



Universidad
Zaragoza

Trabajo Fin de Máster

Análisis de sentimiento: Influencia de las noticias en
el Mercado Alternativo Bursátil

Autora

Lorena Corbacho Ballesteros

Directora

Cristina Ortiz Lázaro

Facultad de economía y empresa / Universidad de Zaragoza
2020

RESUMEN

Actualmente se tiene disponible en la web un gran volumen de información accesible para todos los públicos. Es por ello que muchas empresas publican sus noticias económico/financieras en prensa digital o blogs de opinión. El objetivo de esta investigación es analizar si la evolución de unas variables financieras como son la rentabilidad o el volumen de títulos, están influenciadas por las publicaciones que realizan las empresas que cotizan en el Mercado Alternativo Bursátil, desde sus inicios en 2008 hasta diez años después. Se comprueba que las publicaciones de noticias no afectan tanto a la rentabilidad, pero si al volumen de títulos que cotizan en este mercado.

Palabras clave: Mercado Alternativo Bursátil (MAB), minería de datos, análisis de sentimiento, titulares de noticias y web scraping.

ABSTRACT

Currently, a large volume of information is available on the web, accessible to all audiences. That is why many companies publish their economic / financial news in the digital press or opinion blogs. The objective of this research is to analyze whether the evolution of financial variables such as profitability or the volume of securities, are influenced by the publications made by companies listed on the Alternative Stock Market, from its inception in 2008 to ten years later. . It is found that news publications do not affect profitability as much, but the volume of titles listed in this market.

Keywords: Alternative Stock Market (MAB), data mining, sentiment analysis, news headlines and web scraping.

INDICE

1. INTRODUCCIÓN	3
2. REVISIÓN DE LA LITERATURA	7
3. DATOS	9
3.1. EMPRESAS	9
3.2. FINANCIEROS	11
3.3. NOTICIAS	17
4. METODOLOGÍA	20
4.1. ANÁLISIS DE SENTIMIENTO	20
4.2. ANÁLISIS EMPÍRICO	24
4.2.1. EFECTOS SOBRE LA RENTABILIDAD	25
4.2.2. EFECTO SOBRE ROTACIÓN: VOLUMEN	31
4.2.3. REGRESIÓN DEL MERCADO	34
5. CONCLUSIONES	36
6. BIBLIOGRAFÍA	37
ANEXO 1	41
ANEXO 2	46

1. INTRODUCCIÓN

De acuerdo con la teoría clásica y la hipótesis de mercados eficientes, para la toma de decisiones los agentes actuaban de manera racional dada la información perfecta que el mercado ofrecía (Fama, 1970). La teoría clásica, por tanto, se fundamenta en tres hipótesis básicas: racionalidad de los individuos, información perfecta y mercados eficientes que incorporan de manera inmediata la nueva información. Por tanto, se entiende como inversor aquel que toma decisiones racionales, teniendo libertad de información y en mercados libres de fricciones (Elton et al., 2004).

Sin embargo, posteriormente se ha ido demostrando que el mercado no funciona de tal manera. Más tarde nace una nueva corriente de pensamiento basada en las finanzas del comportamiento (“Behavioral Finance”), que demuestra que el ser humano no actúa de manera racional y que se cometen errores a la hora de procesar la información. Es importante destacar que esta nueva corriente de pensamiento no rechaza la teoría clásica, sino que se nutre de estas finanzas tradicionales. De manera que las tres hipótesis básicas se debilitan, porque el comportamiento humano no es perfecto, no siempre se actúa de manera racional y se cometen errores a la hora de procesar la información.

Nace en este punto la teoría propuesta por Kahneman y Tversky (1979): Teoría Prospectiva. En dicha propuesta, estos autores encuentran evidencia de que existen ciertos comportamientos humanos que no cumplen la racionalidad. Esta teoría afirma que cada decisión que se toma tiene un modelo mental (*mental frame*), es decir, que cada decisión puede ser manipulada por unos sesgos y hacer cambiar la decisión.

Muchos autores han investigado que las decisiones económicas pueden estar influenciadas por estos sesgos comentados. Actualmente, existen numerosos economistas que demuestran que los inversores no se comportan de manera totalmente racional, sino que existen una serie de sesgos cognitivos que hacen que las decisiones se vean influenciadas. Conocer estos sesgos a la hora de invertir en los mercados puede ayudar a evitar esos “errores”. El cerebro humano trata de acortar lo máximo posible el tiempo de análisis de la información, y para ello toma unos atajos y sesgos emocionales (Ferrer de Lucas, 2015). Estos sesgos también se denominan heurísticos, aspecto que fue muy estudiado por el nobel de economía de 2002 Daniel Kahneman, puesto que defiende que los seres humanos no toman sus decisiones de manera racional (Daniel Kahneman et al., 1982).

Como se ha comentado, la información que ofrece el mercado no es perfecta y en numerosas ocasiones tiene un trasfondo. Son muchas las empresas que ofrecen información para que los inversores la procesen y actúen en consonancia a ella. Ahora bien, ¿realmente los inversores tienen en cuenta la información disponible cuando toman sus decisiones de inversión?

En este aspecto, cabe mencionar el heurístico de representatividad (Tversky y Kahneman, 1974), que se basa en realizar juicios en la medida que unas situaciones son

semejantes a otras. Dicho de otra forma, se trata de observar la probabilidad de que un individuo pertenezca a una categoría por la semejanza que tienen el resto de individuos.

Actualmente gracias a la evolución tecnológica se tiene un gran volumen de información al alcance de todos los agentes, disponible en múltiples fuentes de información y sobre cualquier aspecto. No obstante, lo complicado está en la manera de procesar el gran volumen de información del que se dispone, ya que esta crece a un ritmo desenfrenado. En este punto es donde cabe señalar el concepto clave de “Big Data”, que hace referencia a las 3 gran V: gran Volumen, gran Velocidad y gran Variedad de datos (Gartner IT Glossary, 2018).

La disponibilidad de tal volumen de información requiere herramientas computacionales para extraer, almacenar y procesar esta misma. Aquí es donde surge el concepto de KDD (*Knowledge Discovery in Databases*), que fue acuñado en 1989 y se refiere a todo ese proceso de extracción de información de una base de datos (Riquelme et al., 2006). Este proceso comprende las siguientes fases:

- Crear la base de datos objetivo, extrayendo los datos necesarios.
- Preparar toda la información disponible, adaptar los datos a lo que realmente se necesita, obviando aquellos que sean innecesarios, es decir, proceso de limpieza de datos.
- Minería de datos, fase clave de todo el proceso que se explicará con más detalle posteriormente.
- Interpretación final de la muestra obtenida, resultados que se obtienen.

La minería de datos es la etapa más importante, que consiste en la aplicación de algoritmos determinados para extraer una serie de modelos (Pérez y Santin, 2008). Este proceso puede abarcar dos grandes ámbitos: puede estar enfocado para tareas descriptivas (con una sucesión de datos, descubrir unos patrones concretos) o tareas más predictivas (teniendo unos datos disponibles averiguar un patrón de comportamiento). Para ello se utilizan una sucesión de herramientas como, por ejemplo:

- Clasificación: agrupar los datos según unas categorías.
- Regresión: estimar relaciones entre variables, enlazar un dato a un valor real de una variable.
- Clustering: agrupación de casos que comparten un aspecto común.

Ahora bien, la minería de datos tiene muchos ámbitos de aplicación, y a raíz de eso han surgido nuevos conceptos que provienen de este método. Uno de ellos es la minería de texto (“Text Mining”), que nace con el objetivo de descubrir tendencias, agrupaciones y desviaciones de la gran cantidad textual de la que se dispone (Montes - y - Gómez 2001). Es decir, se trata de un proceso por el cual se obtiene nuevo conocimiento de un conjunto de datos en formato texto. Esta técnica tiene dos fases primordiales: la primera consiste en un preprocesamiento, donde se transforman una serie de datos para facilitar su análisis, y una segunda etapa de descubrimiento, que trata de investigar patrones comunes o nuevo conocimiento.

Del mismo modo, la minería de texto incluye a su vez diversas aplicaciones, entre ellas el análisis de sentimiento o minería de opiniones (Ruiz, 2018). Esta herramienta cada vez está tomando más relevancia debido a la gran cantidad de opiniones que se pueden extraer de las múltiples fuentes de información de las que se dispone. El análisis de sentimiento consiste en intentar clasificar los textos de acuerdo a lo que quiera expresar su autor. En otras palabras, otorgar un valor a cada texto con la carga emocional que intenta trasladar su escritor. Existen diferentes variables que se pueden obtener en un análisis de sentimiento (Bravo-Marquez et al. 2014):

- Polaridad: una variable categórica con valores positivos, negativos o neutros. Esta variable se nutre de léxicos de palabras positivos y negativos.
- Intensidad: se refiere a la intensidad del sentimiento positivo o negativo.
- Emoción: extraer el estado de ánimo de un texto, como podría ser la alegría, la sorpresa o la tristeza.

Una vez se dispone de esta herramienta tan concreta, se puede retomar la pregunta anteriormente planteada: ¿son los inversores conscientes del volumen de información que tienen a su disposición? Esta información, ¿la tienen en cuenta a la hora de tomar sus decisiones de inversión?

De acuerdo con autores como Isen (1987) o Schwarz (2002), los inversores incorporan las emociones como información en sus elecciones, por tanto, estos negocian de forma diferente en función de esta variable.

En este proyecto, se va a disponer de información de carácter de opinión para tratar de responder a estas preguntas. Este tipo de información que ofrecen los medios de comunicación, de acuerdo con Schiller (2000) juegan un papel importante para los movimientos del mercado. En este caso, se dispondrá de información basada en noticias encontradas en la web sobre un mercado concreto. Este va a ser el Mercado Alternativo Bursátil (en adelante MAB). Se analizarán por tanto todas aquellas noticias que hayan aparecido en relación a las empresas que componen este mercado.

Respecto a la investigación que se va a llevar a la cabo, se ha elegido este mercado debido a la volatilidad del mismo. Es un mercado muy volátil, puesto que está compuesto entre otros, por empresas en expansión, con previsión de crecimiento, y esto hace que los movimientos sean extremos, tanto al alza como a la baja. Y este tipo de empresas son las que van a formar parte de la muestra a analizar.

El MAB se constituyó oficialmente en febrero de 2006, aunque no fue hasta 2008 cuando empezó a estar operativo. Este mercado apareció con la finalidad de ofrecer una alternativa a empresas de dimensión reducida. En otras palabras, quisieron ofrecer una oportunidad de negociación a las PYMES (Pequeñas y Medianas empresas) y crear un mercado similar a Alternext o AIM¹ pero en España (Sánchez, 2008). De acuerdo con

¹ AIM (Alternative Investment Market) en Londres y Alternext en París, son mercados enfocados para pequeñas y medianas empresas.

BME (Bolsas y Mercados Españoles, 2008) las razones para la creación de este mercado se resumen en tres:

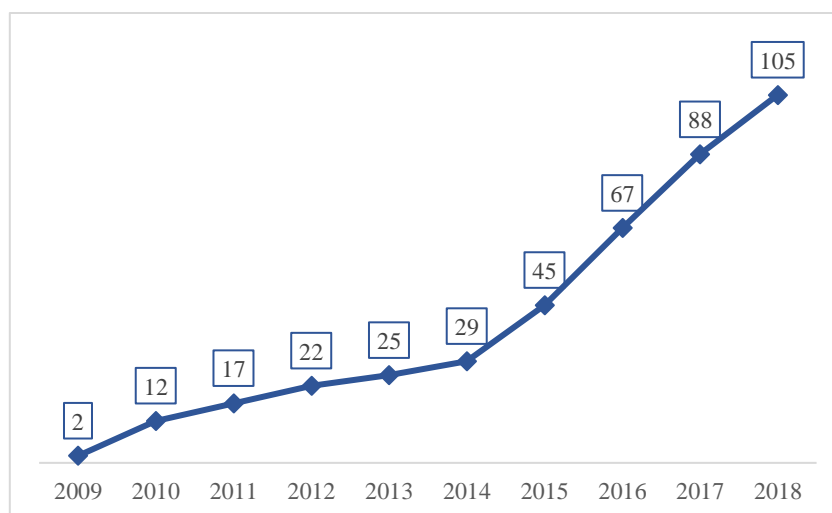
- Facilitar el acceso al mercado a ciertas empresas en un contexto europeo con ciertas regulaciones más rígidas cada vez.
- Conceder al mercado una alternativa bursátil que se autorregule y con capacidad de adaptación.
- Plantear un mercado a medida de las empresas de reducida capitalización con la adecuada transparencia para los inversores.

Dado que para la fecha de creación del MAB, el tejido empresarial del tipo de empresas que podrían cotizar era de aproximadamente el 99%, teniendo una fuerza laboral del 79% y aportaban un 65% del PIB, resultaba un mercado muy interesante (SUBDIRECCION GENERAL DEL APOYO A LAS PYME, 2009). Además, si las empresas lograban entrar en este mercado también obtendrían una serie de beneficios como, por ejemplo: ganarían notoriedad, se financiarían su crecimiento, y les podría servir como un entrenamiento para alcanzar el Mercado Continuo.

Los requisitos para poder acceder al MAB eran sencillos y los cumplían un gran volumen de empresas como se ha visto, y eran los siguientes: poseer constitución de Sociedad Anónima, tener una actividad de negocio demostrada a través de la entrega de una información descriptiva y financiera de la empresa, obtener un valor mínimo de *free float*² de dos millones de euros, y, por último, disponer de un Asesor Registrado durante todo el proceso de colocación y cotización en el mercado.

Esto llevó a que muchas empresas se plantearan esta opción, y a lo largo de la historia del mercado son muchas las que han pasado por este mercado, tal y como se muestra en el Gráfico 1:

Gráfico 1: Evolución número de empresas que integran el MAB



Fuente: Bolsas y Mercados, 2019

² Free float se refiere a las acciones que se tienen en circulación, capital que está cotizado.

Se observa como en 2015 se produce una aceleración en el número de empresas que integran el MAB (González Nieto-Márquez, 2019).

El trabajo presenta la siguiente estructura: en primer lugar, se realizará un breve repaso de la literatura previa, es decir, se hará un recorrido sobre lo que han investigado otros autores en el ámbito; posteriormente se analizarán los datos obtenidos, así como los métodos utilizados para su extracción, donde se encuentran datos de las empresas analizadas y aquellos de carácter financiero. Se continuará con la metodología que se ha utilizado para comprobar las hipótesis básicas de la investigación, con sus posteriores resultados y conclusiones.

2. REVISIÓN DE LA LITERATURA

Existe una extensa literatura que analiza la influencia que una variable de sentimiento del inversor pueda tener sobre aspectos económicos y/o.

El enfoque tradicional se basaba en la realización de encuestas a pequeña escala. Sin embargo, este método era realmente costoso para las agencias que lo realizaban. El alcance y la oportunidad del mismo no era el más idóneo, puesto que las encuestas a menudo se basaban en muestras de individuos relativamente pequeñas, lo que implicaba algún problema de muestreo (Ludvigson, 2004). Además, la frecuencia de publicación de este método era lento, de manera mensual o incluso trimestral.

Nace así otro enfoque basado en el procesamiento de publicaciones gracias a métodos computacionales. Debido al gran volumen de información que se maneja, en los últimos años muchos autores se han visto obligados a desarrollar estos métodos para poder procesar la información y extraer variables de sentimiento. Herramientas con este enfoque son utilizadas en muchas aplicaciones comerciales, como pueden ser a través de las redes sociales.

No obstante, este enfoque tradicional se sigue utilizando para elaborar índices basados en encuestas. En EEUU destacan los dos siguientes: *Índice de confianza del consumidor de Michigan* y el *Índice de confianza del consumidor ("Conference Board")*³ (Shapiro et al. 2020). Para el caso de la Zona Euro, una agencia llamada Sentix, también elabora un índice de confianza del inversor que recibe el mismo nombre que la agencia. Este *Índice de confianza del inversor Sentix* es un indicador que se elabora desde 2001 con una periodicidad mensual y mide la confianza que tienen los inversores en la actividad económica de la Zona Euro. Esta agencia realiza una encuesta a más de 3.100 inversores, tanto particulares como institucionales en 20 países diferentes (AndBank, 2014).

³ Estos índices se basan en encuestas donde preguntan a los hogares sobre su situación y su perspectiva con respecto a las finanzas sociales. También existen otros índices similares que se basan en encuestas a grandes directivos que son los encargados de tomar decisiones empresariales.

Para perfilar este tema se inicia el recorrido temporal con el estudio realizado por Ricciardi (2008), donde este autor examina estudios y analiza el efecto negativo que tienen la toma de decisiones en cuestiones de género. Parte de la hipótesis de que las mujeres van a manifestar una mayor preocupación que los hombres por sus inversiones. Por su parte, Cruz (2009) analiza cómo los cambios en la confianza del inversor se ven reflejados en las crisis financieras de los mercados emergentes de México.

En el caso de García (2013), estudia el efecto que puede tener el sentimiento sobre el precio de los activos bursátiles a lo largo del siglo XX. Para ello, obtiene una variable de sentimiento calculada como una fracción de palabras positivas y negativas de las noticias financieras del periódico emblemático “The New York Times”. La conclusión más clara que obtiene es que el contenido de las noticias predice el rendimiento de las acciones a frecuencia diaria en el periodo analizado de 1905 a 2005, sobre todo durante los periodos de recesión en los EEUU.

Un año más tarde, Li et al. (2014) evalúan el impacto que pueden tener las noticias financieras sobre el rendimiento del precio de las acciones. Para ello, utilizan una base de datos formada por noticias financieras en Hong Kong de 2003 a 2008 y a través de un diccionario psicológico de Harvard, miden cuantitativamente los artículos proporcionándoles un valor de sentimiento.

Otro autor como Fraiberger (2016) realiza una investigación sobre cómo una variable de sentimiento extraída de los artículos de noticias puede influir en el crecimiento del PIB. Para ello utiliza un conjunto de datos formado por noticias económicas de todo el mundo y mide la cantidad neta de expresiones positivas en el conjunto de datos para un periodo desde 1987 a 2013 en una serie de países. Concluye que para 10 de los 12 países analizados las medidas de opinión obtenidas de las noticias influyen en las fluctuaciones del PIB. En la misma línea, Baker et al. (2016) elaboran un índice de incertidumbre de política económica basado en más de 12.000 artículos de periódicos en el marco de los Estados Unidos. Lo más interesante de este estudio es que el índice representa grandes movimientos de incertidumbre en fechas muy destacadas de la economía estadounidense, como pueden ser: elecciones presidenciales, ataque del 11 de septiembre o la caída de Lehman Brothers.

Otros autores como Nyman et al. (2018), investigan sobre medidas de sentimiento y confianza, cómo estas juegan un papel importante para impulsar los desarrollos del sistema financiero, en el contexto de la crisis financiera de 2007. Obtienen un conjunto de datos formado por noticias económicas de varias fuentes e informes realizados por brokers. Y concluyen que cambios en las variables analizadas se correlacionan con los eventos financieros del mercado. También Thorsrud (2017) examina cómo los agentes económicos utilizan una gran cantidad de información incluida en los medios de comunicación para tomar sus decisiones y con ellas dar forma a las fluctuaciones agregadas de la economía. Para ello utiliza un periódico diario de negocios de Noruega obteniendo información durante un largo periodo. Demuestra que con la información que ofrecen las noticias económicas del periódico es posible predecir el crecimiento trimestral del PIB de su país.

En la misma línea aparece la publicación de otros autores (Calomiris and Mamaysky, 2019) que desarrollan una metodología de clasificación para el contexto y el contenido de un corpus de artículos de noticias extrayendo una variable de sentimiento. El objetivo que mantienen estos autores es predecir el riesgo y el rendimiento de mercados bursátiles en 51 economías desarrolladas y emergentes mediante la metodología que han desarrollado con la variable de sentimiento.

Uno de los trabajos más recientes es el de Shapiro et al. (2020) que elaboran una herramienta de análisis de sentimiento para los artículos de periódicos económicos y financieros. Demuestran en primer lugar, que el sentimiento de las noticias diarias es capaz de predecir los movimientos de las mediciones de los consumidores basadas en encuestas. Y, en segundo lugar, relacionan esta variable con variables macroeconómicas, obteniendo la conclusión de que sentimientos positivos aumentan el consumo, la producción y las tasas de intereses, amortiguando el impacto de la inflación.

3. DATOS

3.1. EMPRESAS

Se han analizado un total de 50 empresas, aquellas que han formado parte del MAB en el periodo que comprende el análisis (2008-2018), clasificadas dentro del mismo como *empresas en expansión*. En la Tabla 1 se puede observar la relación de todas las empresas con su ISIN.

A su vez, se incluye en la tabla la fecha de introducción en el MAB para cada una de las empresas, y la fecha de exclusión del mismo si se hubiera dado el caso.

Tabla 1: Empresas que componen el MAB, 2008-2018

<i>ISIN</i>	<i>EMPRESA</i>	<i>INTRODUCCIÓN</i>	<i>EXCLUSIÓN</i>
ES0156303002	INKEMIA	21/12/2012	
ES0109659013	AB-BIOTICS SA	19/07/2010	
ES0184980003	ADL BIONATUR	26/01/2012	
ES0105102000	AGILE CONTENT	23/11/2015	
ES0105366001	ALQUIBER QUALITY SA	11/07/2018	
ES0113312005	ALTIA	01/12/2010	
ES0105148003	ATRY'S HEALTH SA	22/07/2016	
ES0171957014	BODACLICK	30/06/2010	28/10/2014
ES0116162068	CARBURES	23/03/2012	16/10/2018
ES0112320009	CATENON SA	06/06/2011	
ES0105038006	CERBIUM HOLDING	28/07/2014	
ES0105121000	CLERHP ESTRUCTURAS	10/03/2016	
ES0105152005	CLEVER GLOBAL SA.	22/12/2016	

Fuente: Elaboración propia a partir de la web Bolsas y Mercados

Tabla 1: Empresas que componen el MAB, 2008-2018 (continuación)

ISIN	EMPRESA	INTRODUCCIÓN	EXCLUSIÓN
ES0121071007	COMMCENTER	30/12/2010	
BG1100005179	EBIOSS ENERGY	05/07/2013	
ES0105227005	ELZINC (ASTURIANA DE LAMINADOS)	19/06/2017	
ES0105042008	EUROCONSULT ENGINEERING CONSULTING GROUP SA.	05/11/2014	07/03/2018
ES0133421000	EUROESPES, SA	16/02/2011	
ES0133443004	EURONA WIRELESS TELECOM SA	15/12/2010	
ES0105029005	FACEPHI BIOMETRIA, S.A.	01/07/2014	
ES0105093001	GIGAS HOSTING SA	03/11/2015	
ES0158252033	GOWEX	15/06/2010	02/07/2014
ES0105293007	GREENALIA	01/12/2017	
ES0105079000	GREENERGY RENOVABLES SA	08/07/2015	
ES0143328005	GRINÓ ECOLOGIC	29/07/2011	
ES0162603007	GRUPO NOSTRUM RNL, S.A	10/11/2010	05/03/2014
ES0105049003	HOME MEAL REPLACEMENT SA	04/12/2014	15/11/2018
ES0171613005	IFFE FUTURA	04/12/2013	
ES0147791018	IMAGINARIUM	01/12/2009	
ES0105083002	INCLAM SOCIEDAD ANONIMA	29/07/2015	
ES0105089009	LLEIDANETWORKS SERVEIS TELEMATICS SA	09/10/2015	
ES0184696013 //ES0184696104	MASMOVIL	30/03/2012	22/06/2017
ES0161857018	MEDCOM TECH, SA	26/03/2010	
ES0105216008	MONDO TV IBEROAMERICA	23/12/2016	
ES0105062022	NBI BEARINGS EUROPE SA	08/04/2015	
ES0165946015	NEGOCIO & ESTILO DE VIDA	07/06/2010	15/02/2012
ES0105081006	NEOL BIO	29/10/2015	30/08/2017
ES0105270005	NETEX KNOWLEDGE FACTORY SA	31/10/2017	
ES0166198012	NEURON BIO	02/07/2010	25/09/2017
ES0105020004	NPG TECHNOLOGY	22/04/2014	30/04/2015
ES0105221008	PANGAEA ONCOLOGY SA	29/12/2016	
ES0105225009	ROBOT SA	27/02/2018	
ES0158746026	SEAFOOD GLOBAL PROCESSOR // LUMAR	06/07/2011	26/01/2015
ES0131703003	SECUOYA, GRUPO DE COMUNICACION SA	28/07/2011	
ES0176406009	SUAVITAS // VOUSSE	12/04/2012	30/01/2017
ES0105076006	TECNOQUARK	21/12/2015	29/10/2018
ES0105097002	THINK SMART SA	17/11/2015	14/05/2018
ES0105344008	TIER 1 TECHNOLOGY SOCIEDAD ANONIMA	26/06/2018	
ES0105156006	VOZTELECOM OIGAA360 SA.	28/07/2016	
ES0184849018	ZINKIA ENTERTAINMENT, SA	01/07/2009	

Fuente: Elaboración propia a partir de la web Bolsas y Mercados⁴

⁴ <https://www.bolsasymercados.es/mab/esp/Home.aspx>

3.2. FINANCIEROS

Para la extracción de la información bursátil se ha utilizado la base de datos *Eikon Datastream*⁵. Esta plataforma se define como un recurso financiero, ofrecido por la empresa de información Thomson Reuters.

De esta plataforma se han extraído variables bursátiles que permiten el análisis financiero. En primer lugar, se ha obtenido para cada una de las empresas la variable TOTAL RETURN INDEX, un índice de rendimiento que es bastante fiel para mostrar el comportamiento de los activos. Este índice tiene en cuenta la reinversión de los dividendos que reparten las empresas, y es por ello que refleja el rendimiento de los activos de manera más exacta. Con este índice se ha realizado el cálculo de la rentabilidad para cada una de las empresas de acuerdo a la ecuación (1):

$$Rentabilidad_t = \frac{TIR_t - TIR_{t-1}}{TIR_{t-1}} \quad (1)$$

Para el caso del MAB, se ha extraído la variable PRICE INDEX, y se ha calculado la rentabilidad de la misma manera que se muestra anteriormente.

Para el cálculo de la rentabilidad de cada una de las empresas, se ha corregido esta misma con la rentabilidad de dos índices fundamentales para las PYMES de nuestro país: Índice General de la Bolsa de Madrid (IGBM) y IBEX Small Cap. El primero de ellos comprende un número de valores no fijo, que cumplen dos requisitos de liquidez: el volumen de contratación tiene que ser superior a tres millones de euros en el semestre anterior a la entrada, la rotación del capital tiene que ser superior al 10% de su capitalización en base anual, y la frecuencia de contratación tiene que ser superior al 50%. Por su parte, el IBEX Small Cap congrega las cuatro bolsas españolas que cotizan en el Sistema de Interconexión Bursátil Electrónico (SIBE), y en concreto agrupa las empresas cotizadas de pequeña capitalización. Los requisitos para formar parte de este índice son los siguientes: no formar parte del IBEX 35, tener una ratio free float superior al 15% y que la rotación anualizada sobre el capital flotante real tiene que ser superior al 15%⁶.

En el Gráfico 2 se muestra la rentabilidad media anual de cada una de las empresas analizadas, mientras que en el Gráfico 3 y el Gráfico 4 muestran la rentabilidad media anual corregida por el IGBM y el IBEX Small Cap, respectivamente. Se observa que patrón de comportamiento es similar en los tres casos, resaltando tres momentos que parecen ser claves:

- En el año 2014, la empresa VOUSSE (antigua SUAVITAS) incrementa su rentabilidad media anual, para normalizarse de nuevo al año siguiente. Es

⁵ Se agradece al Departamento de Contabilidad y Finanzas de la Universidad de Zaragoza la cesión de la licencia para su uso académico (<https://eikon.thomsonreuters.com/index.html>).

⁶ Consultar normativa Bolsas y Mercados disponible en la web: <https://www.bolsasymercados.es/mab/esp/ECR/Normativa.aspx>

posible que este incremento tenga que ver con dos hechos clave: el primero es que durante ese año la sociedad cambió su denominación, y el segundo es que en el mismo año adquirió a la sociedad HEDONAI⁷.

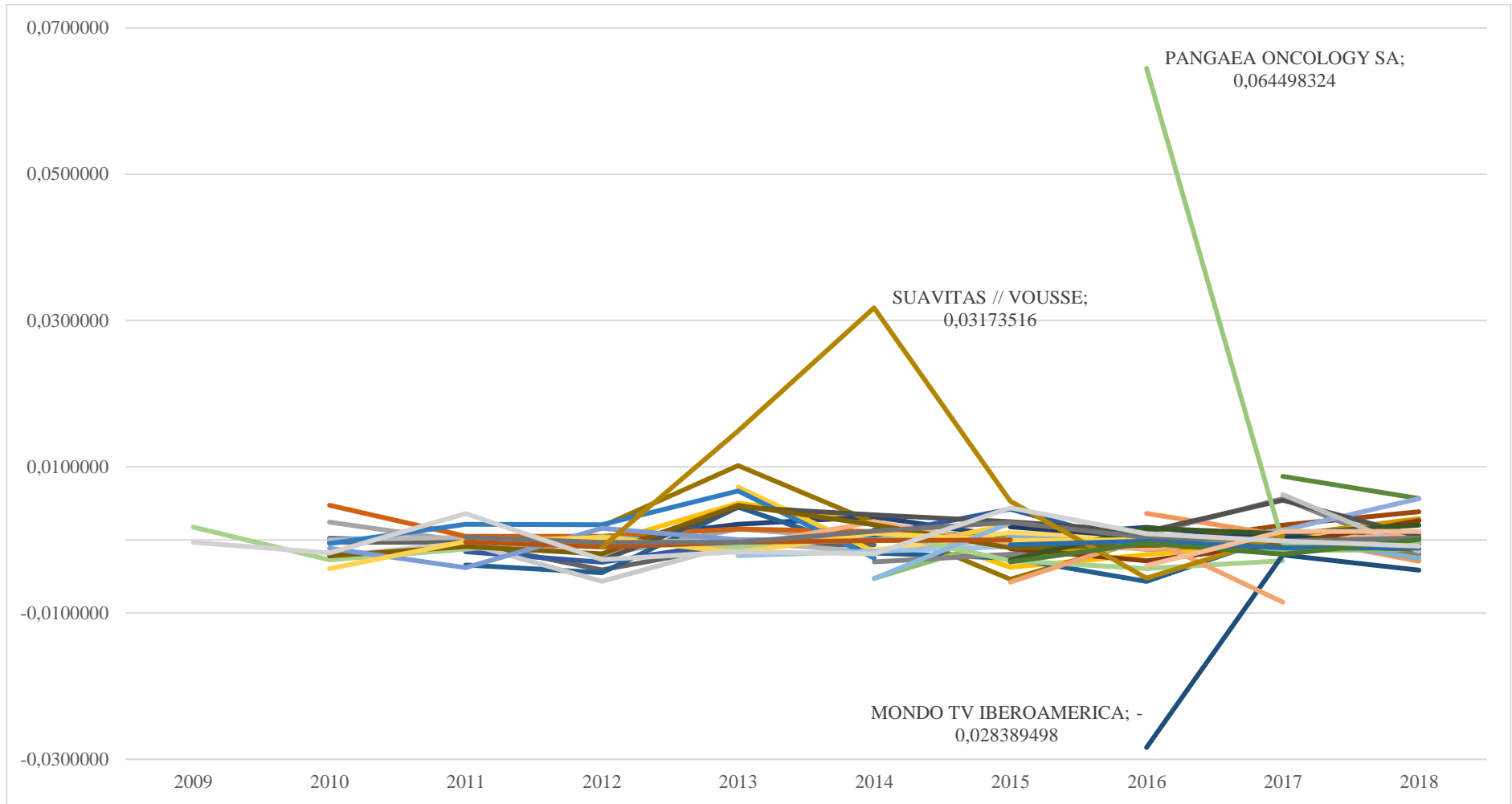
- En 2016 la empresa MONDO TV IBEROAMERICA comienza su camino con una rentabilidad negativa, bastante más elevada a otras empresas en su misma situación de entrada.
- En el mismo año, la empresa PANGAEA ONCOLOGY es introducida en el MAB con una rentabilidad bastante alta en comparación al resto de empresas cuando han sido introducidas.

Por otra parte, también se va a realizar un análisis sobre el volumen negociado en cada una de las empresas analizadas para el periodo correspondiente a 2008 – 2018. Para ello se han descargado de la misma plataforma de *Datastream* la ratio de rotación: *Turnover by Volumen* (VO), que representa el número de acciones negociadas en un día.

De acuerdo con el Gráfico 5, la empresa CARBURES es la que alcanza un valor máximo en comparación con el resto de empresas en cuanto a rotación de volumen. Por su parte, GOWEX y MASMOVIL también tienen una gran cifra de rotación en el volumen.

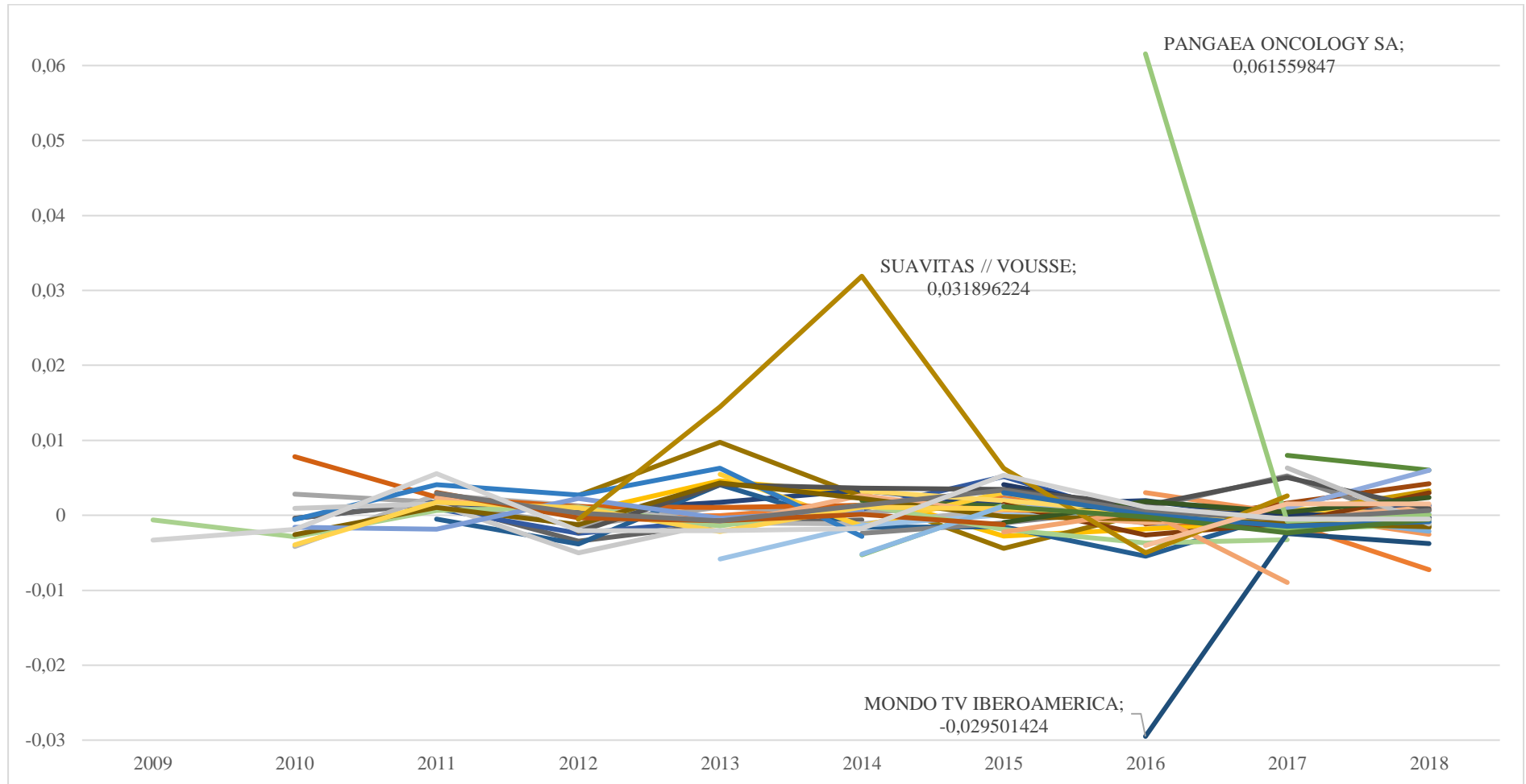
⁷ Comunicaciones de hechos relevantes de VOUSSE publicadas en Bolsas y Mercados a 2 de octubre de 2014 y 26 de noviembre de 2014.

Gráfico 2: Rentabilidad media anual, 2008 – 2018



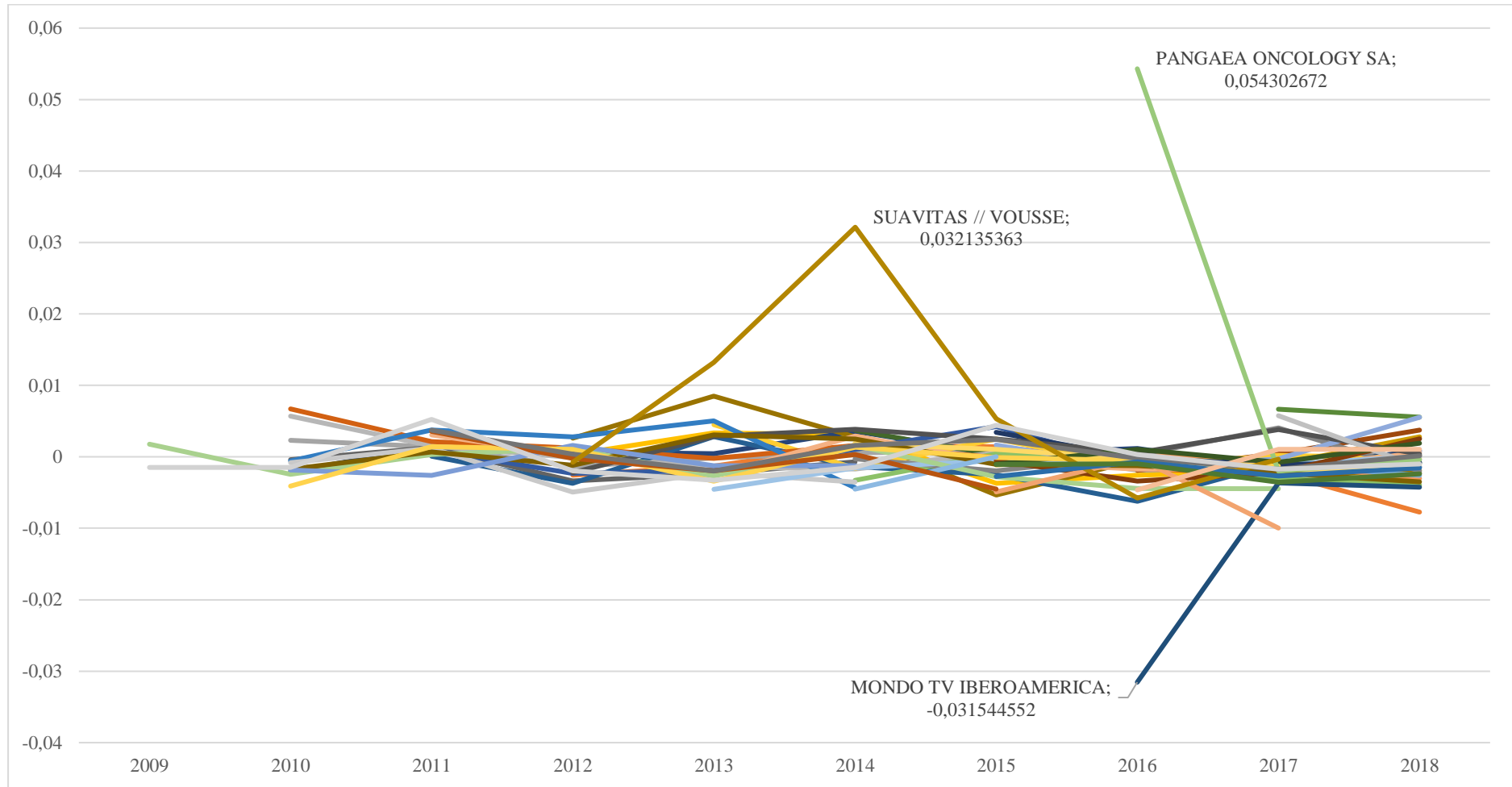
Fuente: Elaboración propia a partir de los datos extraídos de Datastream

Gráfico 3: Rentabilidad media anual corregida por el IGBM, 2008 – 2018



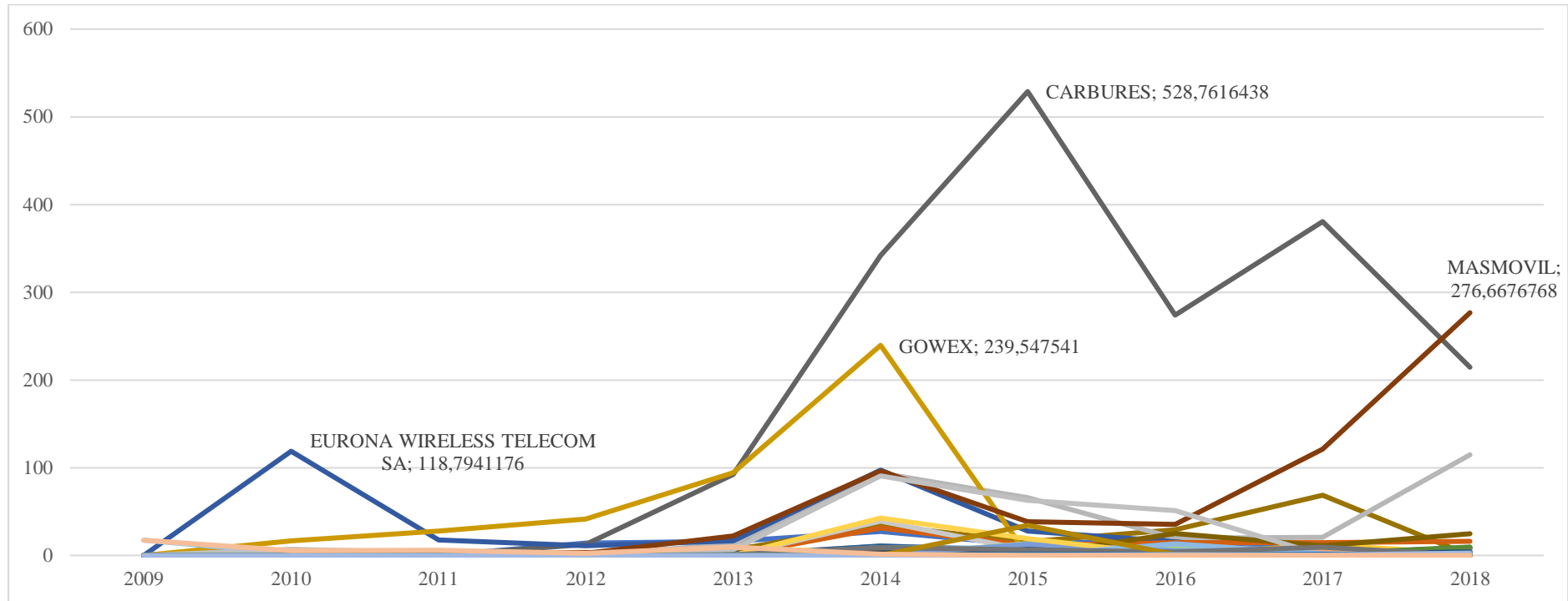
Fuente: Elaboración propia a partir de los datos extraídos de Datastream

Gráfico 4: Rentabilidad media anual corregida por el IBEX Small Cap, 2008 – 2018



Fuente: Elaboración propia a partir de los datos extraídos de Datastream

Gráfico 5: Rotación Volumen medio anual, 2008 – 2018 (miles de acciones)



Fuente: Elaboración propia a partir de los datos extraídos de Datastream

3.3. NOTICIAS

Se elabora una base de datos que recopila todas las noticias extraídas de *Google News*. Esta herramienta consiste en un buscador de noticias automatizado de información de los principales medios de comunicación que publican en sus webs. En suma, no solo se realizan búsquedas en los periódicos con más repercusión, sino que permite obtener información también de blogs de opinión. Otra de las ventajas de este buscador es que permite filtrar las noticias a un periodo concreto, por tanto, solo se han extraído las noticias que han sido publicadas en el periodo analizado de 2008 a 2018.

Mediante la metodología de *Web Scraping*, se han conseguido recopilar un conjunto de noticias para el análisis. Este método consiste en extraer información de páginas webs de manera automática. Por tanto, ha sido una herramienta realmente útil para crear la base de datos. El procedimiento para crear dicha base de datos ha sido el que se describe a continuación:

- En primer lugar, se realiza una búsqueda para cada una de las cincuenta empresas analizadas en el buscador Google News, para el periodo que se investiga.
- Posteriormente se implementa el link de cada una de las búsquedas en el código elaborado anteriormente⁸ mediante el lenguaje de programación Python⁹. Este código permite capturar los elementos necesarios para elaborar la base de datos: fecha de publicación de la noticia, fuente en la que se ha publicado, titular y cuerpo de la misma.

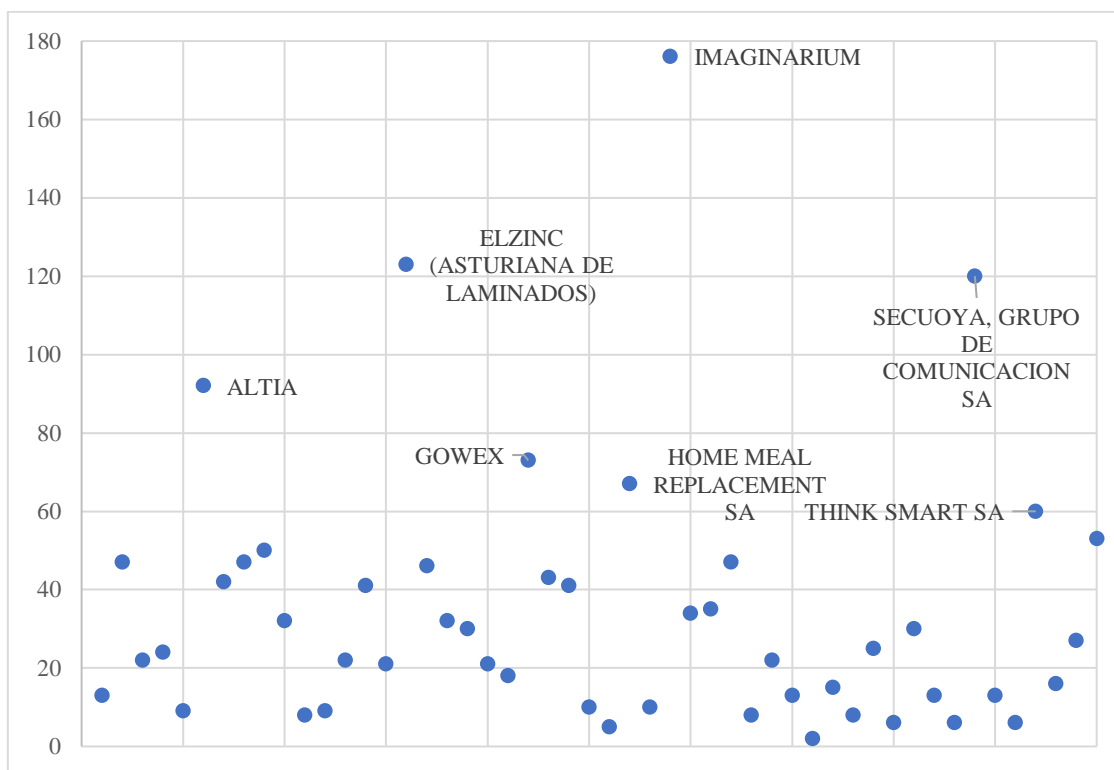
Para cada una de las empresas analizadas se han extraído noticias donde cada una de esas empresas haya sido mencionada en el cuerpo del artículo. Exceptuando el caso del mercado, que se han realizado búsquedas para que solo se extrajeran noticias en las que se mencionara al MAB en el titular de las noticias y blogs, puesto que, si se hubiera realizado una búsqueda de la misma manera que para las empresas, el volumen de información hubiera sido demasiado elevado en comparación con las noticias extraídas de cada una de las empresas.

Tal y como muestra el Gráfico 6, un aspecto importante ha sido que no existe el mismo número de noticias para todas las empresas, hay empresas con gran repercusión y permiten extraer más volumen de noticias que otras empresas menos conocidas y con menos noticias disponibles. De manera que no se tienen el mismo número de noticias para cada una de las empresas analizadas.

⁸ Puede comprobarse este código en el ANEXO 1

⁹ Se ha utilizado el entorno del software de Spyder para visualizar los códigos. Este entorno de desarrollo multiplataforma se permite utilizar en código abierto para el lenguaje de Python.

Gráfico 6: Número de noticias analizadas de cada empresa, 2008 – 2018¹⁰



Fuente: elaboración propia a través de la base de datos obtenida

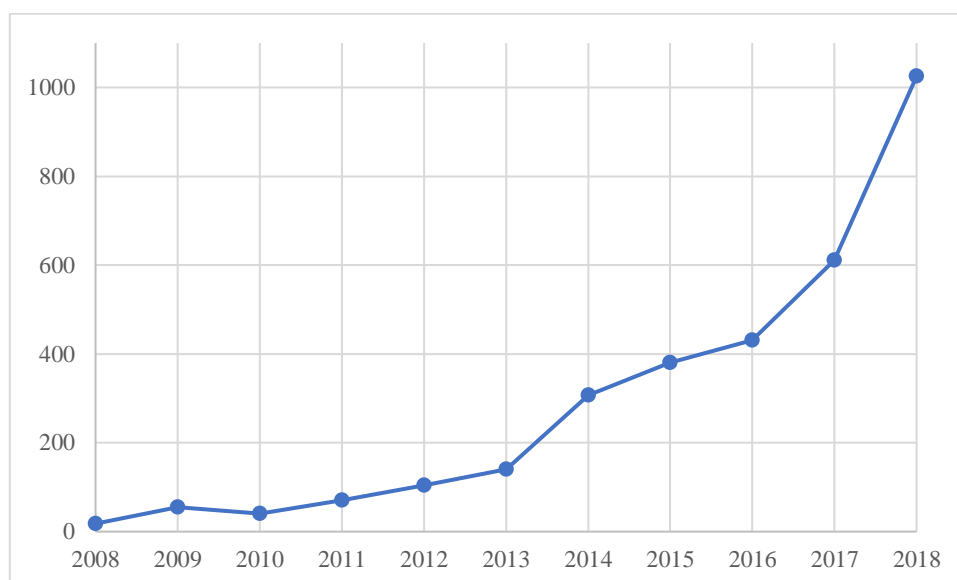
Empresas como Imaginarium (176 noticias para analizar) o Elzinc (123 noticias) tienen un mayor número de noticias disponibles en la web.

Además, se ha observado que existen mayor número de noticias en los periodos más actuales, es decir, se encuentran muchas más noticias al final del periodo analizado que al principio del mismo, como se puede comprobar en el Gráfico 7. Esto último tiene lógica, puesto que conforme ha ido pasando el tiempo, son más las empresas de comunicación que se han sumado a realizar las publicaciones de manera digital, y existen muchos más autores que de manera personal se dedican a publicar artículos en blogs de opinión.

Respecto a las noticias disponibles, una variable importante a analizar es la fuente de la que proceden las mismas. Por ello se creará una variable que se introducirá en los modelos a analizar para poder contemplar este aspecto.

¹⁰ Más información disponible en ANEXO 2

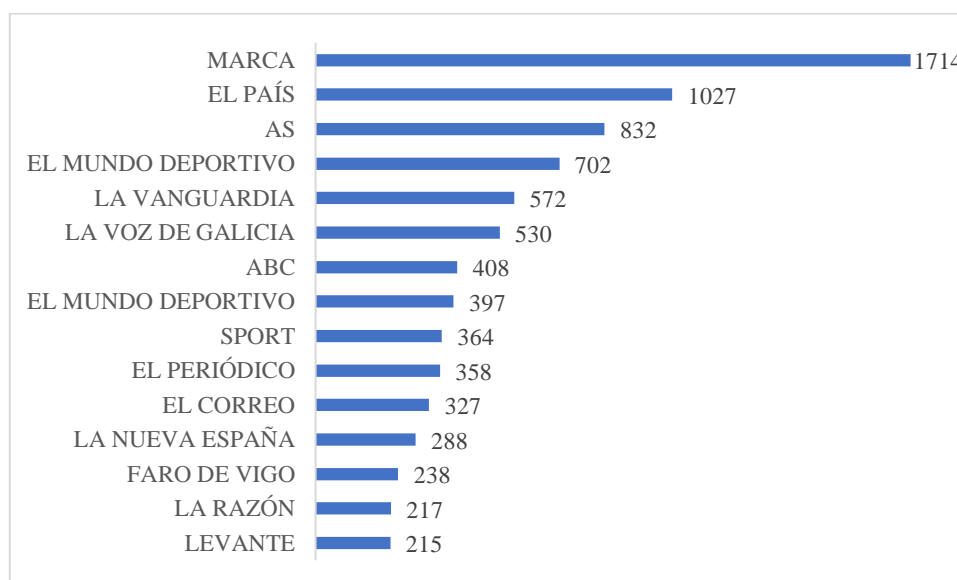
Gráfico 7: Evolución número de noticias analizadas, 2008-2018



Fuente: elaboración propia

De acuerdo con el gráfico 8, se tienen en cuenta los 15 periódicos con más número de lectores diarios en 2018 para crear esta variable dummy:

Gráfico 8: Número de lectores diarios de los principales periódicos españoles en 2018



Fuente: Statista

Además, también se consideran 8 fuentes emblemáticas de economía y finanzas como son: Expansión, El Economista, Cinco Días, Bolsamanía, Fortune, Aquí Europa, El Confidencial y Forbes¹¹.

¹¹ <https://www.coinc.es/blog/noticia/mejores-revistas-periodicos-economia>

4. METODOLOGÍA

4.1. ANÁLISIS DE SENTIMIENTO

De la misma manera que los titulares de las noticias ofrecen un contenido resumido de la noticia, el análisis de sentimiento muestra una variable óptima de la opinión que exprese la noticia. Actualmente, las opiniones son de gran ayuda, cada vez que se quiera tomar una decisión, se tienen muy en cuenta las opiniones del resto.

La posibilidad de obtener una variable de sentimiento de forma óptima es relativamente reciente. Hace unas décadas, si alguien a nivel particular quería tomar una decisión, preguntaba a amigos o familiares. Si era una empresa la que quería información para tomar una decisión, realizaba encuestas de opinión a una muestra de la población. Hoy en día, gracias al volumen de información publicada en la web y el desarrollo tecnológico, resulta más sencillo obtener opiniones sobre cualquier producto o activo. Este desarrollo tecnológico ha hecho que muchos autores comiencen a implicarse en investigar sobre métodos de extracción de una variable de sentimiento.

A grandes rasgos, existen dos grandes metodologías para cuantificar el análisis de sentimiento: aprendizaje automático (*Machine Learning*) y análisis basado en léxicos.

El primero de ellos es una rama de la inteligencia artificial¹², consiste en elaborar un modelo que permita identificar patrones y tomar decisiones con un mínimo de intervención humana. Generalmente se dispone de un gran conjunto de datos ya identificados, que contienen un mapeo entre expresiones textuales y clasificaciones de textos asignadas por el humano (Shapiro, Sudhof y Wilson, 2020)). Una vez se ha entrenado el modelo, se puede proceder a introducir una nueva base de datos que permita predecir sobre las variables deseadas.

Esta metodología se ha desarrollado principalmente en campos como la publicación de opiniones en redes sociales o páginas webs de revisión de películas (Pang, Lee y Vaithyanathan, 2002). Esto es porque son sitios webs donde el autor ya ha mostrado su opinión, y se identifica fácilmente la positividad o negatividad de la publicación. No obstante, el principal inconveniente de esta metodología es el volumen de información necesaria para entrenar el modelo inicial, requieren un aprendizaje muy profundo para aprender los léxicos de las palabras con la mayor precisión posible. Esto conlleva que sea una metodología muy costosa en tiempo y difícil de construir. Al contrario, su principal ventaja es que se construye un modelo basado en las características necesarias para predecir, logrando una precisión mucho mayor que con otras metodologías.

La segunda metodología es la que se va a utilizar en este proyecto. Se trata de un análisis de sentimiento que se basa en léxicos ya entrenados. Este enfoque se basa en diccionarios o listas de palabras predeterminadas, que tienen una puntuación de sentimiento preasignada. El más usado en esta metodología es el diccionario

¹² https://www.sas.com/es_es/insights/analytics/machine-learning.html (29/5/2020)

psicosociológico de Harvard (*Harvard General Inquirer*¹³). En el ámbito de economía y finanzas, son múltiples los autores que han utilizado este enfoque.

Loughran y McDonald (2011) se basan en el diccionario de Harvard creando una nueva versión, puesto que tres cuartas partes de las palabras negativas, se consideran realmente neutrales en un contexto financiero. Este nuevo diccionario lo relacionan con variables financieras como la volatilidad, el volumen de operaciones o las ganancias inesperadas. Unos años después, Heston y Sinha (2015) emplearon el diccionario propuesto por los autores mencionados anteriormente para investigar si las noticias podrían predecir el rendimiento de las acciones. Posteriormente, Shapiro y Wilson (2020), también utilizaron el léxico creado por Loughran y McDonald (2011), pero estos lo hicieron para estimar las preferencias del Banco Central mediante el sentimiento expresado por los responsables políticos en sus reuniones internas.

Otro de los diccionarios más usados es el que desarrollaron Hutto y Gilbert (2014) al que bautizaron como VADER (*Valence Aware Dictionary for sentiment Reasoning*). Se define como un clasificador de sentimientos que se utiliza mediante el lenguaje de Python. Consiste en un léxico compuesto de miles de palabras que están etiquetadas de 4 a -4 que corresponden del sentimiento más positivo al más negativo.

Sin embargo, Google Cloud también ha lanzado un apartado de productos de inteligencia artificial y aprendizaje automático. Dentro de estos encontramos una aplicación para el tema que abarca en esta investigación: *Natural Language*¹⁴, una herramienta de procesamiento de textos. Tiene dos funciones básicas y de diferente alcance:

- *AutoML Natural Language*. Ofrece la posibilidad de crear modelos de aprendizaje automático y entrenarlos de acuerdo a las preferencias que cada uno necesite. Permite extraer, clasificar y detectar opiniones.
- *API Natural Language*. Engloba modelos entrenados previamente para ofrecer análisis de opinión, de entidades, clasificación sintáctica o análisis sintáctico.

De esta manera, para la creación de la variable sentimiento en la investigación que se lleva a cabo en este proyecto, se ha utilizado la herramienta *API Natural Language*, elegida por su sencillez a la hora de trabajar con ella.

Una de las posibilidades de análisis que ofrece esta herramienta es el análisis de opiniones. Utilizando el método de análisis de sentimiento, se determina la actitud general de un escritor entre positiva, negativa o neutra. Esto se identifica con la variable polaridad (mencionada anteriormente). Esta variable está normalizada, de manera que toma valores entre -1 y 1. Un texto que obtenga valor -1 tiene sentimiento negativo, mientras que si obtiene valor 1 tendrá un sentimiento positivo. Aquellos que obtengan valor cercano a 0, tendrán una opinión neutral, indicando que tiene opiniones mixtas.

¹³ Puede consultarse en <http://www.wjh.harvard.edu/~inquirer/homecat.htm>

¹⁴ <https://cloud.google.com/natural-language?hl=es>

En la Tabla 2 se realiza un análisis descriptivo inicial de esta variable, donde se recogen todas las noticias extraídas de todo el periodo analizado:

Tabla 2: Descriptivos principales variable sentimiento, 2008 – 2018

	MEDIA	VARIANZA	DESVIACION	MINIMO	MAXIMO
MAB	0,0031279	0,0004413	0,0210061	-0,2083333	0,4000000
INKEMIA	0,0004219	0,0001169	0,0108106	-0,2000000	0,3000000
AB-BIOTICS SA	0,0040402	0,0012909	0,0359294	-0,3000000	0,4000000
ADL BIONATUR	0,0014730	0,0004485	0,0211767	-0,3000000	0,7000000
AGILE CONTENT	0,0005885	0,0002259	0,0150304	-0,6000000	0,3520370
ALQUIBER QUALITY SA	0,0001328	0,0000260	0,0051032	0	0,2470280
ALTIA	0,0001634	0,0000237	0,0048697	-0,0006944	0,1898816
ATRYA HEALTH SA	0,0015702	0,0004095	0,0202360	0	0,6000000
BODACLICK	0,0024261	0,0004085	0,0202122	-0,0062500	0,3087500
CARBURES	0,0036202	0,0005426	0,0232928	-0,1000000	0,3125000
CATENON SA	0,0028588	0,0006940	0,0263433	-0,3000000	0,4000000
CERBIUM HOLDING	0,0001316	0,0000309	0,0055600	-0,2000000	0,1035656
CLERHP ESTRUCTURAS	0,0008129	0,0002285	0,0151148	0	0,4000000
CLEVER GLOBAL SA.	0,0009089	0,0002134	0,0146088	0	0,6000000
COMMCENTER	0,0007900	0,0004954	0,0222573	-0,4000000	0,9000000
EBOSS ENERGY	0,0005324	0,0000675	0,0082138	-0,0458333	0,2096771
ELZINC (ASTURIANA DE LAMINADOS)	0,0045824	0,0012711	0,0356523	-0,1000000	0,9000000
EUROCONSULT ENGINEERING CONSULTING GROUP SA.	0,0013425	0,0001965	0,0140183	-0,2000000	0,2744444
EUROESPES, SA	0,0023509	0,0006519	0,0255319	-0,1000000	0,7000000
EURONA WIRELESS TELECOM SA	0,0003515	0,0001688	0,0129930	-0,4000000	0,3000000
FACEPHI BIOMETRIA, S.A.	0,0004476	0,0001283	0,0113261	-0,3000000	0,3000000
GIGAS HOSTING SA	0,0008081	0,0002174	0,0147436	0	0,7000000
GOWEX	0,0029935	0,0004205	0,0205069	-0,1077381	0,4170455
GREENALIA	0,0017865	0,0004265	0,0206521	-0,2000000	0,4000000
GREENERGY RENOVABLES SA	0,0013807	0,0003608	0,0189948	-0,2000000	0,4000000
GRINÓ ECOLOGIC	0,0001530	0,0000520	0,0072142	-0,3000000	0,2000000
GRUPO NOSTRUM RNL, S.A	0,0001990	0,0000218	0,0046697	0	0,1574405
HOME MEAL REPLACEMENT SA	0,0029693	0,0007991	0,0282682	-0,3000000	0,5000000
IFFE FUTURA	0,0009071	0,0001249	0,0111775	-0,0233333	0,2899908

Fuente: Elaboración propia a través de los resultados obtenidos de la variable sentimiento

Tabla 2: Descriptivos principales variable sentimiento, 2008 – 2018 (continuación)

	MEDIA	VARIANZA	DESVIACION	MINIMO	MAXIMO
IMAGINARIUM	0,0163740	0,0054366	0,0737336	-0,6000000	0,9000000
INCLAM SOCIEDAD ANONIMA	0,0047627	0,0017496	0,0418286	-0,4000000	0,7000000
LLEIDANETWORKS SERVEIS TELEMATICS SA	0,0014435	0,0002005	0,0141610	-0,1416667	0,2219449
MASMOVIL	0,0030397	0,0004653	0,0215706	0	0,3270513
MEDCOM TECH, SA	0,0005643	0,0001454	0,0120589	0	0,4000000
MONDO TV IBEROAMERICA	0,0065997	0,0019364	0,0440047	-0,1000000	0,7000000
NBI BEARINGS EUROPE SA	0,0002831	0,0000391	0,0062566	0	0,1723030
NEGOCIO & ESTILO DE VIDA	0,0000333	0,0000023	0,0015028	0	0,0755084
NEOL BIO	0,0006761	0,0002529	0,0159040	-0,4000000	0,4000000
NETEX KNOWLEDGE FACTORY SA	0,0000218	0,0000019	0,0013806	0	0,0875216
NEURON BIO	0,0017966	0,0004053	0,0201313	-0,2000000	0,4000000
NPG TECHNOLOGY	0,0001819	0,0000208	0,0045597	-0,0312500	0,1377650
PANGAEA ONCOLOGY SA	0,0008687	0,0002532	0,0159110	-0,4000000	0,4000000
ROBOT SA	0,0004235	0,0001069	0,0103415	0	0,3000000
SEAFOOD GLOBAL PROCESSOR // LUMAR	0,0001123	0,0000153	0,0039149	-0,0583333	0,1300902
SECUOYA, GRUPO DE COMUNICACION SA	0,0096558	0,0033582	0,0579502	-0,8000000	0,7000000
SUAVITAS // VOUSSE	0,0003677	0,0000411	0,0064080	-0,0481481	0,1669886
TECNOQUARK	0,0001704	0,0000374	0,0061135	-0,0964815	0,2750000
THINK SMART SA	0,0030170	0,0010398	0,0322458	-0,1000000	0,8000000
TIER 1 TECHNOLOGY SOCIEDAD ANONIMA	0,0002145	0,0000308	0,0055516	-0,0537997	0,2137131
VOZTELECOM OIGAA360 SA.	0,0006253	0,0001160	0,0107688	-0,0878788	0,2769762
ZINKIA ENTERTAINMENT, SA	0,0054284	0,0020390	0,0451548	-0,2000000	0,7000000

Fuente: Elaboración propia a través de los resultados obtenidos de la variable sentimiento

En primer lugar, la media de toda la muestra indica que las noticias son mayoritariamente positivas en una gran parte de los casos. Exceptuando valores medios de empresas como NEGOCIO & ESTILO DE VIDA y NETEX KNOWLEDGE FACTORY SA, que son más cercanos a cero, indicando que la distribución de noticias es más amplia entre positivas y negativas.

Gracias a la varianza y la desviación se logra analizar la dispersión de la variable. De esta manera, CARBURES e IMAGINARIUM son las empresas que mayor dispersión obtienen respecto a su media.

Por su parte, si se analiza el mínimo de cada empresa, todas aquellas que obtienen un valor igual a 0 indican que no poseen noticias negativas. En suma, IMAGINARIUM es la empresa que obtiene un valor máximo de 0,9.

4.2. ANÁLISIS EMPÍRICO

En este apartado se va a explicar los modelos que se han utilizado para estimar los efectos individuales y conjuntos que tienen las noticias sobre las variables financieras que se han extraído, además se realizan para cada modelo dos tipos de ajustes: el primero se basa en comprobar si la noticia en el momento t tiene efecto sobre las variables financieras del mismo momento, mientras que el segundo ajuste se comprobará el posible efecto sobre el momento $t + 1$. La razón de analizar los dos momentos temporales proviene de que no se tiene constancia de la hora en la que se comparte la publicación, de manera que no es lo mismo que sea publicada a primera hora de la mañana, que afectaría a la variable financiera de ese día, a que sea publicada a última hora de la tarde, que ya afectaría financieramente al día siguiente.

Finalmente se comprobará también en una regresión a parte el efecto que tienen las noticias del MAB en su propia rentabilidad. Para estimar estos modelos se va a utilizar la herramienta de RStudio.

Cabe señalar que para la realización de estos modelos se ha reducido la muestra de datos, puesto que solo se tiene en cuenta la información de cada empresa desde que empieza a cotizar en el MAB. Esto es debido a que las variables financieras son inexistentes antes de que se produzca la introducción en el MAB de cada una de las empresas. Es por ello que el periodo muestral no es el mismo para todas las empresas analizadas.

Las variables que forman parte de los modelos son las siguientes:

- *RentCorrIGBM*: variable que expresa la rentabilidad corregida por el benchmark del IGBM. Calculada a través de la TIR.
- *RentCorrIBEXSC*: variable que expresa la rentabilidad corregida por el benchmark del IBEX Small Cap. Calculada a través de la TIR.
- *RentIGBM*: variable que expresa la rentabilidad del IGBM, también calculada a través de la TIR.
- *RentIBEXSC*: variable que expresa la rentabilidad del IBEX Small Cap, también calculada a través de la TIR.
- *VO*: variable que expresa la rotación de volumen de cada empresa.
- *Sent*: variable que recoge el sentimiento de cada una de las noticias, tomando valores desde -1 a 1. Además, se parte de la hipótesis de que, si no existe noticia, es como si fuera una noticia neutral, es decir, con valor 0.

- *SentMAB*: variable que recoge el sentimiento de las noticias del mercado (MAB).
- *Periodico*: variable de tipo *dummy*, toma valor 1 cuando la noticia proviene de una fuente de carácter emblemático, y 0 para el resto de casos¹⁵.
- *Entrada*: variable de tipo *dummy*, toma valor 1 en la fecha de introducción al MAB de cada empresa, y 0 para el resto de casos.
- *Salida*: variable de tipo *dummy*, toma valor 1 en la fecha de exclusión del MAB para cada empresa (en el caso de que haya sido excluida), y 0 para el resto de casos.

Para todas estas variables, el subíndice i recoge las 50 empresas analizadas, mientras que el subíndice t expresa el momento temporal. Además, la razón principal de corregir la rentabilidad con los benchmark de cada una de las empresas es tratar de no perder información. Es decir, se está buscando el exceso de rentabilidad con el mercado, y no la rentabilidad diaria en términos absolutos. De esta manera se pueden comparar circunstancias ocurridas en el mercado y el valor va a ser comparable para todas las empresas.

4.2.1. EFECTOS SOBRE LA RENTABILIDAD

AJUSTE 1 - EFECTO SOBRE LA RENTABILIDAD EN t

En este ajuste se va a comprobar el efecto que tienen las variables explicativas en el momento t sobre la rentabilidad en el mismo momento, en este caso corregida por el IGBM, tal y como muestra la ecuación (2). Para el caso de una noticia que aparezca en sábado o domingo, se tendrá en cuenta la rentabilidad del lunes. De la misma manera ocurrirá con los festivos, que se tendrá en cuenta la rentabilidad del día siguiente laboral.

$$\begin{aligned}
 RentCorrIGBM_{it} = \beta_0 + \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it} \\
 t = 1, 2, \dots, T \\
 i = 1, 2, \dots, 50
 \end{aligned}
 \tag{2}$$

Para este caso se estima un modelo **MCO** (Mínimos Cuadrados Ordinarios), suponiendo que los efectos individuales son comunes entre todas las empresas. Este modelo agrega la dimensión i y la t , sin tener en cuenta las peculiaridades de cada una de las empresas. Los resultados se muestran en la Tabla 3.

El modelo 1.1. se corresponde con el efecto sobre la rentabilidad corregida a través del IGBM. Como se puede apreciar, el coeficiente estimado de la variable SALIDA es significativo al nivel de confianza 5% y el que se corresponde con la variable PERIODICO es significativo al 10%. Por su parte, el estimador de la constante es significativo al 1%.

¹⁵ Recordar Gráfico 8 con principales periódicos.

Tabla 3: Modelos efecto sobre la rentabilidad corregida por IGBM en t

	Modelo 1.1.		Modelo 1.2.		Modelo 1.3.	
Beta 0	0,0004699	(0,002465)			0,0003868	(0,23471)
Sentimiento	-	(0,693362)	-0,0002471	(0,95692)	-0,0005989	(0,89560)
	0,0017893					
Periodico	0,0030058	(0,088589)	0,0027052	(0,12609)	0,0028520	(0,10652)
Entrada	-	(0,840996)	-0,0010779	(0,86190)	-0,0011739	(0,84973)
	0,0012440					
Salida	-	(0,019707)	-0,0264615	(0,01922)	-0,0266106	(0,01854)
	0,0263709					
R2	0,0001023		-0,0005557		0,0000974	

Fuente: Elaboración propia a través de la herramienta RStudio

Posteriormente, se realiza un modelo para comprobar los **efectos fijos**. Este método resulta interesante cuando los individuos que representa tienen características y comportamientos diferentes. Como se están analizando empresas diferentes, es interesante realizarlo. Su especificación difiere de la anterior en el término de error:

$$\begin{aligned}
 RentCorrIBEXSC_{it} &= \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it} \\
 \text{Donde } u_{it} &= \alpha_i + \varepsilon_{it} \\
 t &= 1, 2, \dots, T \\
 i &= 1, 2, \dots, 50
 \end{aligned}
 \tag{3}$$

Antes de analizar los resultados del modelo, hay que comprobar si los efectos fijos son significativos, es decir, si existe una constante común para todas las empresas o es diferente. Esto se lleva a cabo a través del contraste F, obteniendo el siguiente:

Contraste 1: Test F para efectos individuales

F = 4,068	p-value < 2,2e-16
-----------	-------------------

Fuente: Elaboración propia a través de la herramienta RStudio

De acuerdo con este, se rechaza la hipótesis nula de que los efectos son comunes para las empresas, es decir, la constante es diferente para cada empresa. En el Modelo 1.2. de la Tabla 3 se muestran los resultados de un modelo de efectos fijos. Al igual que anteriormente, el estimador de la variable SALIDA es significativo al 5%.

Otra estimación que se puede realizar es la de **efectos aleatorios**. Este estimador será consistente cuando los efectos individuales no estén correlacionados con las variables explicativas. Para comprobarlo se realiza el test de Hausman.

Contraste 2: Test Hausman

chisq = 6,8077

p-value = 0,1464

Fuente: Elaboración propia a través de la herramienta RStudio

En este caso no se rechaza la hipótesis nula, el modelo es consistente con efectos aleatorios. Para este, se observan los resultados de la Tabla 3 correspondientes con el Modelo 1.3. De nuevo, tan solo la variable SALIDA mantiene un coeficiente significativo al 5%.

Siguiendo la misma línea del ajuste, se comprueba el efecto que tienen las variables explicativas sobre la rentabilidad corregida por el IBEX Small Cap, tal y como se muestra en la ecuación (4):

$$\begin{aligned}
 RentCorrIBEXSC_{it} = \beta_0 + \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it} \\
 t = 1, 2, \dots, T \\
 i = 1, 2, \dots, 50
 \end{aligned} \tag{4}$$

Para el caso del modelo **MCO** se obtienen los resultados recogidos en la tabla 4, correspondientes al Modelo 2.1. Tal y como se ha ido observando en el modelo anterior, en este caso ocurre lo mismo, el estimador de la variable SALIDA es significativo al 5%, y el de la variable PERIODICO al 10%.

En el caso de analizar los **efectos fijos**, la ecuación (5) muestra la regresión del modelo:

$$\begin{aligned}
 RentCorrIBEXSC_{it} = \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it} \\
 \text{Donde } u_{it} = \alpha_i + \varepsilon_{it} \\
 t = 1, 2, \dots, T \\
 i = 1, 2, \dots, 50
 \end{aligned} \tag{5}$$

Tabla 4: Modelos efecto sobre la rentabilidad corregida por IBEX Small Cap en t

	Modelo 2.1.		Modelo 2.2.		Modelo 2.3.	
Beta 0	-0,0001401	(0,36189)			-0,0002543	(0,43602)
Sentimiento	-0,0005286	(0,90632)	0,0011076	(0,80672)	0,0007052	(0,87595)
Periodico	0,0031020	(0,07581)	0,0027930	(0,11055)	0,0029382	(0,09295)
Entrada	0,0022385	(0,71530)	0,0024053	(0,69490)	0,0023399	(0,70277)
Salida	-0,0264903	(0,01795)	-0,0265445	(0,01765)	-0,0266900	(0,01702)
R2	0,0001116		0,0001094		0,0001160	

Fuente: Elaboración propia a través de la herramienta RStudio

Y de nuevo se comprueba a través del contraste F si los efectos fijos son significativos:

Contraste 3: Test F para efectos individuales

F = 4,2056	p-value < 2,2e-16
------------	-------------------

Fuente: Elaboración propia a través de la herramienta RStudio

La constante es diferente para cada individuo de acuerdo con este contraste, puesto que se rechaza la hipótesis nula de que existe una constante común para todas las empresas. La estimación del modelo de efectos fijos resultante aparece reflejada en la Tabla 4, y los resultados se corresponden con el Modelo 2.2. En este caso, al igual que anteriormente el estimador de la variable SALIDA es significativo al 5%.

Para realizar la estimación mediante **efectos aleatorios**, se comprueba el Test de Hausman:

Contraste 4: Test Hausman

Chisq = 7,1845	p-value = 0,1265
----------------	------------------

Fuente: Elaboración propia a través de la herramienta RStudio

De la misma manera que anteriormente, no se rechaza la hipótesis nula y, por tanto, el modelo es consistente con efectos aleatorios. Los resultados del mismo se pueden observar en la Tabla 4, y en concreto, en el Modelo 3.3. En este caso, el estimador de la variable SALIDA es significativo al 5% y el de la variable PERIODICO lo es al 10%.

AJUSTE 2 - EFECTO SOBRE LA RENTABILIDAD EN $t + 1$

Por su parte, en este ajuste se va a comprobar el efecto que tienen las variables explicativas en el momento t sobre la rentabilidad en el momento $t + 1$, es decir, la noticia que se produce en el momento t , se comprueba el efecto que tiene sobre la rentabilidad del día siguiente. Para la noticia que aparezca en viernes, sábado o domingo, se tendrá en cuenta la rentabilidad del lunes, y del mismo modo, para los festivos, será la rentabilidad del día siguiente laboral.

En primer lugar, se plantea un modelo **MCO** sobre la rentabilidad corregida por el IGBM, tal y como muestra la ecuación (6).

$$\begin{aligned} \text{RentCorrIGBM}_{it+1} &= \beta_0 + \beta_1 \text{Sent}_{it} + \beta_2 \text{Periodico}_{it} + \beta_3 \text{Entrada}_{it} + \beta_4 \text{Salida}_{it} + u_{it} \\ t &= 1, 2, \dots, T \\ i &= 1, 2, \dots, 50 \end{aligned} \tag{6}$$

Los resultados que se obtienen se muestran en la Tabla 5, en referencia al Modelo 3.1. Para este caso, resalta de nuevo la significatividad del estimador de la variable SALIDA, a un nivel de confianza del 0,1%, al igual que la variable constatae.

Tabla 5: Modelos efecto sobre la rentabilidad corregida por IGBM en $t + 1$

	Modelo 3.1.		Modelo 3.2.		Modelo 3.3.	
Beta 0	0,0004117	(0,006642)			0,0003319	(0,253451)
Sentimiento	-	(0,458954)	-	(0,622430)	-	(0,583547)
	0,0031447		0,0021083		0,0023412	
Periodico	0,0020658	(0,212338)	0,0018583	(0,263193)	0,0019951	(0,229173)
Entrada	0,0003068	(0,959550)	0,0005089	(0,932922)	0,0003903	(0,948521)
Salida	-	(0,008966)	-	(0,008986)	-	(0,008534)
	0,0288336		0,0288142		0,0290042	
R2	0,0001076		0,0001011		0,0001009	

Fuente: Elaboración propia a través de la herramienta RStudio

Para el análisis de los **efectos fijos**, primero se realiza el contraste F:

Contraste 5: Test F para efectos individuales

F = 3,4112	p-value = 8,564e-15
------------	---------------------

Fuente: Elaboración propia a través de la herramienta RStudio

Se rechaza la hipótesis nula de que los efectos son comunes para las empresas, es decir, la constante es diferente para cada empresa.

La estimación del modelo de efectos fijos se corresponde con la ecuación (7).

$$\begin{aligned}
 RentCorrIGBM_{it+1} &= \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it} \\
 \text{Donde } u_{it} &= \alpha_i + \varepsilon_{it} \\
 t &= 1, 2, \dots, T \\
 i &= 1, 2, \dots, 50
 \end{aligned} \tag{6}$$

Los resultados de este se modelo se corresponden con el Modelo 3.2 de la Tabla 5. En esta estimación, tan solo el estimador de la variable SALIDA es significativo, al nivel del 0,1%.

En cuanto a la estimación de **efectos aleatorios**, se comprueba el Test de Hausman:

Contraste 6: Test Hausman

chisq = 8,0763	p-value = 0,08882
----------------	-------------------

Fuente: Elaboración propia a través de la herramienta RStudio

En este caso se acepta la hipótesis nula de que el modelo con efectos aleatorios es consistente.

Por tanto, se plantea una estimación con **efectos aleatorios**, reflejada en el Modelo 3.3. que aparece en la Tabla 5. En este caso, el estimador de la variable SALIDA vuelve a ser significativo.

A continuación, se plantea el mismo ajuste, pero esta vez con la rentabilidad corregida por el IBEX Small Cap. En primer lugar, el modelo **MCO** que se estima se muestra en la ecuación (7). Los resultados del mismo aparecen en la Tabla 6. Si se observa el Modelo 4.1. que se corresponde con el MCO estimado, tan solo el estimador de la variable SALIDA es significativo.

$$\begin{aligned}
 RentCorrIBEXSC_{it+1} &= \beta_0 + \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it} \\
 t &= 1, 2, \dots, T \\
 i &= 1, 2, \dots, 50
 \end{aligned}
 \tag{7}$$

Tabla 6: Modelos efecto sobre la rentabilidad corregida por IBEX Small Cap en $t + 1$

	Modelo 4.1.		Modelo 4.2.		Modelo 4.3.	
Beta 0	-0,0001854	(0,215640)			-	(0,314096)
					0,0003000	
Sentimiento	-0,0040906	(0,329063)	-0,0030194	(0,474899)	-	(0,437352)
					0,0032741	
Periodico	0,0025853	(0,113804)	0,0023672	(0,148697)	0,0024962	(0,127412)
Entrada	0,0009221	(0,877266)	0,0011254	(0,850413)	0,0010432	(0,861204)
Salida	-0,0303730	(0,005286)	-0,0303886	(0,005241)	-	(0,005006)
					0,0305456	
R2	0,0001348		0,0001260		0,0001364	

Fuente: Elaboración propia a través de la herramienta RStudio

Antes de estimar un modelo con **efectos fijos**, se debe comprobar el contraste F. La ecuación (8) muestra la estimación del modelo.

$$\begin{aligned}
 RentCorrIBEXSC_{it+1} &= \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it} \\
 \text{Donde } u_{it} &= \alpha_i + \varepsilon_{it} \\
 t &= 1, 2, \dots, T \\
 i &= 1, 2, \dots, 50
 \end{aligned}
 \tag{8}$$

Contraste 7: Test F para efectos individuales

F = 3,6967	p-value < 2,2e-16
------------	-------------------

Fuente: Elaboración propia a través de la herramienta RStudio

De acuerdo con este contraste, la constante es diferente para cada individuo. Y el modelo con efecto fijos resultante aparece en la Tabla 6, Modelo 4.2. Al igual que en los casos anteriores, tan solo el estimador de la variable SALIDA es significativo al nivel de confianza del 0,1%.

Para realizar una estimación de **efectos aleatorios**, antes hay que comprobar el test de Hausman:

Contraste 8: Test Hausman

chisq = 7,7818	p-value = 0,09991
----------------	-------------------

Fuente: Elaboración propia a través de la herramienta RStudio

Este test indica que el modelo con efectos aleatorios es consistente, puesto que se acepta la hipótesis nula. Los resultados de la estimación de efectos aleatorios aparecen en la Tabla 6, y en concreto, se corresponden con el modelo 4.3. De la misma manera, el estimador de la variable SALIDA es significativo al nivel de confianza del 0,1%.

4.2.2. EFECTO SOBRE ROTACIÓN: VOLUMEN

En este caso, se va a comprobar el efecto que tienen las variables explicativas sobre la ratio de rotación que se ha extraído. De la misma manera que con la rentabilidad, se realizan dos tipos de ajuste.

AJUSTE 1 - EFECTO SOBRE LA ROTACION DE VOLUMEN EN t

En el primer ajuste, se analizar el efecto de las variables explicativas en el momento t sobre la rotación de volumen en el mismo momento. Exceptuando los fines de semana, puesto que una noticia en sábado o domingo, se comprueba el efecto sobre el volumen del lunes. Lo mismo ocurre con los festivos. La ecuación (9) muestra la estimación de un modelo **MCO**, y posteriormente se muestran los resultados obtenidos.

$$\begin{aligned}
 VO_{it} &= \beta_0 + \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it} \\
 & \quad t = 1, 2, \dots, T \\
 & \quad i = 1, 2, \dots, 50
 \end{aligned} \tag{9}$$

Para este caso, los estimadores de las variables PERIODICO, ENTRADA y SALIDA son significativos. Y lo mismo ocurre con la constante. Estos resultados aparecen en la Tabla 7, correspondientes al modelo 5.1.

Por su parte, para la elaboración estimación de los **efectos fijos**, antes se realiza un contraste F. La estimación del modelo se corresponde con la ecuación (10).

$$\begin{aligned}
 VO_{it} &= \beta_0 + \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it} \\
 & \quad \text{Donde } u_{it} = \alpha_i + \varepsilon_{it} \\
 & \quad t = 1, 2, \dots, T \\
 & \quad i = 1, 2, \dots, 50
 \end{aligned} \tag{10}$$

Contraste 9: Test F para efectos individuales

F = 378,12	p-value < 2,2e-16
------------	-------------------

Fuente: Elaboración propia a través de la herramienta RStudio

De acuerdo con este contraste, la constante es diferente para cada empresa. El modelo con efectos fijos se puede encontrar en la Tabla 7, concretamente el Modelo 5.2. Para

este, todas las variables son significativas, aunque la variable SENTIMIENTO lo es en menor nivel.

Tabla 7: Modelos efecto sobre el volumen en t

	Modelo 5.1.		Modelo 5.2.		Modelo 5.3.	
Beta 0	19113,74	(0,000000)			14694,10	(0,03023)
Sentimiento	7230,04	(0,5476)	20724,90	(0,05836)	20703,00	(0,05860)
Periodico	72830,42	(0,000000)	57827,90	(0,000000)	57857,20	(0,000000)
Entrada	85900,42	(0,000000)	92324,90	(0,000000)	92289,10	(0,000000)
Salida	45833,50	(0,000000)	437484,30	(0,000000)	437523,20	(0,000000)
R2	0,0066132		0,0065860		0,0065722	

Fuente: Elaboración propia a través de la herramienta RStudio

Antes de realizar un modelo con **efectos aleatorios**, se debe comprobar el Test de Hausman:

Contraste 10: Test Hausman

Chisq = 0,35836	p-value = 0,9857
-----------------	------------------

Fuente: Elaboración propia a través de la herramienta RStudio

En este caso, se acepta la hipótesis nula, indicando que el modelo con efectos aleatorios es consistente.

En la misma línea, los resultados aparecen reflejados en la Tabla 7 y el Modelo 5.3., donde todos los estimadores de las variables son significativos, aunque a diferentes niveles de confianza.

AJUSTE 2 - EFECTO SOBRE LA ROTACION DE VOLUMEN EN t + 1

En este ajuste se comprueba el efecto que tienen las variables explicativas en el momento t sobre la rotación de volumen en el momento $t + 1$. Para el caso de una noticia en viernes, sábado o domingo, se comprobar el efecto sobre el volumen del lunes, y lo mismo ocurre con los festivos.

El primer modelo a analizar es un MCO, especificado en la ecuación (11).

$$\begin{aligned}
 VO_{it+1} &= \beta_0 + \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it} \\
 t &= 1, 2, \dots, T \\
 i &= 1, 2, \dots, 50
 \end{aligned}
 \tag{11}$$

Tabla 8: Modelos efecto sobre el volumen en $t + 1$

	Modelo 6.1.		Modelo 6.2.		Modelo 6.3.	
Beta 0	19375,45	(0,000000)			14910,40	(0,03092)
Sentimiento	14054,67	(0,2560)	27683	(0,0142)	27660,30	(0,01427)
Periodico	66105,94	(0,000000)	50571,00	(0,000000)	50782,00	(0,000000)
Entrada	14358,49	(0,3957)	20881,60	(0,1721)	20844,50	(0,17286)
Salida	-4659,26	(0,8799)	-20370,60	(0,4652)	-20330,80	(0,46608)
R2	0,0027119		0,0021755		0,0021630	

Fuente: Elaboración propia a través de la herramienta RStudio

Los resultados de la estimación mediante MCO aparecen reflejados en la Tabla 8, y se corresponden con el Modelo 6.1. En este caso, el estimador de la variable PERIODICO es significativo, al igual que la constante.

Para realizar el modelo con **efectos fijos** especificado en la ecuación (12), antes se debe realizar el contraste F:

$$VO_{it+1} = \beta_1 Sent_{it} + \beta_2 Periodico_{it} + \beta_3 Entrada_{it} + \beta_4 Salida_{it} + u_{it}$$

Donde $u_{it} = \alpha_i + \varepsilon_{it}$
 $t = 1, 2, \dots, T$
 $i = 1, 2, \dots, 50$

(11)

Contraste 11: Test F para efectos individuales

F = 369,62	p-value < 2,2e-16
------------	-------------------

Fuente: Elaboración propia a través de la herramienta RStudio

Según este, la constante es diferente para cada empresa. Los resultados obtenidos de la estimación mediante efecto fijos aparecen en la Tabla 8 y en concreto, en el Modelo 6.2. De nuevo, el estimador de la variable PERIODICO vuelve a ser significativo.

Para realizar el modelo con **efectos aleatorios**, primero se tiene que realizar el Test de Hausman:

Contraste 12: Test Hausman

chisq = 0,36846	p-value = 0,985
-----------------	-----------------

Fuente: Elaboración propia a través de la herramienta RStudio

Para este caso, no se acepta la hipótesis nula, indicando que el modelo con efectos aleatorios es consistente. Este aparece reflejado en la Tabla 8, y se corresponde con el

Modelo 6.3., donde los estimadores de las variables SENTIMIENTO y PERIODICO son significativos.

4.2.3. REGRESIÓN DEL MERCADO

Siguiendo la misma línea que con los modelos anteriores, se realizan dos ajustes. El primer de ellos se corresponde con el efecto de las variables explicativas en t , sobre la rentabilidad en el mismo momento. Teniendo en cuenta el ajuste de los fines de semanas y festivos. Y, por otra parte, se comprueba el efecto de las variables explicativas en t , sobre la rentabilidad en el momento $t + 1$, teniendo en cuenta también el ajuste sobre fines de semana y festivos.

AJUSTE 1 - EFECTO SOBRE LA RENTABILIDAD EN t

En primer lugar, se realiza una regresión MCO con la rentabilidad obtenida del IGBM, la especificación de la misma se muestra en la ecuación (12):

$$RentIGBM_t = \beta_0 + \beta_1 Periodico_t + \beta_2 SentMAB_t + u_t \quad (12)$$

$$t = 1, 2, \dots, T$$

Tabla 9: Regresiones del mercado en t

	Modelo 7.1.	Modelo 7.2.
Beta 0	-0,0004384 (0,0847)	-0,0001733 (0,359)
Periodico	-0,0009665 (0,6398)	-0,0000977 (0,949)
Sentimiento	0,0121646 (0,3637)	0,0071957 (0,470)
R2	0,0002065	0,0001522

Fuente: Elaboración propia a través de la herramienta RStudio

Para este caso, los resultados del modelo aparecen detallados en la Tabla 9, concretamente en el Modelo 7.1. En este caso, la constante es significativa al 1%.

Posteriormente, se analiza el efecto sobre la rentabilidad del IBEX Small Cap, regresión especificada en la ecuación (13):

$$RentIBEXSC_t = \beta_0 + \beta_1 Periodico_t + \beta_2 SentMAB_t + u_t \quad (13)$$

$$t = 1, 2, \dots, T$$

Los resultados de este modelo aparecen en la Tabla 9, y se corresponden con el Modelo 7.2., donde ninguno de los estimadores de las variables es significativos.

AJUSTE 2 - EFECTO SOBRE LA RENTABILIDAD EN $t + 1$

Primero se estima un modelo MCO, correspondiente con la ecuación (14), si se tiene en cuenta la rentabilidad del IGBM.

$$RentIGBM_{t+1} = \beta_0 + \beta_1 Periodico_t + \beta_2 SentMAB_t + u_t \quad (14)$$

$$t = 1, 2, \dots, T$$

Para este caso, los resultados aparecen detallados en la Tabla 10, y en concreto, el Modelo 8.1. y ninguno de los estimadores es significativo.

Tabla 10: Regresiones del mercado en $t + 1$

	Modelo 8.1.	Modelo 8.2.
Beta 0	-0,0004267 (0,0933)	-0,0001690 (0,371)
Periodico	0,0018182 (0,3786)	0,0013395 (0,383)
Sentimiento	-0,0075897 (0,5709)	-0,0040604 (0,683)
R2	0,0002019	0,0001899

Fuente: Elaboración propia a través de la herramienta RStudio

A continuación, se estima la regresión con la rentabilidad del IBEX Small Cap, expresada en la ecuación (15).

$$RentIBEXSC_{t+1} = \beta_0 + \beta_1 Periodico_t + \beta_2 SentMAB_t + u_t \quad (14)$$

$$t = 1, 2, \dots, T$$

Para este modelo, que aparece detallado en la Tabla 10, en el Modelo 8.2., tampoco se encuentra ningún estimador que sea significativo.

5. CONCLUSIONES

El objetivo que se ha mantenido a lo largo de toda la investigación ha sido analizar la evolución de unas variables financieras como son la rentabilidad y el volumen de los títulos del MAB, y cómo influyen las publicaciones que se realizan en web sobre ellas.

Se cree que este razonamiento podía ser cierto debido a que actualmente la información que se tiene disponible en la web es de uso público y de fácil acceso. Por tanto, es una fuente de información fácil de consultar y que puede influenciar en las decisiones de inversión.

Para comprobar el objetivo base del trabajo se han elaborado una serie de modelos, partiendo de la base de datos que se ha creado. Ahora bien, de acuerdo con los resultados que se han obtenido, se puede afirmar que el sentimiento de las publicaciones no influye en la rentabilidad de las empresas. No obstante, otras como la variable SALIDA influye en la rentabilidad de las empresas, aspecto lógico puesto que, si una empresa se sale del mercado, deja de cotizar en el mismo, influyendo así en su rentabilidad.

No obstante, si se comprueba el efecto que tienen el sentimiento de las publicaciones sobre la variable de volumen, los resultados son algo más concluyentes. La variable sentimiento sí que influye en las decisiones de rotación de acciones, ya que un sentimiento positivo, indica un mayor volumen, es decir, una mayor rotación en el número de acciones. De la misma manera, la variable PERIODICO también influye en ese volumen, si la publicación proviene de una fuente de información emblemática, el volumen de rotación de las acciones es mayor.

Por su parte, si se analiza solo el mercado, la variable que muestra el sentimiento de las publicaciones sobre el mercado, deja de tener significatividad, mostrando que no tiene ningún efecto sobre la rentabilidad del mercado.

En resumen, es posible que los inversores tengan en cuenta las noticias que aparecen publicadas en la web a la hora de tomar las decisiones de inversión, puesto que existe una mayor rotación de acciones cuando aparece una publicación con sentimiento positivo, aunque esta no tenga efecto significativo sobre la rentabilidad de cada empresa. Se podría llegar a pensar que quizá el MAB está más enfocado a inversores profesionales, por el nivel de riesgo que asumen, y estos se ven menos influenciados por las publicaciones que aparecen en la web, de manera que, el efecto que tendrían estas sobre la rentabilidad del título no es significativo.

Sin embargo, el análisis que se ha realizado a lo largo de todo el trabajo puede ir mucho más allá, se podría realizar a mayor escala, analizando más variables financieras y teniendo en cuenta otros mercados más grandes. En definitiva, realizar un análisis a mayor escala con más títulos y más variables.

6. BIBLIOGRAFÍA

- Andbank, (2014). “¿Qué es el índice sentix?” [Disponible en: <https://www.andbank.es/observatoriodelinversor/que-es-el-indice-sentix/> (ultimo acceso: 6/02/2020)].
- Baker, Scott, Nicholas Bloom, and Steven Davis. (2016). “*Measuring Economic Policy Uncertainty Scott.*” Lecture Notes in Electrical Engineering 87 LNEE(VOL. 2): 383–91.
- Bravo-Marquez, Felipe, Marcelo Mendoza, and Barbara Poblete. (2014). “*Meta-Level Sentiment Models for Big Social Data Analysis.*” Knowledge-Based Systems 69(1): 86–99. <http://dx.doi.org/10.1016/j.knosys.2014.05.016>.
- Brown, G. y Cliff, M (2005). “*Investor sentiment and asset valuation*”. Journal of Business, 78 (2), 405-440.
- Bolsas y Mercados. (2019). “*Mercado Alternativo Bursatil: Empresas en expansión.*” [Disponible en: https://www.bolsasymercados.es/mab/docs/docsSubidos/MaB/PRESENTACI%C3%93N_EE_MAB_FEBRERO_2019.pdf (ultimo acceso: 23/02/2020)]
- Calomiris, Charles W., and Harry Mamaysky. (2019). “*How News and Its Context Drive Risk and Returns around the World.*” Journal of Financial Economics 133(2): 299–336.
- Corredor, P., Ferrer, E. y Santamaría, R. (2013). “*El sentimiento del inversor y las rentabilidades de las acciones. El caso español*”. International review of economics and finance, Vol. 27. 572-591.
- Correa, R., Garud, K., Londono, J. and Misláng, N. (2017). “*Sentiment in Central Banks’ Financial Stability Reports*”. SSRN 3091943.
- Cruz, Moritz. (2009). “*Liberalización Financiera y El Sentimiento Del Mercado: El Caso de La Economía Mexicana.*” EconoQuantum vol.5 no.2 Zapopan.
- Davis, A., Piger, J., and Sedor, L. (2006). “*Beyond the Numbers: An Analysis of Optimistic and Pessimistic Language in Earnings Press Releases*”. Federal Reserve Bank of St. Louis Working Paper Series.
- Dimpfl, T. (2011). “*The impacto f US news on the German stock market – An evento study análisis*”. The Quarterly Review of Economics and Finance 51, 389-398.
- Elton, Edwin J., Martin J. Gruber, and Jeffrey A. Busse. (2004). “*Are Investors Rational? Choices among Index Funds.*” Investments and Portfolio Performance: 145–72.

- Fama, Eugene. (1970). "*Efficient Capital Markets: A Review of Theory and Empirical Work.*" *The Journal Of Finance*: 383–417.
- Ferrer de Lucas, Sandra. (2015). "Finanzas Del Comportamiento. Revisión Bibliografica." Universidad de Comillas.
- Fraiberger, Samuel P. (2016). "*News Sentiment and Cross-Country Fluctuations.*" SSRN Electronic Journal: 125–31.
- Gaofeng, H., Ginger, J. and Wu, C. (2019). "*Asymmetric Effect of Media Tone on IPO Underpricing and Volatility*". *Emerging Markets Finance and Trade*. ISSN: 1540-496X.
- García, Diego. 2013. "*Sentiment during Recessions*". 68 *Journal of Finance*.
- Gatnet IT Glossary. (2018) "*What Is Big Data?*" [Disponible en: <https://www.gartner.com/en/information-technology/glossary/big-data> (ultimo acceso: 23/01/2020)]
- Gonzalez Nieto - Márquez, J. (2019). "*Mercado Alternativo Bursatil: Empresas en expansión*". Bolsas y Mercados.
- Heston, S. and Sinha, N (2015). "*News versus Sentiment: Predicting Stock Returns from News Stories,*" *Finance and Economics Discussion Series 2016-048*. Washington: Board of Governors of the Federal Reserve System,
- Hutto, C. and Gilbert, E. (2014). "*VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*". Georgia Institute of Technology.
- Isen, Alice M. (1987). "*Positive Affect, Cognitive Processes, and Social Behavior.*" In ed. Leonard B T - *Advances in Experimental Social Psychology* Berkowitz. Academic Press, 203–53.
- Kahneman, D., Slovic, P., and Tversky, A. (1982). "*Judgment under uncertainty: Heuristics and biases*". New York: Cambridge University Press.
- Liu, B. (2010). "*Sentiment Analysis and Subjectivity*". *Handbook of Natural language processing*, 2, 627-666.
- Liviano, D y Pujol, M. (2013). "*Modelos econométricos avanzados con R*". Universitat Oberta de Catalunya.
- Loughran, T. and Mcdonald, B. (2011). "*When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks*". *The Journal of Finance*. Vol. LXVI, nº1. Pp. 35-65.

- Ludvigson, S.C. (2004). "Consumer confidence and consumer spending". *The Journal of Economic Perspectives*, 18 (2), 29-50.
- Mileydi, D y Pinto, D. (2018). "Influencia de Variables de Sentimiento y Expectativas de los Inversionistas sobre el Comportamiento del índice S&P 500". Colegio de Estudios Superiores de Administración (CESA).
- Montes-y-gómez, Manuel. (2001). "Minería de Textos: Un Nuevo Reto Computacional." 3rd International Workshop on Data Mining MINDAT2001. <http://ccc.inaoep.mx/~mmontesg/publicaciones/2001/MineriaTexto-md01.pdf>.
- Niederhoffer, V. (1971) "The análisis of words events and stock prices". *J. Business* 44, 193-219.
- Nyman, Rickard et al. (2018). "News and Narratives in Financial Systems: Exploiting Big Data for Systemic Risk Assessment." *SSRN Electronic Journal* (704).
- Pang, B. and Lee, L. (2005). "Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales". *Proceedings of the 43rd anual meeting on association for computational linguistics*, pp. 115-124. Association for Computational Linguistics.
- Pang, B, Lee, L and Vaithyanathan, S. (2002). "Thumbs up? Sentiment Classification using Machine Learning Techniques". *Association for Computational Linguistics*, 79-86.
- Perez, C. y Santin, D. (2008). "Minería de datos. Técnicas y herramientas". Ed. Paraninfo.
- Qiu, L and Welch, L (2006). "Investor sentiment mesures". *NBER Working Paper No. 10794*.
- Ricciardi, Victor. (2008). "The Psychology of Risk: The Behavioral Finance Perspective." *Handbook of Finance* (May): 269–92.
- Riquelme, José C., Roberto Ruiz, and Karina Gilbert. (2006). "Minería de Datos: Conceptos y Tendencias." *Inteligencia Artificial* 10(29): 11–18.
- Ruiz Alonso, Marcos. (2018). "El impacto de los titulares de las noticias en el mercado bursátil estadounidense." *Universidad de Alicante*.
- Sánchez, José Ramón. (2008). "El Mercado Alternativo Bursátil (MAB): El Mercado de Las PYMES". *Estrategia financiera* (251): 64–68.
<http://dialnet.unirioja.es/servlet/articulo?codigo=2652786&info=resumen&idoma=SPA>.
- Schiller, Robert. (2000). "Irrational Exhuberance". Ed. Princeton University Press.

- Schumaker, R. and Chen, H. (2009). “*Textual análisis of stock market prediction using breaking financial news: the AZfin text system*”. ACM Trans. Inform. Syst. 27, 12.
- Schwarz, M, (2002). "*Situated Cognition and the Wisdom of Feelings: Cognitive Tuning*". Guilford Press, pp. 144-166
- Shapiro, Adam Hale, Moritz Sudhof, Daniel Wilson. (2020). “*Measuring News Sentiment*”. Federal Reserve Bank of San Francisco Working Paper 2017-01.
- Subdirección general del apoyo a las PYME, Ministerio de industria. (2009). “*Retrato de La PYME 2009.*”: 1–7.
- Tetlock, P., Saar-Tsechansky, M. and Macskassy, S. (2008) “*More than words: quantifying language to measure firm’ fundamentals*”. J. Finance 63. 1437-1467.
- Tetlock, P. (2007). “*Giving content to investor sentiment: the role of media in the stock market*”. J. Finance 62. 1139-1168.
- Thorsrud, Leif Anders. (2017). "*Nowcasting Using News Topics. Big Data versus Big Bank*". SSRN Electronic Journal (1): 1–60.
- Tversky, A, Kahneman, D, (1974). "*Judgment under Uncertainty: Heuristics and Biases*". Science. 185 (4157): 1124-1131.
- Xiaodong Li, Hoaran Xie, Li Chen, Jianping Wang and Xiaotie Deng, (2014). “*News Impact on Stock Price Return via Sentiment Analysis*”. Knowledge-Based Systems 69(1): 14–23. <http://dx.doi.org/10.1016/j.knosys.2014.04.022>.

ANEXO 1

In [1]:

```
#Código para poner el notebook a pantalla completa
from IPython.core.display import display, HTML
display(HTML("<style>.container { width:87% !important; }</style>"))
```

In [2]:

```
import logging
import json
import pickle
from selenium import webdriver

from urllib.parse import urldefrag, urljoin
import urllib.request
from collections import deque
from bs4 import BeautifulSoup
from bs4.element import Comment
```

```
import re
```

```
import time
from tqdm import trange, tqdm_notebook
```

```
# import warnings
# warnings.filterwarnings('ignore')
```

In [3]:

```
class SeleniumCrawler(object):

    def __init__(self, base_url, wait_for = 4,
chromedriver_path='../scripts/chromedriver.exe'):
        '''
        Inicializacion
        '''
        self.browser = webdriver.Chrome(chromedriver_path) #Add path
to your Chromedriver
        self.base_url = base_url
        self.wait_for = wait_for
        self.crawled_urls = [] #List to keep track of URLs we have
already visited
        self.url_new = {} #JSON para guardar datos
        self.url_date_source = {} #Diccionario para obtener fecha y
fuente correspondientes de una url
        self.url_queue = deque() #Add the start URL to our list of
URLs to crawl

        #Nota: Un objeto deque es un contenedor de datos del módulo
collections similar a una lista o una cola que permite añadir
        # o suprimir elementos por sus dos extremos.

    def get_page(self, url):
        '''
        Método que obtiene la página web para ser crawleada.
        '''
        try:
            self.browser.get(url)
```

```

        self.browser.execute_script("window.scrollTo(0,
document.body.scrollHeight);") #Para páginas como Facebook
    try:
        self.browser.implicitly_wait(self.wait_for)
    except:
        logging.warning('I can not get coordenates for %s',
url)
        return (self.browser.page_source)
    except Exception as e:
        logging.exception(e)
    return

def get_soup(self, html):
    """
    Obtener de BeautifulSoup el html de la página web con la que
se va a trabajar.
    """
    if html is not None:
        soup = BeautifulSoup(html, 'html.parser') #lxml
        return soup
    else:
        return

def get_links(self, soup):
    """
    Método para obtener los links de una búsqueda (en este caso
Google). Se encuentra con BeautifulSoup mediante
'a' y la clase a buscar 'l lLrAF'.

    Notas:
    - If url contains a fragment identifier, return a modified
version of url with no fragment identifier, and the
    fragment identifier as a separate string. If there is no
fragment identifier in url, return url unmodified and
    an empty string.
    """
    #links
    for link, date, source in zip(soup.find_all('a', 'l lLrAF'),
soup.find_all('span', class_='f
nsa fwzPFf'),
soup.find_all('span',
class_='xQ82C e8fRJf')):
        url = link['href']
        if url not in self.url_queue and url not in
self.crawled_urls:

            #dates
            matched = re.search(r'<span class="f nsa
fwzPFf">(.*?)</span>', str(date))
            if matched:
                fecha = matched.group(1) ## 'found word:cat'
            else:
                fecha = None

            #sources
            matched = re.search(r'<span class="xQ82C
e8fRJf">(.*?)</span>', str(source))
            if matched:
                fuente = matched.group(1) ## 'found word:cat'

```

```

        else:
            fuente = None

        self.url_queue.append(url)
        self.url_date_source[url] = {'date': fecha,
'source': fuente}

def title_visible(self, element):
    """
    Encontrar únicamente los tags relevantes: titulo.
    source-code:
https://stackoverflow.com/questions/1936466/beautifulsoup-grab-visible-webpage-text
    """
    if element.parent.name in ['title']:
        return True

def p_visible(self, element):
    """
    Encontrar únicamente los tags relevantes: texto principal.
    source-code:
https://stackoverflow.com/questions/1936466/beautifulsoup-grab-visible-webpage-text
    """
    if element.parent.name in ['p']:
        return True

def text_from_html(self, html):
    """
    Método para la búsqueda de título y texto en un html.
    """
    soup = self.get_soup(html)
    texts = soup.findAll(text=True)
    visible_title = filter(self.title_visible, texts)
    visible_p = filter(self.p_visible, texts)
    return (" ".join(t.strip() for t in visible_title), "
".join(t.strip() for t in visible_p))

def get_data(self, url):
    """
    Método para obtener el título y la noticia desde una url.
    """
    try:
        html = urllib.request.urlopen(url)
        title, new = self.text_from_html(html)
        time.sleep(5)
    except:
        title = None
        new = None
        time.sleep(1.5)
    return title, new

def json_output(self, data, filename):
    """
    Método para guardar los datos en formato json.
    """

```

```

    try:
        with open('../data/'+str(filename)+'.json', 'w',
encoding='utf8') as fp:
            json.dump(self.url_new, fp, indent=4,
ensure_ascii=False)
    except:
        with open('../data/'+str(filename)+'.pickle', 'wb') as fp:
            pickle.dump(self.url_new, f)

def run_crawler(self, key_word:str = None, npags=500):
'''
Método que sirve para correr el crawler. A partir de una web
base de Google (la cual debe ser la segunda página para
que en el link este incluido la cadena "start"), se realiza un
bucle sobre todas las paginas para coger todos los enlaces
de cada una de ellas. Después, sobre cada uno de estos enlaces
se guarda el título, la noticia, etc. Todo se guarda en un
diccionario de diccionario, donde la clave principal es el tã-
tulo, y el diccionario interno contiene los metadatos nombrados.

Finalmente se guarda en un JSON (por defecto
../data/name.json)

Argumentos:
- key_word: palabra clave para la que identificar el conjunto
de noticias. Tambien sera el nombre del json a guardar.
- npags: numero de paginas a recorrer.
'''
count_nones = 1 #Para añadir aquellos enlaces que no se han
podido extraer la informaciã³n.
try:
    for j in tnrange(npags, desc='Pã³ginas'):
        nlinks = str(j*10)
        l =
re.compile("(.*start=)\d{1,3}(&.*)").split(self.base_url)

        try:
            html = self.get_page(l[1]+nlinks) #De la base
donde se obtienen todas las URLs con get_links
        except:
            if l[0] not in self.crawled_urls: #En caso de que
sólo sea 1 página de links la que exista
                html = self.get_page(l[0])
                self.crawled_urls.append(l[0])
            else:
                self.json_output(self.url_new, key_word)
                return(None)

        soup = self.get_soup(html)
        if soup is not None:
            self.get_links(soup) #Obtiene todos los links de
la pagina

        if len(self.url_queue) == 0: #Caso de que no haya más
links a scrapear
            self.json_output(self.url_new, key_word)
            return(None)
        for i in tqdm_notebook(range(len(self.url_queue)),
desc='Enlaces', leave=False):
            current_url = self.url_queue.popleft()

```

```

        if (title is not None) & (new is not None):
            self.url_new[current_url] = {'title':title,
                                         'new':new,

'date':self.url_date_source[current_url]['date'],

'source':self.url_date_source[current_url]['source'],

'keyword':key_word}
        else:
            self.url_new[current_url] =
{'title':'WARNING_'+str(count_nones)+'-'+key_word,
                                         'new':None,
                                         'date':None,
                                         'source':None,

'keyword':key_word}
            count_nones+=1
            time.sleep(0.1)
            self.json_output(self.url_new, key_word)
    except:
        self.json_output(self.url_new, key_word)
    return (None)

```

ANEXO 2

Tabla 11: Número de fuentes consultadas para cada empresa, 2008 – 2018

EMPRESA	NÚMERO DE FUENTES CONSULTADAS
INKEMIA	13
AB-BIOTICS SA	47
ADL BIONATUR	22
AGILE CONTENT	24
ALQUIBER QUALITY SA	9
ALTIA	92
ATRYS HEALTH SA	42
BODACLICK	47
CARBURES	50
CATENON SA	32
CERBIUM HOLDING	8
CLERHP ESTRUCTURAS	9
CLEVER GLOBAL SA.	22
COMMCENTER	41
EBOSS ENERGY	21
ELZINC (ASTURIANA DE LAMINADOS)	123
EUROCONSULT ENGINEERING CONSULTING GROUP SA.	46
EUROESPES, SA	32
EURONA WIRELESS TELECOM SA	30
FACEPHI BIOMETRIA, S.A.	21
GIGAS HOSTING SA	18
GOWEX	73
GREENALIA	43
GREENERGY RENOVABLES SA	41
GRINÓ ECOLOGIC	10
GRUPO NOSTRUM RNL, S.A	5
HOME MEAL REPLACEMENT SA	67
IFFE FUTURA	10
IMAGINARIUM	176
INCLAM SOCIEDAD ANONIMA	34
LLEIDANETWORKS SERVEIS TELEMATICS SA	35
MASMOVIL	47
MEDCOM TECH, SA	8
MONDO TV IBEROAMERICA	22
NBI BEARINGS EUROPE SA	13
NEGOCIO & ESTILO DE VIDA	2
NEOL BIO	15
NETEX KNOWLEDGE FACTORY SA	8

NEURON BIO	25
NPG TECHNOLOGY	6
PANGAEA ONCOLOGY SA	30
ROBOT SA	13
SEAFOOD GLOBAL PROCESSOR (LUMAR)	6
SECUOYA, GRUPO DE COMUNICACION SA	120
SUAVITAS // VOUSSE	13
TECNOQUARK	6
THINK SMART SA	60
TIER 1 TECHNOLOGY SOCIEDAD ANONIMA	16
VOZTELECOM OIGAA360 SA.	27
ZINKIA ENTERTAINMENT, SA	53

Fuente: Elaboración propia a partir de la información obtenida mediante webscraping