



Facultad de Ciencias Económicas y Empresariales

# **EQUITY VS. DEBT: IMPACTO DE LEHMAN BROTHERS EN EL JOURNAL OF FINANCE**

Autor: Guillermo Pedrosa Díez  
Director: Alejandro Rodríguez Gallego

MADRID | Junio 2022

## **EQUITY VS. DEBT: IMPACTO DE LEHMAN BROTHERS EN EL JOURNAL OF FINANCE**

Este trabajo busca ofrecer una imagen sintética, pero a la vez integral, sobre la evolución de la literatura financiera, acerca de los recursos propios (*Equity*) y ajenos (Deuda), recogida en la revista *The Journal of Finance*, con especial foco en el impacto que haya podido tener la Gran Crisis Financiera tras la caída de Lehman Brothers. Por lo tanto, se llevarán a cabo dos particiones distintas de la muestra, la primera y de mayor entidad será la que divida la base de datos entre el período previo y posterior a la caída de Lehman, mientras que, por otro lado, se llevará a cabo una contraposición de términos, en particular, las palabras *Equity* y *Debt*. Finalmente, se ofrecerán los resultados del estudio de investigación, dónde se mostrarán los efectos, hallados tras analizar la base de datos creada a partir de los documentos de la revista.

This paper aims to provide a synthetic, yet comprehensive, picture of the evolution of the financial literature on equity and debt in *The Journal of Finance*, with a special focus on the impact of the Great Financial Crisis after the collapse of Lehman Brothers. Therefore, two different partitions of the sample will be carried out, the first and larger one dividing the database between the period before and after the fall of Lehman Brothers, while on the other hand, a contrast of terms will be carried out, specifically with the words *Equity* and *Debt*. Finally, the results of the research study will be presented, showing the effects found after analysing the database created from the journal documents.

Palabras clave: Análisis de Textos, Equity, Debt, Lehman Brothers, Journal of Finance

Keywords: Textual Analysis, Equity, Debt, Lehman Brothers, Journal of Finance

## **ÍNDICE**

**1.-INTRODUCCIÓN**

**2.- HIPÓTESIS INICIAL**

**3.- REVISIÓN DEL ESTADO DE LA LITERATURA**

**4.- BASE DE DATOS Y METODOLOGÍA**

**5.- RESULTADOS**

**6.- CONCLUSIONES**

## ÍNDICE DE FIGURAS

- *Figura 1. Evolución de la emisión de préstamos apalancados tanto a nivel de Estados Unidos, como del resto del mundo. Unidades en miles de millones de dólares. Fuente: Standard and Poor's Leveraged Commentary and data and IMF staff calculations.*
- *Figura 2. Evolución del mercado hipotecario por volumen de las emisiones (garantizas o no) de derivados financieros, junto con el porcentaje de hipotecas subprime sobre el total. Fuente: Inside Mortgage Finance.*
- *Figura 3. Volumen de las operaciones de salida a bolsa o IPO (initial public offering por sus siglas en inglés) desde 1995 hasta 2020, segmentados por geolocalización y tipo. Fuente: Refinitiv y Bloomberg Data.*
- *Figura 4. Evolución histórica del coste medio de la deuda y del equity, junto con el incremento del endeudamiento de las empresas no financieras. Fuente: Federal Reserve Economic data.*
- *Figura 5. Conjunto de métricas clave para medir la relevancia de un journal, siendo cuanto mayor el Impact Factor mejor; cuando menor el Citation Indicator, mejor; y cuanto mayor el número de citaciones, mejor. Fuente: Journal of Citation Reports, 2020.*
- *Figura 6. Top 10 palabras más utilizadas por los investigadores de The Journal of Finance a lo largo de los veinte años de muestra.*
- *Figura 7. Diferencias entre las posiciones relativas de las 100 palabras más repetidas por conteo entre períodos. Léase cómo que "financing" aparece post Lehman treinta puestos por encima de su posición pre Lehman en la clasificación de palabras más repetidas.*
- *Figura 8. Top 15 palabras con mayor relevancia según TF-IDF a lo largo de los veinte años de muestra.*
- *Figura 9. Diferencias entre las posiciones relativas de las 100 palabras con mayor factor TF-IDF entre períodos. Léase cómo que "financing" aparece post Lehman veinticinco puestos por encima de su posición pre Lehman en la clasificación de palabras más valoradas.*
- *Figura 10. Tracking de la evolución de ciertas palabras relacionadas con el término "Debt" a lo largo del período de la muestra, 2000-2019. Representado mediante gráficos de densidad.*
- *Figura 11. Tracking de la evolución de ciertas palabras relacionadas con el término "Equity" a lo largo del período de la muestra, 2000-2019. Representado mediante gráficos de densidad.*

## INTRODUCCIÓN

En este trabajo se pretende estudiar la evolución de dos figuras contrapuestas en el haber de los balances de situación de las empresas, como son los recursos propios (“*equity*” en su versión inglesa) y los recursos ajenos (“*debt*” en su versión inglesa).

Por lo general, ambas figuras son fundamentales en tanto y cuanto representan la estructura de capital de una compañía, que define la forma en la que la empresa financia sus activos. Es, por tanto, un elemento primordial en el mundo financiero a la hora de realizar tanto análisis, como valoraciones, como tomar decisiones de inversión.

Es relevante señalar, aunque posteriormente se profundizará más en este punto, la principal diferencia entre ambos recursos: el coste del capital para la compañía. Por lo general, y salvo raras excepciones<sup>1</sup>, los recursos propios tienen un coste mayor que los recursos ajenos, o financieros. Esto se entiende desde el punto de vista del inversor, que es quién exige una rentabilidad mayor (un coste mayor para la compañía) a sus inversiones en *equity* que a sus inversiones en deuda. La razón última de esta diferencia está en el riesgo que conllevan ambas inversiones, ya que, a mayor riesgo, mayor rentabilidad (esperada y exigida); los socios requerirán una mayor retribución, ya que, en caso de concurso, sólo cobrarán una vez satisfechos los intereses de los acreedores.

El estudio contrapuesto de ambas figuras se llevará a cabo mediante el análisis textual de los artículos de la revista titulada *Journal of Finance*, editada y dirigida por la American Finance Association, y publicada por Wiley-Blackwell; uno de los *journals* de mayor prestigio en el ámbito de las finanzas, y de la economía también<sup>2</sup>.

De cada artículo del *journal* se extraerán las palabras que lo componen, y se buscará establecer una relación entre ambos términos a través del estudio de su conteo, así como otras métricas relevantes en el ámbito del análisis de textos, que se implementarán usando las distintas herramientas y paquetes que existen, basados en el lenguaje de programación R. Todo el análisis se llevará a cabo desde la plataforma de R Studio.

---

<sup>1</sup> La teoría asume mercados de deuda infinitos, dónde no se plantean restricciones de acceso a la financiación, ni la imposibilidad de los socios de levantar más capital. En estos casos, muy parecidos a los actuales, se puede llegar a dar la situación en que la financiación exija un mayor retorno del capital que el actual retorno exigido por los socios; también en casos de compras hiper-apalancadas (LBOs, por sus siglas inglesas de Leveraged Buy Out). Véase ampliamente Modigliani, F., & Miller, M. H. (1958). The Cost of Capital, Corporation Finance and the Theory of Investment. *American Economic Review*, 48(3), 261.

<sup>2</sup> Véase el *Journal Citation Reports*, 2020 metrics.

Para realizar la extracción de los datos, aunque se profundizará más adelante en el apartado de Metodología, se ha empleado la herramienta NVivo, que extrae automáticamente el texto de cada uno de los archivos en los que se guardan cada uno de los artículos.

Como parte de la investigación, se tratará de mostrar la evolución de ambos términos en particular junto con otros términos relacionados con los mismos, es decir, por un lado, todos aquellos que se relacionan con la parte del balance de recursos propios, y por otro, todos aquellos relacionados con los recursos ajenos.

Además, se pondrá el foco en el impacto que tuvo la caída del gigante bancario Lehman Brothers<sup>3</sup>, así como la consecuente crisis económica que se inició, si bien este trabajo de investigación se centrará únicamente en su carácter financiero. Con el tiempo, y debido a sus devastadores efectos, hoy en día es conocida como la Gran Crisis Financiera, a la que este trabajo se referirá como GCF.

Esta investigación parte de la hipótesis de que existió un cambio de paradigma tras la caída del conocido banco de inversión, que propició, junto con la GCF, una revisión acerca del papel de la deuda en la economía. Cabe señalar, que una de las principales razones de la caída de Lehman Brothers fue la enorme exposición que tenía el banco a obligaciones de deuda garantizada (CDOs, por sus siglas inglesas de *Collateralized Debt Obligations*) y valores respaldados por hipotecas (MBSs, por sus siglas inglesas de *Mortgage-Backed Securities*), por lo que sería de esperar que gran parte de la literatura centrara el foco al menos momentáneamente en la deuda y el estudio de las causas de dicha crisis financiera.

En este sentido, y dado que el *Journal of Finance* es un *journal* bimensual, junto con el hecho de que los autores e investigadores, tardan cierto tiempo en desarrollar y llevar a término sus artículos, y que el proceso editorial requiere de un proceso de revisión hasta que se culmina con la publicación, para el estudio del impacto que tuvo la caída de Lehman Brothers en estos términos (*equity* y deuda) se usará como fecha de corte el año 2009.

---

<sup>3</sup> El día 15 de septiembre de 2008, el banco Lehman Brothers “*filed for bankruptcy*”, es decir, inició formalmente su concurso de acreedores que acabaría en la liquidación de los activos, parte de los cuales fue vendido a Barclays, quién se había postulado como uno de los posibles compradores de la entidad previamente al concurso de acreedores.

Junto con el estudio llevado a cabo, se buscará ofrecer, en base a la amplia literatura disponible, razones que apoyen las hipótesis que se describirán en los apartados siguientes, así como posibles explicaciones a los resultados obtenidos en el estudio. En cualquier caso, el grueso del trabajo se restringirá a ofrecer los resultados de forma descriptiva, haciendo hincapié en las ya mencionadas áreas de recursos ajenos y propios.

Únicamente a propósito de ofrecer una descripción sucinta del trabajo, se procederá a exponer de forma ordenada los capítulos de este estudio. En primer lugar, se realizará un análisis de los elementos que conforman la hipótesis de partida que suscita la realización de esta investigación. Cabe señalar, que es a raíz del trabajo realizado por D. Ramón Hernán López en el ámbito de su Doctorado en *Behavioral Finance*<sup>4</sup>, o finanzas del comportamiento, que se origina la idea de estudiar en profundidad si las situaciones previa y posterior a la caída de Lehman Brothers, habrían afectado a los investigadores y los trabajos que realizaron, así como su posible influencia en las decisiones de la junta editorial de la revista *Journal of Finance*.

En segundo lugar, se describirá en profundidad la metodología usada, así como una revisión de la literatura científica dónde se haya aplicado el análisis textual para otorgar al lector un contexto en el que poder asentarse, a la vez que comparar y descubrir nuevas aplicaciones de dicha metodología. En tercer lugar, se mostrarán los resultados de la investigación, dónde, además, se analizarán con la ayuda de nuevas formas de visualización de los datos, dichos resultados.

Por último, se realizará un compendio de todo lo expuesto, así como de los resultados obtenidos y se verificará si la hipótesis inicial se cumple o no; junto con las conclusiones de la investigación.

## **HIPÓTESIS INICIAL**

El presente trabajo se inicia bajo el supuesto de que, tras la GCF, el paradigma financiero en la economía, a nivel general, y en lo académico, a nivel particular, cambió. Específicamente, se vio afectado en tanto y cuanto, se puso el foco en la deuda que había

---

<sup>4</sup> Véase como parte de dicha investigación el artículo de Hernán, R., Corzo, T. and Pedrosa, G., “Behavioral Finance in a hundred Keywords”, (Unpublished).

causado tal colapso, tanto en la primera economía mundial, Estados Unidos, como a nivel europeo, y, aunque en menor medida, también a nivel global.

Esta hipótesis se basa en la idea de que, debido a que fue la deuda la causante de la crisis, tal y como se explicará posteriormente, tendría sentido que los investigadores económico-financieros centrasen sus estudios en analizar los tipos de deuda y las razones que llevaron al colapso. En pro de justificar y apoyar el análisis de la hipótesis descrita, se procederá a exponer las razones que sitúan a la deuda como responsable de la GCF. También es verdad, que, desde otro punto de vista, la deuda podría ser categorizada, no sólo como causante de la crisis, sino también, como el medio que sirvió de canal a través del cual el resto de factores económicos provocaron la GCF.

Cabe señalar, que, en parte, el exceso de deuda disponible era debido al apetito de riesgo desmesurado de los bancos, ya que estaban en búsqueda de grandes rentabilidades, en unos entornos macroeconómicos favorables<sup>5</sup>.

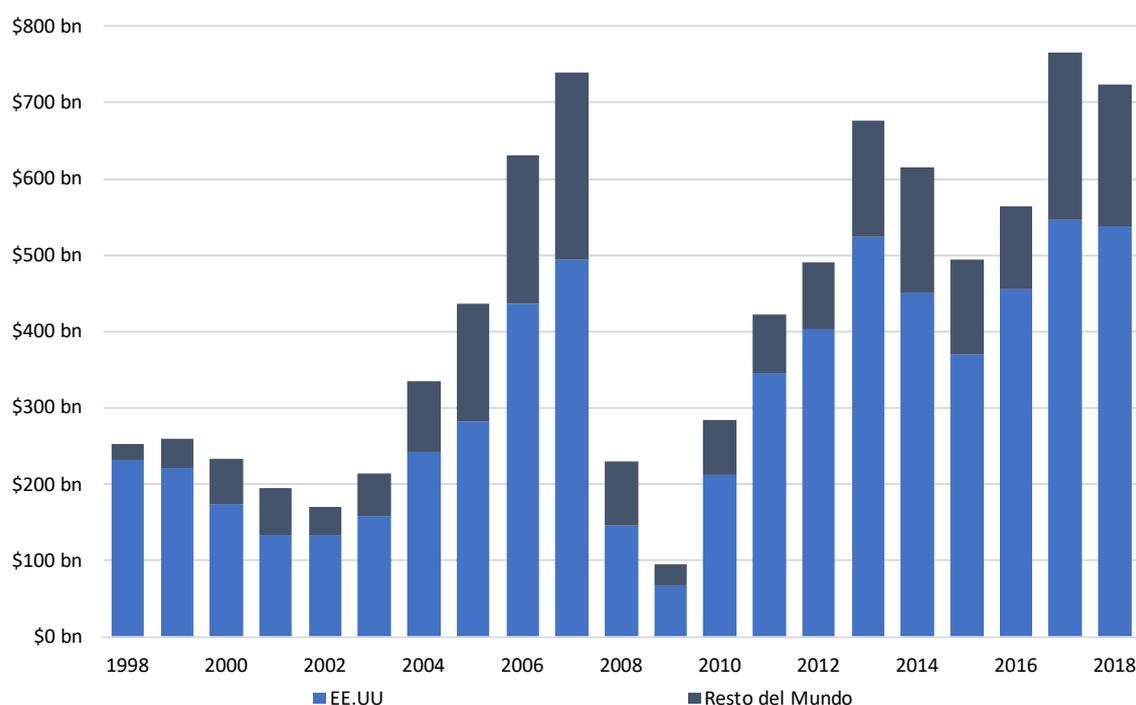


Figura 1. Evolución de la emisión de préstamos apalancados tanto a nivel de Estados Unidos, como del resto del mundo. Unidades en miles de millones de dólares. Fuente: Standard and Poor's Leveraged Commentary and data and IMF staff calculations.

<sup>5</sup> De todas formas, en 2007, Estados Unidos tuvo el menor crecimiento del PIB en cuatro años con un discreto 0,9%, según datos del Banco Mundial (véase <https://datos.bancomundial.org/indicador/NY.GDP.PCAP.KD.ZG?locations=US>).

Como se observa en el gráfico, la emisión de préstamos apalancados se había disparado justo antes del estallido de la GCF y la caída de Lehman Brothers en 2008. Esto provocó que el shock inicial fuera mucho peor de lo que hubiera sido sin tanta deuda apalancada. Aunque, por otro lado, también es destacable el rápido ritmo de recuperación tras tocar fondo en el año 2009, volviendo en 2017 a niveles previos al estallido de la crisis. En cualquier caso, tal y cómo se explicará posteriormente, la literatura mantiene el criterio de considerar que la gran crisis económica fue consecuencia de una sucesión de crisis menores, como la financiera, la de deuda soberana o la hipotecaria (Thornton, 2021); y que sus efectos, lejos de afectar únicamente a Estados Unidos, fueron afectando al resto del mundo, desde Europa como principal perjudicado, hasta las economías emergentes, incluida China (Levinson, 2009). Pero si se ha de señalar un punto de partida, se ha de referir a un principal desencadenante de la crisis: las hipotecas.

Tal y cómo se ha indicado sucintamente supra, una de las principales causas de la GCF, fue la conocida como crisis de las *subprime*, en referencia a las hipotecas basura. Las hipotecas basura se conocen como tal debido a que se concedieron a personas con perfiles crediticios muy dudosos, con niveles de ingresos muy reducidos, y altas probabilidades de impago o *default*. A cambio, las entidades financieras mantenían políticas de tipos de interés elevados, que fomentaban la agrupación de hipotecas en instrumentos financieros derivados, conocidos como CDOs. El proceso a seguir, desde que se concedía la hipoteca al consumidor, hasta que se vendía como parte de un conjunto de activos agrupados bajo derivados financieros, era el siguiente:

- En primer lugar, los empleados del departamento de ventas, también conocidos como *brokers*, de los bancos “*retail*”, es decir, de la banca comercial, sugerían al individuo contratar una hipoteca. En este proceso, los mecanismos de control del riesgo se omitían en pro de un mayor volumen de negocio, que consiguiera mantener los altos niveles de retorno de los bancos.
- En segundo lugar, los bancos comerciales, conscientes de que dichas obligaciones representaban un riesgo importante en sus balances, procedían a su venta a descuento. En particular, los derechos de crédito que representaban cada hipoteca se titularizaban para emitir instrumentos representativos de deuda en bloque. Es aquí donde el papel de los bancos de inversión se tornaba fundamental, pues entraban en juego para estructurar y confeccionar las

titulizaciones de los créditos que posteriormente conformarían los paquetes de deuda colateralizada, o CDOs.

- Una vez creados, estos paquetes eran vendidos por los bancos de inversión a diferentes tipos de compradores, desde inversores institucionales como *brókers* de seguros o fondos de pensiones, como a otros bancos de inversión. Los paquetes estaban formados por distintos tramos que diferían en calidad crediticia. Aquí se ha de desarrollar la explicación de un elemento clave para que nadie desconfiara de la calidad de dichos activos: la calificación crediticia.

Las agencias de rating debían emitir sus opiniones y juicios sobre si estos paquetes entrañaban o no algún riesgo de cara a cobrar los créditos. En contra de lo que se pudiera entender razonable (si se realizara un análisis pormenorizado de las hipotecas que conforman el CDO se podría observar la enorme cantidad de riesgo que entraña por el bajo perfil crediticio de los deudores), las agencias de rating otorgaron calificación de AAA, a cada uno de los tramos superiores de dichos instrumentos (GRIFFIN & TANG, 2012)<sup>6</sup>. Las hipotecas *subprime* difieren en un aspecto importante de los créditos basados en el resto de las hipotecas. En el caso de los valores respaldados por hipotecas “conformes” o solventes, la calificación crediticia, así como la seguridad de la misma, está fijada de antemano. En cambio, en el caso de los valores respaldados por hipotecas *subprime*, el grado de protección crediticia para aquellos inversores que se fiaron de la calificación de AAA era reducido, y, de hecho, se fue acumulando a lo largo de la vida inicial de los créditos, siempre que no hubiera mora.

Fue en el punto álgido del boom inmobiliario, que esta forma de estructuración y revisión crediticia de los CDOs, propició el derrumbe en cadena de dichas obligaciones. El sustento de estas se encontraba en el incremento de los precios de la vivienda, por eso, el boom inmobiliario fue una *conditio sine qua non* para que se diera esta situación. Esto se debía a que, a mayores precios de vivienda, mayores hipotecas se podrían solicitar, por lo que personas que no pudieran pagar su primera hipoteca, dado que la casa (garantía hipotecaria) había aumentado su valor, podrían solicitar una ampliación de la hipoteca, o bien una nueva que le permitiera diferir el pago de intereses, a la vez que otorgarle la

---

<sup>6</sup> Según el estudio llevado a cabo por ambos autores, los ratings crediticios otorgados por las agencias de rating eran el resultado de multitud de ajustes de difícil justificación. Así, de 195 CDOs analizados, el 91,2% de los mismos únicamente se situaba en el rating AA de las propias agencias. Los autores llegan a demostrar que eliminando los ajustes realizados y la desviación de criterio implementada, esos ratings de AAA, estaban estructurados para soportar ratings de BBB.

entrada de *fresh money*. Como se puede observar, la economía había entrado en un círculo vicioso que se rompería cuando la burbuja inmobiliaria explotó. Dicha situación conllevó una rebaja del precio de las casas, que, en el corto plazo, supuso la imposibilidad de aumentar el endeudamiento para hacer frente a las obligaciones crediticias y, por tanto, al inicio, de forma masiva, de default en los CDOs. Esto afectó sobre todo a personas con bajos recursos económicos, que eran quienes más se endeudaban por encima de sus capacidades, sin que nadie impidiera que pudieran solicitar más deuda (Baker & Diaz, 2004)<sup>7</sup>.

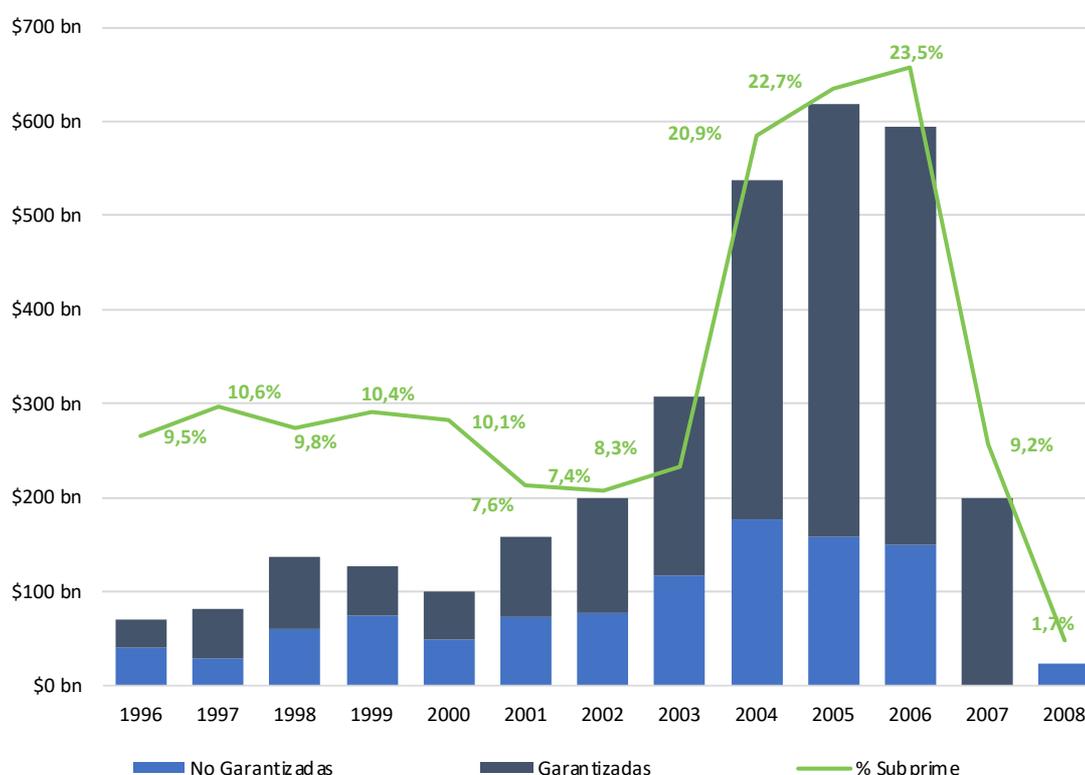


Figura 2. Evolución del mercado hipotecario por volumen de las emisiones (garantizadas o no) de derivados financieros, junto con el porcentaje de hipotecas subprime sobre el total. Fuente: Inside Mortgage Finance.

<sup>7</sup> Tal y cómo señalaban Baker & Diaz, ya en el año 2004, con tres años de antelación al estallido de la burbuja, el incremento en precios, que llegaba a ser del 70% en algunas ciudades del estado de California, no era sostenible. De hecho, en su artículo, indican que el gran incremento en las tasas de consumo observado en los años anteriores a la realización del estudio era en gran parte explicable porque se había financiado contra las viviendas de los ciudadanos. El hecho de poner en garantía casas con precios inflados a un ritmo anual de cuatrocientos mil millones de dólares, suponía un grandísimo riesgo para la economía, pues el mercado inmobiliario representaba el 5% del producto interior bruto.

Cómo se puede observar en el gráfico, en los años previos al *crash* de la burbuja inmobiliaria en el año 2007, que provocó la crisis hipotecaria y posteriormente la financiera, el volumen de hipotecas *subprime* concedidas llegó a suponer casi una cuarta parte del total del mercado hipotecario. Esto implica que, de cada cuatro hipotecas concedidas, una tenía un alto riesgo de impago, que, si se lleva el plano teórico, sería asumir unos niveles de morosidad del 25%, algo desorbitado si se compara con los niveles actuales<sup>8</sup>. Además, se manifiesta la relación directamente proporcional entre el aumento del porcentaje de hipoteca *subprime*, con el incremento en la emisión de derivados financieros, como las titulizaciones de los CDOs.

Al compararlo con la situación actual, la regulación de los límites a las hipotecas en cuanto al riesgo total que puede asumir una misma entidad, se observan grandes diferencias. Para comenzar, en Estados Unidos, el presidente Barack Obama impulsó y aprobó la conocida como, y todavía vigente, Ley Dodd-Frank<sup>9</sup>. Esta ley, concebida precisamente para evitar los conocidos casos de *too big to fail*, elevó los requisitos de capital de los bancos, así como los ratios de consumo de capital o los coeficientes de los *risk weighted assets* (RWAs), según los cuales, si se conceden financiación a ciertos tipos de activos o industrias, el banco deberá provisionar un porcentaje de capital en función del volumen prestado.

Estas medidas, entrañan que si, por ejemplo, se financiaran proyectos inmobiliarios, dicha financiación, consuma un mayor porcentaje de capital, lo que desincentiva el recurso a ciertos sectores (como el inmobiliario) y productos (como las CDOs), por parte de los bancos<sup>10</sup>. En Europa, la normativa denominada Basilea III introdujo otros límites similares, entre los cuales se pueden encontrar los conocidos como *Common Equity Tier I* o CET1, que forman parte del capital Tier I que se exige que los bancos y entidades financieras mantengan por encima de un nivel respectivo al total del capital, y que se calcula también en referencia a los RWAs mencionados *supra*.

---

<sup>8</sup> Tal y cómo indica el informe 'EY European Bank Lending Forecast Spring', la tasa de morosidad de bancos españoles se situará en torno a los 4,5 puntos porcentuales.

<sup>9</sup> La ley se denomina: "Dodd-Frank Wall Street Reform and Consumer Protection Act", y fue aprobada el 21 de julio de 2010.

<sup>10</sup> En la actualidad se están usando estos mecanismos para favorecer la transición energética y penalizar la financiación a lo que se conoce como *vice o sin stocks*. Entre las consideradas como tales, se encuentran empresas dedicadas al petróleo, al juego, a la industria armamentística, al tabaco, etc. Véase ampliamente Fabozzi, J., & Blitz, D., "Sin Stocks Revisited: Resolving the Sin Stock Anomaly", *The Journal Of Portfolio Management*, 2017.

Así, con este nuevo cambio regulatorio, que no fue fácil de sacar adelante pues encontró mucha resistencia en el ala conservadora, o republicana, del parlamento de los Estados Unidos, se elevaron los ratios de capital desde niveles inferiores al cinco por ciento, a máximos cercanos al veinte por ciento. El sistema se aseguraba con ello, que los bancos dispusieran de suficiente liquidez para poder evitar shocks, como el que sufrió Lehman. Para muestra de los efectos negativos de dichos límites en la rentabilidad de las entidades financieras, un botón. Si en niveles previos a la gran crisis financiera, los bancos obtenían rentabilidades de doble dígito (*low double digit*), en los años posteriores, estos pasaron a obtener unos crecimientos del entorno del dos por ciento anual (Thornton, 2021). Esto afectó mucho al sector bancario español, que comenzó a reflotar su expansión internacional, tras la GCF, y todavía no ha vuelto a los niveles previos a la pandemia.

Estas consecuencias sobre la deuda, también se vieron reflejadas sobre el *equity*. En particular, debido a la crisis financiera, los mercados bursátiles frenaron drásticamente su actividad, dada la enorme volatilidad causada, así como inseguridad sobre el alcance real que podía llegar a tener la crisis.

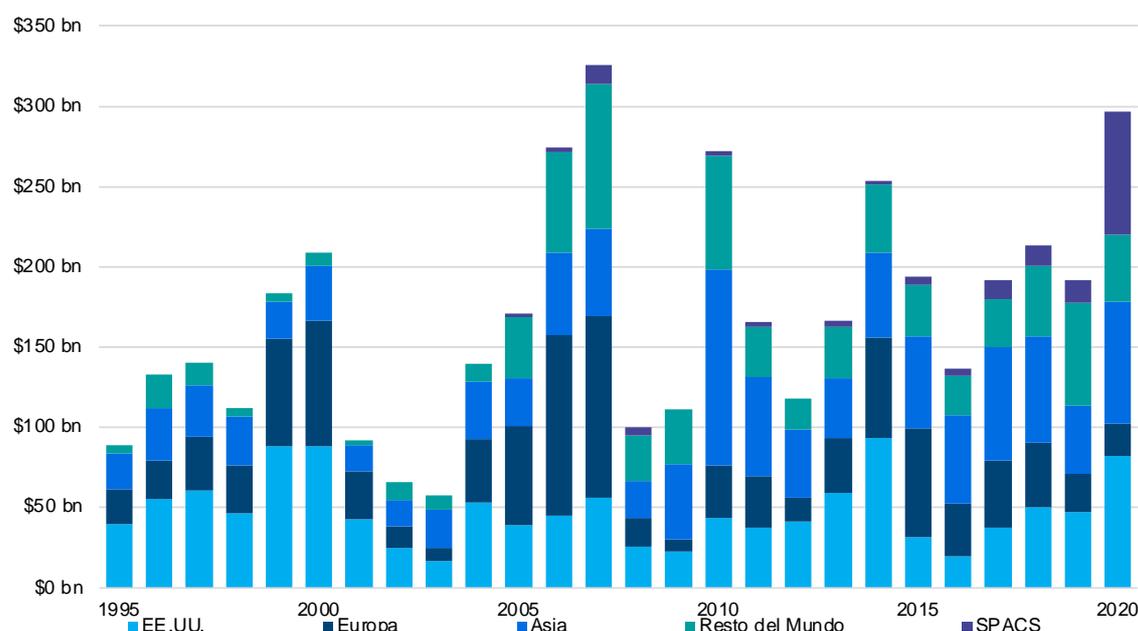


Figura 3. Volumen de las operaciones de salida a bolsa o IPO (inicial public offering por sus siglas en inglés) desde 1995 hasta 2020, segmentados por geolocalización y tipo. Fuente: Refinitiv y Bloomberg Data.

Como se puede apreciar en el gráfico expuesto *supra*, en el año 2007 se marcó un récord de volumen (capital levantado) de operaciones de salidas a bolsa, sobrepasando los

trescientos mil millones de dólares. Además, se aprecia como la caída posterior afectó especialmente a Europa y el resto del mundo, lejos de quedarse como un efecto aislado de los mercados americanos sin consecuencias globales. Una de las razones de esta diferencia se puede encontrar en el hecho de que los mercados financieros de Estado Unidos son muy maduros y disponen de una mayor base de inversores tanto privados como institucionales, mientras que en Europa falta capital y cultura de inversión, que junto con la interdependencia que existe en sus mercados, dan lugar a un sistema bursátil mucho más endeble (Fernandez, 2006).

Además, también afecta mucho la enorme dependencia del sistema financiero europeo ante los bancos. Se estima que el 80% del acceso al capital se realiza a través de entidades financieras, y sólo el 20% restante a través de los mercados bursátiles de *equity* y deuda (Domingos & Dalhuisen, 2020). Esto supone, además, que ante la crisis provocada por la caída de Lehman Brothers los efectos en Europa fueran mayores debido a su alta exposición a los créditos bancarios.

Esto sumado a circunstancias como que el mercado de capitales europeo todavía no atrae tanta inversión ni interés de los inversores institucionales, convierte la bolsa europea en un mercado falto de estabilidad, y que todavía no sirve de mercado de referencia a inversores. Esto implica que cuando las cosas no evolucionan de forma correcta, los inversores se retiran bien a Estados Unidos, bien a Reino Unido, en vez de permanecer, por ejemplo, en Francia, Alemania o España (Dirk Schoenmaker et al., 2018).

En otro orden de cosas, en la introducción de este trabajo se había mencionado uno de los principales elementos diferenciadores entre la deuda y el *equity*: el coste del capital. En esto, como en casi todo, el “precio” es fundamental a la hora de tomar decisiones. Aplicado al objeto de este trabajo, el inversor se fijará por tanto en el precio, a la hora de levantar capital, como uno de los factores que influyan en su decisión; en líneas generales se deberá decidir por si realizarlo a través de mercados de capitales (normalmente más caros, aunque a cambio permiten mayores importes) frente a financiaciones bancarias (más baratas, pero de menor entidad y más exigentes en cuanto a requisitos o *covenants*), y en el primero de los casos, si elegir deuda o capital.

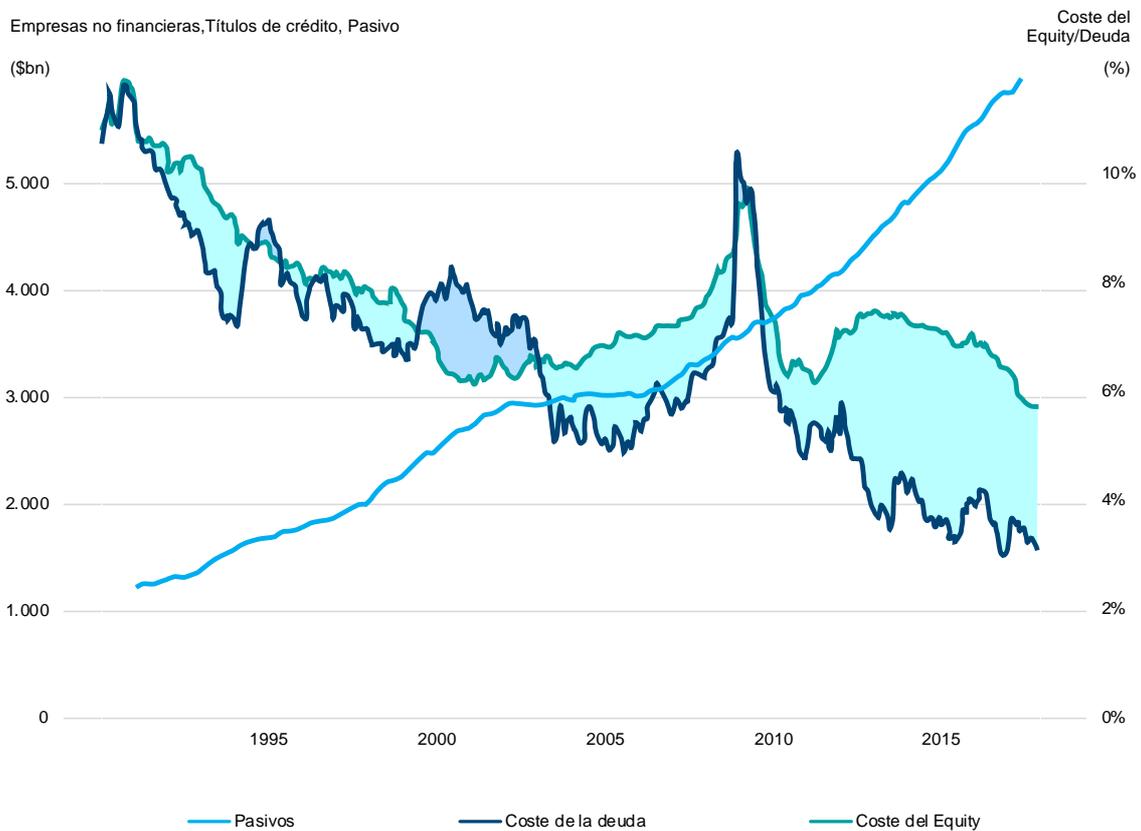


Figura 4. Evolución histórica del coste medio de la deuda y del equity, junto con el incremento del endeudamiento de las empresas no financieras. Fuente: Federal Reserve Economic data.

Tal y cómo se puede comprobar en el gráfico, y en línea con lo que se indicaba al inicio de este trabajo, el coste de la deuda tiende a ser inferior al del *equity*, dado que este último conlleva un mayor riesgo al estar más atrás en el orden de prelación de pagos en caso de concurso de acreedores. Sin embargo, también existen ciertas razones que pueden generar situaciones en las que el coste del *equity* sea inferior al de la deuda. Es el caso de lo ocurrido en los años 1995 y 2000, y suele corresponderse con situaciones en las que la economía sufre *shocks*, y de hecho también se aprecia, dentro del incremento del coste de capital durante la GCF (a mayor riesgo, mayor precio exigido por el inversor) que en cierto momento se llegó a dar dicha situación.

En cualquier caso, si se atiende a los momentos post Lehman, en que el coste de la deuda se aleja enormemente del coste del *equity*, lo que, a su vez, favorece también el incremento del endeudamiento de las empresas no financieras (como se observa en la línea azul clarito del gráfico), es de esperar que el foco general de la sociedad, el particular de la economía, y, más específicamente, el de la investigación en el ámbito de las finanzas

(representado en el articulado de la revista *Journal of Finance*), se centre en la deuda, y por tanto, se aparte del *equity*.

Con lo expuesto hasta ahora, se ha ofrecido una imagen global del contexto en que se enmarca la hipótesis y la ratio de esta investigación: una crisis financiera provocada por el estallido de una burbuja inmobiliaria intoxicada por la falta de exigencia de garantías crediticias a los deudores, que resultó en el colapso de un gigante financiero, como lo era Lehman Brothers, y que consecuentemente arrastró al resto de la economía, y del mundo, a una crisis económica, dónde se alteraron los equilibrios existentes en cuanto a las fuentes de financiación de las compañías (*debt vs. equity*), a la vez que surgieron nuevas reformas que buscaban evitar que una situación así se volviera a repetir.

Si bien puntualmente se continuará profundizando en cada uno de estos ámbitos, seguidamente se procederá a realizar una revisión del estado de la literatura sobre dos aspectos, por un lado, el empleo de técnicas de análisis textual en artículos de investigación, y, por otro lado, el impacto de la crisis financiera en distintos campos de la economía. Todo ello con el objetivo de dotar al lector de un contexto variado que le permita enmarcar mejor este trabajo.

## **REVISIÓN DEL ESTADO DE LA LITERATURA**

En esta sección se comenzará por realizar un estudio de la literatura sobre cuáles han sido, y son, los avances más importantes en el mundo del análisis de textos, así como en un segundo lugar, exponer la literatura que haya tratado el impacto de la caída de Lehman Brothers, con la intención de llevar a cabo un trabajo de síntesis entre ambas líneas de estudio, que puedan enfocar mejor nuestra hipótesis.

En primer lugar, cabe destacar que el análisis de textos ha sido una herramienta muy usada, a la vez que desarrollada, en los últimos años. En una época, como la actual, en la que se busca extraer valor de cualquier tipo de dato que se haya recogido, y en la que abundan las inmensas cantidades de datos con las que se trabaja cada día, el análisis textual juega un papel muy relevante, en especial, por el ingente contenido que se crea cada día en las redes sociales como Twitter, por ejemplo.

Así pues, como punto de partida, es de interés comenzar dicha revisión de la literatura señalando los progresos que, en este campo, se fueron realizando incluso antes de que comenzara el siglo actual. Inicialmente, el análisis de textos fue empleado en el ámbito de la esteganografía (del griego *steganos*, que quiere decir oculto; y *graphos*, que quiere decir escritura). Esta ciencia, que fue de vital importancia durante la Segunda Guerra Mundial (IIGM) pues permitía el intercambio de comunicaciones secretas a través de canales de comunicación inseguros, ha experimentado un avance estratosférico gracias a las herramientas digitales de la actualidad. Entiéndase por análisis de textos, no la mera intención de descifrar mensajes ocultos, o interpretarlos<sup>11</sup>, sino al uso de elementos estadísticos relevantes que sitúen al investigador ante una verdadera estadística de textos (Lebart et al., 1991).

Precisamente, tal y como indican Lebart et al. (1991), el análisis de textos también fue relevante en la “psicobiología de textos”, ciencia que se encarga del estudio del comportamiento humano reflejado en la escritura desde un enfoque biológico (J. C. Eccles, 2006)<sup>12</sup>. Dentro de este ámbito, hay multitud de estudios muy relevantes basados en el *textual analysis*, como, por ejemplo, el llevado a cabo por I-Mei Lin y Erik Peper (2009), en el que estudian patrones psicobiológicos en base a los textos recibidos por mensajes de teléfono.

Siguiendo con la evolución de este tipo de análisis, cabe señalar que si bien se comenzó a estudiar unidades lingüísticas dentro de un texto (Yule, 1944), hoy en día lo que se busca es conseguir que programas informáticos, o la propia inteligencia artificial, comprendan de forma autónoma los significados de los textos. A este campo se le conoce como *natural language processing* (NLP, por sus siglas en inglés) y es muy prolífico en cuanto a literatura. Cabe destacar que, si bien ha sido últimamente cuando más se ha estudiado sobre este ámbito, lo inició tras la IIGM, el matemático Alan (TURING, 1950), que precisamente se hizo famoso por su máquina enigma, capaz de encriptar y desencriptar mensajes secretos.

En los últimos años, ha sido muy usado por distintos autores para estudiar el comportamiento de las redes sociales, y analizar el estado de ánimo de sus usuarios, como por ejemplo en un reciente estudio sobre la red social Twitter en Turquía (Balli et al.,

---

<sup>11</sup> Como podría ser el caso de la hermenéutica aplicada al conocimiento en profundidad de la Biblia.

<sup>12</sup> Véase para mayor contenido y aplicación de la psicobiología el artículo de Clark, M.S. (1991).

2022). Además, las técnicas de análisis de datos con distinto número de variables han demostrado su eficacia en la investigación basada en textos, o *textual analysis*. En particular, mediante la visualización de los datos basados en el conteo de palabras (Lebart et al., 1991). Las herramientas necesarias para llevar a cabo un análisis similar al descrito pueden comparar varios textos, que conjuntamente formen un *corpus*; que será, como se verá más adelante, la forma de análisis a emplear.

El análisis textual ha sido empleado en muchas disciplinas para el estudio de muy diversas hipótesis. Véase, por ejemplo, estudios aplicados en el ámbito de la medicina para detectar el verdadero interés de futuros médicos en el campo de la medicina rural a través del análisis de sus respuestas a preguntas abiertas durante el proceso de admisión (Carol L. Elam et al., 2015), o para descubrir diferencias relevantes en los informes de medicina interna de los pacientes, en función de su género (Osman et al., 2015). En otros ámbitos, cómo las finanzas, esta técnica ha sido muy empleada, debido al potencial que ofrece para sacar rédito de los datos. Por ejemplo, para el análisis de las transcripciones de las presentaciones de resultados para poder saber en qué estado real se encontraban las empresas (Mayew & Venkatachalam, 2012), o para estudiar la información transmitida en la negociación de la compraventa de títulos de deuda del Tesoro, en la Bolsa de Chicago (Coval & Shumway, 2001).

Una vez comprobado el elevado grado de uso y desarrollo que ha tenido la disciplina del análisis textual en la literatura científica, se procederá a comprobar si la caída de Lehman Brothers ha sido y fue tan intensamente objeto de investigación.

Para encontrar una de las primeras referencias (previas al colapso de la burbuja inmobiliaria y del sistema financiero), cabe señalar el informe bancario de (R. K., 2007) que ya advertía de que ciertos inversores no acudían a las operaciones encargadas a los bancos por el incremento del apalancamiento, así como que los CDO's, en particular por las *subprime*, estaban causando pérdidas mayores de las esperadas en las cuentas de resultado de los bancos. Es curioso ver cómo el propio informe habla de que Lehman Brothers había reducido su riesgo ante dichos activos, al haber disminuido su exposición contrarrestando dichas posiciones con estrategias de *hedging*, esto es, de cobertura, a través de derivados financieros.

Es oportuno señalar un estudio llevado a cabo por investigadores británicos, en el que ponen de manifiesto los principales problemas a los que se enfrenta un analista financiero que busca cubrirse, como hizo Lehman, de los CDO's (Cousin & Laurent, 2008). En particular, el trabajo describe tres principales dificultades a la hora de cubrir el riesgo derivado de los CDOs:

- En primer lugar, exponen la cuestión relacionada con la elección y la liquidez de los instrumentos de cobertura. Depende del tipo de instrumento, algunos serán más razonables que otros para cubrir este tipo de activos. Por lo general, se podría pensar en *credit default swaps*, o instrumentos de cobertura por incumplimiento (CDS, por sus siglas en inglés), CDO sintéticos estandarizados de un solo tramo e incluso otros productos como opciones de venta de acciones. Explican, además, que la utilización de acciones, índices u otros productos de renta variable sólo serían aplicables en las partes con mejor solvencia del CDO, y que, sin embargo, no servirían para los estratos inferiores más *subprime*.
- En segundo lugar, mencionan la cuestión relativa a los productos a cubrir, en particular, las distintas formas de CDOs que había en el mercado. La investigación, sugiere evitar cubrirse ante híbridos de tipos de interés o de divisas, las opciones de diferencial de crédito y los derivados exóticos, como los tramos apalancados.
- Por último, los autores hablan de una tercera cuestión que se refiere a la elección del método de cobertura. Así, indican que el enfoque teórico dominante en el ámbito de las finanzas matemáticas favorece la idea de cubrir productos complejos mediante estrategias de cobertura dinámicas basadas en instrumentos subyacentes simples.

Para profundizar más sobre las formas y estructuras de cobertura que se usaron durante esos años existe abundante teoría que aborda el tema, pero que excede el ámbito de este trabajo, como el llevado a cabo por (Giambona & Wang, 2020)

Cabe señalar que el estudio enfocado en la cobertura del riesgo inherente a los CDOs fue llevado a cabo de forma previa a la caída de Lehman Brothers, aunque ya recogía y mencionaba la crisis de crédito que se estaba empezando a gestar por el *crash* de la burbuja inmobiliaria y el consiguiente impago de las hipotecas.

La literatura abarca este tema de conversación desde diversos puntos de vista, partiendo casi siempre de la caída de Lehman, o de forma más reciente, que todo se ha revolucionado desde el punto de vista de analizar el verdadero impacto que supuso la insolvencia de Lehman Brothers en sus acreedores, con porcentajes medios de repago inicialmente estimados en el 25%, que posteriormente han ido incrementándose hasta llegar casi al 50% (de Paoli & Hill, 2022).

Muchos de los análisis se centran en estudiar el impacto de Lehman en otras economías, como por ejemplo la economía de China (Ranjeeni & Sharma, 2015), si bien lo suelen realizar a futuro, centrándose en la caída de Lehman (entendida como suceso abstracto) como catalizador de nuevos procesos. Sin embargo, también se puede encontrar estudios que se enfoquen en el elemento disruptivo que supuso el colapso de Lehman Brothers, analizando el contexto, tanto anterior como posterior (Nhung & Okuda, 2015), en línea con lo que se trata en este trabajo.

Esto permite obtener una imagen mucho más amplia de la discusión objeto de investigación, y a futuro, una mejor perspectiva en cuanto a escalabilidad y solidez de la base de datos.

Tras haber estudiado la literatura disponible, tanto por el lado del análisis textual, como por el del análisis financiero (en particular de la caída de Lehman y la crisis financiera que desató); y tras haber descrito la hipótesis del trabajo (el cambio de paradigma causado por la crisis en el foco de estudio de los investigadores), se procederá a describir la base de datos empleada, así como la metodología implementada en el análisis textual.

## **BASE DE DATOS Y METODOLOGÍA**

### ***BASE DE DATOS***

Para el desarrollo de la investigación objeto de este trabajo se ha construido una base de datos que compilara los artículos de la revista *Journal of Finance*. Se ha seleccionado esta revista por dos razones principales, siendo la primera su ámbito de investigación, es decir, las finanzas, y la segunda, su enorme relevancia en el ámbito académico. Si atendemos a la clasificación del *Journal of Citations Report*, el *Journal of Finance* es el número dos por *impact factor* en el ámbito denominado como *Business, Finance* (si bien ha sido el

primero desde 2012 a 2018), y el número diez en la categoría de *Economics* (aunque su mejor posición la consiguió por primera vez en 2013 como segundo). Estamos por tanto ante uno de los referentes del mundo académico financiero.

Además, cómo se puede comprobar en el gráfico inferior, el Journal of Finance aparte de ser una revista líder en su ámbito de investigación, cada año sigue mejorando sus métricas de relevancia.

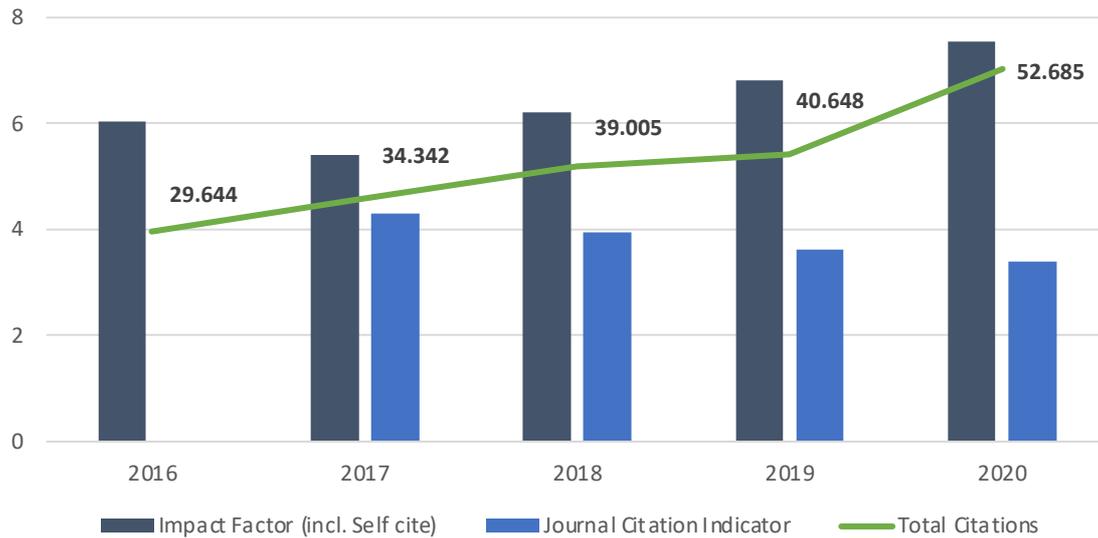


Figura 5. Conjunto de métricas clave para medir la relevancia de un journal, siendo cuanto mayor el Impact Factor mejor; cuando menor el Citation Indicator, mejor; y cuanto mayor el número de citaciones, mejor. Fuente: Journal of Citation Reports, 2020.

Aunque la base de datos se podría haber transferido a un sistema de bases de datos como puede ser MySQL (bases de datos relacionales), se ha preferido mantenerla en Excel para una mejor compatibilidad con el análisis posterior que se llevó a cabo. Este análisis se ha realizado usando distintos paquetes de programación en R, en el *framework* de R Studio, que para la carga y descarga de archivos, se compaginan mejor con archivos de Excel (ya sea en formatos .xlsx o .csv).

Para componer la base de datos, se realizaron distintos procesos de descarga, compilación, y tratamiento de archivos, obteniendo dos sets de archivos como paso previo a la base de datos:

- En primer lugar, se llevó a cabo una descarga manual de cada uno de los artículos que conforman el *journal*, en formato .pdf, para que después fueran

analizados en su conjunto, y además permitir su consulta individual en caso de que fuera necesaria. Tal y como se ha explicado anteriormente, se seleccionó una muestra que va desde el año 2000 (inclusive) hasta el año 2019 (inclusive), es decir, veinte años de estudio de la revista como muestra del presente análisis. Cabe señalar que únicamente se descargaron lo que se considera por la propia editorial de la revista como artículo (de contenido académico), dejando fuera de nuestra muestra documentos tales como, anexos, presentaciones, agradecimientos, etc. El resultado obtenido mostró un total de 1.513 artículos.

- En segundo lugar, una vez obtenidos los documentos de los artículos en formato .pdf, se empleó la herramienta NVivo para la extracción de las palabras que conformaban el texto de cada uno de los artículos. Esta herramienta, que es utilizada en muchos ámbitos como soporte para el análisis posterior de la información (Tokarz, 2020), permitió extraer el texto articulado de la muestra.<sup>13</sup> El *output* de la herramienta consiste en 1.513 archivos en formato .csv que posteriormente fueron consolidados en un solo fichero .xlsx para obtener la base de datos del presente trabajo. A la hora de recabar las palabras de cada artículo, se restringió esta búsqueda a las primeras doscientas cincuenta palabras más repetidas de cada artículo. Junto con la extracción de cada una de las palabras que conformaban el texto, la herramienta permite realizar un ejercicio de *lemmatisation*. Esto se basa en unos diccionarios internos de la NVivo, que permiten agrupar palabras de una misma familia, aunque se escriban de forma distinta, por ejemplo, el caso de los participios de verbos irregulares (ir y fui; ó en inglés, *go* y *went*). Sin embargo, no se debe confundir con otras técnicas como las empleadas en el ámbito del *topic analysis* que permiten ir un paso más allá e identificar temas o *topics*, dentro de una revista, un artículo, un párrafo, etc. Por último, también se empleó lo que se conoce en terminología anglosajona como *stopwords*. Esto consiste en crear una lista con palabras no relevantes para el estudio pretendido, y evitar así que añadan ruido a la muestra. En el caso del presente trabajo, se introdujeron como límites, que las palabras tuvieran más de tres letras, que no

---

<sup>13</sup> La herramienta realiza la extracción de palabras de un texto cada vez, por lo que se empleó un sistema de macros que automatizaran el proceso, cargando la herramienta de NVivo con un Excel que ejecutaba la macro autónomamente y permitió ahorrar una enorme cantidad de tiempo.

fueran números, y que no se trataran de determinantes, pronombres, artículos, meses, días de la semana, años, etc.

Expuesto todo lo anterior, la base de datos que contiene la muestra y objeto de este análisis consta de dos partes, por un lado, la base de datos con los elementos bibliométricos de los artículos tales como su título, su fecha, su resumen, sus autores y sus palabras clave, cinco variables. Sin embargo, el número de variables es fácilmente ampliable con información acerca de la revista, de las citaciones de cada artículo, el género de sus autores, etc. Este tipo de datos se han explotado en muy distintos ámbitos que van desde la enfermería (Cant et al., 2022), hasta el marketing digital (Tosun, 2022), pasando por la educación en tecnologías de la información (Le Thi Thu et al., 2021). En cualquier caso, si bien es verdad que es escalable, también es cierto que cuantas más variables se quieran añadir, junto con más observaciones, se corre el riesgo de que la base de datos pierda cierta manejabilidad a causa de su tamaño.

Por otro lado, contiene el conjunto de palabras extraídas a través de NVivo, clasificadas de la forma que se expondrá a continuación. Dado que el análisis realizado se llevó a cabo artículo por artículo, las palabras están agrupadas en función de la referencia interna otorgada a cada uno de los textos. Aparece el conteo de la palabra dentro del artículo, su longitud, y el porcentaje relativo al total que representa cada palabra en cada artículo; las palabras que conforman el lema (que se presentan bajo una misma palabra de forma agregada), el título del artículo y el año de publicación; en total, ocho variables.

Como ya se ha indicado, por la configuración de esta base de datos, sería muy interesante llevar a cabo una reorganización de la misma, para conformar una única base de datos, en sistemas relacionales. Estos sistemas, entre los que podemos encontrar MongoDB, MySQL, o SQLite, se organizan como conjuntos de tablas, de columnas y filas, independientes, pero conectadas entre sí, a través de los que se conocen como variables clave. De todas formas, en pro de una mejor combinación de trabajo con el lenguaje de programación R, se ha considerado pertinente, mantenerlas en formato Excel.

Descrita la base de datos, se continuará seguidamente con la explicación de la metodología empleada para el análisis de la misma, en particular de las herramientas utilizadas, su descripción, explicación y motivación, así como de las fases en las que se ha dividido este proyecto de investigación.

## METODOLOGÍA

Antes de comenzar, tal y como se ha indicado *supra*, se ha de señalar que para el estudio de esta información se han empleado distintos paquetes de R que han permitido su manejo (*data wrangling*, en inglés), limpieza, y muestra de resultados a través también de paquetes de visualización y herramientas externas como Tableau.

En una primera fase, seguramente la más extensa en tiempo, se llevaron a cabo labores de limpieza de la base de datos, en particular, porque el diccionario de la herramienta NVivo no es extremadamente amplio, por lo que existían todavía ciertas familias de palabras desagregadas, o agregadas en plurales y singulares simultáneamente. Por otro lado, se volvió a realizar una segunda ronda de eliminación de palabras irrelevantes para el objeto de la muestra, principalmente de combinaciones alfanuméricas.

En una segunda fase, se llevó a cabo una agregación de variables numéricas cuantitativas tomando como variable independiente la propia palabra o término. Esta agrupación se llevó a cabo también por año, lo que permitió comparar el resultado absoluto en general, junto con el respectivo total por cada año, en particular.

A partir de aquí, se llevó a cabo una “tercera” fase, si cabe, más de creatividad, para extraer los resultados del análisis. Por la forma en que está presentado este trabajo, como contraposición de dos períodos divididos por los efectos de la caída de Lehman Brothers, con el año 2009 como punto de inflexión, se llevó a cabo una partición de la base de datos desde el 2000 hasta el 2008 (inclusive), y desde el 2009 al 2019 (inclusive ambos). Ambas particiones, o *sets*, representarán el período pre y post Lehman, respectivamente.

Una vez separados los *datasets*, el análisis puro se realizará a través de dos vías, la primera, estudiando los distintos términos a través de su mero conteo, esto es, de la suma agregada de su número de apariciones; y, por otro lado, a través de la métrica que se conoce como TF-IDF (o en inglés, *term frequency - inverse document frequency*).

Esta medida se presenta como un cociente, en la que, por un lado, en el numerador, se dispone el número de veces que aparece el término en el conjunto de la base de datos, y, por otro lado, en el denominador, se observa la frecuencia con la que aparece en un documento distinto, esto es, el número de documentos que contienen el término dividido

por el número total de documentos. Así, la TF-IDF ofrece una prima cuantas más veces sea repetida la palabra, a la vez que una penalización en función de lo “común” que sea. Esta metodología ha sido empleada con multitud de fines, entre ellos la clasificación de textos literarios (Lin Xiang, 2022), u otros más peculiares como, el *sentiment analysis* de los diputados británicos en base a sus *tweets* (Mee et al., 2021), o la recomendación de recetas de cocina (Chhipa Shubham et al., 2022).

De cara al presente estudio, en particular, para profundizar todo lo posible en la metodología, en aras de favorecer la replicación de este trabajo aplicado a otras materias, es relevante señalar que hace falta la creación de un *corpus*. El *corpus* es un conjunto de largos textos o *sets*, que se conforma a partir de un subconjunto de muestras, o documentos (en este caso los artículos de la revista), que permiten posteriormente, calcular la parte del IDF.

Una vez obtenido el corpus, se ha de montar la matriz que contenga todas las apariciones de cada palabra en cada documento. Esta matriz es conocida por su denominación inglesa como DTM (*document term matrix*). Tanto la matriz, DTM, como el *corpus*, se han obtenido usando paquetes específicos de R, en particular, el conocido paquete “*quanteda*”.

El paquete de *quanteda*, que proviene de las iniciales de *quantitative analysis of textual data*, se encuentra provisto de numerosas herramientas muy útiles para el análisis de textos. Fue desarrollado por Kenneth Benoit, Kohei Watanabe, entre otros, y se ha convertido en uno de los principales recursos en este ámbito (Benoit et al., 2018).

Aun así, el paquete es un poco rígido en cuanto a variedad de inputs por lo que se tuvo que realizar un trabajo previo de preparación de los datos. Dada nuestra base de datos formada con las variables expuestas *supra*, se realizó una agregación de datos mediante el uso de tablas dinámicas para obtener algo parecido a la *term document matrix* (TDM), con los documentos por columnas, y las palabras por filas, que posteriormente se transpone para llegar a la DTM, o *document term matrix*, con los documentos por filas, y las palabras por columnas.

Posteriormente, ese fichero se transformaba coercitivamente en un archivo de clase *dataframe* para su posterior empleo, y al mismo tiempo, se buscaba realizar un conjunto de bucles *for* con sentencias lógicas *if* para conseguir un archivo en formato csv.

compatible con la carga de ficheros en la función “corpus”. A partir de aquí, el proceso es relativamente sencillo, pues el paquete de *quanteda* ofrece todas las funciones y herramientas necesarias para analizar los textos, en este caso, desde la perspectiva que ofrece el TF-IDF.

En cualquier caso, debido a la enorme cantidad de observaciones contenidas en el *corpus* los tiempos de análisis de los datos se extendieron de manera relevante. Específicamente, se usó la función “dfm” que, a partir del *corpus*, genera un fichero del tipo *document-feature matrix*, de ahí el nombre de la función. Ese fichero, será el que permita mostrar con la función “dfm\_tfidf”, los valores otorgados a cada palabra<sup>14</sup>.

A modo resumen, cabe indicar que se han obtenido en total 18.544 palabras únicas, con una repetición media por artículo de 24,1 palabras; apareciendo 376.929 palabras únicas por documento, que en total suman 9.037.454 apariciones contenidas en la muestra.

Si se atiende a los *datasets* intermedios, en el período previo a la caída de Lehman Brothers se obtiene una repetición media por artículo de 21,8 palabras apareciendo 181.229 palabras únicas por documento, que en total suman 3.935.422 apariciones contenidas en el *subset*. En el caso del período post caída de Lehman Brothers el estudio muestra una repetición media por artículo de 26,2 palabras apareciendo 195.700 palabras únicas por documento, que en total suman 5.102.032 apariciones contenidas en el *subset*.

Por último, en cuanto a la metodología también se ha aprovechado el conteo de las propias palabras para mostrar un *tracking* de su evolución a lo largo del período de la muestra, para el cual se han empleado paquetes de visualización de datos programados en R. En este sentido, el potencial de visualización de las herramientas diseñadas en R, es superior, al menos en cuanto al grado de personalización, que lo que pueden ofrecer otros sistemas como Python.

---

<sup>14</sup> A título explicativo, cabe mencionar los argumentos utilizados para la obtención de dichos valores. Específicamente son dos los *inputs* que se requieren para ajustar el algoritmo, *scheme\_tf* y *scheme\_df*. La primera de ellas se encarga de seleccionar la forma de agregación del numerador, que se conoce como *dfm\_weight*, y que en este trabajo se ha decidido que se lleve a cabo mediante conteo. El segundo ajuste, si el primero modifica la prima que se otorga a cada palabra, se encarga de la penalización, es decir, ajusta como se ha de tratar la frecuencia en cada documento, habiéndose elegido en este caso que afecte de forma inversamente proporcional.

## RESULTADOS

Previamente a comenzar la presentación de los resultados del trabajo realizado, se ha de señalar que se ha buscado exponer los resultados mediante el uso de herramientas gráficas como Tableau, así como de paquete gráficos de R, cómo *ggplot*.

Tableau es una herramienta de visualización de datos que permite, entre otras cosas la visualización dinámica de bases de datos, para la interpretación de patrones y tendencias en estos (Batt et al., 2020). La posibilidad de calcular nuevas variables que representar en los gráficos, así como los sistemas de filtrado especializados que ofrece la herramienta, han sido muy útiles para la realización de este análisis.

Por su lado, el paquete *ggplot* de R, es seguramente uno de los más usados, por el alto grado de personalización que permite. En este sentido, cabe señalar que, debido a su flexibilidad, se han desarrollado multitud de herramientas y materiales que se basan en *ggplot* para generar una gran variedad de gráficos como pueden ser los gráficos de redes o *network diagrams* (Tyner et al., 2017), o nuevos tipos de gráfico cómo los anatogramas para la visualización y estudio del cuerpo humano (Maag, 2018), e incluso el diseño de aplicaciones web a través de la herramienta *Shiny* (Mouksassi & Pastoor, 2015).

Dada la importancia de presentar los resultados de forma directa pero interpretable, sólida pero flexible, y consistente pero también llamativa, y dentro de los límites que impone el formato de este trabajo, se ha optado por llevar a cabo la representación de los resultados a través de las dos herramientas ya mencionadas y explicadas.

Para comenzar, se mostrará las palabras más usadas en los veinte años que componen la muestra. Posteriormente, se tratarán las diferencias entre los dos términos objeto de debate cómo son el *equity* y el *debt*. Y para finalizar, se expondrán los resultados obtenidos a la hora de medir el impacto de la caída de Lehman en dichos términos.

## Top 10 Palabras más Utilizadas

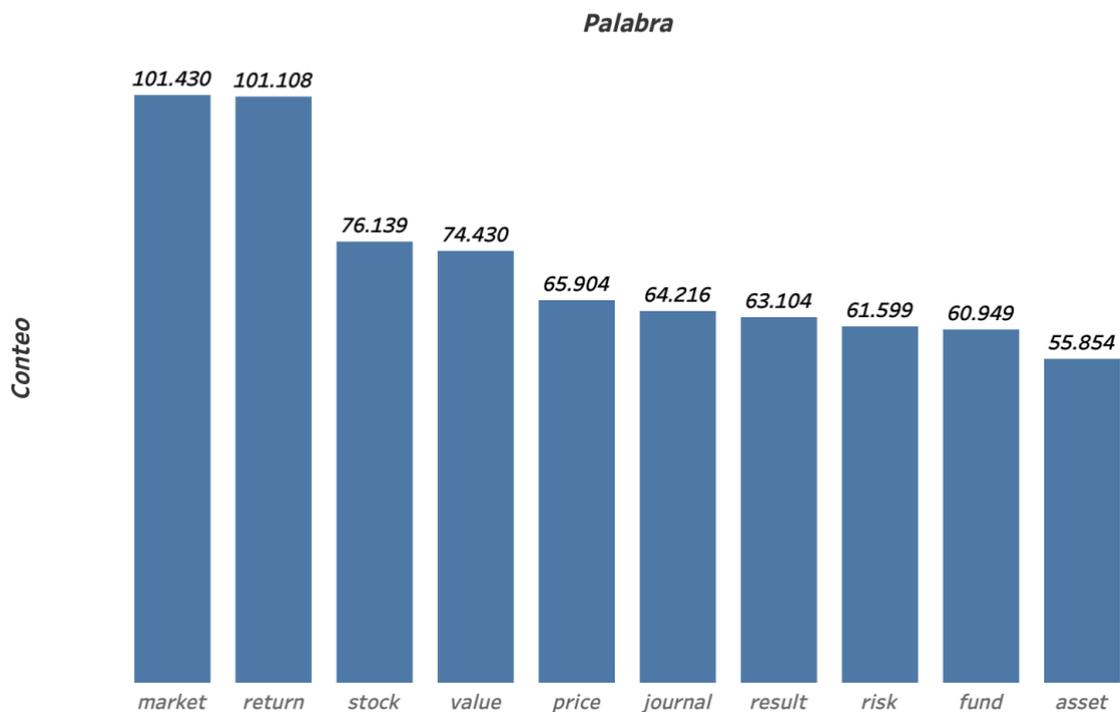


Figura 6. Top 10 palabras más utilizadas por los investigadores de *The Journal of Finance* a lo largo de los veinte años de muestra.

En primer lugar, tal y como se puede observar, las dos palabras más utilizadas a lo largo de la muestra han sido *market* y *return*, además con bastante diferencia respecto al resto. Sin embargo, aquí entra en juego la métrica que se expuso *supra*, la TF-IDF, ya que en caso de que aparecieran en muchos de los artículos analizados, la palabra pasaría a ser considerada como una “*commodity*” en el sentido de que no aporta valor adicional alguno. Entre las cinco palabras más repetidas se podría realizar una asociación entre *market* y *value*, así como entre el *return* y el *price*, y después los cuatro términos juntos, si eso juntándolo con *stock*, que sería un poco el *tertium genus*. Aunque no aparezcan entre los que más se repiten a lo largo de la muestra, no quiere decir que no estén las palabras de gran relevancia como por ejemplo *loan*, *equity* o *debt*.

Sin necesidad de acudir a ella de momento, y solamente analizando los datos en base a estadísticos conocidos por todos como la media, se observa como ambas palabras se sitúan en dos planos distintos. Por un lado, *market* aparece en 1.399 documentos, lo que supone un 92% del total, con una media de 72,55 repeticiones en cada documento en el que aparece. Por otro lado, *return* se menciona en 1.056 artículos (70% del total), pero

con una media mayor de apariciones de 95,82 por artículo. Como se puede observar, es estadísticamente más relevante la palabra *return*, que la palabra *market*, pues lo que indican los datos es que no se suele hablar específicamente de los mercados, y, sin embargo, sí que existe un mayor foco por las rentabilidades.

Aplicando esta medida a las dos palabras contrapuestas, objeto de este estudio, *equity* y *debt*, se puede observar lo siguiente:

- *Equity*: aparece en un total de 32 mil veces, en 848 artículos, con una media de 39,1 veces repetidas por artículo.
- *Debt*: aparece en un total de 29 mil veces, en 443 artículos, con una media de 67,0 veces repetidas por artículo.

Por tanto, se constata que, al igual que con el par anterior, en este caso, *debt*, es mucho más relevante que *equity*, o, al menos, supone un tema de estudio central en los artículos en los que aparece. Posteriormente, con el estudio de los valores de TF-IDF se tratará de confirmar esta presunción. Aunque cómo ya se ha explicado, la parte fundamental no es tanto la función de agregación de las funciones usadas para los Excel, si no el valor que se le dé a la penalización por el número de documentos que contengan la palabra.

## Impacto de la caída de Lehman Brothers

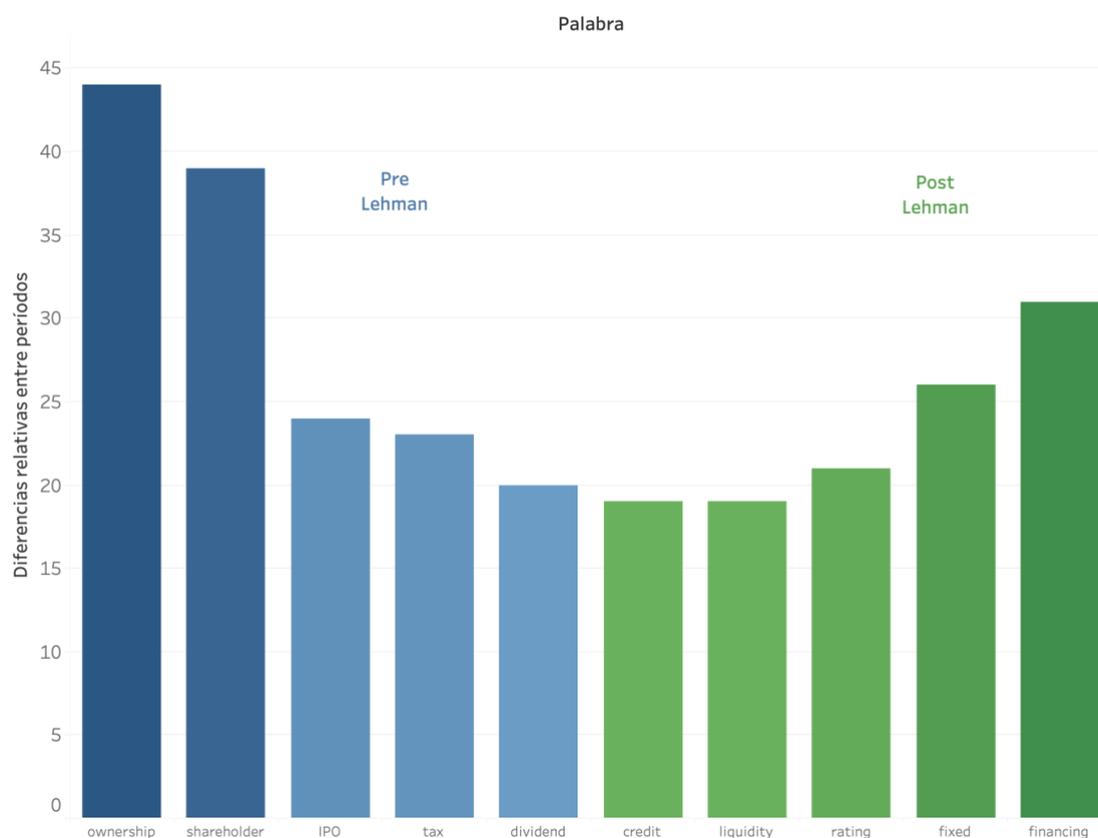


Figura 7. Diferencias entre las posiciones relativas de las 100 palabras más repetidas por conteo entre períodos. Léase cómo que “financing” aparece post Lehman treinta puestos por encima de su posición pre Lehman en la clasificación de palabras más repetidas.

En el gráfico expuesto sobre estas líneas, se pretende mostrar cómo han variado las posiciones relativas de cada una de las cien palabras más repetidas en cada uno de los períodos. Para ello, se calcularon las cien palabras con más observaciones en cada período, y se ordenaron, en función de tal conteo, en una clasificación a la que se mencionará aquí en adelante como *ranking*. Pues bien, en dicha imagen se puede apreciar de forma clara las diez palabras que más han modificado su posición en el *ranking*. Esto indica, por ejemplo, que la palabra *dividend* ocupaba una posición relativa veinte veces superior en el *ranking* pre Lehman, esto es, veinte puestos por encima, en comparación con la etapa post Lehman.

Es manifiesto como tras la caída de Lehman Brothers, en verde, las palabras que aumentan su posición en el ranking con más fuerza son aquellas relacionadas con el término *debt*, cómo pueden ser *credit*, *liquidity*, *rating*, *fixed* (asumiendo que es en

referencia al tipo de interés) y *financing*. Y al mismo tiempo, entre las palabras que más bajan en el ranking respecto a su situación pre Lehman, se encuentren *ownership*, *shareholder*, *IPO* y *dividend*, todas ellas relacionadas con el *equity*, o la parte de recursos propios de una empresa.

Además, si se atiende al estudio particular de los propios términos en sí, se puede observar que *debt*, aumenta trece posiciones respecto a su situación en el *ranking* pre Lehman; mientras que *equity*, reduce doce posiciones. Esto da una buena imagen, no sólo de los cambios en cuanto a la investigación alrededor de estos dos términos, sino también al hecho de que el cambio que se produce es sustitutivo, pues ocurre en igual forma e intensidad, e incluso, si se equipararan en significado el resto de los términos que aparecen en el gráfico, también en contenido.

Por otro lado, de enfocarse específicamente en los resultados obtenidos tras realizar la ponderación que entraña la herramienta de TF-IDF, la matriz resultante aporta unos valores que buscan representar la importancia de la palabra, directamente proporcional a su presencia en el conjunto de textos (cuantas más observaciones mejor), e inversamente proporcional a su frecuencia en los documentos (en cuantos más documentos aparezca peor).

En este sentido, las palabras más relevantes según el factor generado por el TF-IDF, varían respecto de las señaladas previamente en función de su conteo.

Top 15 Palabras más Valoradas por TF-IDF

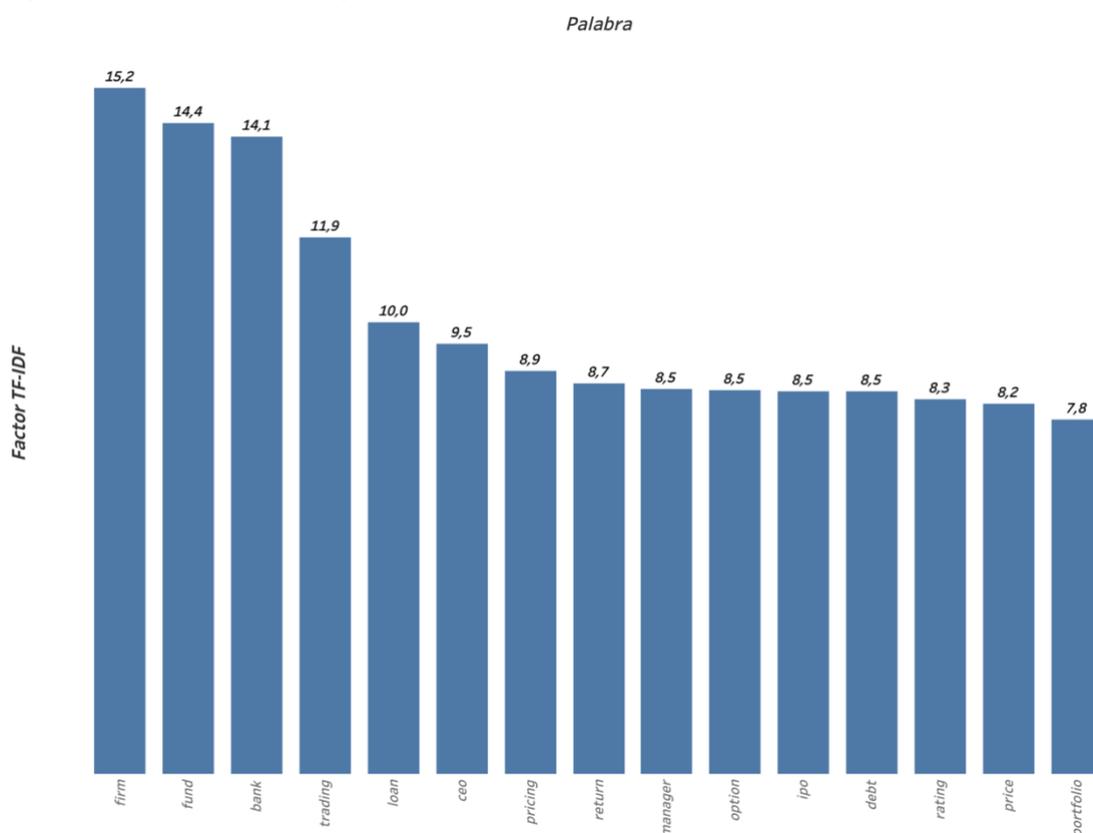


Figura 8. Top 15 palabras con mayor relevancia según TF-IDF a lo largo de los veinte años de muestra.

A la hora de comparar los resultados obtenidos por conteo, frente a los obtenidos en términos de valoración a través del TF-IDF, si bien ambas variables están relacionadas (el numerador del TF-IDF es precisamente el conteo de dicha palabra), se pueden vislumbrar numerosos cambios entre las primeras filas. Así, se puede observar cómo términos de alta relevancia financiera como *risk*, *stock*, o *value*, desaparecen. Esto claramente se debe al hecho de que estas palabras se encuentran fácilmente en la mayoría de los artículos publicados por The Journal of Finance, por lo que la penalización que les conlleva el IDF (*inverse document frequency*) es muy alta. Además, también se confirma el análisis preliminar realizado en base a la media de palabras por artículo, y porcentaje de artículos en los que aparecían, respecto de los pares, *market* y *return*; y del par objeto de este trabajo que contrapone al *equity* frente a la *debt*. Ni la palabra *market*, ni la palabra *equity* aparecen entre las quince primeras palabras por valoración TF-IDF, mientras que sus pares, que son *return* y *debt*, aparecen en la posición octava y décimo segunda, respectivamente.

## Impacto de la caída de Lehman Brothers

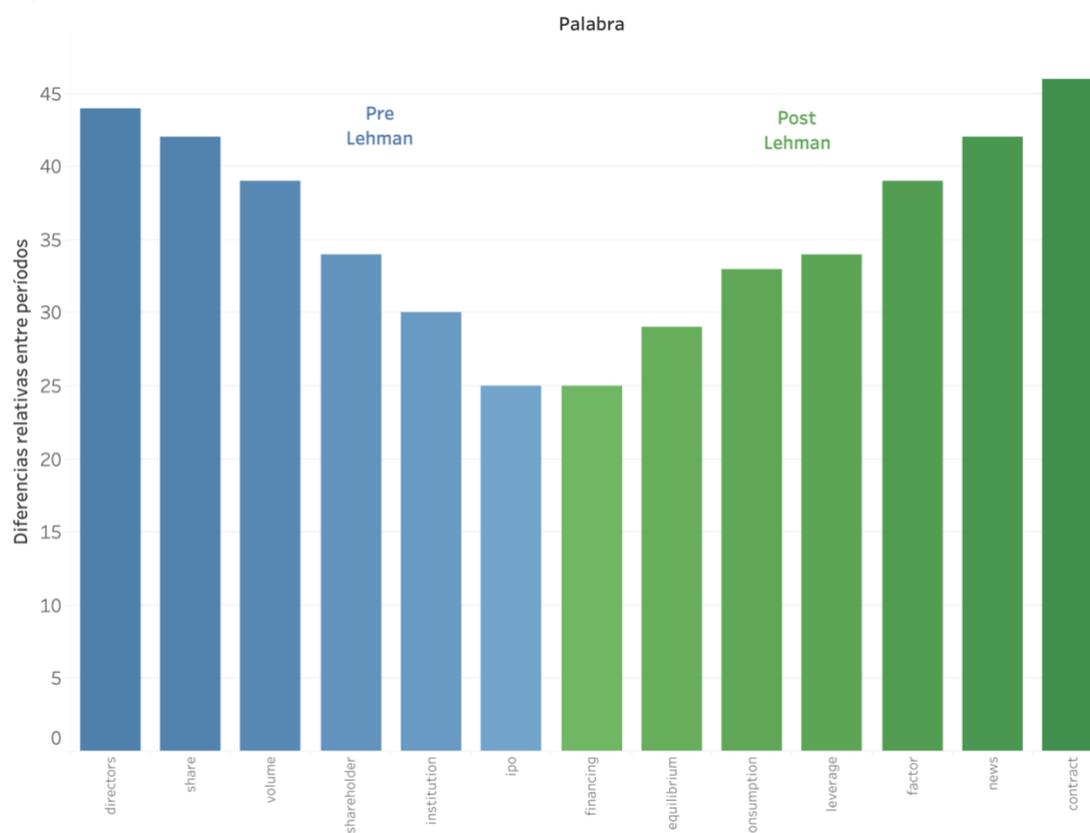


Figura 9. Diferencias entre las posiciones relativas de las 100 palabras con mayor factor TF-IDF entre periodos. Léase cómo que “financing” aparece post Lehman veinticinco puestos por encima de su posición pre Lehman en la clasificación de palabras más valoradas.

Obsérvese en el gráfico superior cómo las palabras que más mejoran o empeoran su posición en el *ranking*, entre el período pre Lehman y el período post Lehman son ligeramente diferentes que las que reflejaba el conteo. Aun así, se puede seguir identificando al menos dos términos, de entre los que más relevancia cobran tras la caída de Lehman Brothers, relacionados con el término *debt* como son *financing* y *leverage*. En la otra cara de la moneda, entre los términos que más relevancia pierden, están precisamente tres relacionados con el *equity*, y son *share*, *shareholder* y *IPO*<sup>15</sup>.

A continuación, también se estudiará la evolución en términos absolutos de las palabras que se han obtenido con el análisis efectuado *supra*. De cara al estudio de la evolución de

<sup>15</sup> Cómo se puede ver en este último ejemplo, el *lemmatisation* de NVivo no es perfecto, pues agrupó bajo dos familias distintas la palabra *share* (formada por *share*, *shares*, *sharing*), y *shareholder* (formada por *sharehold*, *shareholder*, *shareholding*, y sus respectivos plurales).

estas palabras relacionadas con cada una de las partes de los recursos de una empresa, ya sean propios o ajenos, se han seleccionado cuatro palabras que guardan estrecha relación con cada uno de los términos en cuestión.

Cabe señalar, que se han elegido de entre las palabras obtenidas en el análisis comparativo de las etapas pre y post Lehman, precisamente como muestra de su relevancia. Así, en relación con el término *debt*, se han seleccionado las siguientes palabras: *financing*, *rating*, *liquidity* y *credit*; mientras que en referencia al término *equity* se eligieron: *shareholder*, *ownership*, *IPO* y *dividend*.

El análisis efectuado es por tanto un *tracking* de la evolución del conteo registrado de estas palabras durante el período de estudio de la muestra (2000-2019). Este *tracking* se ha llevado a cabo mediante el uso de tablas dinámicas en Excel para la parte cuantitativa, mientras que, para la parte cualitativa, se decidió por representarlas a través de unos gráficos de densidad que muestran de forma muy clara las distribuciones del conteo a lo largo de los años. Para el desarrollo de estos gráficos se ha empleado el paquete de R de visualización de datos *ggplot*, ampliamente conocido por la literatura, y muy versátil a la hora de realizar gráficos con diseños nuevos. En particular, la función o herramienta usada se corresponde con la de *geom density ridges*.

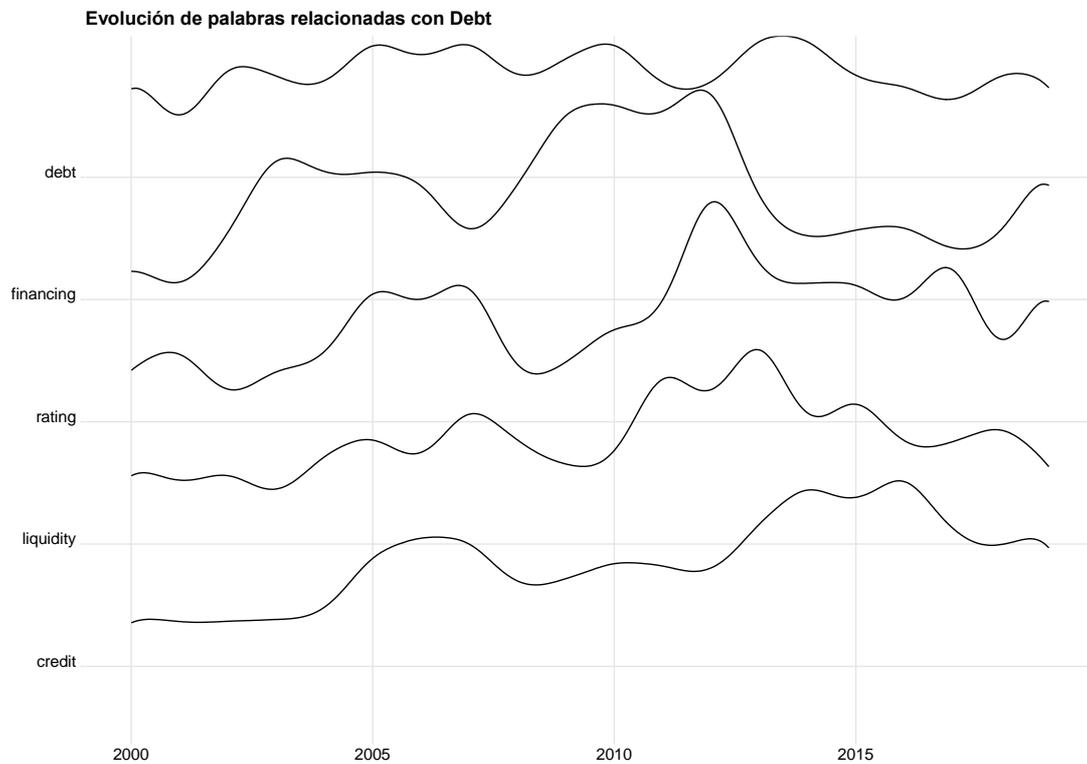


Figura 10. Tracking de la evolución de ciertas palabras relacionadas con el término "Debt" a lo largo del período de la muestra, 2000-2019. Representado mediante gráficos de densidad.

Tanto el gráfico superior, como el gráfico inferior a estas líneas, confirman la hipótesis de partida: a raíz de la caída de Lehman Brothers y el estallido de la GCF en 2008, el término deuda y todos los relacionados con el mismo, fueron sustituyendo gradualmente al término *equity* y sus similares. Además, al mismo tiempo, siguen confirmando la misma línea trazada por los análisis realizados de forma previa, que mostraban estos resultados. Mientras que, por un lado, los términos más estrechamente relacionados con *debt* incrementan sus observaciones a partir de la caída de Lehman, especialmente fuerte hasta el año 2012, las palabras más relacionadas con el término *equity*, se ven especialmente castigadas a partir de la caída del mercado, con algún pequeño amago de rebote en los últimos años.

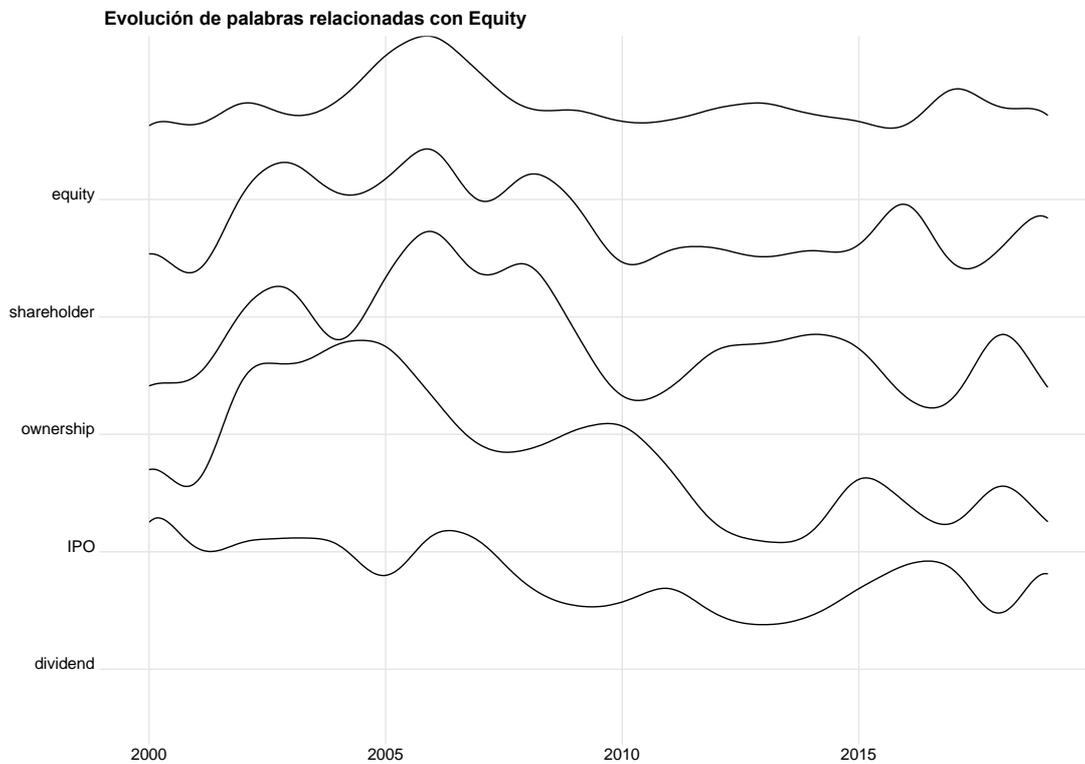


Figura 11. Tracking de la evolución de ciertas palabras relacionadas con el término "Equity" a lo largo del período de la muestra, 2000-2019. Representado mediante gráficos de densidad.

Como resultado adicional, si bien no guarda relación con la hipótesis de este trabajo, es reseñable la relación inversa que mantienen las palabras IPO y *ownership*. En el gráfico se puede observar claramente como en la gran mayoría de los años, aunque no en todos, cuando aumenta la visibilidad de una de las palabras, la de la otra disminuye y viceversa. Se aprecia muy bien en los años 2012-2013, cuándo parece que la investigación en IPOs, o el interés por los mismos, está en niveles mínimos en los ámbitos de mercado de capitales, mientras que, para la contraparte, se da comienzo a la “recuperación” de la crisis financiera.

Si bien se expondrá más adelante en las conclusiones, para rematar el apartado de resultados, es oportuno considerar confirmada la hipótesis inicial, acerca del impacto de la caída de Lehman Brothers (y, por supuesto, las crisis que lo acompañaron) en el cambio de foco experimentado en las publicaciones del Journal of Finance, desde una perspectiva más basada en el *equity*, a otra más apoyada en la deuda.

## CONCLUSIONES

Previamente a realizar conclusiones taxativas sobre el resultado y propósito de este trabajo, se va a intentar ofrecer una imagen consolidada que permita al lector recordar a través de un pequeño resumen, la línea de investigación que se ha seguido a lo largo del texto.

Al inicio del documento, en la introducción, se puso en contexto el objeto de este trabajo, con la exposición de la hipótesis que se iba a resolver, las contraposiciones que se iban a dar, no sólo entre *debt* y *equity*, sino también la de los períodos pre y post Lehman, y los principales temas que se iban a tratar.

Entre esos temas a tratar, cabría realizar una separación en dos principales ramas de conocimiento o interés investigador. Por un lado, la parte más económico-financiera del trabajo, en particular el foco situado sobre Lehman Brothers, su caída y desaparición; junto con el resto de las crisis que ha habido, y las causas y efectos de las mismas. Por otro lado, la parte más de análisis, específicamente, el análisis textual, junto con la visualización de datos que permitieran ofrecer la mejor forma de presentar los resultados de dicho análisis.

La revisión de la literatura ha tornado ser muy beneficiosa, sobre todo a la hora de poner en contexto la hipótesis de este trabajo. Al fin y al cabo, comprobar que ya se habían realizado análisis textuales, bibliométricos, y con el uso de herramientas como NVivo otorga un plus de consistencia al presente documento. Al mismo tiempo, la extensa literatura acerca de la caída de Lehman Brothers, o las distintas crisis que se fueron sucediendo, permite focalizarse aún más en las razones que pudieran explicar por qué se llegó a tal situación y qué se puede hacer para evitarlo de nuevo.

La descripción, tanto de la metodología, como de la base de datos, se ha buscado mantener lo más breve posible, a la vez que no dejar nada sin explicar ni justificar. El formato de la base de datos, así como la idea sobre la que se asienta, es muy eficiente, pues se ha estructurado de tal forma que sea escalable. Así, a la hora de actualizarla según vayan saliendo nuevos artículos, o si se quiere añadir nuevas revistas al análisis, el proceso sea fácil, sencillo y rápido.

Para finalmente llegar a los resultados dónde se ha podido ofrecer una reconciliación de las dos ramas, financiera y analítica, motivadas por la confirmación de la hipótesis inicial, de que tanto el *equity* como el *debt* se vieron perjudicados, si bien en términos relativos, el segundo pasó a reemplazar en su puesto al primero.

En vista de dichos resultados obtenidos, se puede concluir que efectivamente, la caída de Lehman Brothers, junto con el resto de los factores estudiados, ha implicado un cambio de foco en el mundo académico (que seguramente refleje otro de igual magnitud en el mundo económico en general), en el que al menos hasta la pandemia del Covid-19, se ha primado el estudio de la deuda frente al *equity*, sugiriendo, tal vez, que actualmente esté considerada como la forma de financiación preferente.

El análisis pormenorizado y detallado de este cambio de paradigma, seguramente requiera de explicaciones muy variadas y complejas, que irán desde los sesgos del comportamiento a los que esté sometida la mesa editorial de la revista, hasta los intereses externos de aquellas personas que la financien. Todo esto son métricas y suposiciones que se exceden del ámbito de estudio de un trabajo de fin de grado, y además, conllevan un alto componente subjetivo y personal que difícilmente será analizable desde fuera de la organización.

En cualquier caso, este trabajo ha sido capaz de confirmar el cambio de paradigma, desde un enfoque centrado en *equity* a uno centrado en deuda; en una de las revistas más importantes a nivel mundial en el ámbito financiero, siendo incluso atrevido a la hora de indicar las razones de tal cambio (los *drivers*, que se diría en inglés); y confirmando la caída de Lehman Brothers junto con el resto de las crisis subyacentes, como punto de inflexión de tal cambio.

Ya que el mercado siempre está en movimiento, y que este tiene repercusiones directas en el ámbito académico tal y como se acaba de exponer, dada la situación actual, habrá que esperar a ver cómo se traduce la crisis causada por la pandemia en el ámbito social, laboral y mercantil, para poder estimar el impacto que tendrá sobre el resto de los artículos de la revista.

Dentro de esta línea de investigación, aparte de poder aplicar este tipo de análisis a otras revistas, o enfocados, en otros términos, hay múltiples análisis a realizar siendo un esfuerzo marginal cercano a cero, salvo la recopilación de los documentos pues toda la

parte de análisis está programada y únicamente requiere de ciertos ajustes en función de las particularidades de cada caso.

Aparte, sobre el tema elegido como objeto de este trabajo, sería interesante investigar posibles razones que justificaran los cambios observados, tal vez adentrándose en el ámbito del *behavioral finance*, e incluso analizar posibles sesgos, o si existieran relaciones directamente proporcionales con otras revistas en cuanto a los términos más usados. Esto serviría también para comparar líneas editoriales, y sería aplicable, por ejemplo, a periódicos y editoriales, de los cuales se pudieran extraer connotaciones políticas y sociales.

Por último, y de cara a concluir este trabajo, se ha de señalar que un claro ejemplo de línea futura de investigación pasaría por añadir, a este estudio, las citas de cada artículo, así como el género de cada autor. Por un lado, se busca entender qué artículos han sido más relevantes para según que tipos de audiencia (los portales específicos requieren vínculo académico para poder citar, mientras que otras herramientas como Google Scholar no), y en función de eso, intentar o bien predecir futuras líneas de investigación, o bien adaptarse a ellas. Y, en segundo lugar, es muy interesante poder analizar los sesgos de género que afectan a las mujeres y hombres investigadores, para también visibilizar más el rol de la mujer investigadora.

## BIBLIOGRAFÍA

- Baker, D., & Diaz, M. M. (2004). Bubble Bath: Will a Crashing Housing Market Sink Low-Income Homebuyers? *Journal of Housing & Community Development*, 61(5), 6.
- Balli, C., Guzel, M. S., Bostanci, E., & Mishra, A. (2022). Sentimental Analysis of Twitter Users from Turkish Content with Natural Language Processing. *Computational Intelligence and Neuroscience*, 2022, 2455160.
- Batt, S., Grealis, T., Harmon, O., & Tomolonis, P. (2020). Learning Tableau: A Data Visualization Tool. *Journal of Economic Education*, 51(3-4), 317.
- Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). quanteda: An R package for the quantitative analysis of textual data.
- Cant, R., Ryan, C., & Kardong-Edgren, S. (2022). Virtual simulation studies in nursing education: A bibliometric analysis of the top 100 cited studies, 2021. *Nurse Education Today*, 114
- Carol L. Elam, Anthony D. Weaver, Elmer T. Whittler, Terry D. Stratton, Linda M. Asher, Kimberly L. Scott, & Emery A. Wilson. (2015). Discerning applicants' interests in rural medicine: a textual analysis of admission essays. *Medical Education Online*, 20(0), 1.
- Chhipa Shubham, Berwal Vishal, Hirapure Tushar, & Banerjee Soumi. (2022). Recipe Recommendation System Using TF-IDF. *ITM Web of Conferences*, 44, 02006.
- Cousin, A., & Laurent, J. (2008). Hedging issues for CDOs.

Coval, J. D., & Shumway, T. (2001). Is Sound Just Noise? *Journal of Finance* (Wiley-Blackwell), 56(5), 1887.

de Paoli, L., & Hill, J. (2022). The Very Last of Lehman Brothers. *Bloomberg Businessweek*, (4742), 32.

Dirk Schoenmaker, Zsolt Darvas, Rotterdam School of Management (RSM), Erasmus University, Erasmus Research Institute of Management, & Centre for Economic Policy Research, London. (2018). Institutional Investors and Development of Europe's Capital Markets.

Domingos, P. I. P., & Dalhuisen, J. H. (2020). Why the European market depend so much on the banking system when compared to capital markets? A comparison with the United States reality.

Fernandez, V. (2006). Extremal Dependence In European Capital Markets. *Journal of Applied Economics*, 9(2), 275.

Giambona, E., & Wang, Y. (2020). Derivatives Supply and Corporate Hedging: Evidence from the Safe Harbor Reform of 2005. *Review of Financial Studies*, 33(11), 5015.

Griffin, J. M., & tang, D. Y. (2012). Did Subjectivity Play a Role in CDO Credit Ratings? *The Journal of Finance*, 67(4), 1293-1328.

<http://www.jstor.org/stable/23261360>

J. C. Eccles. (2006). New Text Asserts Psychobiology as an Independent Science The Psychobiology of Mind William R. Uttal. *Bioscience*, 29, 482.

- Le Thi Thu, H., Tran, T., Trinh Thi Phuong, T., Le Thi Tuyet, T., Le Huy, H., & Vu Thi, T. (2021). Two Decades of STEM Education Research in Middle School: A Bibliometrics Analysis in Scopus Database (2000-2020). *Education Sciences, 11*
- Lebart, L., Salem, A., & Berry, L. (1991). Recent developments in the statistical processing of textual data. *Applied Stochastic Models and Data Analysis, 7*(1), 47.
- Levinson, M. (2009). The Economic Collapse. *Dissent (00123846), 56*(1), 61.
- Lin Xiang. (2022). Application of an Improved TF-IDF Method in Literary Text Classification. *Advances in Multimedia, 2022*
- Maag, J. L. V. (2018). Gganatogram: An R package for modular visualisation of anagrams and tissues based on ggplot2 [version 2; peer review: 2 approved]. *F1000Research, 7*
- Mayew, W. J., & Venkatachalam, M. (2012). 3: Research on Nonverbal Communication in Accounting and Finance: 3.1 Why Should Managers Exhibit Affective States During Earnings Conference Calls? *Foundations & Trends in Accounting, 7*(2), 84.
- Mee, A., Homapour, E., Chiclana, F., & Engel, O. (2021). Sentiment analysis using TF-IDF weighting of UK MPs' tweets on Brexit. *Knowledge-Based Systems, 228*
- Modigliani, F., & Miller, M. H. (1958). The Cost Of Capital, Corporation Finance And The Theory Of Investment. *American Economic Review, 48*(3), 261.

- Mouksassi, M., & Pastoor, D. (2015). GGplot-Shiny: A Shiny App That Facilitates Data Manipulation And Exploration. *Journal of Pharmacokinetics and Pharmacodynamics*, 42, S70.
- Nhung, L. T. P., & Okuda, H. (2015). Effects of state ownership on companies' capital structure and profitability: Estimation analysis before and after the Lehman shock. *Journal of Asian Economics*, 38, 64.
- Osman, N. Y., Schonhardt-Bailey, C., Walling, J. L., Katz, J. T., & Alexander, E. K. (2015). Textual analysis of internal medicine residency personal statements: themes and gender differences. *Medical Education*, 49(1), 93.
- R. K. (2007). Leverage Causes Investors To Think Twice. *Bank Loan Report*, 22(44), 3.
- Ranjeeni, K., & Sharma, S. S. (2015). The Effect of the Lehman Brothers' Bankruptcy on the Performance of Chinese Sectors. *Emerging Markets Finance & Trade*, 51(5), 904.
- Thornton, D. L. (2021). The financial crisis: what caused it and when and why it ended. *Applied Economics*, 53(33), 3854.
- Tokarz, R. (2020). Identifying criminal justice faculty research interests using Voyant and NVivo. *Behavioral and Social Sciences Librarian*, 36(3), 113.
- Tosun, C. (2022). Bibliometric and Content Analyses of Articles Related to Science Education for Special Education Students. *International Journal of Disability, Development & Education*, 69(1), 352.

Turing, A. M. (1950). I.— Computing machinery and intelligence. *Mind*, *LIX*(236), 433-460. <https://10.1093/mind/LIX.236.433>

Tyner, S., Hofmann, H., & Briatte, F. (2017). Network visualization with ggplot2. *R Journal*, *9*(1), 27.

Yule, G. U. (1944). *The statistical study of literary vocabulary*.