



Facultad de Ciencias Económicas y
Empresariales Universidad de León

Grado en Finanzas
Curso 2019/2020

**¿INFLUYE LA CORRUPCIÓN EN EL RIESGO SOBERANO EN
ESPAÑA? UN ANÁLISIS MEDIANTE LA HERRAMIENTA
GOOGLE TRENDS.**

(DOES CORRUPTION AFFECT SOVEREIGN RISK IN SPAIN? AN
ANALYSIS THROUGH GOOGLE TRENDS TOOL.)

Realizado por Judit Naveira Gallego

Tutelado por Marcos González Fernández

León, a 14 de Julio de 2020

MODALIDAD DE DEFENSA PÚBLICA:

Tribunal

Póster

AGRADECIMIENTOS

En primer lugar, me gustaría agradecer a mi tutor, Marcos, su paciencia y ayuda, tanto como en su postura de profesor como desde su posición de guía de este “TFG”.

También me gustaría agradecer a mi familia y amigas, y a todas esas personas que han estado presentes en mi camino en la Universidad de León, su gran apoyo a lo largo de estos cuatro años.

¡Muchas gracias!

ÍNDICE DE CONTENIDOS

| | |
|---|----|
| RESUMEN..... | 9 |
| PALABRAS CLAVE..... | 9 |
| ABSTRACT | 9 |
| KEY WORDS | 9 |
| 1. INTRODUCCIÓN..... | 11 |
| 2. METODOLOGÍA | 15 |
| 3. REVISIÓN DE LA LITERATURA | 17 |
| 3.1. LA CORRUPCIÓN Y SU RELACIÓN CON EL RIESGO SOBERANO..... | 17 |
| 3.1.1. El fenómeno de la corrupción: ¿Qué es y cómo se mide?..... | 17 |
| 3.1.2. Efectos de la corrupción sobre el riesgo soberano | 18 |
| 3.2. <i>GOOGLE TRENDS</i> | 21 |
| 4. DATOS Y METODOLOGÍA | 23 |
| 4.1. CORRUPCIÓN..... | 23 |
| 4.2. RIESGO SOBERANO (CDSs) | 27 |
| 5. ANÁLISIS EMPÍRICO..... | 29 |
| 5.1. ANÁLISIS DESCRIPTIVO | 29 |
| 5.2. ANÁLISIS DE REGRESIÓN | 30 |
| 5.3. ANÁLISIS <i>OUT-OF-SAMPLE</i> | 32 |
| 6. CONCLUSIONES..... | 37 |
| 7. REFERENCIAS BIBLIOGRÁFICAS..... | 39 |

ÍNDICE DE GRÁFICOS

| | |
|--|----|
| Gráfico 3.1. Evolución de los principales índices de percepción de la corrupción..... | 18 |
| Gráfico 4.1. Evolución del GSVI para la palabra "corrupcion" | 24 |
| Gráfico 4.2. Evolución de la preocupación de la población por la corrupción | 25 |
| Gráfico 4.3. Evolución anual de los índices de percepción de la corrupción de Transparencia Internacional y Heritage Foundation junto con el GSVI para la palabra "corrupcion" | 26 |
| Gráfico 4.4. Búsquedas en <i>Google Trends</i> de los conceptos: "corrupcion", "españa corrupcion" y "corrupcion política" | 27 |
| Gráfico 4.5. Evolución de los CDSs a 5 años..... | 28 |
| Gráfico 4.6. Evolución de CDSs a 5 años y el GSVI para la palabra "corrupcion" | 28 |
| Gráfico 5.1. Evolución de los CDSs en subperiodos..... | 34 |

ÍNDICE DE TABLAS

| | |
|--|----|
| Tabla 5.1. Estadísticos descriptivos..... | |
| | 29 |
| Tabla 5.2. Correlación entre GSVI "corrupcion" y los CDSs | |
| | 29 |
| Tabla 5.3- Análisis de regresión del GSVI de la palabra “corrupcion” sobre los CDSs.... | |
| | 30 |
| Tabla 5.4. Análisis de regresión de los CDSs retardados y del GSVI de la palabra “corrupcion” sobre los CDSs..... | |
| | 31 |
| Tabla 5.5. Análisis de regresión de los CDSs retardados sobre los CDSs | |
| | 33 |
| Tabla 5.6. Análisis out-of-sample | |
| | 35 |

ÍNDICE DE ABREVIATURAS (en orden alfabético)

CDSs- *Credit Default Swaps.*

CIS- Centro de Investigaciones Sociológicas.

EAM- Error Absoluto Medio.

EPAM- Error Porcentual Absoluto Medio.

GSVI- *Google Search Volume Index.*

MAE- *Mean Absolute Error.*

MAPE- *Mean Absolute Percentage Error*

P.b- Puntos básicos.

RECM- Raíz del Error Cuadrático Medio.

RMSE- *Root Mean Square Error.*

RESUMEN

En este trabajo, se realiza un análisis del efecto de la corrupción en el riesgo soberano en España en el periodo comprendido entre los años 2008 y 2018. Para medir la corrupción se utiliza la herramienta *Google Trends* para la palabra “corrupción”. Por su parte, para medir el riesgo soberano, se utilizan los *Credit Default Swaps* o CDSs. Para analizar cómo la primera influye sobre la segunda, se realiza un análisis empírico a través de la metodología de regresión y el análisis *out-of-sample* para comprobar la capacidad predictiva del modelo. Los resultados muestran que la corrupción es influyente en la evolución de los CDSs, y que su influencia es mayor en los años en los que la economía del país no presenta crisis.

PALABRAS CLAVE

Corrupción | *Google Trends* | Riesgo soberano | *Credit Default Swaps*

ABSTRACT

In this paper, we perform an analysis about the effect of corruption on sovereign risk in Spain in the period between 2008 and 2018 through the Google Trends tool. Corruption and sovereign risk (through the Credit Default Swap) have been studied, as well as the effect that the former has on this risk and, subsequently, we have perform an empirical analysis through regression analysis and Out of Sample analysis. The results show that corruption is influential in the evolution of the CDSs, and that its influence is greater in the years in which the country's economy does not present crisis.

KEY WORDS

Corruption | *Google Trends* / Sovereign risk | *Credit Default Swaps*

*“The evil of corruption reaches into every corner of the world. It lies at the heart of the most urgent problems we face: from economic uncertainty, to endemic poverty, to the ever-present threat of radicalisation and extremism”.*¹ (Cameron, 2016)

1. INTRODUCCIÓN

En España existe una elevada percepción de la corrupción (Villoria Mendieta y Jiménez Sánchez, 2012). Además, ésta es un problema sustancial para el crecimiento económico, así como un obstáculo para la inversión (Mauro, 1995; Shleifer y Vishny, 1993). Este fenómeno detrae multitud de recursos e impide que se empleen en aquellos sectores más productivos de la economía. En este sentido, el efecto de la corrupción en el crecimiento económico se ha investigado exhaustivamente a lo largo de las últimas décadas (Mauro, 1995; Mo, 2001). Si bien estos estudios no niegan que en algunos momentos la corrupción haya jugado un papel positivo en determinados países concretos, los principales hallazgos indican que la corrupción tiende a conducir a un menor crecimiento, obstaculizando el gasto de la inversión, tanto privado como gubernamental, e inhibiendo la eficiencia de servicios públicos (Aidt, 2003; Svensson, 2005).

Por otro lado, si la corrupción ha sido y es un perjuicio para el desarrollo económico, el riesgo soberano, es decir, el posible riesgo de que un estado quiebre en relación a su deuda soberana, no lo ha sido menos. A raíz de la reciente crisis de deuda soberana han sido varios los trabajos que han analizado esta variable (Arghyrou y Kontonikas, 2012; Huixin, 2012; Sgherri y Zoli, 2009).

Aunando estos dos conceptos, algunos estudios han analizado el impacto de la corrupción sobre el riesgo soberano (Chakrabarti y Zeaiter, 2014; Paserman, 2017). Sin embargo, las variables que se utilizan para medir la corrupción suelen ser variables de baja frecuencia, generalmente índices de percepción de la corrupción de frecuencia anual, lo que limita el análisis. En este sentido, en este Trabajo Fin de Grado (en adelante TFG) se aborda el impacto que tiene la percepción de la corrupción sobre el

¹ El demonio de la corrupción llega a todos los rincones del mundo. Se encuentra en el corazón de los problemas más urgentes a los que nos enfrentamos: desde la incertidumbre económica hasta la pobreza endémica y la amenaza siempre presente de la radicalización y el extremismo.

riesgo soberano. Para ello, se utilizan datos relativos a los *Credit Default Swaps*² (en adelante CDSs) para medir el riesgo soberano en España. Los CDSs son el instrumento que mejor refleja el riesgo soberano y se utilizan frecuentemente en la literatura (Ammer y Cai, 2011; Ejsing y Lemke, 2011, entre otros). No obstante, la alta frecuencia de esta variable impide que pueda relacionarse con las medidas tradicionales de corrupción anteriormente indicadas. Es por ello que, a diferencia de en estudios previos, en este TFG se utiliza una variable de mayor frecuencia para medir la corrupción y que se basa en una herramienta frecuentemente usada en la literatura como son los datos provenientes de *Google Trends*, que se han llegado a denominar por algunos autores como *Google Econometrics* (Fondeur y Karamé, 2013). Esta herramienta permite sintetizar en un único indicador un gran volumen de datos sobre las búsquedas de Google, constituyendo de esta forma una nueva fuente de *big data* (Jun et al., 2018). Concretamente, se obtiene el índice de búsquedas de la palabra “corrupcion” como medida de la preocupación o atención sobre la misma con una frecuencia semanal.

Por tanto, el objetivo fundamental de este TFG es el análisis de la relación entre la corrupción y el riesgo soberano, proporcionando un nuevo *proxy* sobre la corrupción de alta frecuencia. Asimismo, se plantean otros objetivos secundarios: i) se pretende obtener una mejor comprensión sobre la evolución del riesgo soberano en España; ii) conocer los principales indicadores del riesgo soberano; iii) aprender más sobre el fenómeno de la corrupción a través de una revisión de la literatura sobre la materia; iv) aprender a utilizar la herramienta *Google Trends*; v) llevar a cabo un análisis de regresión y un análisis *out- of-sample* para determinar la capacidad predictiva de los datos de Google.

Los resultados indican que existe una relación positiva entre la corrupción y el riesgo soberano. Además, con el análisis se demuestra la capacidad predictiva del modelo. Por tanto, en este TFG se ha demostrado que la herramienta *Google Trends* sirve como indicador del comportamiento económico (McLaren y Shanbhogue, 2011) y que, en concreto, los datos de *Google Trends* de nuestra palabra clave, “corrupcion”, pueden ayudar a predecir la evolución del riesgo soberano en España, incluso mejorando la predicción realizada de la variable CDSs retardada un periodo.

² Los CDS son derivados financieros que se utilizan como un seguro contra el riesgo de incumplimiento y que se materializan a través de un contrato de *swap* sobre un instrumento financiero (normalmente un bono).

El resto del TFG se estructura de la siguiente manera: en la segunda sección se explica la metodología utilizada en la realización del mismo. En la tercera sección, se describe la evolución y la importancia de la corrupción, así como del riesgo soberano. Posteriormente, en la sección 4, se resumen los datos y la metodología empleada. En la sección 5 se realiza el análisis empírico. Finalmente se presentan las principales conclusiones en la sección 6.

2. METODOLOGÍA

En este TFG se estudia la relación existente entre la corrupción y el riesgo soberano. Para llevar a cabo este estudio, el TFG se estructura en dos partes claramente diferenciadas en función de la metodología aplicada.

La primera de esas partes hace referencia al análisis de la literatura y del marco teórico, para conocer cuál es el conocimiento previo sobre el objeto del TFG. Para ello, se ha llevado a cabo una revisión bibliográfica, que consiste en sintetizar la información de distintos artículos en los que se estudia la corrupción, el riesgo soberano y la relación entre ambas variables. Además, se estudia *Google Trends* como herramienta de medida de la atención que la población centra en un fenómeno concreto, en este caso, en la atención que la misma presta sobre la corrupción.

La segunda de las partes del trabajo se centra en el análisis empírico, que consiste en estudiar la relación entre la corrupción y el riesgo soberano. A este respecto, los datos para llevar a cabo esta parte empírica del TFG se han obtenido de *Datastream* (CDSs) y *Google Trends* (Corrupción). Se han utilizado los datos semanales de ambas variables desde el año 2008, debido a que es el año en el que estalló la gran crisis financiera. Además, el número de búsquedas que la población realizaba en los años anteriores a 2008 es mucho menor, por lo que los datos procedentes de *Google Trends* son más estables desde esa fecha. Asimismo, los datos relativos a los CDSs comienzan a estar disponibles también a mediados del año 2008. Por estos motivos, el horizonte temporal analizado comprende el periodo entre el 25 de julio del año 2008 y el 31 de diciembre del año 2018. Posteriormente, para estudiar la relación entre estas variables, se utiliza el análisis de regresión. Esta metodología es un método de dependencia que analiza la relación entre una variable dependiente (métrica) y una o varias independientes (métricas), y tiene como objetivos principales: i) obtener la expresión matemática que relaciona las variables; ii) cuantificar la relación de dependencia; iii) determinar el grado de confianza con el que se puede afirmar que la relación es válida; iv) predecir resultados y realizar simulaciones. Es decir, el análisis de regresión consiste en analizar el comportamiento de una variable explicada o dependiente (riesgo soberano, medido a través de los CDSs), en función de una o varias variables independientes o explicativas (corrupción, medida a través de las búsquedas de *Google Trends*).

Finalmente, se ha realizado un análisis *out-of-sample*. Este método permite determinar la capacidad predictiva de los modelos de regresión utilizados para pronosticar los valores de la variable dependiente. Concretamente, se utiliza este método prediciendo una única observación de forma iterativa, es decir, se estima un modelo de regresión hasta el instante t y se predice el valor que proporciona el modelo para la variable dependiente en el instante $t+1$, a continuación, el modelo de regresión se estima hasta $t+1$ y se predice el valor de la variable dependiente para el instante $t+2$ y así sucesivamente. Posteriormente, para comprobar si la precisión de la predicción mejora o empeora al añadir la variable “corrupción” se utiliza el test de Diebold y Mariano (1995), que compara la variable real y los valores pronosticados en el *out-of-sample*.

3. REVISIÓN DE LA LITERATURA

En esta sección del trabajo se realiza un breve repaso de la literatura en relación a los temas que se abordan en este TFG. En concreto, se resumen los principales trabajos que analizan la relación entre la corrupción y el riesgo soberano, así como aquellos estudios que justifican la utilización de la herramienta *Google Trends*.

3.1. LA CORRUPCIÓN Y SU RELACIÓN CON EL RIESGO SOBERANO

A continuación, se resume brevemente en qué consiste el fenómeno de la corrupción, así como las formas en las que se puede medir este fenómeno. Asimismo, se sintetizan aquellos trabajos que vinculan la corrupción con el riesgo soberano.

3.1.1. El fenómeno de la corrupción: ¿Qué es y cómo se mide?

A pesar de que es difícil de determinar una definición concreta de la corrupción, existe un consenso en que la corrupción es un acto en el que se emplean los poderes públicos para obtener ganancias personales eludiendo las reglas de juego (Jain, 2001). Basándose en esta definición, Aidt (2003) establece que se han de cumplir las siguientes condiciones para que se lleve a cabo la corrupción: el uso discrecional del poder, la existencia de rendimientos económicos y la debilidad de las instituciones públicas. Existe también un consenso entre los autores que han investigado este tema respecto a los determinantes de la corrupción. Éstos se agrupan en diferentes categorías: de tipo económico, de tipo cultural y de tipo político legal, siendo esta última la que mayor correlación presenta con la corrupción. Así, el nivel de corrupción depende del equilibrio existente entre los costes asociados al acto de corrupción³ y el beneficio obtenido del mismo. Sus resultados indican que los países protestantes, de origen legal inglés o antecedentes culturales británicos, así como los estados con un mayor nivel de desarrollo económico presentan menores índices de corrupción (Treisman, 2000).

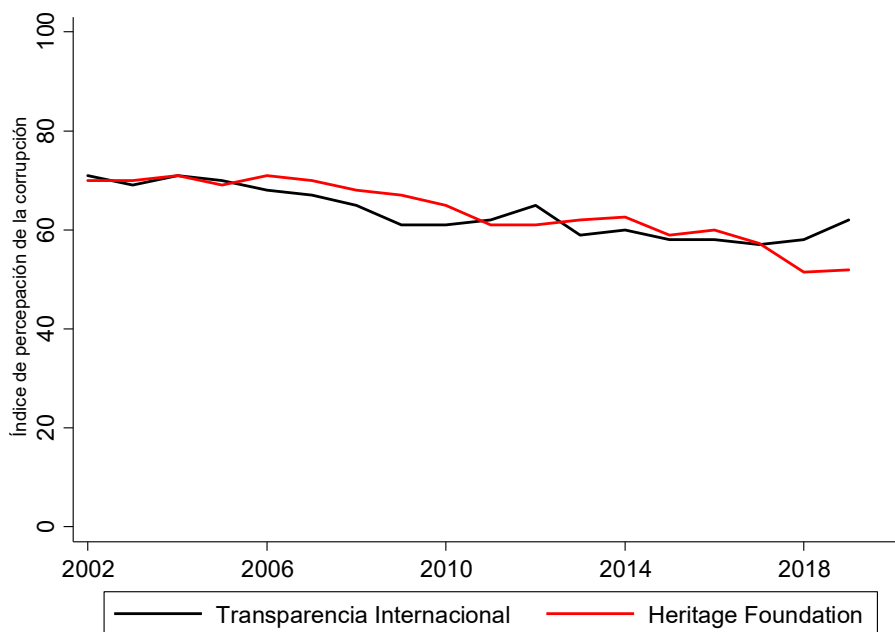
Para medir la corrupción se utilizan índices de percepción de la corrupción. Posiblemente, el más conocido sea el creado por Transparencia internacional⁴, que sirve para medir las percepciones sobre corrupción en el sector público en distintos países del

³ Considerándose dentro de éstos los costes sociales y psicológicos: cultura, religión, etc., además del castigo o sanción en caso de ser detectados.

⁴ Transparencia Internacional es la única organización no gubernamental a escala universal dedicada a combatir la corrupción, congregando a la sociedad civil, sector privado y los gobiernos en una amplia coalición global (Transparency International, 2019).

mundo. Este índice se mueve en una escala de 0 a 100, donde 0 equivale al nivel más alto de percepción de corrupción, y 100 implica el nivel más bajo de percepción de la corrupción. Teniendo esto en cuenta, Transparencia Internacional indica que un país de las características económicas de España, que se sitúa entre las primeras 15 economías del mundo, debería tener un valor por encima de 70 puntos en el Índice de Percepción de la Corrupción si quiere mantener su imagen y su competitividad (Transparency International, 2019). Otro índice de percepción de la corrupción, es el elaborado por la *Heritage Foundation*. Concretamente esta fundación mide lo que denominan *government integrity*, cuyos datos son similares a los del índice de Transparencia Internacional. En el Gráfico 3.1 se puede observar la evolución de estos dos índices. Se aprecia que España lleva más de 10 años sin alcanzar el valor recomendado por Transparencia Internacional.

Gráfico 3.1. Evolución de los principales índices de percepción de la corrupción



Fuente: elaboración propia a partir de datos de Transparencia Internacional y de *Heritage Foundation*

3.1.2. Efectos de la corrupción sobre el riesgo soberano

Antes de analizar el efecto que la corrupción causa sobre el riesgo soberano, primero se ha de tener en cuenta el efecto que ésta tiene sobre la economía de un país en general. En este sentido, el Banco Mundial estima que en el mundo se pagan unos mil millones de dólares en sobornos con fines corruptos cada año. Las consecuencias económicas y

sociales de esta situación son considerables. Por esta razón, James D. Wolfensohn convirtió la corrupción en una de las mayores prioridades durante su paso por la presidencia del Banco Mundial en 1996. Los estudios han examinado repetidamente la relación de la corrupción y el desarrollo económico. En este sentido, si se dan las condiciones adecuadas para un buen gobierno, mediante el adecuado control de la corrupción, se promueve el desarrollo económico. Por otro lado, un gobierno que no previene la corrupción puede perjudicar el progreso económico del país. Como resultado, la corrupción se ha considerado una barrera para la economía del desarrollo y ha sido abordada predominantemente por organizaciones dedicadas a los desafíos de las políticas de desarrollo (Vargas y Sommer, 2014). Estos autores sugieren que se pueden formular cuatro hipótesis basadas en el impacto de la corrupción en la política, la economía, el sistema legal y la sociedad: i) la corrupción es un indicador de una economía sumergida y de fraude fiscal; ii) la corrupción es un signo de inmadurez socioeconómica, capacidad insuficiente para la reforma política y falta de cohesión social; iii) la corrupción es un signo de mala gestión económica; iv) la corrupción indica un riesgo de manipulación estadística. Un ejemplo muy claro de la influencia de la corrupción en la economía de un país es Grecia, al que muchos expertos han tomado como ejemplo de mala gestión económica, atribuyéndole a la corrupción un gran efecto sobre la crisis que sufrió el país heleno.

Asimismo, la corrupción debilita a las instituciones responsables de la gestión macroeconómica. En este sentido, es necesario que una sociedad apoye las reformas políticas y económicas necesarias para reducir los niveles de deuda y fomentar el crecimiento económico. La alta corrupción limita la capacidad de un país para implementar tales reformas porque socava la confianza pública en su liderazgo. Por tanto, la corrupción afecta al riesgo soberano a través de su implicación sobre el crecimiento económico y la capacidad del país para financiarse. A lo largo de los años, se ha desarrollado un gran debate acerca de los efectos de la corrupción en la inversión, desarrollo y crecimiento económico (Mauro, 1995; Shleifer y Vishny, 1993), y se considera que la corrupción juega un importante papel en las crisis financieras (Ciocchini et al., 2003).

También se ha establecido que la solvencia crediticia de un país está vinculada a la corrupción a través de la capacidad del gobierno y su disposición a pagar deuda. Es decir, una mayor corrupción disminuye la recaudación de impuestos, los ingresos del

gobierno y, por tanto, su capacidad para pagar su deuda (Johnson et al., 1997). La corrupción también está asociada con la mala ubicación de los recursos, que puede ser perjudicial para el crecimiento económico y afectar a la solvencia (Depken et al., 2011).

Debido a la falta de un código de bancarrota que proteja a los compradores de bonos soberanos en caso de incumplimiento, canjear la deuda del país es principalmente una decisión política y depende de la voluntad de las personas que tienen el poder para pagarla. Los países más corruptos ofrecen menos protección legal durante los períodos de normalidad y son más propensos a violar los derechos de los inversores durante las crisis⁵. Adama (2013) estudia el efecto de la corrupción en el reembolso de la deuda y el incumplimiento bajo diversas condiciones del mercado. Ese estudio muestra que las fluctuaciones del nivel de corrupción y el ciclo económico en su conjunto afectan a las decisiones de endeudamiento y de incumplimiento y esa corrupción amplifica el efecto de los golpes negativos. Además, los funcionarios corruptos pueden confiscar préstamos u otras fuentes de ingresos del gobierno, lo que limita la capacidad de éste para satisfacer sus obligaciones de deuda. Los funcionarios pueden estar dispuestos a pedir prestados fondos sustanciales (incluso a elevados tipos de interés) para tener la oportunidad de robar (Ciocchini et al., 2003; Shleifer y Vishny, 1993).

En este sentido, las preocupaciones con respecto a los impagos de deuda soberana han ido en aumento durante este milenio. Aunque la crisis financiera de 2008 golpeó sobre las posibles causas y remedios del incumplimiento de deuda, el primer caso registrado se remonta al siglo IV a. C., cuando más de las tres cuartas partes de los municipios griegos no fueron capaces de cumplir con sus obligaciones de deuda con el templo de Delos (Winkler, 1933). Desde entonces, muchos países se han enfrentado a las dificultades de pagar su deuda. Sin embargo, la década pasada fue testigo de una gran acumulación de deuda externa global, que se disparó de 2 billones de dólares en el año 2000 a 60 billones de dólares en el año 2011. En consecuencia, la proporción que la deuda externa representaba en el PIB aumentó de 4.57 % en el año 2000 a 72.5 % en el año 2011 y la deuda externa per cápita aumentó de 329 dólares a 8.700 en los citados años. Durante este período, el mayor incremento anual de la deuda externa fue en 2005, multiplicando más de 5 veces el de 2004. Con el fuerte aumento del riesgo de incumplimiento de deuda de muchos países del mundo, no es sorprendente que una gran

⁵ Ver Lee y Ng (2009), que analizan el efecto de la corrupción en el valor de la empresa.

parte de la investigación continúe analizando los posibles determinantes de este riesgo, pero existe una falta de consenso general sobre las conclusiones alcanzadas por la amplia gama de estudios empíricos sobre la importancia relativa y la dirección del impacto de los determinantes potenciales del incumplimiento soberano, debido a las amplias diferencias de perspectivas, metodologías, selección de muestras y herramientas analíticas. No obstante, la mayor parte de los expertos coinciden en manifestar la existencia de una relación entre la corrupción y el riesgo soberano.

Por tanto, los estudios existentes establecen que la corrupción afecta al riesgo soberano a través de dos relaciones: i) afectando negativamente al crecimiento económico y, de esta forma, empeorando las perspectivas económicas del país y ii) limitando la capacidad del gobierno para financiarse y aumentando los problemas para devolver su deuda.

3.2. GOOGLE TRENDS

Google Trends es una herramienta de Google que proporciona un índice de actividad de búsquedas por consulta o categoría de consulta. Esta herramienta, creada por Google en el año 2006, permite analizar la atención de la población en relación a un término, así como la evolución de la atención con respecto a ese término a lo largo del tiempo (Stephens-Davidowitz y Varian, 2015). Esto hace que numerosos autores hayan utilizado esta herramienta como predicción de algún factor concreto, no solo dentro del ámbito financiero (Da et al., 2011, 2015; Dergiades et al., 2015; Dzielinski, 2012; Vozlyublennaia, 2014) sino para otros campos de estudio, como medicina⁶ o psiquiatría (Cervellin et al., 2017; Ginsberg et al., 2009; Solano et al., 2016).

Este índice no mide el número total de búsquedas, sino la fracción de esas búsquedas que contienen el término escogido en el espacio geográfico elegido en un momento determinado en relación con el número total de consultas en ese momento en ese mismo espacio geográfico. El valor máximo de este índice se establece en 100. Si *Google Trends* muestra que un término de búsqueda ha disminuido con el tiempo, esto no significa necesariamente que haya menos búsquedas ahora de las que hubo previamente, sino que hay menos búsquedas en relación al total de búsquedas ya que en términos

⁶ Por ejemplo, cabe destacar el reciente trabajo de Stephens-Davidowitz (2020), en el que estudia cómo *Google Trends* puede ayudar a predecir los rebotes de COVID-19 a través de las búsquedas de la palabra “anosmia”, ya que es uno de los principales síntomas asociados a la enfermedad.

absolutos las búsquedas han aumentado con el tiempo en prácticamente todos los temas. Además, *Google Trends* tiene un umbral de relevancia por debajo del cual la herramienta reporta el valor 0. Por lo tanto, los términos menos buscados mostrarán ceros más frecuentemente, como también lo harán los períodos de tiempo más antiguos. En este sentido, el período utilizado para la realización del TFG se ha escogido teniendo también esto en cuenta, ya que el periodo más significativo se inicia tras el año 2008 (Stephens-Davidowitz y Varian, 2015).

4. DATOS Y METODOLOGÍA

A continuación, se resumen brevemente los datos empleados y se realiza un análisis preliminar de los mismos. Asimismo, se describe la metodología que se aplicará posteriormente en el análisis empírico.

4.1. CORRUPCIÓN

Para medir la corrupción, se han utilizado los datos extraídos de la herramienta *Google Trends*, disponible en <http://www.google.com/trends/>. Concretamente, se ha obtenido lo que se conoce como *Google Search Volume Index* (en adelante GSVI), que es el índice que proporciona Google. En este sentido, es importante resaltar que el GSVI no representa el número total de búsquedas, sino que es un índice relativo que oscila entre 0 y 100, indicando el valor 100 la máxima intensidad de búsqueda de un término. Para obtener este índice, Google divide el volumen de búsquedas que contienen el término j en el instante t ($V_{j,t}$) entre una muestra aleatoria de todas las búsquedas a para el mismo período t ($V_{a,t}$). Posteriormente esa ratio es normalizada para lo cual se multiplica por un factor de normalización en el que r^* representa el valor máximo del primer cociente (Dergiades et al., 2015), tal y como se muestra en la ecuación (1).

$$GSVI_t = \frac{V_{j,t}^s}{V_{a,t}^s} \times \frac{100}{r^*} \quad (1)$$

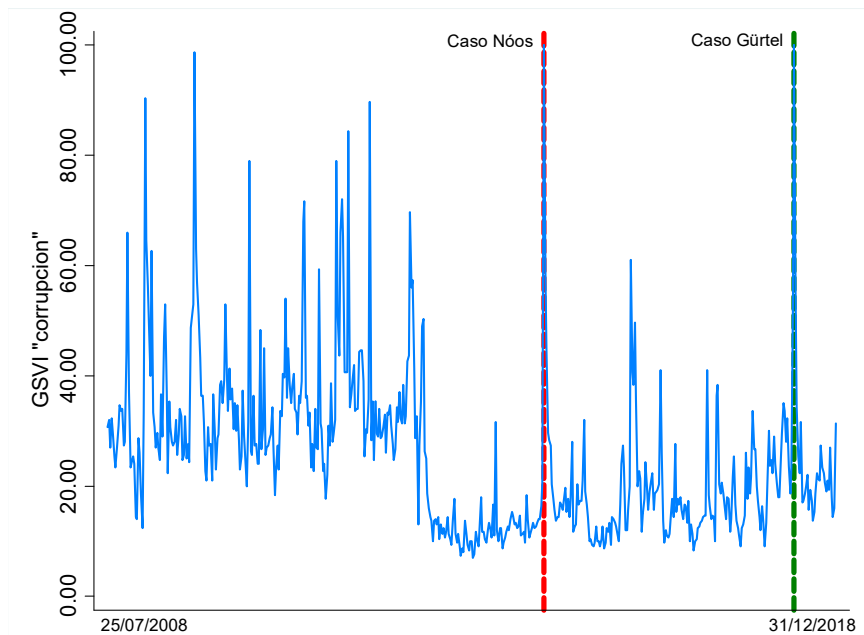
Obviamente, un elemento fundamental en el proceso de obtención de los datos de *Google Trends* es la elección de la palabra o palabras clave. Para este trabajo, se ha seleccionado la palabra “*corrupcion*”⁷. Se han descargado los datos de *Google Trends* con una frecuencia semanal entre enero del año 2008 y la actualidad⁸. Además, los datos se han filtrado por territorio y se refieren únicamente a las búsquedas realizadas en España⁹. En el Gráfico 4.1, se puede observar la evolución de las búsquedas semanales de la palabra “*corrupción*” procedente de *Google Trends* desde el año 2008.

⁷ Se ha buscado sin tilde debido a que es frecuente que los usuarios omitan tildes al realizar las búsquedas al priorizar la velocidad (Lyddy et al., 2014). Los resultados usando la tilde son prácticamente idénticos.

⁸ Para aumentar la velocidad de respuesta, Google utiliza una muestra aleatoria en lugar del número total de búsquedas (Da et al., 2015). Por lo tanto, los datos de Google descargados en diferentes momentos del tiempo pueden diferir ligeramente. Para solucionar este problema, se han descargado los datos en tres fechas diferentes (26, 27 y 28 de marzo de 2020) y se ha usado el valor promedio. Estos valores tienen unas correlaciones superiores al 94%, por lo que la variación entre los mismos es mínima.

⁹ Google permite filtrar los resultados geográficamente basándose en la dirección IP de los usuarios.

Gráfico 4.1. Evolución del GSVI para la palabra "corrupcion"



Fuente: elaboración propia a partir de datos de *Google Trends*

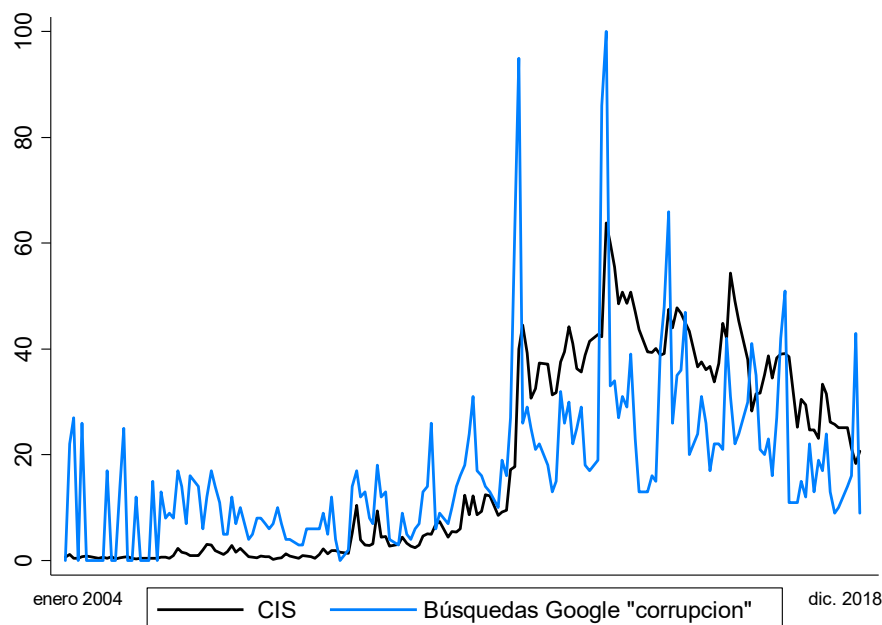
El gráfico muestra como existen diferentes picos en las búsquedas, que coinciden con eventos relacionados con casos de corrupción en España. Así, en el año 2009, fueron detenidas 252 personas por corrupción en España, con estallidos de casos como “Pretoria” o “Palma Arena”, además del caso “Gürtel”. Posteriormente, el pico en las búsquedas en el año 2015 se debe a declaraciones previstas a lo largo de dicho año en varios de los casos de corrupción más importantes en España, tales como el caso “Nóos”, el caso “Rato” o el caso de “Los Pujol”. El otro máximo de la gráfica, el 27 de mayo de 2018, se explica gracias a la sentencia de la red corrupta “Gürtel”, que tuvo lugar en la citada fecha.

Cabe preguntarse si esta medida de la corrupción a través de *Google Trends* realmente mide la corrupción en España. Para comprobar si esto es así, en el Gráfico 4.2 se ha representado la serie mensual¹⁰ del GSVI mensual de la palabra “corrupcion” junto con otra forma de medir la preocupación de la población por la corrupción en España: el informe del CIS (2020). Este informe señala los problemas que la sociedad española considera más importantes en cada momento. Como se puede comprobar, la tendencia de esta preocupación es muy similar a la del GSVI para la palabra

¹⁰ La encuesta del CIS está disponible con periodicidad mensual con la excepción del mes de agosto de cada año. Para poder comparar ambas series se ha eliminado también el dato del mes de agosto de cada año referente al GSVI.

“corrupción”. Concretamente, la correlación entre las series es del 94%. Por ello, se puede determinar que la herramienta *Google Trends* nos proporciona datos fiables de la preocupación de la población en materia de corrupción.

Gráfico 4.2. Evolución de la preocupación de la población por la corrupción

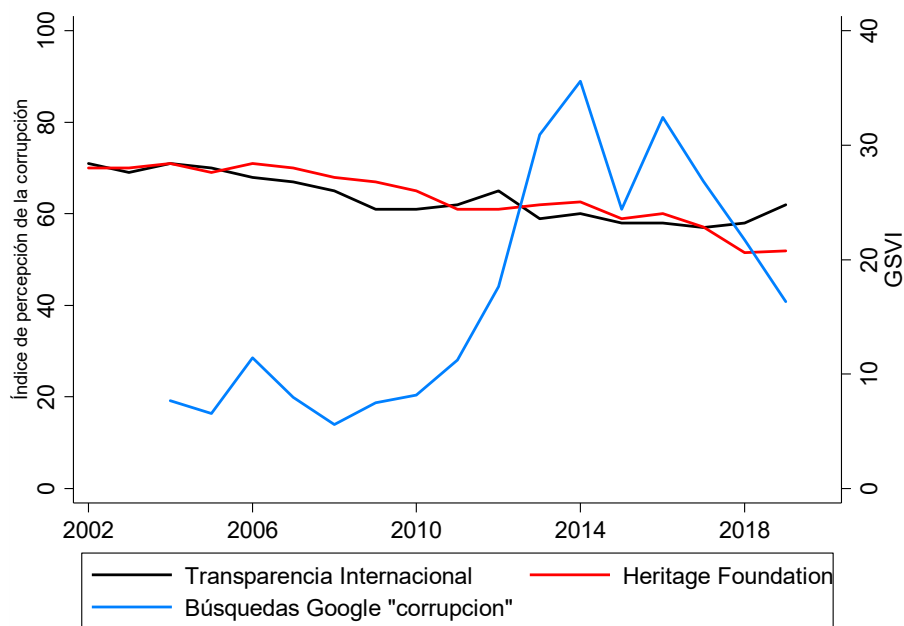


Fuente: elaboración propia a partir de datos de *Google Trends* y del CIS

Además, en el Gráfico 4.3 se muestra la evolución anual del GSVI¹¹ para la palabra corrupción (eje derecho) junto con los índices de percepción de la corrupción elaborados por Transparencia Internacional y por *Heritage Foundation* (eje izquierdo). Es necesario puntualizar que la interpretación de estos índices implica que un mayor valor en el índice representa una menor corrupción, mientras que un mayor GSVI indica una mayor atención o preocupación por la corrupción. Por tanto, a priori se espera una relación opuesta entre los índices y el GSVI. La evolución de las series confirma esta hipótesis. Concretamente, las correlaciones entre el GSVI y los índices de percepción de la corrupción de Transparencia Internacional y de *Heritage Foundation* son del -71% y -57% respectivamente.

¹¹ Para la comparación se ha calculado el GSVI anual como el promedio de los datos mensuales. Ya que no es posible obtener el GSVI en periodicidad anual directamente de *Google Trends*.

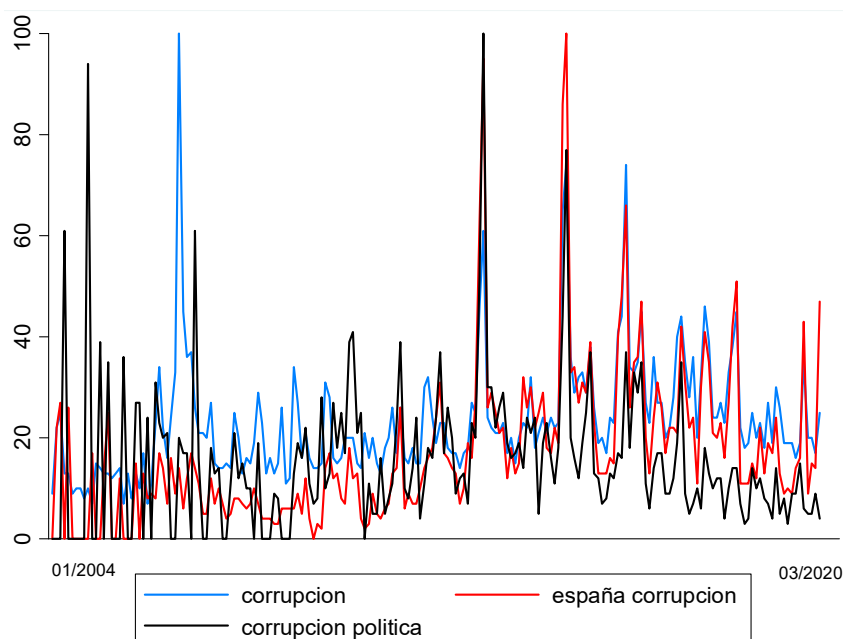
Gráfico 4.3. Evolución anual de los índices de percepción de la corrupción de Transparencia Internacional y Heritage Foundation junto con el GSVI para la palabra “corrupcion”



Fuente: elaboración propia a partir de datos de Transparencia Internacional, *Heritage Foundation* y *Google Trends*

Es necesario destacar que se ha considerado utilizar otras palabras en lugar de “corrupcion”. En este sentido, se planteó la posibilidad de usar las palabras “españa corrupcion” o “corrupcion politica”. Aunque, como se puede comprobar en el Gráfico 4.4, los tres conceptos tienen una tendencia similar, “españa corrupcion” y “corrupcion politica” muestran un elevado número de ceros en la serie de datos procedente de *Google Trends*. Por tanto, se ha considerado que la palabra “corrupcion” es la más adecuada y, además, es la que presenta una mayor correlación a priori con los CDSs. Además, el término “corrupcion” es lo suficientemente concreto y su significado es inequívoco e invariante a lo largo del tiempo, que son los principales requisitos que ha de tener la palabra clave cuando se utilizan datos de *Google Trends* (Dzielinski, 2012).

Gráfico 4.4. Búsquedas en *Google Trends* de los conceptos: "corrupcion", "españa corrupcion" y "corrupcion política"



Fuente: elaboración propia a partir de datos de *Google Trends*

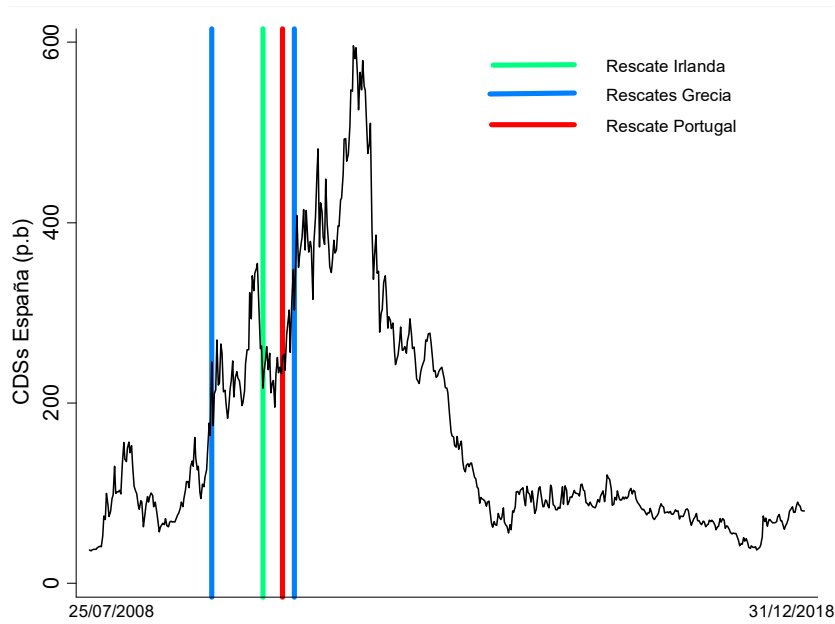
4.2. RIESGO SOBERANO (CDSs)

Los bonos soberanos son la principal forma de financiación para los estados y se han convertido en una importante clase de activos en carteras. Muchos mercados emergentes dependen del capital extranjero para su desarrollo y para cubrir necesidades financieras, confiando en entradas de fondos a través de los mercados internacionales de bonos (Panizza et al., 2009).

El riesgo soberano hace referencia al riesgo de que se produzca un impago de esos bonos soberanos. Para medir el riesgo soberano se utilizan los datos relativos a los CDSs. Concretamente, se han usado los CDSs a 5 años, ya que son los más comúnmente usados en la literatura (Ammer y Cai, 2011; Arce et al., 2013; Kalbaska y Gatkowski, 2012, entre otros). Los datos se han obtenido de *Datastream* con una frecuencia semanal, debido a que es la frecuencia temporal utilizada en otros estudios¹² para el horizonte temporal desde el 25 de julio del año 2008 hasta el 31 de diciembre del año 2018.

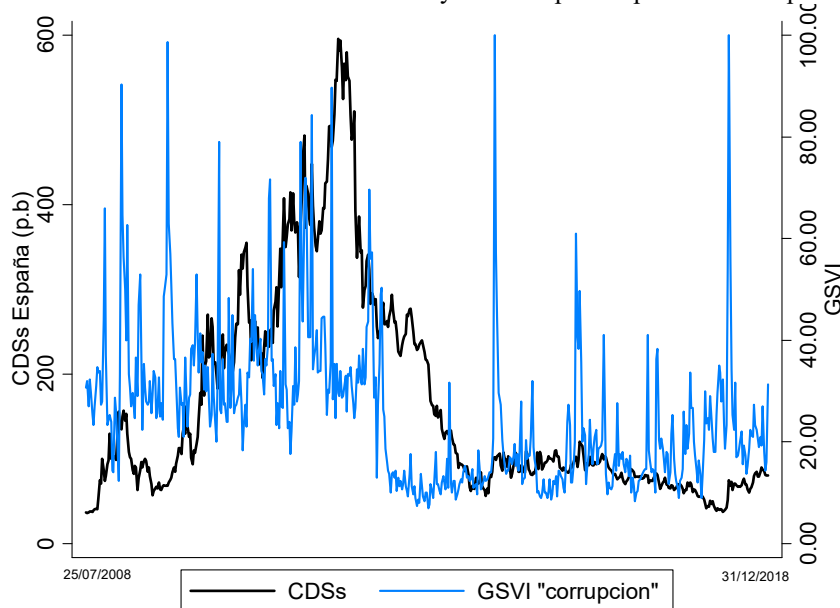
¹² Ver Paserman (2017), que estudia la relación entre la corrupción y el riesgo soberano a través de datos semanales.

Gráfico 4.5. Evolución de los CDSs a 5 años



En el Gráfico 4.5 se muestra la evolución de los CDSs a 5 años españoles durante el período analizado. Como se puede observar, el precio de los CDSs comenzó a aumentar con el estallido de la crisis de 2008 y, especialmente, tras la crisis de deuda soberana que afectó principalmente a Grecia, Portugal, Irlanda y España en torno a los años 2010 y 2011. Desde ese momento, se observa una reducción en el valor de los CDSs en consonancia con la recuperación del país. Finalmente, en el Gráfico 4.6 se muestran las dos series objeto de estudio de forma conjunta.

Gráfico 4.6. Evolución de CDSs a 5 años y el GSVI para la palabra “corrupcion”



5. ANÁLISIS EMPÍRICO

En este apartado, se analiza empíricamente cómo afecta la corrupción al riesgo soberano a través de los CDSs. Para ello, se realiza, en primer lugar, un análisis descriptivo y, posteriormente, un análisis de regresión y un análisis *Out of Sample*.

5.1. ANÁLISIS DESCRIPTIVO

Antes de realizar el análisis empírico, en la Tabla 5.1 se muestran los principales estadísticos descriptivos de las variables objeto del estudio. Se observa que la muestra cuenta con un total de 545 observaciones semanales. Los CDSs han tenido un valor medio de 163,68 p.b. con una elevada desviación típica, probablemente como consecuencia de la crisis financiera y la crisis de deuda soberana que provocaron una elevada volatilidad entre los años 2008 y 2012. El punto máximo se alcanza el 1 de junio del año 2012 con 595,94 p.b., a partir de entonces la crisis de deuda soberana se relaja hasta valores más habituales.

Tabla 5.1. Estadísticos descriptivos

| Variable | N | Media | Desv. Típica | Min. | Max. | Mediana |
|-------------------|-----|--------|--------------|------|--------|---------|
| GSVI "corrupcion" | 545 | 25,65 | 14,89 | 7 | 100 | 23,33 |
| CDSs | 545 | 163,68 | 125,67 | 36 | 595,94 | 100 |

La tabla muestra los estadísticos descriptivos de CDSs soberanos de España a 5 años y el Google Search Volume Index (GSVI) semanal para la palabra "corrupción".

Fuente: elaboración propia a partir de Datos de *Datastream* y *Google Trends*.

En la Tabla 5.2 se muestra la correlación entre el precio de los CDSs y el GSVI para nuestra palabra "corrupcion". Tal y como se esperaba, la correlación entre ambas variables es positiva y significativa. Por tanto, el signo positivo de la correlación indica que, a priori, un incremento de las búsquedas de la palabra "corrupcion" implica un incremento del riesgo soberano.

Tabla 5.2. Correlación entre GSVI "corrupcion" y los CDSs

| | GSVI "corrupcion" | CDSs |
|-------------------|-------------------|------|
| GSVI "corrupcion" | 1.00 | |
| CDSs | 0,2946*** | 1.00 |

La tabla muestra la correlación entre los valores semanales de los CDSs soberanos de España a 5 años y el Google Search Volume Index (GSVI) semanal para la palabra "corrupción". El horizonte temporal comprende el periodo desde el 25 de julio de 2008 hasta el 31 de diciembre de 2018.

*** Significativo al 1%

** Significativo al 5%

* Significativo al 10%

Fuente: elaboración propia a partir de Datos de *Datastream* y *Google Trends*.

5.2. ANÁLISIS DE REGRESIÓN

Para profundizar en el análisis de la relación entre el GSVI de la palabra “corrupcion” y el riesgo soberano, a continuación se ha realizado un análisis de regresión sobre la rentabilidad de los CDSs teniendo en cuenta las búsquedas de la palabra “corrupción” como variable explicativa. El modelo inicial tiene la siguiente forma:

$$CDSs_t = \alpha + \beta_1 GSVI \text{ "corrupcion"}_t + \varepsilon_t \quad (2)$$

donde $CDSs_t$ es la variable dependiente y sobre la que se va a realizar la regresión, α es la constante del modelo, $\beta_1 GSVI \text{ "corrupcion"}_t$ representa el GSVI de la palabra “corrupcion” y ε_t es el término de error. Una vez se ha establecido el modelo, se aplica a los datos existentes y se obtienen los resultados que se muestran en la Tabla 5.3.

| Tabla 5.3- Análisis de regresión del GSVI de la palabra “corrupcion” sobre los CDSs | |
|---|---------------------|
| Variable dependiente: CDSs soberanos de España a 5 años | Modelo |
| GSVI “corrupción” | 2,485*** (0,042) |
| Constante | 99,92 (10,04) |
| R^2 | 0,0868 |
| Estadístico F | 34,54*** |
| RECM | 120,21 |
| Nº observaciones | 545 |

La tabla muestra los resultados de la regresión del GSVI de la palabra “corrupcion” sobre los CDSs soberanos de España a 5 años. Los errores robustos estándar se muestran en paréntesis. RECM indica la Raíz del Error Cuadrático Medio.

*** Significativo al 1%

** Significativo al 5%

* Significativo al 10%

Fuente: elaboración propia a partir de Datos de *Datastream* y *Google Trends*.

Se observa como el GSVI de la palabra “corrupcion” muestra un coeficiente positivo y significativo, como se esperaba. De esta forma, se confirma que un aumento en las búsquedas de Google conlleva un incremento de los CDSs soberanos y, por extensión, del riesgo soberano. Se puede decir, por tanto, que un incremento en la preocupación por la corrupción, que se traduce en un mayor número de búsquedas, aumenta el riesgo soberano. Concretamente, de acuerdo con el modelo, un incremento en una desviación típica en las búsquedas de Google implica un incremento de los CDSs de 37 p.b. Además, la capacidad explicativa del modelo queda contrastada con el estadístico F, que contrasta la validez del modelo en su conjunto, y que muestra un valor altamente significativo.

Sin embargo, esta relación positiva puede ser una relación espuria. El bajo valor del R^2 y el valor elevado de la raíz del error cuadrático *medio* (generalmente conocido por RECM o también por sus siglas en inglés *RMSE*, *Root Mean Square Error*) son indicativos de que el GSVI de la palabra “corrupcion” por sí solo no tiene un gran poder explicativo sobre los CDSs. Para contrastar este extremo, se ha añadido en el modelo el primer retardo de la variable dependiente, es decir, la variable CDSs retardada un periodo, tal y como se indica en la ecuación 3. De esta forma, es posible que el retardo de la variable dependiente absorba toda la influencia y desaparezca el efecto del GSVI de la palabra “corrupcion”. Se utiliza la siguiente ecuación:

$$CDSs_t = \alpha + \beta_1 CDSs_{t-1} + \beta_2 GSVI \text{ "corrupcion"}_t + \varepsilon_t \quad (3)$$

donde $CDSs_t$ es la variable dependiente y sobre la que se va a realizar la regresión, α es la constante del modelo, $\beta_1 CDSs_{t-1}$ representa la variable CDSs retardada un periodo, $\beta_2 GSVI \text{ "corrupcion"}_t$ representa el GSVI de la palabra “corrupcion” y ε_t es el término de error. Una vez se ha establecido el modelo, se aplica a los datos existentes y se obtienen los resultados que se muestran en la Tabla 5.4.

Tabla 5.4. Análisis de regresión de los CDSs retardados y del GSVI de la palabra “corrupcion” sobre los CDSs

| Variable dependiente: CDSs soberanos de España a 5 años | Modelo |
|---|---------------------|
| CDSs _{t-1} | 0,983*** (0,010) |
| GSVI “corrupción” | 0,136* (0,070) |
| Constante | -0,710 (2,010) |
| R^2 | 0,9777 |
| Estadístico F | 4.514,75 *** |
| RECM | 18,821 |
| Nº observaciones | 544 |

La tabla muestra los resultados de la regresión de los CDSs soberanos de España a 5 años retardados y el GSVI de la palabra “corrupcion” sobre los CDSs soberanos de España a 5 años. Los errores robustos estándar se muestran en paréntesis. RECM indica la Raíz del Error Cuadrático Medio.

*** Significativo al 1%

** Significativo al 5%

* Significativo al 10%

Fuente: elaboración propia a partir de Datos de *Datastream* y *Google Trends*.

Se observa que la relación positiva y significativa del GSVI con los CDSs no desaparece, si bien la significatividad se reduce ya que pasa a ser significativo tan sólo al 10%. Asimismo, el valor del coeficiente disminuye. Todo esto se debe a que el

retardo de los CDSs absorbe gran parte de la influencia sobre la variable dependiente. Esto suele ser habitual en series financieras en las que el valor del activo depende en gran medida del valor de ese mismo activo en el periodo inmediatamente anterior. Por ello, el valor del R^2 se incrementa de forma sustancial, como era previsible, y se reduce el *RMSE*. Sin embargo, a pesar de que el retardo de los CDSs reduce la significatividad y el impacto del GSVI de la palabra “corrupcion”, su efecto no desaparece. Se puede decir, entonces, que cuanto mayor corrupción exista y, por tanto, cuanto mayor preocupación tenga la población respecto a este fenómeno, mayor es el precio de los CDSs.

5.3. ANÁLISIS *OUT-OF-SAMPLE*

En el análisis de regresión anterior (*in-sample*) se ha demostrado que existe una relación positiva y significativa entre los CDSs y el GSVI de la palabra “corrupcion”. Sin embargo, para completar el análisis empírico, se realiza un análisis *out-of-sample* para determinar si la inclusión del GSVI permite mejorar la predicción del valor de los CDSs. En primer lugar se plantea un modelo AR(1) en el que se realiza una regresión de la variable dependiente sobre sí misma retardada un periodo tal y como se muestra en la ecuación 4:

$$CDSs_t = \alpha + \beta_1 CDSs_{t-1} + \varepsilon_t \quad (4)$$

donde $CDSs_t$ es la variable dependiente y sobre la que se va a realizar la regresión, α es la constante del modelo, $\beta_1 CDSs_{t-1}$ representa la variable CDSs retardada un periodo y ε_t es el término de error. Una vez se ha establecido el modelo, se aplica a los datos existentes.

Los resultados de la Tabla 5.5 muestran que el coeficiente de los CDSs retardados un periodo es muy próximo a 1. Esto suele ser indicativo de que este modelo es el que proporciona la mejor predicción (Choi y Varian, 2012; González-Fernández y González-Velasco, 2019; McLaren y Shanbhogue, 2012). No obstante, lo que se plantea a continuación es comparar la predicción de este modelo de la ecuación 4 con la predicción obtenida del modelo de la ecuación 3, que incluye además el GSVI de la palabra “corrupción”. Si la predicción del modelo de la ecuación 3 es superior, eso implica que el GSVI de la palabra “corrupcion” es útil para predecir los CDSs.

Tabla 5.5. Análisis de regresión de los CDSs retardados sobre los CDSs

| Variable dependiente: CDSs soberanos de España a 5 años | Modelo |
|---|---------------------|
| CDSs $t-1$ | 0,988*** (0,011) |
| Constante | 2,027 (1,247) |
| R^2 | 0,9774 |
| Estadístico F | 8.657,72*** |
| RECM | 18,904 |
| Nº observaciones | 544 |

La tabla muestra los resultados de la regresión de los CDSs soberanos de España a 5 años retardados sobre los CDSs soberanos de España a 5 años. Los errores robustos estándar se muestran en paréntesis. RECM indica la Raíz del Error Cuadrático Medio.

*** Significativo al 1%

** Significativo al 5%

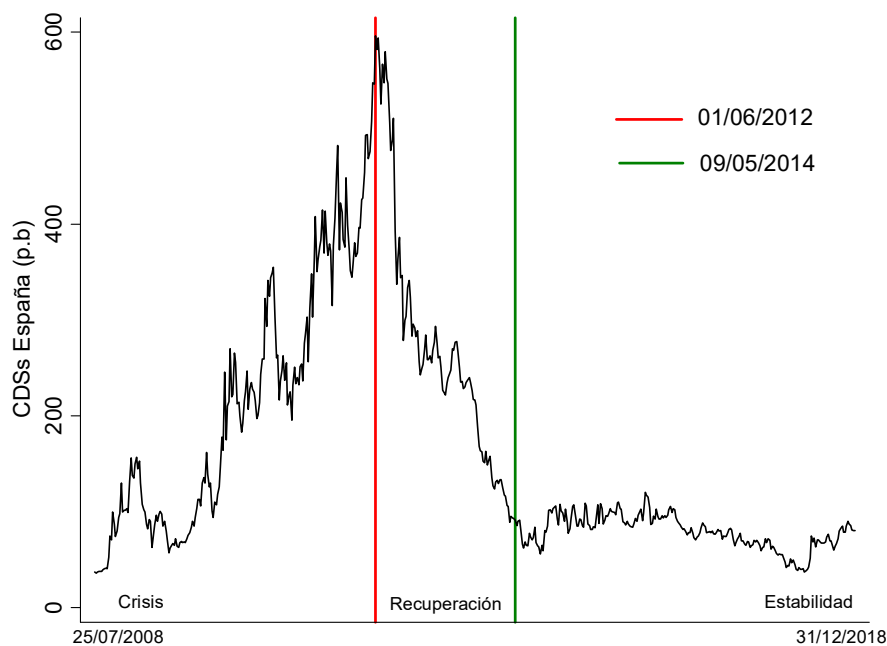
* Significativo al 10%

Fuente: elaboración propia a partir de Datos de *Datastream* y *Google Trends*.

Específicamente, para llevar a cabo este análisis, cada uno de estos modelos se estima hasta el 18 de junio de 2010¹³, y luego, se predicen los datos de la semana posterior. A continuación, se estima el modelo hasta el 25 de junio de 2010, y se predicen los datos de la semana posterior. Se repite este procedimiento cada semana hasta el final de la muestra. Finalmente, se recopilan las predicciones para los dos modelos y se compara la predicción con los datos reales de los CDSs. Esta comparación tiene en cuenta cuál de los dos modelos comete mayores errores al predecir al compararlos con los datos reales. Para medir este error se suele usar generalmente la RECM. Sin embargo, un menor valor en la RECM no implica que el modelo sea mejor. En este sentido, hay que verificar que las diferencias entre los modelos sean estadísticamente significativas. Para ello se utiliza el test de Diebold y Mariano (1995) que verifica si las predicciones son significativamente diferentes. Concretamente, este test plantea la hipótesis nula de que la capacidad predictiva de los dos modelos analizados es la misma. Por tanto, un rechazo de la hipótesis nula indica que aquel modelo que muestre unas menores diferencias con los datos reales es el que predice mejor. Generalmente, este test usa por defecto la RECM para comprobar este extremo, aunque también pueden usarse otras medidas alternativas como el Error Porcentual Absoluto Medio (EPAM), también conocido por sus siglas en inglés MAPE (*Mean Absolute Percentage Error*), o el Error Absoluto Medio (EAM) conocido en inglés como MAE (*Mean Absolute Error*).

¹³ Se ha escogido esta fecha debido a que representa la observación número 100 y ya se considera un número lo suficientemente elevado para realizar la predicción.

Gráfico 5.1. Evolución de los CDSs en subperiodos



Dado que la predicción se genera desde el 25 de junio hasta el 31 de diciembre de 2018, se han distinguido tres subperiodos dentro de ese horizonte temporal, que, como se puede observar en el gráfico 5.1, son las tres etapas de la economía española en los últimos años. El primero de ellos (Panel A) comprende el periodo de crisis de deuda soberana, es decir, desde el 25 de junio de 2010 hasta el valor máximo alcanzado por los CDSs: el 1 de junio de 2012. Por otro lado, el segundo subperiodo (Panel B) abarca el periodo posterior a la crisis, en el que España sufre un periodo de recuperación, y comprende desde junio de 2012 hasta mayo de 2014, que es donde los datos empiezan a ser más estables, en consonancia con la economía general del país. Finalmente, el tercer subperiodo (Panel C) abarca el periodo de estabilización de la economía española, desde mayo de 2014 hasta diciembre de 2018. Los resultados del análisis *out-of-sample* se muestra en la Tabla 5.6. Además, en la tabla se presentan la RECM, el EPAM y el EAM del modelo base (ecuación 4) mientras que para el modelo con GSVI (ecuación 3) se muestra la ratio con respecto a los valores del modelo base. Por tanto, un valor inferior (superior) a 1 indica que el modelo que incluye el GSVI de la palabra “corrupcion” mejora (empeora) la predicción del modelo base.

Tabla 5.6. Análisis out-of-sample

| Panel A (Periodo crisis: 25 junio 2010 - 1 de junio de 2012) | | |
|---|-------------|-----------------|
| Variable dependiente: CDSs | Modelo base | Modelo con GSVI |
| RECM | 32,35 | 1,006* |
| EPAM | 0,0818 | 1,005* |
| EAM | 25,72 | 1,005 |
| Panel B (Periodo de recuperación: 8 de junio de 2012 –2 de mayo de 2014) | | |
| Variable dependiente: CDSs | Modelo base | Modelo con GSVI |
| RECM | 21,67 | 0,989*** |
| EPAM | 0,232 | 0,969 |
| EAM | 3,786 | 0,985 |
| Panel C (Periodo de estabilidad: 9 de mayo de 2014 – 31 de diciembre de 2018) | | |
| Variable dependiente: CDSs | Modelo base | Modelo con GSVI |
| RECM | 6,924 | 0,968** |
| EPAM | 0,254 | 0,978*** |
| EAM | 2,251 | 0,980*** |

La tabla muestra los valores de la Raíz del Error Cuadrático Medio (RECM), del Error porcentual Absoluto Medio (EPAM) y del Error Absoluto Medio (EAM) para las estimaciones *out-of-sample*. En primer lugar, se realiza una regresión hasta el 25 de junio de 2010 y se predicen los datos de la semana posterior. Se repite este procedimiento cada semana hasta el final de la muestra. El modelo base es un modelo AR(1) en el que la variable explicativa son los CDSs retardados un periodo (ecuación 4). Por su parte, el modelo con GSVI incluye las búsquedas de Google de la palabra “*corrupcion*” (ecuación 3). Al igual que en D’Amuri y Marcucci (2017), se muestran los valores para el modelo base, mientras que para el modelo con GSVI se informa de la ratio de éste en relación al modelo base. Por tanto, un valor inferior a 1 sugiere que el modelo mejora la predicción de referencia y viceversa.

* Significativo al 10%

** Significativo al 5%

*** Significativo al 1%

Fuente: elaboración propia a partir de Datos de *Datastream* y *Google Trends*.

A la vista de los resultados, el análisis *out-of-sample* revela que el modelo que incluye el GSVI no supera al modelo base en el periodo de crisis de deuda soberana, ya que todos los valores para el modelo con GSVI son superiores a la unidad, independientemente del criterio que se tenga en cuenta para medir el error (RECM, EPAM o EAM). Además, la prueba de Diebold y Mariano (1995) muestra que la precisión de la predicción es diferente entre los modelos, lo que indica que la inclusión de GSVI no mejora significativamente la predicción, es decir, no reduce de forma significativa los errores del pronóstico. Por tanto, durante la época de la crisis de deuda soberana, el modelo que mejor predice los CDSs soberanos en España es el modelo base, ya que, como se ha dicho anteriormente, introducir la palabra “*corrupcion*” no mejora la predicción en este caso.

Sin embargo, como se puede observar en el Panel B de la Tabla 5.6, el análisis del período de recuperación arroja resultados totalmente opuestos. Durante esta etapa, se

observa que la predicción del modelo que incluye los datos del GSVI es superior a la del modelo base, con independencia del criterio utilizado (RECM, EPAM o EAM). Cabe destacar, también, que el test de Diebold y Mariano (1995) indica que la diferencia de la RECM es estadísticamente significativa.

El análisis del periodo posterior, en el que se puede apreciar la estabilización de la economía española, muestra que la influencia de la corrupción en el riesgo soberano (a través de los CDSs) es aún mayor que en la anterior estimación, así como su significación. Durante este periodo se observa que la predicción del modelo que incluye los datos del GSVI es superior a la del modelo base, con independencia del criterio utilizado (RECM, EPAM o EAM). Asimismo, el test de Diebold y Mariano (1995) indica que las diferencias son estadísticamente significativas al 5% en todos los casos. Además, se ha comprobado que esta capacidad predictiva del modelo para el periodo posterior a la crisis se mantiene para otros periodos de inicio diferentes a las fechas seleccionadas. Parece ser, por tanto, que el GSVI de la palabra “corrupcion” incrementa la capacidad predictiva cuando no existe una crisis de deuda soberana. Durante ese periodo de crisis (hasta junio de 2012) el análisis *out-of-sample* arroja diferencias significativas en favor del modelo base.

Del análisis realizado se puede afirmar que los datos del GSVI de la palabra corrupción, no sólo tienen una relación directa y significativa con los CDSs, sino que esta relación es útil para mejorar las predicciones de los CDSs. Además, del análisis se concluye que estos resultados son más relevantes cuando no existe una crisis de deuda como la que tuvo lugar hasta el año 2012. Durante esta etapa, los datos del GSVI no permiten mejorar las predicciones, ya que la corrupción supone una preocupación menor para la población en época de crisis y, también, porque la evolución de los CDSs a 5 años en España tiene unas fluctuaciones mucho más elevadas en estos periodos.

6. CONCLUSIONES

En este TFG, se ha analizado la relación entre la corrupción, o la preocupación que presenta la población sobre la corrupción y el riesgo soberano en España en el periodo 2008-2018. Para medir la corrupción se ha utilizado la herramienta *Google Trends*, que permite obtener un indicador de alta frecuencia (semanal) de la corrupción, frente a otros indicadores tradicionales cuya frecuencia es anual. Para aproximar el riesgo soberano se han utilizado los CDSs soberanos a 5 años.

En la primera parte del TFG se ha realizado un análisis introductorio sobre las variables objeto de estudio. Concretamente, se ha mostrado la preocupación que existe en la actualidad, y que ha existido a lo largo de los últimos años, respecto al fenómeno de la corrupción en España, mediante la observación de varios índices que miden esta percepción. Se ha observado como la corrupción juega un papel importante en España, manteniéndose entre las preocupaciones de la población de nuestro país a lo largo de todos los años de la muestra utilizada para la realización de este TFG. También se ha estudiado el efecto que esta puede tener sobre la economía de un país, llegando a la conclusión de que la corrupción es un signo de presencia de economía sumergida y de fraude fiscal en un país, así como una representación de inmadurez socioeconómica, de una insuficiente capacidad para llevar a cabo reformas políticas y de una falta de cohesión social. Además, la corrupción es un signo de mala gestión económica en el gobierno, ya que un gobierno que no previene la corrupción está poniéndole una barrera al progreso económico de su país. Respecto al riesgo soberano, cabe destacar que se ve afectado por la corrupción debido al efecto que ésta tiene en la economía en general y por la limitación que causa en la capacidad del gobierno para financiarse. Esto provoca que los problemas para devolver la deuda se vean aumentados. Además, también se ha investigado acerca del funcionamiento de *Google Trends* y se ha analizado su utilidad como herramienta de predicción y, en concreto, de predicción del riesgo soberano español mediante las búsquedas de la palabra “corrupcion”.

En la segunda parte de este TFG, se ha llevado a cabo un análisis empírico para comprobar si existe una relación entre la corrupción y el riesgo soberano del país en cuestión. Para ello, se ha utilizado el análisis de regresión y un análisis *out-of-sample*. Los resultados obtenidos para el periodo 2008-2018 mediante estos análisis muestran que el nivel de corrupción percibido en nuestro país tiene un efecto positivo y

significativo en el riesgo soberano del mismo. Esta afirmación va en consonancia con otros estudios previos y respalda la importancia de los datos de Google como una herramienta alternativa para medir la atención de los inversores en el ámbito económico. Además, la capacidad predictiva del modelo que incluye los datos de Google mejora las predicciones de forma significativa.

Los resultados van en la línea de estudios previos que ya han demostrado la capacidad de los datos de Google para predecir variables económicas y financieras. Estos resultados pueden tener diferentes implicaciones, teniendo en cuenta, además, que los datos de Google se pueden obtener con facilidad y proporcionan información más transparente que otras fuentes (Da et al., 2015). En este sentido, un análisis pormenorizado de los datos facilitados por *Google Trends* puede proporcionar información importante a los inversores que quieran predecir el precio de los CDSs.

7. REFERENCIAS BIBLIOGRÁFICAS

- Adama, A. (2013). Corruption and Political Turnover in a Sovereign Default Model Government Households Recursive Equilibrium Calibration. *Working Papers, School of Social Sciences- University of Manchester*.
- Aidt, T. S. (2003). Economic analysis of corruption: A survey. *Economic Journal*, 113(491), F632-F652. <https://doi.org/10.1046/j.0013-0133.2003.00171.x>
- Ammer, J., y Cai, F. (2011). Sovereign CDS and bond pricing dynamics in emerging markets: Does the cheapest-to-deliver option matter? *Journal of International Financial Markets, Institutions and Money*, 21(3), 369-387. <https://doi.org/10.1016/j.intfin.2011.01.001>
- Arce, O., Mayordomo, S., y Peña, J. I. (2013). Credit-risk valuation in the sovereign CDS and bonds markets: Evidence from the euro area crisis. *Journal of International Money and Finance*, 35(0), 124-145.
- Arghyrou, M. G., y Ktonikas, A. (2012). The EMU sovereign-debt crisis: Fundamentals, expectations and contagion. *Journal of International Financial Markets, Institutions and Money*, 22(4), 658-677. <https://doi.org/10.1016/J.INTFIN.2012.03.003>
- Cameron, D. (2016). *Anti-Corruption Summit*.
- Cervellin, G., Comelli, I., y Lippi, G. (2017). Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *Journal of Epidemiology and Global Health*, 7(3), 185-189. <https://doi.org/10.1016/j.jegh.2017.06.001>
- Chakrabarti, A., y Zeaiter, H. (2014). The determinants of sovereign default: A sensitivity analysis. *International Review of Economics and Finance*, 33, 300-318. <https://doi.org/10.1016/j.iref.2014.06.003>
- Choi, H., y Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88(s1), 2-9. <https://doi.org/10.1111/j.1475-4932.2012.00809.x>
- Ciocchini, F., Durbin, E., y Ng, D. T. C. (2003). Does corruption increase emerging market bond spreads? *Journal of economics and business*, 55(5-6), 503-528.
- CIS. (2020). *Three principal problems*. http://www.cis.es/cis/export/sites/default/-Archivos/Indicadores/documentos_html/TresProblemas.html
- D'Amuri, F., y Marcucci, J. (2017). The predictive power of Google searches in forecasting US unemployment. *International Journal of Forecasting*, 33(4), 801-816. <https://doi.org/10.1016/J.IJFORECAST.2017.03.004>

- Da, Z., Engelberg, J., y Gao, P. (2011). In Search of Attention. *The Journal of Finance*, 66(5), 1461-1499. <https://doi.org/10.1111/j.1540-6261.2011.01679.x>
- Da, Z., Engelberg, J., y Gao, P. (2015). The Sum of All FEARS Investor Sentiment and Asset Prices. *Review of Financial Studies*, 28(1), 1-32. <https://doi.org/10.1093/rfs/hhu072>
- Depken, C. A., Lafountain, C. L., y Butters, R. B. (2011). Corruption and Creditworthiness: Evidence from Sovereign Credit Ratings. En *Sovereign Debt: From Safety to Default* (pp. 79-87). John Wiley and Sons. <https://doi.org/10.1002/9781118267073.ch9>
- Dergiades, T., Milas, C., y Panagiotidis, T. (2015). Tweets, Google trends, and sovereign spreads in the GIIPS. *Oxford Economic Papers*, 67(2), 406-432. <https://doi.org/10.1093/oep/gpu046>
- Diebold, F. X., y Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13(3), 253-263. <https://doi.org/10.1080/07350015.1995.10524599>
- Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters*, 9(3), 167-175. <https://doi.org/10.1016/j.frl.2011.10.003>
- Ejsing, J., y Lemke, W. (2011). The Janus-headed salvation: Sovereign and bank credit risk premia during 2008-2009. *Economics Letters*, 110(1), 28-31.
- Fondeur, Y., y Karamé, F. (2013). Can Google data help predict French youth unemployment? *Economic Modelling*, 30, 117-125. <https://doi.org/10.1016/j.econmod.2012.07.017>
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., y Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014. <https://doi.org/10.1038/nature07634>
- González-Fernández, M., y González-Velasco, C. (2019). An approach to predict Spanish mortgage market activity using Google data. *Economics and Business Letters*, 8(4), 209. <https://doi.org/10.17811/eb1.8.4.2019.209-214>
- Huixin, B. (2012). Sovereign default risk premia, fiscal limits, and fiscal policy. *European Economic Review*, 56(3), 389-410. <https://doi.org/10.1016/j.euroecorev.2011.11.001>
- Jain, A. K. (2001). Corruption: A Review. *Journal of Economic Surveys*, 15(1), 71-121. <https://doi.org/10.1111/1467-6419.00133>

- Johnson, S., Kaufmann, D., Shleifer, A., Goldman, M. I., y Weitzman, M. L. (1997). *The Unofficial Economy in Transition*. Brookings Papers on Economic Activity. <https://www.brookings.edu/bpea-articles/the-unofficial-economy-in-transition/>
- Jun, S. P., Yoo, H. S., y Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change*, 130, 69-87. <https://doi.org/10.1016/j.techfore.2017.11.009>
- Kalbaska, A., y Gatkowski, M. (2012). Eurozone sovereign contagion: Evidence from the CDS market (2005–2010). *Journal of Economic Behavior & Organization*, 83(3), 657-673.
- Lee, C. M. C., y Ng, D. (2009). Corruption and International Valuation: Does Virtue Pay? *The Journal of Investing*, 18(4), 23-41. <https://doi.org/10.3905/joi.2009.18.4.023>
- Lyddy, F., Farina, F., Hanney, J., Farrell, L., y Kelly O'Neill, N. (2014). An Analysis of Language in University Students' Text Messages. *Journal of Computer-Mediated Communication*, 19(3), 546-561. <https://doi.org/10.1111/jcc4.12045>
- Mauro, P. (1995). Corruption and growth. *The Quarterly Journal of Economics*, 110(3), 681-712.
- McLaren, N., y Shanbhogue, R. (2012). Using Internet Search Data as Economic Indicators. *SSRN Electronic Journal*, 1, 134-140. <https://doi.org/10.2139/ssrn.1865276>
- Mo, P. H. (2001). Corruption and Economic Growth. *Journal of Comparative Economics*, 29, 66-79.
- Panizza, U., Sturzenegger, F., y Zettelmeyer, J. (2009). The economics and law of sovereign debt and default. *Journal of Economic Literature*, 47(3), 651-698. <https://doi.org/10.1257/jel.47.3.651>
- Paserman, M. (2017). Comovement or safe haven? The effect of corruption on the market risk of sovereign bonds of emerging economies during financial crises. *Journal of International Money and Finance*, 76, 106-132. <https://doi.org/10.1016/j.jimonfin.2017.06.001>
- Sgherri, S., y Zoli, E. (2009). Euro area sovereign risk during the crisis. *IMF Working Paper*, 09(222).
- Shleifer, A., y Vishny, R. W. (1993). Corruption. *Quarterly Journal of Economics*, 108(3), 599-617.

- Solano, P., Ustulin, M., Pizzorno, E., Vichi, M., Pompili, M., Serafini, G., y Amore, M. (2016). A Google-based approach for monitoring suicide risk. *Psychiatry Research*, 246, 581-586. <https://doi.org/10.1016/j.psychres.2016.10.030>
- Stephens-Davidowitz, S. (2020). *Google Searches Can Help Us Find Emerging Covid-19 Outbreaks* - *The New York Times*. <https://www.nytimes.com/2020/04/05/opinion/coronavirus-google-searches.html>
- Stephens-Davidowitz, S., y Varian, H. (2015). A Hands-on Guide to Google Data. *Google, Inc.*, 1-25.
- Svensson, J. (2005). Eight questions about corruption. *Journal of Economic Perspectives*, 19(3), 19-42. <https://doi.org/10.1257/089533005774357860>
- Transparency International. (2019). *Transparencia Internacional España*. <https://transparencia.org.es/>
- Treisman, D. (2000). The causes of corruption: a cross-national study. En *Journal of Public Economics* (Vol. 76). www.elsevier.nl/locate/econbase
- Vargas, M., y Sommer, F. (2014). *Corruption and the risks of losses on government bonds*. April.
- Villoria Mendieta, M., y Jiménez Sánchez, F. (2012). La corrupción en España (2004-2010): datos, percepción y efectos. *REIS: Revista Española de Investigaciones Sociológicas*, 138, 109-134.
- Vozlyublennai, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking & Finance*, 41, 17-35. <https://doi.org/10.1016/j.jbankfin.2013.12.010>
- Winkler, M. (1933). *Foreign Bonds: An Autopsy* (BeardBooks). https://books.google.es/books?hl=es&lr=&id=6n1G7loyyAYC&oi=fnd&pg=PR9&ots=I5i6tDqFSe&sig=0LIS6tUETdimn5GAi4i5hQ7cVrg&redir_esc=y#v=snippet&q=delos&f=false