



UNIVERSIDAD DEL ROSARIO

Sentimiento del mercado y minutas del banco central

Luis Felipe Reyes López

Tutor: Jesús Gilberto Otero Cardona

Trabajo de Grado
Facultad de Economía

Sentimiento del mercado y minutas del banco central

Resumen

Los anuncios del Banco de la República constituyen un elemento determinante en la toma de decisiones de diferentes actores en la economía colombiana, especialmente bancos, inversionistas institucionales y fondos de inversión internacionales. El comportamiento de los retornos de las acciones que cotizan en la bolsa de valores de Colombia (BVC) ofrece una aproximación al nivel de crecimiento, inversión, estabilidad de los precios y al estado general de la economía.

Esta investigación analiza el vínculo entre los anuncios del banco central y el comportamiento de los retornos del principal indicador bursátil y las acciones más dinámicas de la BVC. Para esto, se utilizó el sentimiento del mercado, un concepto perteneciente al campo del Procesamiento del Lenguaje Natural (PLN). La fuente para estimar el sentimiento del mercado fueron las minutas del Banco de la República y se consideraron los retornos mensuales del COLCAP, y los de las acciones de Ecopetrol y Bancolombia para el periodo comprendido entre 2012 y 2022.

Utilizando un modelo ARDL, los resultados muestran una relación entre los retornos del COLCAP y el sentimiento de mercado.

Palabras claves: Procesamiento del lenguaje natural (PNL), sentimiento del mercado, Banco de la República, Bolsa de valores de Colombia.

Abstract

The Banco de la República's announcements are a determining element in the decision-making process of different agents in the Colombian economy, especially banks, institutional investors and international investment funds. The behavior of stock returns listed on the Colombian stock exchange (BVC) offer an approximation to the level of growth, investment, price stability and the general state of the economy.

This research analyzes the link between the central bank's announcements and the behavior of the returns of the main stock market indicator and the most dynamic shares listed on the BVC. For this purpose, market sentiment, a concept belonging to the field of Natural Language Processing (NLP), was used. The sources for estimating market sentiment were the minutes of Banco de la República and the monthly returns of COLCAP, and those of Ecopetrol and Bancolombia shares for the period between 2012 and 2022 were considered.

Using an ARDL model, the results show a relationship between COLCAP returns and the market sentiment.

Keywords: Natural Language Processing (NLP), market sentiment, Banco de la República, Colombian Stock Exchange.

Tabla de contenido

1. Introducción	5
2. Estado del arte	7
3. Metodología	12
3.1 Bases de datos y fuentes de información	12
3.2 Sentimiento de mercado	13
3.3 Modelo econométrico	17
3.4 Modelos autorregresivos	17
3.5 Modelo ARDL	18
3.7 Diagrama técnico	19
4. Evaluación empírica	20
5. Conclusiones y recomendaciones	22
Referencias	24
Anexos	26
Resultados econométricos	26
Programa de R	26
Lista de minutas	33

1. Introducción

El análisis sobre el comportamiento de los agentes económicos, es un área fundamental de la economía, gobiernos y bancos centrales, quienes mediante medidas económicas asumen una reacción esperada por parte del mercado. En este contexto, es necesario entender y profundizar en el estudio del comportamiento de los agentes económicos y en particular, en cómo éstos toman sus decisiones. El concepto de “sentimiento de mercado” permite utilizar información que no es considerada en el análisis económico tradicional. Su enfoque gira en torno al “sentimiento” que puede generar una noticia en los agentes y no únicamente en la información que transmite.

Las noticias económicas son evaluadas desde un análisis fundamental (el cual se basa principalmente en el análisis de datos y cómo estos pueden anticipar futuros eventos) que resulta incompleto al momento de predecir el comportamiento de la economía. Por esta razón, y gracias al avance tecnológico, se han creado herramientas que ayudan a determinar cómo, las noticias y diferentes datos, afectan las decisiones que toman los agentes económicos.

Según estudios realizados por Tetlock (2007, 2008) las noticias pueden contener información precisa para predecir tanto ganancias como pérdidas. Sin embargo, dicho efecto o previsibilidad no suele durar más de dos días. En su análisis, Tetlock llega a la conclusión de que las noticias o artículos negativos conllevan a que los agentes se motiven a seguir leyendo e investigando en noticias diarias. Su análisis se basó en el impacto que tienen las palabras negativas, de los periódicos del Wall

Street Journal (WSJ), compañías que publican información financiera como el Dow Jones New Service (DJNS) y artículos del S&P 500 (índice bursátil de EEUU), en los precios de las acciones en Estados Unidos.

De forma similar, en Colombia la noticia del retiro de la reforma tributaria que pretendía recaudar 23,4 billones de pesos, generó gran incertidumbre y ocasionó caídas en la inversión extranjera, alerta en fondos de inversión y acentuó la depreciación del peso.

La intención de este estudio, es ofrecer un índice del sentimiento del mercado utilizando las minutas emitidas por el Banco de la República y un análisis sobre la relación entre este índice y varios indicadores bursátiles. Taborda (2015) abordó esta problemática utilizando principalmente el número de palabras positivas y negativas en los anuncios del banco central. Esta investigación busca complementar ese análisis, considerando cada una de las frases en las minutas emitidas por el Banco de la República y asignándoles un valor positivo o negativo.

Un estudio similar, fue realizado en el Reino Unido y en los Estados Unidos por Bai (2014), quien analizó el impacto del sentimiento generado por ambos países, sobre distintos mercados del Reino Unido. Se encontró, que el sentimiento de mercado tiende a tener un mayor efecto sobre el rendimiento de las acciones, en periodos de estabilidad económica. Este estudio también mostró que el sentimiento generado por los Estados Unidos se puede transmitir e influir en el Reino Unido. Sin embargo, también se aclara que, dependiendo de las variables consideradas, las estimaciones pueden ser poco relevantes o no significativas en lo absoluto.

Este documento consta de cinco secciones incluyendo la introducción. En la segunda sección, se presenta de manera general el concepto de la programación del lenguaje natural, su historia e importancia actual. La siguiente sección incluye la metodología que se empleó para el estudio y la base de datos utilizada. La sección cuarta, corresponde a la evaluación empírica y los análisis de los resultados obtenidos. Finalmente, en la quinta sección, se presentan las conclusiones y se proponen algunas recomendaciones.

2.Estado del arte

Al estudiar el sentimiento del mercado, es fundamental que exista una estructura fuerte y transparente de la institucionalidad, con el fin de que los diferentes agentes crean en las minutas, artículos o ruedas de prensa que proporcione el banco central.

De acuerdo con Taborda (2015) las actas son un elemento central en el proceso de transparencia, claridad, redacción y pueden tener efectos sobre la política para controlar la inflación. Sin embargo, en la vida real, los agentes no son completamente racionales, por lo que estarán influenciados por la forma y fondo de dichas minutas o artículos antes de tomar cualquier decisión.

Cabe constatar que todos los días, cientos de noticias son producidas. Para los agentes, esta cantidad enorme de información es difícil de procesar, ya sea por la cantidad, extensión o dificultad gramatical de los textos.

A pesar de la crisis actual en Colombia, las instituciones luchan por controlar tanto el déficit fiscal como la inflación que, para el 2022 alcanzó el 8,53%, a todo esto, se suma igualmente la fuerte devaluación del peso colombiano. Ante esta situación según el Ministerio de Hacienda para 2022 el déficit fiscal continuaría su reducción y se ubicaría en 6,2% del PIB, “...como consecuencia de un incremento en los ingresos corrientes, derivados del proceso de reactivación económica, que se compensa con un mayor gasto en intereses...”. (Boletín N°4, Ministerio de Hacienda,2022). Éste es un claro ejemplo de la transparencia institucional, donde la fortaleza de las instituciones hace posible que una política monetaria o informes económicos surtan efecto sobre la economía.

Para Apel y Grimaldi (2012), si bien la transparencia de cualquier política monetaria debería considerarse como crucial, aún existe una intensa discusión sobre si publicar las actas y los resultados de las votaciones de las reuniones de política monetaria es beneficioso.

Un ejemplo extremo es Venezuela, en donde la destrucción institucional, entre otros, ocasionó una catástrofe económica y social. En este contexto, es claro como la legitimidad de un gobierno e independencia del Banco Central son realmente necesarios, de lo contrario, cualquier artículo, texto, minuta o anuncio de una política monetaria no tendría mayor efecto en los mercados, pues la credibilidad de los distintos agentes en el Estado e instituciones se habría perdido.

En concordancia con lo propuesto por Andersen (2007), deberíamos cuestionarnos cómo se incorporan las noticias o, en este caso, las minutas en los precios del

mercado. Si los mercados son eficientes, se espera que el efecto sea momentáneo y esperado. De ser así, entonces ¿Por qué se presentaron las crisis de las burbujas en los mercados en 1990 o en el 2008? ¿Tiene algún efecto mayor sobre el mercado, una noticia buena o una noticia mala? ¿Podríamos esperar que ante noticias alentadoras los rendimientos de los activos deberían tender a subir?, ¿Cuál sería dicho impacto en el corto o largo plazo?

Para Kurihara (2014), los efectos de la política monetaria sobre el precio de las acciones y la tasa de cambio existen solo en un corto periodo de tiempo. En el largo plazo, el mercado tiende a estabilizarse y dichos efectos serían simplemente parte de un ciclo sin fin de subidas y bajadas.

Teniendo en cuenta lo anterior, para poder estudiar cómo el sentimiento de mercado afecta el precio de las acciones es necesario ampliar el análisis e incluir el Procesamiento del Lenguaje Natural (PLN) o por sus siglas en inglés NLP.

El análisis del PLN, empezó alrededor de 1950, con el surgimiento de las computadoras que se utilizaron para realizar las primeras traducciones del lenguaje ruso al inglés. Aunque solo fueron algunas frases, aquí empezó el interés por intentar entender el lenguaje natural de las personas a través de la programación. Si bien el análisis del lenguaje empezó hace varias décadas, éste no se desarrolló sino hasta mediados de los 70, y para Leiva y Muñoz (1996), tuvo su gran auge en los 80.

Sin embargo, en la actualidad el PLN es un campo relativamente desconocido aun cuando se ha convertido en algo intrínseco en la en la vida cotidiana. Google, Yahoo o Ecosia usan el PLN para codificar, a través de programación, lo que deseamos buscar y nos arroja una cantidad de resultados mediante el análisis del lenguaje. Igualmente, la revolución del “internet de las cosas” o IoT por sus siglas en inglés ha impulsado la comprensión del lenguaje por voz de nuestros teléfonos y aparatos electrónicos. De la misma manera y debido a la pandemia mundial, los chatbots o asistentes virtuales han tenido un crecimiento exponencial desde abril de 2020.

En el Diagrama 1° podemos evidenciar como todo tipo de información, textual, de voz u otros, puede ser analizada a través del PLN y construir ciertas predicciones o alertarnos de cambios en el futuro cercano.

Diagrama 1°



Fuente: Data Management Inteligencia Artificial. Datia.<http://dat-ia.com/>

Según Verdejo (1994) existen varias diferencias entre el lenguaje natural y artificial. Primero, encontramos la riqueza de tener un vocabulario extenso en el lenguaje natural, opuesto al lenguaje artificial, que es más rígido y necesita establecer parámetros que no permiten excepciones. Así mismo, en el lenguaje natural las

palabras por sí solas tienden a tener varios significados dependiendo del contexto. Éstas, son algunas de las dificultades que enfrenta el PLN, al recolectar datos, analizarlos y generar un resultado o interpretación.

Si bien es posible omitir algunas palabras en una oración gracias a la elipsis, la oración no deja de perder sentido y se reduce su ambigüedad si se consideran más palabras simultáneamente. Esta es la razón principal por la cual es completo realizar un análisis de PLN mediante frases y no palabras.

La idea principal del PLN es combinar el campo de la inteligencia artificial y la informática. Actualmente, existen diferentes lenguajes de programación y herramientas (R, Python, Java) que permiten analizar textos en muy poco tiempo.

Aunque ningún software es capaz de entender todos los matices conceptuales que puede tener la interpretación de una oración, es cierto que la capacidad computacional permite procesar cantidades enormes de información y reduce el riesgo de la subjetividad en la interpretación.

De acuerdo con Moreno (2021), podemos encontrar distintos niveles del PLN:

1. Morfológico: Se examina cada vocablo con el fin de obtener toda la información gramatical del texto.
2. Analizador Sintáctico: Se busca desambiguar aquellas palabras del morfológico tomando oraciones.
3. Semántico: Dar un sentido e interpretación a las oraciones.
4. Pragmático: Interpretación final.

De los 4 componentes principales del procesamiento del lenguaje natural, este estudio se enfoca en el análisis semántico, donde se da la interpretación de las oraciones.

Con este fin, el PLN junto con herramientas computacionales, aparte de reducir en gran medida los costes tanto humanos como monetarios, también se garantiza que, al analizar un texto o en nuestro caso las minutas del Banco Central, se obtendrá un valor al escrito en el menor tiempo posible ayudando a inversionistas, empresas o cualquier persona interesada en un análisis económico, a tomar una mejor decisión o acción.

3. Metodología

3.1 Bases de datos y fuentes de información

Para estudiar el efecto de las minutas¹ emitidas por el Banco de la República, se escogió un periodo de 10 años, partiendo desde el 2012 hasta el 2022. Se seleccionaron las minutas con una periodicidad mensual, se obtuvo una base de datos de 105 textos.

Según Taborda (2015) *“la estructura de las actas es bastante estática en el sentido de que típicamente se revisan desarrollos económicos recientes, tendencias de la inflación, problemas de mercado laboral, comercio exterior, comportamiento de los*

¹Página web donde se encuentran las minutas del Banco de la República: <https://www.banrep.gov.co/es/minutas>

mercados monetarios entre otros”, es por esta razón que la base de datos viene únicamente de las minutas emitidas por el Banco de la República, ya que los textos, al salir de una misma fuente hacen más preciso el estudio.

Como variables dependientes, se eligieron los retornos del índice Colcap y el de las acciones de Ecopetrol y Bancolombia del mercado bursátil colombiano, cada retorno en una estimación independiente.

El índice Colcap refleja las variaciones de las acciones más líquidas y está compuesto por las 20 principales compañías que cotizan en la bolsa de valores, entre las que se encuentran Bancolombia, Cemento Argos, Grupo Nutresa, Davivienda, Grupo Aval entre otras.² Todos los retornos se extrajeron con periodicidad mensual.

3.2 Sentimiento de mercado

La construcción del índice del sentimiento del mercado se hizo a través del programa R y el paquete Syuzhet³. Este paquete fue desarrollado por Matthew Jockers en el 2015, cofundador del Laboratorio Literario de la Universidad de Stanford en Silicon Valley y profesor de inglés en la Universidad de Nebraska en Lincon. Este paquete permite seleccionar una lista de frases que mantienen el sentido de la oración y evita ambigüedades en la interpretación, cuenta con 4 diccionarios: Bing, Afinn, Stanford y NRC y fue producido por el grupo del PLN de Stanford.

²Los retornos fueron obtenidos de la base disponible en la página web: <http://es.investing.com/equities>.

³En la página web :<https://programminghistorian.org/es/lecciones/analisis-de-sentimientos-r#objetivos> podemos encontrar un ejemplo de cómo se usa el paquete para un mayor entendimiento del mismo, además de varios ejemplos y sus resultados, aunque sólo de pequeños párrafos.

La ventaja de trabajar con el diccionario NRC es que incluye el idioma español, además según el artículo de Jennifer Isasi (2021), cuenta con 14.182 palabras con las categorías de sentimiento positivo y negativo y emociones como el enfado, sorpresa, tristeza entre otras.

El paquete Syuzhet genera una lista de frases con un respectivo valor del sentimiento (valores entre 1 y -1), en donde, 0 es un valor neutro, es decir que no le otorga un sentimiento determinado al texto; >0 , la frase tiene un sentimiento positivo; <0 la frase tiene un sentimiento negativo. En este estudio, se tomó el promedio de la puntuación de todas las frases de una minuta para determinar el sentimiento total de la publicación.

La Tabla 1° muestra las frases de la minuta del 28 de enero del 2022, este es el primer paso para determinar el sentimiento del mercado de esta minuta.

Tabla 1°

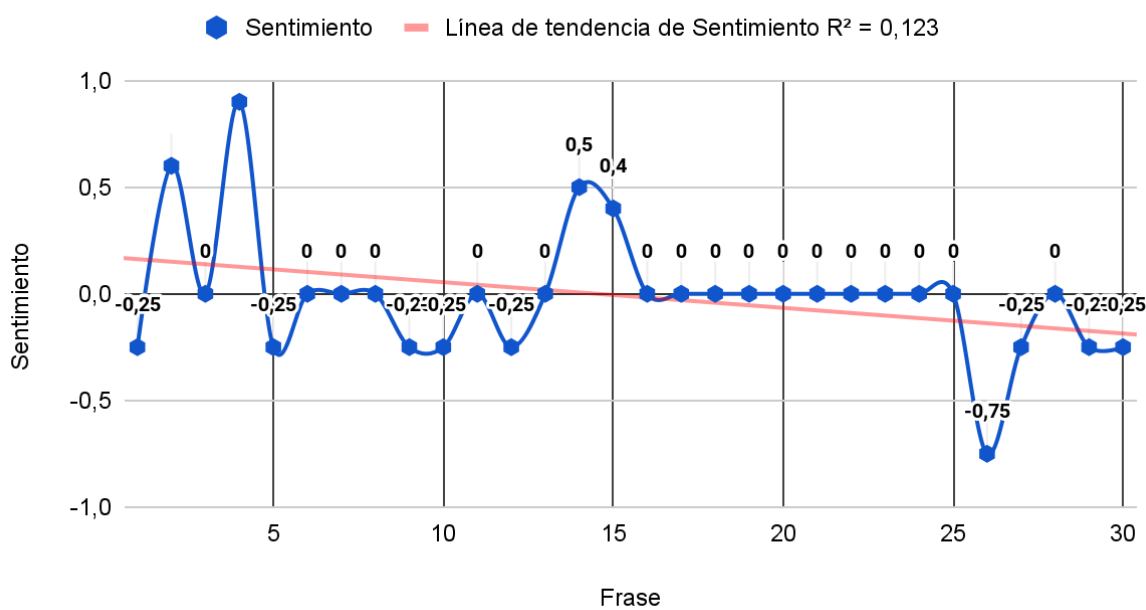
Ejemplo frases de la minuta

[1] "En su sesión del 28 de enero, la Junta Directiva del Banco de la República incrementó la tasa de interés de política en 100 puntos básicos (p.b.) en una decisión por mayoría de 5 a 2, llevándola a 4,0%."
[2] "Dos miembros votaron a favor de un aumento de 75 p.b."
[3] "La decisión tuvo en cuenta los aspectos que se resumen a continuación."
[4] "La inflación cerró el año con una tendencia alcista y superior a la anticipada, tanto en el caso del índice general, como en el caso de la inflación básica."
[5] "Las expectativas de inflación son altas y terminaron el año con tendencia alcista."
[6] "El equipo técnico revisó al alza sus pronósticos de inflación los cuales prevén que la inflación total y la inflación básica se ubicarían por encima de la meta al final de 2022 (4,3% y 4,5% respectivamente) y de 2023 (3,4% y 3,6%)."

La Tabla 2° muestra la trayectoria del valor asignado a cada una de estas frases.

Tabla 2°

Ejemplo Trayectoria Minuta



En el eje horizontal, se encuentran todas las frases de la minuta, en el eje vertical, el sentimiento de cada frase. Para esta minuta, el programa encuentra 30 frases, las primeras 5 tienden a sembrar un sentimiento positivo, luego, de la 5 a la 13 uno negativo, de la 14 a la 15 vuelve a generar un sentimiento positivo y a partir de la 16 neutral hasta la 24; por último, de la 25 a la 30, se transmite un sentimiento negativamente al lector.

Al analizar algunas frases se encuentra lo siguiente: la primera frase, con un valor de -0.25, indica un sentimiento negativo, la misma contiene: "... incrementó la tasa de interés...". Para el segundo valor, 0.60, en la frase se encuentra: "...a favor de un aumento..." dándole un sentimiento positivo a la oración y vemos como en la

trayectoria de la narrativa la línea sube esporádicamente. Estos dos ejemplos ilustran el funcionamiento del programa elegido.

En la Tabla 3° se presenta la sumatoria y el resumen de los valores de todas las frases a lo largo de la minuta, el promedio es de -0.011666, lo cual sugiere que el texto generará en el lector un sentimiento negativo en general.

Tabla 3°

<i>Resumen Comportamiento de la Minuta</i>					
Min	1st Qu	Median	Mean	3rd Qu	Max
-0.75000	-0.25000	0.00000	-0.01167	0.00000	0.90000

Al considerar los resultados de Taborda y como lo muestra la Tabla 4°, la estructura general de las minutas se ha mantenido estable a través del tiempo. Por ejemplo, en las primeras doce frases de la minuta del 28 de enero del 2022, se encuentran las palabras como: año, crecimiento, inflación, tasa de interés y mes. La estabilidad en la forma en la que son escritas las minutas le brinda robustez al presente estudio.

Tabla 4°

Frecuencia de palabras de las minutas del Banco de la República

Colombia	
Term	Freq.
año	2281
crecimiento	1615
inflación	1537
tasa	1453
mes	1295
precios	929
trimestre	809
mundo	739
alimentos	673
tasa_de_interes	623
junta	620
comportamiento	581
consumo	569
menor	531
crédito	505
indicadores	496
economía	494
banco	493
demanda	489
parte	482

Fuente: R. Taborda / *Ensayos sobre Política Económica* 33 (2015) 76–92

3.3 Modelo econométrico

Para la estimación empírica se utilizó un modelo ARDL autorregresivo y de rezagos distribuidos. La ventaja de este modelo es que permite incluir los rezagos de la variable dependiente y una variable exógena al proceso, características que coinciden con los datos a analizar dado que los retornos de las acciones usualmente siguen un proceso autorregresivo y el sentimiento de mercado es una interpretación exógena por parte del Banco de la República.

3.4 Modelos autorregresivos

Un ejemplo de modelos autorregresivos (AR) según Gujarati y Porter (2011) , es:

$$y_t = \alpha + \gamma y_{t-1} + u_t \quad (1)$$

donde la regresión sólo incluye valores pasados de las variables dependiente, α es una constante, γ es el parámetro autorregresivo y u_t es ruido blanco.

En el caso de un AR (1) el valor del proceso en el periodo t , dependen del valor en $t-1$ y así sucesivamente al aumentar el orden p del proceso.

3.5 Modelo ARDL

Un modelo ARDL, es un proceso autorregresivo de rezagos distribuidos, donde y_t se explica en parte por sus valores rezagados ($y_{t-1}, y_{t-2}, y_{t-3} \dots y_{t-n}$) con $n > 0$ y por los rezagos de la variable explicativa X_t ($X_{t-1}, X_{t-2}, X_{t-3} \dots X_{t-m}$), con $m > 0$.

Siguiendo a Chen (2010). el modelo se representa de la siguiente manera:

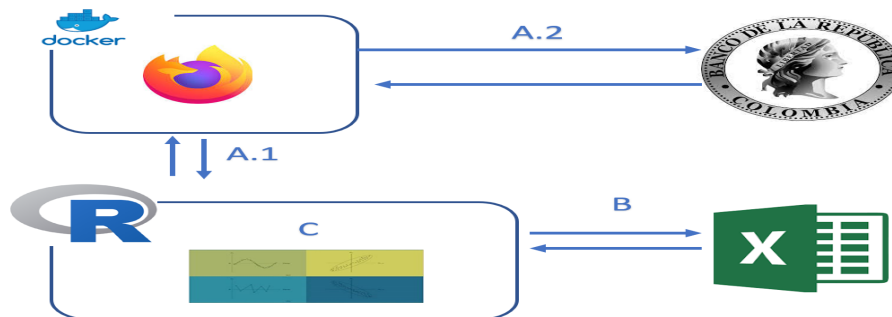
$$y_t = \alpha + \gamma_1 y_{t-1} + \dots + \gamma_n y_{t-n} + b_0 X_t + b_1 X_{t-1} + \dots + b_m X_{t-m} + u_t \quad (2)$$

donde b_0 es el multiplicador de impacto del modelo, u_t es un ruido blanco y n y m son los rezagos de y_t y X_t .

3.7 Diagrama técnico

El Diagrama 2° presenta la relación técnica entre los diferentes componentes utilizados en la recolección de los datos, consolidación y estimación econométrica.

Diagrama 2°



A1: R se conecta a Docker⁴ que contiene una imagen del navegador Firefox⁵ y le solicita visitar la página del Banco de la República en donde está publicada una minuta.

A2: Firefox visita la página y envía el resultado a R, en donde es analizada para determinar el valor correspondiente al sentimiento de mercado.

R tiene paquetes como tidyverse que directamente pueden hacer llamadas a páginas web, sin embargo, muchos sitios web tiene restricciones cuando esas llamadas no son realizadas por un navegador estándar, este es el caso para la

⁴ “Docker es un proyecto de código abierto que automatiza el despliegue de aplicaciones dentro de contenedores de software, proporcionando una capa adicional de abstracción y automatización de virtualización de aplicaciones en múltiples sistemas operativos.” [https://es.wikipedia.org/wiki/Docker_\(software\)](https://es.wikipedia.org/wiki/Docker_(software))

⁵ “Mozilla Firefox (o simplemente Firefox) es un navegador web libre y de código abierto desarrollado para distintas plataformas, está coordinado por la Corporación Mozilla y la Fundación Mozilla.” [https://es.wikipedia.org/wiki/Docker_\(software\)](https://es.wikipedia.org/wiki/Docker_(software))

página del Banco de la República y es la razón principal por la cual es necesario utilizar Docker.

B: Los datos de los retornos son cargados en R desde Excel; al ejecutar el programa que hace las llamadas al Banco de la República, existe una restricción técnica que no permite realizar todas las llamadas en una sola ejecución, es necesario hacerla por bloques, por esta razón, los resultados del sentimiento del mercado están también incluidos manualmente en el archivo de Excel.

C: Estimación econométrica con varios paquetes estadísticos de R, la lista completa se encuentra en el anexo, sección “Lista de minutas”.

4. Evaluación empírica

Al estimar el modelo ARDL se obtuvieron los siguientes resultados: Significancia del intercepto, de los retornos y del sentimiento. Se presenta también el mejor modelo según AIC⁶, la significancia de los rezagos correspondientes, el R ajustado, el error estándar y el P-valor en la tabla 6°.

El criterio de Akaike, nos proporciona la información necesaria para estimar el modelo con los rezagos necesarios de cada variable. La tabla 5° nos indica que para la variable Rcolcap (Rendimientos de Colcap), el mejor modelo que puede

⁶AIC es el criterio de información de Akaike donde según Racine, J. S. (2019) “el criterio de log-verosimilitud es penalizado y proporciona un equilibrio entre un buen ajuste en la muestra (log-verosimilitud) y la complejidad del modelo. Un buen modelo tiene un valor AIC pequeño”.

explicar los rendimientos, es el estimado con el primer rezago de Rcolcap y no incluir ningún rezago de la variable sentimiento⁷, es decir un modelo ARDL(1,0).

Tabla 5°

Mejor Modelo AIC

<i>Mejores Modelos</i>	Rcolcap	Sentimiento	AIC
1	1	0	-368.3396
2	1	1	-367.3305
3	2	0	-363.0266
4	2	1	-361.9738
5	2	2	-360.3738
6	3	0	-358.3686
7	3	1	-357.4899
8	3	2	-355.8675
9	3	3	-355.1336
10	4	4	-350.2066
11	4	3	-349.2971
12	5	5	-347.3795

Tabla 6°

RESULTADOS DE LA REGRESIÓN

	<i>Variables Dependientes</i>		
	Colcap ARDL(1,0)	Bancolombia ARDL(1,0)	Ecopetrol ARDL(1,1)
(Intercept)	0.01 (0.01)	0.02* (0.01)	-0.00 (0.01)
L(RColcap, 1)	0.05 (0.09)		
Sentimiento	0.18* (0.08)	0.29* (0.12)	0.30 (0.16)
L(RBancolombia, 1)		0.06 (0.09)	
L(REcopetrol, 1)			0.15 (0.09)
L(Sentimiento, 1)			-0.33* (0.16)
Error Estándar	0.05	0.07	0.09
R ²	0.04	0.05	0.06
Adj. R ²	0.02	0.04	0.04
F.Estadístico	2.41	3.18	2.50
P.Valor	0.09	0.04	0.06
Num. obs.	118	118	118

****p* < 0.001; ***p* < 0.01; **p* < 0.05

8

⁷ En los anexos se encuentran las tablas correspondientes a las 2 variables faltantes (bancolombia y ecopetrol), en donde podemos evidenciar que para la variable Ecopetrol, sí es significativo incluir el primer rezago del sentimiento.

⁸ RColcap, RBancolombia, REcopetrol , corresponden a los rendimientos de las variables del Colcap, Bancolombia y Ecopetrol respectivamente.

Para el índice Colcap⁹, el sentimiento de mercado influye en la variabilidad de sus rendimientos con un R ajustado de 2%, es decir que tan solo un 2% de los rendimientos del Colcap son explicados por el sentimiento de mercado calculado a partir de las minutas del Banco de la República.

Este resultado bajo es confirmado por el P. valor del 9%. Aun cuando el coeficiente de significancia no es muy alto, esto indica que existen más variables que afectan en mayor medida el Colcap y que no fueron consideradas en este modelo.

Bancolombia¹⁰

Para Bancolombia el R ajustado fue un poco mayor, de 4% y tanto el intercepto como el sentimiento fueron significativas. En cuanto al sentimiento, cuanto mayor sea éste, existirá mayor probabilidad de que afecte los rendimientos de Bancolombia, mostrándonos que las variaciones en los rendimientos esperados son explicados por el sentimiento de mercado.

Ecopetrol¹¹

Los resultados muestran que el primer rezago del sentimiento de mercado (-0.33) afectan los rendimientos de la acción ecopetrol, con un R ajustado fue de 4% y un P.valor del 6%.

⁹ Detalle de los resultados incluidos en el anexo.

¹⁰ Detalle de los resultados incluidos en el anexo.

¹¹ Detalle de los resultados incluidos en el anexo.

5. Conclusiones y recomendaciones

El análisis e implementación del PLN es fundamental hoy en día para la interpretación de noticias y textos, más allá de una ayuda, se ha convertido en una base indispensable para nuestra comunicación, en un mundo donde la globalización y digitalización demandan un procesamiento rápido y eficiente de la información. Esta investigación también muestra cómo el sentimiento de mercado siempre estará ligado a la institucionalidad de un país, pues cada individuo, siendo racional, tiene perspectivas distintas de cómo interpretar una noticia, si creer o no en ella, dependiendo de la legitimidad y confianza en las instituciones.

Finalmente el peso del valor positivo o negativo, es más significativo a partir de las frases intermedias, pues en las frases iniciales suele sugerirse una introducción para poner en contexto al lector, pero es en la parte central del cuerpo del texto donde se desarrollan las ideas y argumentos.

Como recomendación, la calidad de los datos es esencial para saber si una noticia, mala o buena, influirá no solo en el corto plazo, como por ejemplo el conflicto actual que se vive entre Rusia y Ucrania, en donde a pesar de que lleguen a una posible tregua, el efecto de la noticia del conflicto sobrepasa el efecto de cualquier noticia positiva para Rusia, pues son más fuertes las consecuencias (sanciones sobre Rusia) de la guerra que la ganancia que se obtiene de la misma.

En complemento a esta investigación, sería necesario incluir diferentes fuentes de noticias económicas y no solo las minutas del Banco de la República, estas fuentes podrían ser, prensa económica, plataformas de información y negociación como Bloomberg o Reuters entre otras. Una evaluación ponderada del sentimiento de estas fuentes aportaría más información y robustez a este tipo de investigaciones.

Igualmente, este tipo de análisis podría incluir fuentes en diferentes idiomas, especialmente el inglés, dado que muchos eventos económicos internacionales afectan al mercado local colombiano y esta información, mayoritariamente, está disponible en inglés.

Referencias

*Andersen, T. G., Bollerslev, T., Diebold, F. X., & Vega, C. (2007). *Real-time price discovery in global stock, bond and foreign exchange markets*. Journal of International Economics. 7373(1), 251–277.

*Apel, M. and Grimaldi, M. (2012). *The information content of central bank minutes*. Riksbank Research Paper Series.No. 92.
http://archive.riksbank.se/Documents/Rapporter/Working_papers/2012/rap_wp261_120426.pdf. Acceso 12 enero 2022.

*Bai, Y. (2014). Cross-border sentiment: An empirical analysis on EU stock markets. Applied Financial Economics, 24(1), 259–290.

*Chen, Y. Y. (2010).Autoregressive distributed lag (ADL) model (2018). Acceso web: <http://mail.tku.edu.tw/chenyiyi/ADL.pdf>. Acceso 12 de Mayo de 2022.

*Jennifer Isasi (2021). "Análisis de sentimientos en R con 'syuzhet'". *Programming Historian en español* 5, Acceso web: <https://doi.org/10.46430/phes0051>. Acceso 26 de abril de 2022

*Moreno, A (junio 2021). *Aplicaciones del Procesamiento del Lenguaje Natural*. Instituto de Ingeniería del Conocimiento. Acceso web: <https://www.iic.uam.es/inteligencia/que-es-procesamiento-del-lenguaje-natural/>. Acceso 22 octubre 2021.

* Ministerio de Hacienda (4 de febrero de 2022) *El cierre fiscal preliminar de 2021 y el Plan Financiero de 2022 traen excelentes noticias en materia de crecimiento económico y sostenibilidad fiscal*. Acceso web: https://www.minhacienda.gov.co/webcenter/ShowProperty?nodeId=/ConexionContent/WCC_CLUSTER-185368//idcPrimaryFile&revision=latestreleased. Acceso 9 noviembre 2021.

* Racine, J. S. (2019). *Reproducible econometrics using R*. Oxford University Press, USA. 175-196.

*Taborda, R. (2015). *Procedural transparency in Latin American central banks under inflation targeting schemes*. A text analysis of the minutes of the Boards of Directors. *Ensayos sobre Política Económica* 33(76), 76-92.

*Tetlock, P. C. (2007). *Giving content to investor sentiment: The role of media in the stock market*. *Journal of Finance*, 62(3), 1139–1168.

*Tetlock, P., Saar-Tsechansky, M., &Macskassy, S. (2008). *More than words: Quantifying language to measure firms' fundamentals*. *Journal of Finance*, 63(3), 1437–1467.

*Gil Leiva, I; Rodriguez Muñoz, J.V. (1996). *El procesamiento del lenguaje natural aplicado al análisis del contenido de los documentos*. *Revista general de Información y Documentación*, 6 (2), 205-218.

*Gujarati, D. N., & Porter, D. C. (2011). *Econometría básica*. ed. Porto Alegre: AMGH, 617-670

*Verdejo M. (1994) *Procesamiento del lenguaje natural: fundamentos y aplicaciones*. UNED, Curso de Verano, julio 1994.

Anexos

Resultados según AIC para el mejor modelo ARDL

- RBancolombia ARDL(1,0)

<i>Mejores Modelos</i>	RBancolombia	Sentimiento	AIC
1	1	0	-288.2354
2	1	1	-286.4677
3	2	0	-283.4486
4	2	1	-281.7105
5	2	2	-280.1998
6	3	3	-276.7255
7	3	2	-275.8064
8	4	4	-272.6565
9	4	3	-272.1277
10	5	5	-268.0596

- REcopetrol ARDL(1,1)

<i>Mejores Modelos</i>	REcopetrol	Sentimiento	AIC
1	1	1	-225.4908
2	1	0	-223.2836
3	2	1	-220.9638
4	2	2	-219.8304
5	3	1	-216.8672
6	3	2	-215.8509
7	3	3	-213.8511
8	4	1	-212.9985
9	4	2	-211.9776
10	4	3	-209.9778
11	4	4	-208.2488
12	5	5	-203.5678

Resultados econométricos de R

A continuación, se presentan las gráficas de la evidencia empírica para Bancolombia y Ecopetrol. De igual manera la forma como se emplearon los diferentes paquetes en R.

● Colcap ARDL

```
Residuals:
  Min       1Q   Median       3Q      Max
-0.233840 -0.033391  0.001687  0.035596  0.121591

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.011034   0.007005   1.575   0.1179
L(RColcap, 1) 0.047896   0.090346   0.530   0.5970
Sentimiento  0.178184   0.082407   2.162   0.0327 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04975 on 115 degrees of freedom
(0 observations deleted due to missingness)
Multiple R-squared:  0.04029, Adjusted R-squared:  0.0236
F-statistic: 2.414 on 2 and 115 DF, p-value: 0.09399
```

● Bancolombia ARDL

```
Residuals:
  Min       1Q   Median       3Q      Max
-0.32317 -0.03950 -0.00298  0.04506  0.22719

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.022536   0.009826   2.294   0.0236 *
L(RBancolombia, 1) 0.059266   0.089568   0.662   0.5095
Sentimiento  0.285203   0.115629   2.467   0.0151 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06986 on 115 degrees of freedom
(0 observations deleted due to missingness)
Multiple R-squared:  0.05245, Adjusted R-squared:  0.03598
F-statistic: 3.183 on 2 and 115 DF, p-value: 0.04513
```

● Ecopetrol ARDL

```
Residuals:
  Min       1Q   Median       3Q      Max
-0.309288 -0.063781  0.001907  0.050496  0.244277

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.002549   0.014364  -0.177   0.8594
L(REcopetrol, 1) 0.150074   0.091350   1.643   0.1032
Sentimiento  0.301238   0.160480   1.877   0.0631 .
L(Sentimiento, 1) -0.329222   0.161844  -2.034   0.0443 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.09076 on 114 degrees of freedom
(0 observations deleted due to missingness)
Multiple R-squared:  0.06189, Adjusted R-squared:  0.0372
F-statistic: 2.507 on 3 and 114 DF, p-value: 0.06251
```

Programa de R

1. Como primer paso se instalan los siguientes paquetes:

```
#install.packages("RSelenium")
```

```
#install.packages("Rtools")
```

```
#install.packages("syuzhet")
#install.packages("remotes")
#remotes::install_github("jeffwong/fastVAR")
#install.packages("gets")
#install.packages("olsrr")
#install.packages("writexl")
#install.packages("SciViews")
#install.packages("ARDL")
#install.packages("zoo")
#install.packages("xts")
#Load packages
library(zoo)
library(xts)
library(SciViews)
library(gets)
library(syuzhet)
library(readxl)
library(RSelenium)
library(stringi)
library(stringr)
library(xml2)
library(tidyverse)
library(purrr)
library(pdftools)
library(httr)
```

```
library(sysid)
library(stats)
library(aTSA)
library(olsrr)
library(writexl)
library(ARDL)
```

2. En este paso realizamos la construcción del sentimiento de mercado a partir de las minutas del Banco de la República copiando las URL en un archivo.
3. Luego tenemos la conexión al servidor docker -- run -d -p 4445:4444 selenium/standalone-firefox:2.53.1 mediante:
 - `remDr <- remoteDriver(remoteServerAddr = "localhost", port = 4445L, browserName = "firefox")`
4. Cargamos la URL de las minutas desde el archivo local donde se halla guardado, para nuestro caso se hizo por partes por la extensión de los archivos que dificulta el rendimiento del programa, utilizando las siguientes funciones:
 - `dfm<-data.frame(read_excel("C:\\Users\\morok\\OneDrive\\Documents\\Minutas.xlsx"))`
`dfm_list <- pull(dfm)`
 - `dfm1<-data.frame(read_excel("C:\\Users\\morok\\OneDrive\\Documents\\Minutas1.xlsx"))`
`dfm_list1 <- pull(dfm1)`
 - `dfm2<-data.frame(read_excel("C:\\Users\\morok\\OneDrive\\Documents\\Minutas2.xlsx"))`

```
dfm_list2 <- pull(dfm2)
```

5. Procedemos con definir la función para extraer el texto de cada minuta y calcular la media del sentimiento de las mismas.

- ```
f_sentimiento <- function(url) {remDr$navigate(url) webElem <-
remDr$findElement(using = "xpath", value =
'//*[@id="block-banrep-content"]/div/div[3]/div')
minuta <-webElem$getElementText()[[1]]
s_v <- get_sentences(minuta)
s_v_sentiment <- get_sentiment(s_v,lang = "spanish")
result <- mean (s_v_sentiment)
result_df <- as.data.frame(result)}
```
- ```
f_sentimiento1 <- function(url) { remDr$navigate(url)  
webElem <- remDr$findElement( using = "xpath", value =  
'//*[@id="block-banrep-content"]/div/div[3]/div/div' )  
minuta <-webElem$getElementText()[[1]]  
s_v <- get_sentences(minuta)  
s_v_sentiment <- get_sentiment(s_v,lang = "spanish")  
result <- mean (s_v_sentiment)  
result_df <- as.data.frame(result) }
```
- ```
f_sentimiento2 <- function(url) { remDr$navigate(url)
webElem <- remDr$findElement(using = "xpath", value =
'//*[@id="block-banrep-content"]/div/div[2]/div')
minuta <-webElem$getElementText()[[1]]
```

```

s_v <- get_sentences(minuta)
s_v_sentiment <- get_sentiment(s_v,lang = "spanish")
result <- mean (s_v_sentiment)
result_df <- as.data.frame(result) }

```

6. Hallamos los sentimientos para las tres secciones en las que se dividió el total de minutas, es necesario que R abra el navegador y se dirija a la página asignada mediante los siguientes comandos:

- remDr\$open()
 

```

results <- dfm_list %>%
map_df(f_sentimiento)
remDr$close()

```
- remDr\$open()
 

```

results1 <- dfm_list1 %>%
map_df(f_sentimiento1)
remDr$close()

```
- remDr\$open()
 

```

results2 <- dfm_list2 %>%
map_df(f_sentimiento2)
remDr$close()

```
- write\_xlsx(results2,"C:\\Users\\morok\\OneDrive\\Documentos\\Resultados.xlsx")

7. Luego se realizó un ejemplo del sentimiento de mercado para una minuta

- `remDr$open()`

```
remDr$navigate(dfm_list[1])
```

```
webElem <- remDr$findElement(using = "xpath", value =
'//*[@id="block-banrep-content"]/div/div[3]/div')
```

```
minuta <- webElem$getElementText()[[1]]
```

```
s_v <- get_sentences(minuta)
```

```
s_v_sentiment <- get_sentiment(s_v, lang = "spanish")
```

8. Finalmente llegamos a la estimación econométrica mediante el modelo ARDL

que podemos encontrarlo en el siguiente link

web: <https://www.rdocumentation.org/packages/ARDL/versions/0.1.1>

- Hacemos que R lea el archivo de excel donde se encuentra la base de datos de los rendimientos de las acciones y los valores de los sentimiento de mercado obtenidos anteriormente:

```
dfRetornos <- data.frame(read_excel("C:\\Users\\-_-\\Desktop\\TESIS\\D
ata.xlsx"))
```

- Analizamos los datos para cada variable:

```
models <- auto_ardl(RColcap ~ Sentimiento , data = dfRetornos,
max_order = 5)
```

```
models <- auto_ardl(RBancolombia ~ Sentimiento , data = dfRetornos,
max_order = 5)
```

```
models <- auto_ardl(REcopetrol ~ Sentimiento , data = dfRetornos,
max_order = 5)
```

- Escogemos el mejor modelo entre el Top 20 de acuerdo con AIC para la variable RColcap (repetimos este procedimiento para las otras dos) :  
`models$top_orders`
- Aplicamos el mejor modelo ARDL(1,0) según R para el índice de los rendimientos de Colcap  
`ardl_10 <- models$best_model`
- Finalmente obtenemos los resultados con la función:  
`summary(ardl_10)`

## Lista de minutas

A continuación, se encuentra las URL de cada minuto que conformó la base de datos:

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-28-enero-2022>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-17-diciembre-2021>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-29-octubre-2021>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-30-septiembre-2021>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-30-julio-2021>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-28-junio-2021>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-30-abril-2021>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-26-marzo-2021>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-29-enero-2021>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-18-diciembre-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-27-noviembre-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-de-l-30-octubre-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-septiembre-25-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-agosto-31-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-julio-31-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-junio-30-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-mayo-29-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-abril-30-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-marzo-27-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-enero-31-2020>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-octubre-31-2019>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-septiembre-23-2019>

<https://www.banrep.gov.co/es/minutas-reunion-junta-directiva-del-banco-republica-julio-26-2019>

<https://www.banrep.gov.co/es/minutas-06-2019>

<https://www.banrep.gov.co/es/minutas-04-2019>

<https://www.banrep.gov.co/es/minutas-03-2019>

<https://www.banrep.gov.co/es/minutas-01-2019>

<https://www.banrep.gov.co/es/minutas-12-2018>

<https://www.banrep.gov.co/es/minutas-10-2018>

<https://www.banrep.gov.co/es/minutas-09-2018>

<https://www.banrep.gov.co/es/minutas-07-2018>

<https://www.banrep.gov.co/es/minutas-06-2018>

<https://www.banrep.gov.co/es/minutas-04-2018>

<https://www.banrep.gov.co/es/calendario-reuniones-y-minutas/minutas-reunion-junta-directiva-del-banco-republica-marzo-20-2018-0>

<https://www.banrep.gov.co/es/minutas-01-2018>

<https://www.banrep.gov.co/es/minutas-12-2017>

<https://www.banrep.gov.co/es/minutas-11-2017>

<https://www.banrep.gov.co/es/minutas-10-2017>  
<https://www.banrep.gov.co/es/minutas-09-2017>  
<https://www.banrep.gov.co/es/minutas-08-2017>  
<https://www.banrep.gov.co/es/minutas-07-2017>  
<https://www.banrep.gov.co/es/minutas-06-2017>  
<https://www.banrep.gov.co/es/minutas-05-2017>  
<https://www.banrep.gov.co/es/minutas-04-2017>  
<https://www.banrep.gov.co/es/minutas-03-2017>  
<https://www.banrep.gov.co/es/minutas-02-2017>  
<https://www.banrep.gov.co/es/minutas-01-2017>  
<https://www.banrep.gov.co/es/minutas-12-2016>  
<https://www.banrep.gov.co/es/minutas-11-2016>  
<https://www.banrep.gov.co/es/minutas-10-2016>  
<https://www.banrep.gov.co/es/minutas-09-2016>  
<https://www.banrep.gov.co/es/minutas-08-2016>  
<https://www.banrep.gov.co/es/minutas-07-2016>  
<https://www.banrep.gov.co/es/minutas-06-2016>  
<https://www.banrep.gov.co/es/minutas-05-2016>  
<https://www.banrep.gov.co/es/minutas-04-2016>  
<https://www.banrep.gov.co/es/minutas-03-2016>  
<https://www.banrep.gov.co/es/minutas-02-2016>  
<https://www.banrep.gov.co/es/minutas-01-2016>  
<https://www.banrep.gov.co/es/minutas-12-2015>  
<https://www.banrep.gov.co/es/minutas-11-2015>  
<https://www.banrep.gov.co/es/minutas-10-2015>

<https://www.banrep.gov.co/es/minutas-09-2015>  
<https://www.banrep.gov.co/es/minutas-08-2015>  
<https://www.banrep.gov.co/es/minutas-07-2015>  
<https://www.banrep.gov.co/es/minutas-06-2015>  
<https://www.banrep.gov.co/es/minutas-05-2015>  
<https://www.banrep.gov.co/es/minutas-04-2015>  
<https://www.banrep.gov.co/es/minutas-03-2015>  
<https://www.banrep.gov.co/es/minutas-02-2015>  
<https://www.banrep.gov.co/es/minutas-01-2015>  
<https://www.banrep.gov.co/es/minutas-12-2014>  
<https://www.banrep.gov.co/es/minutas-11-2014>  
<https://www.banrep.gov.co/es/minutas-10-2014>  
<https://www.banrep.gov.co/es/minutas-09-2014>  
<https://www.banrep.gov.co/es/minutas-08-2014>  
<https://www.banrep.gov.co/es/minutas-07-2014>  
<https://www.banrep.gov.co/es/minutas-06-2014>  
<https://www.banrep.gov.co/es/minutas-05-2014>  
<https://www.banrep.gov.co/es/minutas-04-2014>  
<https://www.banrep.gov.co/es/minutas-03-2014>  
<https://www.banrep.gov.co/es/minutas-02-2014>  
<https://www.banrep.gov.co/es/node/34441>  
<https://www.banrep.gov.co/es/node/34218>  
<https://www.banrep.gov.co/es/node/34063>  
<https://www.banrep.gov.co/es/node/33553>  
<https://www.banrep.gov.co/es/node/33459>

<https://www.banrep.gov.co/es/node/33114>

<https://www.banrep.gov.co/es/node/32767>

<https://www.banrep.gov.co/es/node/32454>

<https://www.banrep.gov.co/es/node/32238>

<https://www.banrep.gov.co/es/node/32239>

<https://www.banrep.gov.co/es/node/32082>

<https://www.banrep.gov.co/es/node/31593>

<https://www.banrep.gov.co/es/node/31313>

<https://www.banrep.gov.co/es/node/31020>

<https://www.banrep.gov.co/es/node/30847>

<https://www.banrep.gov.co/es/node/29484>

<https://www.banrep.gov.co/es/node/29097>

<https://www.banrep.gov.co/es/node/28177>

<https://www.banrep.gov.co/es/node/28042>

<https://www.banrep.gov.co/es/node/27053>

<https://www.banrep.gov.co/es/node/26463>

<https://www.banrep.gov.co/es/node/25145>

<https://www.banrep.gov.co/es/node/29116>