News Flow, Web Attention and Extreme Returns in the European Financial Crisis

Chouliaras, Andreas and Grammatikos, Theoharry

Luxembourg School of Finance

31 October 2013

News Flow, Web Attention and Extreme Returns in the European Financial Crisis

Andreas Chouliaras*, Theoharry Grammatikos
Luxembourg School of Finance**

June 2015

Abstract

We attempt a connection between three entities: Extreme Stock Market Returns, the Web Attention factor and a set of News Flow factors, for three groups of countries during the European Financial Crisis: the Euro-periphery countries, the Euro-core countries, and the major European Union -but not euro- countries. Using daily stock market data from January 2004 till March 2013 and textual analysis on more than 24,000 news articles from seven leading international news providers, we find that the Euro-periphery Web Attention (SVI) and News Flow variables significantly affect the probabilities of extreme bottom returns for the Euro-periphery, the Non-euro and the Euro-core groups. More Web Attention and more bad news for the Euro-periphery in times of crisis are associated with higher probabilities of extreme bottom returns within and across groups.

JEL classification: G01, G14, G15, D83.

Keywords: Financial Crisis, Textual Analysis, Web Attention, SVI, News Flow, Financial Sentiment.

* Corresponding author. E-mail address: andreas.chouliaras@uni.lu
** Luxembourg School of Finance, 4, rue Albert Borschette, 1246 Luxembourg.
We thank Joakim Westerholm, Xiangkang Yin, Yu-Chen Wei, Sebastian Bethke, Helena Chulia, seminar participants at the 2013 Auckland Finance Meeting, the 26th Australasian Finance & Banking Conference, MSF Symposium 2014, 2014 Financial Management Association (FMA) Europe Conference, 2014 European Financial Management Association (EFMA) conference, World Finance Conference 2014, International Finance and Banking Society 2014 conference for helpful comments.

1. Introduction

One topic that dominated the financial press over the past few years has undoubtedly been the Euro-crisis. At the center of the Euro-crisis were the Euro-periphery countries (Portugal, Ireland, Italy, Greece, Spain), whose vulnerabilities led to bailout packages, either at the sovereign or the banking level. Scores of news stories were written on whether these countries could make it or not, how bad their finances and how uncompetitive they were, whether they should part ways with the common currency area, and whether bad news about them might propagate and affect the rest of the European countries, causing a domino effect.

Classic asset pricing theory uses financial factors to price assets. But the original source of the markets' reactions and hence of the crisis spillover must be traced to relevant information about the underlying financial entities, and the way investors process and interpret the content of this information. Previous research regarding news and events concerning the Euro-crisis mainly dealt with the impact of official news announcements such as sovereign debt rating changes, using dummy variables to denote the occurrence of events, or arbitrarily defining events and news as "good" or "bad" (e.g. Arezki, Candelon, and Sy (2011), Beetsma, Giuliodori, de Jong, and Widijanto (2013), Mink and De Haan (2013)). Nevertheless, such specific announcements give at best a partial and at worst a biased view of the impact of information on market prices since they do not reflect all available news and in many cases they are anticipated by market participants. Classifying events arbitrarily is problematic because it depends entirely on the perceptions and beliefs of the researcher(s) who classify the news, while it also neglects the degree of negative (or positive, or uncertain) information, dealing only with the extreme parts of the sentiment spectrum, and completely ignoring all values in between. One can conjecture that changes in relevant information flows is one of the significant factors that affect stock prices. Obtaining usable indicators of information flows and their sentiment and relating them directly to the market returns is the focus of this paper. We attempt to shed light on the impact of two main information sources on market prices and crisis transmission; the News Flow imbedded in newspaper articles and newswires, and investors' attention about the crisis captured by their Web search activities. The main contributions of this paper are the following: first, we incorporate a broad selection of news sources and we use a rather elaborate method to select news items relevant for the topic of the Euro-crisis under investigation; second, we develop and test different metrics of news relevance; third, we specifically examine the impact of News Flow (and Web Attention) about the peripheral countries on their own stock markets as well as on the stock markets of other European countries; This way we answer the question of whether after all the Euro-periphery financial sentiment does indeed significantly affect the other countries financial asset returns, which is the main concern of financial discussions and policy making since the advent of the Euro-crisis.

We find that the News Flow and Web Attention about the financial crisis significantly affect not only the Euro-periphery but also the Euro-core and the Non-euro country groups. During the Euro-crisis higher values for the News Flow and Web Attention factors about the Euro-periphery crisis are associated with higher probabilities of extreme bottom returns for all three country groups.

The rest of the paper is organised as follows. Section 2 presents the related textual analysis and Web Attention literature. Section 3 presents the data and the model specification. The empirical findings are shown in Section 4. Finally, Section 5 provides a set of robustness and alternative specifications and Section 6 concludes.

2. Related Literature

The study of News Flow has attracted the researchers' interest rather recently with the advent of Data Mining and Sentiment Analysis techniques. The strong interest in this area has been demonstrated by the recent creation of companies and commercial products specialized in the production of financial sentiment (see e.g., RavenPack[1] and Thomson Reuters News Analytics[2]). As far as the finance literature is concerned, the pioneering work of Tetlock (2007) uses textual analysis (based on the Harvard psychosocial dictionary) of a Wall Street Journal column, and associates the content of the news with the Dow Jones returns, using vector autoregressions (VARs). He finds that media pessimism has predictive power on market returns, while reversion effects occur and extreme absolute values of pessimism predict higher trading volumes. Loughran and McDonald (2011) develop finance-oriented word lists by fine-tuning the Harvard dictionary, and correlate textual analysis variables with stock returns, volatility and trading volume after 10-K filings dates. Other studies report evidence of predictive power of stock message boards and major financial columns on volatility, returns and volume (Antweiler and Frank (2004), Chen, De, Hu, and Hwang (2013)). The related literature also studies the effect of returns on media content Garcıa (2012), the effect of media content on returns during recessions and expansions (Garcia (2013)), while a high level of similarity in firm-specific news is found to provoke higher trading aggressiveness of individual investors (Tetlock (2011)). Boudoukh, Feldman, Kogan, and Richardson (2013) find that news that can be identified and classified in certain categories have a higher impact on stock markets than unidentified news. Another area of research has been the field of corporate earnings, where Tetlock, Saar-Tsechansky, and Macskassy (2008) find that a higher percentage of negative words in news about specific firms predicts lower quarterly earnings. Furthermore, textual analysis has been used for the study of initial public offerings (IPOs). Loughran and McDonald (2013) find that higher uncertainty in filings affect first-day returns and ex post volatility, Jegadeesh and Wu (2013) give different weights on words based on the market reactions that they caused and Li (2010) studies the effect of forward-looking statements in corporate filings on future earnings and liquidity. Chouliaras (2015c) studies the effect of newswire intraday high-frequency (30-minutes) news on international stock markets during the European financial crisis. Ahern and Sosyura (2014) show evidence of firms manipulating media coverage to achieve better stock prices during mergers and acquisitions negotiations. Chouliaras (2015b) finds that monthly portfolios based on the product of annual pessimism change and the previous period returns generate returns in excess of previous winners/losers. Finally, Chouliaras (2015a) finds that 10-K pessimism negatively affects stock holdings after the filing, while institutions do not appear to have forecasting power as to how pessimistic the annual report will be.

As far as the Web Attention literature is concerned, Varian and Choi (2009) use the Google Trends Search Volume Index (SVI) to forecast economic indicators, such as car sales and unemployment claims[3]. Da, Engelberg, and Gao (2011) find that a higher SVI for stocks in Russell 3000 forecasts higher returns in the next two weeks, an effect which reverses within one year. Da, Engelberg, and Gao (2015) use queries that may concern households, such as "recession", "unemployment", "bankruptcy" and create an investor sentiment index which can forecast return reversals, volatility spikes and mutual fund movements from the stock to the bond market.

Regarding financial returns, we use the approach proposed by Bae, Karolyi, and Stulz (2003) and

[1] `http://www.ravenpack.com/`
[2] `http://www.machinereadablenews.com/`
[3] Google Trends can be found at: `http://www.google.com/trends/explore#cmpt=q`

Boyson, Stahel, and Stulz (2010) who examine the coincidence of extreme return shocks across groups of countries. A number of authors have used this methodology[4]. We mostly study extreme returns days, because these are the days where the biggest potential losses (and gains) occur for investors, and this is where one would expect the effects to be most powerful. Nevertheless, our findings are robust even when taking all days into account in a regression framework.

Our study is related to Tetlock (2007), Garcia (2013) and Garcıa (2012) as far as the analysis of News Flow is concerned and to Da et al. (2015) regarding the Web Attention (SVI). The main contributions to the previous literature are: first and foremost, we employ a database of over 24,000 news articles from some of the biggest international news sources. We perform a cross-media analysis since we take into account all relevant news items from the entire news sources selected, while the previous studies typically use one or two columns from one or two newspapers; second, we study local and cross-country effects, while the previous literature mainly deals with the effects of specific financial columns on the US stock exchange; third, we investigate the interplay between financial returns, News Flow and Web Attention (SVI). Blending these research strands allows interesting new insights about the generation and the impact of new information. Since it is well accepted that the most vulnerable eurozone countries -the Euro-periphery group- were the most badly hit by the Euro-crisis, our main interest is to study for the crisis transmission from the Euro-periphery group to the other two groups (Euro-periphery vs. Euro-core, Euro-periphery vs. Non-euro). Additional questions that we are addressing include: Is it the news media and their tone that draws the attention of individuals, or is the Web Attention (SVI) that precedes and then the media catch up with the stories? Do financial returns lead or lag the News Flow and the Web Attention (SVI) factors? Or are these three entities inseparably intertwined, where each of them provides feedback loops and affects the other two?

3. The Data

The main area of study for this paper is the European Union area. Thus, we create three country groups: the Euro-periphery group contains the periphery eurozone countries (Portugal, Ireland, Italy, Greece, Spain), the Euro-core group contains the core countries of the Eurozone (Germany, France, the Netherlands, Finland, Belgium), and the Non-euro group contains the major European Union (but not Euro) countries (Poland, Sweden, Czech Republic, UK, Denmark)[5]. The rationale behind these three groups is the following: the Euro-periphery group consists of the five Eurozone countries that were most severly hit during the Euro-crisis. The Euro-core group consists of the biggest countries (in terms of market capitalisation) that are member-countries of the Eurozone. The last group, the Non-euro group, consists of countries that are not part of the Eurozone, but are part of the European Union, which is a major trade union, with free movement of capital and individuals throughout all member states. We want to examine the three groups separately because they may have different degrees of integration and dependence among them. We examine the period from 01/01/2004 till 13/03/2013 using daily financial data obtained from the Thomson Reuters Datastream, News Flow data from Dow Jones Factiva and Web Attention data from the Google Trends. We also split our sample in three subperiods (Pre-crisis, US-crisis and Euro-crisis) to be able to make comparisons between normal and abnormal times in the financial markets:

[4]See, for example, Christiansen and Ranaldo (2009), Gropp, Duca, and Vesala (2009), Chouliaras and Grammatikos (2015).

[5]We take the biggest five stock markets from each group using the market capitalisation ranking (as of 2011) from `http://www.indexmundi.com/facts/indicators/CM.MKT.LCAP.CD/rankings`

- the Pre-crisis period (from 1 January 2004 till 26 February 2007)
- the US-crisis period (from 27 February 2007 till 7 December 2009).[6]
- the Euro-crisis period (from 8 December 2009 till the end of our sample period, 13 March 2013).

On 27 February 2007, the Federal Home Loan Mortgage Corporation (Freddie Mac) announced that it will no longer buy the most risky subprime mortgages and mortgage-related securities. On 8 December 2009, the Greek debt was downgraded by Fitch from A- to BBB+, with a negative outlook.

3.1. News Flow and Web Attention (SVI)

One can conjecture that changes in relevant information flows is one of the significant factors that affect stock prices. Obtaining usable indicators of information flows and their sentiment about the evolving Euro-crisis and relating them directly to the market returns is the focus of this paper.

We extract and analyse from Dow Jones Factiva[7] news articles covering the test period from January 1st, 2004 till March 13th, 2013. We collect news articles from seven sources: *Dow Jones Newswires, Thomson Reuters, Financial Times, The Wall Street Journal, The New York Times, The Telegraph and The Times.* We use these seven sources because first of all they returned the greatest number of news items for our queries and secondly because they are undoubtedly among the most popular news sources worldwide. Dow Jones Newswires and Thomson Reuters give news items in newswires form, capturing news in real time. The Wall Street Journal and The New York Times are the main points of reference from the United States, and The Financial Times, The Telegraph and The Times are the main European news papers for the financial markets. We include both content from the print and the online editions (where available) from all our seven sources. For each Euro-periphery country the relevant stories are obtained by a query searching for news that include the name of the country plus one of the following terms each time: crisis, debt, economy, deficit, default. For example, for Greece the news were retrieved by searching for news stories containing any of the terms:

- "greek crisis"
- "greek debt"
- "greek economy"
- "greek deficit"
- "greek default"

The same applies to all five Euro-periphery countries. The importance of these search terms in the period examined is obvious and follows closely the search terms used in Google Trends (see below). A news item that contains the term Greek crisis is certainly related to the crisis in Greece. "Greek debt" is relevant since the european crisis is also a debt crisis. The search term "greek economy" is included in order to capture the stories about the nation's economy. The "greek deficit" component is included since a lot of discussion is made around the deficits of the countries and the deficit is obviously one of the main factors to assess the financial performance of a nation. Finally, the "greek default" component captures the sovereign default risk debate, since the fear of countries defaulting elevated at various time points during the crisis. These five search terms were also found to be the most relevant key words used in Google searches.

[6]We use 27 February 2007 as the start of the financial crisis, as used by the Federal Reserve Bank of St. Louis in their Timeline of Events and Policy Actions. The timeline can be found at `http://timeline.stlouisfed.org/index.cfm?p=timeline`.

[7]Dow Jones Factiva can be found in `http://www.dowjones.com/factiva/index.asp`

3.1.1. Preprocessing the news data

These selection criteria result in a total of 110,800 news articles. As a first step we exclude duplicate articles that can reach very high numbers. Especially in newswires (*Dow Jones Newswires* and *Thomson Reuters*), it is very common that the same (or highly similar) pieces of information are redistributed, even up to ten times or more. This can cause problems since there is the possibility that a small number of news dominates the news sample, simply because it is being delivered multiple times, with none (or insignificant) changes. Furthermore, newspapers have print and online editions (which we both take into account), and it is very common that the same information is first uploaded on the Web (online edition), and then printed on the regular newspaper edition. Another problem that this repeated information can cause is that information that is available on the Web at day t, might be released at the press edition at time t+1. But in reality, this information belongs to time t, not at both time periods. Moreover, various news sources (such as the Wall Street Journal) have multiple editions (WSJ US, WSJ Asia, WSJ Europe). Many times the same information is published in these editions, even with different dates, since the time zone differences can be quite significant. Thus, since we want to study the unique impact of information at the day it was first released, we keep the news item published first and discard all duplicates after the distribution of the first news item. For each country, for the subset of news that are timestamped, we keep the news that were released until the stock market was operating on that day. If a news item was released after the stock market has closed for the day, we set the day of this news item to t+1, because the effect of this news on the stock market will at best be at t+1. After this preprocessing step, the number of news articles falls to 58,741.

3.1.2. Keywords in titles as a determinant of news items relevance

Another issue of concern has to do with the fact that news items that contain a set of keywords, do not necessarily concern only this topic. This is true especially for newswires. Very often multiple pieces of information are released through newswires in the same news item, covering multiple topics, each one of them occupying no more than a few lines of the overall news item. To deal with this problem, we select only items that contain one or more of a set of keywords in their title since the title is perhaps the best signal of the article content. The title keywords for Greece are the following: "greece", "greece's", "greek", "greeks", "hellas", "hellenic". Similar keywords are used for the other four Euro-periphery countries. After these preprocessing steps are applied, the news sources and the total number of news items appear in Table 1, having a total of 24,402 news items:

Insert Table 1 here

We see from Table 1 that Pre-crisis most news are about Italy and Spain, but in the Euro-crisis period most news concern Greece. Most news overall are obtained from the newswires (Dow Jones and Thomson Reuters: 6,536 and 6,609 respectively out of a total of 18,786 news articles in the Euro-crisis). One can easily notice that the amount of news surrounding each country increased dramatically from the Pre-crisis to the Euro-crisis periods. For example, there were a total of 307 articles regarding Greece in the Pre-crisis period, and this number jumped to 11,483 for the Euro-crisis period.

3.1.3. *Textual Analysis and Web Attention (SVI)*

Using textual analysis, based on the Loughran and McDonald (2011) dictionary[8], we measure the positive media content as in Garcıa (2012) and Garcia (2013): $G_t = \sum_i \frac{g_{it}}{w_{it}}$, calculated as the percentage of positive words over the total number of words of day t. The symbol g_{it} stands for the number of positive words in all relevant articles on day t, and w_{it} stands for the total number of words in all relevant articles on day t. We do the same for the negative words, obtaining the negative media content as $B_t = \sum_i \frac{b_{it}}{w_{it}}$, with b_{it} denoting the negative words in all articles of day t. Thus, we obtain the *Pessimism* on day t as the difference between the negative and the positive media measures:

$$Pessimism_t = B_t - G_t \tag{1}$$

The *Pessimism* is calculated for every Euro-periphery country, for every day. Then, the Euro-periphery *Pessimism* (P_t) is calculated as the average of the Euro-periphery pessimism factors on every day:

$$P_t = \frac{\sum_{j=1}^{5} Pessimism_{j,t}}{5} \tag{2}$$

where $Pessimism_{j,t}$ is the *pessimism factor* for the Euro-periphery country j (j takes values 1 to 5, one for each of the Euro-periphery countries).

A second way to measure news pessimism is by calculating the *Weighted Pessimism* (WP_t), defined as the weighted average of the pessimism of the five Euro-periphery countries:

$$WP_t = \frac{\sum_{j=1}^{5} Pessimism_{j,t} N_{j,t}}{\sum_{j=1}^{5} N_{j,t}} \tag{3}$$

where $N_{j,t}$ stands for the number of relevant news (on day t) for country j.

Another metric we use is the *News Count* (N_t) which is the total number of articles written in a day regarding any of the Euro-periphery countries.

During the Pre-crisis period, the average pessimism of news was 0.399%, while the weighted pessimism was 1.067%. To account for the fact that news stories tend to measure higher on pessimism even during normal periods, we estimate another metric, which we call the *Abnormal Pessimism Count* (AP_t) which measures the number of articles for all Euro-periphery countries with a pessimistic content which is higher than the Pre-crisis average pessimistic content (0.399%). A final metric is the *Abnormal Weighted Pessimism Count* (AWP_t) where instead of the Pre-crisis pessimism, we use the Pre-crisis Weighted pessimism (1.067%) as a threshold to count the number of pessimistic news.

Finally, we attempt a connection between the financial data and the investor attention as measured by the search frequency of Google Trends via the Search Volume Index (SVI). Google Trends provides weekly (and for some frequent terms daily) time series that depict how much a key term (or terms) was searched for via the Google Search Engine for a certain period of time. Our SVI data are daily, since the crisis queries were searched in high volumes during the crisis periods. Google is by far the most popular search engine in the world, with an 88.8% market share as of June 2013[9]. Thus, it is

[8]The dictionary can be found at `http://www3.nd.edu/~mcdonald/Word_Lists.html`

[9]Source: `http://www.karmasnack.com/about/search-engine-market-share/`

safe to assume that it captures the worlwide interest of the (Internet) population as measured by the searches the individuals perform worldwide. Moreover, as mentioned in Da et al. (2011), when someone searches for something on Google (be it a stock, a bond or information about the crisis), he certainly is interested in it. Thus, Google Trends provide a direct measure of Web Attention. Especially in crisis times, one could argue that the SVI can capture the uncertainty and the interest over topics and issues that trouble the markets and the nations and attract the investors' interest worldwide. We hasten to add, however, that the information about the SVI itself is not publicly available in real time. Investors can know about it only with a time delay. The Google Trends SVI is:

$$\textit{Web Search Volume Index: } SVI_{jt} = k, k = 0, ..., 100 \tag{4}$$

SVI_{jt} is a scaled time series taking a discrete value (0 to 100) for time t (0 meaning the query was not searched at all on time t, and a 100 when it was most searched for in the given time frame), based on the number of searches made via the Google Search Engine for a specific query and time period, with j once more taking values 1 to 5 for each one of the five Euro-periphery countries. We then calculate the average SVI for the five (5) Euro-periphery countries. Depending on the popularity of the query, Google Trends provides a time series of monthly (least searched), weekly, or daily (most searched) frequency. If a query is not searched enough for Google's threshold, no results are returned for this period[10]. Since we are mainly interested in the Euro-crisis period, and especially the Euro-periphery countries, we decided to proceed with the same sets of key search terms that we used in our News Flow analysis before, each one corresponding to a country[11,12].

Table 2 summarizes the News Flow and Web Attention (SVI) variable definitions:

Insert Table 2 here

The graphical illustrations for the SVI for Greece appears in Figure 1:

Insert Figure 1 here

One can clearly see that there are time periods where the Web Attention spikes, in other words periods where people were searching a lot in the Web (via the Google Search Engine) using Greek crisis related queries.

The summary statistics for the News Flow and Web attention factors for the Pre-crisis, the US-crisis and the Euro-crisis subperiods appear in Table 3:

Insert Table 3 here

[10]The maximum time period for which daily data can be obtained by Google Trends is three months for each query. And every Google Trends time series returned is scaled with the maximum value for the specified timeperiod. For this reason, we scaled all three months time intervals for each country with a common scaling factor which was the day with the most searches in the entire time period, thus obtaining a homogenized scaling for the entire period.

[11]For each query, for example "Greek crisis", Google Trends provides 5 related search terms and their popularity. The choice of the 5 terms used in Google Trends and the News Flow was partially influenced by this popularity. Moreover the syntax of the Google queries was modified slightly also in function of their popularity. For example we used "Greece crisis" instead of "Greek crisis", because "Greece crisis" was much more searched for. Of course the two queries are referring to the same entities and thus it is safe to claim the two queries are equivalent.

[12]No daily data were available for these search queries for the Pre-crisis and the US-crisis period, thus the SVI analysis is done only for the Euro-crisis period.

We notice that the News Flow factors significantly increase from the Pre-crisis to the Euro-crisis period. More specifically, the mean value of the Pessimism factor increased from 0.399% Pre-crisis to the value of 1.574% in the Euro-crisis period. As far as the News Count factor is concerned, it increased from an average of 3.279 Euro-periphery articles per day during the Pre-crisis period to an average of 22.093 articles per day during the Euro-crisis period.

3.2. Stock Returns and Extreme Returns

We employ the classical measures of stock returns, but we mostly focus on extreme stock returns as a measure better tailored to capture "exceptional" performance typical in a financial crisis. Table 4 shows the summary statistics of the percentage (%) log-returns of the major stock market country indices[13].

Insert Table 4 here

For the Pre-crisis period all groups of countries had positive mean stock returns, consistent with the overall optimism in the financial markets. The best performing markets were firstly the Non-euro countries (+0.101%) followed by the Euro-periphery countries (+0.089%). Regarding the standard deviation we see that we have rather low values for all country groups as this was a period of relative calmness for the financial markets. During the US-crisis period all country groups had a negative mean return. The Euro-periphery countries were the most badly hit with a mean (daily) return of -0.074%, followed by the Euro-core countries which had a mean (daily) return of -0.049%, then the Non-euro with a -0.024%. Compared to the Pre-crisis period, the standard deviations have increased significantly in the crisis periods for all three country groups. The descriptive statistics for the Euro-crisis period show that once more the Euro-periphery countries were the most severely affected from the financial crisis (mean daily return of -0.017%). The other two groups have positive mean returns for this period, indicating that they were able to better cope with the crisis. The standard deviations were lower than in the US-crisis period but still higher than the Pre-crisis period, especially for the Euro-periphery group.

The correlations among the information variables and the country group stock indices appear in Table 5.

Insert Table 5 here

There exist some significant changes when comparing the three subperiods. The correlations among the stock indexes of the three country groups generally increased during the US-crisis period, declined slightly in the Euro-crisis period but remained at higher levels than the Pre-crisis period. The correlations between the information variables and the group stock returns Pre-crisis were small (and positive for the Euro-core and the Non-euro groups) during the US-crisis and the Euro-crisis periods almost all correlations become negative and much larger in magnitude. For example, the correlation between the Weighted Pessimism and the Euro-core group, has a value of 0.003 in the Pre-crisis period. This value becomes 0.008 in the US-crisis period, but becomes negative and equal to -0.06 in the Euro-crisis period. The SVI is also negatively correlated with all three group stock indexes during the Euro-crisis period. As a matter of fact, all correlations of the information variables with

[13] All stock indices used are the Thomson Reuters Datastream indices created for each country

the stock markets are negative during the Euro-crisis period, which indicates that higher news flow values during the Euro-crisis period were associated with negative stock market returns for all three country groups[14]. Naturally, the information variables are quite significantly (positively) correlated among themselves.

3.2.1. Extreme Returns

According to Bae et al. (2003) an extreme return is one that lies below (or above) the lowest (or the highest) quantile of the marginal return distribution respectively. The method deals with the counts of joint occurrences of extreme returns within a group on a particular day. The original approach studies the extreme returns counts for the entire test period, taking as thresholds for extreme returns the 5th and the 95th percentiles. In our case, since we are mostly interested in the dynamics in the Pre-crisis and the Euro-crisis periods, we choose as thresholds the 10th and the 90th percentiles in order to have a sufficient number of observations, as in Boyson et al. (2010) (our findings are robust to the 5th and 95th percentiles). For each country we consider returns below the 10th percentile as extreme bottom returns and those above the 90th percentile as extreme top returns for this country.

This procedure is followed for all countries in all groups. Top extreme returns are treated separately from bottom extreme returns. Bottom and top extreme returns counts for the entire period (1/1/2004-13/3/2013) are reported in Table 6. For each country we calculate the days for which it had an extreme (bottom or top) return separately. Then, the extreme returns count for each group and day is given as the number of countries of the group that have extreme returns on that specific day.

Insert Table 6 here

The left side of Table 6 presents bottom return counts and the right side shows top return counts. A count of i units for bottom returns is the joint occurrence of i extreme bottom returns on a particular day for a specific group. By counting the total number of days with extreme returns of a given count and identifying which countries participate in those events and how often we have a good overview of the extreme returns for each country and group of countries.

The Greek stock market had the most days (106) on which it was the only country experiencing a bottom extreme return, followed by Ireland (56 days) and Portugal (37 days). A total of 54 days are reported for the Euro-periphery countries on which all of them experienced extreme bottom returns (109 days for the Euro-core, 55 days for the Non-euro). There were 40 days where all Euro-periphery countries experienced an extreme top return (91 days for the Euro-core, 28 days for the Non-euro). The Czech Republic had the most days (84) as the only country experiencing an extreme bottom return for the Non-euro group and once more had the most days (95) with extreme top returns.

3.2.2. News Flow, Web Attention and Extreme Returns

The graphical illustrations of the information variables during the Euro-crisis period along with the bottom extreme returns count for the Euro-periphery group appear in Figure 8[15].

[14]Correlations with the extreme bottom returns (see Section 3.2.1) are significant and much higher than these of raw returns.

[15]the two Abnormal Pessimism Counts (not presented) are very similar to the overall News Count (albeit at a smaller scale)

Insert Figure 8 here

There exists a correlation between the bottom extreme returns count of the Euro-periphery group and the four News Flow factors during the Euro-crisis. "Spikes" of extreme bottom returns (or "extreme bottom returns clustering") notably in the periods April-June 2010 and May-December 2011 seem to be related to the evolution of the information variables.

In order to examine the effects of the information variables on the probability of extreme returns of all three country groups, we consider a polychotomous variable, like Bae et al. (2003) and Boyson et al. (2010). In the theory of multinomial logistic regression models, if P_i is the probability of an event category i out of m possible categories, a multinomial distribution can be defined by

$$P_i = P(Y_t = i|x_j) = \frac{G(\beta_i' x_j)}{1 + \sum_{j=1}^{m-1} G(\beta_j' x_j)}, \tag{5}$$

where x is the vector of covariates and β_i the vector of coefficients associated with the covariates. The function $G(\beta_i' x)$ many times takes the form of an exponential function $exp(\beta_i' x)$, in which case Equation 5 represents a multinomial logistic (or multinomial logit) model.[16] To capture the range of possible outcomes, and yet have a concrete model, we have a total of six categories: 0, 1, 2, 3, 4, and 5 extreme returns. For a model that has only constants, m-1, or five parameters, need to be estimated. But for every covariate added to the model, such as the daily SVI, five additional parameters need to be estimated, one for each outcome. The top and the bottom extreme returns are estimated separately. Finally, we compute the probability of an extreme return count of a specific level, P_i, by evaluating the covariates at their unconditional values,

$$P_{ij}^* = \frac{exp(\beta_i' x_j^*)}{1 + \sum_{j=1}^{m-1} exp(\beta_j' x_j^*)}, \tag{8}$$

where x_j^* is the unconditional mean value of x_j.
The coefficients that are given by a multinomial logistic regression compare the probability of a given outcome with the base outcome (in our case the outcome 0 is the base outcome - i.e. the outcome where no country has an extreme return). As mentioned in Greene (2003), the coefficients of such a model are not easy to interpret.[17] In multinomial logistic regressions the coefficients correspond to probabilities. Thus, these partial effects give us the marginal change in probability for a unit change in

---

[16]Such models are estimated using maximum likelihood, with the log-likelihood function for a sample of n observations given by

$$logL = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij} log P_{ij}, \tag{6}$$

where I_{ij} is a binary variable that equals one if the ith observation falls in the jth category, and zero otherwise. Goodness-of-fit in these models is measured using the $pseudo-R^2$ approach of McFadden (1974) where the unrestricted (full model) likelihood, L_Ω, and restricted (constants only) likelihood, L_ω, functions are compared:

$$pseudoR^2 = 1 - [logL_\omega / logL_\Omega]. \tag{7}$$

[17]This is why in these models it is necessary to differentiate 5 in order to obtain the partial effects of the covariates on the probabilities

$$\delta_{ij} = \frac{\delta P_{ij}}{\delta \beta_i} = P_{ij}[x_j - \sum_{k=0}^{J} P_{ik}\beta_k] = P_{ij}[\beta_j - \bar{\beta}] \tag{9}$$

where $\bar{\beta} = \sum_{k=0}^{J} P_{ik}\beta_k$, the weighted average of every subvector of β.

the independent covariate. In such models we are interested in seeing whether these marginal effects are statistically significant or not.[18]

Y_t counts the number of extreme returns and takes the value i when i stock market indices have extreme returns (top or bottom) on day t. Y_t is calculated separately for the Euro-core, the Euro-periphery, and the Non-euro groups. Then, in Equation 8, P_i is equal to $P(Y_t = i|x_t)$ where $Y_t = 0, 1, 2, ...k$ is the extreme returns count variable that is created for the Non-euro, Euro-periphery and Euro-core respectively. So, we have k=5 for all three country groups, where x_t is the explanatory variable (covariate), on day t. In Equation 8, the argument of the exponential part (representing the logistic function) is a function of the covariate (x_t) and the coefficient (the beta). This function is a linear expression of the arguments. Let's call it $g_i(t)$. We will use this function to study the effect of information variables on stock returns. For each group, the (daily) stock returns are calculated as the equally weighted average of the stock returns of the countries that belong in each respective group.

4. Empirical Findings

4.1. News Flow, Web Attention and Extreme Returns

In equation 5, the dependent variable is the number of bottom (or top) extreme returns for one of the three country groups while the independent variable is each one of the information variables for the Euro-periphery group. Thus, the logistic regression $G(\beta_i' x) = exp(g_i(x_t))$ of equation 5 has the following form for $g_i(x_t)$:

$$g_i(x_t) = b_{0i} + b_{1i} X_{it} \tag{11}$$

where i=0, 1, 2, 3, 4, 5 for each country group, the extreme returns count for the group. X_{it} takes the values of the six information variables we calculated earlier. The results for the effect of the information variables on extreme returns during the Euro-crisis appear in Table 7.

Insert Table 7 here

One can see that the Web Attention is the most significant variable for all three groups (in terms of the highest R^2). The Weighted Pessimism is the next most significant (see R^2). These two variables seem to be the most relevant for extreme returns during the crisis. For the Euro-periphery group, an increase of one in the Euro-periphery Web Attention increases the probability of all five countries having bottom extreme returns in the same day by 0.2%, while an increase of 1% in the

[18] These marginal effects may even have different signs than the corresponding coefficients, since the derivative $\frac{\delta P_{ij}}{\beta_{ik}}$ can have a different sign than the coefficient β_{jk}. To elaborate a little further on why it is crucial that marginal effects are calculated for such models, it is known that the coefficients of a multinomial logistic are obtained from comparing the probability of a given outcome with the base outcome. In our case, the outcome is 0, in other words, no extreme returns in the group. Thus, the estimated coefficient for covariate x_{13} for outcome 3, which is β_{13} and is the coefficient for the 1st covariate, calculated for the 3rd outcome, measures the probability of having an outcome equal to 3 (3 extreme returns in the group), instead of an outcome 0 (no extreme returns in the group), for a unit change in the covariate x_{13}. But in reality, there is also the possibility of having the outcome 2 instead of 0 for a unit change in covariate x_{13}. This is exactly why we need the marginal effects, to calculate the probabilities associated with a unitary covariate change in adjacent categories, and not taking as an alternative only the base outcome (0 in our study). This happens because the coefficients of a multinomial logistic regression model exhibit what is known as the "log odds ratio" property:

$$\ln \frac{P_{ij}}{P_{i0}} = \beta_i' x_j \tag{10}$$

Euro-periphery Weighted Pessimism increases the probability of this outcome by 1.4%. For the Non-euro group, an increase of one in the Euro-periphery Web Attention increases the probability of all five countries having bottom extreme returns in the same day by 0.2%, while an increase of 1% in the Euro-periphery Weighted Pessimism increases the probability of this outcome by 1.3%. An increase in the Euro-periphery Web Attention is associated with an increase in the probability of all five Euro-core countries having extreme bottom returns on the same day by 0.5%, and an increase of 1% in the Euro-periphery Weighted Pessimism increases the probability of this outcome by 2.9%.

The summary results for the three periods and for all the information variables appear in Table 8.

Insert Table 8 here

Each row of this table contains a separate model specification. In other words, a separate model is estimated with *Pessimism* as the independent variable, another with *Weighted Pessimism* as the independent variable et cetera. The symbol "+" denotes a positive statistical significant effect. The number of "+" denotes the number of statistical significant coefficients. As we see in Table 8, for the Pre-crisis period there is no effect for all three country groups and all variables.

During the US-crisis, we see significant effects mostly for the Euro-periphery and the Non-euro groups for the bottom extreme returns. The fact that the coefficients are positive, means that an increase in the variables increases the probabilities of the underlying outcomes. We also see a few significant coefficients for the top extreme returns for the *Pessimism* and the *Weighted Pessimism* (mainly for the Euro-core and Non-euro groups). In a turbulent period, it should be expected that extreme bottom return days are followed by extreme top return days, and vice versa, because of higher uncertainty. Even during the US-crisis, the effects for Euro-periphery are much more significant for the bottom than for the top extreme returns (14 significant coefficients for the bottom returns versus 3 significant coefficients for the top returns), while for the Non-euro and the Euro-core the image is more mixed.

The effects on bottom extreme returns are stronger for the Euro-crisis period. We notice effects for all three groups and all variables. The marginal effects are positive, which means that a higher value in any of the information variables is associated with higher probabilities of extreme bottom returns in the groups' stock markets. Thus, the Euro-periphery information variables do not only affect the probabilities of extreme bottom returns for the Euro-periphery group, but also for the Non-euro and the Euro-core groups. Finally, the *Web Attention*, for which daily data exist only for the Euro-crisis period), exhibits significant and positive marginal effects for twelve out of fifteen bottom extreme returns. In other words, more Web Attention for the Euro-periphery during the Euro-crisis is associated with higher probabilities of extreme returns for all three groups we studied. The effect on the probabilities of top returns is insignificant for most of the cases.

These results bring some useful implications for investors and policymakers: The quantity of news, the tone of news and the Web Attention are closely related with the probabilities of extreme returns (especially in times of crisis). We find that these effects are not only contained within the borders of the group that the news or the Web Attention metrics concern (the Euro-periphery group), but they also spread out across groups. For all three groups the probabilities of extreme bottom returns are affected in a positive way (e.g. more pessimistic news about the periphery are associated with higher probabilities of extreme returns for the other groups). Based on these results, one can argue in favor of "transmission" or "propagation" of News Pessimism and of Web Attention (SVI) across groups during crisis times. Thus it might be useful for investors and policymakers to be aware of these dynamics and effects when making investment or policy decisions.

5. Robustness and alternative specifications

To verify the robustness of our results, as a first robustness check, instead of 10% and 90% extreme returns cutoffs, we used the 5% and 95% percentages. The results are robust in this change. Furthermore as a second robustness check, instead of the raw returns, we calculated extreme returns on the standardized residuals of a GARCH(1,1) model, accounting for the time-varying volatility effects, since in periods of high volatility, extreme returns are more probable. In order to calculate the volatility, we move in line with Christiansen and Ranaldo (2009), estimating a AR(1)-GARCH(1,1) model for each group's average returns:

$$Ret_t^{group} = c_0 + c_1 Ret_{t-1}^{group} + \epsilon_t \tag{12}$$

where ϵ_t $N(0, \sigma_t^2)$ and the variance follows a GARCH(1,1) process:

$$\sigma_t^2 = c_2 + c_3 \sigma_{t-1}^2 + c_4 \epsilon_{t-1}^2 \tag{13}$$

As far as the Google Trends Web Attention (SVI) robustness checks are concerned, apart from the average Euro-periphery Web Attention (SVI), we also calculated the scaled Web Attention (SVI) for the Euro-periphery using the most searched query as a common scaling factor for all country Web Attention (SVI) time series. Combining queries in Google Trends provides a common scaling factor for all the time series (providing a unique maximum equal to 100), instead of scaling each one separately to its own maximum of 100. The results were found to be robust.

Last but not least, on top of the multinomial logistic regressions, Ordinary Least Squares (OLS) regressions as well as quantile regressions verified that our results hold, finding a negative and statistical coefficient (i.e. a higher Pessimism associated with negative stock returns) for the crises periods.

6. Conclusion

We use daily stock market data from January 2004 till March 2013 and 24,402 news articles from seven major international news sources to examine whether six "information variables" related to the Euro-periphery countries affect the probabilities of extreme stock returns in three groups of countries: the Euro-periphery, the Euro-core and the major Non-euro (European Union -but not euro- countries.

We find evidence that the Euro-periphery information variables have a statistically significant and positive effect on the probabilities of extreme returns not only for the Euro-periphery countries but also for the Euro-core and the major European Union -but not euro- countries. The effect in the vast majority of cases is stronger for the bottom extreme returns. The implications of the overall findings are quite significant for investors who may want to diversify their portfolios and should be aware of the stock indices movement dynamics and of how extreme shocks propagate from one group of countries to the others, affecting their portfolios' overall risk. Furthermore, these findings are useful for policy makers who need to assess policy decision making in times of extreme shocks and uncertainty (such as crisis times). Due to the high complexity of financial markets and the extremely high level of available information from the press and the web, agents can incorporate information extracted from textual analysis of news items and trends on the web that may be associated with the market movements.

Future research could study alternative data mining and textual analysis techniques in order to further improve the quality of the information variables. On top of that, the effect of policy making and textual analysis on official meetings and announcements (e.g. bailout announcements) could also be a field of research for subsequent studies.

References

Ahern, K. R., Sosyura, D., 2014. Who writes the news? corporate press releases during merger negotiations. The Journal of Finance 69, 241–291.

Antweiler, W., Frank, M. Z., 2004. Is all that talk just noise? the information content of internet stock message boards. The Journal of Finance 59, 1259–1294.

Arezki, R., Candelon, B., Sy, A., 2011. Sovereign rating news and financial markets spillovers: Evidence from the european debt crisis. IMF working papers pp. 1–27.

Bae, K.-H., Karolyi, G. A., Stulz, R. M., 2003. A new approach to measuring financial contagion. Review of Financial studies 16, 717–763.

Beetsma, R., Giuliodori, M., de Jong, F., Widijanto, D., 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.

Boudoukh, J., Feldman, R., Kogan, S., Richardson, M., 2013. Which news moves stock prices? a textual analysis. Tech. rep., National Bureau of Economic Research.

Boyson, N. M., Stahel, C. W., Stulz, R. M., 2010. Hedge fund contagion and liquidity shocks. The Journal of Finance 65, 1789–1816.

Chen, H., De, P., Hu, J., Hwang, B.-H., 2013. Wisdom of crowds: The value of stock opinions transmitted through social media. Review of Financial Studies .

Chouliaras, A., 2015a. Institutional investors, annual reports, textual analysis and stock returns: Evidence from sec edgar 10-k and 13-f forms .

Chouliaras, A., 2015b. The pessimism factor: Sec edgar form 10-k textual analysis and stock returns .

Chouliaras, A., Grammatikos, T., 2015. Extreme returns in the european financial crisis. Working Paper .

Chouliaras, A. S., 2015c. High frequency newswire textual sentiment: Evidence from international stock markets during the european financial crisis. Available at SSRN 2572597 .

Christiansen, C., Ranaldo, A., 2009. Extreme coexceedances in new eu member states' stock markets. Journal of banking & finance 33, 1048–1057.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. The Journal of Finance 66, 1461–1499.

Da, Z., Engelberg, J., Gao, P., 2015. The sum of all fears investor sentiment and asset prices. Review of Financial Studies 28, 1–32.

Garcıa, D., 2012. The kinks of financial journalism .

Garcia, D., 2013. Sentiment during recessions. The Journal of Finance 68, 1267–1300.

Greene, W. H., 2003. Econometric analysis. Pearson Education India.

Gropp, R., Duca, M. L., Vesala, J., 2009. Cross-border bank contagion in europe. International Journal of Central Banking 5, 97–139.

Jegadeesh, N., Wu, D., 2013. Word power: A new approach for content analysis. Journal of Financial Economics 110, 712–729.

Li, F., 2010. The information content of forward-looking statements in corporate filings—a naïve bayesian machine learning approach. Journal of Accounting Research 48, 1049–1102.

Loughran, T., McDonald, B., 2011. When is a liability not a liability? textual analysis, dictionaries, and 10-ks. The Journal of Finance 66, 35–65.

Loughran, T., McDonald, B., 2013. Ipo first-day returns, offer price revisions, volatility, and form s-1 language. Journal of Financial Economics .

McFadden, D., 1974. The measurement of urban travel demand. Journal of public economics 3, 303–328.

Mink, M., De Haan, J., 2013. Contagion during the greek sovereign debt crisis. Journal of International Money and Finance 34, 102–113.

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

Tetlock, P. C., 2011. All the news that's fit to reprint: Do investors react to stale information? Review of Financial Studies 24, 1481–1512.

Tetlock, P. C., Saar-Tsechansky, M., Macskassy, S., 2008. More than words: Quantifying language to measure firms' fundamentals. The Journal of Finance 63, 1437–1467.

Varian, H. R., Choi, H., 2009. Predicting the present with google trends. Google Research Blog 2.

Table 1: Number of news stories per country and source for the Pre-crisis, US-crisis and Euro-crisis periods.

Pre-crisis: 1 January 2004 - 26 February 2007						
Source	**Portugal**	**Ireland**	**Italy**	**Greece**	**Spain**	**Total**
Dow Jones Newswires	69	132	613	108	176	1098
Thomson Reuters	128	186	386	192	408	1300
Financial Times	8	28	80	0	125	241
The Wall Street Journal	4	5	28	4	14	55
The New York Times	0	9	13	1	4	27
The Telegraph	0	13	10	1	2	26
The Times	2	98	16	1	8	125
Total	211	471	1146	307	737	2972
US-crisis: 27 February 2007 - 7 December 2009						
Dow Jones Newswires	67	190	371	151	223	1002
Thomson Reuters	58	251	247	194	462	1212
Financial Times	1	54	24	37	104	220
The Wall Street Journal	0	32	20	23	45	120
The New York Times	0	12	5	2	4	23
The Telegraph	21	14	5	2	13	55
The Times	0	98	3	1	10	112
Total	147	651	675	410	861	2644
Euro-crisis: 8 December 2009 - 13 March 2013						
Dow Jones Newswires	310	658	497	4342	729	6536
Thomson Reuters	298	661	771	3786	1093	6609
Financial Times	99	177	182	1010	262	1730
The Wall Street Journal	84	178	148	1059	302	1771
The New York Times	15	32	61	355	69	532
The Telegraph	21	81	90	531	128	851
The Times	8	175	44	400	130	757
Total	935	1962	1793	11483	2713	18786

Note: Selected news are news that pass two filters: first, for each country the news item must include the name of the country plus any of the following keywords (the two words must be next to each other): crisis, debt, economy, deficit, default. For example, for Greece the first filter selects the news stories containing any of the terms: "greek crisis", "greek debt", "greek economy", "greek deficit", "greek default"; second, for each country, the news item must contain in it's title a country keyword. For example, for Greece the second filter selects the news that passed the first filter and that furthermore contain any of the following terms in the title: "greece", "greek", "greeks", "greece's", "hellas", "hellenic". The same applies to all five Euro-periphery countries.

Table 2: News Flow and Web Attention (SVI) Variables: Names, Symbols and Definitions

Variable Name	Variable Symbol	Variable Definition
Pessimism (%)	P_t	Average of the Euro-periphery News Pessimism factors.
Weighted Pessimism (%)	WP_t	Weighted average of the Euro-periphery News Pessimism factors.
News Count	N_t	The number of news items that passed the crisis-related filters.
Abnormal Pessimism Count	AP_t	The number of news items that passed the crisis-related filters, and whose pessimism is higher than the Pre-crisis average pessimism.
Abnormal Weighted Pessimism Count	AWP_t	The number of news items that passed the crisis-related filters, and whose pessimism is higher than the Pre-crisis weighted average pessimism.
Web Attention	SVI_t	A time series taking discrete values, from 0 to 100, based on the number of searches made via the Google Search Engine for crisis-related queries, 0 means that the query was not searched at all on time t, and a 100 means that it was most searched for in the given time frame.

Table 3: Euro-periphery News Flow and Web Attention (SVI): Descriptive Statistics

Pre-crisis (1/1/2004 - 26/2/2007)						
(Euro-periphery)	Pessimism (P_t %)	Weighted Pessimism (WP_t %)	News Count (N_t)	AP_t	AWP_t	SVI
Mean	0.399	1.067	3.279	2.14	1.612	–
Std. Dev.	0.487	1.292	2.714	2.121	1.866	–
Minimum	−0.903	-4.07	0	0	0	–
Maximum	3.302	6.707	19	15	15	–
US-crisis (27/2/2007 - 7/12/2009)						
(Euro-periphery)	Pessimism (P_t %)	Weighted Pessimism (WP_t %)	News Count (N_t)	AP_t	AWP_t	SVI
Mean	0.706	1.694	3.594	2.909	2.43	–
Std. Dev.	0.690	1.323	3.117	2.726	2.39	–
Minimum	−1.333	-2.913	0	0	0	–
Maximum	4.337	6.965	18	17	15	–
Euro-crisis (8/12/2009 - 13/3/2013)						
(Euro-periphery)	Pessimism (P_t %)	Weighted Pessimism (WP_t %)	News Count (N_t)	AP_t	AWP_t	SVI
Mean	1.574	2.447	22.093	19.82	17.77	9.682
Std. Dev.	0.720	0.779	21.347	19.741	18.104	5.326
Minimum	−0.104	−0.223	0	0	0	0
Maximum	3.729	5.371	192	179	161	58.4

AP_t = Abnormal Pessimism Count
AWP_t = Abnormal Weighted Pessimism Count
SVI = Web Attention

Table 4: Country Groups Stock Indices: Descriptive Statistics

	Pre-crisis 1/1/2004 - 26/2/2007			US-crisis 27/2/2007 - 7/12/2009			Euro-crisis 8/12/2009 - 13/3/2013		
	Non-euro	Euro-periphery	Euro-core	Non-euro	Euro-periphery	Euro-core	Non-euro	Euro-periphery	Euro-core
Mean (%)	0.101	0.089	0.081	−0.024	−0.074	−0.049	0.030	−0.016	0.03
Median (%)	0.142	0.104	0.108	0.013	0.008	0.013	0.0544	0.015	0.052
Std. Dev. (%)	0.715	0.569	0.703	1.616	1.555	1.619	0.952	1.284	1.189
Minimum (%)	−4.514	−3.405	−3.338	−8.809	−8.118	−7.618	−4.467	−4.929	−5.081
Maximum (%)	3.441	2.79	2.858	8.689	7.838	8.584	5.321	9.118	6.848

Note: European countries are split in three groups: the Euro-periphery countries (Portugal, Ireland, Italy, Greece, Spain), the Euro-core countries (Germany, France, Finland, the Netherlands, Belgium) and the European Union -non Euro- countries (Poland, Czech Republic, Sweden, UK, Denmark). Country group log returns and standard deviations are calculated on the equally weighted mean portfolio of the country stock market daily returns for each group.

Fig. 1. Web Attention (SVI) for Greece.

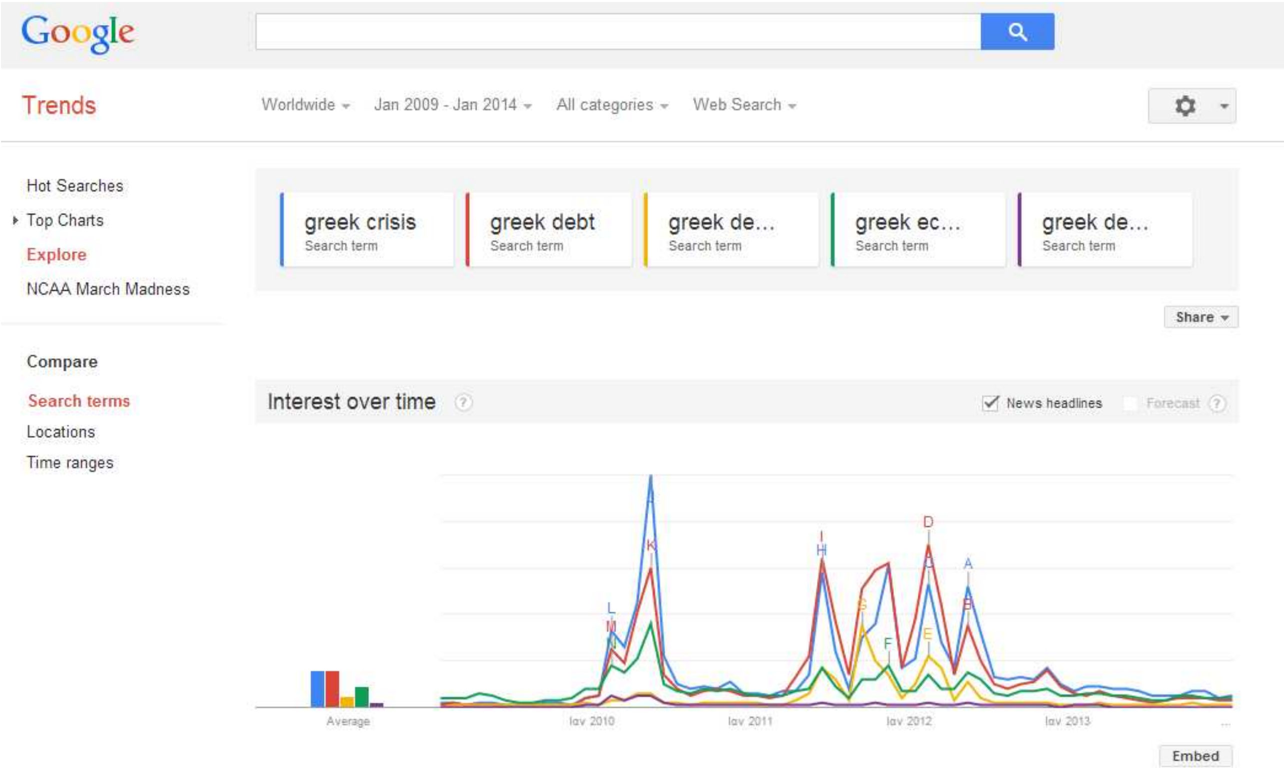

Fig. 2. Euro-periphery bottom extreme returns count, News Flow and Web Attention (SVI) factors

Table 5: Euro-periphery Information variables and Stock Returns Correlation Matrices

Pre-crisis (1/1/2004 - 26/2/2007)

	Pessimism	Weighted Pessimism	News Count	AP_t	AWP_t	SVI	Euro-periphery	Euro-core	Non-euro
Pessimism	1.000								
Weighted Pessimism	0.813	1.000							
News Count	0.463	0.264	1.000						
AP_t	0.644	0.510	0.873	1.000					
AWP_t	0.674	0.569	0.799	0.933	1.000				
Web Attention (SVI)	–	–	–			–	–		
Euro-periphery return	−0.002	−0.004	−0.020	−0.023		–	1.000		
Euro-core return	0.006	0.003	0.007	0.007	0.009	–	0.869	1.000	
Non-euro return	0.042	0.033	0.036	0.028	0.026	–	0.791	0.805	1.000

US-crisis (27/2/2007 - 7/12/2009)

	Pessimism	Weighted Pessimism	News Count	AP_t	AWP_t	SVI	Euro-periphery	Euro-core	Non-euro
Pessimism	1.000								
Weighted Pessimism	0.761	1.000							
News Count	0.589	0.309	1.000						
AP_t	0.710	0.458	0.945	1.000					
AWP_t	0.772	0.540	0.878	0.952	1.000				
Web Attention (SVI)	–	–	–	–	–				
Euro-periphery return	−0.049	−0.011	−0.069	−0.077	−0.070	–	1.000		
Euro-core return	−0.038	0.008	−0.065	−0.068	−0.060	–	0.928	1.000	
Non-euro return	−0.022	−0.006	−0.055	−0.061	−0.056	–	0.919	0.912	1.000

Euro-crisis (8/12/2009 - 13/3/2013)

	Pessimism	Weighted Pessimism	News Count	AP_t	AWP_t	SVI	Euro-periphery	Euro-core	Non-euro
Pessimism	1.000								
Weighted Pessimism	0.614	1.000							
News Count	0.296	0.205	1.000						
AP_t	0.317	0.250	0.994	1.000					
AWP_t	0.337	0.289	0.985	0.994	1.000				
Web Attention (SVI)	0.309	0.252	0.568	0.578	0.587	1.000			
Euro-periphery return	−0.079	−0.086	−0.048	−0.055	−0.068	−0.142	1.000		
Euro-core return	−0.061	−0.060	−0.047	−0.052	−0.062	−0.130	0.876	1.000	
Non-euro return	−0.047	−0.053	−0.040	−0.044	−0.054	−0.139	0.836	0.926	1.000

Note: AP_t=Abnormal Pessimism Count, AWP_t=Abnormal Weighted Pessimism Count.

Table 6: Count of bottom and top extreme daily log returns for country groups' stock indices, January 1st 2004 to March 13th 2013.

	Mean return (%) when $i = 5$	Number of bottom extreme returns						Number of top extreme returns						Mean return (%) when $i = 5$
		5	4	3	2	1	0	0	1	2	3	4	5	
Non-euro														
POL	−3.446	55	41	38	49	57	1847	1783	82	47	40	43	28	3.653
SWE	−3.727	55	45	54	54	32	1847	1783	42	56	61	53	28	4.025
CZE	−3.828	55	27	28	46	84	1847	1783	95	57	31	29	28	4.022
UK	−3.241	55	52	60	43	30	1847	1783	37	56	62	57	28	3.572
DEN	−3.370	55	47	48	42	48	1847	1783	61	42	55	54	28	3.382
Subtotal		55	53	76	117	251	1847	1783	317	129	83	59	28	
Euro-periphery														
POR	−3.253	54	66	44	39	37	1859	1817	61	45	34	60	40	3.008
IRE	−3.944	54	54	26	50	56	1859	1817	69	47	31	53	40	3.522
ITA	−3.636	54	69	45	52	20	1859	1817	20	62	56	62	40	3.678
GRE	−4.160	54	35	22	23	106	1859	1817	108	32	20	40	40	4.200
SPA	−3.503	54	64	49	44	29	1859	1817	29	56	54	61	40	3.652
Subtotal		54	72	62	104	248	1859	1817	287	121	65	69	40	
Euro-core														
GER	−2.855	109	44	32	19	36	1970	1938	46	25	31	47	91	2.530
FRA	−3.137	109	56	46	19	10	1970	1938	13	35	41	60	91	2.842
NL	−3.169	109	51	37	22	21	1970	1938	20	27	50	52	91	2.782
FIN	−3.303	109	38	24	19	50	1970	1938	48	26	26	49	91	3.240
BEL	−2.809	109	35	32	27	37	1970	1938	51	29	29	40	91	2.534
Subtotal		109	56	57	53	154	1970	1938	178	71	59	62	91	

Note: Extreme returns for daily stock index top (bottom) log returns are the ones belonging to the highest (lowest) 10% of all daily returns. The extreme returns count is defined as the joint occurrence of extreme returns (bottom or top) across different country indexes on the same day. For example, out of a total sample of 2399 trading days, there are 104 occurrences of bottom extreme returns for the Euro-periphery countries with 2 countries only, and in 23 of those days Greece is the one of the two countries with bottom extreme returns.

Table 7: The Euro-periphery information variables and the bottom extreme returns count of the three country groups for the Euro-crisis period.

	(1) Margin / SE	(2) Margin / SE	(3) Margin / SE	(4) Margin / SE	(5) Margin / SE	$Pseudo-R^2$
To Non-euro						
Pessimism	-0.014	0.001	0.012**	0.015**	0.013***	0.010
Weighted Pessimism	-0.020	0.013	0.019***	0.015***	0.013***	0.018
News count	0.000	0.000	0.001***	0.000*	0.000**	0.013
Abnormal Pessimism Count	0.000	0.000	0.001***	0.000*	0.000***	0.014
Abnormal Weighted Pessimism Count	0.000	0.001*	0.001***	0.000**	0.000***	0.017
Web Attention (SVI)	0.002	0.004***	0.003***	0.003***	0.002***	0.045
To Euro-core						
Pessimism	0.013	0.011*	-0.008	0.003	0.032***	0.016
Weighted Pessimism	0.019*	0.009	0.008**	0.010	0.029***	0.019
News count	0.000	0.000**	0.000***	0.000	0.001**	0.014
Abnormal Pessimism Count	0.000	0.000**	0.000***	0.000	0.001**	0.012
Abnormal Weighted Pessimism Count	0.000	0.000**	0.000***	0.000	0.001***	0.014
Web Attention (SVI)	-0.001	0.003***	0.002***	0.002**	0.005***	0.044
To Euro-periphery						
Pessimism	0.021	-0.003	0.016***	0.015**	0.015***	0.018
Weighted Pessimism	0.023*	0.015	0.016***	0.014**	0.014***	0.019
News count	0.001**	0.000	0.000*	0.000***	0.000***	0.014
Abnormal Pessimism Count	0.001**	0.000	0.000**	0.001***	0.000***	0.015
Abnormal Weighted Pessimism Count	0.001**	0.000	0.000**	0.001***	0.000***	0.017
Web Attention (SVI)	0.004*	0.004***	0.003***	0.003***	0.002***	0.045

Note: Columns (1) to (5) correspond to bottom extreme returns count (1 to 5). In other words, column (1) presents the marginal effects in the case of one bottom extreme return for the respective group, and columns (2),(3),(4),(5) correspond to two, three, four and five bottom extreme returns for this group. The value of 0.012 for the Non-euro Bottom Extreme Returns Count (column 3) means that an increase of one percent in the *Pessimism* increases the probability of three Non-euro countries having extreme bottom stock returns (i.e. three bottom Euro-core extreme return) by 1.2%. (***), (**), (*) : significance at 1%, 5%, 10% level.

Table 8: The Euro-periphery information variables and extreme returns: Summary results.

	Bottom tail			Top tail		
	Non-euro	Euro-core	Euro-Periphery	Non-euro	Euro-core	Euro-Periphery
Pre-crisis (1/1/2004 - 26/2/2007)						
Pessimism				+	−	
Weighted Pessimism						
News Count						
Abnormal Pessimism Count	−−				+	−
Abnormal Weighted Pessimism Count	−					
US-crisis (27/2/2007 - 7/12/2009)						
Pessimism	++	+++	+++	+++	++	+
Weighted Pessimism	+	+	++	+++	++	+
News Count	++	+	++	++		
Abnormal Pessimism Count	++	+	+++	+	+	
Abnormal Weighted Pessimism Count	++	+	++++	++	+	+
Euro-crisis (8/12/2009 - 13/3/2013)						
Pessimism	+++	++	+++	+	+	++
Weighted Pessimism	+++	+++	++++	++	++	+
News Count	+++	+++	++++	+	++−	+
Web Attention (SVI)	++++	++++	+++++	+	+	
Abnormal Pessimism Count	+++	+++	++++	+	++−	+
Abnormal Weighted Pessimism Count	++++	+++	++++	+	++−	+

Note: The number of "+" (or "-") indicate the number of statistically significant (in the 1%, 5% or 10% levels) and positive (or negative) marginal effects. For the bottom tail and the Euro-crisis period, three out of five Euro-periphery periphery Weighted Pessimism marginal effects were found to be significant and positive for the Non-euro group, i.e. an increase of one percent in the periphery News Pessimism increases the probability of extreme bottom returns for the Non-euro group in three out of five bottom extreme returns counts. The grey area corresponds to the results of Table 7

Table 9: Markov switching model. Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Regime 1				
	Estimate Std.	Error	t value	p rvalue
(Intercept)(S)	1.5307	0.0576	26.5747	$<2e-16^{***}$
Average Pessimism	0.0302	0.0278	1.0863	0.2773
Multiple R-squared:	0.004526			
Regime 2				
	Estimate	Std. Error	t value	p value
(Intercept)(S)	0.9438	0.0169	55.8462	$< 2.2e-16^{***}$
Average Pessimism	0.0327	0.0093	3.5161	0.0004379^{***}
Multiple R-squared:	0.02455			
Transition probabilities:				
	Regime 1	Regime 2		
Regime 1	0.97644897	0.01563271		
Regime 2	0.02355103	0.98436729		

Fig. 3. High/Low Volatility Regimes for Euro-periphery Volatility

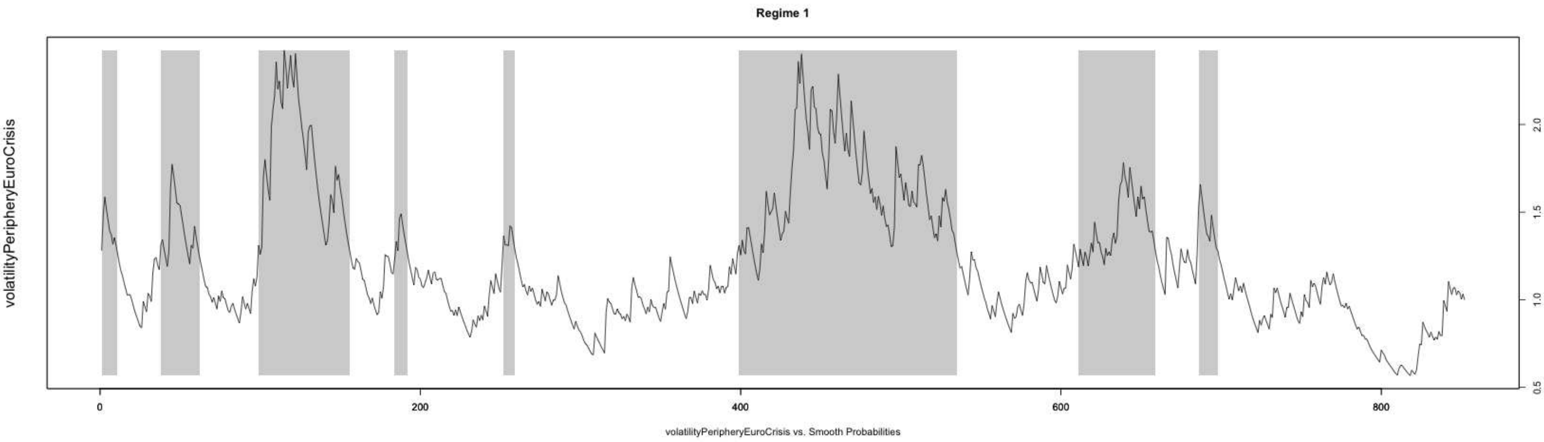

Table 10: Markov switching model. Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Regime 1				
	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)(S)	0.8607	0.0212	40.5991	¡ 2.2e-16 ***
Weighted Average Pessimism	0.0589	0.0084	7.0119	2.351e-12 ***
Multiple R-squared:	0.09475			
Regime 2				
	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)(S)	1.3710	0.0830	16.5181	¡ 2.2e-16 ***
Weighted Average Pessimism	0.0823	0.0282	2.9184	0.003518 **
Multiple R-squared:	0.03117			
Transition probabilities:				
	Regime 1	Regime 2		
Regime 1	0.9840403	0.0252825		
Regime 2	0.0159597	0.9747175		

Fig. 4. High/Low Volatility Regimes for Euro-periphery Volatility

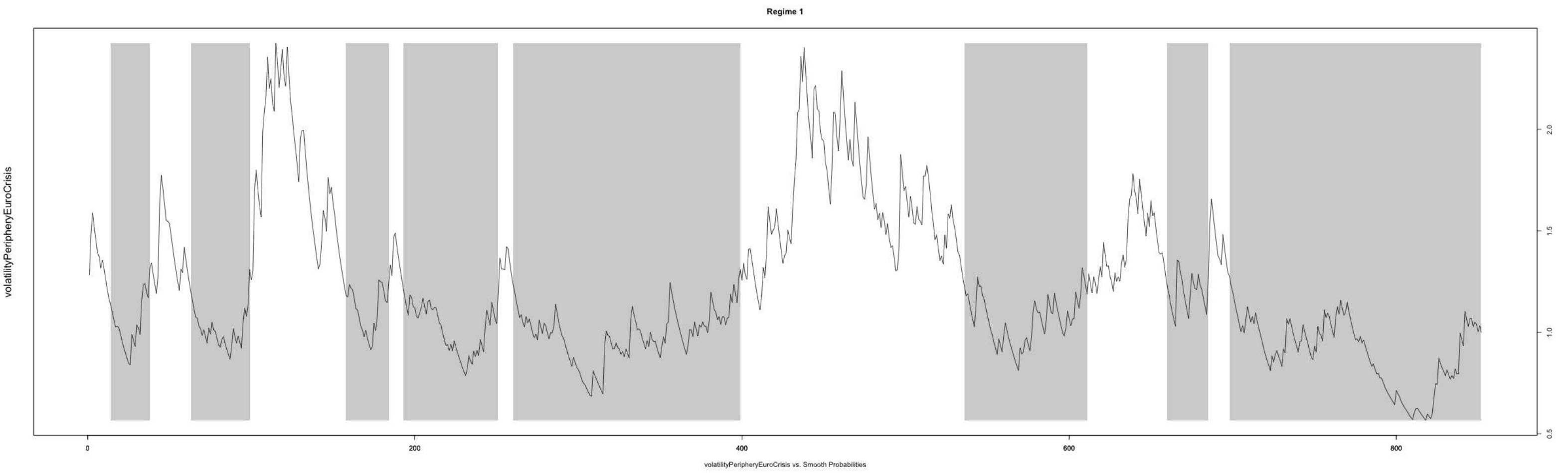

Table 11: Markov switching model. Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Regime 1				
	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)(S)	1.5742	0.0373	42.2038	¡ 2e-16 ***
peripheryNewsCountEuroCrisis(S)	0.0015	0.0009	1.6667	0.09557 .
Multiple R-squared:	0.01039			
Regime 2				
	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)(S)	0.9530	0.0121	78.7603	¡ 2.2e-16 ***
peripheryNewsCountEuroCrisis(S)	0.0025	0.0003	8.3333	¡ 2.2e-16 ***
Multiple R-squared:	0.1063			
Transition probabilities:				
	Regime 1	Regime 2		
Regime 1	0.97479492	0.01476384		
Regime 2	0.02520508	0.98523616		

Fig. 5. High/Low Volatility Regimes for Euro-periphery Volatility

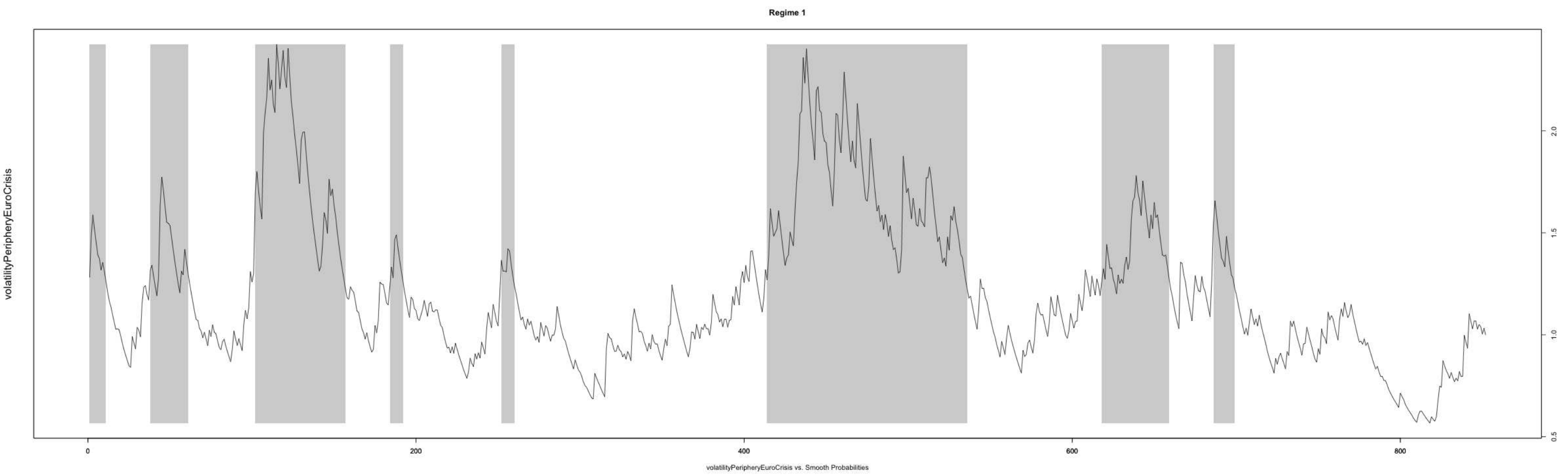

Table 12: Markov switching model. Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Regime 1				
	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)(S)	1.5712	0.0367	42.812	¡ 2e-16 ***
pessimisticCountGtAveraEuCrisis(S)	0.0018	0.0010	1.800	0.07186 .
Multiple R-squared:	0.01289			
Regime 2				
	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)(S)	0.9515	0.0116	82.026	¡ 2.2e-16 ***
pessimisticCountGtAveraEuCrisis(S)	0.0029	0.0004	7.250	4.168e-13 ***
Multiple R-squared:	0.1237			
Transition probabilities:				
	Regime 1	Regime 2		
Regime 1	0.97472112	0.01474346		
Regime 2	0.02527888	0.98525654		

Fig. 6. High/Low Volatility Regimes for Euro-periphery Volatility

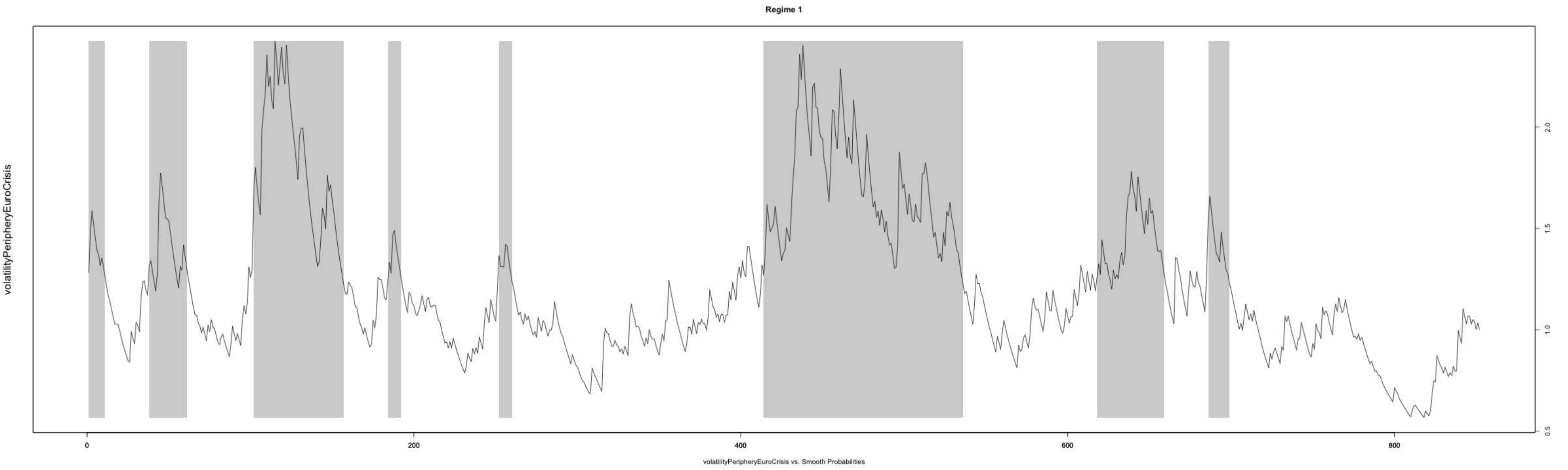

Table 13: Markov switching model. Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Regime 1				
	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)(S)	1.5681	0.0371	42.2668	¡ 2e-16 ***
pessimisticCountGtWeighEuCrisis(S)	0.0021	0.0011	1.9091	0.05625 .
Multiple R-squared:	0.01549			
Regime 2				
	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)(S) 0.9516	0.0115	82.748	¡ 2.2e-16 ***	
pessimisticCountGtWeighEuCrisis(S)	0.0032	0.0004	8.000	1.332e-15 ***
Multiple R-squared:	0.1314			
Transition probabilities:				
	Regime 1	Regime 2		
Regime 1	0.97468955	0.01474063		
Regime 2	0.02531045	0.98525937		

Fig. 7. High/Low Volatility Regimes for Euro-periphery Volatility

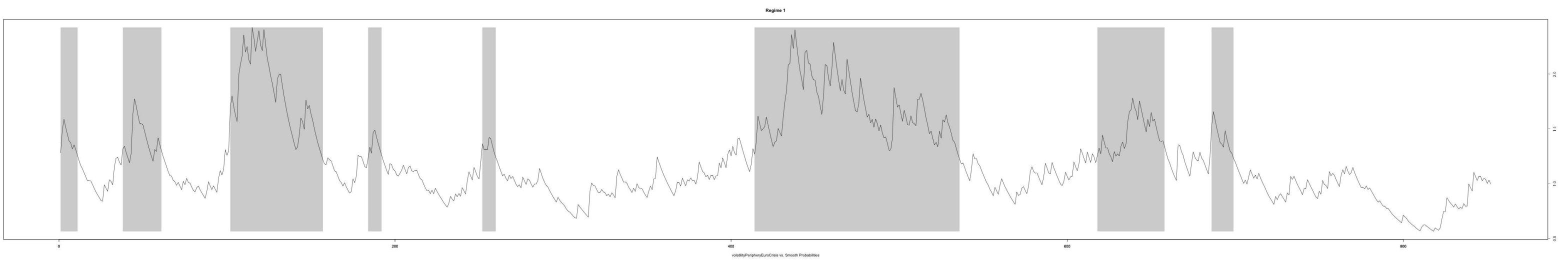

Table 14: Markov switching model. Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Regime 1				
	Estimate	Std. Error	t value Pr(¿—t—)	
(Intercept)(S)	0.8655	0.0187	46.2834	¡ 2.2e-16 ***
sviEuroPeripheryEuroCrisis(S)	0.0160	0.0019	8.4211	¡ 2.2e-16 ***
Multiple R-squared:	0.1595			
Regime 2				
	Estimate	Std. Error	t value	Pr(¿—t—)
(Intercept)(S)	1.4018	0.0413	33.9419	¡ 2.2e-16 ***
sviEuroPeripheryEuroCrisis(S)	0.0181	0.0028	6.4643	1.018e-10 ***
Multiple R-squared:	0.1558			
Transition probabilities:				
	Regime 1	Regime 2		
Regime 1	0.98693811	0.02189085		
Regime 2	0.01306189	0.97810915		

Fig. 8. High/Low Volatility Regimes for Euro-periphery Volatility