HELLENIC REPUBLIC

UNIVERSITY OF MACEDONIA

ISSN 1791-3144

## University of Macedonia
## Department of Economics

## Discussion Paper Series

# Tweets, Google Trends and Sovereign Spreads in the GIIPS

Theologos Dergiades, Costas Milas and Theodore Panagiotidis

## Discussion Paper No. 4/2014

Department of Economics, University of Macedonia, 156 Egnatia str, 540 06
Thessaloniki, Greece, Fax: + 30 (0) 2310 891292
http://www.uom.gr/index.php?newlang=eng&tmima=3&categorymenu=2

# Tweets, Google Trends and Sovereign Spreads in the GIIPS

Theologos Dergiades

School of Science and Technology, International Hellenic University, Greece

Email: t.dergiades@ihu.edu.gr

Costas Milas *

Management School, University of Liverpool, United Kingdom

Rimini Centre for Economic Analysis, Italy

Email: costas.milas@liverpool.ac.uk

and

Theodore Panagiotidis

Department of Economics, University of Macedonia, Greece

Rimini Centre for Economic Analysis, Italy

Email: tpanag@uom.edu.gr

**Abstract**

We examine whether the information contained in social media (Twitter, Facebook & Google Blogs) and web search intensity (Google) influences financial markets. Using a multivariate system and focussing on Eurozone's peripheral countries, the GIIPS (Greece, Ireland, Italy, Portugal and Spain) as well as two of Eurozone's core countries (France and the Nethelands), we show that social media discussion and search-related queries for the Greek debt crisis provide significant short-run information primarily for the Greek-German and Irish-German government bond yield differential even when other financial control variables (international risk, Eurozone's risk, default risk and liquidity risk) are accounted for, and to a much lesser extent for Portuguese, Italian and Spanish sovereign yield differentials. Social media discussion and Google search-related queries for the Greek debt crisis do not affect spreads in France and the Netherlands.

***Keywords***: Google, social media, Greek crisis, frequency domain analysis, GIIPS.

***JEL Classification***: C10, G01, G02

---

* Corresponding author: Costas Milas, Management School, University of Liverpool, Chatham Street, L69 7ZH, Liverpool, UK. Email: costas.milas@liverpool.ac.uk.

# 1. Introduction

Social media have progressively become a popular open forum for analysing economic/financial topics and a field where the public sentiment is reflected in real time. They are widely used by influential economic commentators, policymakers and their followers. For instance, 2008 Nobel Laureate in Economics and *The New York Times* Columnist Paul Krugman had, at the time of writing, 1,198,047 Twitter account followers, whereas International Monetary Fund Managing Director Christine Lagarde had 205,760 Twitter followers. Further, "hot" economic topics like the Eurozone crisis and the Greek debt crisis are covered and discussed in great detail by dedicated websites (for instance in the *The Wall Street Journal* and *The Financial Times*). It has been argued that this "storehouse" of precious information might be contributing to the explanation of upcoming movements in financial markets. Smith (2012) finds evidence for this in currency markets. Da *et al.* (2011) and Joseph *et al.* (2011) find evidence that online search activity predicts price movements in the (very liquid) U.S. equity markets. Even in the less liquid environment of residential home sales, Beracha and Wintoki (2013) find that online search activity predicts price changes. In motivating why an increase (or decrease) in the volume of search intensity predicts changes in asset prices in one direction, Joseph *et al.* (2011) and Beracha and Wintoki (2013) argue that, behaviourally, buyers are more likely to use search engines before making stock or residential real estate purchases than are sellers who already hold the assets and do not need to "search", and as such, one should expect a positive relationship between the prices of those assets and search intensity.

The importance of online search activity has received the attention of the financial press. On April 18th 2013, *Financial Times* commentator Gillian Tett noted that investors can track investment returns with growing precision by plugging into social media.[1] A recent fake tweet reflects the power of social media and the cost of inaccurate information in an era where the speed at which information travels is unprecedented.[2] In fact, it has been suggested that financial markets are more influenced by negative press rumors than by fundamentals (see Hasan *et al.*, 2012).

[1] Gillian Tett on "Markets Insight: Wake up to the Twitter effect on markets", *The Financial Times*, April 18, 2013.

[2] A fake tweet on April 23rd 2013 from a hacked Associated Press account (asserting that explosions at the White House had injured Barack Obama) wiped more than $130 billion off the value of the S&P 500 (*The Economist*; see http://www.economist.com/news/finance-and-economics/21576671-hacked-tweet-briefly-unnerves-stockmarket-newscrashrecover).

Our paper focuses on the determinants of the sovereign cost of borrowing. Being one of the largest capital markets,[3] the sovereign bond market has important implications for the ability of a sovereign as well as the private sector to manage effectively their borrowing needs. Rapidly rising government bond yields add to the burden a country faces to borrow in international markets and therefore undermine its ability to roll existing debt over at a low cost. The fact that the country has to roll its debt over at high interest rates worsens its fiscal prospects, making default more likely. In this sense, the crisis of confidence can become a "self-fulfilling prophecy" (see e.g. Krugman, 2011). At the same time, higher sovereign borrowing costs translate into higher corporate borrowing costs through the sovereign risk effect. For instance, Aguiar *et al.* (2009) and Aguiar and Amador (2011) note that a rise in sovereign debt increases the risk of higher future corporate taxes or the expropriation of private investments at the same time while reducing the ability of the government to offer implicit guarantees to the private sector. Ağca and Celasun (2012) note that since both corporate and sovereign debts are subject to the same country-specific macroeconomic risk factors, creditors seeking portfolio diversification would handle their overall exposure to a given country irrespectively of whether lending is channeled to the public or the private sector.

Consequently, a rise in government debt pushes corporate borrowing costs higher. Indeed, Ağca and Celasun (2012) find that an increase in sovereign debt by one standard deviation from its mean is associated with 9 percent higher loan yield spreads in emerging markets. Using data on 118 non-financial companies in the Eurozone area over the 2008-2011 period, Bedendo and Colla (2013) identify significant spillover effects from the sovereign to the corporate segment; according to their estimates, a 1 percentage point increase in sovereign risk translates into a 50 basis points increase in corporate credit risk. According to a special report by Deutsche Bank (2013), the higher cost of sovereign cost of borrowing is indeed felt by small and medium-sizes enterprises (SMEs) in Eurozone's periphery. The report estimates the average interest rate on loans to "sole proprietors and unincorporated partnerships" (a proxy for small and medium-sized enterprises) in Italy/Spain at 300 to 400 basis points above the German level.

Given the importance of the sovereign debt market for both the public and private sector, the focus of our paper is on the impact of the volume of activity in social media

[3] According to the Global Financial Development database of the World Bank, global outstanding domestic public debt securities accounted for 32.6% of global GDP in 2011 whereas global stock market capitalization accounted for 31.5% of global GDP.

(Twitter, Facebook & Google Blogs) and web search intensity (Google Trends) on the sovereign spread between the GIIPS (Greece, Ireland, Italy, Portugal, Spain) and the German long-term government bond yield during the Greek debt crisis. We believe that the recent Greek crisis provides a natural platform for empirically investigating the role of social media in the debt market over and above the impact of the information provided by other financial control variables (idiosyncratic default risk, liquidity risk, Eurozone's risk and international risk). As the Greek crisis escalated, the term "Grexit" (Greece's exit from the single currency) was added to the financial vocabulary. During this period, the Greek spread rose to unprecedented levels contributing further to the risk of contagion in Eurozone's other peripheral countries. Greece, which was bailed-out twice (for €110bn in 2010 and then again for €109bn in 2011), negotiated, in February 2012, a new €130bn rescue package involving a voluntary haircut of some 53.5% on the face value of its bonds held by the private sector. Eurozone ministers agreed (in November 2012) to cut Greece's debt by a further €40bn. Ireland was bailed-out for €85bn in November 2010. Portugal was bailed-out for €78bn in May 2011. Spain was granted, in July 2012, financial assistance from the European Stability Mechanism (ESM) for up to €100bn. Despite the bail-outs, international markets remain volatile and worried that the debt levels of all GIIPS could be unsustainable (this is reflected, for instance, on Spanish and Italian government yields that are still elevated) posing a risk to the entire Eurozone.[4] These concerns appear justifiable as the GIIPS account for around 34.3% of Eurozone's GDP (Italy is the third and Spain is the fourth largest Eurozone economies) and run both current account deficits and high debt-to-GDP ratios.[5]

It is clear that web search intensity could be linked to both rising and falling spreads. Indeed, search intensity is mainly triggered upon the arrival of news (supplementary information) which can be either good or bad. News for a country's economic fundamentals (or even political stability) can be understood as signals that convey valuable pricing information especially for financial securities issued by the government. Consider the case where a released information set shortens the distance to default for a country (bad news). This will trigger the immediate interest (reflected also in Google's search intensity) of both bondholders and potential bond buyers (bondholders

[4] Following from the pledge of European Central Bank President Mario Draghi to do "whatever it takes" to save the Euro (in July 2012), the European Central Bank approved (in September 2012) a plan paving the way for the bank to make unlimited purchases in struggling Euro members' bond markets (such as Italy and Spain) with the aim of lowering their government bond yields. The plan was conditional on struggling governments to sign up to a Eurozone program of budgetary discipline.

[5] Italy accounts for 17.3% of Eurozone GDP. Spain accounts for 11% of Eurozone GDP, whereas Greece, Portugal and Ireland account for 2.3%, 1.9%, and 1.8% of Eurozone GDP, respectively (*Eurostat*).

from all member-states could be influenced). Assuming that the risk tolerance of individual and institutional investors declines with increasing risk and that bondholders and potential bond buyers update their information set (with bad news), the following occurs. On the one hand, bondholders, who were marginally tolerant with respect to the prior risk level, are now expected to liquefy their assets and, indeed, as the level of risk escalates, an even increasing number of investors are expected to do so. In this case, a growing supply of bonds is observed. On the other hand, faced now with a higher level of risk, potential bond buyers, who marginally accepted the previous risk level, will be less willing to buy, and, indeed, as the level of risk escalates, an even increasing number of potential bond buyers are expected to do so. In this case, a lower demand for bonds is observed. The very demand and supply mismatch in the bond market will only be eliminated by rising spreads. It becomes evident that once we concentrate our attention on a clearly defined sample where bad news persist, then we may argue that search intensity is linked to rising spreads.

To assess the impact of the volume of activity in social media (Twitter, Facebook & Google Blogs) and web search intensity (Google Trends) on the sovereign spreads in the GIIPS, we use the Breitung and Candelon (2006) causality test in the frequency domain (B&C, hereafter). We choose to employ the B&C test because it embraces some features that cannot be traced in the standard linear Granger causality test which is conducted in the time domain. The advantages of the B&C test can be summarized as follows: (i) it distinguishes between short-run and long-run causality, (ii) it allows the identification of causal relationships even if the true interdependence between two variables is non-linear in nature, (iii) it allows us to condition upon a set of relevant variables avoiding potential spurious causality inference, and finally (iv) the test is valid in the presence of volatility clusters, a common characteristic of financial variables. Therefore, in certain circumstances, the B&C test may reveal hidden channels of causality that otherwise would not be identifiable.

We demonstrate that social media discussion and Google search queries for the Greek debt crisis provide significant information for explaining the spread between the cost of borrowing in Eurozone's peripheral bond market and Germany over and above the information provided by other financial control variables (idiosyncratic default risk, liquidity risk, Eurozone's risk and international risk). For comparison reasons, we also report results for France and the Netherlands, two of Eurozone's core countries where

borrowing spreads remained immune to the unprecedented rise recorded in the spreads of the GIIPS.[6]

Our main findings are summarized as follows: First, we identify short-run causality from social media and Google search queries data to the Greek and Irish spreads. Second, there is evidence of short-run causality running from the Greek spread to social media discussion. Third, there is some weak (and information-set sensitive) evidence of predictability of Greek-debt related social media discussion and Google search queries for Portuguese, Italian and Spanish spreads. Arguably then, bad news related to the Greek debt crisis and circulated via online activity, had a negative impact on Ireland (the smaller peripheral countries amongst the GIIPS) especially since the country had a very weak fiscal position (in 2010-2011, Ireland recorded a much higher fiscal deficit than the remaining GIIPS). Furthermore, exposure of banks in peripheral countries to Greek public and private debt might explain why there is some evidence of predictability for the GIIPS. Unsurprisingly, as the Greek debt crisis evolved and Greek creditworthiness took a hit, the market price of Greek debt declined rapidly and consequently, banks exposed to Greek debt witnessed a weakening in their balance sheets. At the same time, the lower market price of Greek debt had an adverse impact on the value of collateral banks needed to secure wholesale funding and triggered margin calls requiring the posting of additional collateral. Fourth, Google carries different short-run predictive information relative to social media. Although Google is used by a wider base, social media (especially Twitter) have become a very popular way of keeping track of news and directing "followers" to news analysis (e.g. in blogs) in an extremely speedy way. Fifth, the borrowing costs in a set of Eurozone's core countries (France and the Netherlands) remained immune to social media/Google queries information related to the Greek crisis.

The structure of the paper is as follows: Section 2 of the paper discusses the implemented methodology. Section 3 reviews briefly the determinants of sovereign spreads and discusses the dataset used in this paper. Section 4 reports our empirical results. Finally, Section 5 discusses our findings and offers suggestions for future research.

[6] France is the second largest Eurozone economy accounting for 21.4% of Eurozone GDP. The Netherlands is the fifth largest Eurozone economy accounting for 6.3% of Eurozone GDP.

## 2. Methodology

Consider the structural bivariate system $\mathbf{z}_t = \mathbf{\Theta}(L)\mathbf{C}^{-1}\mathbf{u}_t = \mathbf{\Psi}(L)\mathbf{u}_t$ with $\mathbf{z}_t = (S_t, G_t)'$. In the context of the Breitung and Candelon (2006) framework (B&C, hereafter), the non-causality hypothesis at frequency $\omega$ is tested by:

$$M_{G\to S}(\omega) = \log\left[1 + \left|\Psi_{12}(e^{-i\omega})\right|^2 \Big/ \left|\Psi_{11}(e^{-i\omega})\right|^2\right] \tag{1}$$

where $\Psi_{11}$ and $\Psi_{12}$ are derived from a VAR moving average representation. In case where $G_t$ does not cause $S_t$ at $\omega$, $M_{G\to S}(\omega)$ is zero and $\left|\Psi_{12}(e^{-i\omega})\right|^2 = 0$. B&C remedy the estimation complexity of $\left|\Psi_{12}(e^{-i\omega})\right|^2$ through a set of restrictions imposed on the VAR coefficients. B&C restate the null based on $\Psi_{12}(L) = -(1/c_{22}) * \left(\Theta_{12}(L)/\left|\Theta(L)\right|\right)$. $1/c_{22}$ is positive[7] element of the $C^{-1}$ and $\left|\mathbf{\Theta}(L)\right|$ is the determinant of $\mathbf{\Theta}(L)$. The null hypothesis of no causality at frequency $\omega$ running from $G_t$ to $S_t$ is not rejected whenever (2) holds:

$$\left|\Theta_{12}(e^{-i\omega})\right| = \left|\sum_{k=1}^{p} \bar{\ }_{12,k} \cos(k\omega) - \sum_{k=1}^{p} \bar{\ }_{12,k} \sin(k\omega) i\right| = 0 \tag{2}$$

with $\bar{\ }_{12,k}$ to be element of the $\mathbf{\Theta}_k$ matrix. The set of restrictions imposed are:[8]

$$\sum_{k=1}^{p} \bar{\ }_{12,k} \cos(k\omega) = 0 \quad \text{and} \quad \sum_{k=1}^{p} \bar{\ }_{12,k} \sin(k\omega) = 0 \tag{3}$$

If we denote $\sim_j = \bar{\ }_{11,j}$ and $\ _j = \bar{\ }_{12,j}$, the VAR equation that corresponds to the $S_t$ variable may be rewritten as:

$$S_t = \sim_1 S_{t-1} + \ldots + \sim_p S_{t-p} + \ _1 G_{t-1} + \ldots + \ _p G_{t-p} + £_{1t} \tag{4}$$

Thus, the null hypothesis is equivalent to the following set of restrictions:

$$R(\omega) \ = 0, \quad \text{where} \quad = \left(\ _1, \ldots, \ _p\right)' \text{ and } R(\omega) = \begin{pmatrix} \cos(\omega) & \ldots & \cos(p\omega) \\ \sin(\omega) & \ldots & \sin(p\omega) \end{pmatrix} \tag{5}$$

To assess the validity of (5) for frequencies $\omega$ that range within $(0, ®)$, B&C compare the obtained statistic with the 0.05 critical value of the ¡ $^2$ distribution with 2 degrees of freedom.

---

[7] This is due to the assumption that the variance-covariance matrix is a positive definite.

[8] Given that $\sin(k\omega) = 0$ in the cases where $\omega = 0$ and $\omega = ®$, it comes that the second restriction in (3) is simply disregarded.

Hosoya (2001) proposed a method that eliminates from the variables of interest the one way feedback due to a third variable, while the initial feedback structure remains invariant. For instance,[9] in the multivariate system $\mathbf{z}_t = (S_t, G_t, \mathbf{m}_t)'$ with $\mathbf{m}_t$ to be the vector that contains the conditioning variables, let $\mathbf{w}_t$ denote the projection residual vector obtained by projecting $\mathbf{m}_t$ into the Hilbert space $H(S_t, G_t, \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, ...)$. Similarly, let $©_t$ and $¢_t$ represent the projection residual vectors obtained by projecting $S_t$ and $G_t$, respectively, into Hilbert space $H(\mathbf{w}_t, \mathbf{w}_{t-1}, ...)$. After the described transformation, Hosoya (2001) argues that a higher order conditional causality measure can be expressed equivalently by the bivariate causality measure $M_{G \to S/\mathbf{m}}(\omega) \equiv M_{¢ \to ©}(\omega)$.

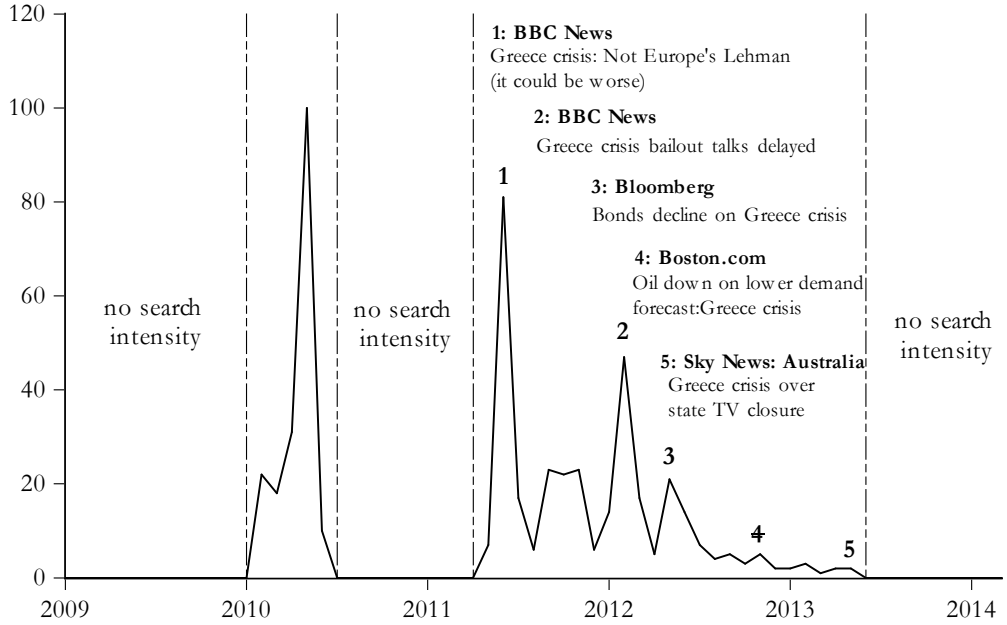
## 3. Data issues and determinants of sovereign spreads

As discussed in the Introduction, to ensure that web search intensity is linked to rising spreads, it is important to select a sample period where released (and related to the Greek case) news disclose predominantly negative information. To determine a period with such a physiognomy in a justifiable way, we utilize as a "guide" the search intensity for news in the Google trends facility. Initially, we select two composite keywords, "*Greece crisis*" and "*Greek debt crisis*".[10] Substantial search intensity for the above keywords can determine candidate periods for investigation. At a second stage, in order to ensure that the events taking place during the determined periods communicate principally bad news, we backdate and examine all major events connected to the Greek crisis. Once we confirm that the period is deluged by bad news, we may then argue with reasonable confidence that the realized search intensity for the specific sample is connected to rising spreads. By entering the above criteria in Google trends facility and focusing on the search intensity for news and headlines, two candidate periods are determined: (i) from January 2010 to June 2010, and (ii) from May 2011 to May 2013 (see Figure 1 below). We focus our attention on the most recent period simply because of its extensive duration.

[9] In this study, we condition upon four variables.

[10] We argue that the resulting search intensity after the joint usage of the above mentioned keywords ("*Greece Crisis*"+"*Greek debt crisis*") is capable of identifying a period dominated by bad news.

**Figure 1.** Google trends search intensity for news related to the Greek crisis




Having established the study period (20/5/2011 to 09/5/2013 - 495 obs), the search intensity index (henceforth, $G_t$) for key-phrases related to the Greek debt crisis is retrieved from the Google Trends facility (http://www.google.com/trends/). To select a cluster of queries capable of capturing the web search interest that is directly related to the Greek debt crisis, it is essential to identify a key-phrase that explicitly describes the subject of interest. Naturally, we choose as a suitable key-phrase the expression *Greek debt crisis*. Armed with that key-phrase, Google trends suggests analogous queries classifying them in terms of relevance (the relevance index receives values from 0 to 100). Based on these suggestions, we select all top queries (with relevance index>90) as directly related searches to the topic of our interest.[11] The recommended queries are the following: *Greece debt crisis*, *Greece debt*, *the Greek crisis* and *Greece crisis*.[12, 13]

Furthermore, special care must be given to the treatment of the search operators. For instance, by querying *Greece debt*, the results may embody supplementary related terms (e.g. *Greece debt crisis*) apart from the inclusion of *Greece* and *debt* (in any sequence). Given a set of queries, overlapping may be a severe threat (e.g. the query *Greece debt* may

[11] All the subsequent top queries (apart from those selected) illustrate relevance index below or equal to 35.

[12] To confirm the validity of the proposed queries, we also make use of an additional Google facility, namely Google correlate. Given that the facility does not offer correlated searches on a global basis, we seek similar queries for the key-phrase *Greek debt crisis* in the US, Canada and UK. The results reveal that the initially proposed queries indeed belong into the top 10 most related searches for all examined countries.

[13] The $G_t$ index is available at a weekly frequency; we convert the series to daily frequency using the quadratic-match average method.

contain the search phrase *Greece debt crisis*). To overcome this concern, all queries are enclosed in quotation marks so only the exact term is used. The usage of quotation marks indirectly permits us to minimize the "contamination" of the $G_t$ index by search queries which are related to incidences of positive news. The indirect obstruction of search intensity related to good incidents, is taking place through the exclusion of terms that can alter the physiognomy of the search query. For example, the phrase *Greece debt* signifies predominantly search intensity for bad news while the phrase *Greece debt deal* suggests mostly search intensity for good news. Thus, for the extraction of the $G_t$ index, the search is specified as follows: "*Greek debt crisis*"+"*Greece debt crisis*"+"*Greece debt*"+"*the Greek crisis*"+"*Greece crisis*". The selected geographic area is worldwide and the language used is English.[14]

We note that for a selected key-phrase or a group of key-phrases, Google trends does not deliver the exact volume of queries but rather a normalized index of search intensity. The normalization is conducted as follows: the search volume associated to the query of interest $V_t^q$ for a given time unit (day, week or month) is expressed as a fraction ($r_t$) of the entire search volume of queries $V_{e,t}^q$ that correspond to the same time unit (day, week or month). At a second stage the data is scaled through the multiplication of every fraction (obtained after the normalization process) by the scaling factor $F = 100/r^*$, where $r^*$ is the fraction with the highest value, that is:

$$\max_{r_t \in R^+}\{r_t\} = \{r^*\} \tag{6}$$

The finally constructed normalized series for a single key-phrase (or *n* multiple key-phrases) can be depicted as follows:

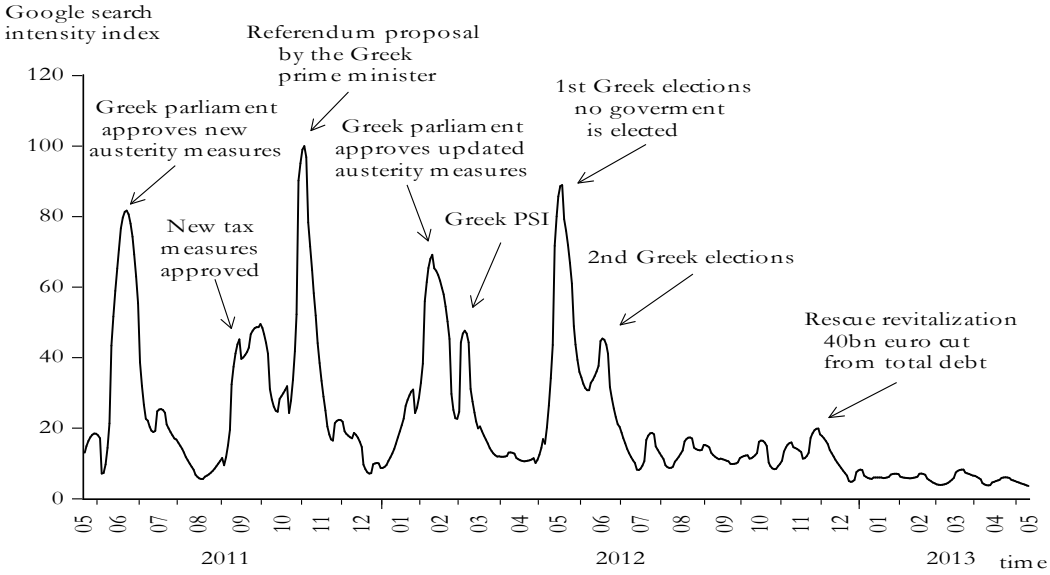
$$S_t = \frac{V_t^q / V_{e,t}^q}{r^*} 100 \quad \text{or} \quad \left( S_t = \frac{\sum_{i=1}^{n} V_{i,t}^q / V_{e,t}^q}{r^*} 100 \right) \tag{7}$$

Naturally, the resulting index of search intensity receives values that range within the interval [0, 100]. The lower limit of the interval signifies inconspicuous search

[14] Undeniably, a more complete treatment demands integration of the related web search activity which is conducted in other European languages (notably because different exposure on Greek debt of the banking sector exists in different countries). As a result, we try to identify web search activity for the translated term "greek debt crisis" in six different languages, namely German (griechische schuldenkrise), French (crise de la dette grecque), Italian (crisi del debito greco), Spanish (crisis de la deuda griega), Greek (ελληνική κρίση χρέους) and Portuguese (crise da dívida grega). For every translated phrase, the Google trends facility indicates not enough search volume in order to deliver data. This may be attributed to the fact that (i) the prevailing language in the web is English and (ii) the majority of web users from these countries have knowledge of the English language and are therefore able to execute Google searches in English.

intensity, while the upper limit of the interval indicates the maximum observed search intensity. Finally, provided that the delivered index is unaffected by $V_{e,t}^{q}$ over time, the comparison of different observations for the same index becomes more meaningful and therefore any subsequent analysis more robust. Figure 2 plots the $G_t$ index.

**Figure 2.** The $G_t$ index



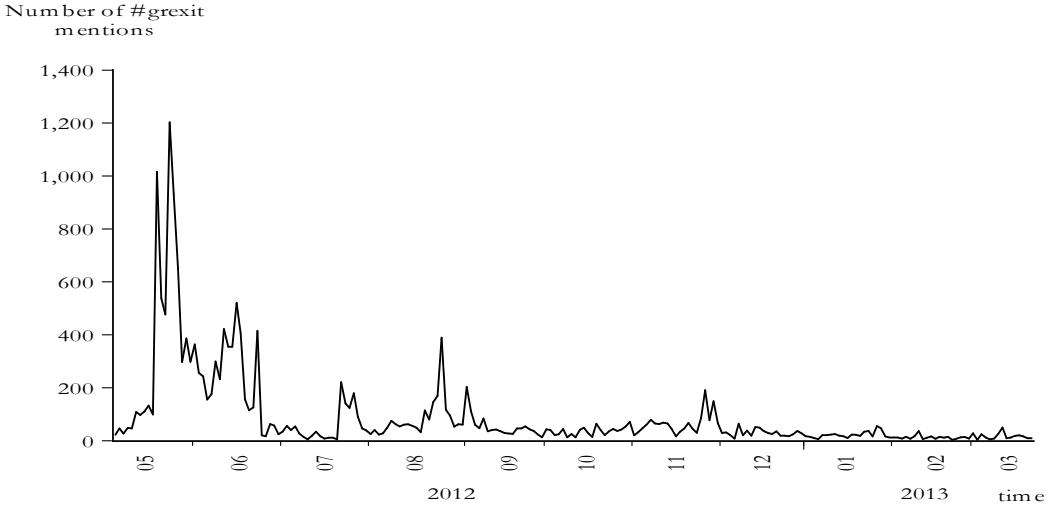

*Note*: PSI refers to Private Sector Involvement in the restructuring of Greek debt (February 2012). Greece, which was bailed-out twice (for €110bn in 2010 and then again for €109bn in 2011) negotiated, in February 2012, a new €130bn rescue package involving a voluntary haircut of some 53.5% on the face value of its bonds held by the private sector. Eurozone ministers agreed (in November 2012) to cut Greece's debt by a further €40bn.

As a second online activity index with respect to the Greek crisis we use the total number of mentions for the keyword #*Grexit* in the Twitter, Facebook and Google blogs (source: http://analytics.peoplebrowsr.com/). To allow direct comparison (apples-to-apples) with the respective Google trends data, the sample size for our analysis is determined based on the Google trends search intensity index for the key-word "*Grexit*". Therefore, the sample period extends from 04/5/2012 to 21/3/2013 (220 observations).[15] Figure 3 plots jointly for the Twitter, Facebook and Google blogs the total number of mentions (henceforth, $T_t$) for the key-word #*Grexit*. The Twitter (Facebook) data refer to the total number of users mentioning in their tweets (posts) the key-word #*Grexit*, including also re-tweets (re-posts). Similarly, the Google blogs data refer to the total number of users mentioning the keyword #*Grexit*. Finally, the selected

[15] Given that the second sample is a subset of the first sample, we argue that this second type of online activity is also linked to rising spreads.

geographic area is worldwide and the dataset includes tweets or posts that can be in any language.

**Figure 3.** Twitter, Facebook & Google blogs #Grexit mentions ($T_t$ index)



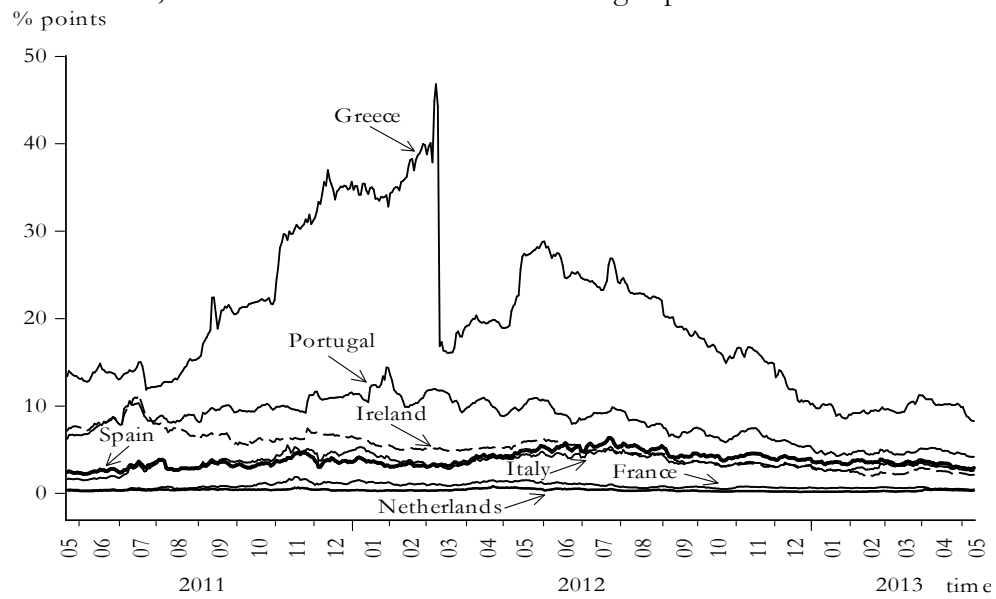

Empirical studies that focus on the determinants of sovereign yield spreads introduce a wide range of explanatory variables. These are classified into two broad categories: international factors and idiosyncratic factors (see Oliveira *et al.* 2012; Hilscher and Nosbusch, 2010). Commonly used international factors include: (i) the international risk aversion level captured by the implied volatility of the S&P 500 - *VIX* index (Beber *et al.* 2009; Gerlach *et al.* 2010; Arghyrou and Kontonikas, 2012) or by the difference between the US corporate bonds yield and the 10-year US sovereign bonds yield (Codogno *et al.* 2003), (ii) global liquidity conditions proxied by the TED spread (Hilscher and Nosbusch, 2010) or by the US Federal Funds Rate (Csonto and Ivaschenko, 2013), and (iii) the global cost of capital proxied by the US bond yields (Hilscher and Nosbusch, 2010; Maltritz, 2012). Frequently implemented idiosyncratic factors refer to macroeconomic variables capturing the country's macroeconomic fundamentals as well as its overall credit risk rating (or probability of default) and liquidity risk (market size and depth). Macroeconomic variables include (amongst others) the debt-to-GDP ratio (Pagano and Von Thadden, 2004), the fiscal balance-to-GDP ratio (Schuknecht *et al.*, 2009), the current account balance (Heinz and Sun, 2014), the inflation rate (Lemmen and Goodhart, 1999), the short-term interest rate (Schuknecht *et al.*, 2009), the credit rating (Manganelli and Wolswijk, 2007) or probability of default (Hilscher and Nosbusch, 2010). The role of liquidity risk, frequently measured by the bid-ask spread differential, is recognized by several studies as limited (Geyer *et al.*, 2004;

Favero *et al.*, 2010). Nevertheless, lack of attention towards liquidity risk has been cited by the President of the Federal Reserve Bank of Boston Eric Rosengren (2010) as one of the reasons explaining why the seriousness of the recent financial crisis was underestimated by economic forecasters; in fact, liquidity considerations have become a central issue in the literature only recently (see e.g. Angelini *et al.*, 2011; Naes *et al.*, 2011).

Failure to account for factors that are crucial in shaping sovereign spreads and, at the same time, may also impact on the online search intensity and social media activity, could provide misleading causality inferences. Therefore, we condition upon a set of relevant variables. In particular, we use two international factors that capture Eurozone and global risk and two idiosyncratic factors that capture country-specific default risk (also reflecting country-specific macroeconomic position) and country-specific liquidity risk. On top of these factors, and given our interest in web activity, we extend previous work by accounting for the effects of web-traced information.

For the sovereign spreads and their determinants discussed above we use daily time-series data over the period 20/5/2011 to 09/5/2013. The sovereign spread ($S_{jt}$) between the 10-year government bond yield in Eurozone peripheral country $j$ ($j =$ Greece, Ireland, Italy, Portugal, Spain, France and the Netherlands) and the German government bond yield is available from Datastream; see Figure 4.[16, 17]

**Figure 4.** GIIPS, France and the Netherlands sovereign spreads

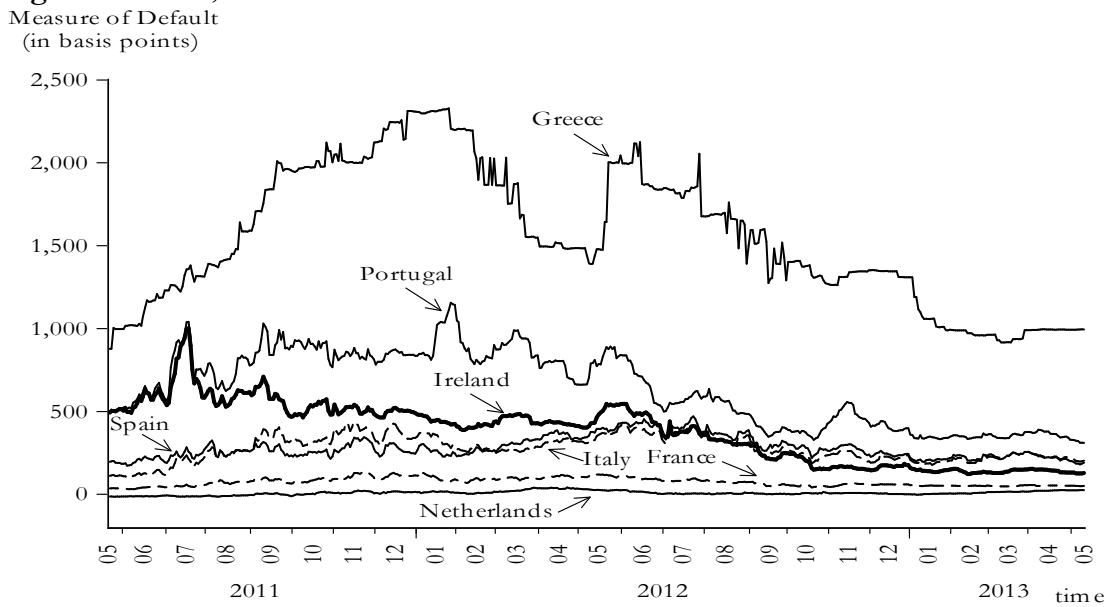



[16] Nonlinear unit root tests (Kapetanios *et al.*, 2003) support stationarity for $S_{jt}$ and the web search indices (detailed results are available upon request).

[17] GARCH effects might be present in the sovereign spreads (Figure 4). Bodart and Candelon (2009) demonstrate that the B&C test is not sensitive to the presence of volatility clustering.

We proxy the default risk ($D_{jt}$) by the difference between the 10-year Credit Default Swap (CDS) premia in country $j$ and the 10-year German CDS premia (available from Datastream; see Figure 5).[18]

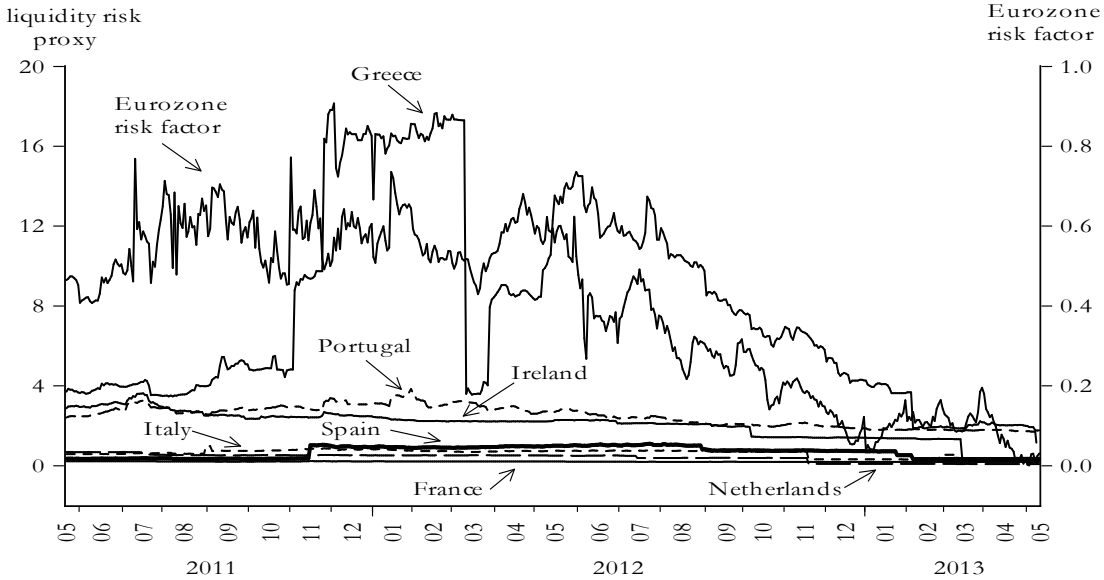
**Figure 5.** GIIPS, France and the Netherlands default indices




We proxy liquidity risk ($L_{jt}$; see Figure 6) using the differential between the bid-ask spread of the 10-year bond in country $j$ and the bid-ask spread of the 10-year German bond (see e.g. De Santis, 2012; Favero *et al.,* 2010). Following De Santis (2012) and Schwarz (2013), we also use (from Bloomberg) the spread between the 10-year KfW (Kreditanstalt für Wiederaufbau) bond yield and the 10-year German government bond yield as a proxy for the Eurozone area common risk factor ($E_t$); see Figure 6. KfW are German agency bonds. These bonds are less liquid than the federal government ones; however, KfW bonds carry the same default risk as they are fully guaranteed by the German federal government. Therefore, any difference should reflect "flight-to-liquidity" and "flight-to-safety" considerations.

[18] To proxy the default risk of Greece, we use the difference between the equally weighted sum of the 10-year Credit Default Swap (CDS) premia for two major Greek banks (Alpha Bank and National Bank of Greece) and the 10-year German CDS premia. Greek debt restructuring (in February 2012) triggered the payment of Greek CDS in March 2012. Since then, these series has been discontinued. For this reason, we use the CDS information on Greece's two major banks which are positively correlated (0.58) with the Greek CDS.

**Figure 6.** GIIPS, France and the Netherlands liquidity risk & Eurozone risk factor



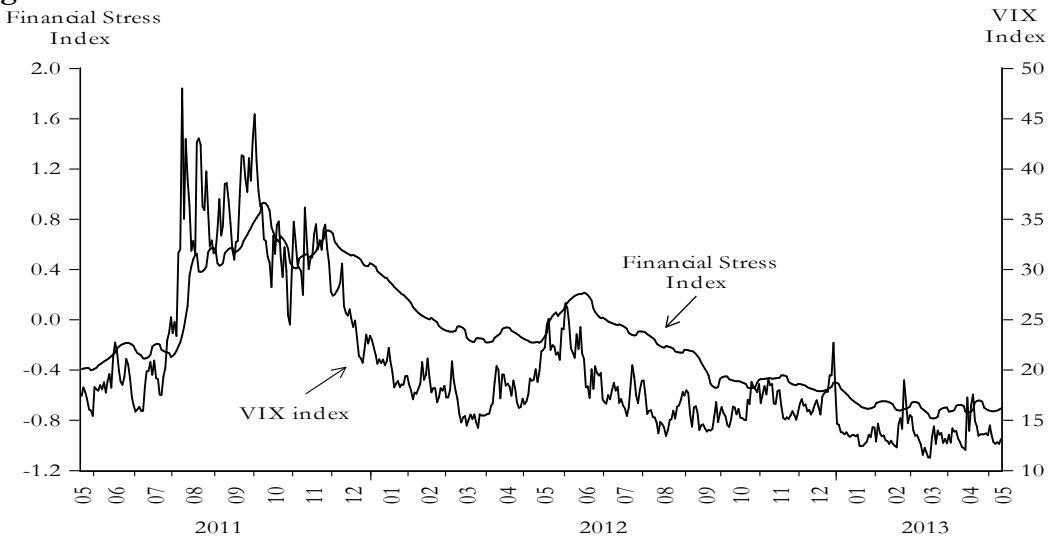

*Note*: Liquidity risk ($L_{jt}$) for country $j$ is defined in terms of the percentage Bid-Ask spread:

$$100*\left(\frac{Ask_{jt}-Bid_{jt}}{0.5*(Ask_{jt}+Bid_{jt})}\right)-100*\left(\frac{Ask_{Germany,t}-Bid_{Germany,t}}{0.5*(Ask_{Germany,t}+Bid_{Germany,t})}\right)$$

, where *Ask* and *Bid* refer to the Ask and Bid price of the 10-year government bond.

We follow Arghyrou and Kontonikas (2012) in proxying international risk by the *VIX* index. As an alternative and broader measure, we use the Federal Reserve Bank of St. Louis Financial Stress Index (*FSI*) which is a composite index of 18 financial variables (including *VIX*); see Figure 7.[19] Finally, the data for France and the Netherlands in Figures 4 to 6 have been included in our analysis for comparison purposes.

**Figure 7.** Proxies for international risk: *FSI* & *VIX*




[19] *FSI* is available at a weekly frequency; we convert the series to daily frequency using the quadratic-match average method.

## 4. Empirical results

Using the B&C test, we disentangle short- and long-run predictability among the variables of interest. To implement the B&C test, depending on the examined hypothesis each time, target or causing variable may be one of the $S_{jt}$, $G_t$ or $T_t$. Following the notation of Section 2, the $\mathbf{m}_t$ vector, upon which we condition, includes four variables and more specifically proxies for idiosyncratic default risk ($D_{jt}$), liquidity risk ($L_{jt}$), Eurozone risk ($E_t$) and international risk ($VIX_t$ or $FSI_t$). For the projection of the $\mathbf{m}_t$ vector as well as for the projection of both the target and the causing variables into the $H(S_t, G_t, \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, ...)$ and the $H(\mathbf{w}_t, \mathbf{w}_{t-1}, ...)$, respectively, the optimal lag-length selection is based on the Schwarz information criterion. Empirical models that use the $G_t$ index are estimated over the period 20/5/2011 to 09/5/2013 whereas empirical models that use the $T_t$ index are estimated over the period 04/5/2012 to 21/3/2013.

The results for Greece (with and without conditioning) are presented in Figures 8 to 11. The null hypothesis of no predictability running from $G_t$ towards $S_{Greece,t}$ is rejected for the bivariate B&C measure, at the 0.05 significance level, when $\omega \in [0.09\pi, 0.74\pi] \cup [0.85\pi, \pi]$, (Figure 8). This implies that medium size and high frequencies of $G_t$, with wave lengths of less than 2.36 days ($2\pi/\omega$=6.28/2.66=2.36) as well as between 2.70 and 21.66 days, are those that offer predictive power with respect to $S_{Greece,t}$. When the B&C test is re-conducted, after the Hosoya's (2001) conditioning approach, the revealed predictability pattern is comparable to the bivariate case irrespective of which international risk proxy is employed ($VIX_t$ or $FSI_t$). The range of frequencies in which predictability is supported in the case where $FSI_t$ is used, correspond to cyclical components with wave lengths of less than 2.07 days ($\omega \in [0.97\pi, \pi]$) and between 2.68 and 22.43 days ($\omega \in [0.09\pi, 0.75\pi]$). When $VIX_t$ is used, predictability is revealed for wave lengths of less than 2.23 days ($\omega \in [0.90\pi, \pi]$) and between 2.62 and 21.66 days ($\omega \in [0.09\pi, 0.76\pi]$). Raggedly, it can be argued that the short-run cyclical components of the $G_t$ index with wave lengths of less than three weeks are capable of offering additional information with respect to the future movements of $S_{Greece,t}$. At the same time, there is no credible evidence of long-run causality. On the other hand, the null hypothesis of no causality at the opposite direction ($S_{Greece,t} \rightarrow G_t$) is rejected for the

entire range of frequencies. The latter also holds when conditioning takes place (see Figure 10).

**Figure 8.** $G_t \rightarrow S_{Greece,\,t}$

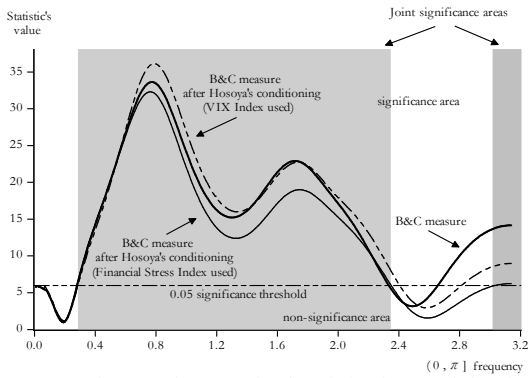
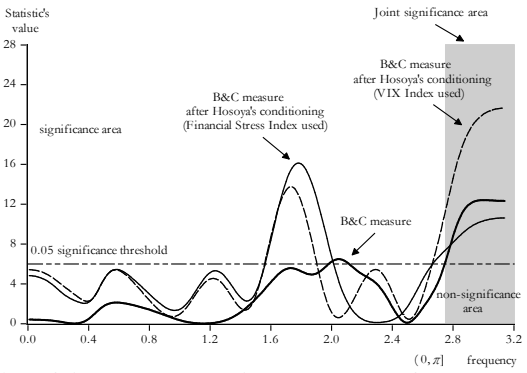



**Figure 9.** $T_t \rightarrow S_{Greece,\,t}$




*Note*: In Figure 8 the VAR lag length implemented for the derivation of the B&C measure, the B&C measure after Hosoya's conditioning using $FSI$ and the B&C measure after Hosoya's conditioning using $VIX$ is 8 in every case. Similarly, in Figure 9 the VAR lag length is 13 in every case. The grey area indicates joint statistical significance of the three plotted measures.

**Figure 10.** $S_{Greece,\,t} \rightarrow G_t$

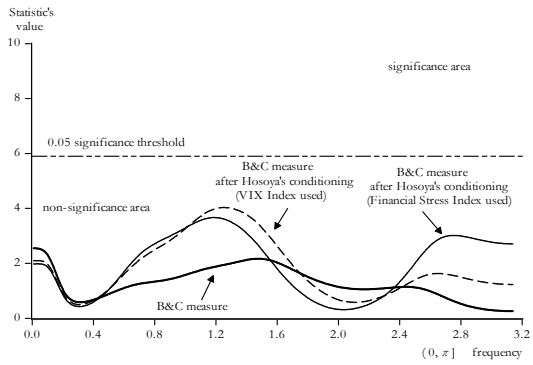
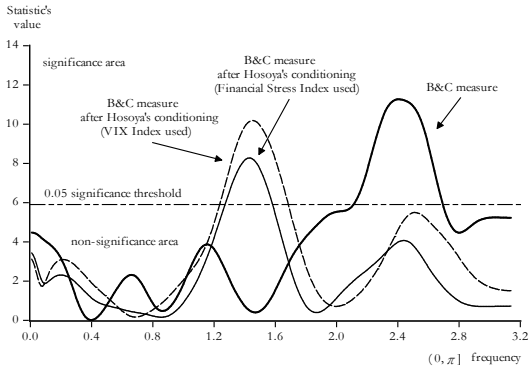



**Figure 11.** $S_{Greece,\,t} \rightarrow T_t$




*Note*: In Figure 10 the VAR lag length implemented for the derivation of the B&C measure, the B&C measure after Hosoya's conditioning using $FSI$ and the B&C measure after Hosoya's conditioning using $VIX$ is 8 in every case. Similarly, in Figure 11 the VAR lag length is 14, 13 and 13, respectively.

The same testing procedure (with and without conditioning; with the same control variables) is implemented to the Twitter, Facebook and Google Blogs index (number of mentions) in order to assess its impact on the Greek spreads. The non-Granger causality hypothesis running from $T_t$ to $S_{Greece,\,t}$ is rejected for the bivariate B&C measure at the 0.05 significance level when $\omega \in [0.63\pi,\ 0.67\pi] \cup [0.88\pi,\ \pi]$; see Figure 9. The above mentioned frequency ranges correspond to wave lengths of less than 2.28 days and between 2.98 and 3.16 days. Therefore, only short-run causality is established. Similar pattern as above, with relatively larger range of frequencies, contributing significantly in the prediction of $S_{Greece,\,t}$, is uncovered under the Hosoya's (2001) conditioning approach. In particular, conditioning under the $FSI_t$ index indicates significant predictability for wave lengths of less than 2.35 days and between 3.12 and 4.00 days.

Similarly, in the case where $VIX_t$ is included in $\mathbf{m}_t$ predictability exists for wave lengths of less than 2.35 days and between 3.31 and 4.03 days. Additionally, we find empirical evidence to support the reverse hypothesis mainly in medium-run frequencies; the same holds even when we condition upon the set of variables included in $\mathbf{m}_t$ (see Figure 11). For the bivariate B&C measure significant predictability is established between 2.33 and 2.96 days ($\omega \in [0.68\pi, 0.86\pi]$). In the case where in $\mathbf{m}_t$ we include $VIX_t$ the predictability period is defined within the range of 3.74 and 5.02 days ($\omega \in [0.40\pi, 0.54\pi]$), while the respective period when $FSI_t$ is used lies between 4.00 and 4.91 days ($\omega \in [0.41\pi, 0.50\pi]$). The results for Greece are summarized in Table 1 below (Table 1 does not present the findings of the reverse hypothesis).

The asymmetry in the findings of the reverse causality ($G_t \not\rightarrow S_{Greece,t}$ and $T_t \rightarrow S_{Greece,t}$) can be characterized at first glance as puzzling. Such evidence may be puzzling if we assume homogeneous profile (in terms of education and capacity to comprehend economic topics) for the users of both data sources. Provided that Google is used by a wide range of users, a small fraction of those is anticipated to comply with the profile described above and therefore their impact on search intensity due to the evolution of spreads is expected to be rather small. On the other hand, Twitter users (the number of #*Grexit* mentions primarily comes from Twitter) are more educated relative to Google users (Mitchell *et al.*, 2012), implying that we expect a larger fraction of users to meet the profile described above. It appears that, for the case of Greece, the size of the fraction of users meeting the profile above is large enough to verify causality from spreads towards the number of #*Grexit* mentions.[20]

Figures 12-19 report our findings for the remaining GIIPS whereas a summary of all results for the GIIPS is reported in Table 1. For the $G_t$ index we find credible evidence of predictability mainly for Ireland (Figure 12) and to a much lesser extent for Italy (Figure 13), while no predictability is confirmed for Portugal (Figure 14) and Spain (Figure 15). For the case of Ireland the range of frequencies for the bivariate B&C measure in which predictability is established, corresponds to cyclical components with wave lengths between 5.66 and 13.08 days. When conditioning is taking place and $FSI_t$

[20] When we examine the reverse causality for the remaining GIIPS, we find evidence of significant short-run predictability only for the case of Italy, but the significance vanishes once conditioning is taking place (especially when the *VIX* index is used). The results are available upon request. We are grateful to an anonymous referee for pointing this dimension.

is utilized as proxy for international risk predictability exists for wave lengths of less than 2.95 days and between 8.16 and 12.08 days. The respective period for $VIX_t$ is between 7.39 and 12.31 days. For Italy predictability exists only in the bivariate B&C measure between 3.19 and 3.76 days; on the other hand, once conditioning takes place, predictability vanishes for the entire range of frequencies.

**Figure 12.** $G_t \rightarrow S_{\text{Ireland},\, t}$

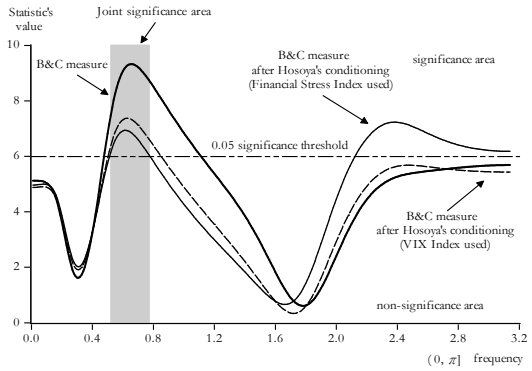
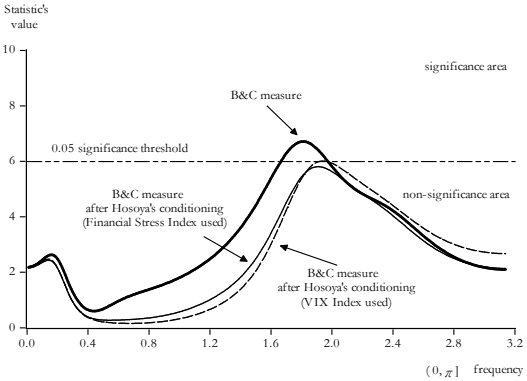



**Figure 13.** $G_t \rightarrow S_{\text{Italy},\, t}$




*Note*: In Figures 12-13, the VAR lag length implemented for the derivation of the B&C measure, the B&C measure after Hosoya's conditioning using $FSI$ and the B&C measure after Hosoya's conditioning using $VIX$ is 7 in every case. The grey area indicates joint statistical significance of the three plotted measures.

**Figure 14.** $G_t \rightarrow S_{\text{Portugal},\, t}$

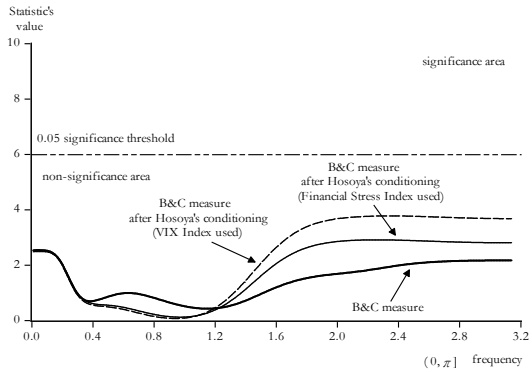
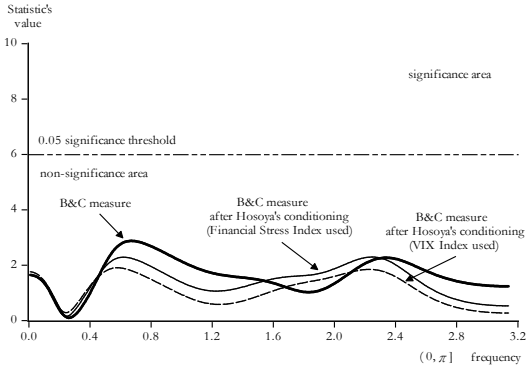



**Figure 15.** $G_t \rightarrow S_{\text{Spain},\, t}$




*Note*: In Figure 14 the VAR lag length implemented for the derivation of the B&C measure, the B&C measure after Hosoya's conditioning using $FSI$ and the B&C measure after Hosoya's conditioning using $VIX$ is 6 in every case.. Similarly, in Figure 15 the VAR lag length is 7 in every case.

For the $T_t$ index, we find convincing evidence of short-run predictability only for Ireland, while the evidence for Italy, Portugal and Spain is feeble and non-systematic. For the case of Ireland (Figure 16) and the bivariate B&C measure the range of frequencies in which predictability is established, correspond to cyclical components with wave lengths between 2.33 and 2.49 days as well as between 3.19 and 4.49 days. In the case where the conditioning process includes $FSI_t$ predictability is verified for wave lengths between 2.38 and 2.71 days as well as between 3.12 and 4.58 days. When $VIX_t$ is used the respective period is defined between 2.28 and 2.66 days as well as between 3.17 and 4.19

days. Concerning Italy (Figure 17), Portugal (Figure 18) and Spain (Figure 19), a common pattern is revealed, since predictability in high frequencies is verified only when the conditioning vector includes the $FSI_t$ variable. For Italy and Spain, evidence of predictability exists for wave lengths between 2.21 and 2.41 days and between 2.32 and 2.53 days, respectively. Finally, for Portugal predictability occurs for wave lengths between 2.26 and 2.80 days and between 4.72 and 5.02 days. In any other case, no significant causality occurs for the entire range of frequencies (see Table 1 for a summary).[21]

**Figure 16.** $T_t \rightarrow S_{\text{Ireland},\, t}$

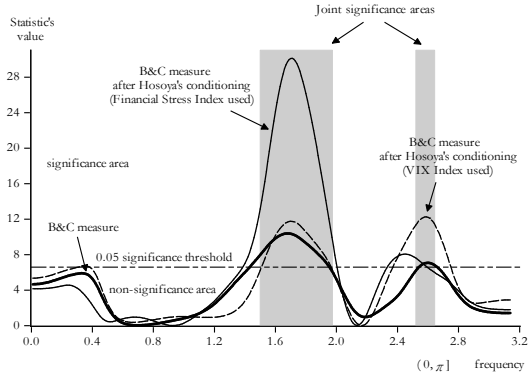
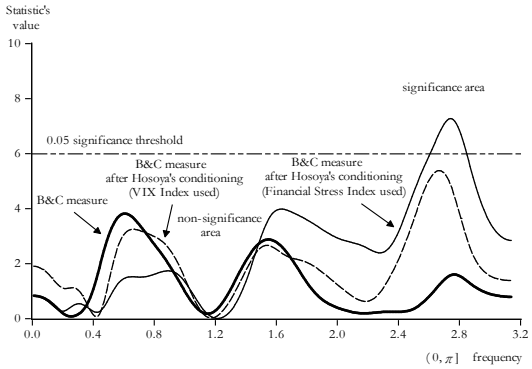



**Figure 17.** $T_t \rightarrow S_{\text{Italy},\, t}$




*Note*: In Figures 16-17, the VAR lag length implemented for the derivation of the B&C measure, the B&C measure after Hosoya's conditioning using $FSI$ and the B&C measure after Hosoya's conditioning using $VIX$ is 13 in every case. The grey area indicates joint statistical significance of the three plotted measures.

**Figure 18.** $T_t \rightarrow S_{\text{Portugal},\, t}$

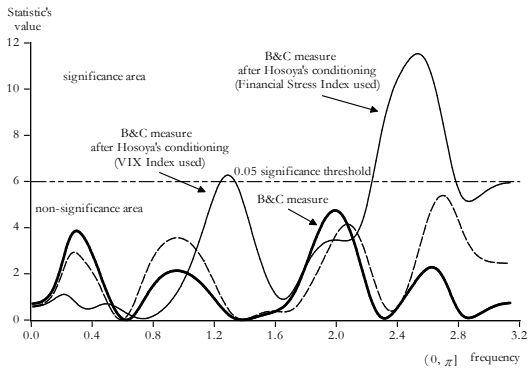
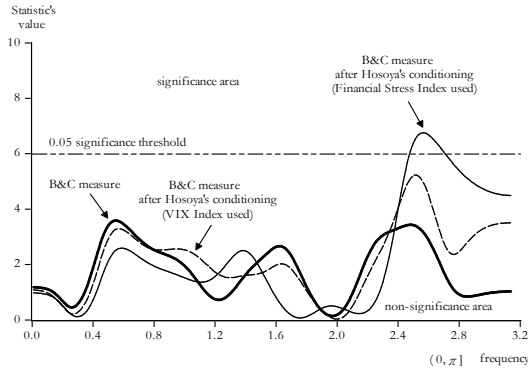



**Figure 19.** $T_t \rightarrow S_{\text{Spain},\, t}$




*Note*: In Figure 18 the VAR lag length implemented for the derivation of the B&C measure, the B&C measure after Hosoya's conditioning using $FSI$ and the B&C measure after Hosoya's conditioning using $VIX$ is 13 in every case. Similarly, in Figure 19 the VAR lag length is 12 in every case.

[21] Given our interest in GIIPS, and following a reviewer's suggestion, we re-conducted the entire empirical part of the paper (without conditioning) by constructing a new search intensity index which has as a starting point the key-phrase *euro crisis*. Adopting the same sample period, the request, which has been performed in the Google trends facility for the formation of the index, is the following: "*euro crisis*"+"*the euro crisis*"+"*euro debt*"+"*euro debt crisis*"+"*euro in crisis*". The derived results revealed that swapping the term Greece (or Greek) with the term Euro does not appear to increase, in any significant way, the predictive power of the search intensity index with respect to the spreads for the GIIPS. Detailed results are available upon request.

A point of caution is that we cannot compare directly for each country the causality results for $G_t$ with those of $T_t$, since the construction of these two indexes is based on a different set of key-words/key-phrases (multiple key-phrases for $G_t$ and a single key-word for $T_t$). To execute apples-to-apples comparison, we need to isolate the same key-word (*Grexit* in our case) from both sources of data and for the same time period. By conducting the above exercise for a sample size that coincides with the period in which significant search volume exists in Google Trends for the key-word *Grexit* (see Section 3), we may argue that the two different sources of data result to qualitative similar causality inference. In particular, no causality exists for Italy, Portugal and Spain while significant causality in relatively high frequencies is established for Greece and Ireland. The identified causality for both countries is clearly more pronounced for the $T_t$ index.[22]

**Table 1**: How Google trends and social media predict 10-year sovereign spreads ($S_{jt}$).

| Country $j$ | Predictability without conditioning | Predictability with conditioning ($FSI$) | Predictability with conditioning ($VIX$) |
|---|---|---|---|
| Panel *A*: Google search intensity predict spreads ($G_t \rightarrow S_{jt}$). All numbers below refer to days. | | | |
| Greece | ≤2.36 & [2.70-21.66] | ≤2.07 & [2.68-22.43] | ≤2.23 & [2.62-21.66] |
| Ireland | [5.66-13.08] | ≤2.95 & [8.16-12.08] | [7.39-12.31] |
| Italy | [3.19-3.76] | $\not\in$ | $\not\in$ |
| Portugal | $\not\in$ | $\not\in$ | $\not\in$ |
| Spain | $\not\in$ | $\not\in$ | $\not\in$ |
| France | $\not\in$ | $\not\in$ | $\not\in$ |
| Netherlands | $\not\in$ | $\not\in$ | $\not\in$ |
| Panel *B*: Twitter, Facebook & Google Blogs number of mentions predict spreads ($T_t \rightarrow S_{jt}$). | | | |
| Greece | ≤2.28 & [2.98-3.16] | ≤2.35 & [3.12-4.00] | ≤2.35 & [3.31-4.03] |
| Ireland | [2.33-2.49] & [3.19-4.49] | [2.38-2.71] & [3.12-4.58] | [2.28-2.66] & [3.17-4.19] |
| Italy | $\not\in$ | [2.21-2.41] | $\not\in$ |
| Portugal | $\not\in$ | [2.26-2.80] & [4.72-5.02] | $\not\in$ |
| Spain | $\not\in$ | [2.32-2.53] | $\not\in$ |
| France | $\not\in$ | $\not\in$ | $\not\in$ |
| Netherlands | $\not\in$ | $\not\in$ | $\not\in$ |

*Note*: The table reports the Breitung and Candelon (2006) predictability test without/with Hosoya's (2001) conditioning. The symbol $\not\in$ denotes absence of predictability.

[22] We are indebted to an anonymous referee for pointing this out. The results are available upon request.

For comparison reasons, Figures 20 to 23 report our results for France and the Netherlands. What we find is no causality of the volume of activity in social media ($T_t$) and web search intensity ($G_t$) on the sovereign spreads of France and the Netherlands. Therefore, our findings suggest that borrowing costs in a set of Eurozone's core countries (in this case France and the Netherlands) remained immune to social media and Google queries information related to the Greek crisis.

**Figure 20.** $G_t \rightarrow S_{\text{France}, t}$

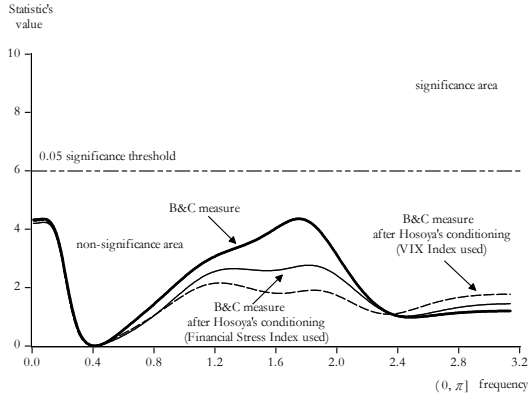
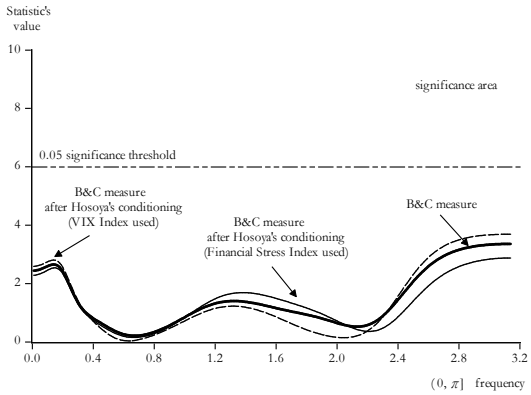



**Figure 21.** $G_t \rightarrow S_{\text{Netherlands}, t}$




In Figures 20-21, the VAR lag length implemented for the derivation of the B&C measure, the B&C measure after Hosoya's conditioning using *FSI* and the B&C measure after Hosoya's conditioning using *VIX* is 7 in every case.

**Figure 22.** $T_t \rightarrow S_{\text{France}, t}$

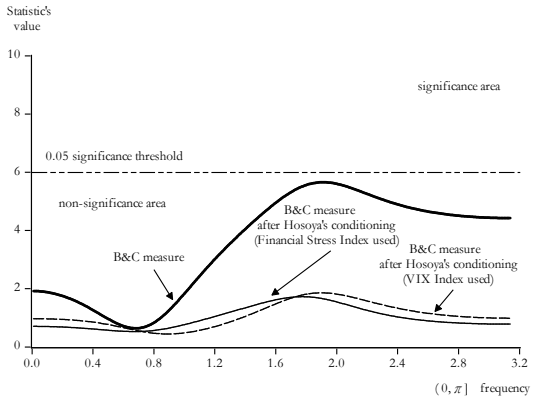
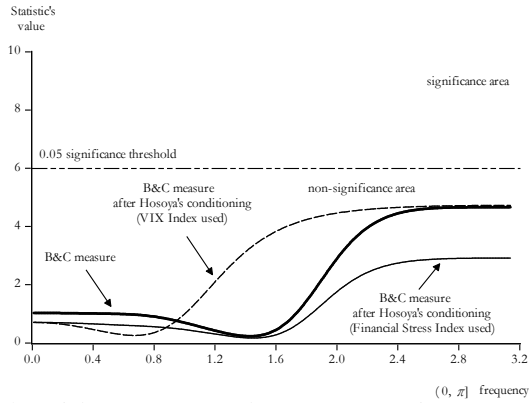



**Figure 23.** $T_t \rightarrow S_{\text{Netherlands}, t}$




*Note*: In Figure 22 the VAR lag length implemented for the derivation of the B&C measure, the B&C measure after Hosoya's conditioning using *FSI* and the B&C measure after Hosoya's conditioning using *VIX* is 4 in every case. Similarly, in Figure 23 the VAR lag length is 4,4 and 3, respectively.

Hence, there is evidence of short-run predictability for Ireland (stemming from both $G_t$ and $T_t$) and sporadic evidence of predictability (sensitive to the information set used) for Portugal (stemming from $T_t$), Italy (stemming from both $G_t$ and $T_t$) and Spain (stemming from $T_t$). It goes without saying that these results should be treated with caution. Ireland, whose economic share in the Eurozone output is smaller than that of Greece was the second peripheral country (after Greece) to be bailed-out. Arguably then,

bad news circulated via online activity is more likely to impact negatively on smaller peripheral countries especially if the latter have very weak fiscal positions. In fact, Ireland recorded a much higher fiscal deficit than Greece, Italy Portugal and Spain during the 2010-2011 period.[23]

In addition, exposure of banks in peripheral countries to Greek public and private debt might also explain why there is some evidence of predictability for the GIIPS. Indeed, *Bank of International Settlements* (BIS) data showed that in June 2011,[24] Portuguese banks had some $10.08bn exposure (or 6.73% of their total exposure around the world) followed by a $3.88bn exposure (or 0.40% of total exposure) for Italian banks, $0.77bn (or 0.21% of total exposure) for Irish banks and $1.22bn (or 0.1% of total exposure) for Spanish banks. Following the Greek debt restructuring, the exposure to Greek debt was reduced; yet, Portuguese banks, followed by Italian banks, remained more exposed to Greek debt than the others. According to the latest BIS data (published in June 2013), the exposure of Portuguese banks to Greek debt had dropped, in December 2012, to $7.34bn (or 6.16% of their total exposure around the world). The exposure of Italian banks to Greek debt had dropped to $1.0bn (or 0.12% of total exposure), for the Irish banks to $0.11bn (or 0.07% of total exposure) and for the Spanish banks to $0.76bn (or 0.05% of total exposure). This argument is in line with the evidence of Mink and de Haan (2013) who rely on an event study approach and employ daily data to identify significant effects of news about the Greek bailout on stock price returns in European banks (irrespective of their exposure to the remaining GIPS; the definition of Mink and De Haan excludes Italy). They also find that that news about the Greek economic situation and the Greek bailout has led to abnormal returns on sovereign bonds for the GIPS with a larger impact in the case of Portugal and a lower impact in the case of Ireland and Spain.

To sum up, we show that Greek debt crisis related information in social media and Google search queries does influence financial markets. This is mainly so for Greece and Ireland, and to a much lesser extent for Italy, Portugal and Spain. This could be viewed as a weak signal of contagion from Greece to (some of) the remaining GIIPS in the sense that social media discussion and Google search queries related to the Greek debt crisis carry some predictive information for the cost of borrowing in other peripheral Eurozone economies. Noting that economists disagree on the definition of contagion

[23] Irish fiscal deficit "hit" 30.5% of GDP in 2010 (this was due to the one-off impact of Government's support to Irish banks; see OECD (2011): *Economic Surveys Ireland*) and then fell to 13.1% of GDP in 2011.
[24] Data from: http://www.bis.org/statistics/consstats.htm (Table 9B).

and how it can be empirically tested (see Corsetti *et al.*, 2011), our Greek-crisis related variables ($G_t$ and $T_t$) arguably comply with the thinking of Mink and de Haan (2013) who refer to contagion in terms of country-specific events and their impact on the asset prices of other countries.

Given that the data used in our case are unwrought, further qualitative elaboration may be of fathomless importance in revealing more clearly the true linkage between Greek-crisis related social media/Google search queries and GIIPS spreads. With this in mind, our work, which relies on the B&C test, differs from recent work by Arghyrou and Kontonikas (2012) who use monthly data to identify contagion effects in terms of a significant direct and positive effect from the Greek spread on other Eurozone sovereign spreads and more so for the remaining GIIPS (with the effect being stronger for Ireland and Portugal), or recent work by De Santis (2012) who identifies contagion effects in terms of the direct impact of a Greek credit rating downgrade on other Eurozone sovereign spreads (the impact is again stronger for Portugal and Ireland). Our work also differs from Beetsma *et al.* (2013) who construct macroeconomic/financial news variables about the GIIPS (using information from the newsflash of *Eurointelligence*; an independent internet-based service which provides daily morning Euro-area news briefings of the European media) to conclude that bad news has increased sovereign yield differentials in the GIIPS and has triggered spillover effects to non-GIIPS countries. Finally, we also note the work of Di Cesare *et al.* (2013) who argue that recent movements in Eurozone spread differentials have increased to levels above those justified by economic fundamentals. Di Cesare *et al.* (2013) construct a monthly index of search volume of Euro break-up keywords ("*end of Euro*", "*end of the Euro*", "*Euro break-up*", "*Euro break*") and note that this index has a strong positive correlation of 0.77 with the residual part of the 10-year Italian spread, that is, the part not explained by economic fundamentals (such as the deficit-to-GDP ratio, expected growth, the volatility of stocks in the banking sector and the volatility of the Italian spread).

## 5. Discussion and concluding remarks

This paper examines the long and short-run causality between Google search queries and social media data related to the Greek debt crisis and sovereign spreads in the GIIPS. We have five main findings. First, we identify short-run causality from social media and

Google search queries data to the Greek and Irish spreads. Our evidence remains strong even when a number of financial controls are accounted for. This might be due to the fact that spreads, social media/Google search queries and controls are all driven by the same unknown underlying process (for instance, expectations could play an important role). Second, there is evidence of short-run causality running from the Greek spread to social media discussion. This finding remains robust even when conditioning takes place. Third, there is some weak (and information-set sensitive) evidence of predictability of Greek-debt related social media discussion and Google search queries for Portuguese, Italian and Spanish spreads. Fourth, $G_t$ carries different short-run predictive information relative to $T_t$. Although Google is used by a wider base, social media (especially Twitter) have become a very popular way of keeping track of news and directing "followers" to news analysis (e.g. in blogs) in an extremely speedy way; this might explain why $T_t$ provides more pronounced evidence of short-run predictability for all the examined Eurozone peripheral countries in comparison to $G_t$. Fifth, the borrowing costs in a set of Eurozone's core countries (in this case France and the Netherlands) remained immune to social media/Google queries information related to the Greek crisis.

A frequently made assumption is that the relationship between information demand and risk aversion is a positive one (see for instance Eeckhoudt and Godfroid 2000). Although opposite views to the latter have been expressed, in this paper we focus on social media attention for the Greek crisis in terms of negative news[25] which drive up spreads on the grounds that it reflects increased concerns about Greek debt sustainability and the future of the Eurozone. The evidence we provide implicitly assumes the following order: increased social media activity (in terms of bad news) implies higher risk which, in turn, results in a higher spread.[26] This very issue is arguably a hostage to fortune. Indeed, it suggests that the more policymakers talk about the Greek debt crisis, and hence the more the social media refer to it, the more spreads will rise.[27]

Our paper looks at the contribution from information contained in social media towards predicting sovereign spreads, as opposed to the contribution of social media

[25] Alternatively, frequently used terms include "animal spirits", "market psychology" and self-reinforcing waves of pessimism and optimism (see the theoretical framework proposed by Angeletos and La'O, 2013).
[26] Vlastakis and Markelos (2012) confirm empirically for US stocks that investors demand more information as their level of risk aversion increases.
[27] A similar argument was put forward by David Smith, Economics Editor of *The Sunday Times*. Commenting on the launch of the Bank of England's new 'uncertainty gauge', which pools information from a set of financial market indicators and the number of press articles citing economic uncertainty, Smith noted that as policymakers intensify their talk about uncertainty, journalists write more about it which, in turn, adds further to uncertainty and damages economic growth (Smith, 2013).

itself. The latter question would however be intellectually interesting: how much mileage do we get from considering social media data in addition to more traditional measures of news intensity (e.g. number of related articles published that day). Indeed, it would be interesting to repeat the exercise conditioning on traditional news intensity measures once such measures will be easily and readily available. In that regard, the frequency of data (daily) we consider is somewhat limiting, as one of the points of a medium like Twitter is ultra-fast communication of very fresh news, and it would be interesting to explore higher frequency impacts, if any. We leave this issue for future research.[28]

Overall, our empirical results suggest that unwrought data, effortlessly traced in social media, enclose valuable information content with respect to the short-run movements of financial markets. We do not argue that the intensity of searches or the number of mentions for a particular term is comparable to a sentiment index; rather, we flag the issue that it offers unexploited information which can be further utilized for improving our understanding of financial markets.

## Acknowledgements

We would like to thank two anonymous reviewers of this Journal for their most useful comments and suggestions. We also thank Richard Baillie (Michigan State University), Roberto De Santis (ECB), Carlo Favero (Bocconi University), Chris Florackis (Liverpool University), Ilias Lekkos (Piraeus Bank), Marcus Miller (University of Warwick) and Thanasis Stengos (University of Guelph) for useful comments. TP acknowledges financial support from the Research Committee of the University of Macedonia. Any remaining errors are our own.

## References

Ağca, Ş. and O. Celasun (2012). Sovereign Debt and Corporate Borrowing Costs in Emerging Markets. *Journal of International Economics* 88, 198-208.

Aguiar, M., Amador, M. and G. Gopinath (2009). Investment Cycles and Sovereign Debt Overhang. *Review of Economic Studies* 76, 1-31.

[28] We are grateful to an anonymous referee for pointing out this future research direction.

Aguiar, M. and M. Amador (2011). Growth in the Shadow of Expropriation. *Quarterly Journal of Economics* 126, 651-697.

Angeletos, G.-M. and J. La'O (2013). Sentiments. *Econometrica* 81, 739-779.

Angelini, P., A. Nobili and C. Picillo (2011). The Interbank Market after August 2007: What Has Changed, and Why? *Journal of Money, Credit and Banking* 43, 924-958.

Arghyrou, M.G. and A. Kontonikas (2012). The EMU Sovereign-Debt Crisis: Fundamentals, Expectations and Contagion. *Journal of International Financial Markets, Institutions and Money* 22, 658-677.

Beber, A., M.W. Brandt and K.A. Kavajecz (2009). Flight-to-quality or flight-to-liquidity? Evidence from the euro area bond market. *Review of Financial Studies* 22, 925-957.

Bedendo, M. and P. Colla (2013). Sovereign and Corporate Credit Risk: Spillover Effects in the Eurozone. 9th CSEF-IGIER Symposium on Economics and Institutions (CISEI), 24-28 June.

Beetsma, R, M. Giuliodori, F. de Jong and D. Widijanto (2013). Spread the News: The Impact of News on the European Sovereign Bond Markets during the Crisis. *Journal of International Money and Finance* 34, 83-101.

Beracha, E. and M.B. Wintoki (2013). Forecasting Residential Real Estate Price Changes from Online Search Activity. *Journal of Real Estate Research* 35, 283-312.

Bodart, V. and B. Candelon (2009). Evidence of Interdependence and Contagion Using Frequency Domain Framework. *Emerging Markets Review* 10, 140-150.

Breitung, J. and B. Candelon (2006). Testing for Short and Long-run Causality: a Frequency Domain Approach. *Journal of Econometrics* 132, 363-378.

Codogno, L., C. Favero and A. Missale (2003). Yield spreads on EMU government bonds. *Economic Policy* 18, 505-532.

Corsetti, G., M. Pericoli and M. Sbracia (2011). Correlation Analysis of Financial Contagion. In: Kolb, R. (Ed.), *Financial Contagion: The Viral Threat to the Wealth of Nations*. John Wiley & Sons, Inc., New York, 11-20.

Csonto, B. and I. Ivaschenko (2013). Determinants of Sovereign Bond Spreads in Emerging Markets: Local Fundamentals and Global Factors vs. Ever-Changing Misalignments. International Monetary Fund Working Paper, No. WP/13/164.

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

De Santis, R.A. (2012). The Euro Area Sovereign Debt Crisis-Safe Haven, Credit Rating Agencies and the Spread of the Fever from Greece, Ireland and Portugal. European Central Bank Working Paper, No. 1419.

Deutsche Bank (2013). Focus Europe. The SME Problem. Special Report, 8 March 2013.

Di Cesare, A., G. Grande, M. Manna and M. Taboga (2013). Recent Estimates of Sovereign Risk Premia for Euro-area Countries. Bank of Italy Occasional Papers, No. 128.

Eeckhoudt L. and P. Godfroid (2000). Risk Aversion and the Value of Information. *Journal of Economic Education* 31, 382-388.

Favero, C.A., M. Pagano and E.L. Von Thadden (2010). How Does Liquidity Affect Bond Yields? *Journal of Financial and Quantitative Analysis* 45, 107-134.

Gerlach, S., A. Schulz and G.B. Wolff (2010). Banking and Sovereign risk in the euro area. Deutsche Bundesbank Discussion Paper, No. 09/2010.

Geyer, A., S. Kossmeier and S. Pichler (2004). Measuring Systematic Risk in EMU Government Yield Spreads. *Review of Finance* 8, 171-197.

Hasan, I., K. Jackowicz, O. Kowalewski and Ł. Kozłowski (2012). Market Discipline during Crisis: Evidence from Bank Depositors in Transition Countries. The Wharton School University of Pennsylvania, Financial Institutions Center Working Paper, No. 12-12.

Heinz, F.F. and Y. Sun (2014). Sovereign CDS Spreads in Europe-The Role of Global Risk Aversion, Economic Fundamentals, Liquidity, and Spillovers. International Monetary Fund Working Paper, No. WP/14/17.

Hilscher, J. and Y. Nosbusch (2010). Determinants of Sovereign Risk: Macroeconomic Fundamentals and the Pricing of Sovereign Debt. *Review of Finance* 14, 235-262.

Hosoya, Y. (2001). Elimination of Third-Series Effect and Defining Partial Measures of Causality. *Journal of Time Series Analysis* 22, 537-554.

Joseph, K., M.B. Wintoki and Z. Zhang (2011). Forecasting Abnormal Stock Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search. *International Journal of Forecasting* 27, 1116-1127.

Kapetanios, G., Y. Shin and A. Snell (2003). Testing for a Unit Root in the Nonlinear STAR Framework. *Journal of Econometrics* 112, 359-379.

Krugman, P. (2011). An Impeccable Disaster. *The New York Times*, September 11.

Lemmen, J. and C. Goodhart (1999). Credit risks and European government bond markets: a panel data econometric analysis. *Eastern Economic Journal* 25, 77-107.

Maltritz, D. (2012). Determinants of sovereign yield spreads in the Eurozone: A Bayesian approach. *Journal of International Money and Finance* 31, 657-672.

Manganelli, S. and G. Wolswijk (2007). Market discipline, financial integration and fiscal rules: what drives spreads in the Euro area government bond market? European Central Bank Working Paper, No. 745.

Mink, M. and J. de Haan (2013). Contagion during the Greek Sovereign Debt Crisis. *Journal of International Money and Finance* 34, 102-113.

Mitchell, A., T. Rosenstiel and L. Christian (2012). What Facebook and Twitter Mean for News. Pew Research Center, Annual Report on American Journalism: The State of the News Media 2012.

Naes, R., J.A. Skjeltorp and B.A. Ødegaard (2011). Stock Market Liquidity and the Business Cycle. *Journal of Finance* 66, 139-176.

Oliveira, L., J.D. Curto and J.P. Nunes (2012). The determinants of sovereign credit spread changes in the Euro-zone. *Journal of International Financial Markets, Institutions and Money* 22, 278-304.

Pagano, M. and E.L. Von Thadden (2004). The European bond markets under EMU. *Oxford Review of Economic Policy* 20, 531-554.

Rosengren, E. (2010). The Impact of Liquidity, Securitization, and Banks on the Real Economy. *Journal of Money, Credit and Banking* 42, 221-228.

Schuknecht, L., J. Von Hagen and G. Wolswijk (2009). Government risk premiums in the bond market: EMU and Canada. *European Journal of Political Economy* 25, 371-384.

Schwarz, K. (2013). Mind the Gap. Disentangling Credit and Liquidity in Risk Spreads. University of Pennsylvania, Wharton School of Business, Working Paper series in Finance.

Smith, G.P. (2012). Google Internet Search Activity and Volatility Prediction in the Market for Foreign Currency. *Finance Research Letters* 9, 103-110.

Smith, D. (2013). Economic Outlook: Productivity is Down but Don't Count it Out. *The Sunday Times*, June 16.

Vlastakis, N. and R.N. Markelos (2012). Information Demand and Stock Market Volatility. *Journal of Banking and Finance* 36, 1808-1821.