

UNIVERSIDAD DE LOS ANDES

PROYECTO FINAL MACHINE LEARNING

Predicción de la Curva de Rendimientos Colombiana empleando redes neuronales recurrentes

Autor:

Santiago Moreno

Diego Salazar

Juan Manuel Mora

Pablo Cabra



UNIVERSIDAD DE LOS ANDES

Abstract

Facultad de Economía

Predicción de la Curva de Rendimientos Colombiana empleando redes neuronales recurrentes

by Santiago Moreno

Diego Salazar

Juan Manuel Mora

Pablo Cabra

El objetivo de este proyecto es realizar un modelo por medio de la metodología Redes Neuronales LSTM que obtenga una predicción robusta de la curva de rendimientos cero cupón de los TES tipo B colombianos, utilizando información histórica entre el periodo comprendido entre 2002 y la actualidad, el cual permita superar la predicción de un modelo de caminata aleatoria en el corto plazo (5-10 días). Este proyecto entra en relevancia debido a la implementación de un ETF (Exchange Traded Fund) de la curva soberana por parte del Gobierno Nacional, por lo cual se espera que bajo nuestro análisis se logren resultados que permitan predecir el comportamiento de la curva soberana y en ese sentido lograr una rentabilidad en el instrumento previamente enunciado.

Palabras clave: Curva de rendimientos, Machine Learning, LSTM, Nelson-Siegel

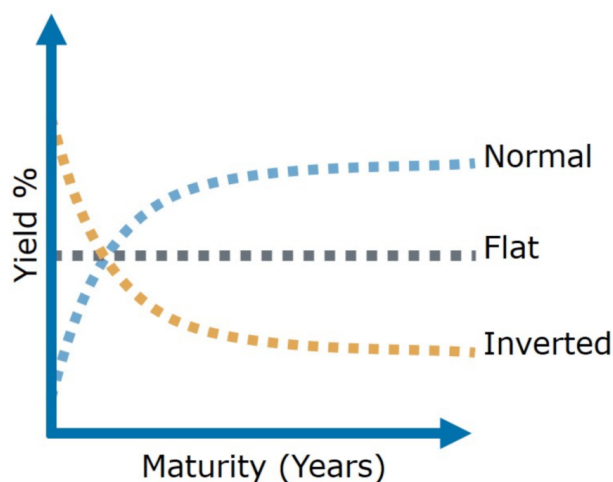
Contents

Abstract	i
1 Introducción	1
2 Datos	3
3 Revisión de la Literatura	5
4 Algoritmo y resultados	7
5 ETF, valoración con nuestras predicciones	11
6 Conclusiones	13
7 Bibliografía	14
8 Anexos	15

Chapter 1

Introducción

La curva de rendimientos o estructura a termino, hace referencia a la relación entre la tasa de interés de los bonos y sus diferentes vencimientos. Estos bonos, TES en el caso colombiano, son títulos de deuda pública del gobierno nacional que son emitidos con el fin de financiarse, por lo cual pagan unos rendimientos al poseedor del mismo en un vencimiento establecido en el momento de la emisión y una tasa de interés durante la vida útil del bono que regularmente se paga semestralmente.



Retomada de: <https://www.sersansistemas.com/la-curva-de-rendimiento-por-lluis-benguere/>

La curva además, permite no sólo evidenciar los retornos acordes a vencimientos no observables, sino que además es una importante fuente de información acerca del estado financiero de los mercados; por ejemplo, es usada en la toma de decisiones de política monetaria debido a que brinda información sobre las expectativas de inflación, crecimiento y estabilidad de la economía en cuestión, resulta también útil para los traders que diariamente venden y compran títulos de deuda pública dado que esta puede presentar oportunidades de arbitraje.

Por otro lado, la curva de rendimientos ha sido objeto de debate sobre su capacidad predictiva de crisis económicas, siendo el caso de Estados Unidos una de las relaciones más evidentes entre una inversión de la estructura a plazos y una futura caída de los mercados; más específicamente, la Reserva Federal de St. Louis afirma que en promedio toma 15 meses entre una inversión de la curva, es decir cuando los intereses de largo plazo son inferiores a los obtenidos en el corto plazo, y una recesión, esto sin causalidad empírica hallada, según Moreno (2018).

Por estas razones la capacidad de lograr una buena estimación y buen pronóstico sobre la curva será de altísima importancia en el sistema financiero moderno, por lo que no es de extrañar que a lo largo del tiempo se hayan desarrollado metodologías que busquen capturar el comportamiento de la curva. Entre las metodologías más relevantes resaltan los procesos autor-regresivos, sin embargo debido a la naturaleza de las series financieras, en específico de los retornos de los activos, muchos de los supuestos sobre normalidad, ergodicidad y heterocedasticidad se ven violados, generando sesgos en las estimaciones por métodos estadísticos convencionales.

Esta problemática brinda la oportunidad de que el aprendizaje de máquinas pueda generar mejores predicciones debido a la capacidad para obtener patrones no observables por medio de las metodologías convencionales y que en el caso de Colombia no han sido extensamente estudiadas pero si podrían brindar aproximaciones diferentes a aquella gran porción de agentes de la economía que realizan operaciones financieras o toman decisiones de inversión basados en el comportamiento futuro de la curva de rendimientos de TES nominales colombianos.

Chapter 2

Datos

Con el fin de generar predicciones sobre el comportamiento futuro de la estructura a término, se utilizará la función paramétrica de Nelson-Siegel para representar la curva de rendimientos colombiana en el tiempo, permitiendo así tener un modelo más parsimonioso al reducir la dimensionalidad y que facilite obtener las diferentes tasas de interés para los distintos vencimientos, además simplificará el entrenamiento del modelo al limitar la predicción a 4 variables, los factores latentes β_0 , β_1 , β_2 y τ que se explicarán a continuación.

Parametrización Nelson-Siegel

$$R(0, t) = \beta_0 + \beta_1 \left(\frac{1 - e^{(-t/\tau)}}{t/\tau} \right) + \beta_2 \left(\frac{1 - e^{(-t/\tau)}}{t/\tau} - e^{(-t/\tau)} \right)$$

Los factores latentes de la curva de rendimientos por medio de esta función serán entonces tres; el factor Nivel o β_0 , será aquel que afectara a todos los vencimientos en la misma proporción y que en la literatura se relaciona con la inflación dado que teóricamente esta afecta las tasas de interés nominal cumpliendo la ecuación de Fisher; el factor pendiente o β_1 , aquel que afecta la curva de rendimientos fuertemente en el corto plazo de la curva pero a medida que aumentan los vencimientos va perdiendo fuerza, este puede también interpretarse como ese trade-off entre el corto y el largo plazo; la curvatura o β_2 , será el que afecte la curva principalmente en el mediano plazo. Mientras que el factor τ brinda información sobre la posición de la curvatura y la pendiente.

Ahora, respecto a la obtención de los datos estos son publicados diariamente en el Banco de la República basados en las transacciones del Sistema Electrónico de Negociación (SEN) y el Mercado Electrónico Colombiano (MEC), estos datos poseen una gran ventaja y es que se estiman bajo la metodología de Nelson-Siegel manteniendo el factor τ constante, reduciendo aún más las variables a usar en el modelo. Son datos de frecuencia diaria, publicados desde el año 2003 y con características puramente transaccionales, poseen una manipulación que afortunadamente va en línea con el objetivo de este documento, siendo adecuados para su uso.

En cuanto a las características propias de estos factores latentes calculados por el Banco de la República se hallaron altos niveles de autocorrelación serial, lo cual supone una ventaja en la predicción de los mismos debido a la naturaleza autoregresiva presente en las series de tiempo de interés y por lo cual no es de extrañar la recurrencia de modelos AR(p) y VAR(p) para su pronóstico por parte de la literatura (Los correlogramas correspondientes se encuentran en la sección Anexos). Además, con el fin de simplificar el entrenamiento de la red neuronal propuesta, se realizó una normalización por medio de la metodología Min-Max a las variables β_0 , β_1 , β_2 .

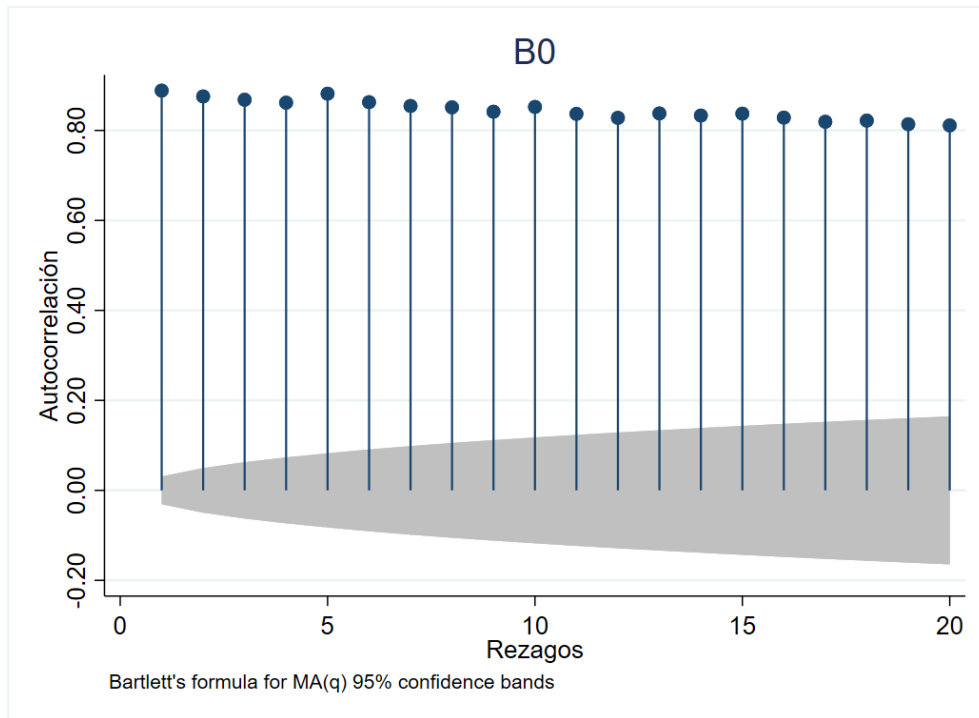


Figura 1. Rezagos para β_0

Chapter 3

Revisión de la Literatura

Desde que apareció la metodología de Nelson y Siegel ha sido la más popular en la literatura y en la práctica. Actualmente sigue vigente y con la aparición del aprendizaje de máquinas se ha visto complementada por distintas herramientas y en particular por redes neuronales, pues pueden considerar no linealidad sin conocer el proceso de generación de los datos (Adhikari, R, 2013). La literatura al respecto no es tan amplia como la que involucra modelos autorregresivos simples AR o vectores autorregresivos VAR, pero ya hay ejemplos valiosos de los que se puede aprender. Una cuestión importante es el hecho de que los diferentes estudios en la literatura no han sido consistentes entre la efectividad del aprendizaje de máquinas y de los modelos econométricos para las series financieras (Adhikari, R, 2013). En ese sentido, este proyecto contribuye a crear una nueva forma de predecir la curva spot que sea tentativamente más eficiente que los modelos AR o VAR, entre otros.

Una de las investigaciones más completas trata el cálculo de la curva spot de Hungría. El paper de Szenczi (2016) usa redes neuronales para predecir las tasas de rendimiento, entrenando su modelo con el algoritmo Levenberg-Marquardt, que se usa para minimizar los errores cuadrados no lineales. Como principal resultado encontró que con el modelo de redes neuronales obtenía una mejor predicción de los rendimientos para 5 y 10 días para todas las ventanas de tiempo contempladas. Ahora bien, cuando la predicción era de solo un día el mejor modelo era un Random Walk. Una segunda investigación relevante es aquella realizada por Daniel Vela, miembro del Departamento de Estudios Económicos del Banco de la República. En este documento, el autor analiza la capacidad de predicción de distintas metodologías para explicar las curvas de rendimiento soberanas de cuatro países latinoamericanos (Colombia, México, Perú y Chile). Para especificar los parámetros, el autor compara dos metodologías similares: la metodología de Nelson-Siegel

y la metodología de Nelson-Siegel-Svensson. Esta segunda metodología es igual a la de Nelson-Siegel (explicada anteriormente), salvo por el hecho que agrega dos parámetros nuevos, γ y τ . En teoría, la inclusión de estos dos parámetros proporciona un mejor ajuste a los datos pues permite que el modelo incorpore una segunda curvatura para explicar la curva de rendimiento. Dicho de otra manera, la metodología Nelson-Siegel-Svensson considera el intercepto (nivel), la pendiente, y hasta dos curvaturas como los factores que determinan la forma de la curva.

El autor contrasta cuatro metodologías de proyección: modelos autoregresivos AR(1), modelos vectoriales autoregresivos VAR(1), redes neuronales artificiales sobre cada factor individual y redes neuronales artificiales sobre todos los factores al tiempo. El análisis del autor muestra resultados diferentes para cada país considerado. En el caso de Colombia, para pronósticos de uno a tres meses, se encontró que la combinación que arroja los pronósticos con la menor variación (de acuerdo con la métrica RMSE) es la metodología de Nelson-Siegel con redes neuronales artificiales sobre cada factor individual. Sin embargo, para horizontes de proyección más extensos (seis meses), el autor concluye que el modelo de caminata aleatoria genera las mejores proyecciones. Vale la pena resaltar que tanto la metodología de Nelson-Siegel como la de Nelson-Siegel-Svensson presentaron valores similares para el nivel y para la pendiente de la curva en todos los países estudiados.

Uno de los principales resultados del documento es que las redes neuronales artificiales, para horizontes de proyección de un mes, tuvieron mejor desempeño fuera de muestra en todas las curvas de rendimiento analizadas, superando a los modelos autoregresivos y divergiendo significativamente de los resultados de una caminata aleatoria. Para el caso particular de Colombia, se identifica que las metodologías que requieren numerosas estimaciones de parámetros son problemáticas pues existe una interacción cruzada muy baja entre los factores. En general, para las curvas de rendimiento latinoamericanas, la metodología de proyección óptima en el corto plazo es la de redes neuronales artificiales sobre cada factor individual de la especificación proporcionada por Nelson-Siegel. Esto contrasta con el caso estadounidense, para el cual las especificaciones de la metodología Nelson-Siegel-Svensson muestran ser más apropiadas.

Chapter 4

Algoritmo y resultados

Para este proyecto se emplearon las redes neuronales artificiales recurrentes (RNN) en dos versiones, la primera denominada Vanilla y la segunda también basada en nodos recurrentes denominada bidireccional. Las redes neuronales recurrentes son un tipo de red capaz de reconocer patrones en datos de tipo secuencial, como una serie de tiempo o genomas, es decir, tienen en cuenta una dimensión temporal. Para lograr esto, la red neuronal recurrente no solo toma como entrada un vector de valores $X = (X_1, \dots, X_n)$ sino que tiene presente lo que percibió en un tiempo anterior cercano. Específicamente, la red que empleamos corresponde una red con nodos de tipo Long-Short Term Memory, las cuales introducen una serie de puertas que permiten que la información perdure o se omita. Inicialmente, la información fluye a través del estado de la celda (C) que recopila la información de los diferentes inputs X_t , a través de operaciones de matrices, es decir, multiplicación por elementos o adición (parte superior en la figura). La información recopilada viene de tres puertas o gates, la primera puerta del olvido ("forget gate layer") decide si el siguiente input $X_{(t)}$ debe ser tomado en cuenta o no. La siguiente puerta decide cual información debe ser incluida en el estado de la celda, primero una función sigmoidea decide que información debe ser actualizada y una función tangente hiperbólica dice cuales serian los valores a actualizar sobre C . Finalmente, realizamos la predicción empleando una función tangente hiperbólica sobre el C actualizado y aplicamos una multiplicación por elementos con la salida de una función sigmoidea para determinar que información hará parte de la predicción. Se debe tener en cuenta que la función sigmoidea en las tres puertas cumple la función de seleccionar la información relevante para ser procesada (Cui et al, 2018). Adicional, en la figura se debe omitir que no estamos generando un h_t , sino que nos interesa mas la predicción de la ultima capa, que corresponde a la predicción del siguiente día. Ver siguiente figura donde se representa la información descrita.

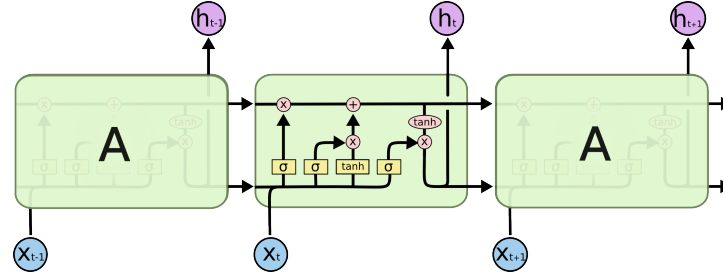


Figura 2. Visualización de una red LSTM (RNN)

Retomada de: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

La red neuronal bidireccional (BRNN) con nodos LSTM, tiene como fin conectar dos capas ocultas que van en sentido contrario. Intuitivamente, una capa va en sentido a la secuencia presentada donde se desarrolla el proceso de la información como se describió arriba; y existe otra capa que realiza el procesamiento de la secuencia en sentido contrario usando igualmente una red LSTM, al final los dos outputs son concatenados. En otras palabras, la red se entrena con estados del pasado y del futuro, simultáneamente. Por tal motivo este tipo de redes son útiles cuando el contexto debe estar presente (Cui et al, 2018). Ver siguiente figura donde se representa la información descrita.

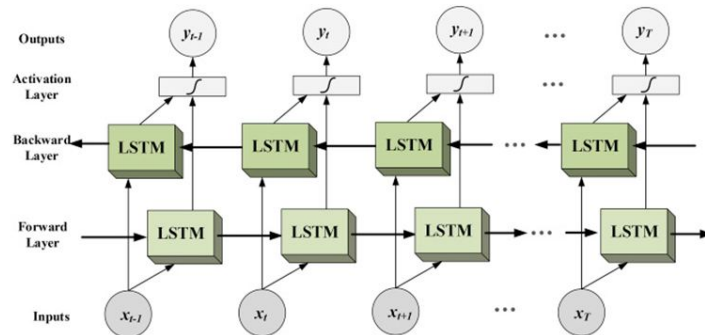


Figura 3. Visualización de una red LSTM Bidireccional (BRNN)

Retomada de: <https://www.i2tutorials.com/technology/deep-dive-into-bidirectional-lstm/>

Recientemente, Althelaya et al en 2018 compararon estas dos estrategias (RNN y BRNN) para problemas de series de tiempo financieras con el fin de predecir los precios de cierre de stock markets. Este problema fue abordado de esta manera debido a la no-linealidad de los datos y por lo tanto los métodos AR, ARIMA y ARMA

no son capaces de extraer estos sucesos. Según reportaron los autores, BRNN tuvo una ventaja sobre LSTM, con un RMSE promedio del testing de 0.015240 contra 0.037634 de una RNN (Athelaya et al, 2018).

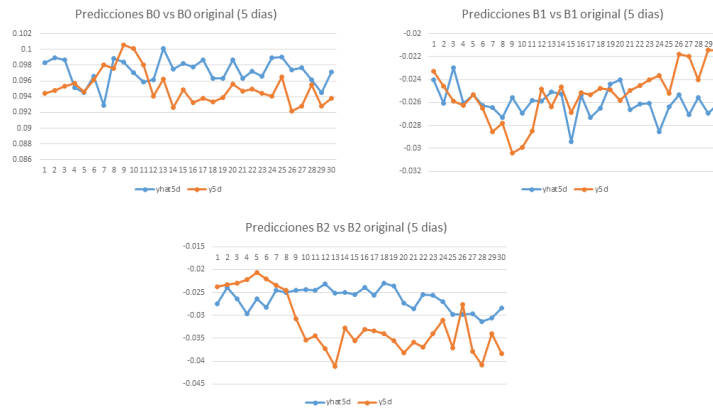
En este sentido, la estrategia para predecir fue la siguiente: se tomaron n datos de un periodo de tiempo de 30 días, y se predecía el siguiente día para cada β , con el valor obtenido para \hat{y}_{n+1} se realizó una nueva predicción para el día $t + 1$ y así hasta 5 y 10 días. Luego en la siguiente iteración se toman X_{n+1} datos y se predice hasta \hat{y}_{n+6} y \hat{y}_{n+11} . De esta manera, vamos obteniendo las predicciones 5 y diez días (Figura 5).

$$\begin{matrix} & it_1 & & \dots & \dots & & & & it_k \\ \begin{bmatrix} x_1 & & & & & & & & x_1 & x_1 \\ \vdots & & & & & & & & \vdots & \vdots \\ x_n & & & & & & & & \vdots & \vdots \\ \vdots & & & & & & & & \vdots & \vdots \\ & & & x_{n+1} & & & & & \vdots & \vdots \\ \hat{y}_{n+5} & & & \vdots & & x_{n+2} & & & \vdots & \vdots \\ \vdots & & & \hat{y}_{n+6} & & \vdots & & & x_{n+3} & \vdots \\ \hat{y}_{n+10} & & & \vdots & & \hat{y}_{n+7} & & & \vdots & \vdots \\ & & & \hat{y}_{n+11} & & \vdots & & & \hat{y}_{n+8} & \vdots \\ & & & & & \hat{y}_{n+12} & & & \vdots & \vdots \\ & & & & & & & & \hat{y}_{n+13} & \vdots \end{bmatrix} \end{matrix}$$

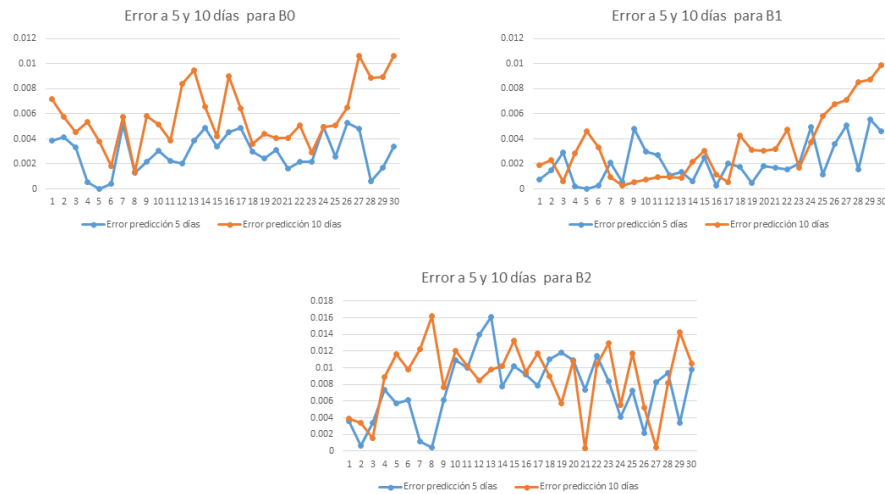
Figura 4. Representación de la matriz creada por el algoritmo propuesto.

Después de las simulaciones (Archivos adjuntos, ver resultados.xlsx y predicciones.xlsx), se encontró que BRNN tuvo un mejor rendimiento para realizar las predicciones, sin embargo, RNN tuvo un mejor rendimiento para predecir a cinco días, esto confirma que una red simple logra identificar mejor patrones a corto plazo. Adicional, los mejores parámetros correspondían a tomar 30 datos originales y hacer grupos de train y test de 5 datos. Se graficó, cada β de los valores estimados contra los originales, y como se puede ver a continuación, las predicciones siguen las tendencias de los β originales (Figura 5).

Por otro lado, el error deja ver que las predicciones en su totalidad se sobrestimaron. A pesar de esto, para todos los casos los errores se mantuvieron en el rango de $[1.44e^{-5} - 1.62e^{-2}]$, por lo que las aproximaciones tienen un error bastante cercano

Figura 5. $\hat{\beta}_t$ para 5 días vs β originales

a cero y las predicciones para la BRNN son extremadamente buenos comparados con RNN (Figura 6).

Figura 6. Errores absolutos para cada β

Chapter 5

ETF, valoración con nuestras predicciones

Para finalizar el análisis, se calculó el comportamiento futuro de la estructura a término, empleando la función de Nelson-Siegel del 17 de junio de 2016 (ver archivo adjunto Nielson-siegel.xlsx). En las figuras 1 y 2 podemos ver como la curva que arroja nuestro modelo se asemeja en gran medida a los datos históricos reales, mostrados para el 17 de junio de 2016. Por un lado, la figura 1 muestra la curva que fue construida con base a una proyección a cinco días. Mientras tanto, la figura 2 muestra la curva construida a partir de proyecciones a diez días. Nótese que, en ambas figuras, el spread entre la curva real y la curva generada va aumentando. Esto señala que el modelo es mejor prediciendo el extremo de corto plazo de la curva que el extremo de largo plazo. Observando la forma de las dos curvas, también resulta evidente que tienen una pendiente positiva, congruente con el ambiente colombiano actual en donde no se esperaría que la curva de rendimientos estuviera invertida o presentara tendencias a disminuir a medida que la madurez de los bonos aumentara.

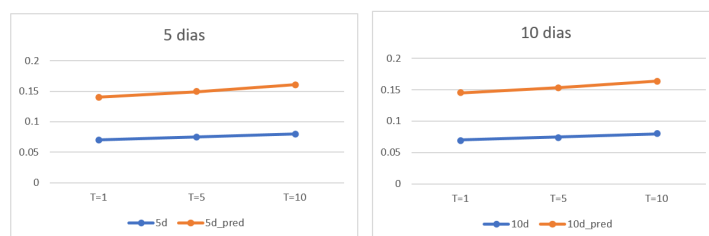


Figura 7. Comparación para 17jun2016 de los valores obtenidos por la formula parametrica Nelson-Siegel estimados y originales

Para determinar la estrategia más rentable, se llevó a cabo un análisis probando diferentes ETFs. Específicamente, se probaron tres ETFs diferentes. Un primer ETF

con 33% en bonos de 1 año, 33% en bonos de 5 años y 33% en bonos de 10 años. Un segundo ETF con 50% en bonos de 1 año, 30% en bonos de 5 años y 20% en bonos de 10 años. Un tercer ETF con 20% en bonos de 1 año, 30% en bonos de 5 años y 50% en bonos de 10 años. Los resultados, aunque no concluyentes, indican que una estrategia donde se utilice la curva de 5 días y se le de un mayor peso a los bonos de 1 año es la mejor estrategia para generar una mayor rentabilidad. Gráficamente, esto puede verse en el hecho de que el spread de la figura 1 cuanto $T=1$ año es el menor de todos los spreads observados.

Chapter 6

Conclusiones

El presente trabajo aporta a las investigaciones económicas del mercado de bonos del tesoro colombiano en tanto presenta una alternativa para predecir el precio de un bono de cualquier duración en el corto plazo (5-10 días).

Respecto a las redes, las RNN fueron redes eficientes para determinar patrones en estas series de tiempo, precisamente porque este tipo de datos requieren de la introducción de estacionalidad. Se encontró que entre mas simple a estructura de las red, mejor era el desempeño a la hora de hacer pronósticos, esto concuerda con lo encontrado por Szenczi et al en 2016, ya que programamos otras redes cuyo desempeño no era óptimo ni comparable con las RNN ni BRNN (datos no mostrados). A futuro se puede indagar porque estos tipos de datos funcionan mejor con este tipo de arquitecturas.

En ese sentido, aunque las redes neuronales parecen brindar mejores resultados en comparación con los modelos utilizados comúnmente por la literatura económica y financiera, aún existen múltiples mejoras en los algoritmos que permitirán reforzar la toma de decisiones dentro de las instituciones que utilizan predicciones sobre la curva de rendimientos a la hora de invertir. No es de extrañar que aunque este tipo de esfuerzos por predecir la estructura a termino no se hayan formalizado dentro de la literatura académica, estos no sean utilizados diariamente por los distintos agentes del mercado, por lo cual será importante continuar evaluando y modificando estas metodologías para brindar no sólo más robustez a los resultados sino que permitirá formalizar el área concerniente a la utilización de algoritmos basados en el aprendizaje de matemáticas dentro del mundo financiero.

Chapter 7

Bibliografía

Szenci, R. (2016). *Forecasting the Hungarian term structure of interest rates with econometrics and neural network methods*, Cornivus University of Budapest

Adhikari, R., Agrawal, R. K. (2013). *A combination of artificial neural network and random walk models for financial time series forecasting.*, Neural Computing and Applications, 24(6), 1441–1449

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

Vela, D. (2013). *Forecasting Latin-American yield curves: An artificial neural network approach.* Borradores de Economía, Banco de la República

Moreno, S. (2018). *Evidencia reciente sobre la relación de la curva de rendimientos de Svensson y factores macroeconómicos.* Memoria de grado, Universidad de los Andes

Althelaya, K. A., El-Alfy, E.-S. M., Mohammed, S. (2018). Evaluation of bidirectional LSTM for short-and long-term stock market prediction. 2018 9th International Conference on Information and Communication Systems (ICICS).

Cui, Z., Ke, R Wang, Y. (2018). Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction. 2018 International Workshop on Urban Computing (UrbComp)

Chapter 8

Anexos

Autocorrelación factores latentes

