Title papge (with autor details)

# A SENTIMENT INDEX TO MEASURE SOVEREIGN RISK USING GOOGLE DATA

Marcos González-Fernández, Carmen González-Velasco*

*Departamento de Dirección y Economía de la Empresa, Facultad de Ciencias Económicas y Empresariales, Universidad de León, Campus de Vegazana s/n, 24071, León, España.*

*e-mail:* carmen.gvelasco@unileon.es; *e-mail*: mgonf@unileon.es

* Corresponding author. Tel. +34 987 291738

E-mail address: carmen.gvelasco@unileon.es

ABSTRACT:

The aim of this paper is to construct an index that reflects investor sentiment regarding sovereign debt markets and to analyze this index to predict the evolution of sovereign risk. This Google Sovereign-Risk Sentiment Index (GSSI) is constructed by aggregating Google search data for a set of keywords related to the sovereign debt crisis that took place in Europe. The results indicate that the GSSI shows a high correlation with other sovereign risk indexes. Moreover, we analyze through panel data regressions its relationship with sovereign Credit Default Swaps (CDSs) for a set of European countries in the period 2008-2017. We determine that the GSSI shows the expected positive relationship with sovereign risk, especially in peripheral countries and during the period of maximum financial distress in sovereign debt markets. Our findings contribute to the investor sentiment literature and provides a novel measure for sovereign risk, which has emerged as one of the main challenges to global financial stability.

**Keywords:** Sovereign risk; Google data; internet activity; investor sentiment; sovereign debt crisis.

**JEL Classification:** G10, G17, G40.

**Funding:** This work was supported by the Ministerio de Economía, Industria y Competitividad, Gobierno de España [project number ECO2017-89715-P], entitled "El Análisis del Riesgo en los Mercados Financieros"].

# A SENTIMENT INDEX TO MEASURE SOVEREIGN RISK USING GOOGLE DATA

## ABSTRACT

The aim of this paper is to construct an index that reflects investor sentiment regarding sovereign debt markets and to analyze this index to predict the evolution of sovereign risk. This Google Sovereign-Risk Sentiment Index (GSSI) is constructed by aggregating Google search data for a set of keywords related to the sovereign debt crisis that took place in Europe. The results indicate that the GSSI shows a high correlation with other sovereign risk indexes. Moreover, we analyze through panel data regressions its relationship with sovereign Credit Default Swaps (CDSs) for a set of European countries in the period 2008-2017. We determine that the GSSI shows the expected positive relationship with sovereign risk, especially in peripheral countries and during the period of maximum financial distress in sovereign debt markets. Our findings contribute to the investor sentiment literature and provide a novel measure of sovereign risk. These results suggest several implications for public authorities and regulators.

**Keywords**: Sovereign risk; Google data; internet activity; investor sentiment; sovereign debt crisis.

**JEL Classification:** G10; G17; G40.

## 1. INTRODUCTION

The internet has become the most important source of information worldwide. Most economic or financial decisions are preceded by a search for information on web browsers, as noted by researchers. The number of papers examining how information searches affect economic and financial variables and assets has increased sharply over the past decade in all fields of academic literature. Hence, almost all areas of research, from medicine to economics and finance, have focused on the internet's ability to predict the evolution of other variables.

Following this trend, in this paper a sentiment index is constructed to serve as a proxy for sovereign risk based on Google data and analyze the ability of internet activity to predict the evolution of sovereign debt market indicators in European countries. It is well known that over the past few years, Europe has faced a critical sovereign debt crisis that has had a great impact on those indicators, specifically, sovereign bond yields, risk premiums and Credit Default Swaps (hereafter, CDSs). This event has particularly impacted peripheral countries, such as Greece, Ireland, Italy, Portugal and Spain, and sovereign risk is a major and pressing issue in the European Union (Panetta and Davies, 2011; Banque de France, 2012; Czech National Bank, 2012/2013; BIS, 2013; Brůha and Kočenda, 2018; Kočenda, 2018).

With this paper which, to the best of our knowledge, is pioneering work in this line of research, along with the paper by Dergiades et al. (2015), we attempt to determine whether internet activity is a useful proxy for investor sentiment regarding sovereign risk. If it is, it will represent a valuable tool that acts as a signal for ups and downs in sovereign debt market indicators and is helpful for financial market participants. To achieve this objective, we use what has been called Google econometrics (Fondeur and Karamé, 2013) to create an index to serve as a proxy for investors' sentiment in sovereign debt markets. Google is undoubtedly the main search engine worldwide. It encompasses more than 90% of all search engine users in the world and registers billions of searches every day. Google econometrics refers to the data obtained from the Google Trends tool. This instrument provides indexed data regarding the number of queries with specific keywords in a certain period. Thus, in relation to the aim of our paper, it is possible to measure internet attention with queries that include keywords such as "*sovereign debt crisis"*, "*European debt crisis, "debt crisis"* and similar words and to analyze whether those queries impact sovereign debt market indicators.

Our paper differs from previous papers which have addressed this topic in several ways. Specifically, the contributions of our paper are threefold. First, we build a Google Sovereign-Risk Sentiment Index (GSSI hereafter) in a similar fashion to that which Da et al. (2015) provided for the stock market. Using a sample of 50 keywords related to the sovereign debt crisis, we select those that better reflect investor sentiment regarding sovereign debt markets. Then, we reduce the selected keywords to a single index which reflects sovereign risk fear. Moreover, we compare the GSSI to other indexes that traditionally reflect sovereign risk.

Second, our paper provides additional evidence to the investor sentiment literature (Barberis et al., 1998), which states that investor sentiment helps to explain price movements beyond macro and traditional variables (Gao and Süss, 2015). In this sense, sentiment indexes have been constructed for commodities (Gao and Süss, 2015) and equities (Da et al., 2015; Xu and Zhou, 2018). Here, we present a new Google-based sentiment index for sovereign risk, which, to the best of our knowledge, is the first of its kind. This sentiment index has the appeal of being more transparent than traditional measures (Da et al., 2015) since it reveals real attitude and attention towards sovereign risk. Moreover, the GSSI allows to proxy investor sentiment with high frequency data, unlike traditional confidence indexes based on surveys.

Third, we run a set of panel data regressions using the GSSI as an explanatory variable of sovereign CDSs for a sample of European countries between 2008 and 2017. The results indicate that the GSSI is highly correlated with other traditional indexes that serve as proxies for sovereign risk. Therefore, the GSSI is a helpful variable, which indicates that Google data are a useful proxy for investor sentiments regarding sovereign debt markets. In addition, as a summary of the panel data regressions indicates, we can state that there is a positive relationship between the GSSI and sovereign risk that is more intense for peripheral countries as well as during periods of financial distress.

The paper unfolds as follows. Section 2 presents the theoretical background on the topic. Section 3 presents the data and methodology used in the study. In section 4, we construct the GSSI. Section 5 presents the panel data regression analysis. Finally, section 6 concludes the paper.

## 2. BACKGROUND LITERATURE

This paper aims to contribute to the investor sentiment literature with the construction of a sentiment index for sovereign risk. According to Da et al. (2015), investor sentiment has been generally measured through market-based variables (Baker and Stein, 2004; Baker and Wurgler, 2006, 2007) or survey-based indexes about consumer concerns (Qiu and Welch, 2006). However, recent literature has focused on the ability of other variables, such as news, Twitter or internet activity to measure investor sentiment (Bollen et al., 2011; Da et al., 2011, 2015; Smales, 2016; Milas et al., 2018).

Here, we use internet search activity, specifically Google data, to measure investor sentiment. The use of Google data as a valuable tool for research is relatively recent in the literature[1], and the interest in this topic among both academics and professionals has been growing over the past decade. During these last years, Google data have gained popularity in economics and finance as a measure of investor sentiment (Ben-Rephael et al., 2017; Da et al., 2011, 2015; Joseph et al., 2011; Siganos, 2013, inter alia). The question that arises here is why Google data should provide information about investor sentiment. Howard and Sheth (1969) shed light on this question through the theory of buyer behavior. According to that theory, people search for information before they purchase an asset. Obviously, that theory arose before the Google era, but the idea remains. In several paper decades later, Barber and Odean (2001; 2008) suggest that investors are buyers of attention-grabbing stocks. In addition, they state that investors are usually reluctant to pay for financial advice and are therefore more likely to use free internet information to drive their decision-making.

In the economics area, the first work on this topic (to the best of our knowledge) is that of Mondria et al. (2010), who use data from America Online to measure investor attention. One of the first studies to use Google data is Da et al. (2011), who use internet activity as a proxy for investor sentiment in the stock market. They use Google data to measure investor attention in the US stock market for more than 3,000 US companies between 2004 and 2008. They use ticker companies as keywords for Google queries, since they argue that company names can be searched for reasons other than investment ones[2]. Their

---

[1] The first examples of the use of Google data as a main variable are not in economics or finance but in medicine. Ginsberg et al. (2009) use Google data, extracted from the Google Trends tool, related to queries that include the word influenza. They attempt to predict influenza epidemics in the United States in the 2003-2008 period. Their results reveal a positive correlation between the number of queries for influenza on a given day and the number of medical center visits the next day, when patients show symptoms of the disease. More recently, in the field of psychiatry, Solano et al. (2016) analyze how Google queries for the keyword *suicide* in Italy during the 2008-2014 period anticipate suicide by three months.

[2] The selection of keywords is a crucial element in using Google data. For instance, people who search for the word *apple* may be referring to the fruit or the company. Therefore, the use of the *AAPL* ticker removes, or at least reduces, that noise in the query.

findings indicate that Google data are a direct measure of investor sentiment that can replace other indirect measures, such as turnover, extreme returns, news or advertising expenses, as this new measure is correlated with them. They also find that increases in Google data predict higher stock returns in the following two weeks. Another example for the stock market includes Vlastakis and Markellos (2012), who analyze the impact of investor sentiment, using Google data, on the stock volatility of 30 US companies from 2004 to 2009. They find that the demand for information, which is proxied by Google data, is related to volatility. In a further contribution considering stocks, Ben Rephael et al. (2017) analyze the impact of abnormal institutional attention (AIA) based on news about selected stocks retrieved from Bloomberg terminals and compare this AIA with retail attention measured through Google searches. Their findings indicate that institutional attention reacts faster to major news events and leads retail attention. Moreover, Joseph et al. (2011) use ticker Google searches for companies to predict abnormal returns in the US stock market. They obtain evidence that increases in the search intensity for their keywords predict abnormal stock returns and trading volumes. Outside the US market, Moussa et al. (2017) in France and Bank et al. (2011) in Germany also obtain evidence of this relationship[3].

There are few papers that link investor sentiment, as measured through Google data, and sovereign risk. To the best of our knowledge, only two papers in the literature address

[3] In addition to stock markets, Google data have been used to predict unemployment. Examples include the studies of Fondeur and Karamé (2013) and D'Amuri and Marcucci (2017), in France and Italy, respectively, who find evidence that Google queries for unemployment-related words forecast unemployment. Furthermore, Google data have been widely used to measure investor sentiment in commodities markets, especially for oil and gold (Li et al., 2015; Peri et al., 2014; Vozlyublennaia, 2014, inter alia). With respect to the exchange markets, Smith (2012) analyzes the ability of Google data to forecast the volatility of several currencies. He finds that Google searches for the keywords *economic crisis*, *financial crisis* and *recession* are related to the volatility of the currencies. Therefore, internet activity can be considered a signal of market volatility, as stated by Dzielinski (2012), who argues that Google queries act as a proxy for market uncertainty and finds evidence of that by comparing Google data to other uncertainty proxies.

this subject. The first, Rose and Spiegel (2012), analyze dollar illiquidity during the global financial crisis. They use Google data for a selection of keywords related to the financial crisis, such as *crisis*, *financial* or *recession,* and check whether they can be used as a proxy for default risk to include in their model. To test that, they run a panel regression in which they use as a dependent variable the change in Fitch's rating for a given country and use Google data as the right-hand-side variable. They find a strong negative relationship between Google data and sovereign ratings. Therefore, an increase in the number of Google queries results in a downgrade of sovereign debt. The authors thus conclude that Google data are a good proxy to track changes in sovereign risk.

The second paper to address this issue is by Dergiades et al. (2015) [4]. They analyze the impact of Google searches on financial markets, focusing on Europe's peripheral countries (Greece, Italy, Ireland, Portugal and Spain) and two core countries (France and the Netherlands). They narrow the period of analysis using the periods in which "*Greece crisis*" and "*Greek debt crisis*" were most searched. They focus on the period between May 2011 and May 2013. Then, they collect the data from Google using combinations of keywords related to sovereign debt crisis using quotation marks to avoid the contamination of searches[5]. Their results indicate that there is short-term causality between Google searches and sovereign spreads but only for Greece and Ireland; there is no impact on core countries (France and the Netherlands).

[4] See also Milas et al. (2018). In their paper, the authors do not study the effects of Google data on sovereign risk but the effects of Twitter and traditional news. Their results indicate that those variables impact not only sovereign spreads, especially in Greece, but also Portuguese and Irish bonds.

[5] For instance, when we search the keywords "*Greece debt crisis*" with quotations, we will obtain results in which those words have been searched together in that particular sequence. If we do not use quotations, we will obtain any sequence or combination of the words.

Regarding the construction of sentiment indexes, several papers have also addressed this issue. Gao and Süss (2015) construct a sentiment index between 1996 and 2013 for commodities using several sentiment proxies from Baker and Wurgler (2006, 2007) to act as a proxy for investor sentiment. Their results indicate that investor sentiment has a strong impact on commodities futures returns, which is stronger for negative events. In a similar fashion, Xu and Zhou (2018) build their sentiment index in the period between February 2015 to March 2017 and determine that changes in sentiment positively impact stock portfolios returns. Regarding sentiment indexes based on Google data, Da et al. (2015) construct the FEARS index for the period between January 2004 and December 2011 based on Google searches for a sample of keywords that reflect investor sentiment. This FEARS index exhibits a positive impact on stock returns. Moreover, Bampinas et al. (2019) create a Google-based index for two commodities: gold and oil. They use keywords related to those commodities for the period from October 2004 to October 2014. Then, they analyze the impact of their index on the conditional volatility of gold and oil. Their findings suggest that Google searches increase the volatility of both commodities. Following this line of research, this paper attempts to contribute to the literature by constructing a sentiment index for sovereign risk based on Google data.

## 3. DATA AND METHODOLOGY

The main variable in our analysis is the data related to Google searches. Google data are obtained from the Google Trends tool (https://www.google.com/trends). The literature usually employs Google rather than other search engines, since it is the main search website in the world (representing about 90% of global internet queries), and it provides data that have been successfully used as a predictor of economic indicators (Gomes and Taamouti, 2016). Specifically, Google Trends calculates the number of searches for a keyword or combination of keywords input by a user in a specific time period in relation to the total number of searches conducted on Google in the same period (Jun et al., 2016). This Google service started in 2006, although the data are available dating back to 2004. The frequency of the data ranges from daily to monthly depending on the time span selected. Data are available on a daily basis for the past 90 days, a weekly basis for the past five years, and a monthly basis for time horizons longer than that.

It is worth noting that Google Trends does not provide the total number of searches for keywords; rather, it provides an index that ranges from 0 to 100, which is usually called Google Search Volume Index (GSVI hereinafter). To build this index, Google starts by dividing the number of searches for a given keyword into the total number of searches for a given time unit (daily, weekly or monthly). Thereby, a ratio is obtained that is subsequently normalized by multiplying it by a scaling factor $F=100/r*$, where r* is the fraction of highest value (Dergiades et al., 2015). Thus, the numbers start at 0 on January 2004, and subsequent values denote changes from the search on that date (Jun et al., 2016), 100 being the point at which the number of queries has achieved the top search intensity. That is, the higher the value of the index, the larger the number of people looking for those terms. Hence, GSVI at time *t* can be denoted as follows:

$$GSVI_t = \frac{N_{j,t}^{s}}{N_{g,t}^{s}} \times \frac{100}{r^*} \tag{1}$$

where $N_{k,t}^{s}$ is the number of searches for a given keyword (j) at time *t*, and $N_{g,t}^{s}$ is the global number (*g*) of searches (*s*) in the same time unit and *r** is the highest value of the GSVI. With this normalization approach, Google reduces the noise and bias in the results that could arise when absolute values are used because, in that case, an increase in the total number of queries could be due to the increase in internet traffic. Furthermore, Google takes into account the fact that users might search for the same keyword repetitively, so those behavior are removed from the data to avoid manipulation (Jun et al., 2016). This GSVI can be narrowed by filtering the information by category or region. The category filter allows us to limit the searches to those in the category or subcategory that we select[6]. The region filter lets us obtain data for the countries or even regions within each country that we choose. In this case, language is another factor to take into account; depending on the geographic area[7] we are interested in the selection of the language may be a key element[8].

Another important aspect to take into account when using the GSVI is that the data can fluctuate depending on the date on which they are obtained (Bampinas et al., 2019; Carrière-Swallow and Labbé, 2013; McLaren and Shanbhogue, 2011). The reason for this is that the denominator from equation 1 is a random sample for the global number of Google searches that have been conducted, which are stored for just one day. Accordingly, when we download the data from the Google Trends tool during the same day, the results should not vary[9]. However, since that random sample changes each day, the same query for the same keyword or combination of keywords in different days can slightly differ. This difference is usually almost residual and does not cause a big change in the data.

Clearly, the selection of the keywords is critical. To do so, we have considered several terms that intuitively reflect investor attention to the European sovereign debt crisis, such

[6] For instance, within the finance category, there are subcategories such as banking, accounting, investment, and insurance, and each of those subcategories also contains subcategories. Obviously, the more we narrow the search, the more difficult it is to obtain results because the number of searches will be smaller, and if there is an insufficient number of searches Google Trends does not provide any data.
[7] We have considered *worldwide* as the geographical area from which we obtain Google data.
[8] In our paper, we assume that English is commonly employed as a universal language in the context of Europe. We have considered other languages such as Spanish or French, but the data series obtained from Google trends were scarce and the amount of data was not sufficient to perform an analysis. As Dergiades et al. (2015) suggest, this scarcity of data for other languages is either due to the fact that English is the prevailing language in Europe or that the majority of web users worried about these keywords know English and perform the queries in that language.
[9] We have even noticed minimal intraday changes in Google data.

as *debt crisis*, *sovereign debt* or *European crisis.* We refer to these as primitive keywords. Then, Google provides the top related searches for these primitive keywords. The top related searches offer information about how users search for information[10] (Da et al., 2015). Subsequently, we collect all the related keywords and remove those that are duplicated or do not offer enough observations. Finally, this procedure results in a list of 50 final keywords, which are summarized in Table 1 along with their descriptive statistics. The GSVI data were collected from January 2008 to December 2017 on a weekly basis[11], which resulted in 520 observations per series. The ADF tests are negative for all the terms, indicating that the series are stationary.

As we have highlighted above, the GSVI can slightly differ based on the day on which we download the data. To address this issue, we gather the GSVI data for several days in order to have more accurate data that reflect those changes. Specifically, we collect the data on February 15th, 16th and 17th, 2019 and we obtain the GSVI average for those three dates:

$$GSVI_{average} = \frac{\sum_{t=1}^{3} GSVI_t}{3} \tag{2}$$

Figure 1 shows an example of monthly data of the GSVI for the primitive keyword *European crisis*[12] along with the Portuguese 10-year bond sovereign yields since Portugal is one of the countries that suffered the sovereign debt crisis to a greater extent. The right

[10] Top related searches indicate that some users who search for the primitive keywords also have searched for any of the related searches.
[11] We use weekly data for several reasons. First, because high-frequency data (weekly or daily) are more appropriate to capture investor sentiment than low-frequency date (monthly or quarterly). Second, because Google daily data is only available for a 3-month span, so each of those data files contains a relative peak that reaches the value 100 according to how Google calculate the GSVI. Thus, it implies that we should concatenate a big number of blocks of data, each of them containing a peak, which would introduce jumps and noise in the data. Weekly data can be obtained for a 5-year span. Therefore, using weekly data we only consider three blocks of data, which reduces this bias.
[12] The GSVI data displayed in Figure 1 were downloaded on 15th February 2019.

axis shows the yields in percentage, and the left axis shows the GSVI ranging from 0 to 100.

Table 1. List of 50 keywords related to sovereign debt crisis and descriptive statistics

| Keywords | Mean | Std. Dev. | Kurtosis | Skewness | JB | ADF |
|---|---|---|---|---|---|---|
| *Bloomberg* | 49.54 | 13.76 | 3.08 | 0.49 | 21.34 | -6.387 |
| *Debt crisis* | 8.26 | 9.72 | 49.61 | 5.93 | 5,000 | -7.566 |
| *Euro* | 14.29 | 10.55 | 39.66 | 5.69 | 3,200 | -5.545 |
| *Euro crisis* | 13.94 | 14.77 | 11.68 | 2.57 | 2,211 | -4.793 |
| *Europe* | 70.74 | 9.65 | 3.08 | -0.13 | 1,856 | -4.982 |
| *Europe crisis* | 17.08 | 16.53 | 9.53 | 2.33 | 1,397 | -4.993 |
| *Europe debt crisis* | 18.27 | 17.48 | 6.33 | 1.58 | 458.0 | -7.960 |
| *European bank crisis* | 18.99 | 14.15 | 5.13 | 1.07 | 199.8 | -15.25 |
| *European countries* | 43.17 | 18.44 | 2.00 | 0.43 | 38.02 | -4.095 |
| *European crisis* | 19.05 | 14.78 | 11.49 | 2.48 | 2,099 | -4.865 |
| *European crisis 2010* | 8.90 | 13.45 | 11.45 | 2.45 | 2,074 | -14.51 |
| *European crisis 2011* | 14.47 | 18.93 | 5.07 | 1.53 | 296.2 | -13.13 |
| *European debt* | 19.61 | 14.49 | 8.68 | 1.75 | 966.7 | -5.875 |
| *European debt crisis* | 18.03 | 17.19 | 6.49 | 1.55 | 475.6 | -5.506 |
| *European debt crisis 2011* | 9.67 | 15.99 | 6.99 | 1.93 | 669.9 | -16.37 |
| *European debt crisis news* | 6.68 | 12.26 | 17.45 | 3.20 | 5,398 | -13.38 |
| *European economic crisis* | 29.39 | 18.65 | 3.54 | 0.79 | 61.9 | -9.417 |
| *European economy* | 45.19 | 13.56 | 4.09 | 0.49 | 46.6 | -10.58 |
| *European financial crisis* | 25.12 | 14.97 | 6.07 | 1.11 | 312.7 | -10.55 |
| *European sovereign debt* | 20.52 | 17.22 | 4.60 | 1.03 | 149.2 | -10.68 |
| *European sovereign debt crisis* | 21.25 | 19.15 | 4.24 | 1.08 | 134.7 | -10.03 |
| *European Union* | 34.53 | 29.50 | 1.62 | 0.38 | 54.12 | -3.019 |
| *European Union crisis* | 28.73 | 17.11 | 3.73 | 0.84 | 73.42 | -9.548 |
| *European Union financial crisis* | 18.34 | 16.34 | 3.50 | 0.86 | 71.11 | -20.15 |
| *Greece* | 31.53 | 12.88 | 6.10 | 1.18 | 330.6 | -4.958 |
| *Greece crisis* | 4.00 | 9.02 | 61.06 | 6.66 | 7,700 | -7.912 |
| *Greece sovereign debt* | 11.48 | 14.05 | 9.26 | 2.12 | 1,245 | -10.77 |
| *Greek sovereign debt* | 14.63 | 15.96 | 7.15 | 1.70 | 652.9 | -11.44 |
| *Greek sovereign debt crisis* | 12.94 | 15.81 | 4.58 | 1.29 | 198.5 | -14.64 |
| *Irish sovereign debt* | 6.43 | 14.37 | 10.85 | 2.73 | 1.982 | -19.35 |
| *S&P sovereign debt rating* | 6.39 | 13.60 | 11.70 | 2.76 | 2,308 | -19.47 |
| *S&P sovereign debt ratings* | 5.74 | 14.78 | 11.68 | 2.94 | 2,388 | -21.85 |
| *Sovereign bank* | 48.07 | 29.55 | 1.47 | -0.32 | 59.01 | -1.355 |
| *Sovereign bonds* | 20.83 | 12.89 | 7.88 | 1.56 | 727.9 | -12.86 |
| *Sovereign debt* | 34.96 | 18.84 | 3.16 | 0.43 | 16.81 | -6.722 |
| *Sovereign debt by country* | 13.31 | 17.26 | 5.17 | 1.48 | 293.1 | -17.29 |
| *Sovereign debt crisis* | 27.37 | 19.17 | 3.58 | 0.66 | 46.03 | -7.418 |
| *Sovereign debt crisis definition* | 11.50 | 16.13 | 4.52 | 1.42 | 225.6 | -19.24 |
| *Sovereign debt default* | 25.36 | 16.68 | 3.89 | 0.71 | 61.97 | -17.18 |
| *Sovereign debt definition* | 20.25 | 16.46 | 3.75 | 0.77 | 63.94 | -16.39 |
| *Sovereign debt rating* | 17.80 | 15.56 | 5.29 | 1.29 | 258.9 | -14.60 |
| *Sovereign debt ratings* | 15.15 | 14.27 | 6.05 | 1.34 | 359.4 | -18.18 |
| *Sovereign definition* | 45.33 | 18.43 | 2.44 | 0.28 | 13.37 | -8.191 |
| *Sovereign risk* | 40.19 | 16.65 | 2.74 | 0.35 | 12.15 | -11.59 |
| *The European crisis* | 28.61 | 16.71 | 4.67 | 1.09 | 165.0 | -6.502 |
| *The European debt crisis* | 16.28 | 15.22 | 7.87 | 1.68 | 760.7 | -7.329 |
| *UK sovereign debt* | 15.37 | 16.53 | 3.55 | 1.01 | 95.9 | -18.12 |
| *US sovereign debt* | 15.79 | 14.64 | 6.35 | 1.40 | 416.6 | -17.27 |
| *What is sovereign debt* | 20.82 | 16.40 | 3.74 | 0.80 | 68.24 | -13.87 |
| *What is sovereign debt crisis* | 10.92 | 14.43 | 5.53 | 1.47 | 329.3 | -18.50 |

The table shows, in alphabetical order, the list of the 50 keywords related to the European sovereign debt crisis that were used to build the GSVI. Following Han et al. (2017), this table shows the following descriptive statistics of these keywords: the mean, standard deviation, skewness, kurtosis, Jarque Bera and Augmented Dickey Fuller.

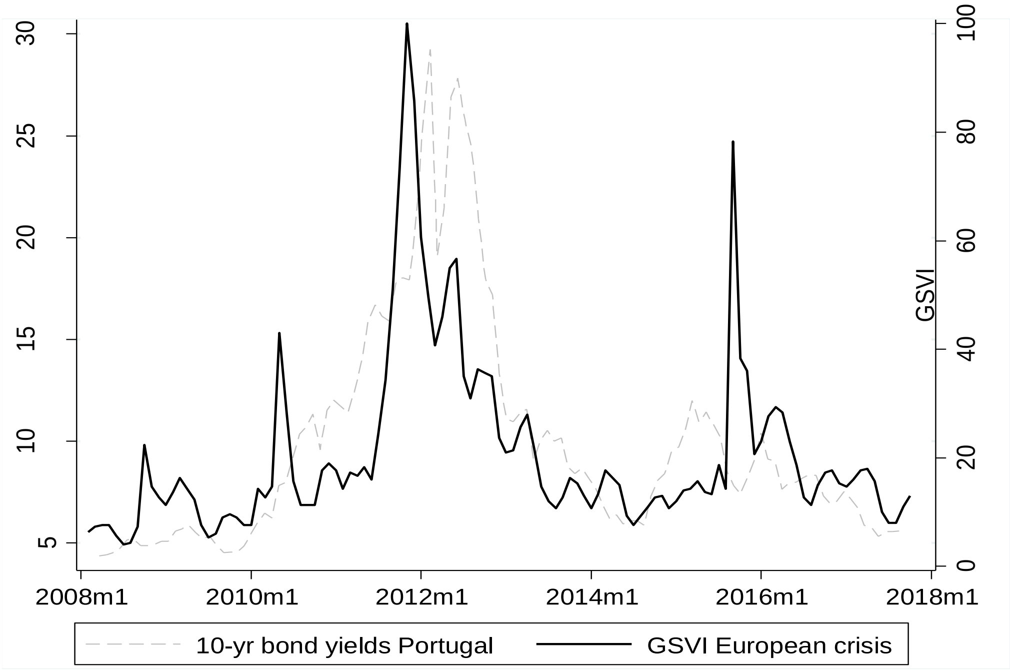

Figure 1. Evolution of GSVI for *European crisis* and Portuguese 10-year sovereign bond yields

There appears to be a similar evolution for both series. The GSVI reaches its highest point surrounding the rescue packages that Portugal and Greece received during 2011 and 2012, which also represents the highest peak in Portuguese sovereign bond yields.

Regarding sovereign risk, we have considered several indicators. We have discarded sovereign ratings from rating agencies (Remolona et al., 2007) since they are low frequency indicators, and we have focused on sovereign CDSs (Ammer and Cai, 2011; Srivastava et al., 2016), which are commonly employed in the literature. Specifically, we use weekly 5-year CDS data obtained from Thomson Reuters Datastream[13]. We have gathered the data for 11 of the most important economies in Europe: Austria, Belgium, Denmark, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain and the United Kingdom[14]. The time horizon extends from January 2008 to December 2017, just as the GSVI data does.

[13] We use 5-year CDSs in the main analysis of the paper since CDSs are the most reliable variable to measure sovereign risk (Ejsing and Lemke, 2011).
[14] We have not included Greece since CDS data obtained from Thomson Reuters Datastream for the Hellenic country present large gaps in the series.

## 4. GOOGLE SOVEREIGN-RISK SENTIMENT INDEX (GSSI)

In this section, we build the GSSI. For this purpose, in a similar fashion as in Da et al., (2015) or Han et al., (2017), we follow several steps. First, we winsorize each $GSVI_{average}$ series at the 5% level (2.5% in each tail) to avoid extreme values or outliers. Then, we identify which of our keywords are more related to sovereign risk and better reflect investor sentiment. For this purpose, we run backward rolling regressions[15] of our $GSVI_{average}$ for the 50 keywords on Euro sovereign 5Y CDS Index (Eq. 3) to determine the historical relationship between them. This index, built by Thomson Reuters Datastream, shows the evolution of an equally weighted portfolio of European CDSs[16]. We fix the window (L) for the backward rolling regressions to be L=52. Therefore, we first perform a regression from January 2008, week 1 to December 2008, week 52, and then we move both extremes one week ahead and repeat the procedure until the end of the series.

$$Euro\ sovereign\ 5Y\ CDS\ Index_t = \alpha + \beta \cdot GSVI_{average_{j,t}} + \epsilon_t \qquad (3)$$

where *j* represents the keywords ($j = 1, \dots 50$) and *t* is the window of 52 weeks. The next step is retaining those keywords whose $t$-statistic is positive and significant since we expect that an increase in internet activity, i.e., in the GSVI of our keywords, will lead to an increase in sovereign risk. The results show that the relationship for 48 out of the 50 keywords is positive, as expected. Thus, the GSSI is calculated as follows:

[15] According to Da et al., (2015), this historical-regression procedure is the most objective way to select the relevant keywords, as opposed to regular regressions. Nevertheless, we also use regular and expansive window regressions for robustness. The results are available upon request to the authors.

[16] According to Thomson Reuters, they base the index on the most liquid term, i.e., 5-year CDSs. The index is equally weighted and reflects an average mid-spread calculation of the index's constituents, which are eurozone countries.

$$GSSI \equiv \sum_{j=1}^{k} GSVI_{AVERAGE}{}_{j,t}^{*} \quad (4)$$

where $k$ represents the total number of positive and significant keywords, which are summarized in Table 2[17].

Therefore, the GSSI is calculated as the sum of the $GSVI_{AVERAGE}$ for the keywords selected following Eq. (4). Nevertheless, for robustness purposes, we have also built the GSSI performing a Principal Components Analysis (PCA hereafter) using the $GSVI_{AVERAGE}$ for those keywords and retaining the scores for the first component. It is worth noting that the pairwise correlation between the keywords in Table 2 retrieved on the aforementioned three dates (February, 15th, 16th and 17th) exceeds 95%. Therefore, we can assume that the differences in Google data, which are a result of the date we downloaded the series, are residual in our analysis.

Table 2. Keywords selected for the construction of the GSSI

| | Keywords | Coefficient (standard errors) | *t*-statistic |
|---|---|---|---|
| *1* | *Debt crisis* | 3.92 (1.47) | 2.67*** |
| *2* | *Euro crisis* | 3.25 (0.90) | 3.61*** |
| *3* | *Europe crisis* | 1.51(0.41) | 3.68*** |
| *4* | *European crisis* | 2.31 (0.55) | 4.20*** |
| *5* | *European debt crisis* | 6.27 (1.96) | 3.19*** |
| *6* | *European debt* | 1.93 (0.94) | 2.05** |
| *7* | *Greece crisis* | 9.05 (3.33) | 2.72*** |
| *8* | *Sovereign debt* | 1.50 (0.47) | 3.19*** |
| *9* | *The European crisis* | 1.05 (0.41) | 2.56** |

The table shows the keywords that show a positive *t*-statistic from the rolling regressions between the Euro sovereign 5Y CDS Index and the 50 keywords. For the rolling regressions, we have fixed the window L=52 weeks. Therefore, the initial regression starts in January 2008, week 1 and runs through December 2008, week 52. We then move both extremes one week ahead. This procedure results in 469 regressions for which *t*-statistic average values are shown. Average robust standard errors are shown in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3 displays the descriptive statistics and the pairwise correlations between the GSSI, based on Google data, and other indexes of sovereign risk. Specifically, as proxies for sovereign risk, we include the aforementioned Euro sovereign 5Y CDS Index, the

[17] The keywords *Europe crisis* and *sovereign debt* show a *t*-statistic above 1.5, and we have considered including them as a robustness measure since they are close to be significative.

Europe banks sector 5Y CDS Index[18] and the aggregated index of the CDSs from peripheral countries. In Panel B, our GSSI shows a high and significant correlation with these other indexes that measure sovereign risk. As shown in Panels C and D, this positive relationship remains stable regardless of whether the period is focused on the zenith of the sovereign debt crisis (Panel C) or when the situation stabilizes (Panel D).

Table 3. Summary statistics and correlations of the GSSI and other indexes of sovereign risk

| Panel A. Summary statistics | Count | Mean | Median | Range | Std. Dev |
|---|---|---|---|---|---|
| Euro sovereign 5Y CDS Index | 520 | 229.93 | 216.95 | 536,10 | 106,52 |
| Europe banks sector 5Y CDS Index | 520 | 106.52 | 75.74 | 406.87 | 88.36 |
| Peripheral countries' CDS aggregate | 491 | 942.80 | 573.78 | 2.973,4 | 750.86 |
| GSSI | 520 | 93.07 | 81.83 | 272.66 | 57.09 |
| GSSI (PCA) | 520 | 2.66e-09 | -0.36 | 10.51 | 2.11 |
| **Panel B. Pairwise correlations (January 2008-December 2017)** | (1) | (2) | (3) | (4) | (5) |
| Euro sovereign 5Y CDS Index (1) | 1.00 | 0.71*** | 0.82*** | 0.48*** | 0.49*** |
| Europe bank sector 5Y CDS Index (2) | | 1.00 | 0.88*** | 0.71*** | 0.69*** |
| Peripheral countries' CDS aggregate (3) | | | 1.00 | 0.66*** | 0.65*** |
| GSSI (4) | | | | 1.00 | 0.99*** |
| GSSI (PCA) (5) | | | | | 1.00 |
| **Panel C. Pairwise correlations (January 2008-December 2012)** | (1) | (2) | (3) | (4) | (5) |
| Euro sovereign 5Y CDS Index (1) | 1.00 | 0.80*** | 0.77*** | 0.74*** | 0.75*** |
| Europe bank sector 5Y CDS Index (2) | | 1.00 | 0.92*** | 0.82*** | 0.81*** |
| Peripheral countries' CDS aggregate (3) | | | 1.00 | 0.85*** | 0.84*** |
| GSSI (4) | | | | 1.00 | 0.99*** |
| GSSI (PCA) (5) | | | | | 1.00 |
| **Panel D. Pairwise correlations (January 2013-December 2017)** | (1) | (2) | (3) | (4) | (5) |
| Euro sovereign 5Y CDS Index (1) | 1.00 | 0.47*** | 0.89*** | 0.53*** | 0.50*** |
| Europe bank sector 5Y CDS Index (2) | | 1.00 | 0.68*** | 0.42*** | 0.39*** |
| Peripheral countries' CDS aggregate (3) | | | 1.00 | 0.44*** | 0.41*** |
| GSSI (4) | | | | 1.00 | 0.98*** |
| GSSI (PCA) (5) | | | | | 1.00 |

The table shows the summary statistics of the GSSI according to equation 4 and the GSSI built performing the PCA analysis (GSSI (PCA)). In addition to the GSSI, other measures of sovereign risk have been included: the Euro sovereign 5Y CDS Index, Europe banks sector 5Y CDS Index and the aggregated value of the peripheral countries' CDS data. The peripheral countries included are Ireland, Italy, Portugal and Spain, but we do not include Greece since there are no CDS data for this country in many dates. The correlations in Panel B cover the entire time horizon, i.e., from January 2008 to December 2017. Panel C displays the correlations during the period in which the financial and sovereign debt crises reached their peaks (January 2008 to December 2012). Panel D displays the correlations from January 2013 to the end of the sample, when the sovereign debt crisis relaxed.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

[18] This index, built by Thomson Reuters Datastream, is an average of European banks 5-year CDSs.

However, it is worth noting that, during peak financial distress (Panel C), the GSSI shows the highest correlation with the rest of measures with a pairwise correlation of 75% to the Euro sovereign 5Y CDS Index and above 80% to Europe bank sector 5Y CDS Index and peripheral countries' CDS aggregate. Graphically, Figure 2 displays the GSSI series along with the Euro sovereign 5Y CDS Index.

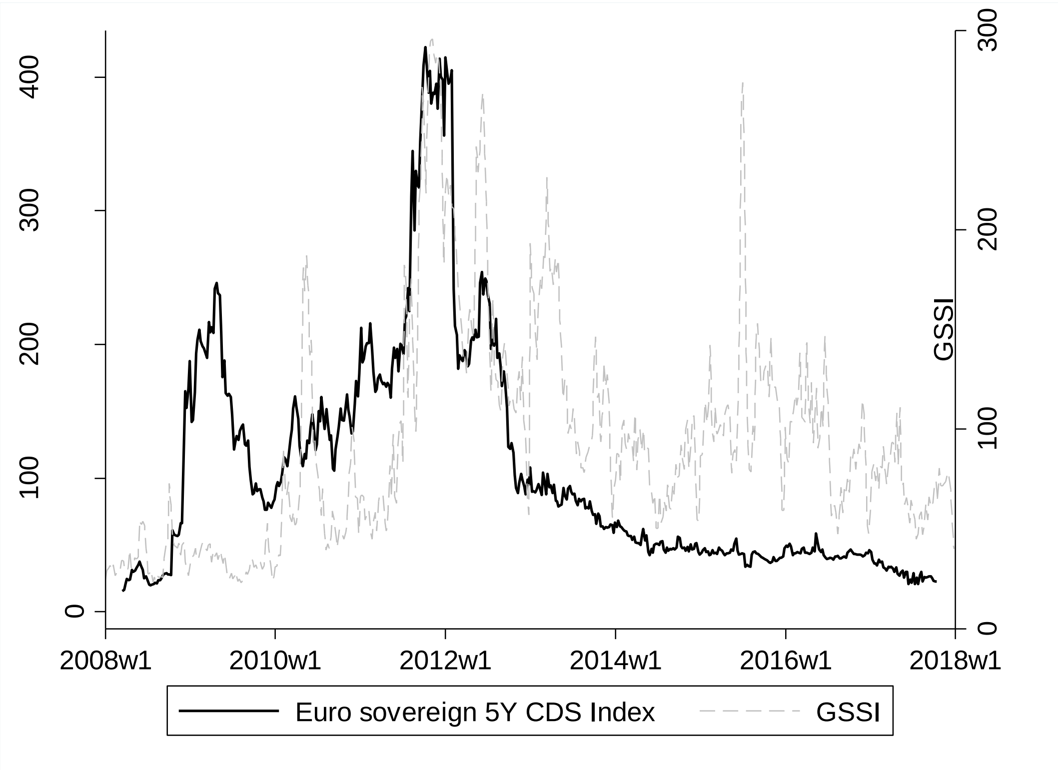


Figure 2. Evolution of GSSI and Euro sovereign 5Y CDS Index

## 5. PANEL DATA REGRESSION ANALYSIS

We analyze the ability of the GSSI to predict the evolution of sovereign risk, using sovereign CDSs as proxies. For this purpose, we run panel data regressions for a sample of European countries in which we have included four peripheral countries, i.e., Ireland, Italy, Portugal and Spain; five core countries, i.e., Austria, Belgium, France, Germany and the Netherlands; and two non-Euro countries, i.e., Denmark and the United Kingdom. We use as a dependent variable the 5-year sovereign CDSs and, as explanatory variables, in addition to the GSSI, the sovereign yields and the VIX index, obtained from Thomson Reuters Datastream. The baseline model specification is provided in Eq. (5):

$$CDS_{it} = \alpha + \beta \cdot GSSI_{it} + \gamma F_{it} + \epsilon_{it} \tag{5}$$

where *CDS* represents the sovereign CDSs; $i = 1, 2, \dots 11$ countries and $t = 1, 2, \dots 520$; *GSSI* is the Google-based sentiment index for sovereign risk; and *F* is a vector of financial control variables: sovereign bond yields and VIX as a proxy for volatility of international financial markets (Arghyrou and Kontonikas, 2012). Ex-ante expectations are for a positive relationship between the GSSI and sovereign CDSs; i.e., a positive coefficient for the GSSI should lead to a surge in sovereign CDSs. For the rest of the control variables, the same positive relationship is assumed.

Table 4 shows the results for the pooled Ordinary Least Squares (OLS) regressions and panel data fixed effects (FE) estimations[19]. In models 1 and 3, we use the GSSI as the index to serve as a proxy for sovereign risk, and in models 2 and 4, we include the GSSI constructed through PCA. In the light of the results, we can observe that both indexes display similar results, indicating a positive relationship between them and sovereign CDSs. Thus, the Google-based index, GSSI, seems to have the ability to influence sovereign risk.

[19] We perform panel data regressions with Driscoll-Kraay (1998) standard errors which suits better with the characteristics of our dataset, i.e., a large T and small N (Hoechle, 2007).

Table 4. Panel data regressions of the GSSI on sovereign CDSs

| Dependent variable: sovereign CDSs | OLS Model 1 | OLS Model 2 | FE Model 3 | FE Model 4 |
|---|---|---|---|---|
| GSSI | 0.228*** | | 0.227*** | |
| | (0.022) | | (0.051) | |
| GSSI (PCA) | | 7.320*** | | 7.281*** |
| | | (0.539) | | (2.199) |
| Sovereign yields | 90.34*** | 90.37*** | 91.51*** | 91.58*** |
| | (1.068) | (1.082) | (3.084) | (3.087) |
| VIX | 1.737*** | 1.929*** | 1.745*** | 1.936*** |
| | (0.119) | (0.125) | (0.545) | (0.579) |
| Constant | -427.43*** | -404.70*** | -432.81*** | -410.37*** |
| | (7.82) | (8.21) | (33.13) | (34.82) |
| N | 5,402 | 5,402 | 5,402 | 5,402 |
| Time dummies | YES | YES | YES | YES |
| $R^2$ | 0.912 | 0.910 | 0.876 | 0.874 |

The table displays the panel data regressions of GSSI and other control variables on sovereign CDSs for 11 European countries: Austria, Belgium, Denmark, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain and the United Kingdom. Models 1 and 2 show pooled OLS regressions, and models 3 and 4 show panel data fixed effects estimations. Robust standard errors are in parentheses.
*** $p$<0.01, ** $p$<0.05, * $p$<0.1.

However, a causal relationship between the explanatory variables and the left-hand side variable is likely to exist. Therefore, a dynamic approach to solve this endogeneity bias is required. For this purpose, an instrumental variables (IV) panel data technique is conducted. The values of the underidentification (Kleibergen-Paap) and overidentification (Hansen) tests are presented are presented at the bottom of Table 5. They all show the expected significance, indicating the validity of the instruments[20]. We also include annual time dummies in the models.

The results indicate a positive influence of the GSSI on sovereign CDSs, as expected. This result remains stable whether we include the GSSI; the GSSI constructed through PCA or the GSSI without winsorizing the data. Other variables, namely, the sovereign yields and the VIX Index, exhibit the expected positive and significant relationship to CDSs. Again, the GSSI develops as an actual sentiment index for investor attention for

[20] Given the endogeneity of the right hand side variables, we use several instruments to solve this issue. Specifically we instrument the VIX thorough the S&P 500 Index, since it is expected that it affects the VIX, but not the CDSs. For the sovereign yields, we use the macroeconomic uncertainty index from Jurado et al. (2015), since yields depend on fundamental conditions of the economy (Capelle-Blancard et al., 2019; Wellmann and Trück, 2018). For the GSSI we use Google correlate tool to find search terms that are related to our keywords but unrelated to CDSs.

sovereign risk. Thus, we demonstrate the ability of our GSSI to act as a determinant of sovereign CDSs.

Table 5. Instrumental variables panel data regressions of GSSI on sovereign CDSs

| Dependent variable: sovereign CDSs | IV Model 1 | IV Model 2 | IV Model 3 |
|---|---|---|---|
| GSSI | 0.555*** | | |
| | (0.048) | | |
| GSSI (PCA) | | 21.90*** | |
| | | (2.228) | |
| GSSI (no winsorization) | | | 0.469*** |
| | | | (0.046) |
| Sovereign yields | 56.36*** | 51.88*** | 44.64** |
| | (15.59) | (16.69) | (18.33) |
| VIX | 0.252 | 0.553** | 0.372 |
| | (0.246) | (0.245) | (0.270) |
| N | 5,402 | 5,402 | 5,402 |
| Time dummies | YES | YES | YES |
| Kleibergen-Paap (*p*-value) | 0.000*** | 0.000*** | 0.000*** |
| Hansen (*p*-value) | 0.263 | 0.131 | 0.393 |

The table presents the instrumental variables panel data regressions of the GSSI and other control variables on sovereign CDSs for 11 European countries: Austria, Belgium, Denmark, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain and the United Kingdom We consider the GSSI, sovereign yields and VIX as endogenous variables. All models include annual time dummies. The Kleibergen-Paap test is the underidentification test, under the null hypothesis that the model is underidentified. Therefore its rejection suggests the model is identified. Hansen test is the overidentification test under the null hypothesis that the instruments are valid. Therefore, a rejection casts doubt on the validity of instruments. Robust standard errors are in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

## 5.1. ROBUSTNESS TESTS

To be thorough in the analysis, we perform some robustness checks to test whether the results remain stable regardless of the sample of countries or the time horizon studied. First, we show the results for different groups of countries. Namely, we split up the sample into peripheral and core countries and Euro and non-Euro countries[21]. These results are shown in Table 6.

[21] In these robustness tests, we run the same instrumental variables methodology shown in Table 5 to avoid the endogeneity bias. To avoid an excessive number of tables, here we show the results only for the GSSI. The results for the GSSI without winsorizing the data, the GSSI constructed through PCA are similar than those for the GSSI. These results are available upon request to the authors.

The GSSI remains significant in all the models regardless of the sample of countries analyzed. It is worth noting that the coefficient's values indicate that in peripheral countries, the GSSI shows the greatest impact on sovereign CDSs, while non-Euro countries are no influenced by our index. This finding is in line with what we expected, since peripheral countries have suffered the sovereign debt crisis to a greater extent.

Table 6. Instrumental variables panel data regressions of GSSI on sovereign CDSs in peripheral and core countries and Euro and non-Euro countries

| Dependent variable: sovereign CDSs | Model 1<br>Peripheral | Model 2<br>Core | Model 3<br>Euro | Model 4<br>Non-Euro |
|---|---|---|---|---|
| GSSI | 0.632*** | 0.377*** | 0.576*** | -0.167 |
| | (0.105) | (0.031) | (0.063) | (0.261) |
| Sovereign yields | 86.14*** | -23.36 | 62.21*** | -106.83 |
| | (12.23) | (23.75) | (14.90) | (77.34) |
| VIX | -0.209 | 0.766 | 0.005 | 2.305*** |
| | (0.425) | (0.232) | (0.271) | (0.544) |
| N | 1,965 | 2,455 | 4,420 | 982 |
| Time dummies | YES | YES | YES | YES |
| Kleibergen-Paap (*p*-value) | 0.000*** | 0.004*** | 0.000*** | 0.611 |
| Hansen (*p*-value) | 0.755 | 0.093* | 0.596 | 0.143 |

The table presents the instrumental variables panel data regressions of GSSI and other control variables on sovereign CDSs for peripheral countries (Ireland, Italy, Portugal and Spain); core countries (Austria, Belgium, France, Germany and the Netherlands); Euro countries (Austria, Belgium, France, Germany Ireland, Italy, the Netherlands, Portugal and Spain) and non-Euro countries (Denmark and the United Kingdom). We consider the GSSI, sovereign yields and VIX Index to be endogenous variables. Annual time dummies are included in all the models. The Kleibergen-Paap test is the underidentification test, under the null hypothesis that the model is underidentified. Therefore its rejection suggests the model is identified. Hansen test is the overidentification test under the null hypothesis that the instruments are valid. Therefore, a rejection casts doubt on the validity of instruments. Robust standard errors are in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Continuing with the robustness tests, we perform some estimations splitting up the sample into two subsamples: the first one starting in January 2008 and ending in December 2012 and the second one starting in January 2013 and finishing in December 2017. Thus, the former encompasses the period of maximum financial distress, and the latter represents a more stable phase. Table 7 displays the results for these estimations.

The GSSI shows a higher coefficient during the period of maximum distress of the sovereign debt crisis, although it remains significant and positive during the stabilization phase. Therefore, the GSSI is a more accurate proxy for investor sentiment regarding sovereign risk during periods of maximum financial distress in sovereign debt markets.

In Table 6, we also present evidence that investor sentiment has greater impact in those countries more affected by the sovereign debt crisis. This asymmetric behavior is in line with the previous results reported in Table 3 and Figure 2 and is similar to previous research that has highlighted that the impact of investor attention is enhanced during negative events (Guo and Ji, 2013; Peri et al., 2014). In this sense, Smales (2016) indicates that the effects of news on bank credit risk, measured through bank CDSs, is negative, and even stronger, with negative news. Here, we show that, similarly, for sovereign risk, the GSSI better reflects investors' sentiment in the peak of the sovereign debt crisis and in highly distressed countries.

Table 7. Instrumental variables data regressions of GSSI on sovereign CDSs in different periods

| Dependent variable: sovereign CDSs | Model 1<br>2008-2012 | Model 2<br>2008-2012 | Model 3<br>2013-2017 | Model 4<br>2013-2017 |
|---|---|---|---|---|
| GSSI | 0.890*** | | 0.163*** | |
| | (0.136) | | (0.023) | |
| GSSI (PCA) | | 32.30*** | | 7.35*** |
| | | (5.530) | | (1.104) |
| Sovereign yields | -92.39 | -111.97* | 19.30*** | 18.46*** |
| | (59.20) | (67.68) | (2.766) | (2.889) |
| VIX | 0.372 | 0.776 | -0.095 | -0.167 |
| | (0.853) | (0.900) | (0.253) | (0.259) |
| N | 2,542 | 2,542 | 2,860 | 2,860 |
| Time dummies | YES | YES | YES | YES |
| Kleibergen-Paap (*p*-value) | 0.010** | 0.014** | 0.000*** | 0.000*** |
| Hansen (*p*-value) | 0.418 | 0.398 | 0.141 | 0.049** |

The table presents the instrumental variables panel data regressions of the GSSI and other control variables on sovereign CDSs for 11 European countries: Austria, Belgium, Denmark, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain and the United Kingdom. We consider the GSSI, sovereign yields and VIX index to be endogenous variables. Models 1 and 2 show the estimations for the period of January 2008 to December 2012. Models 3 and 4 show the estimations for the period of January 2013 to December 2017. Annual time dummies are included in all the models. The Kleibergen-Paap test is the underidentification test, under the null hypothesis that the model is underidentified. Therefore its rejection suggests the model is identified. Hansen test is the overidentification test under the null hypothesis that the instruments are valid. Therefore, a rejection casts doubt on the validity of instruments. Robust standard errors are in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Next, we construct the GSSI considering Google data from the finance category. Thus, we avoid any bias that could affect the results as to whether those searches in the general category were not representative of investor sentiment in financial markets[22]. For this purpose, we build the $GSSI_{FINANCE}$, taking into account the same keywords that for the GSSI, i.e., *debt crisis, Euro crisis, Europe crisis, European crisis, European debt crisis, European debt, Greece crisis, sovereign debt* and *the European crisis* but performing the searches within the finance category from Google trends.

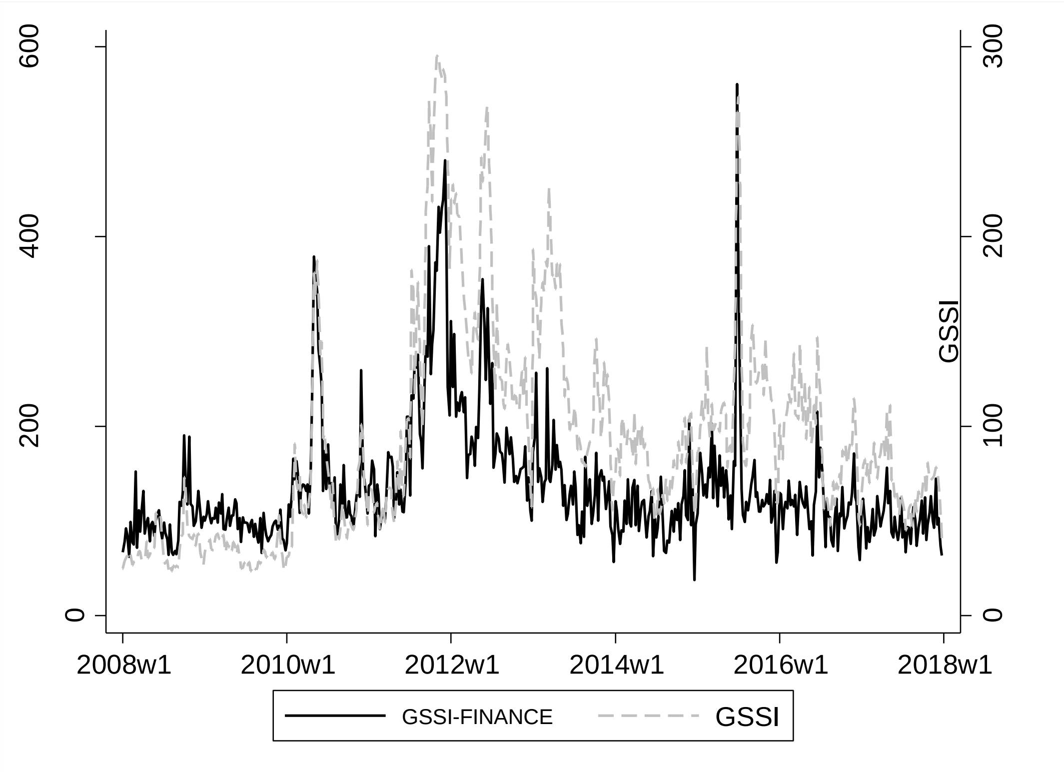


Figure 3. Evolution of GSSI and $GSSI_{FINANCE}$

Figure 3 plots the evolution of the original GSSI and the $GSSI_{FINANCE}$. It shows a similar development in both indexes. Then, we run panel data regressions for the $GSSI_{FINANCE}$. The results are shown in Table 8. As in the previous analysis, the relationship between the $GSSI_{FINANCE}$ and sovereign CDSs is positive and significant. The results remain stable regardless of the countries analyzed, the time horizon considered or the methodology applied. It also shows the asymmetric behavior in peripheral countries

[22] We have downloaded the data on three different dates in the same way as for the GSSI. The correlation between the original GSSI and the $GSSI_{FINANCE}$ is above 80%.

with respect to core countries as well as during the peak of the sovereign debt crisis. Hence, the results confirm that Google data from the finance category report similar results compared to the general category. However, it is worth noting that the coefficients are lower than for the original GSSI, indicating a lower influence on sovereign CDSs.

In short, the robustness tests presented in this section, along with the previous analysis, provide clear evidence in support of a positive relationship between investor sentiment towards sovereign debt markets, using the GSSI as a proxy, and sovereign risk, as measured through sovereign CDSs. This relationship enhances during stress periods and in those countries that have faced more severe constraints in their financial markets.

Table 8. Panel data regressions of $GSSI_{FINANCE}$ on sovereign CDSs

| Dependent variable: sovereign CDSs | OLS | FE | IV<br>All | IV<br>Peripheral | IV<br>Core | IV<br>2008-2012 | IV<br>2013-2017 |
|---|---|---|---|---|---|---|---|
| $GSSI_{FINANCE}$ | 0.166*** | 0.165*** | 0.623*** | 0.728*** | 0.406*** | 0.873*** | 0.324*** |
| | (0.013) | (0.050) | (0.062) | (0.125) | (0.045) | (0.140) | (0.062) |
| Sovereign yields | 90.43*** | 91.70*** | 50.55*** | 82.60*** | -42.38 | -72.36 | 21.29*** |
| | (1.092) | (3.081) | (15.32) | (11.23) | (29.96) | (60.83) | (3.097) |
| VIX | 2.002*** | 2.010*** | 0.335 | -0.255 | 0.950*** | 0.161 | -0.269 |
| | (0.127) | (0.585) | (0.240) | (0.398) | (0.251) | (0.810) | (0.368) |
| N | 5,402 | 5,402 | 5,402 | 1,965 | 2,455 | 2,542 | 2,860 |
| Time dummies | YES | YES | YES | YES | YES | YES | YES |
| Kleibergen-Paap (*p*-value) | | | 0.000*** | 0.000*** | 0.001*** | 0.012** | 0.000*** |
| Hansen (*p*-value) | | | 0.527 | 0.753 | 0.448 | 0.676 | 0.040** |

The table presents the panel data regressions of the $GSSI_{FINANCE}$, and other control variables on sovereign CDSs for 11 European countries: Austria, Belgium, Denmark, France, Germany, Ireland, Italy, the Netherlands, Portugal, Spain and the United Kingdom. We perform a pooled OLS estimation (OLS), fixed effects estimation (FE) and an instrumental variables analysis. We consider the $GSSI_{FINANCE}$, sovereign yields and VIX index to be endogenous variables. The $GSSI_{FINANCE}$ is constructed by aggregating the same keywords as those in the GSSI but considering the queries within the finance category. The model in column 4 includes all countries (All). Columns 5 and 6 include peripheral countries (Peripheral), i.e., Ireland, Italy, Portugal and Spain, and core countries (Core), i.e., Austria, Belgium, France, Germany and the Netherlands, respectively. The last two columns show the estimations for the period of January 2008 to December 2012 and the period of January 2013 to December 2017. Annual time dummies are included. The Kleibergen-Paap test is the underidentification test, under the null hypothesis that the model is underidentified. Therefore its rejection suggests the model is identified. Hansen test is the overidentification test under the null hypothesis that the instruments are valid. Therefore, a rejection casts doubt on the validity of instruments. Robust standard errors are in parentheses.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$.

## 6. CONCLUSIONS

In this paper, we construct a Google Sovereign-Risk Sentiment Index (GSSI) using Google data for a sample of keywords which reflect investor sentiment towards sovereign debt markets and analyze this index to predict the evolution of sovereign risk. The analysis performed shows that the GSSI is highly correlated with other traditional sovereign risk indexes, indicating the ability of Google data to reflect investor sentiment regarding sovereign risk. Moreover, this index shows a positive relationship with CDSs that remains stable regardless of the applied methodology, the time horizon or the countries analyzed. Nonetheless, it is worth noting that the GSSI shows a greater impact on CDSs during times of financial distress and in peripheral countries.

These outcomes are in line with the noise traders' hypothesis (Peri et al., 2014), which indicates that investor sentiment is enhanced after negative shocks in commodity markets. This observation has also been emphasized by Tetlock (2007) and Da et al. (2015), who state that negative events are more helpful to identify investor sentiment for equities, and by Smales (2016) with respect to bank credit risk. Here, we confirm that the same effect occurs in sovereign debt markets. Namely, with the GSSI, we shed light about the predictive power of Google data during negative events regarding sovereign debt markets.

Overall, this paper contributes to the literature by analyzing the role of investor sentiment in financial markets and provides a novel measure of sovereign risk. Our findings can be understood as in accord with the behavioral finance literature and contribute to the discussion on the role of investor sentiment in sovereign debt markets. These results suggest several implications for public authorities and regulators. In this sense, since Google data are easily available and a more transparent means of measuring investor sentiment than other market-based or survey-based sources (Da et al., 2015), these data can be used and monitored by regulators to anticipate the behavior of sovereign debt markets. Moreover, investor sentiment might also play an important role in asset pricing (Han et al., 2017). In this sense, Google-based indexes have been found to be useful tools for asset pricing in stocks (Ben-Rephael et al., 2017; Da et al., 2015) and commodities (Fernandez-Perez et al., 2019; Han et al., 2017). Here, we present a preliminary analysis on the utility of Google data for asset pricing in sovereign debt markets. However, further research is needed in this area. In this sense, future research could consider using out-of-sample procedures or create high-frequency measures of investor sentiment indexes for asset pricing purposes.

**ACKNOWLEDGMENTS**

This work was supported by the Ministerio de Economía, Industria y Competitividad, Gobierno de España [research project number ECO2017-89715-P, entitled “El Análisis del Riesgo en los Mercados Financieros”].

**REFERENCES**

Ammer, J., and 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.

Arghyrou, M. G., and Kontonikas, A. (2012) The EMU sovereign-debt crisis: Fundamentals, expectations and contagion. *Journal of International Financial Markets, Institutions and Money* 22(4), 658–677.

Baker, M., and Stein, J. C. (2004) Market liquidity as a sentiment indicator. *Journal of Financial Markets* 7(3), 271–299.

Baker, M., and Wurgler, J. (2006) Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance* 61(4), 1645–1680.

Baker, M., and Wurgler, J. (2007) Investor Sentiment in the Stock Market. *Journal of Economic Perspectives* 21(2), 129–151.

Bampinas, G., Panagiotidis, T., and Rouska, C. (2019) Volatility persistence and asymmetry under the microscope: the role of information demand for gold and oil. *Scottish Journal of Political Economy* 66(1), 180–197.

Bank, M., Larch, M., and Peter, G. (2011) Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management* 25(3), 239–264.

Barber, B M, and Odean, T. (2001) Boys will be Boys: Gender, Overconfidence, and Common Stock Investment . *The Quarterly Journal of Economics* 116(1), 261–292.

Barber, Brad M., and Odean, T. (2008) All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies* 21(2), 785–818.

Barberis, N., Shleifer, A., and Vishny, R. (1998) A model of investor sentiment. *Journal of Financial Economics* 49(3), 307–343.

Ben-Rephael, A., Da, Z., and Israelsen, R. D. (2017) It Depends on Where You Search: Institutional Investor Attention and Underreaction to News. *The Review of Financial Studies* 30(9), 3009–3047.

Bollen, J., Mao, H., and Zeng, X. (2011) Twitter mood predicts the stock market. *Journal of Computational Science* 2(1), 1–8.

Capelle-Blancard, G., Crifo, P., Diaye, M.-A., Oueghlissi, R., and Scholtens, B. (2019) Sovereign bond yield spreads and sustainability: An empirical analysis of OECD countries. *Journal of Banking & Finance* 98, 156–169.

Carrière-Swallow, Y., and Labbé, F. (2013) Nowcasting with Google Trends in an Emerging Market. *Journal of Forecasting* 32(4), 289–298.

Da, Z., Engelberg, J., and Gao, P. (2011) In Search of Attention. *The Journal of Finance* 66(5), 1461–1499.

Da, Z., Engelberg, J., and Gao, P. (2015) The Sum of All FEARS Investor Sentiment and Asset Prices. *Review of Financial Studies* 28(1), 1–32.

Dergiades, T., Milas, C., and Panagiotidis, T. (2015) Tweets, Google trends, and sovereign spreads in the GIIPS. *Oxford Economic Papers* 67(2), 406–432.

Driscoll, J. C., and Kraay, A. C. (1998) Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics* 80(4), 549–560.

Ejsing, J., and Lemke, W. (2011) The Janus-headed salvation: Sovereign and bank credit risk premia during 2008-2009. *Economics Letters* 110(1), 28–31.

Fernandez-Perez, A., Fuertes, A.-M., González-Fernández, M., and Miffre, J. (2019) Hazard Fear in Commodity Markets. *SSRN Electronic Journal*.

Fondeur, Y., and Karamé, F. (2013) Can Google data help predict French youth unemployment? *Economic Modelling* 30, 117–125.

Gao, L., and Süss, S. (2015) Market sentiment in commodity futures returns. *Journal of Empirical Finance* 33, 84–103.

Gomes, P., and Taamouti, A. (2016) In search of the determinants of European asset market comovements. *International Review of Economics & Finance* 44, 103–117.

Guo, J.-F., and Ji, Q. (2013) How does market concern derived from the Internet affect oil prices? *Applied Energy* 112, 1536–1543.

Han, L., Lv, Q., and Yin, L. (2017) Can investor attention predict oil prices? *Energy Economics* 66, 547–558.

Hoechle, D. (2007) Robust Standard Errors for Panel Regressions with Cross-Sectional Dependence. *The Stata Journal: Promoting communications on statistics and Stata* 7(3), 281–312.

Howard, J. A., and Sheth, J. N. (1969) *The theory of buyer behavior*. New York: Wiley.

Joseph, K., Babajide Wintoki, M., and Zhang, Z. (2011) Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting* 27(4), 1116–1127.

Jun, S.-P., Sung, T.-E., and Park, H.-W. (2016) Forecasting by analogy using the web search traffic. *Technological Forecasting and Social Change* 115, 37–51.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015) Measuring Uncertainty. *American Economic Review* 105(3), 1177–1216.

McLaren, N., and Shanbhogue, R. (2011) Using Internet Search Data as Economic Indicators. *SSRN Electronic Journal*.

Milas, C., Panagiotidis, T., and Dergiades, T. (2018) Twitter versus Traditional News Media: Evidence for the Sovereign Bond Markets. *Working Paper series 18-42, Rimini Centre for Economic Analysis.*

Mondria, J., Wu, T., and Zhang, Y. (2010) The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics* 82(1), 85–95.

Moussa, F., Delhoumi, E., and Ouda, O. Ben (2017) Stock return and volatility reactions to information demand and supply. *Research in International Business and Finance* 39, 54–67.

Peri, M., Vandone, D., and Baldi, L. (2014) Internet, noise trading and commodity futures prices. *International Review of Economics & Finance* 33, 82–89.

Qiu, L., and Welch, I. (2006) Investor Sentiment Measures. *NBER Working paper* w10794.

Remolona, E., Scatigna, M., and Wu, E. (2007) Interpreting sovereign spreads. *BIS Quarterly Review* March, 27–39.

Rose, A. K., and Spiegel, M. M. (2012) Dollar illiquidity and central bank swap arrangements during the global financial crisis. *Journal of International Economics* 88(2), 326–340.

Siganos, A. (2013) Google attention and target price run ups. *International Review of Financial Analysis* 29, 219–226.

Smales, L. A. (2016) News sentiment and bank credit risk. *Journal of Empirical Finance* 38, 37–61.

Srivastava, S., Lin, H., Premachandra, I. M., and Roberts, H. (2016) Global risk spillover and the predictability of sovereign CDS spread: International evidence. *International Review of Economics & Finance* 41, 371–390.

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

Vlastakis, N., and Markellos, R. N. (2012) Information demand and stock market volatility. *Journal of Banking & Finance* 36(6), 1808–1821.

Wellmann, D., and Trück, S. (2018) Factors of the term structure of sovereign yield spreads. *Journal of International Money and Finance* 81, 56–75.

Xu, H.-C., and Zhou, W.-X. (2018) A weekly sentiment index and the cross-section of stock returns. *Finance Research Letters* 27, 135–139.

Click here to access/download
**RDM Data Profile XML**
DataProfile_4123306.xml

# AUTHORSHIP STATEMENT

**Manuscript title**: A SENTIMENT INDEX TO MEASURE SOVEREIGN RISK USING GOOGLE DATA

All persons who meet authorship criteria are listed as authors, and all authors certify that they have participated sufficiently in the work to take public responsibility for the content, including participation in the concept, design, analysis, writing, or revision of the manuscript. Furthermore, each author certifies that this material or similar material has not been and will not be submitted to or published in any other publication before its appearance in the *International Review of Economics & Finance*.

**Authorship contributions**

***Category 1***
Conception and design of study: Marcos González-Fernández, Carmen González-Velasco;
acquisition of data: Marcos González-Fernández, Carmen González-Velasco;
analysis and/or interpretation of data: Marcos González-Fernández, Carmen González-Velasco.

***Category 2***
Drafting the manuscript: Marcos González-Fernández, Carmen González-Velasco;
revising the manuscript critically for important intellectual content: Marcos González-Fernández, Carmen González-Velasco.

***Category 3***
Approval of the version of the manuscript to be published: Marcos González-Fernández, Carmen González-Velasco.

**Acknowledgements**
This work was supported by the Ministerio de Economía, Industria y Competitividad, Gobierno de España [research project number ECO2017-89715-P, entitled "El Análisis del Riesgo en los Mercados Financieros"].

**This statement is signed by all the authors**

| Author's name (typed) | Author's signature | Date |
|---|---|---|
| Marcos González-Fernández | | February, 20 |
| Carmen González-Velasco | Carmen S. Velasco | February, 20 |