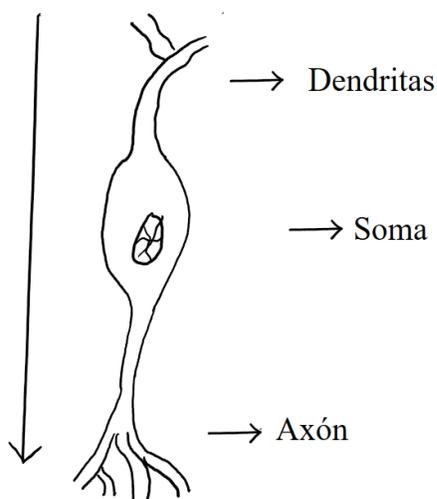
Las redes neuronales artificiales son una clase de modelos paramétricos no lineales. Este modelo recibe su nombre por su similitud con la forma en que el cerebro procesa la información mediante la recepción, procesamiento y envío de señales entre neuronas interconectadas que juntas forman un sistema neuronal. Como lo mencionan Swanson y White (1997), las redes neuronales son un modelo matemático que hace parte de una clase de modelos flexibles propuestos por científicos cognitivos, capaces de aprender mediante prueba y error. Particularmente, este modelo ha sido fuente de atención por sus aplicaciones a pronósticos relativamente acertados en las finanzas y economía, se menciona también que este éxito está

Pronóstico del spread de la curva cero cupón

dado por su capacidad de aproximarse a cualquier función arbitraria con un número de parámetros k lo suficientemente grande.

Redes Neuronales Biológicas

Así, la inspiración de este modelo matemático nace de la forma en que se procesa información en el cerebro. Para entender esta comparación, partimos de explicar los componentes de una neurona biológica sencilla, conocida como la neurona pseudounipolar. Esta célula tiene tres componentes principales bajo las definiciones de Hall (2015):



Neurona biológica, elaborado por Misas, Lopez y Querubin (2002)

I. Dendritas: Ramas del soma receptoras de impulsos eléctricos y químicos proporcionados por el axón de otras neuronas, a las cuales la neurona está conectada. Este proceso de envío y recepción de señales se conoce como la sinapsis.

II. Soma: Es el cuerpo de la neurona, el cual recibe información de las dendritas. Funciona mediante un proceso eléctrico y químico determinado por la genética y las experiencias pasadas de un individuo. El

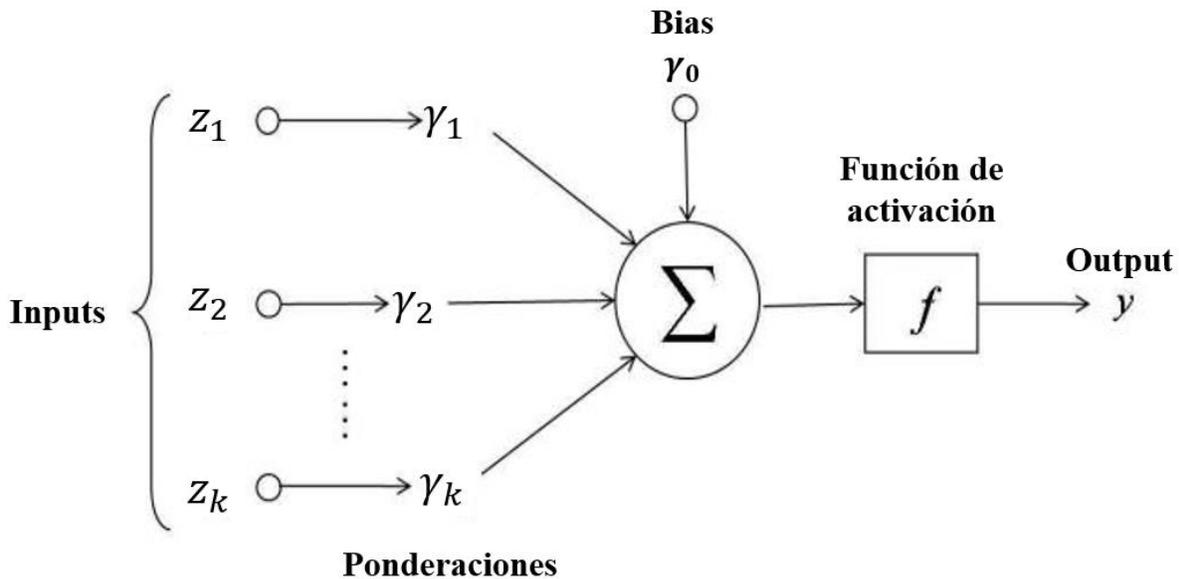
Soma se activa si el impulso recibido supera una intensidad establecida, a lo que, si se activa, culmina en el envío de una señal a otras neuronas mediante el Axón.

III. Axón: Es el canal que envía una señal ante el impulso generado por el soma a otras neuronas, las cuales reciben estas señales con sus respectivas dendritas.

Bajo este esquema, el modelo matemático para una sola neurona, conocido como el perceptrón, funciona de igual manera: (1) Unas variables *input* (z_i) envían un impulso a una función de activación; (2) Similar a las dendritas, esta función de activación recibe estos impulsos y los pondera bajo unos parámetros $\gamma_i \{\forall i = 1, \dots, k\}$; (3) Similar al Soma, si la suma de las señales recibidas ponderadas por esos parámetros es mayor a un intercepto o 'bias' (γ_0), el nodo, que contiene la función de activación, se activará y enviará una señal, dando como

Pronóstico del spread de la curva cero cupón

resultado un *output*. En agregado, la unión de estos componentes se considera como la neurona. La siguiente imagen ilustra este proceso:



Perceptrón y sus componentes, elaborado por: Hay (2019), edición propia.

Arquitectura de las Redes Neuronales, Interpretación y Análisis De Series De Tiempo

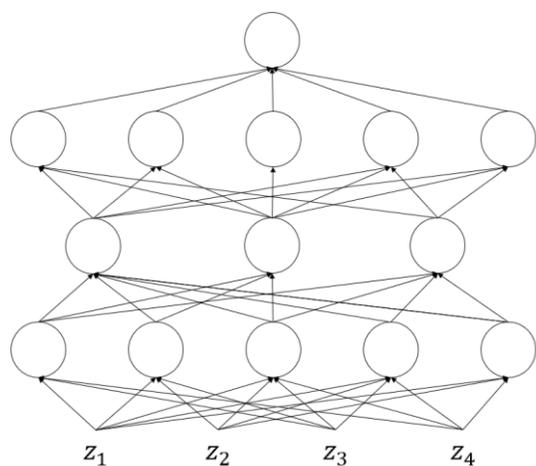
A partir de esta comparación de las redes neuronales biológicas con las artificiales, el siguiente paso es profundizar en la arquitectura del modelo. La agregación de neuronas de forma paralela, secuencial e interconectada permite, al igual que un cerebro, un aprendizaje más flexible capaz de realizar tareas cada vez más complejas, así, estas conexiones forman una estructura neuronal que especifica una arquitectura.

Por esta razón, al construir un modelo de redes neuronales es preciso también considerar la forma en que se envían las señales en la red, el número de nodos y capas ocultas que afectaran el resultado. Por un lado, para propósitos de este trabajo, se va a considerar una red que se alimenta hacia adelante, conocida en la literatura como *feedforward*, la cual tiene como mecanismo la recepción de señales de las variables de entrada que, después de haber atravesado una transformación no lineal, dan un resultado. El mecanismo de una red *feedforward* no permite

Pronóstico del spread de la curva cero cupón

que el resultado de esta transformación no lineal sea insumo para ella misma como input o para otros nodos de la misma capa. No obstante, vale aclarar que, para el periodo de entrenamiento, el error producido por la red en este proceso sí se envía hacia atrás en la red neuronal para que ella misma corrija sus errores, este mecanismo de aprendizaje se conoce como *backpropagation*.

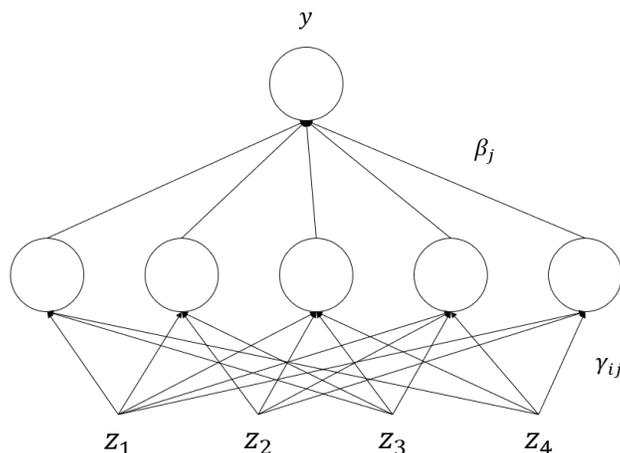
Por otro lado, se definen las capas ocultas como el conjunto de nodos organizados de forma paralela que no están conectados entre sí, donde cada uno recibe unos valores de entrada, ya sea de unas variables *input* o de otros nodos $k-1$, y dan como resultado una salida hacia un *output* o unos nodos $k+1$.



Arquitectura de una red neuronal multicapa feedforward, elaboración propia.

No obstante, en el contexto de series de tiempo, es común utilizar redes neuronales de una sola capa por su gran capacidad de reconocimiento de patrones, incurriendo en un riesgo de sobre-entrenamiento a la muestra presentada. Concretamente, Franses y Van-Dijk (2003) mencionan que la red es capaz de capturar en cierta medida el ruido blanco de las series, lo que provoca que a la hora de pronosticar por fuera de muestra resulte inconveniente este sobreajuste. Esto es debido a que el modelo incluye como parte de sus explicativas una variable aleatoria como un ruido blanco que no necesariamente es útil para pronosticar observaciones futuras. Esta idea la comparte Enders (2014), el cual también menciona que en la medida que se añaden parámetros, el R^2 tiende a 1 por la captura del ruido blanco.

Pronóstico del spread de la curva cero cupón



Arquitectura de red neuronal de una capa, elaboración propia.

Además de esto, la captura de este componente estocástico es una de las razones que impide la interpretación de la red, dando razón a que estos modelos se consideren una ‘caja negra’, pues no es posible dar interpretación a los coeficientes que entran en las funciones de activación de la red, por esta razón, las capas de neuronas se denominan capas ocultas. Sin embargo, Misas, Lopez y Querubin (2002) interpretan estas capas ocultas como la modelación implícita de las relaciones no lineales de las variables de entrada en la red. A su vez, los resultados de estas funciones de activación pueden estar aproximando variables latentes que gobiernan el proceso generador de datos y que, al capturarlas con la red, se consigue una mejora en la calidad de pronóstico del modelo.

Anteriormente, se mencionó sobre la capacidad de las redes neuronales de aproximarse a cualquier función arbitraria. En cuanto a esta capacidad, Franses y Van-Dijk (2003) ponen a prueba a las redes neuronales para capturar datos atípicos y modelar no-linealidades bajo diferentes simulaciones de procesos generadores de datos. Los autores concluyen que el modelo logra capturar dentro de muestra datos atípicos aditivos, inovativos y de cambio de nivel. Asimismo, mediante simulaciones muestran que la capacidad de capturar no linealidades de la red se da sobre la media, donde, evaluando la calidad de pronóstico de procesos generadores de datos tipo Markov Switching y GARCH, se concluye que la red no presenta mejores pronósticos comparados con un modelo lineal. Así, teniendo en cuenta los resultados de estos experimentos,

Pronóstico del spread de la curva cero cupón

se enmarcan los escenarios en los que la red neuronal resulta conveniente para el pronóstico de variables de series de tiempo.

3.3 Estimación

La aplicación de redes neuronales para el ámbito de series de tiempo es popularmente usada bajo la especificación de una sola capa oculta, además de esto, es común añadir un término lineal en el modelo. El siguiente diagrama ilustra esta estructura:

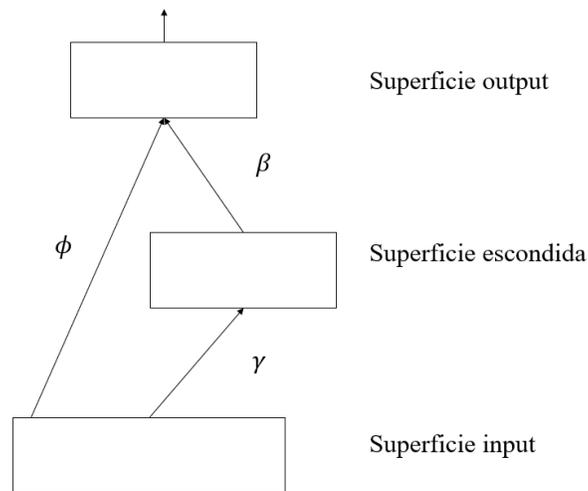


Diagrama de red neuronal incluyendo un término lineal, elaborado por: Misas, Lopez y Querubin (2002)

En términos algebraicos, la anterior imagen se expresa como:

$$y_t = x_t^T \phi + \sum_{j=1}^q \beta_j G(z_t^T \gamma_j) + \varepsilon_t$$

Donde y_t es la variable dependiente en el momento t ; q es el número de nodos que se especifican en el modelo; ϕ es un vector de coeficientes lineales de dimensiones $((1 + mp) \times 1)$ definidos como $\phi^T = [\phi_0 \ \phi_{1,1} \ \dots \ \phi_{m,p}]$, con m actuando como indicador de variable y p como indicador de rezago; x_t es una matriz de diseño para las variables en el término lineal del modelo de dimensiones $((1 + mp) \times 1)$ definida como $x_t^T = [1 \ x_{1,t-1} \ \dots \ x_{m,t-p}]$; β_j es un coeficiente lineal que pondera el resultado de la función de activación j ; $G(\cdot)$ es la función de activación que para propósitos de este trabajo está definida como una función logística $G(z_t^T \gamma_j) = (1 + e^{-z_t^T \gamma_j})^{-1}$; γ_j es el vector de coeficientes no lineales para el nodo j con

Pronóstico del spread de la curva cero cupón

dimensiones $((1 + mp) \times 1)$ definidos como $\gamma_j^T = [\gamma_{0j} \ \gamma_{1,1j} \ \cdots \ \gamma_{m,pj}]$; z_t es otra matriz de diseño que incluye las variables de entrada en los nodos, de dimensión $((1 + mp) \times 1)$ y está definida como $z_t^T = [1 \ z_{1,t-1} \ \cdots \ z_{m,t-p}]$; finalmente, ε_t es el término de error.

El proceso de estimación se realiza mediante mínimos cuadrados no lineales, donde se minimiza la siguiente función objetivo mediante un algoritmo de optimización no lineal:

$$Q_n(\theta) = \sum_{t=1}^n (y_t - F(x_t; z_t; \theta))^2$$

Donde,

$$F(x_t; z_t; \theta) = x_t^T \phi + \sum_{j=1}^q \beta_j G(z_t^T \gamma_j)$$

No obstante, Franses y Van-Dijk (2003) recomiendan prevenir que los parámetros tomen magnitudes muy grandes añadiendo términos de penalización sobre estos parámetros, este método es conocido como *Weight Decay*:

$$Q_n(\theta) = \sum_{t=1}^n (y_t - F(x_t; z_t; \theta))^2 + r_\phi \sum_{i=0}^{m \cdot p} \phi_i^2 + r_\beta \sum_{j=1}^q \beta_j^2 + r_\gamma \sum_{j=1}^q \sum_{i=0}^{m \cdot p} \gamma_{j,i}^2$$

Con $r_\phi = 0.01$ y $r_\beta = r_\gamma = 0.0001$.

Además de esto, para que el *Weight Decay* cumpla su objetivo es necesario normalizar las variables para que puedan estar en una escala comparable entre sí:

$$z_{i,t}^* = \frac{z_{i,t} - \bar{z}_i}{\sigma_{z_i}}$$

Pronóstico del spread de la curva cero cupón

Habiéndose aclarado la estructura de la función objetivo, el entrenamiento de la red neuronal se realiza bajo un algoritmo de optimización no lineal. En este trabajo se utiliza Newton-Raphson con una intervención en el paso denotada lambda (λ), definida desde un algoritmo de búsqueda inexacta de línea. La optimización de Newton-Raphson considera como insumo: (1) un vector de parámetros $\hat{\theta}_n^{(r)}$; (2) un vector score de derivadas parciales $\nabla Q_n(\hat{\theta}_n^{(r)})$, el cual marca la dirección en que se deben mover los parámetros para minimizar la función objetivo; (3) la inversa de la matriz Hessiana $H(\hat{\theta}_n^{(r)})^{-1}$, la cual proporciona de igual forma dirección y magnitud de desplazamiento; y (4) un factor λ que marca el "paso" o la magnitud en que se desplaza el algoritmo en cada iteración, en la literatura se conoce también como el *factor de aprendizaje*.

Con los elementos previamente mencionados, el aprendizaje de la red se da a partir de la actualización de los parámetros bajo la forma:

$$\hat{\theta}_n^{(r+1)} = \hat{\theta}_n^{(r)} - \lambda H(\hat{\theta}_n^{(r)})^{-1} \nabla Q_n(\hat{\theta}_n^{(r)})$$

Habiendo señalado la estimación de una sola red neuronal, es importante considerar que Franses y Van Dijk (2003) mencionan la no existencia de mínimos globales en la función objetivo de la red neuronal. Por esta razón, dentro de la metodología de estimación, es preciso replicar el proceso de estimación explicado anteriormente N veces y seleccionar la réplica cuyo valor de la función objetivo sea el mínimo del conjunto de replicas. La determinación de estas replicas está sujeta a la capacidad computacional que esté al alcance y al tiempo disponible para este proceso de estimación. En esta investigación se utiliza $N = 30$.

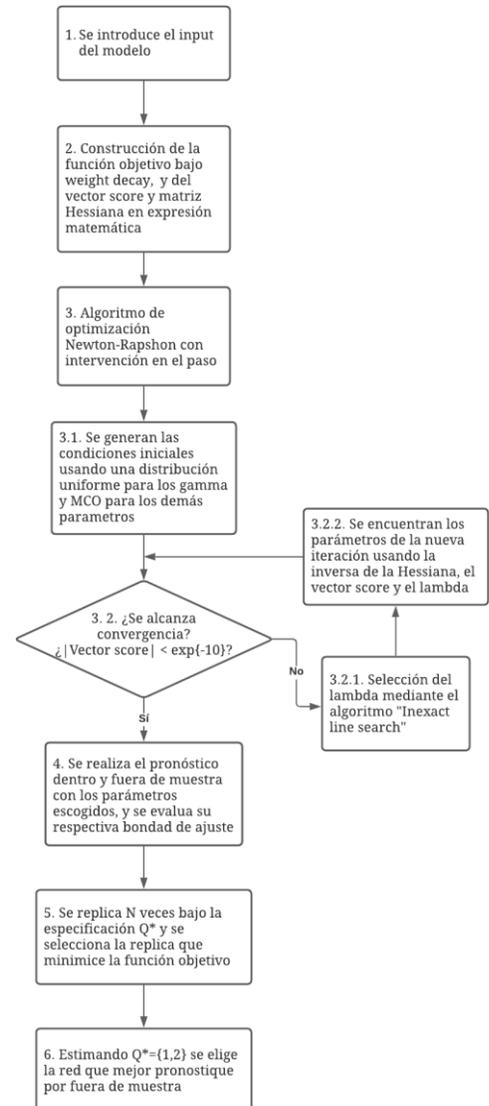


Diagrama de algoritmo de aprendizaje de la red neuronal, elaboración propia.

Pronóstico del spread de la curva cero cupón

4. Datos

Los datos extraídos son de periodicidad diaria y se obtienen para un periodo desde enero de 2010 hasta diciembre de 2019, utilizando la información desde febrero de 2019 para evaluar la calidad de los pronóstico de ambos modelos. Por un lado, se obtienen las tasas cero cupón de 1 y 10 años junto con los coeficientes β_1 y β_2 estimados bajo la metodología de Nelson y Siegel (1987) por el Banco de la República, utilizando como insumo la curva de rendimientos de los TES en pesos. Se utiliza también la tasa interbancaria overnight del Banco de la República para préstamos en pesos. Por otro lado, se obtiene de Bloomberg la tasa LIBOR para préstamos en dólares, los rendimientos de los tesoros a 10 años de Estados Unidos y los precios en puntos básicos de los Credit Default Swaps de los bonos a 10 años de Colombia en dólares.

Tabla 1

Estadísticas descriptivas de las variables

Estadísticas Descriptivas						
	Mínimo	Máximo	Mediana	Media	Desv. Est.	Tamaño
Spread 1/10 COP	-0,021	5,620	2,210	2,289	1,056	2319
IB Overnight COP	2,970	7,826	4,426	4,595	1,263	2319
LIBOR Overnight USD	0,081	2,403	0,218	0,631	0,745	2319
Tasa TES COP 1A	3,579	7,359	4,780	5,011	0,857	2319
β_1	-0,083	-0,010	-0,042	-0,042	0,012	2319
β_2	-0,077	0,144	-0,006	0,005	0,043	2319
CDS COL USD 10A	118,231	386,687	180,645	190,056	48,826	2319
BOND USD 10A	1,359	3,952	2,343	2,395	0,554	2319

COP: Tasa para préstamos en pesos colombianos; USD: tasa para préstamos en dólares estadounidenses; 1A: tasa para títulos con vencimiento a 1 año; 10A tasa para títulos con vencimiento a 10 años; CDS: Credit Default Swap; COL: título colombiano. IB: Tasa para préstamos interbancarios. Elaboración propia.

Inicialmente, se define el spread entre el corto y largo plazo como:

$$spread_t = i_{10años,t} - i_{1año,t} \quad \forall t = 1, \dots, 2319$$

Ahora, se incluyen los factores del modelo de Nelson y Siegel (1987) de pendiente (β_1) y de la curvatura (β_2), la tasa interbancaria para reflejar las condiciones de liquidez local y de política monetaria y la tasa de corto plazo de 1 año siguiendo los resultados de Carriero et al. (2006) donde la inclusión de estas variables lleva al cumplimiento de la teoría de las expectativas. También se utilizan las tasas de 1 año como indicador de expectativas de inflación en el corto

Pronóstico del spread de la curva cero cupón

plazo y de política monetaria siguiendo a Christophers (2017) y el canal de transmisión explicado en la revisión de bibliográfica.

Se aproxima también a las condiciones de liquidez internacional utilizando la tasa LIBOR para préstamos en dólares overnight. Esta variable se utiliza como indicador de liquidez de mercado debido a que, a mayor liquidez, las tasas de interés van a ser más bajas dado que existe una menor demanda de préstamos entre bancos, los cuales son solventes para cumplir con su operación. De forma contraria, si en el mercado se percibe iliquidez, aquellos en capacidad de ser acreedores prestan su dinero a una tasa mayor supliendo el costo de oportunidad que representa y aprovechando el aumento de la demanda de dinero. Teniendo esto en cuenta, la validez de la tasa LIBOR como proxy de la liquidez del mercado internacional nace de la importancia de Londres para el mercado global debido al volumen transaccional que representa en el mercado de divisas, el cual abarcó el 43% del total en 2019 según el Bank of International Settlements (2019).

En cuanto al factor de riesgo internacional, se utilizan las tasas de los tesoros de Estados Unidos a 10 años. Estos títulos se caracterizan como activos refugio o “safe heaven” debido a que están menos expuestos a la volatilidad del mercado, haciéndolos poco riesgosos a nivel crediticio y con un elevado volumen de transacciones en mercado secundario. Teniendo esto en cuenta, las tasas de los tesoros son útiles como proxy del apetito al riesgo debido a que un aumento en su tasa es causada por una menor demanda de estos títulos, y esto puede ocasionarse por un cambio en el comportamiento del mercado global en donde haya un aumento en el apetito al riesgo y una búsqueda de opciones más rentables.

Por último, considerando el factor de riesgo de crédito, las tasas de las permutas de incumplimiento crediticio en dólares estadounidenses para los títulos a 10 años de Colombia se utilizan como proxy de la percepción de riesgo sobre el gobierno colombiano. En términos generales, estos títulos mejor conocidos por su nombre en inglés “Credit Default Swaps” o CDS, son títulos en donde el comprador debe hacer un pago periódico al vendedor de este derivado, en retorno, el comprador recibirá una cantidad de dinero del vendedor en caso de que el activo subyacente, en este caso, los títulos a 10 años de Colombia, incumplan con sus pagos o la entidad emisora de este subyacente se declare en bancarrota.

Pronóstico del spread de la curva cero cupón

Con base en esto, los precios de los CDS de los bonos colombianos son un indicador de volatilidad sobre la deuda colombiana porque una mayor percepción de riesgo de impago conlleva a un aumento en la demanda de estos títulos y en consecuencia a un aumento de sus precios. Vale aclarar que los precios de este derivado reciben influencias de los precios del dólar debido a que si la moneda se devalúa, la deuda no cubierta se encarece, por lo que aumenta la probabilidad de incumplimiento y finalmente aumenta la demanda de estos títulos.

5. Resultados

Dado el conjunto de información previamente descrito, se procede a realizar las pruebas Dickey-Fuller o Dickey-Fuller Aumentada (1979) de acuerdo con la significancia de los rezagos de cada serie. De esta forma, se concluye que todas las variables anteriormente mencionadas son estacionarias en su primera diferencia, vale aclarar que es con estas variables que se realiza la estimación del modelo ARIMAX.

Como se mencionó en la sección 3, se estimaron las redes neuronales considerando $q = \{1, 2\}$ nodos buscando encontrar la mejor red, haciendo la aclaración que se eligió un número de rezagos $p = 1$ para todas las variables; se denota este modelo como $RNA(p, q)$. De esta forma, la estimación de la red neuronal se desarrolla con un código de la elaboración de los investigadores utilizando el lenguaje de programación R. La tabla 2 y 3 muestran los indicadores RMSE y MAE para los pronósticos dentro de muestra y fuera de muestra correspondientes a la información de cada mes de 2019.

Pronóstico del spread de la curva cero cupón

Tabla 2

Indicador RMSE de bondad de ajuste para el año 2019.

Bondad de Ajuste 2019: RMSE				
Horizonte	RNA(1,1)	RNA(1,2)	ARIMA(7,1,8)	ARIMAX(7,1,8)
Dentro de Muestra	0.08384	0.08384	0.07703	0.06877
Febrero	0.03473	0.03472	0.03603	0.02922
Marzo	0.05540	0.05540	0.08266	0.08146
Abril	0.08666	0.08665	0.10290	0.10649
Mayo	0.06272	0.06271	0.07345	0.06569
Junio	0.05761	0.05760	0.24778	0.25729
Julio	0.06788	0.06787	0.49706	0.50593
Agosto	0.09050	0.09049	0.60706	0.61347
Septiembre	0.09447	0.09445	0.54785	0.55711
Octubre	0.07197	0.07195	0.59192	0.59878
Noviembre	0.13117	0.13116	0.42263	0.42933
Diciembre	0.09379	0.09378	0.39374	0.40288

Elaboración propia.

Tabla 3

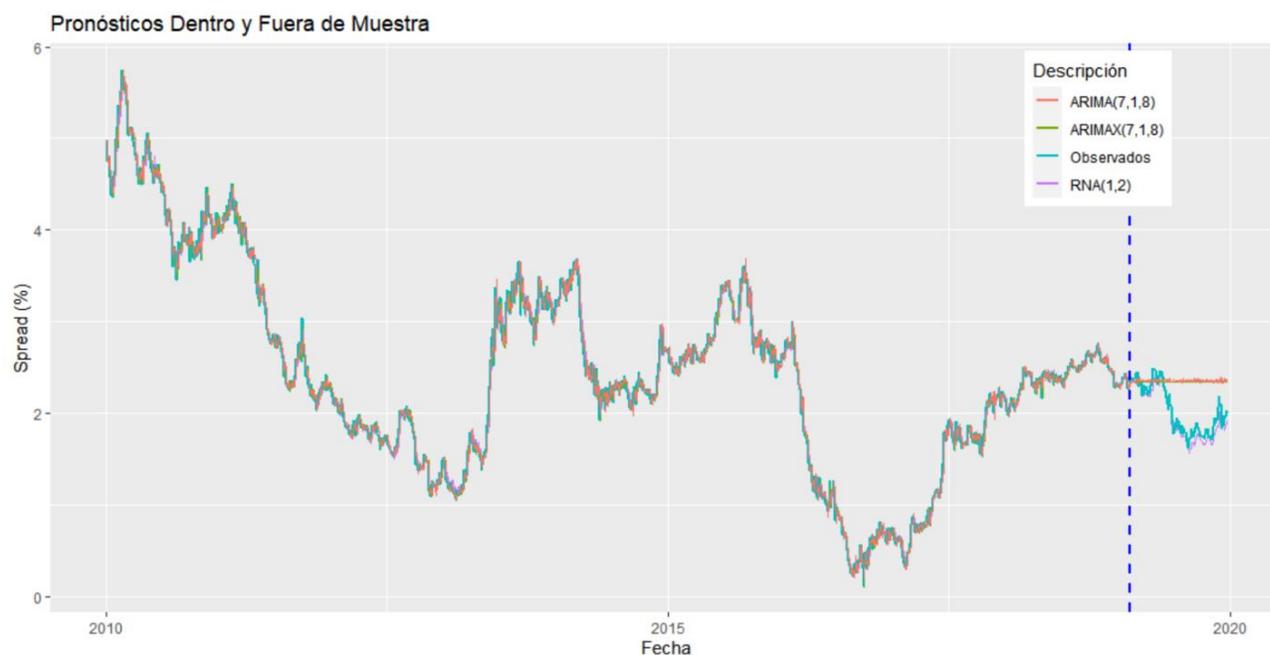
Indicador MAE de bondad de ajuste para el año 2019.

Bondad de Ajuste 2019: MAE				
Horizonte	RNA(1,1)	RNA(1,2)	ARIMA(7,1,8)	ARIMAX(7,1,8)
Dentro de Muestra	0.06115	0.06115	0.05503	0.04941
Febrero	0.02669	0.02668	0.03031	0.02417
Marzo	0.04439	0.04438	0.06962	0.06615
Abril	0.07385	0.07384	0.09426	0.09807
Mayo	0.05349	0.05348	0.06434	0.05834
Junio	0.04035	0.04035	0.23039	0.24172
Julio	0.05624	0.05623	0.49455	0.50334
Agosto	0.07148	0.07146	0.60349	0.60983
Septiembre	0.08478	0.08476	0.54565	0.55521
Octubre	0.06519	0.06518	0.59069	0.59764
Noviembre	0.11943	0.11942	0.40267	0.41012
Diciembre	0.08348	0.08347	0.38756	0.39767

Elaboración propia.

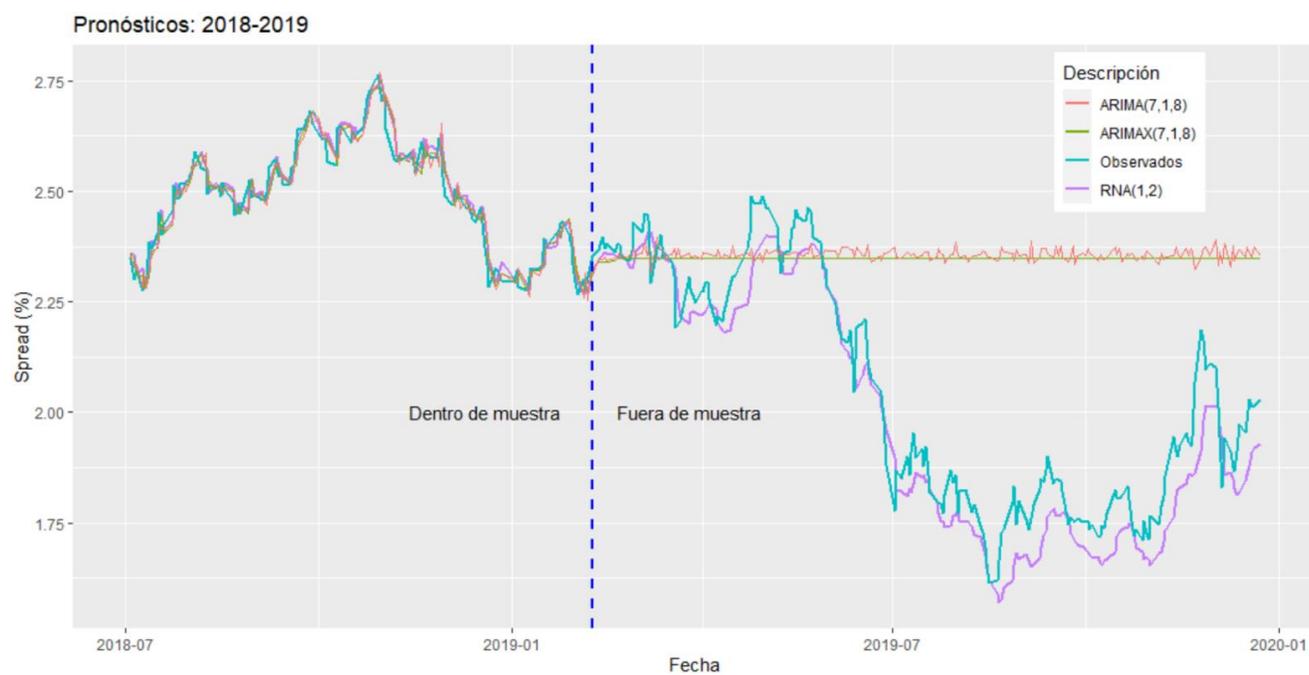
Pronóstico del spread de la curva cero cupón

Gráfica 1



Gráfica de los datos observados, los pronósticos dentro y fuera de muestra, la línea vertical marca el periodo de entrenamiento y de pronóstico. Elaboración propia.

Gráfica 2



Gráfica ampliada de los datos observados, los pronósticos dentro y fuera de muestra entre los años 2018 y 2019, la línea vertical marca el periodo de entrenamiento y de pronóstico. Elaboración propia.

Pronóstico del spread de la curva cero cupón

6. Discusión

En las tablas 2 y 3 se observa que tanto en el indicador RMSE como en el MAE, los modelos ARIMA y ARIMAX superan a las redes neuronales artificiales para los pronósticos dentro de muestra. Así mismo, se hace evidente que, para el primer mes, el modelo ARIMAX minimiza el error de pronóstico por encima de los demás modelos, no obstante, es de notar que su resultado es marginalmente superior al de las redes neuronales. Ahora, considerando horizontes más alejados, ambas redes neuronales logran capturar los movimientos del spread superando a los modelos ARIMA y ARIMAX para este caso particular, siendo que, las redes neuronales sí logran seguir la senda que presenta la variable.

Asimismo, es importante tener en cuenta las diferencias en la información que poseen los modelos. Por un lado, dado el costo en procesamiento computacional que representa añadir rezagos como insumo para el entrenamiento de la red neuronal, solo se tuvo en cuenta el primer rezago de cada variable. Además, no se consideraron más nodos y réplicas para su entrenamiento por el mismo motivo. Por otro lado, los modelos ARIMA y ARIMAX, siendo que, al ser modelos flexibles y no presentar las dificultades de estimación que sí poseen las redes neuronales artificiales, fue posible encontrar aquella combinación de rezagos en el autorregresivo y en las innovaciones que mejorara la calidad de pronóstico. De esta forma, se contrasta el costo computacional de estimar un modelo de redes neuronales en el que se consiga la combinación de variables y rezagos que minimizan el error de pronóstico, similar a lo conseguido con el ARIMAX, dado que, si se quisiera aproximar a este objetivo con el modelo no lineal, se requeriría una inversión en infraestructura tecnológica y una ventana de tiempo para la estimación más amplia.

Por otro lado, retomando de la revisión bibliográfica a Christophers (2017), la autoridad monetaria observa las repercusiones de sus políticas en el mercado, vigilando diferentes indicadores financieros y económicos, entre ellos, el spread. Los pronósticos del modelo de redes neuronales seleccionado consiguen hacer previsión de los movimientos futuros del spread, lo cual, otorga perspectivas e indicios sobre la inflación en el corto plazo, permitiendo tomar decisiones de política monetaria sin importar que en la evaluación por fuera de muestra, existan ocasiones en donde se presenta una diferencia de nivel entre el dato real y el estimado.

Pronóstico del spread de la curva cero cupón

De igual forma, un administrador de portafolios de renta fija encuentra utilidad al observar estos pronósticos porque puede tomar decisiones relevantes sobre la duración de su portafolio. Particularmente, en la gráfica 2, donde se observa una reducción en el spread según el pronóstico, se pueden tener las siguientes consideraciones: (1) Si la reducción en el spread se da por una disminución en la TIR de la tasa de largo plazo, potencialmente habrían rentabilidades a futuro si en el presente se alarga la duración del portafolio; (2) Si el movimiento está dado por un aumento en la TIR de la tasa de corto plazo, los títulos de baja duración del portafolio reducirían su precio, por lo que, de igual forma, sería beneficioso alargar la duración del portafolio para evitar esa pérdida.

7. Conclusión

En resumen, se encuentra que, para el primer mes de pronóstico, el modelo ARIMAX supera marginalmente a las redes neuronales artificiales. No obstante, las redes neuronales obtienen mejores resultados en los horizontes más alejados siendo estas capaces de predecir la senda futura del spread. Además, se destaca a la red neuronal con dos nodos por encima a la de un nodo, teniendo un desempeño equivalente o superior según los indicadores de bondad de ajuste para los periodos posteriores al mes de marzo.

En consecuencia, el modelo de redes neuronales bajo la especificación de dos nodos estimado en esta investigación, puede llegar a ser de gran utilidad para un administrador de portafolios o para una autoridad monetaria, logrando seguir los movimientos del spread en un contexto de alta frecuencia. Es importante recalcar la ventana de tiempo y el costo en procesamiento computacional necesario para estimar este modelo, siendo que, se estableció como estrategia para atenuar estas dificultades, la inclusión de un rezago y dos nodos en el modelo, al igual que se consideró únicamente un número de réplicas igual a 30. Por lo que, para futuras investigaciones o aplicaciones prácticas, si se quiere reducir la discrepancia presentada en la evaluación por fuera de muestra, se recomienda utilizar una estrategia stepwise para la elección de las variables y el número de rezagos, incluyendo también, un mayor número de nodos y de réplicas.

8. Referencias

- Afonso, A., Furceri, D., & Gomes, P. (2011). Sovereign Credit Ratings and Financial Markets Linkages: Application to European Data. Retrieved 25 September 2021, from <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1347.pdf>.
- Álvarez, M. R., Ramírez, A., Rendón, A. (2010). La curva de rendimientos como un indicador adelantado de la actividad económica, el caso colombiano: Período 2001-2009. *Ecos de Economía*, 31(14), 36-63.
- Andrews, B., Dean, M., Swain, R., & Cole, C. (2013). Building ARIMA and ARIMAX models for predicting long-term disability benefit application rates in the public/private sectors. Retrieved 31 October 2021, from <https://www.soa.org/globalassets/assets/files/research/projects/research-2013-arima-arimax-ben-appl-rates.pdf>.
- Bank of International Settlements. (2019). Triennial Central Bank Survey [Video]. Recuperado de: https://www.youtube.com/watch?v=_WMYs4Whzms&t=63s.
- Barrios, S., Iversen, P., Lewandowska, M., & Setzer, R. (2009). Determinants of intra-euro Area Government Bond Spreads During the Financial Crisis. Retrieved 25 September 2021, from https://ec.europa.eu/economy_finance/publications/pages/publication_summary16253_en.htm.
- Beber, A., Brandt, M., & Kavajecz, K. (2006). Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.891736>
- Campbell, J. Y., Shiller, R. J. (1987). Cointegration and tests of present value models. *Journal of Political Economy* 95(5): 1062-1088. doi: /10.1086/261502.
- Campbell, L. (2014) 'Wealth Management: Recession and the Predictive Yield Curve', [Online] Available at: <http://www.dailyrecord.co.uk/business/business-features/wealth-management-recession-predictive-yield3085386#hA5UTw8EQsiqT8ow.97>

Pronóstico del spread de la curva cero cupón

- Carriero, A., Favero, C. A., & Kaminska, I. (2006). Financial factors, macroeconomic information and the Expectations Theory of the term structure of interest rates. *Journal of Econometrics*, 131(1–2), 339–358. Doi: /10.1016/J.JECONOM.2005.01.019.
- Christophers, B. (2017). The performativity of the yield curve. *Journal of Cultural Economy*, 10(1), 63-80. doi: 10.1080/17530350.2016.1236031.
- Dickey, D. A., Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74(366a), 427–431.
- Enders, W. (2014). *Applied Econometric Time Series*, 4th Edition. Hoboken: Wiley.
- Estrella, A., Hardouvelis, G. A. (1991). The Term Structure as a Predictor of Real Economic Activity. *The Journal of Finance*, 46(1), 555-576. Doi: 10.1111/j.1540-6261.1991.tb02674.x
- Franses, P. H. & Van-Dijk, D. (2003). *Nonlinear Time Series Models in Empirical Finance*. Cambridge: Cambridge University Press.
- Geyer, A., Kossmeier, S., & Pichler, S. (2003). Measuring Systematic Risk in EMU Government Yield Spreads. SSRN Electronic Journal. <https://doi.org/10.2139/ssrn.394522>
- Hall, J. E. (2015). *Guyton and hall textbook of medical physiology* (13th ed.). W B Saunders.
- Haugh, D., Ollivaud, P., & Turner, D. (2009). What Drives Sovereign Risk Premiums?. OECD Economics Department Working Papers. <https://doi.org/10.1787/222675756166>
- Hay, A. (2019). Perceptron Model. [Imagen]. Github. <https://alexanderhay2020.github.io/blog/2019/perceptrons/>
- Hochberg, Y. (1988). A Sharper Bonferroni Test for Multiple Tests of Significance. *Biometrika*, 75(1), 800-802. Doi: 10.1093/biomet/75.4.800
- Holló, D., Kremer, M., & Lo Duca, M. (2012). CISS - A Composite Indicator of Systemic Stress in the Financial System. Retrieved 25 September 2021, from <https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1426.pdf>.

Pronóstico del spread de la curva cero cupón

Hull, J. C. (2015). *Options, Futures, And Other Derivatives*. Pearson.

Jankowitsch, R., Mösenbacher, H., & Pichler, S. (2002). Measuring the Liquidity Impact on EMU Government Bond Prices. *SSRN Electronic Journal*.

<https://doi.org/10.2139/ssrn.302795>.

Lee, T. H., White, H., Granger, C. W. J. (1993). Testing for neglected nonlinearity in time series models: A comparison of neural network methods and alternative tests. *Journal of Econometrics*, 56(3), 269-290. Doi: 10.1016/0304-4076(93)90122-L.

Lee, T.H., White, H. C., Granger W.J. (1993). Testing for neglected nonlinearity in time series models. *Journal of Econometrics*, 56(1), 269-290.

Misas, M., Lopez, E., Querubin, P. (2002). La inflación en Colombia: una aproximación desde las redes neuronales. *Revista Ensayos Sobre Política Económica*, 20(41-42), 143-214. doi: 10.32468/Espe.41-4203.

Nelson, C. R., Siegel, A. F. (1987). Parsimonious modelling of yield curves. *Journal of Business* 60(4): 473–489.

Sgherri, S., & Zoli, E. (2009). Euro Area Sovereign Risk During the Crisis. Retrieved 25 September 2021, from <https://www.imf.org/en/Publications/WP/Issues/2016/12/31/Euro-Area-Sovereign-Risk-During-the-Crisis-23325>.

Shatbolt, J., Taylor, J. G. (2002). *Neural Networks and the Financial Markets*. London: Springer Publishing Company Limited.

Shiller, R. J., Campbell, J. Y., Shoenholtz, K. L. (1983). Forward rates and future policy: interpreting the term structure of interest rates. *Brookings Papers Economic Activity*, 1(1): 173-217.

Swanson, N.R., White, H. (1997). A models selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *The Review of Economics and Statistics* 79(4), 540–550. doi: 10.1162/003465397557123.

Anexo 1: Notación Matricial de los Componentes del Optimizador

$$\hat{\theta}_n^{(r)} = \begin{bmatrix} \phi \\ \beta \\ \gamma \end{bmatrix}_{((mp+1)+q(mp+2)) \times 1} \quad \nabla Q_n(\hat{\theta}_n^{(r)}) = \begin{bmatrix} \frac{\partial}{\partial \phi} Q_n(\hat{\theta}_n^{(r)}) \\ \frac{\partial}{\partial \beta} Q_n(\hat{\theta}_n^{(r)}) \\ \frac{\partial}{\partial \gamma} Q_n(\hat{\theta}_n^{(r)}) \end{bmatrix}_{((mp+1)+q(mp+2)) \times 1}$$

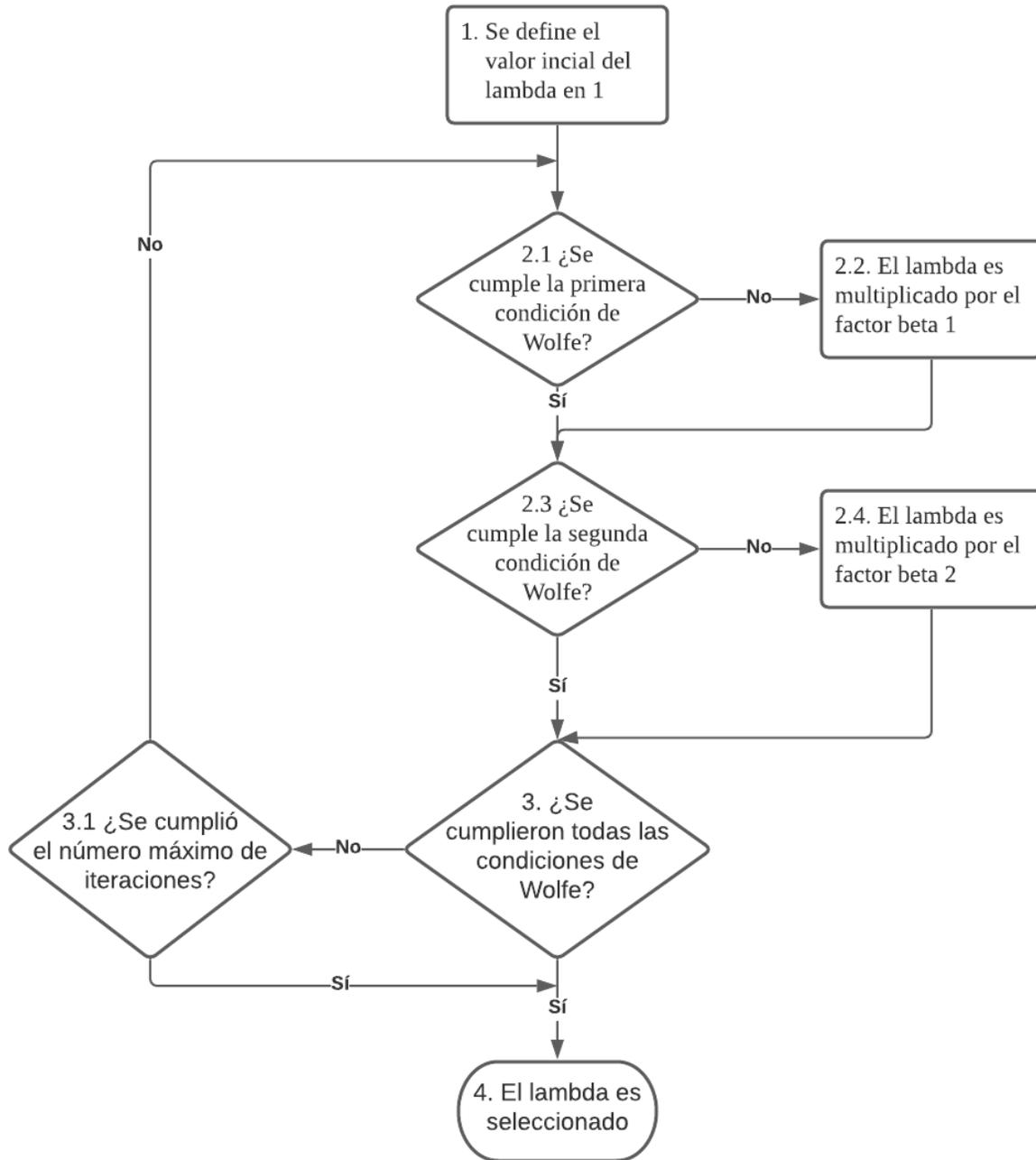
$$H(\hat{\theta}_n^{(r)}) = \begin{bmatrix} \frac{\partial^2}{\partial \phi \partial \phi} Q_n(\hat{\theta}_n^{(r)}) & \frac{\partial^2}{\partial \beta \partial \phi} Q_n(\hat{\theta}_n^{(r)}) & \frac{\partial^2}{\partial \gamma \partial \phi} Q_n(\hat{\theta}_n^{(r)}) \\ \frac{\partial^2}{\partial \phi \partial \beta} Q_n(\hat{\theta}_n^{(r)}) & \frac{\partial^2}{\partial \beta \partial \beta} Q_n(\hat{\theta}_n^{(r)}) & \frac{\partial^2}{\partial \gamma \partial \beta} Q_n(\hat{\theta}_n^{(r)}) \\ \frac{\partial^2}{\partial \phi \partial \gamma} Q_n(\hat{\theta}_n^{(r)}) & \frac{\partial^2}{\partial \beta \partial \gamma} Q_n(\hat{\theta}_n^{(r)}) & \frac{\partial^2}{\partial \gamma \partial \gamma} Q_n(\hat{\theta}_n^{(r)}) \end{bmatrix}_{((mp+1)+q(mp+2)) \times ((mp+1)+q(mp+2))}$$

Anexo 2: Condiciones de Wolfe para el algoritmo “Inexact Line Search”

1. Primera condición de Wolfe: $f(x + \lambda H^{-1} \nabla(x)) \leq f(x) + \lambda \alpha_1 \nabla(x)^T H^{-1} \nabla(x)$

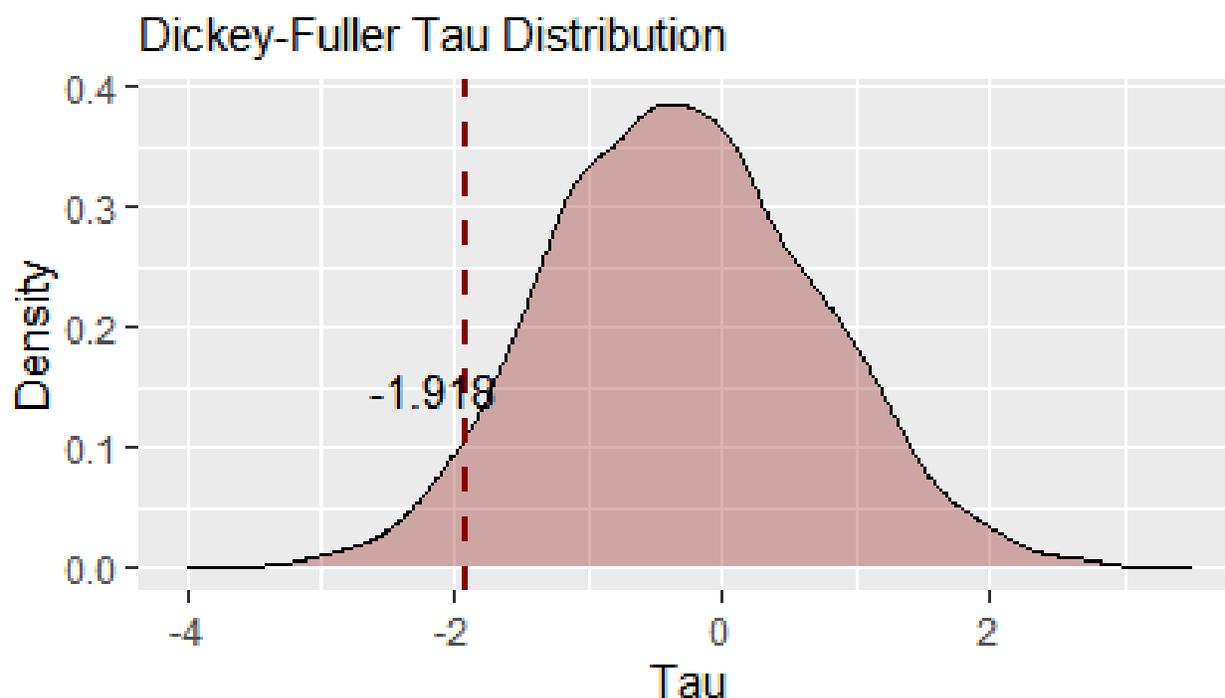
2. Segunda condición de Wolfe: $\frac{\nabla(x + \lambda \nabla(x))^T H^{-1} \nabla(x)}{\nabla(x)^T H^{-1} \nabla(x)} \leq \alpha_2$

Anexo 3: Diagrama del algoritmo de “Inexact Line Search”



Pronóstico del spread de la curva cero cupón

Anexo 4: Distribución τ simulada para un tamaño de muestra de 2319



Anexo 5: Pruebas de Estacionariedad

Pruebas de Estacionariedad					
Niveles			Diferencias		
Prueba	Ljung-Box		Prueba	Ljung-Box	
<i>Spread</i>					
$t = -1.6628$	8 rezagos	Revisión con 36 lags	$t = -15.6772$	7 rezagos	Revisión con 36 lags
$\tau = -1.917626$	ADF	P valor = 0.4602051	$\tau = -1.917626$	ADF	P valor = 0.4704109
<i>Interbancaria Overnight</i>					
$t = 0.0816$	5 rezagos	Revisión con 36 lags	$t = -29.9825$	4 rezagos	Revisión con 36 lags
$\tau = -1.917626$	ADF	P valor = 0.3107552	$\tau = -1.917626$	ADF	P valor = 0.3114121
<i>Libor Overnight Dólar</i>					
$t = 1.0219$	0 rezagos	Revisión con 36 lags	$t = -46.3682$	0 rezagos	Revisión con 36 lags
$\tau = -1.917626$	DF	P valor = 0.6801604	$\tau = -1.917626$	DF	P valor = 0.8164607
<i>Tasas Cero Cupón de 1 año</i>					
$t = -0.2859$	10 rezagos	Revisión con 36 lags	$t = -15.3785$	9 rezagos	Revisión con 36 lags
$\tau = -1.917626$	ADF	P valor = 0.4952582	$\tau = -1.917626$	ADF	P valor = 0.4951808
β_1					
$t = -0.9696$	15 rezagos	Revisión con 36 lags	$t = -12.9111$	14 rezagos	Revisión con 36 lags
$\tau = -1.917626$	ADF	P valor = 0.1198267	$\tau = -1.917626$	ADF	P valor = 0.1198125
β_2					

Pronóstico del spread de la curva cero cupón

$t = -2.2271$	8 rezagos	Revisión con 36 lags	$t = -20.1325$	7 rezagos	Revisión con 36 lags
$\tau = -1.917626$	ADF	P valor = 0.1717125	$\tau = -1.917626$	ADF	P valor = 0.1708797
CDS 10Y USD					
$t = -0.6158$	14 rezagos	Revisión con 36 lags	$t = -14.8537$	13 rezagos	Revisión con 36 lags
$\tau = -1.917626$	ADF	P valor = 0.2788552	$\tau = -1.917626$	ADF	P valor = 0.2794144
Tasas Títulos a 10 Años de Estados Unidos					
$t = -1.3903$	5 rezagos	Revisión con 36 lags	$t = -22.681$	4 rezagos	Revisión con 36 lags
$\tau = -1.917626$	ADF	P valor = 0.5851818	$\tau = -1.917626$	ADF	P valor = 0.5824938

La tabla muestra los resultados de las pruebas Dickey Fuller o Dickey Fuller Aumentado para las variables seleccionadas, en la sección Ljung-Box se evidencia el numero de rezagos para la obtención de ruido blanco y el p valor de la prueba para el último lag (36) revisado para esta.

Anexo 6: Resultado Modelo ARIMA

Resultados ARIMA				
	Coficiente	Error Est.	Z	P-Valor
Δy_{t-1}	-0,142379	0,021609	-6,5888	4,43E-11
Δy_{t-4}	0,067718	0,02159	3,1366	0,0017093
Δy_{t-7}	0,080407	0,021565	3,7286	0,0001925
a_{t-5}	-0,03744	0,021872	-1,7118	0,0869401
a_{t-8}	0,068741	0,02223	3,0923	0,0019862
Ruido blanco				
	Prueba	P-Valor	Lag	
	Ljung-Box	0,623027	36	

Anexo 7: Resultado Modelo ARIMAX

Resultados ARIMAX				
	Coficiente	Error Est.	Z	P-Valor
Δy_{t-4}	0,0497	0,0217	2,2863	0,0222
Δy_{t-7}	0,0760	0,0219	3,4705	0,0005
a_{t-5}	-0,0382	0,0220	-1,7367	0,0824
a_{t-8}	0,0741	0,0222	3,3400	0,0008
$\Delta i_{TES1Y,t-1}$	-0,3930	0,0163	-24,1625	0,0000
$\Delta P_{CDS COL 10Y,t-1}$	0,0012	0,0003	4,6223	0,0000
Ruido blanco				
	Prueba	P-Valor	Lag	
	Ljung-Box	0,4288137	36	

Pronóstico del spread de la curva cero cupón

Anexo 8: Coeficientes estimados del Modelo RNA(1,1)

RESULTADOS RNA (1,1)			
Variable	Coeficiente	Variable	Coeficiente
Intercepto (ϕ_0)	-3,165833e-03	Sesgo (γ_0)	-2,316608e-05
y_{t-1} (ϕ_1)	5,656113e-01	y_{t-1} (γ_1)	8,203896e-03
$i_{IB,t-1}$ (ϕ_2)	-2,779579e-02	$i_{IB,t-1}$ (γ_2)	-4,025959e-04
$i_{LIBOR,t-1}$ (ϕ_3)	-1,650279e-02	$i_{LIBOR,t-1}$ (γ_3)	-2,393766e-04
$i_{TES1Y,t-1}$ (ϕ_4)	-1,442480e-02	$i_{TES1Y,t-1}$ (γ_4)	-2,088135e-04
$\beta_{1,t-1}$ (ϕ_5)	-2,365689e-01	$\beta_{1,t-1}$ (γ_5)	-3,431116e-03
$\beta_{2,t-1}$ (ϕ_6)	2,482484e-01	$\beta_{2,t-1}$ (γ_6)	3,600081e-03
$P_{CDS\ COL,t-1}$ (ϕ_7)	2,204857e-02	$P_{CDS\ COL,t-1}$ (γ_7)	3,198658e-04
$i_{USA10Y,t-1}$ (ϕ_8)	2,885654e-02	$i_{USA10Y,t-1}$ (γ_8)	4,173619e-04
$\beta_{Nodo\ 1}$	5,802818e-04		

Anexo 9: Coeficientes estimados del Modelo RNA(1,2)

RESULTADOS RNA (1,2)					
Variable	Coeficiente	Variable	Coeficiente	Variable	Coeficiente
Intercepto (ϕ_0)	-3.513031e-03	Sesgo ($\gamma_{1,0}$)	-5.830208e-05	Sesgo ($\gamma_{2,0}$)	-5.835821e-05
y_{t-1} (ϕ_1)	5.656105e-01	y_{t-1} ($\gamma_{1,1}$)	9.062032e-03	y_{t-1} ($\gamma_{2,1}$)	9.063201e-03
$i_{IB,t-1}$ (ϕ_2)	-2.779549e-02	$i_{IB,t-1}$ ($\gamma_{1,2}$)	-4.449613e-04	$i_{IB,t-1}$ ($\gamma_{2,2}$)	-4.446685e-04
$i_{LIBOR,t-1}$ (ϕ_3)	-1.650159e-02	$i_{LIBOR,t-1}$ ($\gamma_{1,3}$)	-2.659812e-04	$i_{LIBOR,t-1}$ ($\gamma_{2,3}$)	-2.643657e-04
$i_{TES1Y,t-1}$ (ϕ_4)	-1.442456e-02	$i_{TES1Y,t-1}$ ($\gamma_{1,4}$)	-2.300721e-04	$i_{TES1Y,t-1}$ ($\gamma_{2,4}$)	-2.308029e-04
$\beta_{1,t-1}$ (ϕ_5)	-2.365683e-01	$\beta_{1,t-1}$ ($\gamma_{1,5}$)	-3.792506e-03	$\beta_{1,t-1}$ ($\gamma_{2,5}$)	-3.790751e-03
$\beta_{2,t-1}$ (ϕ_6)	2.482479e-01	$\beta_{2,t-1}$ ($\gamma_{1,6}$)	3.975583e-03	$\beta_{2,t-1}$ ($\gamma_{2,6}$)	3.976403e-03
$P_{CDS\ COL,t-1}$ (ϕ_7)	2.204784e-02	$P_{CDS\ COL,t-1}$ ($\gamma_{1,7}$)	3.550537e-04	$P_{CDS\ COL,t-1}$ ($\gamma_{2,7}$)	3.531550e-04
$i_{USA10Y,t-1}$ (ϕ_8)	2.885518e-02	$i_{USA10Y,t-1}$ ($\gamma_{1,8}$)	4.607088e-04	$i_{USA10Y,t-1}$ ($\gamma_{2,8}$)	4.603201e-04
$\beta_{Nodo\ 1}$	6.409175e-04	$\beta_{Nodo\ 2}$	6.411510e-04		