



Munich Personal RePEc Archive

**High Frequency Newswire Textual
Sentiment: Evidence from international
stock markets during the European
Financial Crisis**

Chouliaras, Andreas

2 March 2015

Online at <https://mpra.ub.uni-muenchen.de/66948/>
MPRA Paper No. 66948, posted 29 Sep 2015 06:34 UTC

High Frequency Newswire Textual Sentiment Analysis: Evidence from international stock markets during the European Financial Crisis

Andreas Chouliaras*
Luxembourg School of Finance**

Abstract

I perform textual analysis in a total of 13145 high frequency (intraday) news articles, 6536 of which are from the Dow Jones Newswires and 6609 from the Thomson Reuters Newswires. Selected articles are Euro-periphery (Portugal, Ireland, Italy, Greece, Spain) crisis-related articles which contain a number of keywords in their content and their title. News pessimism as a product of textual analysis sentiment affects stock returns negatively and volatility positively (an increase in pessimism is associated with lower stock prices and higher volatility). Media pessimism does not only affect the crisis-hit Euro-periphery countries but also European (Germany, France, UK, Switzerland) and overseas (US, Japan, China) stock markets. Stock markets can be very fast when "absorbing" the shocks of media pessimism. Even small time frames such as 5-minutes and 30-minutes can be enough for stock prices to be negatively affected by a higher media pessimism. The media (and especially newswires which release articles with extreme speeds and coverage) provide a channel through which "bad" news are instantaneously circulated and provide worldwide "shocks" to stock prices in extremely small time windows (even 5-minutes).

JEL classification: G01, G14, G15, D83.

Keywords: Financial Crisis, Textual Analysis, News Flow, Financial Sentiment, High Frequency, Intraday, Dow Jones, Thomson Reuters.

* Corresponding author. E-mail address: andreas.chouliaras@uni.lu

** Luxembourg School of Finance, 4, rue Albert Borschette, 1246 Luxembourg.

1. Introduction

The Euro-crisis has been at the centre of attention for the financial press since the Euro-crisis started, around the end of 2009. Thousands of news articles were written about the Euro-periphery countries (Portugal, Ireland, Italy, Greece, Spain), whose vulnerabilities led to bailout packages, either at the sovereign or the banking level. There has been fierce debate 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 (Eurozone), and whether bad news about them might propagate and affect the rest of the European countries, causing a domino effect.

Previous research about Euro-crisis news and events (e.g. Arezki, Candelon, and Sy (2011), Beetsma, Giuliadori, de Jong, and Widijanto (2013), Mink and De Haan (2013)) 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 as "good" or "bad". 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 information and in many cases they are anticipated by market participants. Classifying events manually is problematic because it depends entirely on the perceptions and beliefs of the researcher(s) who do the classification, 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. I attempt to shed light on the impact of the News Flow imbedded in newswires articles. Chouliaras and Grammatikos (2013) study the effect of news sentiment on the probability of extreme stock market returns on a daily basis. But a more high-frequency analysis is also important, especially nowadays that market participants make decisions in a much faster pace. Is a whole trading day necessary for markets to "absorb" the negative media content (pessimism), or are smaller time frames enough? According to the Kyle (1985) model, the speed with which information is incorporated into market prices is essential. The contributions of this article are the following: to the best of my knowledge, it is the first paper to study the high-frequency effects of the Euro-crisis using data with are timestamped up to milliseconds; I incorporate the two biggest international newswires (Thomson Reuters and Dow Jones) and I use a rather elaborate method to select articles relevant for the topic of the Euro-crisis under investigation; I examine the impact of News Flow about the peripheral countries on international intraday stock returns. This way I answer the question of whether after all the Euro-periphery financial sentiment does indeed significantly affect stock returns and volatility, a topic which lies at the centre of financial discussions and policy making since the advent of the Euro-crisis. Such an analysis also contributes to policy decision making, in the sense that huge amounts of money had to be disbursed by

European Union countries (in the form of bailout packages) in the countries that found themselves in serious financial troubles. In case that news affect not only the countries in trouble, the contribution in the policy decision making has to do with the fact that apart from an action of solidarity, providing financial assistance was also an action that had to do with mitigating the risk of crisis transmission and panic in the financial markets.

I find that the News Flow about the financial crisis significantly affect not only the Euro-periphery but also international stock markets. The effect is not only limited in European Union countries, but also overseas financial markets are affected (USA, Japan, China). A higher negative (and pessimistic) content regarding the Euro-crisis is associated with lower stock returns even when considering a very small time frame (5 minutes and 30 minutes after the articles were released). My findings provide support for theoretical models such as De Long, Shleifer, Summers, and Waldmann (1990) which predict that low sentiment will generate downward price pressure. If one considers the media pessimism variables as a proxy of sentiment or risk aversion, one can interpret these findings in the context of a model such as Campbell, Grossman and Wang (1993) which state that changes in the level of risk aversion for a sufficient amount of investors can influence returns in the very short-term. Finally, in the spirit of Daniel, Hirshleifer, and Subrahmanyam (1998), beliefs lead investors to overweight or underweight public signals which are interpreted to confirm or contradict their private information, causing fluctuations in asset returns. As in Shleifer and Summers (1990), there can be incidents where investors are not fully rational and their demand for assets is affected by beliefs or sentiments. The central economic questions I am addressing are threefold: do articles about the crisis hit countries affect the perceived to be financially sound ones? Are the effects contained within Europe? Or do they affect countries geographically distant as well (US, China, Japan)? How fast are these effects? Are multiple hours or days needed, or are small time frames such as 5 minutes and 30 minutes enough to start feeling the effects of these pessimistic articles? The theoretical models about noise traders, beliefs and sentiment provide an interesting framework to examine these questions.

The study of News Flow has become rather popular recently with the advances in the Data Mining and Sentiment Analysis techniques. The strong interest in this area has been demonstrated by the recent creation of companies and commercial products specialised in the production of financial sentiment (see e.g., RavenPack¹ and Thomson Reuters News Analytics²). 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 articles with the Dow Jones returns, using vector auto regressions (VARs). He finds that media pessimism has

¹<http://www.ravenpack.com/>

²<http://www.machinereadablenews.com/>

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 10-Ks filing returns, trading volume, volatility and other characteristics. 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., while Chouliaras (2015a) finds that 10-K pessimism negatively affects stock holdings after the filing. 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 Garcia (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. Finally, Ahern and Sosyura (2014) show evidence of firms manipulating media coverage to achieve better stock prices during mergers and acquisitions negotiations.

Our study is related to Tetlock (2007), Garcia (2013), Garcia (2012) and Chouliaras and Grammatikos (2013) as far as the analysis of News Flow is concerned. The main contributions to the previous literature are: first and foremost, I employ a database of 13145 articles from the two biggest international newswires (Dow Jones and Thomson Reuters), while none of the previous studies deals with the high frequency effects of news during the Euro-crisis. With the exception of Chouliaras and Grammatikos (2013), the other studies typically examine only a column of a newspaper per day, while in this research paper the articles are released at any point during the 24 hours of a day. The fact that every article and every stock price is timestamped (at least on the level of minutes), I am able to "align" the time zone of every article and every stock price, making sure there is a causal link since, in this setting, articles precede stock prices. Furthermore, the selected articles have to do with multiple countries, while I also study their effects on multiple countries, while the previous literature typically

studied the effects on one stock index. In addition, this research contributes to the policy debate, in the sense that it deals with the question of whether news about a group of vulnerable countries can affect financially "stronger" ones, and how fast this effect becomes visible.

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 effect of Euro-periphery news sentiment to international stock markets.

The rest of this paper is organized as follows. Section 2 presents the data. Section 3 presents the model specification and I explain how I study the effect of News Sentiment in a cross-country setting. Section 4 provides the empirical results. Section 5 concludes.

2. The Data

The financial data I employ come from Thomson Reuters Tick History (TRTH)³. I use data on stock indices for various European countries (Portugal, Ireland, Italy, Greece, Spain, Austria, Belgium, Finland, Germany, United Kingdom, Switzerland, Norway) but also overseas stock markets (US S&P, US Dow Jones, China, Japan, Brazil, Canada). The exact stock indexes used in this paper appear in Table 1:

Insert Table 1 here

As far as the news articles data are concerned, I employ the Dow Jones Factiva news database⁴. The time period under examination starts from 1 December 2009, and it ends on 13 March 2013. Europe slipped into recession in 2009⁵, while by December 2009 it was clear something was seriously wrong with the finances of Greece⁶, which makes December 2009 a reasonable point to start our analysis. I end the analysis due to data availability as far as the Thomson Reuters Tick History (TRTH) data are concerned, but by this point the Eurozone countries have intervened and were heavily invested in sustaining the cohesion of the European union and the Eurozone monetary union.

³Thomson Reuters Tick History can be found at: <http://thomsonreuters.com/tick-history/>

⁴Dow Jones Factiva can be found at <http://new.dowjones.com/products/factiva/>

⁵http://ec.europa.eu/economy_finance/explained/the_financial_and_economic_crisis/why_did_the_crisis_happen/index_en.htm

⁶<http://news.bbc.co.uk/2/hi/business/8406665.stm>, <http://news.bbc.co.uk/2/hi/business/8407605.stm>

3. The Methodology

3.1. Preprocessing the financial data

As far as the processing of the financial data is concerned, I create MySQL databases in order to be able to handle the huge amount of data. Furthermore, I process the data so as to keep one entry per minute. Specifically, I keep the first entry for every minute. Then, since I am working in 5-minute (and 30-minute) intervals, I keep one entry per five (and thirty) minutes. Since our news data are time-stamped in a minute basis, I first of all make sure both news data and financial data are in the same time zone. I bring all data in the Greenwich Mean Time (GMT) timezone. The news data were downloaded in a time zone of GMT+1, thus I subtract one hour from every news item. Some financial data are in GMT+2 (or for some countries from GMT-9 to GMT+9), and I subtract 2 hours from this news item, so that every time series is finally in a GMT timezone. Moreover, I make sure I have an aggregate news metric for every financial data point. Thus, if in the interval between two stock prices more than one news items were released, I calculate the average percentages of these news items, so that for every financial price there is one value for the respective articles that fall within this time interval. The table with the performed "time alignments" follows:

Insert Table 2 here

After this time alignment is performed, every time series is brought to a common time zone (GMT). Thus, I am able to proceed studying the 5-minute (and 30-minute) effect of media content on high frequency stock returns. After I perform this "time zone alignment" between news and stock markets (and bring every time series in a GMT time zone), I am able to start combining the news with the financial data. For a trader/investor, to read a news item, process it, place an order, and for this order to be executed and affect stock prices, a reasonable time frame is required. Thus, I keep prices every 5 minutes (and 30 minutes). If between the 5-minutes (and 30-minutes) time intervals more than one stories are released, I calculate the average positive, negative and pessimism factors for each one of these items. If news items were released during non-operating hours, I "attach" these news to the next price available (i.e. the first price of the next trading day). In the era of algorithmic trading, some traders are even faster than that. For this reason, I repeat the analysis using 5-minute time intervals for stock prices, trying to examine even higher frequency dynamics.

3.2. News Flow

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 (extreme) market returns is the focus of this paper. This approach should provide interesting contributions to the relatively new strand in the finance literature that examines directly how information is assimilated by and affects market prices.

In this context I analyse the content of news stories in the newswires. As far as the news stories are concerned, I extract and analyse news articles covering the time period from 1 December 2009 till 28 November 2012, from two sources: *the Dow Jones Newswires* and *the Thomson Reuters Newswire*. I use these two newswires because they are undoubtedly among the most popular providers of real-time news worldwide. Dow Jones Newswires and Thomson Reuters provide stories in real time. For each country the relevant stories are obtained by a query searching for articles that include the name of the Euro-periphery country plus one of the following terms each time: crisis, debt, economy, deficit, default. For example, for Greece the articles 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 is straightforward: An article that contains the term Greek crisis is related (with a high probability) 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 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.

3.3. Preprocessing the news data

As a first step I 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 articles dominates the sample, simply because it is being delivered multiple times, with none (or insignificant) changes. 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 I want to study the unique impact of information at the day it was first released, I keep the article published first and discard all duplicates after the distribution of the first story.⁷

3.4. 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, I select only items that contain one or more of a set of keywords in their title. The title is signalling the content of the article. Thus, I expect articles that contain one or more of these keywords in the title, to be of higher relevance to our topic, than selecting all articles that contain one or more keywords in the text, but may refer to many other topics as well. 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 3, having a total of 13145 news articles passing the filters:

Insert Table 3 here

I notice that most of the stories concern Greece: Out of 13145, a total of 8128 concern the greek crisis. This is well understood since Greece is the country that was most badly hurt, and naturally the financial press took a great interest in covering events concerning the greek financial troubles as they evolved through time. After Greece, the most articles were about Spain (1822), followed by Ireland (1319), Italy (1268) and finally, Portugal (608). Surprisingly, the overall number of articles released from Reuters and Dow Jones are very close to each other (6609 and 6536 respectively). This is an indicator that more or less these two newswires covered in a similar fashion (in terms of the number

⁷I have repeated the analysis without removing the duplicates, constructing a new variable that picks up the repeats, and added this to the regressions. The impact of repeat articles does not appear to be significant. More on this on Section 4.3.

of stories) the Euro-crisis, and are also an indicator that our keyword filters are not biased in terms of selecting stories from the one versus the other newswire.

The graphical plot of the number of articles during each time period is shown in Figure 1:

Insert Figure 1 here

Naturally, the number of stories is not constant through times, as the Euro-crisis "escalates" and "calms down" throughout various points in time. One can see "spikes" in terms of the number of stories. At least in four or five points a much higher number of articles are released: the biggest is on May 2010 (when the Greek bailout was announced) with a maximum of 140 articles released in one single day. This number is very high given the fact that I already performed two series of filtering for the articles (both for the content and the titles of the stories – only the title filtering removes more than half of the stories, while the duplicate filter also removes around half of the remaining articles, thus in this day actually more than 500 crisis-related stories were released, a factor indicating the extremely high interest the Euro-crisis drew in the financial press). Another spike occurs around July 2011, an important period given the Euro-summit which took place in 22 July 2011 to guarantee a new bailout plan for Greece⁸. Another spike occurs around March 2012, when the new bailout of Greece was finally granted⁹. Finally, another spike occurs at November 2012, when Eurozone ministers agreed to cut the Greek debt by another 40 billion euros and to release 44 billions in bailout money and aid¹⁰.

3.5. Textual Analysis

As a next step, using textual analysis, based on the Loughran and McDonald (2011) dictionary¹¹, I measure the positive media content as in Garcia (2012) and Garcia (2013): $G_i = \sum_i \frac{g_i}{w_i}$, calculated as the percentage of positive words over the total number of words of every article. The symbol g_i stands for the number of positive words in the article, and w_i stands for the total number of words in the article. I do the same for the negative words, obtaining the negative media content as $B_t = \sum_i \frac{b_t}{w_t}$, with b_i denoting the negative words in the article. Thus, I obtain the *Pessimism* of article i :

$$Pessimism_i = B_i - G_i \tag{1}$$

⁸For example, an article covering the event from Guardian: <http://www.theguardian.com/business/2011/jul/21/european-debt-crisis-summit-euro>

⁹An article by BBC covering the event: <http://www.bbc.com/news/business-17338100>

¹⁰Another article by BBC covering the event: <http://www.bbc.com/news/business-13798000>

¹¹The dictionary can be found at http://www3.nd.edu/~mcdonald/Word_Lists.html

The *Pessimism* is calculated for every article. If more than one stories are released between two financial prices (30 minute and 5 minute intervals), I also calculate the Average Pessimism from all the articles that fall in this interval. Another metric I use is the *News Count* (N_t): This is the total number of articles released between two financial prices (5 minute and 30 minute intervals) regarding any of the queries I am interested in.

The summary statistics for the media content (positive, negative, pessimism) and the stock markets appear in Table 4:

Insert Table 4 here

I see that stock markets have mean returns of 0 (normal since I am talking about 5-minute and 30-minute returns), while the media have positive means (3.2% for the negative, 0.7% for the positive and 2.4% for the pessimism media content). In order to see how the algorithm works, some selected articles appear in Tables 5, 6, 7, 8 and 9:

Insert Tables 5, 6, 7, 8 and 9 here

For each of the five (5) Euro-periphery countries (Portugal, Ireland, Italy, Greece and Spain), two messages are shown: one from the Dow Jones Newswire and one from the Thomson Reuters Newswire. As one can see, each of these articles contain one or more of the keywords in its title (Portuguese, Portugal, Irish, Italy, Greek, Spain, Spanish), as well as one or more of the crisis keywords in its content (Portuguese Default, Portugal Crisis, Irish Debt, Irish Crisis, Italian Economy, Greek Default, Spanish Economy). The underlined words are the keywords for the titles and the content. The words painted in red are words that appear in the "negative" word list, while green words belong in the "positive" word list. The articles I selected are among the most pessimistic articles for every country. Thus, it is natural that most of the painted words are red. An example of an article that has more positive than negative words, appears in Table 10:

Insert Table 10 here

Especially during a financial crisis, articles are mostly negative. On top of that, journalists may prefer to depict news using a dramatic language, which could draw readers' interest easier than "good

news” would.

4. Empirical Findings

4.1. Effects on stock market returns

The model I employ to study the effect of the content of news in high frequency stock returns is the following:

$$R_t = a_1 + b_1 M_t + \sum_{i=1}^5 c_i R_{t-i} \quad (2)$$

where M_t takes the value of:

- the positive (G_t)
- the negative (B_t)
- the pessimism (P_t) and
- the news count (N_t)

of the previous 5-minute and 30-minute interval respectively. I am primarily interested in the sign, the coefficient and the statistical significance of the b_1 coefficient. I control for five lags of returns (i.e. five 5-minute and 30-minute lagged returns for every stock market) to deal with autocorrelation in the returns. The regressions I perform are robust, using the Huber-White sandwich estimators (Huber (1967), White (1980)) to deal with autocorrelation, heteroskedasticity, heterogeneity and lack of normality.

The results of the robust regressions appear in Table 11:

Insert Table 11 here

Table 11 consists of two panels. The above Panel, A, contains the results for 30-minute intervals between stock returns, while Panel B contains the results for 5-minute intervals. This way I study the very short term (5-minutes) and the short term (30-minute) effect of high frequency newswire articles on stock returns. The first three columns of the Table concern the effect on three of the five Euro-periphery countries (Portugal, Greece, Italy), which are the most badly hit European countries during the Euro-crisis. I do not report the full results due to space constraints, but the full Tables are available in the Online Appendix. In the Online Appendix one can find full results for multiple countries that are not presented here: Ireland, Spain, Austria, Finland, Belgium, Norway, Brazil, Canada. The results for these countries are qualitatively the same as the ones on the sample of countries presented

on this paper. Columns four (4) to six (6) contain the results for four (4) of the European Union countries (Germany, France, Switzerland and the UK), while columns eight (8) to ten (10) contain the results for the overseas financial markets (US, Japan and China). The "pessimism" variable (1st row of the Table) is significant for all ten countries (at the 10% significance level for Germany, at the 5% level for France, UK, Switzerland, Italy and China), and at the 1% level for Portugal, Greece, US and Japan, and has a negative sign. The t-statistics are also quite high, ranging from -3.17 for Japan, to -1.70 for Germany. This means that a higher news pessimism is associated with lower stock prices 30 minutes later. In terms of magnitude, the coefficient is higher for Greece (-0.0194 with a t-stat of -2.99), followed by the United States (-0.0181 with a t-stat of -2.75), which means that a 1% increase in news pessimism, leads to stock prices which are lower by 0.0194% in the case of Greece, and 0.0181 in the case of the US (S&P 500 stock index). It is quite interesting that all three overseas markets are influenced by the Euro-periphery news (for China the coefficient is significant at the 5% level, while for Japan and the US S&P the coefficient is significant at the 1% level). The effect is not contained within the crisis-hit countries, not even within Europe, but it quickly expands to American and Asian stock markets, affecting them very fast.

As far as the "Positive" factor is concerned (2nd row of Panel A), the coefficients are significant for five (5) out of ten (10) countries: Portugal, Greece, Germany, France and Japan. In all these cases, the coefficient is positive, which means that a higher positive percentage in the news article leads to higher stock returns. For the "Negative" factor (line 3 of Panel A), the effect is significant in eight out of ten countries. I notice that the effects are stronger for the "Pessimism" factor (which is constructed as the difference between the Negative and the Positive factors). This indicates that indeed investors incorporate both the negative and the positive information that is embedded in articles, and process it according to whether there is more negative than positive information in every article.

The "Average Pessimism" (4th row of Panel A) is significant for six (6) out of ten (10) countries, with a negative sign. Once more, the "Pessimism" is less significant than the "Pessimism" factor, which provides evidence that investors are more affected by the pessimistic content of the last article, than by the average of the articles that fall within the last time interval (30-minutes in the case of Panel A). Finally, the "News Count" factor does not appear to be significant in this case, which means that high frequency stock returns are affected by the actual pessimism in the content of news articles, and not their number.

"Panel B" provides the results in an even higher frequency (5-minutes intervals). Studying both the 30-minute and the 5-minute intervals increases the strength of our results, since it both serves as a robustness check, but also it gives us a better indication of the speed of adjustment of prices to

high frequency news. The "Pessimism" factor is significant for nine out of ten countries (only for the UK stock market is the coefficient not significant). It is quite interesting that markets react so fast to information: only five minutes are enough for stock markets to start experiencing losses, when pessimistic information is released through newswires. A comparison between the coefficients of 5-minutes versus the coefficients of 30-minutes reveals that the 30-minutes coefficients are significantly stronger: for Portugal the coefficient starts from -0.00374 (and a t-stat of -2.84) to -0.00812 (and a t-stat of -2.99). This is 117% higher, for a time difference of 25 minutes. The difference of -0.00438 (-0.00812-(-0.00374)) becomes quite big if one takes into account the fact that these stock markets have aggregate capitalisations of hundreds of billions (trillions actually) Euros. Since both coefficients are negative, the findings indicate that prices fall within 5 minutes of the news release, and they fall even further 30 minutes after the news release. Accordingly, for Greece, the coefficient becomes higher (in absolute terms), from -0.00689 to -0.0194. There are two countries (Germany and Switzerland), for which the coefficients are more significant for the five minute intervals (significant at the 5% for Germany and 1% for Switzerland) than the thirty minute intervals (10% for Germany and 5% for Switzerland). On the other hand, the coefficients are more significant for four (4) countries in the 30 minute intervals (Italy, France at 5% and UK at 5%, US at 1%) versus the 5 minute intervals (Italy, France at 10%, insignificant for UK, 10% for US). The results show significant effects (in most of the cases) only 5 minutes after stories are released, which become much more severe 25 minutes later. One more, for the five minute intervals, the "Pessimism" factor is more significant than the individual "Positive" and "Negative" factors, once more more significant (in most cases) than the "Average Pessimism" factor (for example, a coefficient of -0.00689 with a t-stat of -3.11 for Greece for the Pessimism factor, versus a coefficient of -0.00522 and a t-stat of -2.58 for the Average Pessimism). This is not true for Japan, for which the coefficient of -0.0683 (t-stat of -2.48) is higher for the Average Pessimism than the Pessimism factor (coefficient of -0.0580 with a t-stat of -3.18). The t-stat is nevertheless higher for the Pessimism factor.

4.2. *Effects on stock market volatility*

Apart from stock returns, it is interesting to examine whether the content of high frequency news affect the volatility of stock markets. For this reason, I employ the following model:

$$Volatility_t = a_1 + b_1 M_t + \sum_{i=1}^5 c_i R_{t-i} \quad (3)$$

where once more M_t takes the value of the previously defined factors (Positive, Negative, Pessimism, Average Pessimism and News Count). I define volatility as the standard deviation of intraday stock returns.

The results appear in Table 12:

Insert Table 12 here

Once more the results show a high number of significant coefficients. Once more, Panel A concerns 30-minute intervals, while Panel B concerns 5-minute intervals.

For the "Pessimism" factor, all ten coefficients are significant (in seven out of ten countries they are significant in the 1% level), and all coefficients are positive. This means that an increase in "Pessimism" increases the volatility of stock markets. An increase of 1% in the "Pessimism" of newswire news articles, increases volatility by 0.003% for Portugal (t-statistic of 2.81), 0.00409 for Greece (t-statistic of 1.69), 0.00755 for Germany (t-statistic of 5.30), and 0.00947 (t-statistic of 2.22), 0.0156 (t-statistic of 2.22) and 0.00960 (t-statistic of 2.34) for the US S&P, Japan's Nikkei 225 and China's Hong Kong Hang Seng stock indices. Quite interestingly the result is stronger (in terms of statistical significance) for Germany, France, UK, US, Japan and China than for Greece (although this was not true for stock returns). An increase in the percentage of positive words decreases volatility for all ten stock markets, while an increase in the percentage of negative words and an increase in the "Average Pessimism" increases the volatility in nine out of ten stock markets. Once more, the "Pessimism" effect is stronger than the individual "Positive" and "Negative" factors. Finally, the "News Count" is significant in seven out of ten stock markets. "News Count" was found to be non significant as far as high frequency stock returns are concerned, but it appears to be significantly affecting the volatility of stock markets. In other words, although stock prices do not appear to fall in a high frequency manner because more stories are released, but they appear to become more volatile when more crisis-related stories appear on newswires.

One can see a similar story on Panel B, once more noticing that effects are stronger in the 30-minute intervals than the 5-minute intervals. It seems as it takes a bit of time (from 5-minutes to 30-minutes) for stock prices and stock volatility to process and assimilate the information embedded in news articles, with more pessimistic news leading to lower stock prices and higher volatility even 5-minutes after the news items are released, with the effects becoming much stronger 25 minutes later. For example, the coefficient for the volatility of the Portuguese stock market (PSI 20) is 0.00126 for the 5-minute interval (t-statistic equal to 3.15), while it increases to 0.00343 (t-statistic equal to 2.81)

25 minutes later. This is an increase 172% in only 25 minutes. In all ten countries in our sample, the coefficient of "Pessimism" is higher in the 30-minutes intervals than the 5-minute intervals. One can claim that, as is the case with stock returns, the effect which starts 5 minutes after the news article is released, further builds on and becomes more significant 25 minutes later. The "News Count" factor is significant in five out of ten countries for the 5-minute intervals, while it is significant in seven out of ten stock markets for the 30-minute intervals.

4.3. *The impact of repeat articles*

In Section 3.3, I mentioned that I exclude duplicate (i.e. highly similar news articles). This is an option offered by Dow Jones Factiva, in case one wants to have articles appearing once in the overall sample. But, it could be that traders are triggered by repeated messages. Therefore, I provide a further check to study whether the number of highly similar messages indeed affects stock prices and volatility. For this purpose, I construct a new variable that picks up the repeats, and add this to the regression. The model thus becomes

$$R_t = a_1 + b_1 Repeat_t + \sum_{i=1}^5 c_i R_{t-i} \quad (4)$$

The variable $Repeat_t$ capturing the number of similar messages that were released in the previous time interval (5-minutes or 30-minutes respectively, as in the previous analysis).

To calculate the similarity of articles, I compare the content of every story with articles released in the previous 30-minutes (or 5-minutes), using the Ratcliff/Obershelp (Ratcliff and Metzener (1988)) pattern recognition algorithm which calculates the similarity of two strings. I define repeat articles using five different cutoff similarity percentages: articles that have a similarity ratio that is higher than 50%, 60%, 70%, 80% and 90% respectively. A visual inspection shows that stories that are 50% or more similar to each other, are already quite similar.

The results appear in Table 13:

Insert Table 13 here

"Panel A" contains the results for the effect on stock returns, while "Panel B" contains the results for the effect on volatility. The effect on stock returns is virtually non-existent. The only exception is Greece, for stories that have a similarity between 50% and 70%, for which statistically significant coefficients appear (at the 5% and 10% levels). These coefficients indicate that a higher number of similar crisis stories negatively affects stock markets only in the case of Greece. The coefficient is

rather small (-0.000140 with a t-statistic of -2.19). In Portugal, Italy, Germany and the US, the effect is insignificant. This indicates that discarding the repeated stories as explained in Section 3.3, does not significantly affect our results.

In "Panel B", I examine the effect of repeated stories on volatility using the following specification:

$$Volatility_t = a_1 + b_1 Repeat_t + \sum_{i=1}^5 c_i R_{t-i} \quad (5)$$

The results show statistically significant coefficients for four out of five countries, but in the opposite direction: Repeated stories tend to *decrease* volatility, not increase it, as was the case with pessimism in our previous findings. The only exception is, quite interestingly, Greece, for which an increase in the number of repeated stories does not affect volatility (which the exception of articles that are more than 90% similar, in which case there is an increase in volatility). For the other countries, coefficients are negative and significant (in most of the cases). One can interpret this finding as repeat news resolving volatility (i.e. uncertainty in the markets). In any case, as these results show, discarding repeat news (and keeping only the first message from every series) does not affect our results.

5. Conclusion

Textual analysis is performed in a total of 13145 news articles in total, 6536 of which come from the Dow Jones Newswires and 6609 from Thomson Reuters. Selected articles are crisis-related stories which contain a number of keywords in their content and their title. News pessimism affects stock returns negatively and volatility positively (an increase in news pessimism is associated with lower stock prices and higher volatility). Media pessimism does not only affect the crisis-hit Euro-periphery countries (Portugal, Italy, Greece) but also European (Germany, France, UK, Switzerland) and overseas (US, China, Japan) stock markets. Stock markets can be very fast when "absorbing" the shocks of media pessimism. Even small time frames such as 5-minutes and 30-minutes can be enough for stock prices and volatility to be affected by a higher media pessimism. Euro-crisis related stories also affect markets which are very far away geographically, such as the United States, Japan and China. Indeed the significance of the Euro-crisis spans far across the "narrow" borders of the Euro-periphery countries, and even the borders of Europe, since markets all over the globe are affected on "real time" by high frequency stories released through newswires. Repeated stories do not seem to influence stock markets, while they seem to decrease volatility.

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.
- Beetsma, R., Giuliadori, 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.
- 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., 2013. News flow, web attention and extreme returns in the european financial crisis .
- Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. *the Journal of Finance* 53, 1839–1885.
- De Long, J. B., Shleifer, A., Summers, L. H., Waldmann, R. J., 1990. Noise trader risk in financial markets. *Journal of political Economy* pp. 703–738.
- Garcia, D., 2012. The kinks of financial journalism .
- Garcia, D., 2013. Sentiment during recessions. *The Journal of Finance* 68, 1267–1300.
- Huber, P. J., 1967. The behavior of maximum likelihood estimates under nonstandard conditions. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1, pp. 221–233.

- Jegadeesh, N., Wu, D., 2013. Word power: A new approach for content analysis. *Journal of Financial Economics* 110, 712–729.
- Kyle, A. S., 1985. Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society* pp. 1315–1335.
- 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* .
- Mink, M., De Haan, J., 2013. Contagion during the greek sovereign debt crisis. *Journal of International Money and Finance* 34, 102–113.
- Ratcliff, J. W., Metzener, D. E., 1988. Pattern-matching-the gestalt approach. *Dr Dobbs Journal* 13, 46.
- Shleifer, A., Summers, L. H., 1990. The noise trader approach to finance. *The Journal of Economic Perspectives* pp. 19–33.
- 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.
- White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society* pp. 817–838.

Table 1: Stock index used for every country

Country	Stock Index used
Portugal	PSI 20
Ireland	ISEQ
Italy	FTSE MIB
Greece	ASE
Spain	IBEX 35
Germany	XETRA DAX
France	CAC 40
Austria	ATX
Finland	OMX
Belgium	BEL 20
UK	FTSE 100
Switzerland	SMI
Norway	OBX
Brazil	BOVESPA
Canada	S&P TSX 60
USA	Dow Jones
USA	S&P 500
Japan	Nikkei 225
China	Hong Kong Hang Seng

Table 2: Timezones for intraday news articles and financial data. Stock markets and news articles are in different time zones. To be able to examine the effect of stories on an intraday setting, all time series must be brought on a common time zone, making the necessary adjustments (adding or subtracting hours respectively).

Source	Original Timezone	Add/Subtract	Timezone after processing
Stories from Dow Jones Factiva	GMT+1	Subtract one hour (-1)	GMT
Portugal	GMT+1	Subtract one hour (-1)	GMT
Ireland	GMT	No action needed	GMT
Italy	GMT	No action needed	GMT
Greece	GMT+2	Subtract two hours (-2)	GMT
Spain	GMT+1	Subtract one hour (-1)	GMT
Germany	GMT+1	Subtract one hour (-1)	GMT
France	GMT+1	Subtract one hour (-1)	GMT
Austria	GMT+1	Subtract one hour (-1)	GMT
Finland	GMT+2	Subtract two hours (-2)	GMT
Belgium	GMT+1	Subtract one hour (-1)	GMT
UK	GMT	No action needed	GMT
Switzerland	GMT+1	Subtract one hour (-1)	GMT
Norway	GMT+1	Subtract one hour (-1)	GMT
Brazil	GMT-2	Add two hours (+2)	GMT
Canada	GMT-5	Add five hours (+5)	GMT
USA (S&P)	GMT-6	Add six hours (+6)	GMT
USA (Dow Jones)	GMT-5	Add five hours (+5)	GMT
Japan	GMT+9	Subtract nine hours (-9)	GMT
China	GMT+8	Subtract eight hours (-8)	GMT

Table 3: Number of stories per country/source. Selected stories pass through two filters: first, for each country the story must include the name of the country plus any of the following keywords: crisis, debt, economy, deficit, default. For example, for Greece the first filter selects the stories containing any of the terms: “greek crisis”, “greek debt”, “greek economy”, “greek deficit”, “greek default”; second, for each country, the story must contain in it’s title a country keyword. For example, for Greece the second filter selects the stories 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.

Source	Portugal	Ireland	Italy	Greece	Spain	Total
Dow Jones Newswires	310	658	497	4342	729	6536
Thomson Reuters	298	661	771	3786	1093	6609
Total	608	1319	1268	8128	1822	13145

Fig. 1.

Number of newswires articles

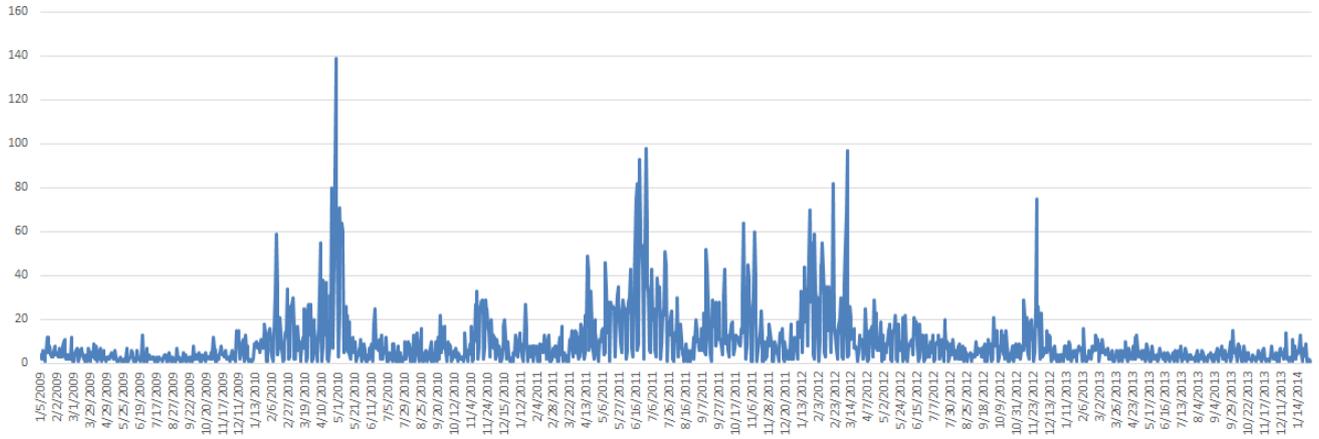


Table 4: Summary statistics for news articles and stock markets. The construction of the media content variables ("Negative", "Positive", "Pessimism") are defined in the Section 3.5

Variable	Mean	Std. Dev.	Min.	Max.
Negative	0.032	0.016	0	0.136
Positive	0.007	0.007	0	0.062
Pessimism	0.024	0.018	-0.046	0.136
Portugal Return	0	0.003	-0.049	0.068
Ireland Return	0	0.003	-0.031	0.04
Italy Return	0	0.004	-0.061	0.046
Greece Return	0	0.006	-0.079	0.072
Spain Return	0	0.004	-0.041	0.098
Germany Return	0	0.003	-0.043	0.039
France Return	0	0.003	-0.039	0.037
Austria Return	0	0.003	-0.056	0.062
Finland Return	0	0.003	-0.05	0.04
Belgium Return	0	0.003	-0.033	0.064
UK Return	0	0.003	-0.034	0.029
Switzerland Return	0	0.002	-0.03	0.04
Norway Return	0	0.003	-0.034	0.048
Brazil Return	0	0.003	-0.039	0.049
Canada Return	0	0.002	-0.031	0.027
US S&P Return	0	0.003	-0.091	0.091
US Dow Jones Return	0	0.003	-0.035	0.037
Japan Return	0	0.004	-0.075	0.052
China Return	0	0.004	-0.072	0.05

Table 5: Examples of selected stories for Portugal. Two stories are provided, one from Dow Jones and one from Reuters. Selected articles must contain a number keywords in their title and their content. Painted words belong in the Negative (red colour) and Positive (green colour) word lists of Loughran and McDonald (2011).

Country	Source	Date	Title	Text
Portugal	Dow Jones	8 April 2013	EURO GOVT- Portuguese default insurance costs rise	EURO GOVT-Portuguese default insurance costs rise- The cost of insuring against a Portuguese sovereign default rose on Monday following a constitutional court decision to reject some austerity measures, raising doubt over its ability to keep to a bailout programme. Five-year credit default swaps (CDS) on Portuguese government debt rose 17 basis points to 430 basis points, according to data monitor Markit. This means it costs \$430,000 annually to buy \$10 million of protection against a Portuguese default using a five-year CDS contract. Portugal’s constitutional court on Friday rejected four out of nine contested austerity measures from this year’s budget. (Reporting by William James; Editing by Marius Zaharia)
Portugal	Reuters	24 March 2011	Portugal Crisis Overshadows EU Summit To Seal Key Reforms	Portugal Crisis Overshadows EU Summit To Seal Key Reforms –Portugal’s political breakdown and disagreement over financing the region’s bailout funds threaten to overshadow a key summit meeting of European Union leaders beginning Thursday. The two-day summit - the culmination of a string of meetings across the continent in recent weeks between leaders and finance officials - has been seen as a key chance for the region to turn its back on the debt crisis by embracing deep reforms. But political turmoil in Portugal has added a new stumbling block. Portugal’s parliament late Wednesday rejected a new government austerity plan, moving Prime Minister Jose Socrates to submit his resignation and setting off a new phase in Europe’s sovereign debt crisis . The failure to pass the measure threatened to push already-high government borrowing costs to unaffordable levels and force Lisbon to seek a bailout . That would make Portugal the third among the 17 nations that use the euro to apply for help from other members of the European Union and the International Monetary Fund. Greece and Ireland went first. Socrates’ resignation forces President Anibal Cavaco Silva to consult with political parties over whether they are willing and able to form a new government. Socrates is expected to remain as caretaker prime minister once the president accepts his resignation . He will head to Brussels Thursday for the summit. Some senior euro-zone government officials now see a high probability that Portugal may have to apply for a bailout under the European Financial Stability Facility. ”The big Euro-zone countries have been pressing Portugal to seek a bailout for a long time. It won’t be long before it happens” one official told Dow Jones Newswires, adding that a bailout would likely be between EUR50 billion-EUR100 billion. The political crisis comes at a tricky time for Portugal. The country faces debt repayments of EUR4.23 billion (\$6 billion) next month and has around EUR4 billion in cash reserves at present. For European leaders, the Portuguese crisis is both an untimely distraction as they try to seal the ”comprehensive package” of reforms the March 24-25 summit was supposed to deliver.

Table 6: Examples of selected stories for Ireland. Two stories are provided, one from Dow Jones and one from Reuters. Selected articles must contain a number keywords in their title and their content. Painted words belong in the Negative (red colour) and Positive (green colour) word lists of Loughran and McDonald (2011).

Country	Source	Date	Title	Text
Ireland	Dow Jones		MARKET TALK: Fears Of <u>Irish</u> Debt Restructuring Rising	MARKET TALK: Fears Of <u>Irish</u> Debt Restructuring Rising 144 words 24 November 2010 09:54 Dow Jones International News DJI English (c) 2010 Dow Jones & Company, Inc. 0854 GMT [Dow Jones] Apart from the lack of detail about the Irish rescue plan and S&P downgrade of both long and short-term sovereign debt, there is rising concern about the threat of restructuring , notes Credit Agricole. Although the Irish Prime Minister is resisting opposition pressure to call an election now, latest polls suggest that there is a good chance of a change in government when the election is held in January. "Given such indicators, latest debt restructuring threats from possible victors will not be ignored by markets," the bank warns , pointing to the ramifications that a debt restructuring by Ireland would have for all other European debtor nations. (nick.hastings@dowjones.com)
Ireland	Reuters	23 November 2010	Merkel says <u>Irish</u> crisis just as worrying as Greece	Merkel says <u>Irish</u> crisis just as worrying as Greece 100 words 23 November 2010 15:00 Reuters News LBA English (c) 2010 Reuters Limited BERLIN, Nov 23 (Reuters) - German Chancellor Angela Merkel said on Tuesday Ireland's crisis was different to Greece's but just as worrying and the euro was in an " exceptionally serious " situation. Merkel added however that the best European was not always the one who helped first. Germany has faced strong criticism for mulling over decisions during the euro zone debt crisis . (Reporting by Gernot Heller, Writing by Sarah Marsh)

Table 7: Examples of selected stories for Italy. Two stories are provided, one from Dow Jones and one from Reuters. Selected articles must contain a number keywords in their title and their content. Painted words belong in the Negative (red colour) and Positive (green colour) word lists of Loughran and McDonald (2011).

Country	Source	Date	Title	Text
Italy	Dow Jones	20 June 2011	Moody's Analyst: <u>Italy</u> Facing Multiple Challenges	Moody's Analyst: Italy Facing Multiple Challenges 163 words 17 June 2011 22:24 Dow Jones International News DJI English (c) 2011 Dow Jones & Company, Inc. NEW YORK (Dow Jones)–Italy's Aa2 rating may be imperiled by a host of factors, including a weak economy, rising interest rates and overall nervousness sparked by the debt turmoil battering Greece, a Moody's Investors Services analyst told Dow Jones Newswires on Friday. Alexander Kockerbeck, an analyst based in Frankfurt, said that the euro zone's third-largest economy faces a raft of challenges both within and outside of Italy's control. These include fiscal reform plans that could be derailed by political instability , and an environment in which investors are losing patience with countries that suffer high debt loads. "The coexistence of some of the weaknesses of the <u>Italian economy</u> , together with the uncertainty in market sentiment due to the euro zone debt crisis ...creates further risks that I need to review," Kockerbeck said. -By Javier E. David, Dow Jones Newswires
Italy	Reuters	5 November 2010	Draghi warns on <u>Italy</u> economy's sluggish growth	Draghi warns on Italy economy's sluggish growth 142 words 5 November 2010 11:16 Reuters News LBA English (c) 2010 Reuters Limited ANCONA, Italy, Nov 5 (Reuters) - The <u>Italian economy</u> 's difficulty in growing and creating income must remain a source of concern , Bank of Italy chief Mario Draghi said, warning that Italy had lost competitive ground with European peers. The impact of the recession on Italian productivity is still unclear and may have helped speed up restructuring in some parts of the system, but could also have had a more negative impact, Draghi said in prepared remarks to a conference on Friday. He also said that including the high costs of servicing public debt in evaluating the country's well-being would clearly worsen the Italian scenario that benefits from a relatively high level of personal savings and wealth. (Reporting by James Mackenzie)

Table 8: Example of selected stories for Greece. Two stories are provided, one from Dow Jones and one from Reuters. Selected articles must contain a number keywords in their title and their content. Painted words belong in the Negative (red colour) and Positive (green colour) word lists of Loughran and McDonald (2011).

Country	Source	Date	Title	Text
Greece	Dow Jones	20 July 2011	Fitch: Disorderly <u>Greek</u> De-fault Would Create Severe Volatility DJ Fitch: Material Threat Of Contagion From Disorderly Greek Default	Fitch: Disorderly <u>Greek</u> Default Would Create Severe Volatility 140 words 20 July 2011 14:55 Dow Jones Global FX & Fixed Income News CM English 2011 Dow Jones & Company, Inc. LONDON (Dow Jones)– The risk of contagion to other distressed and vulnerable euro-zone sovereigns and their banking systems from a disorderly <u>Greek</u> default ”is material,” Fitch Ratings said Wednesday. A disorderly Greek de-fault would create severe market volatility and put pressure on bank and sovereign funding and liquidity, Fitch said. ”Resolution of the Greek crisis is therefore a necessary, though not a sufficient condition for preventing a broader systemic threat to the euro area,” Fitch said in its semi-annual Global Credit Outlook publication. Fitch also said that default by the U.S. is a ”remote” possibility, but if it occurred, a U.S. default would threaten global financial stability . -By Mark Brown, Dow Jones Newswires
Greece	Reuters	20 October 2011	ECB’s Stark warns against <u>Greek</u> default	ECB’s Stark warns against <u>Greek</u> default 161 words 20 October 2011 11:16 Reuters News LBA English (c) 2011 Reuters Limited FRANK-FURT, Oct 20 (Reuters) - Greek bankruptcy or forced private sector involvement would only make Europe’s crisis worse and increase the overall costs of getting Greece back on a sustainable growth path, a European Central Bank policymaker was quoted on Thursday as saying. ”I warn against a default and also against forced private sector involvement,” ECB Executive Board member Juergen Stark told newspaper VDI Nachrichten. ”A haircut as well as an insolvency of Greece would become even more expensive for European taxpayers than the path taken so far.” He also said discussion over private sector involvement in Greece had served to slow the pace of reforms in the country, adding that hardly any progress had been made since the end of last year. In general, one should expect that the problems will last for several years even after the worst has passed, Stark said. (Reporting by Sakari Suoninen)

Table 9: Examples of selected stories. Two stories are provided, one from Dow Jones and one from Reuters. Selected articles must contain a number keywords in their title and their content. Painted words belong in the Negative (red colour) and Positive (green colour) word lists of Loughran and McDonald (2011).

Country	Source	Date	Title	Text
Spain	Dow Jones	14 October 2011	S&P Downgrades <u>Spain</u> One Notch, Citing Economic Woes	S&P Downgrades Spain One Notch, Citing Economic Woes 354 words 14 October 2011 01:13 Dow Jones News Service DJ English (c) 2011 Dow Jones & Company, Inc. DOW JONES NEWSWIRES Standard & Poor's Ratings Services on Thursday downgraded Spain a notch, citing increasingly unpredictable financing conditions that could squeeze a private sector already pressured by lackluster economic growth. The ratings cut is the latest blow to a large European sovereign's credit status after S&P last month downgraded Italy a notch, citing many of the same problems afflicting euro-zone economies. S&P expects the <u>Spanish economy</u> will grow at about 1% in real terms next year, a drop from the 1.5% pace it forecast in February. In downgrading the Iberian nation, S&P cited growing challenges for Spain's private sector as it seeks fresh external financing to roll over high levels of external debt. S&P now rates Spain at double A-minus, three steps below the top triple A rating. Its outlook is negative. S&P also sees the quality of assets held in Spain's financial institutions deteriorating after a separate review of its banking system found it posed additional challenges to the broader economy.
Spain	Reuters	16 January 2014	RPT- <u>Spanish</u> protesters riot in Madrid, 11 people hurt	RPT-Spanish protesters riot in Madrid, 11 people hurt 359 words 15 January 2014 23:44 Reuters News LBA English (c) 2014 Reuters Limited MADRID, Jan 15 (Reuters) - A Madrid demonstration in sympathy with protests in the northern Spanish city of Burgos against a local government plan to convert a street into a tree-lined boulevard turned violent on Wednesday, leading to 11 arrests and 11 injuries . Rioters tossed smoke bombs, threw chairs from street terraces and burned garbage containers in central Madrid after a march that began in the capital's Puerta del Sol square and ended near the ruling conservative People's Party (PP) central headquarters. Police and emergency service sources said 11 protesters were arrested and 11 people, including five police officers, were injured during the riots. It was one of 46 protests across Spanish cities on Wednesday against the state-financed project in Burgos that has stoked public fury. Critics say widespread corruption has plunged Spain into an economic crisis that has lasted for years, leaving one in four workers unemployed . The <u>Spanish economy</u> emerged from recession in the third quarter of last year but state finances are still under scrutiny by the European Union .

Table 10: Example of selected stories for Greece. Typically, crisis-related news articles are more pessimistic than optimistic. This is an example of a story that is more optimistic than pessimistic. Painted words belong in the Positive (green colour) word list of Loughran and McDonald (2011).

Country	Source	Date	Title	Text
Greece	Reuters	8 March 2011	IMF believes <u>Greek</u> debt sustainable	IMF believes Greek debt sustainable - IMF official 227 words 7 March 2011 17:22 Reuters News LBA English (c) 2011 Reuters Limited (Adds quotes, details) WASHINGTON, March 7 (Reuters) - The International Monetary Fund is confident Greece's economic program will succeed and believes its debt is sustainable, an IMF official said Monday after Moody's slashed Greece's credit rating by three notches. "I am confident that I will succeed, that the Greek debt is sustainable, and therefore our program will be successful ," IMF European Director Antonio Borges told reporters on the sidelines of a conference. "I maintain our confidence in what is happening in Greece." For more on Greece see [ID:nLDE7260IG] and [ID:nLDE7260NG]. Borges urged patience as Greece implements an IMF-EU supported program. "These are not programs that deliver miracles in a few months, therefore I am at the moment when economic consequences are more difficult to accept and people become more sceptical," Borges said, adding: "But this is a long-term program, I will have to be a little patient." Borges said Greece was not currently in the markets and could benefit from IMF and European assistance while being immune to market pressures. "I just have to make sure the banking system remains strong, and the Greeks have made quite a bit of progress on their banks," he added. (Reporting by Lesley Wroughton; Editing by James Dalglish) IMF-GREECE/BORGES (UPDATE 1) Reuters Limited Document LBA0000020110310e737001eb

Table 11: Effect of high frequency intraday news articles on stock returns. The model I employ the study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^5 c_i R_{t-i}$ where M_t takes the value of, the positive (G_t), the negative (B_t), the pessimism (P_t) and the news count (N_t), of the previous 5-minute and 30-minute interval respectively, as defined in Section 3.5. I control for five lags of returns (i.e. five 5-minute and 30-minute lagged returns for every stock market) to deal with autocorrelation in the returns. The regressions I perform are robust, using the Huber-White sandwich estimators (Huber (1967), White (1980)) to deal with autocorrelation, heteroskedasticity, heterogeneity and lack of normality.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Portugal	Greece	Italy	Germany	France	UK	Switzerland	US	Japan	China
Panel A: 30-minutes intervals										
Pessimism	-0.00812*** (-2.99)	-0.0194*** (-2.99)	-0.00886** (-2.14)	-0.00596* (-1.70)	-0.00757** (-2.29)	-0.00618** (-2.51)	-0.00536** (-2.15)	-0.0181*** (-2.75)	-0.0767*** (-3.17)	-0.0471** (-2.31)
Positive	0.0131* (1.71)	0.0296* (1.74)	0.0169 (1.50)	0.0147* (1.72)	0.0218*** (2.65)	0.00951 (1.53)	0.00927 (1.32)	0.0189 (1.03)	0.132** (2.16)	0.0937 (1.54)
Negative	-0.00768*** (-2.61)	-0.0193*** (-2.73)	-0.00784* (-1.72)	-0.00475 (-1.23)	-0.00555 (-1.56)	-0.00593** (-2.17)	-0.00499* (-1.86)	-0.0194*** (-2.82)	-0.0736*** (-2.71)	-0.0438* (-1.93)
Average Pessimism	-0.00531** (-2.09)	-0.0139** (-2.17)	-0.0106** (-2.49)	-0.00609* (-1.82)	-0.00501 (-1.60)	-0.00455* (-1.81)	-0.00332 (-1.41)	-0.00978 (-1.33)	-0.129*** (-2.72)	-0.0537 (-1.42)
News Count	-0.0000111 (-0.17)	0.0000579 (1.54)	-0.00000382 (-0.10)	0.0000197 (0.58)	-0.0000228 (-0.56)	0.0000130 (0.53)	-0.00000604 (-0.23)	-0.00000460 (-0.49)	-0.0000432 (-0.07)	0.0000166 (0.78)
Panel B: 5-minutes intervals										
Pessimism	-0.00374*** (-2.84)	-0.00689*** (-3.11)	-0.00353* (-1.93)	-0.00432** (-2.34)	-0.00330* (-1.95)	-0.000935 (-0.79)	-0.00478*** (-3.45)	-0.00728* (-1.74)	-0.0580*** (-3.18)	-0.0501** (-2.52)
Positive	0.00867** (2.37)	0.0138** (2.25)	0.00638 (1.30)	0.0122*** (2.66)	0.00658 (1.62)	0.00258 (0.91)	0.0102** (2.58)	0.00868 (0.78)	0.0960* (1.95)	0.0756 (1.30)
Negative	-0.00311** (-2.31)	-0.00634** (-2.55)	-0.00322 (-1.58)	-0.00322 (-1.62)	-0.00293 (-1.64)	-0.000705 (-0.52)	-0.00416*** (-2.83)	-0.00757* (-1.74)	-0.0567*** (-2.70)	-0.0511** (-2.24)
Average Pessimism	-0.00200** (-2.22)	-0.00522*** (-2.58)	-0.00263* (-1.75)	-0.00272** (-2.00)	-0.00197 (-1.57)	-0.000293 (-0.32)	-0.00290*** (-2.83)	-0.00104 (-0.26)	-0.0683** (-2.48)	-0.0505* (-1.83)
News Count	-0.0000473 (-0.36)	0.0000274 (1.36)	-0.0000549 (-1.54)	0.0000293 (0.75)	-0.0000223 (-0.37)	-0.0000207 (-1.01)	-0.0000552** (-2.29)	-0.000000619 (-0.09)	0.00000551 (0.23)	0.0000147 (0.66)

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 12: Effect of high frequency intraday stories on volatility. The model I employ the study the effect of the content of news in high frequency stock returns is the following: $Volatility_t = a_1 + b_1 M_t + \sum_{i=1}^5 c_i R_{t-i}$ where M_t takes the value of, the positive (G_t), the negative (B_t), the pessimism (P_t) and the news count (N_t), of the previous 5-minute and 30-minute interval respectively, as defined in Section 3.5. I define volatility as the standard deviation of intraday stock returns.. control for five lags of returns (i.e. five 5-minute and 30-minute lagged returns for every stock market) to deal with autocorrelation in the returns. The regressions I perform are robust, using the Huber-White sandwich estimators (Huber (1967), White (1980)) to deal with autocorrelation, heteroskedasticity, heterogeneity and lack of normality.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Portugal	Greece	Italy	Germany	France	UK	Switzerland	US	Japan	China
Panel A: 30-minutes intervals										
Pessimism	0.00343*** (2.81)	0.00409* (1.69)	0.00514*** (2.83)	0.00755*** (5.30)	0.00662*** (5.16)	0.00673*** (6.58)	0.00467*** (4.46)	0.00947** (2.22)	0.0156*** (2.90)	0.00960** (2.34)
Positive	-0.00994*** (-2.64)	-0.0137** (-2.15)	-0.0143*** (-2.97)	-0.0147*** (-3.94)	-0.0117*** (-3.51)	-0.0118*** (-4.45)	-0.00659** (-2.27)	-0.0143* (-1.91)	-0.0236* (-1.82)	-0.0187 (-1.48)
Negative	0.00249* (1.93)	0.00285 (1.05)	0.00382* (1.90)	0.00672*** (4.19)	0.00610*** (4.31)	0.00620*** (5.50)	0.00462*** (3.99)	0.00936** (2.06)	0.0156*** (2.65)	0.00901** (1.98)
Average Pessimism	0.00504*** (3.75)	0.00706** (1.95)	0.00828*** (3.07)	0.0104*** (3.46)	0.00778*** (1.75)	0.00871*** (3.26)	0.00604*** (5.06)	0.0142*** (1.35)	0.0141 (0.57)	0.0227** (2.02)
News Count	0.0000440*** (3.54)	0.0000272*** (3.80)	0.0000221** (2.53)	0.0000121** (2.01)	0.0000163** (2.12)	0.00000554 (1.30)	0.0000123*** (2.73)	0.000000255 (0.07)	0.00000317 (0.72)	0.0000141*** (3.11)
Panel B: 5-minutes intervals										
Pessimism	0.00126*** (3.15)	0.00172** (2.24)	0.00280*** (5.05)	0.00306*** (6.13)	0.00250*** (5.43)	0.00299*** (8.76)	0.00210*** (5.85)	0.00297*** (3.29)	0.00451*** (3.18)	0.00402** (2.40)
Positive	-0.00291** (-2.40)	-0.00266 (-1.22)	-0.00759*** (-5.00)	-0.00726*** (-5.61)	-0.00536*** (-4.43)	-0.00546*** (-6.42)	-0.00465*** (-4.77)	-0.00314 (-1.19)	-0.00656* (-1.68)	-0.00994** (-2.03)
Negative	0.00104** (2.31)	0.00170** (1.98)	0.00213*** (3.44)	0.00251*** (4.53)	0.00215*** (4.22)	0.00272*** (7.08)	0.00180*** (4.47)	0.00315*** (3.24)	0.00456*** (2.87)	0.00344* (1.76)
Average Pessimism	0.00147*** (3.54)	0.00213** (1.40)	0.00319*** (2.44)	0.00344*** (6.46)	0.00279*** (5.75)	0.00332*** (9.14)	0.00229*** (6.07)	0.00459*** (3.86)	0.00496 (1.48)	0.00983** (2.12)
News Count	0.0000117** (2.01)	0.00000565** (2.06)	0.00000656* (1.93)	0.00000335 (1.32)	0.00000585* (1.84)	0.00000184 (1.05)	0.00000286 (1.56)	-0.000000368 (-0.31)	0.000000866 (0.56)	0.00000547*** (2.93)

t statistics in parentheses

* p<.1, ** p<.05, *** p<.01

Table 13: Effect of high frequency intraday news articles similarity on stock returns and volatility - 30 minutes intervals. It could be the case that traders are triggered by repeated messages. Therefore, I provide a further check to study whether the number of highly similar messages indeed affects stock prices and volatility. For this purpose, I construct a new variable that picks up the repeats, and add this to the regression. The model thus becomes $R_t = a_1 + b_1 Repeat_t + \sum_{i=1}^5 c_i R_{t-i}$ for the effect of repeat articles on stock returns (Panel A), and $Volatility_t = a_1 + b_1 Repeat_t + \sum_{i=1}^5 c_i R_{t-i}$ for the effect on volatility (Panel B). The variable $Repeat_t$ captures the number of similar messages that were released in the previous time interval. To calculate the similarity of articles, I compare the content of every story with articles released in the previous 30-minutes (or 5-minutes), using the Ratcliff/Obershelp (Ratcliff and Metzner (1988)) pattern recognition algorithm which calculates the similarity of two strings. I define repeat articles using five different cutoff similarity percentages: articles that have a similarity ratio that is higher than 50%, 60%, 70%, 80% and 90% respectively.

	(1)	(2)	(3)	(4)	(5)
	Portugal	Greece	Italy	Germany	US
Panel A: Effect on stock returns					
<i>Similarity > 50%</i>	-0.00000119 (-0.05)	-0.000140** (-2.19)	-0.00000637 (-0.15)	0.0000245 (0.64)	-0.00000736 (-0.17)
<i>Similarity > 60%</i>	-0.00000955 (-0.43)	-0.000129** (-2.00)	0.0000198 (0.42)	0.0000306 (0.78)	0.00000709 (0.02)
<i>Similarity > 70%</i>	0.00000165 (0.07)	-0.000119* (-1.79)	0.0000296 (0.62)	0.0000266 (0.68)	-0.00000120 (-0.03)
<i>Similarity > 80%</i>	-0.00000505 (-0.22)	-0.0000932 (-1.37)	0.0000220 (0.47)	0.0000241 (0.58)	-0.00000421 (-0.09)
<i>Similarity > 90%</i>	-0.0000313 (-0.37)	-0.0000234 (-0.08)	-0.000238 (-1.39)	0.0000735 (0.62)	-0.0000538 (-0.27)
Panel B: Effects on volatility					
<i>Similarity > 50%</i>	-0.0000424** (-3.43)	0.0000107 (0.29)	-0.0000898*** (-5.39)	-0.0000445** (-2.34)	-0.0000490* (-1.65)
<i>Similarity > 60%</i>	-0.0000551*** (-4.20)	0.00000784 (0.21)	-0.0000986*** (-5.44)	-0.0000489** (-2.49)	-0.0000578* (-1.88)
<i>Similarity > 70%</i>	-0.0000610*** (-4.45)	-0.0000108 (-0.28)	-0.0000974*** (-5.09)	-0.0000511** (-2.48)	-0.0000588* (-1.91)
<i>Similarity > 80%</i>	-0.0000613*** (-4.32)	0.0000147 (0.40)	-0.0000937*** (-4.87)	-0.0000511** (-2.58)	-0.0000603** (-1.98)
<i>Similarity > 90%</i>	-0.0000811 (-1.34)	0.000234* (1.88)	-0.000109 (-1.42)	-0.0000458 (-0.66)	-0.000312 (-1.31)

t statistics in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$