PhD-FDEF-2016-08
The Faculty of Law, Economics and Finance

DISSERTATION

Presented on 07/06/2016 in Luxembourg
to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN SCIENCES FINANCIÈRES

by

Andreas CHOULIARAS

Born on 17 January 1987 in Ioannina (Greece)

ESSAYS ON THE IMPACT OF INFORMATION ON FINANCIAL MARKETS: EVIDENCE FROM THE EUROPEAN FINANCIAL CRISIS

Dissertation Defence Committee:

Dr. Theoharry Grammatikos, dissertation supervisor
Professor, *Université du Luxembourg, Luxembourg*

Dr. Thorsten Lehnert, Chairman
Professor, *Université du Luxembourg, Luxembourg*

Dr. Christoph Schommer, Deputy Chairman
Professor, *Université du Luxembourg, Luxembourg*

Dr. Dimitrios Gounopoulos
Professor, *Newcastle University Business School, United Kingdom*

Dr. Konstantinos Tsiveriotis
Chief Executive Officer (CEO), *Eurobank Private Bank, Luxembourg*

Contents

Chapter 1 Introduction/Acknowledgements **4**

1.1 Introduction . . . 4

1.2 Acknowledgements . . . 8

Chapter 2 Extreme Returns in the European Financial Crisis **10**

2.1 Introduction . . . 1

2.2 The Data . . . 3

2.3 Extreme Returns . . . 4

2.3.1 The Base Model . . . 4

2.3.2 Examining the presence of extreme returns transmission . . . 6

2.3.3 Effects within groups . . . 7

2.3.4 Effects across groups . . . 10

2.3.5 Alternative specification of the cutoffs that define ”extreme returns” . . . 13

2.4 Robustness and alternative specifications . . . 14

2.5 Conclusion . . . 15

Chapter 3 News Flow, Web Attention and Extreme Returns in the European Financial Crisis **30**

3.1 Introduction . . . 31

3.2 Related Literature . . . 32

3.3 The Data . . . 33

3.3.1 News Flow and Web Attention (SVI) . . . 34

3.3.2 Stock Returns and Extreme Returns . . . 40

3.4 Empirical Findings . . . 43

3.4.1 News Flow, Web Attention and Extreme Returns . . . 43

3.5 Robustness and alternative specifications . . . 45

3.6 Conclusion . . . 46

Chapter 4 High Frequency News in the European Financial Crisis **60**

4.1 Introduction . . . 61

4.2 The Data . . . 63

4.3 The Methodology . . . 64

4.3.1 Preprocessing the financial data . . . 64

4.3.2 News Flow . . . 64

4.3.3 Preprocessing the news data . . . 65

4.3.4 Keywords in titles as a determinant of news items relevance . . . 65

4.3.5 Textual Analysis . . . 66

4.4 Empirical Findings . . . 67

4.4.1 Effects on stock market returns . . . 67

4.4.2 The impact of repeat articles . . . 70

4.4.3 Overnight Returns and News Similarity . . . 71

4.4.4 Effects on stock market volatility . . . 72

4.5 Conclusion . . . 74

Chapter 5 Conclusions and Future Work **75**

5.1 Conclusions and Future Work . . . 75

Chapter 6 Appendix **79**

Chapter 1

Introduction/Acknowledgements

1.1. Introduction

The European Union (EU) started as a project after the World War II[1]. The idea was that countries with strong trade links become more interdependent, and thus should find it more difficult to enter into conflict. What started from six countries (Belgium, Germany, France, Italy, Luxembourg and the Netherlands) in 1958, known as the European Economic Community (EEC), transformed into the European Union (EU) in 1993, with twenty-eight (28) member states as of 2016. The next level of the European integration came in 1999, with the introduction of the euro, the common currency, which was initially introduced as a virtual currency, while banknotes and coins were launched on 1 January 2002[2]. As of 2016, nineteen (19) countries have adopted the euro as their currency[3]. More than five decades of growth, prosperity and convergence were interrupted by the advent of the Global Financial Crisis. The crisis in Europe followed the dramatic repercussions of events in the American banking sector of trouble in the American mortgage market. European banks that had invested in the American market were hit so hard, that governments felt driven to intervene and bail them out, often at great cost. Investors became worried, and began scrutinizing government finances more carefully than in previous years[4]. The most vulnerable European countries were the five Euro-periphery countries (Portugal, Ireland, Italy, Greece and Spain), of which Greece had the biggest problems, mainly due to the existing high debt and deficit, which made markets less willing to keep buying their sovereign debt bonds. The current PhD thesis consists of three (3) research papers, all of which concern the "European Financial Crisis". The Euro-crisis has been one of the most significant and dramatic events of the past decades. It drew worldwide attention, and has had severe repercussions in Europe and has shaken political dynamics, provoking a series of global debates. A few of the most dramatic events/dates during this period are the

[1] http://europa.eu/about-eu/basic-information/about/index_en.htm

[2] http://europa.eu/about-eu/basic-information/money/euro/index_en.htm

[3] http://ec.europa.eu/economy_finance/euro/index_en.htm

[4] http://ec.europa.eu/economy_finance/explained/the_financial_and_economic_crisis/why_did_the_crisis_happen/index_en.htm

following:

- **May 2010:** 110 billion euro bailout package for **Greece**[5]
- **November 2010:** 85 billion euros bailout package for **Ireland**[6]
- **May 2011:** 78 billion euro bailout for **Portugal**[7]
- **January 2012:** Credit rating agency Standard & Poor's downgrades **France** and **eight other eurozone countries**, blaming the failure of eurozone leaders to deal with the debt crisis[8]
- **June 2012**, **Spain's** Economy Minister Luis de Guindos says that the country will make a formal request for 100 billion euros in loans from eurozone funds to help its banks[9]
- **July 2012 Spanish** 10 year bond yields rise above 7 percent[10]

Each of the three papers contributes in these debates: the magnitude of the transmitted financial shocks; the daily effects of media pessimism and Web Attention on stock markets; and the high frequency (intraday) effects of media on international (European and overseas) stock markets. The first paper, "Extreme Returns in the European Financial Crisis" deals with the dynamics of extreme equity returns during the European Financial Crisis. The paper uses a setting of fifteen (15) European Union countries, ten (10) of which are Eurozone countries. The countries are split into three groups: the Euro-core countries (Germany, France, the Netherlands, Finland, Belgium), the Non-euro countries (Sweden, UK, Poland, Czech Republic, Denmark) and the Euro-periphery countries (Portugal, Ireland, Italy, Greece, Spain). Using extreme returns on daily stock market data from January 2004 until March 2013, we find that transmission effects are present from the Euro-periphery group to the Non-euro and the Euro-core groups for the tails of the returns distributions for the Pre-crisis, the US-crisis and the Euro-crisis periods. During the crises, the shocks transmitted were more substantial (in some cases, extreme bottom returns doubled). As extreme returns increased during the financial crisis periods, the expected losses on extreme return days have increased significantly. Given the fact that stock market capitalizations in these country groups are trillions of Euros, a 1% or 2% increase in extreme bottom returns (in crisis periods) can lead to aggregate losses of tens of billions Euros in one single trading day.

The second paper, "News Flow, Web Attention and Extreme Returns in the European Financial Crisis" studies the daily effect of News Sentiment and Web Attention, quantified through the use of Textual Analysis and Google Trends (Web Attention) data, in the probabilities of extreme equity returns in three groups of European countries. Using daily stock market data from January 2004 until March 2013 and textual analysis on more than 24,000 news articles from seven leading international news providers (*Dow Jones Newswires, Thomson Reuters, Financial*

[5] http://www.reuters.com/article/us-eurozone-idUSTRE6400PJ20100502

[6] http://www.theguardian.com/business/ireland-business-blog-with-lisa-ocarroll/2010/nov/28/ireland-bailout-full-government-statement,http://www.telegraph.co.uk/finance/financialcrisis/8150137/Ireland-forced-to-take-EU-and-IMF-bail-out-package.html

[7] http://www.reuters.com/article/us-portugal-bailout-idUSTRE7425UP20110503,http://www.bbc.com/news/business-13275470

[8] http://www.reuters.com/article/us-eurozone-sp-idUSTRE80C1BC20120113

[9] http://www.theguardian.com/world/2012/jun/09/spain-bank-bailout-eurozone-crisis

[10] https://www.washingtonpost.com/blogs/ezra-klein/wp/2012/07/19/why-spains-7-percent-bond-yields-have-the-whole-world-on-edge/

Times, The Wall Street Journal, The New York Times, The Telegraph and The Times - all extracted from the *Dow Jones Factiva*), 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.

Finally, the third paper, "High Frequency Newswire Textual Sentiment Analysis: Evidence from International Stock Markets during the European Financial Crisis" deals with the high-frequency (5 minutes and 30 minutes) effects of intraday news (provided by the Thomson Reuters and the Dow Jones Newswires) on international equity markets returns and volatility, during the European Financial Markets. The paper employs high-frequency "tick data", timestamped to a frequency of milliseconds, and along with high frequency (intraday) news from the *Dow Jones Newswires* and the *Thomson Reuters Newswires*, the paper finds that 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 extensive coverage) provide a channel through which "bad" news is instantaneously circulated and provide worldwide "shocks" to stock prices in extremely small time windows (even 5-minutes).

The tools used in this PhD thesis provide a possibility to link financial news to stock markets, but there may exist alternatives which can be explored in future research. As with every model, there is room for further enhancements and alternatives. One thing is certain: the tools and techniques that are used in this thesis have stood out well in the literature, and have resulted in multiple publications in leading finance journals in the recent years (Bae et al. (2003) (RFS), Boyson et al. (2010) (JF), Loughran and McDonald (2011) (JF), Chen et al. (2013) (RFS), Garcia (2014), Garcia (2013) (JF), Loughran and McDonald (2013) (JFE), Jegadeesh and Wu (2013) (JFE), Ahern and Sosyura (2014) (JF), Dougal et al. (2012) (RFS), Kelley and Tetlock (2013), Tetlock (2007) (JF), Tetlock et al. (2008) (JF), Da et al. (2011) (JF), Da et al. (2015) (RFS), Bodnaruk et al. (2015) (JFQA)).

This PhD Thesis contributes to the field of "Textual Analysis", which has gained a significant amount of attention in Finance in the past few years (Loughran and McDonald (2011) (JF), Tetlock (2007) (JF)). Textual analysis attempts to convert qualitative information that is contained in the textual content of news and corporate announcements, into a quantified measure of its content, by calculating the amount of "negative" and "positive" information that is included in the text. Furthermore, it contributes to the "Web Attention" literature (Da et al. (2011) (JF), Da et al. (2015) (RFS)), which studies how Web Searches of investors (captured through Google Trends data) affect movements in the stock market. In addition, it contributes to the high frequency literature, as it studies the speed of adjustment in financial markets, examining how fast information is assimilated in stock prices (even small time frames such as 5 minutes and 30 minutes are enough for stock markets to move after information is released). Finally, it contributes to the policy debate regarding the European Financial Crisis, that was concerned with whether bad news about crisis-hit countries can indeed affect other countries, European and overseas (North America, Asia).

1.2. Acknowledgements

This PhD Thesis would have not be successful without the guidance, the support and the help of my PhD Committee. My advisor, Professor Theoharry Grammatikos, helped me in every step of the process. He introduced me to an extremely interesting research field, and spent a lot of time guiding me and discussing with me a series of research questions. I would also like to thank Professor Thorsten Lehnert who was always there to provide guidance, feedback and help, whenever needed, in a very kind, supportive and productive manner. Finally, we had many interesting discussions with Professor Christoph Schommer, as the project we undertook lies in the intersection of finance and computer science, and his opinions and knowledge provided me with interesting ideas and feedback throughout these years. Finally, I would like to thank Dr. Konstantinos Tsiveriotis (CEO of Eurobank Private Bank Luxembourg), and Professor Dimitrios Gounopoulos (Professor of Finance, Newcastle University Business School) for agreeing to be on my PhD Thesis committee.

The PhD is a constant journey towards more knowledge. While at the LSF, I was lucky to take a series of interesting courses, which helped me shape my knowledge and formulate my research agenda. I would like to deeply thank Professors Gautam Tripath, Matti Suominen, Giovanni-Barone Adesi, Rajnish Mehra, Roman Kraussl, Christian Wolff and Pierre Picard.

Attending conferences and interacting with colleagues has been a very helpful experience. I would like to thank the following for their helpful comments and suggestions: 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 (IFABS) 2014 conference, as well as the finance seminars of the Peking University HSBC Business School and the ICMA Centre, Henley Business School, University of Reading.

Throughout these years, I had the pleasure and fortune, to meet and interact with some very interesting people in the finance academia. For our communication, their advice, and their help, which has shaped my research thinking and approach tremendously, I would like to thank Professors Gerard Hoberg, Bill McDonald, and Tim Loughran.

My colleagues and friends in LSF made this journey more enjoyable. I would like to thank Tim, Rene, Fanou, Ali and Yuehao.

I would also like to thank my good friends from my NTUA (ECE) days, Kostas, Ioannis, Dimitris, Panos, Nikos, Ilias, for our continuing friendship, which has endured throughout these years, even though we are all in different countries.

The education I received in the National Technical University of Athens (National Metsovion Polytechnic), at the School of Electrical and Computer Engineering (undergraduate studies) and at the School of Applied Mathematical and Physical Sciences (graduate studies) has proven to be my strongest asset in this journey. I am grateful to my NTUA Professors and to my fellow NTUA students, for setting the bar so high. The very strong association between Electrical and Computer Engineering, Mathematics, Physics and Finance, fascinates me every day. The endless hours of studying at the "Prytaneia" and at the "Anagnostirio" have (at least partially) paid off.

A wise man once told me that "research is a wonderful hobby, but teaching is where you'll impact lives". I would like to thank Apostolos Christopoulos (UOA), Ioannis Leventidis (UOA) and Leonidas Rompolis (AUEB) for teaching me a lot during my graduate studies. Athanasios (Sakis) Mpasogiannis taught me Mathematics, and Dimitrios (Takis) Efthymiou taught me Physics, the two sciences that are the basis, the core and the foundation upon which all other sciences lay.

Finally, I would like to thank my family (Sara, Aigli, Andreas, Cleopatra, Alexandros) for their unconditional love and support. I would probably not be writing these words without them in my life.

Chapter 2

Extreme Returns in the European Financial Crisis

We examine the transmission of financial shocks among three groups of countries: the Euro-periphery countries (Portugal, Ireland, Italy, Greece, Spain), the Euro-core countries (Germany, France, the Netherlands, Finland, Belgium), and the major European Union -but not euro- countries (Sweden, UK, Poland, Czech Republic, Denmark). Using extreme returns on daily stock market data from January 2004 until March 2013, we find that transmission effects are present for the tails of the returns distributions for the Pre-crisis, the US-crisis and the Euro-crisis periods from the Euro-periphery group to the Non-euro and the Euro-core groups. Within-group effects are stronger in the crisis periods. Even before the two crises there was a significant shock transmission channel from the Euro-periphery to the Euro-core and the Non-euro. During the crises, the shocks transmitted were more substantial (in some cases, extreme bottom returns doubled). As extreme returns have become much more severe during the financial crisis periods, the expected losses on extreme return days have increased significantly. Given the fact that stock market capitalizations in these country groups are in trillions of Euros, a 1% or 2% increase in extreme bottom returns (in crisis periods) can lead to aggregate losses of tens of billions Euros in one single trading day.

JEL classification: G01, G15.

Keywords: Financial Crisis, Financial Contagion, Spillover, Euro-crisis, Stock Markets.

2.1. Introduction

The recent global financial crisis began as a crisis in the subprime mortgage loan business in the United States of America in 2007, and continued with multiple waves of financial distress that hit the global financial markets. Since the beginning of 2010 the Euro area faces a severe financial crisis. What started off as a sovereign debt crisis in Greece soon transmitted itself to Portugal, Ireland, Cyprus and, at least partially, to Spain and Italy. Pretty soon it became clear for Europe that beneath the sovereign debt crisis surface there also existed a severe banking crisis. The propagation of financial distress from one country to another, with stock markets, bond yields and CDS spreads being affected, makes the case of studying the transmission of extreme returns more pertinent than ever[2].

A number of researchers investigate the recent eurozone financial crisis and its transmission effects, giving particular emphasis on the sovereign debt and the Credit Default Swaps (CDS) markets. Missio and Watzka (2011) report the existence of contagion effects using dynamic conditional correlation models. Metiu (2012) employs a simultaneous equations model and examines the tails of bond yield distributions, an approach derived from the Extreme Value Theory and Value-at-Risk, and finds structural shift contagion effects for the crisis periods. Other papers, however, do not find contagion effects for the sovereign bond and the credit default swaps markets. See, for example, Caporin et al. (2013) and Bhanot et al. (2012).

The study of stock markets during financial crises has not been examined sufficiently in the previous literature despite them being the most liquid markets. In this paper we investigate the stock market financial transmission effects of the European financial crisis (and the US-crisis) for three groups of countries: two groups of eurozone countries, the Euro-core eurozone countries (Germany, France, Netherlands, Belgium, Finland) and the Euro-periphery eurozone countries (Portugal, Ireland, Italy, Greece, Spain), and finally a group for European Union (EU) but non-euro countries (Sweden, UK, Poland, Czech Republic, Denmark). The creation of these three groups is justified mainly by the existence of the European Union and the Eurozone. The European Union is primarily a free-trade union, in which free movement of capital, labor and tradable goods take place. This has resulted in strong ties which go well beyond these trade relationships, taking also the form of a primary political union (with the existence of European Union legislation which applies to all member countries, the European Parliament, and various political and administrative authorities such as the European Commission). The Non-euro group is heterogeneous, but their participation in the European Union justifies grouping these countries together. A subset of the European Union countries have formed the Eurozone which as well as being a free trade area, is a monetary union as well, sharing the Euro as a common currency. The heads of state of the member countries of the Eurozone meet regularly in order to coordinate policy and take decisions. The agreed measures affect all Eurozone countries. This justifies the inclusion of the Euro-core in our sample, which are the five countries with the highest market capitalizations among all Eurozone countries. Finally, the five countries that were most badly hit by the recent

[2]The shock transmission literature is extensive. See, for example Allen and Gale (2000), Rigobon (2002), Kaminsky et al. (2003), Pericoli and Sbracia (2003), Bekaert et al. (2005), Forbes and Rigobon (2001), Aït-Sahalia et al. (2010), Dungey et al. (2005), Corsetti et al. (2005).

crisis, are Portugal, Ireland, Italy, Greece and Spain. These five countries are part of the European Union and the Eurozone, which is why we group them together in the Euro-periphery group. The transmission of shocks between the Euro-core and the Euro-periphery is interesting, because, being part of the same trading union and the same monetary union, financial problems in one group may indeed also affect the other groups. The Euro-core countries were called upon to provide financial aid (along with the other Eurozone countries, the European Central Bank and the International Monetary Fund) to the Euro-periphery countries that were in need. Since hundreds of billions Euros were provided in assistance, it is worth examining the effect of extreme stock market shocks to the three country groups for one more reason: the provision of financial assistance may not only have been an act of solidarity, but also an act of self interest for the Euro-core and the Non-euro group, if this financial assistance was able to mitigate the transmitted shocks[3].

The correlations framework has been widely used by previous authors in related studies but there is no consensus in the literature as to how to best define contagion when using that framework. Forbes and Rigobon (2001) claim that heteroskedasticity biases correlation tests for contagion[4]. To avoid this problem we follow the extreme returns approach proposed mainly in two papers, Bae et al. (2003) and Boyson et al. (2010). Bae et al. (2003) examine the coincidence of extreme return shocks across countries within a group and across groups, while Boyson et al. (2010) study hedge funds contagion. A number of other studies have also used this methodology[5]. Moving in line with multinomial logistic analysis, as proposed by Bae et al. (2003) and Boyson et al. (2010), we can use control variables (covariates) in order to justify the characteristics of extreme returns. Furthermore, this approach allows us to study the effects within groups, and the crisis transmission across groups. 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 the crisis transmission from the Euro-periphery group to the other two groups (Euro-periphery vs. Euro-core, Euro-periphery vs. Non-euro)[6],[7]. We find that extreme returns in Euro-periphery countries are related to extreme returns in the Euro-core and the Non-euro country groups. In order to test if the crises result in a fundamental shift in the transmission mechanism the extreme returns methodology is applied not only on the entire period (as in previous studies), but also separately on each of the three subperiods (Pre-crisis, US-crisis, Euro-crisis).

We find that even before the two crises there was a significant shock transmission channel from the Euro-

[3]Although the Non-euro countries did not directly contribute to the financial assistance packages, they are indirectly affected through the IMF contributions. On top of that, decisions to create the European Financial Stability Facility (EFSF) and the European Stability Mechanism (ESM) in order to deal with the Euro-crisis were taken by all European Union member states.

[4]By applying a correction they find no contagion for the 1997 Asian crisis. On the other hand, Corsetti et al. (2005) claim that the variance restrictions imposed by Forbes and Rigobon (2001) are "arbitrary and unrealistic". They find evidence for at least five countries facing contagion effects during the Hong Kong stock market crisis of 1997.

[5]See, for example, Markwat et al. (2009), Lucey and Sevic (2010), Christiansen and Ranaldo (2009), Gropp et al. (2009), Chouliaras and Grammatikos (2015)

[6]Another study that uses this approach is Thomadakis (2012), but our main difference is that he considers the Eurozone countries as one group, thus not studying the within eurozone dynamics of the various subgroups, and he studies the interactions mainly with the USA, for the industrials sectors of the stock exchanges.

[7]One critique on Bae et al. (2003) and Boyson et al. (2010) is that they arbitrarily pick the top and the bottom 5% from the sample of returns to examine the joint occurrence of extreme returns. This critique has indeed some merit but a choice of cutoff points is a necessary decision in order to proceed with this methodology and to study the tails of the marginal return distributions in order to see what happens in the presence of extreme returns. The results of our study were found to be robust in the change of the percentiles.

periphery to the Euro-core and the Non-euro groups. During the crises the shocks transmitted were more substantial, not only for the Euro-periphery countries, but also for the Euro-core group and the Non-euro group. Given the extremely big size of the European equity markets, a 1% or 2% increase in the magnitude of extreme bottom returns can lead to aggregate losses of tens of billions Euros in one single trading day, resulting in very important implications for investors and policy makers. The differences in the models of the three periods are further verified by using likelihood ratio tests.

The remainder of this paper is organized as follows. Section 4.2 presents the data. Section 4.3 presents the basic model and we explain how we study the crisis transmission within and across groups. Section 3.5 provides a set of robustness and alternative specifications. Section 4.5 is a conclusion.

2.2. The Data

The main area of study for this paper is the European Union. 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)[8]. We examine the period from 01/01/2004 until 13/03/2013 using daily financial data obtained from the Thomson Reuters Datastream. Our selection of countries for the Eurozone follows to a large extent the studies of Missio and Watzka (2011), Caporin et al. (2013), Bhanot et al. (2012), and Metiu (2012). Country group log returns (expressed in local currency) and standard deviations are calculated on the equally weighted portfolio of the country stock market daily returns (expressed in local currency) for each group. Table 2.1 shows the summary statistics and correlation matrices of the percentage returns of the major stock market indices (Panel A and Panel B)[9]. To be able to make comparisons between normal and abnormal times in the financial markets, we split our sample in three subperiods (Pre-crisis, US-crisis Euro-crisis):

- the Pre-crisis period (from 1 January 2004 until 26 February 2007)
- the US-crisis period (from 27 February 2007 until 7 December 2009).[10]
- the Euro-crisis period (from 8 December 2009 until 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.

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

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

[10]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`.

Insert Table 2.1 here

For the Pre-crisis period all groups of countries had positive mean 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%). These numbers may appear to be high (a mean of +0.101% per day leads to almost 25% per year), but one has to take into account the rally that was observed in the stock markets during the Pre-crisis period. For example, the London Stock Exchange index had a value of 273 in January 2004, and it climbed up to 1197 in February 2007, which means that in the end of the Pre-crisis period its price was almost four times higher than the beginning of the Pre-crisis period. 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 -0.016%). 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. As far as the correlations are concerned the main remark is that they increased between the Pre-crisis and the crisis periods. The correlation between the Non-euro and the Euro-periphery group grew from the Pre-crisis value of 0.791 to 0.919 in the US-crisis period, then went down to 0.836 at the Euro-crisis period. The correlation between the Euro-periphery and the Euro-core followed a similar pattern, increasing from 0.870 Pre-crisis to 0.928 in the US-crisis, then declined to 0.876 in the Euro-crisis period, which is still much higher than the Pre-crisis period. The correlation between the Non-euro and the Euro-core group increased from 0.806 Pre-crisis to 0.912 in the US-crisis period, remaining at the elevated level of 0.926 in the Euro-crisis. To summarize, for both crisis periods (US and Euro-crisis) the correlations are higher than what they were in the Pre-crisis period, and the most hit group is found to be the Euro-periphery group, which has negative mean returns in both crisis periods.

2.3. Extreme Returns

2.3.1. The Base Model

According to Bae et al. (2003) and Boyson et al. (2010), an extreme return is one that lies below (or above) the lowest (or the highest) quantile of the marginal return distribution. This methodology concerns the counts (i.e. instances) of joint occurrences of extreme returns within a group on a particular day. The original approach studies the number of extreme returns for the entire test period, taking as thresholds for extreme returns the 5th and the 95th percentiles. In our case, and in order to have a sufficient number of observations, we choose as thresholds the 10th and the 90th percentiles, as in Boyson et al. (2010) (our findings are robust to the 5th and 95th percentiles). Thus, 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. To demonstrate the application of the Bae et al. (2003) model, extreme bottom and top counts are reported in Table 2.2, using the one cutoff for the overall sample. 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 sum of the extreme returns for all countries that belong to that group for that specific day.

Insert Table 2.2 here

The left side of Table 2.2 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 noting 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.

We notice that out of the 10% lowest returns for all Euro-periphery 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. For the Euro-core countries, 109 days are identified in which all five countries experienced an extreme bottom return shock. On 55 days all five Non-euro countries experienced bottom extreme returns, with the Czech Republic having the most days (84) as the only country experiencing an extreme bottom return. On the other hand, from the top 10% distribution, all Euro-periphery countries experienced an extreme top return on 40 days. There are a total of 91 days during which five Euro-core countries experienced extreme top returns. On a total of 28 days, all Non-euro countries had an extreme top return, with the Czech Republic once more having the most days (95) with extreme top returns.

The graphical illustrations of bottom extreme return counts for the three groups appear in the following Figure:

Insert Figure 2.1 here

It is obvious that extreme bottom returns have a much higher density in the crisis periods. What we observe is "bottom extreme returns clustering", since as one would expect most of the extreme bottom returns fall within the crisis periods. This happens for all the three (3) groups of the fifteen (15) European countries we study, and provides a visual confirmation of the quantitative result we found as far as the intensification of extreme returns is concerned.

The methodology of Bae et al. (2003) can be applied to study two types of spill-over effect: within groups and across groups. In this paper we mainly focus on effects across groups.

2.3.2. Examining the presence of extreme returns transmission

In order to capture the effects within a group 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

$$P_i = P(Y_t = i|x_j) = \frac{G(\beta_i' x_j)}{1 + \sum_{j=1}^{m-1} G(\beta_j' x_j)}, \tag{2.1}$$

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 3.5 represents a multinomial logistic (or multinomial logit) model. Such models are estimated using maximum likelihood, with the log-likelihood function for a sample of n observations given by

$$logL = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij} log P_{ij}, \tag{2.2}$$

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 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:

$$pseudoR^2 = 1 - [logL_\omega / logL_\Omega]. \tag{2.3}$$

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 return counts. 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 average exchange rate changes, 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 a count of a specific level, P_i, by evaluating the covariates at their unconditional values,

$$P_{ij}^* = \frac{exp(\beta_i' x_j^*)}{1 + \sum_{j=1}^{m-1} exp(\beta_j' x_j^*)}, \tag{2.4}$$

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. This is why it is necessary to differentiate 3.5 in order to obtain the partial effects of the covariates on the probabilities

$$\delta_{ij} = \frac{\delta P_{ij}}{\delta \beta_i} = P_{ij}[x_j - \sum_{k=0}^{J} P_{ik}\beta_k] = P_{ij}[\beta_j - \bar{\beta}] \tag{2.5}$$

where $\bar{\beta} = \sum_{k=0}^{J} P_{ik}\beta_k$, the weighted average of every subvector of β. 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 the independent covariate. In such models we are interested in seeing whether these marginal effects are statistically significant or not. 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}. [11]

In our case, we have a variable Y_t that counts the number of extreme returns and takes the value i when extreme returns (top or bottom) occur for the same day in i stock market indices on day t. This variable is calculated separately for the Euro-core, the Euro-periphery, and the Non-euro groups. Then, in the multinomial logistic regression Equation 3.8 P_i is equal to $P(Y_t = i|x_t)$ where $Y_t = 0, 1, 2, ...k$ is the extreme return count variable that is created for the Non-euro, and for each of the country groups we defined (Euro-periphery vs. Euro-core etc.). So, we have k=5 for all three country groups, where x_t is a vector of explanatory variables (covariates), on day t. In Equation 3.8, the argument of the exponential part (representing the logistic function) is a function of the covariates (x_t) and the coefficients(the betas). This function is a linear expression of the arguments. Let's call it $g_i(t)$. We will use this function (which will take different forms) to study both the "within" and "across" groups extreme returns effects.

2.3.3. Effects within groups

In this section, we study the three country groups to determine whether there exist effects within them. Each of these groups has it's own set of covariates. In line with Bae et al. (2003) and Gropp et al. (2009), as independent

[11] 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:

$$\ln \frac{P_{ij}}{P_{i0}} = \beta_i' x_j \tag{2.6}$$

variables incorporated in $g_i(t)$ we have the intercept, the conditional volatility of the group stock index at time t (h_t)[12], the average exchange rate change (per US dollars) in the group (e_t), the average short term (ST) interest rate level in the group (i_t) as a proxy for the interbank short term liquidity risk[13], and the average long term (LT) spread change (b_t) vis-à-vis Germany as a proxy for the sovereign risk change[14].

We include exchange rate changes following Bae et al. (2003) who find that when currencies fall on average (which means that e_{it} rises) extreme returns are more common. Thus, the logistic regression $G(\beta_i' x) = exp(g_i(x_t))$ of equation 3.5 has the following form for $g_i(x_t)$:

$$g_i(x_t) = b_{0i} + b_{1i}h_t + b_{2i}e_t + b_{3i}i_t + b_{4i}b_t \tag{2.7}$$

where i=0, 1, 2, 3, 4, 5 for each country group, the extreme return count for the group. Equation 3.11 represents the inter-group effects formula for the three groups examined. For each group we calculate the equally weighted average group values, on a daily basis, of the conditional volatility (h_t), the exchange rate change (e_t), the short term interest rates levels (i_t), and the long term spread change vis-à-vis Germany (b_t).

We estimate these models for each group, for the entire sample and for each of the three time periods. It is worth noting that, in the second case, the extreme return counts are calculated separately for each of the three periods. In other words, in each of the three periods the bottom and top extreme values correspond to the respective 10% and 90% threshold points of each period. For the entire sample, we calculate the sum of the three subsamples (with three cutoffs). Otherwise, we would have observations in the subsamples that might not be in the entire sample (or vice versa). As a robustness test, we also calculated the entire sample using one cutoff (see Section 3.5). We first present in Table 2.3 the detailed findings for the Euro-periphery group for bottom extreme returns and for the entire period.

Insert Table 2.3 here

All the coefficients are the marginal effects, calculated as described in Equation 3.9[15]. The probability that no Euro-periphery country has a bottom tail return is equal to 77.49%. This is calculated as the fraction of the number of 0 extreme returns divided by the total days $\frac{1859}{2399} = 0.774$. The coefficient β_{01} corresponds to the event Y=1, in other words the event where only one Euro-periphery country has an extreme return (an exceedance) on that day, and the probability of this event is calculated as $P_1 = \frac{\exp(\beta_{01})}{1+\sum_{i=0}^{5}\exp(\beta_{0i})}$. This probability is found to be equal

[12]The conditional volatility is estimated using an EGARCH(1,1) model to the equally-weighted group indexes.

[13]Short term interest rates are available in Datastream (3-month Interbank interest rates).

[14]Spreads are calculated as the difference between the yield of the 10 year government bond of country i's debt and the yield of the 10 year German government bond. Naturally, for the Euro-core group, one of the five countries is Germany, so, for Germany, the LT Spread Change will be zero, but the other four Euro-core countries will have their respective LT daily spread change.

[15]There are 23 more Tables like Table 2.3 (24 in total - 12 for bottom and 12 for top extreme returns. We do not include them in this paper due to space constraints. These tables are available upon request.

to 12.3% (see baseline predicted probability of Table 2.3, for column (1), i.e. for one bottom extreme return). If currencies in the group fall on average (in which case e_{it} rises), the probability of extreme returns increases, since the signs of the exchange rate marginal effects are positive, and statistically significant at the 5% level for the first exceedance, and at the 1% level for the coincidence of two, three, four and five bottom extreme returns. In their study Bae et al. (2003) measured returns in dollars and the fact that they came up with very similar results made them wonder whether the stock return contagion they measured was actually foreign exchange contagion. Thus, they also estimated their models in local currencies, but the results were similar to the dollar returns. But we estimate our models in local currencies from the beginning, so we do not face such an issue. The results for ST interest rates are mixed since two marginal effects are statistically significant, for the outcomes of one and five bottom counts in the group, but with contradictory signs. Regarding the LT spread change in the group vis-à-vis Germany, we find positive and statistically significant marginal effects. For all extreme bottom outcomes except for the second, the marginal effects are significant at the 1% level. The positive sign of the coefficient indicates that a 1% increase in the average Euro-periphery LT spread increases the probability of extreme bottom returns in the group. A change of 1% in the average LT spread of the Euro-periphery group, increases the probability of two bottom extreme returns by 14.7%. To simplify the presentation, in Table 2.4 we show a summary for the within groups results, for the entire sample, and the three periods separately. The number of “+” (or “-”) indicates the number of statistically significant (in the 1% or 5% levels) and positive (or negative) marginal effects.

Insert Table 2.4 here

Pre-crisis, the effect of the covariates on the probability of extreme returns is rather weak, while the role of covariates increases significantly in the crisis periods in most of the cases. The results are even stronger when we take the extreme returns over the entire period which notably includes the US-crisis as well. The effect of volatility is somehow smaller in the Euro-crisis period compared to the Pre-crisis period, for the bottom tail. Exchange rate changes are not significant for the bottom count of the Non-euro group in the Pre-crisis period, with zero significant marginal effects, while they became significant in four of the five cases in the Euro-crisis period. Exchange rate changes have a positive coefficient for the bottom tail, and a negative coefficient for the top tail. This means that higher exchange rates (i.e. weaker currencies) lead to a higher probability of extreme bottom returns and a lower probability of extreme top returns. Average ST interest rates are not significant in most of the cases, while average LT spread changes become more significant in the Euro-crisis period, especially for the Euro-core and the Euro-periphery group as far as the bottom tails are concerned, and the Non-euro and Euro-core groups for the top tail returns. For the bottom tail, higher average group LT spread changes (i.e. an increase in the average group sovereign risk) lead to higher probabilities of extreme group bottom returns, while they decrease the probability of extreme top returns.

In summary, the findings so far indicate a much tighter relationship between the fundamental factors (covariates) affecting the extreme stock market movements within each group during the Euro-crisis and US-crisis periods compared to the Pre-crisis period.

2.3.4. Effects across groups

Next we test for across-groups effects. This deals with the question of whether the number of extreme return counts in one group (the Euro-periphery group) can help predict the number of extreme returns in other groups (the Euro-core and the Non-euro groups). According to Bae et al. (2003) and Boyson et al. (2010), if a fraction of extreme returns in one group is unexplained by its own covariates, but can be explained by extreme returns in another area, this can be interpreted as evidence of transmission of extreme returns across groups[16].

Our primary interest is to study for across-groups effects from the Euro-periphery group to the Non-euro and the Euro-core groups. To examine this question, we reestimate the models of Table 2.4 for the Euro-core and Non-euro groups respectively, adding a covariate related to the extreme return count (Y^*_{jt}) from the Euro-periphery. In this case the equations for the across groups examination take the following shape:

$$g_i(x_t) = b_0 + b_1 h_{it} + b_2 e_{it} + b_3 i_{it} + b_4 b_{it} + b_5 Y^*_{jt} \tag{2.8}$$

For example, to examine if the Euro-periphery group provokes transmission effects in the Euro-core group the dependent variable is the number of extreme returns in the Euro-core group and the first three covariates of the right hand side concern the Euro-core group, while the last covariate is related to the count of extreme returns of the Euro-periphery group on that day. The null hypothesis of no transmission effects can be rejected in case the coefficient of Y^*_{jt} is found to be statistically significant.

In Table 2.5 we present the detailed across groups effects from the Euro-periphery extreme bottom returns to the Euro-core group for the entire period:

Insert Table 2.5 here

The main variable of interest is the "Bottom Count Euro-periphery" variable, which is the Y^*_{jt} in equation 2.8. We see that this variable is positive and significant for all five Euro-core bottom outcomes. In other words a higher value of bottom Euro-periphery extreme returns increases the probability of bottom extreme returns for the Euro-core group as well. For one more Euro-periphery country having extreme bottom returns, there is an increase of

[16] In general, the definition of contagion is far from being simple and commonly accepted. Pericoli and Sbracia (2003) provide five of the most widely accepted definitions of financial contagion. According to one of their definitions "Contagion is a significant increase in the probability of a crisis in one country, conditional on a crisis occurring in another country.". According to another of their definitions, "Contagion occurs when cross-country comovements of asset prices cannot be explained by fundamentals". Hence, these definitions are consistent with Bae et al. (2003) and Boyson et al. (2010).

7.8% in the probability of one Euro-core country having extreme bottom returns. Given the fact that the baseline predicted probability of one Euro-core country having an extreme bottom return is 6.9%, this marginal effect is very significant both economically and statistically.

Table 2.6 reports the marginal effects of Y^*_{jt} for all five extreme outcomes, for all groups and all periods:

Insert Table 2.6 here

The gray line corresponds to the gray line in Table 2.5. Indeed we see that the statistical significance of Y^*_{jt} is very strong and quite stable throughout the entire period and the three subperiods (Pre-crisis, US-crisis and Euro-crisis). The marginal effects are quite significant, with one more Euro-periphery country having extreme bottom returns increasing by 9.8% the probability of one Non-euro country having extreme bottom returns in the entire period, 10.5% in the Pre-crisis period, 10.5% in the US-crisis period and 6.4% in the Euro-crisis period. In some cases the coefficients intensify during the US-crisis: a 4.8% increase for two (2) Non-euro bottom extreme returns in the Pre-crisis period becomes 6.3% in the US-crisis, an 8.5% increase for the one (1) Euro-core bottom extreme return becomes 11% during the US-crisis et cetera. We notice that the marginal effects are sometimes lower in the Euro-crisis period (6.4% instead of 10.5%, 3.6% instead of 6.3%, 1.3% instead of 2.6% for the Non-euro group for the outcomes of one, two and three bottom extreme returns). But one has to take into account that an extreme return is much more "extreme" during the Euro-crisis period compared to the Pre-crisis period, in terms of the magnitude of expected losses. We discuss this further later in this section.

Table 2.7 provides the summary results for the across groups effects, for the entire period and the three time periods we examine (the Pre-crisis, the US-crisis and the Euro-crisis periods) separately.

Insert Table 2.7 here

The evidence supports the hypothesis that there are important (positive) effects from the Euro-periphery to the other two groups. Extreme bottom (or top) return counts in the Euro-periphery group have a significant (and positive) impact on the extreme return counts of the Euro-core and Non-euro groups in most of the cases. The results are stronger for the entire sample period (which also includes the US-crisis) but the number of statistically significant coefficients does not change between the Pre-crisis and the crisis periods. Counting the number of statistically significant parameters provides an indication of the evolving dynamics of the transmission of extreme shocks from the Euro-periphery to the Non-euro and the Euro-core groups, but, to add rigor on top of this approach, we estimate the parameters on the three subsamples together, and then conduct likelihood ratio tests to see whether separate parameters for the subsamples are needed. The null hypothesis is that models under examination are nested to each other, which means that estimating separate parameters does not create a statistically significant improvement in the fit of the model. The null hypothesis is rejected in a statistical significance level α if:

$$\text{Reject } H_0 \text{ if: P-value} < \alpha \tag{2.9}$$

In case the null hypothesis is rejected, the two models are not nested, which means that statistically significant differences exist between the examined subperiods. The results of the likelihood ratio tests appear in Table 2.8:

Insert Table 2.8 here

The likelihood ratio test results of Tables 2.8 indeed confirm significant statistical differences of the relationships affecting bottom extreme returns count for almost all periods. We are testing all possible four (4) assumptions: 1) that the Pre-crisis, US-crisis and Euro-crisis models are nested in the Entire model, that the Pre-crisis and US-crisis period models are nested in the combined Pre-crisis and US-crisis models, that the US-crisis and Euro-crisis period models are nested in the combined US-crisis and Euro-crisis model, and that the Pre-crisis and Euro-crisis period models are nested in the combined Pre-crisis and Euro-crisis models. Since we are studying the across group effects on two groups (Euro-core and Non-euro), we have in total eight (8) different model specifications to compare. Interestingly, out of these eight (8) specifications, two models appear to not have significantly changed (at the 1% level): those comparing the individual US-crisis and Euro-crisis models to the combined US-crisis and Euro-crisis model. Therefore, the evidence is strong that the relationships affecting bottom extreme returns have indeed changed during the two crises periods compared to the before crisis period, but may have been similar between the two crises. Moreover, the absolute size of the effects is stronger since the same 10% cutoff values are higher during the crisis periods, as Table 2.9 shows:

Insert Table 2.9 here

In other words, even before the two crises there was a significant shock transmission channel from the Euro-periphery to the Euro-core and the Non-euro, but the shocks became deeper in terms of the expected losses for all groups, indicating an intensification of the effects in terms of the actual stock market losses they incur. Studying only the coefficients of the extreme returns during the three periods can be misleading in the sense that the underlying extreme returns are sometimes significant higher in the crisis periods, which means the expected losses are significantly higher, a fact that should not be neglected. This is also verified by the average returns on the days with extreme bottom outcomes which appear on Table 2.10:

Insert Table 2.10 here

Indeed, one can easily compare the Pre-crisis with the crisis periods and see the evident intensification of extreme bottom returns. For example, the average return on days where four (4) Non-euro country had an extreme bottom return (column 4) is -1.694% in the Pre-crisis period, while it decreases to -3.346% in the US-crisis and to -2.565% in the Euro-crisis period. For the outcome where all five (5) Euro-core countries had an extreme bottom return on the same day (column 5) the average return decreased from -1.776% in the Pre-crisis to -4.236% in the US-crisis and to -2.802% in the Euro-crisis period. Thus, an "extreme bottom return" in the Euro-crisis period, is much more intense than an "extreme bottom return" in the Pre-crisis period. Although the percentile does not change (10% of the marginal distribution in both periods), the actual returns themselves are much more negative. This results in higher expected losses for investors in the occurrence of an "extreme event" (which by definition happens in 10% of the days for all countries). Given the fact that stock market capitalizations in these country groups are trillions of Euros, a 1% or 2% increase in extreme bottom returns (in crisis periods) can lead to aggregate losses of tens of billions Euros in one single trading day.

2.3.5. Alternative specification of the cutoffs that define "extreme returns"

An alternative specification of the cutoffs that define "extreme returns" would be to fix the levels of the cutoffs in the pre-crisis period and to apply the same cutoffs in the US-crisis and the Euro-crisis period. The results appear in Table 2.11:

Insert Table 2.11 here

One can see from the results of Table 2.11 that the effects are maximized during the US-crisis period. For example, for the Euro-core, for the outcome of one (1) bottom count, the increase by one (1) in the number of Euro-periphery countries experiencing a bottom extreme return, is associated with an increase of 0.055 (5.5%) in the probability of one (1) Euro-core country having extreme an bottom return on the same day. This probability increases to 0.017 (17%) during the US-crisis, while it decreases to 0.050 (5%) in the Euro-crisis. If one was to compare the coefficients of the Pre-crisis to the ones of the Euro-crisis, there seem to exist some significant differences, especially for the Euro-core group. The coefficients of one (1) and two (2) Euro-core bottom counts have lower values in the Euro-crisis versus their Pre-crisis values; 0.050 versus 0.055 and 0.010 versus 0.014. The coefficient for two bottom counts is only significant in the 5% significance level for the Pre-crisis period, while it becomes significant in the 1% level for the Euro-crisis period), the coefficients for three (3), four (4) and five (5) Euro-core bottom counts are higher than their Pre-crisis values (0.014 versus 0.002, 0.010 versus 0.008 and 0.004 versus 0.000 respectively). Not only are the values higher, but the significance is much higher: The coefficient for the three (3) extreme bottom returns is insignificant in the Pre-crisis period, while it becomes significant in the 1% significance level in the Euro-crisis period. The coefficient for the four (4) extreme bottom returns is significant in the 10% level in the Pre-crisis period, while it becomes significant in the 5% level in the Euro-crisis period. Finally, the coefficient for the five (5) extreme bottom returns is insignificant in the Pre-crisis period, while it becomes significant in the 10% level during the Euro-crisis period.

For a formal way to compare the coefficients, we use a series of t-tests, the results of which appear in Table 2.12:

Insert Table 2.12 here

The table provides a total of thirty (30) coefficients comparison, fifteen (15) for each of the two regions effects (Euro-periphery to Euro-core, Euro-periphery to Non-euro), five for every combination of periods (Pre-crisis/US-crisis, Pre-crisis/Euro-crisis, US-crisis/Euro-crisis, Pre-crisis/US-crisis, Pre-crisis/Euro-crisis, US-crisis/Euro-crisis). Regarding the "Euro-periphery To Euro-core" hypothesis: the null hypothesis that the two coefficients are equal is rejected for all five cases in the Pre-crisis/US-crisis comparison (at least in the 5% significance level), which means that coefficients between these two periods are significantly different from each other. Regarding the Pre-crisis/US-crisis periods comparison, the coefficients are difference for the three bottom extreme returns outcome, while regarding the US-crisis/Euro-crisis period comparison, the coefficients are significantly different in three cases: the one (1), two (2) and four (4) bottom extreme returns. As far as the "Euro-periphery To Non-euro" outcome is concerned, four (4) out of five (5) coefficients are significantly different for the Pre-crisis/US-crisis periods (the coefficients of one, two, three and four bottom extreme returns), while two coefficients are significantly different when comparing the Pre-crisis/Euro-crisis period (three and four bottom extreme returns). Finally, all five (5) coefficients are significantly different for the US-crisis/Euro-crisis periods comparison.

2.4. 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:

$$Ret_t^{group} = c_0 + c_1 Ret_{t-1}^{group} + \epsilon_t \tag{2.10}$$

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

$$\sigma_t^2 = c_2 + c_3\sigma_{t-1}^2 + c_4\epsilon_{t-1}^2 \tag{2.11}$$

The volatilities are then obtained as the estimated $\hat{\sigma_t}$ from the AR(1)-GARCH(1,1) model. We notice that for the extreme returns counts filtered by a GARCH the effect of volatility is not significant (for the raw returns all volatility coefficients were found to be positive and statistically significant - in other words an increase in volatility increases the probability of extreme bottom returns).
As a final robustness check, we re-estimated the models for the Entire period, using one cutoff, instead of separate cutoffs for the subperiods.

The results are found to be robust when compared with the results of the three cutoffs previously examined.[17]

2.5. Conclusion

We examine the transmission of financial shocks among three groups of countries: the Euro-periphery countries (Portugal, Ireland, Italy, Greece, Spain), the Euro-core countries (Germany, France, the Netherlands, Finland, Belgium), and the major European Union -but not euro- countries (Sweden, UK, Poland, Czech Republic, Denmark), using daily stock market data from January 2004 until March 2013. The entire period is further split in three sub periods, the Pre-crisis period (1/1/2004-26/2/2007), the US-crisis period (27/2/2007-7/12/2009) and the Euro-crisis period (8/12/2009-13/3/2013). The creation of the three groups is justified by the existence of the European Union (which is mainly a free-trade area), and the Eurozone (monetary union). The five Euro-periphery countries were the most badly hit during the recent crises periods. Our analysis is split in two parts: the first part concerns extreme stock index returns, controlling for various fundamentals derived from financial market data (volatility, exchange rate change, short interest rates, long term spread change). We find that even before the two crises periods there was a significant shock transmission channel from the Euro-periphery to the Euro-core and the Non-euro groups. During the crises the shocks transmitted were more substantial. Thus, expected losses from extreme returns have increased in the crises periods, being evidence of an intensification of the effects during the recent financial crises. The fact that indeed the models in the different periods exhibit differences is verified using likelihood ratio tests.

The implications of the overall findings are quite significant for investors who may want to diversify their portfolios, and they should be aware of the stock indices movement dynamics and of how extreme shocks propagate from one group of countries to the others, thus affecting their portfolios' overall risk profiles. Furthermore, the findings would be useful for policy makers in order to assess policy decision making in times of extreme shocks (such as crisis periods). The fact that the European financial markets affect one another provides evidence that a failure to properly confront the crisis could see the propagation and clustering of extreme returns, leading to

[17] The results are available upon request.

significant losses for investors and institutions. Future research could also take into account different models or move in the direction of studying higher frequency (intraday) financial markets dynamics.

Table 2.1: Descriptive Statistics and Correlations for Stock Indices

Panel A: 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	0.030
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	2.790	2.858	8.689	7.838	8.584	5.321	9.118	6.848

Panel B: Correlations

	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
Non-euro	1.000			1.000			1.000		
Euro-periphery	0.791	1.000		0.919	1.000		0.836	1.000	
Euro-core	0.806	0.870	1.000	0.912	0.928	1.000	0.926	0.876	1.000

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.

Table 2.2: Counts of extreme bottom (and top) log returns for daily country group stock indices, January 1st 2004 to March 13th 2013.

	Mean return (%) when $i = 5$	Number of bottom counts						Number of top counts						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 counts are defined as the joint occurrence of extreme returns across different country indexes on the same day. For example, out of a total sample of 2399 trading days, there are 104 days where exactly two Euro-periphery countries had extreme bottom returns on the same day, and in 23 of those days Greece is the one of the two countries having extreme bottom returns.

Table 2.3: Within the Euro-periphery group bottom extreme counts of log returns for the entire period. The bottom extreme counts for the entire period are calculated as the sum of the bottom extreme counts for the three subperiods. All reported coefficients are marginal effects.

	(1)	(2)	(3)	(4)	(5)
	Margin / SE	Margin / SE	Margin / SE	Margin / SE	Margin / SE
Constant	-0.157^{***}	-0.103^{***}	-0.085^{***}	-0.067^{***}	-0.066^{***}
	(0.017)	(0.011)	(0.009)	(0.009)	(0.009)
Volatility	0.014	0.006	0.008^{*}	0.015^{***}	0.012^{***}
	(0.013)	(0.007)	(0.005)	(0.003)	(0.003)
Exchange Rate Change	0.043^{***}	0.024^{***}	0.010^{**}	0.012^{***}	0.009^{***}
	(0.011)	(0.006)	(0.004)	(0.003)	(0.003)
ST Interest Rate	-0.009^{*}	0.001	0.002	-0.001	0.004^{***}
	(0.005)	(0.003)	(0.002)	(0.002)	(0.001)
LT Spread Change	0.065	0.147^{***}	0.132^{***}	0.075^{***}	0.076^{***}
	(0.072)	(0.034)	(0.023)	(0.018)	(0.016)
Observations	2399	2399	2399	2399	2399
Baseline predicted probability	0.123	0.043	0.029	0.024	0.022
$Pseudo - R^2$	0.063				

Note: Columns (1) to (5) correspond to bottom extreme counts 1 to 5. In other words, column (1) presents the marginal effects in the case of one bottom count for the Euro-periphery group, and columns (2),(3),(4),(5) correspond to two, three, four and five bottom counts for this group. The value of 0.147 for the Euro-periphery LT spread changes (column 2) means that an increase of 1 percent in the average Euro-periphery long term spread (vis-à-vis Germany) increases the probability of two Euro-periphery countries having extreme bottom stock returns by 14.7%, while the value of 0.043 for the average exchange rate change (column 1) means that a one percent increase in the average Euro-periphery exchange rate increases the probability of one bottom Euro-periphery extreme return by 4.3%.

(***) : significance at 1% level

(**) : significance at 5% level

(*) : significance at 10% level

Table 2.4: Within-groups summary results for bottom and top extreme return counts.

	Bottom tail			Top tail		
Entire Period						
	Non-euro	Euro-periphery	Euro-core	Non-euro	Euro-periphery	Euro-core
Volatility	+ + ++	+ + +	+ + +	+ + ++	+ + ++	+ + ++
Exchange Rate Ch.	+ + + + +	+ + + + +	+ + +	− − − − −	− − − − −	− − −−
ST Interest Rate		−+	+	−−	−	−−
LT Spread Change	+ + + + +	+ + ++	+ + +	− − − − −	− − − − −	− − −
$Pseudo - R^2$	0.081	0.063	0.052	0.086	0.049	0.052
Pre-crisis Period						
	Non-euro	Euro-periphery	Euro-core	Non-euro	Euro-periphery	Euro-core
Volatility	+ + +	+ + +	++	+ + +	+	++
Exchange Rate Ch.		−	−		+	+
ST Interest Rate	−−			−	+	
LT Spread Change	+	+		−		
$Pseudo - R^2$	0.040	0.039	0.036	0.044	0.016	0.038
US-crisis Period						
	Non-euro	Euro-periphery	Euro-core	Non-euro	Euro-periphery	Euro-core
Volatility	+ + +	+ + ++	+ + ++	+ + ++	+ + ++	+ + + + +
Exchange Rate Ch.	+ + + + +	++	+	− − −−	−	−−
ST Interest Rate				−−	−	−−
LT Spread Change	+	++		−	−−	
$Pseudo - R^2$	0.122	0.085	0.084	0.137	0.084	0.079
Euro-crisis Period						
	Non-euro	Euro-periphery	Euro-core	Non-euro	Euro-periphery	Euro-core
Volatility	+	++	++	+ + ++	−	+ + ++
Exchange Rate Ch.	+ + ++	+ + ++	+ + +	− − −−	− − −−	−−
ST Interest Rate	+					
LT Spread Change	+	+ + +	+ + ++	− − −−	−	− − −−
$Pseudo - R^2$	0.194	0.176	0.181	0.172	0.119	0.167

Note: The number of "+" (or "-") indicate the number of statistically significant (in the 1% or 5% levels) and positive (or negative) marginal effects. For example, for the bottom tail returns and the entire period sample, all five volatility marginal effects are significant and positive for the Non-euro group (this is why we have five plus symbols at the Non-euro column), indicating that an increase in volatility increases the probability of extreme bottom returns in all five bottom extreme outcomes. For the top tail returns, the number of statistically significant marginal effects are five for the average exchange rate change in the Non-euro group, meaning that an increase in the average group's exchange rates (i.e. weaker group currencies on average) leads to lower probabilities of top Non-euro counts (for all five possible outcomes).

Table 2.5: Across-groups effects: Euro-periphery to Euro-core bottom extreme returns for the entire period. All reported coefficients are marginal effects.

	(1)	(2)	(3)	(4)	(5)
	Margin / SE	Margin / SE	Margin / SE	Margin / SE	Margin / SE
Constant	-0.228^{***}	-0.121^{***}	-0.072^{***}	-0.040^{***}	-0.014^{***}
	(0.017)	(0.013)	(0.011)	(0.009)	(0.005)
Volatility	0.007	0.007	0.007^{**}	0.003	0.002^{**}
	(0.012)	(0.007)	(0.003)	(0.002)	(0.001)
Exchange Rate Change	-0.006	-0.002	0.002	-0.001	0.000
	(0.010)	(0.006)	(0.003)	(0.002)	(0.000)
ST Interest Rate	0.009^{**}	0.002	0.001	-0.000	-0.000
	(0.004)	(0.003)	(0.001)	(0.001)	(0.000)
LT Spread Change	0.847^{***}	0.115	0.044	0.056	0.013
	(0.285)	(0.157)	(0.077)	(0.044)	(0.011)
Bottom Count Euro-periphery	0.078^{***}	0.041^{***}	0.024^{***}	0.014^{***}	0.005^{***}
	(0.009)	(0.005)	(0.004)	(0.003)	(0.002)
Observations	2399	2399	2399	2399	2399
Baseline predicted probability	0.069	0.030	0.023	0.020	0.045
$Pseudo-R^2$	0.318				

Note: Columns (1) to (5) correspond to bottom counts 1 to 5. In other words, column (1) presents the marginal effects in the case of one bottom count for the Euro-periphery group, and columns (2),(3),(4),(5) correspond to two, three, four and five bottom counts for this group. The value of 0.78 for the Euro-periphery bottom count (column 1) means that an increase of 1 Euro-periphery countries having extreme bottom returns increases the probability of one Euro-core country having extreme bottom stock returns (i.e. one bottom Euro-core count) by 7.8%. The gray line corresponds to the gray line in Table 2.6.

(***) : significance at 1% level

(**) : significance at 5% level

(*) : significance at 10% level

Table 2.6: Across groups effects: Euro-periphery to all groups for all periods. All reported coefficients are marginal effects.

	(1) Margin / SE	(2) Margin / SE	(3) Margin / SE	(4) Margin / SE	(5) Margin / SE	$Pseudo-R^2$
Entire Period						
To Non-euro (Bottom)	0.098***	0.052***	0.024***	0.005***	0.000	0.258
To Euro-core (Bottom)	0.078***	0.041***	0.024***	0.014***	0.005***	0.318
To Non-euro (Top)	0.099***	0.060***	0.021***	0.004***	0.000	0.218
To Euro-core (Top)	0.094***	0.045***	0.025***	0.018***	0.007***	0.345
Pre-crisis Period						
To Non-euro (Bottom)	0.105***	0.048***	0.027***	0.005*	0.000	0.223
To Euro-core (Bottom)	0.085***	0.044***	0.027***	0.009**	0.003	0.281
To Non-euro (Top)	0.104***	0.055***	0.020***	0.001	0.000	0.159
To Euro-core (Top)	0.092***	0.041***	0.025***	0.016***	0.002	0.246
US-crisis Period						
To Non-euro (Bottom)	0.105***	0.063***	0.026***	0.004	0.000***	0.341
To Euro-core (Bottom)	0.110***	0.036***	0.020**	0.013**	0.001	0.383
To Non-euro (Top)	0.092***	0.068***	0.018***	0.008**	0.000	0.316
To Euro-core (Top)	0.102***	0.043***	0.020***	0.014**	0.004	0.345
Euro-crisis Period						
To Non-euro (Bottom)	0.064***	0.036***	0.013***	0.002	0.000	0.290
To Euro-core (Bottom)	0.044***	0.022***	0.014**	0.014**	0.004	0.373
To Non-euro (Top)	0.042**	0.046***	0.011***	0.002	0.000	0.262
To Euro-core (Top)	0.063***	0.031***	0.016***	0.013**	0.003*	0.343

Note: Columns (1) to (5) correspond to bottom counts 1 to 5. In other words, column (1) presents the marginal effects in the case of one bottom count for the Euro-periphery group, and columns (2),(3),(4),(5) correspond to two, three, four and five bottom counts for this group. The value of 0.078 for the Euro-periphery bottom count (column 1) means that an increase of 1 in the number of Euro-periphery countries having extreme bottom returns increases the probability of one Euro-core country having extreme bottom stock returns (i.e. one bottom Euro-core count) by 7.8% for the Entire period, 8.5% in the Pre-crisis period, 11.0% in the US-crisis period and 4.4% in the Euro-crisis period. The gray line corresponds to the gray line in Table 2.5.

(***) : significance at 1% level

(**) : significance at 5% level

(*) : significance at 10% level

Table 2.7: Across groups summary results for bottom and top extreme counts.

	Bottom tail		Top tail	
Entire Period				
	Non-euro	Euro-core	Non-euro	Euro-core
(from Euro-periphery)				
Volatility	+ + +	++	+ + +	+ + + + +
Exchange Rate Change	+ + ++		− − −−	
ST Interest Rate		+	+ + ++	+
LT Spread Change	+	+	− − −	−
Euro-periphery Bottom Extreme Count	+ + + + +	+ + + + +	+ + ++	+ + + + +
$Pseudo-R^2$	0.257	0.318	0.218	0.291
Pre-crisis Period				
	Non-euro	Euro-core	Non-euro	Euro-core
(from Euro-periphery)				
Volatility			++	++
Exchange Rate Change		−		+
ST Interest Rate			−	
LT Spread Change			−	
Euro-periphery Bottom Extreme Count	+ + +	+ + ++	+ + +	+ + ++
$Pseudo-R^2$	0.223	0.281	0.159	0.245
US-crisis Period				
	Non-euro	Euro-core	Non-euro	Euro-core
(from Euro-periphery)				
Volatility	++	++	+ + +	+ + ++
Exchange Rate Change	+ + +		− − −	−
ST Interest Rate			−	−
LT Spread Change	+			
Euro-periphery Bottom Extreme Count	+ + ++	+ + ++	+ + ++	+ + ++
$Pseudo-R^2$	0.341	0.384		0.345
Euro-crisis Period				
	Non-euro	Euro-core	Non-euro	Euro-core
(from Euro-periphery)				
Volatility			+ + +	+ + +
Exchange Rate Change	+ + +		− − −	
ST Interest Rate		+		
LT Spread Change		+	−	−−
Euro-periphery Bottom Extreme Count	+ + +	+ + ++	+ + +	+ + ++
$Pseudo-R^2$	0.290	0.373	0.262	0.342

Note: The number of "+" (or "-") indicate the number of statistically significant (in the 1% or 5% levels) and positive (or negative) marginal effects. For example, for the bottom tail returns and the entire period sample, three out of five volatility marginal effects are significant and positive for the Non-euro group, indicating that an increase in volatility increases the probability of bottom Non-euro extreme counts in three out of five outcomes. For the top tail returns, the number of statistically significant marginal effects are four for the average exchange rate change, and have a negative sign in all four cases, meaning that an increase in the average group's exchange rates (i.e. weaker group currencies on average) lead to lower probabilities of extreme top returns in four out of five top Non-euro outcomes.

Table 2.8: Likelihood ratio tests.

Euro-periphery to Euro-core				
Model		Obs	df	Prob > chi2
Nested	Entire period	2399	30	0.000
Non-nested	Pre-crisis	822	30	
	US-crisis	725	30	
	Euro-crisis	852	30	
Nested	Pre-crisis, US-crisis	1547	30	0.0077
Non-nested	Pre-crisis	822	30	
	US-crisis	725	30	
Nested	US-crisis, Euro-crisis	1577	30	0.0128
Non-nested	US-crisis	725	30	
	Euro-crisis	852	30	
Nested	Pre-crisis, Euro-crisis	1674	30	0.0077
Non-nested	Pre-crisis	822	30	
	Euro-crisis	852	30	
Euro-periphery to Non-euro				
Model		Obs	df	Prob > chi2
Nested	Entire period	2399	30	0.000
Non-nested	Pre-crisis	822	30	
	US-crisis	725	30	
	Euro-crisis	852	30	
Nested	Pre-crisis, US-crisis	1547	30	0.000
Non-nested	Pre-crisis	822	30	
	US-crisis	725	30	
Nested	US-crisis, Euro-crisis	725	30	1.0000
Non-nested	US-crisis	725	30	
	Euro-crisis	852	30	
Nested	Pre-crisis, Euro-crisis	1674	30	0.0010
Non-nested	Pre-crisis	822	30	
	Euro-crisis	852	30	

Note: Likelihood ratio tests for the Entire Period, the Pre-crisis Period, the US-crisis Period and the Euro-crisis Period models. Obs denotes the number of observations for each model, df is the number of degrees of freedom.

Table 2.9: 10% percentiles for the extreme bottom returns (%) of the three country-groups for all subperiods.

	Non-euro	Euro-periphery	Euro-core
Pre-crisis	-0.957	-0.742	-0.831
US-crisis	-2.037	-2.060	-2.011
Euro-crisis	-1.262	-1.903	-1.450

Table 2.10: Average returns (%) on days with extreme bottom outcomes, for all subperiods.

	(1)	(2)	(3)	(4)	(5)
Pre-crisis					
Non-euro	-1.341	-1.514	-1.501	-1.694	-2.624
Euro-periphery	-1.076	-1.108	-1.128	-1.512	-1.843
Euro-core	-1.154	-1.179	-1.141	-1.214	-1.776
US-crisis					
Non-euro	-2.570	-2.829	-3.222	-3.346	-5.156
Euro-periphery	-2.691	-2.978	-2.912	-3.484	-4.238
Euro-core	-2.732	-2.709	-2.734	-2.819	-4.236
Euro-crisis					
Non-euro	-1.606	-1.793	-1.894	-2.565	-2.852
Euro-periphery	-2.759	-2.574	-2.718	-2.888	-3.749
Euro-core	-1.848	-1.689	-1.759	-2.162	-2.802

Note: Columns one (1) to five (5) correspond to the count of extreme bottom returns. For example, column one (1) corresponds to one country in the group having an extreme bottom return on this day, while column five (5) corresponds to all five (5) countries in the group having extreme bottom returns on the same day.

Table 2.11: Across-groups effects summary results: Euro-periphery to Non-euro and Euro-core for the Pre-crisis, the US-crisis and the Euro-crisis period. In this case, the "extreme returns" are defined by fixing the levels of the cutoffs in the Pre-crisis period, and applying the same cutoffs to all periods. All reported coefficients are marginal effects.

	(1)	(2)	(3)	(4)	(5)
	Margin / SE	Margin / SE	Margin / SE	Margin / SE	Margin / SE
To Euro-core					
Pre-crisis	0.055***	0.014**	0.002	0.008*	0.000
	(0.014)	(0.006)	(0.002)	(0.005)	(0.000)
US-crisis	0.170***	0.090***	0.030***	0.039***	0.012**
	(0.028)	(0.017)	(0.010)	(0.012)	(0.006)
Euro-crisis	0.050***	0.010***	0.014***	0.010**	0.004*
	(0.011)	(0.004)	(0.005)	(0.004)	(0.002)
To Non-euro					
Pre-crisis	0.052***	0.020***	0.026***	0.006	0.000
	(0.012)	(0.007)	(0.007)	(0.004)	(0.000)
US-crisis	0.094***	0.072***	0.060***	0.024***	0.008**
	(0.016)	(0.012)	(0.011)	(0.007)	(0.004)
Euro-crisis	0.034***	0.023***	0.009***	0.001	0.000
	(0.007)	(0.005)	(0.003)	(0.001)	(0.000)

Note: Columns (1) to (5) correspond to bottom counts 1 to 5. In other words, column (1) presents the marginal effects in the case of one bottom count for the Euro-periphery group, and columns (2),(3),(4),(5) correspond to two, three, four and five bottom counts for this group. The value of 0.055 for the Euro-periphery bottom count (column 1) means that an increase of 1 in the number of Euro-periphery countries having extreme bottom returns increases the probability of one Euro-core country having extreme bottom stock returns (i.e. one bottom Euro-core count) by 5.5% in the Pre-crisis period.

(***) : significance at 1% level

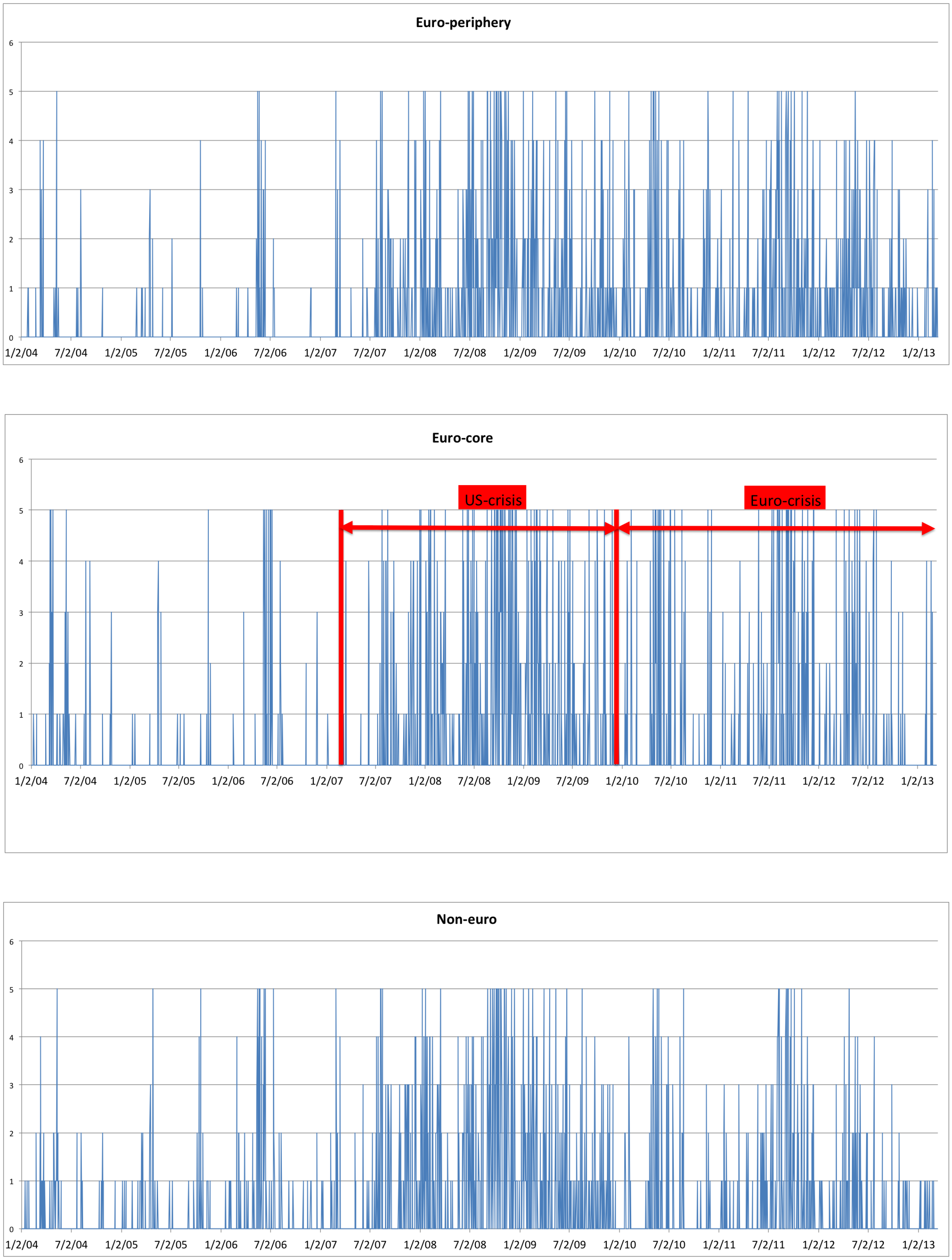
(**) : significance at 5% level

(*) : significance at 10% level

Table 2.12: T-tests for coefficients equality, for three periods.

	(1) p-value	(2) p-value	(3) p-value	(4) p-value	(5) p-value
Euro-periphery To Euro-core					
Pre-crisis/US-crisis	0.0002	0.0001	0.0062	0.0404	0.0446
Pre-crisis/Euro-crisis	0.8766	0.5512	0.0181	0.9549	0.1080
US-crisis/Euro-crisis	0.0001	0.0000	0.1461	0.0384	0.2132
Euro-periphery To Non-euro					
Pre-crisis/US-crisis	0.0370	0.0004	0.0114	0.0592	0.1726
Pre-crisis/Euro-crisis	0.3395	0.4863	0.0267	0.0783	0.3351
US-crisis/Euro-crisis	0.0013	0.0009	0.0000	0.0025	0.0540

Fig. 2.1. Extreme bottom return counts for all three country groups, for the entire period.



Chapter 3

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

We attempt to seek a connection between three phenomena: Extreme Stock Market Returns, the Web Attention factor and a set of News Flow factors. We do this 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 until 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. Higher levels of 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.

3.1. Introduction

One topic that has dominated the financial press over the past few years has undoubtedly been the Euro-crisis. A major feature 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 avoid default, on the state of their finances, their economic uncompetitiveness, whether they should leave the common currency area, and whether bad news about them might propagate and affect other 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. Dummy variables were used to denote the occurrence of events, or arbitrarily defining events and news as "good" or "bad" (e.g. Arezki et al. (2011), Beetsma et al. (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 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 Euro-periphery financial sentiment does indeed significantly affect the other countries financial asset returns. This after all is the main concern of financial discussions and policy making since the advent of the Euro-crisis[1].

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

[1]The topic of this paper is the European Financial Crisis. The countries with the biggest financial difficulties during the Euro-crisis were the Euro-periphery countries. The policy actions/news during this period, were about Europe. Studying the impact of non European markets and news would be of interest for another research paper.

extreme bottom returns for all three country groups.

The rest of the paper is organized as follows. Section 3.2 presents the related textual analysis and Web Attention literature. Section 3.3 presents the data and the model specification. The empirical findings are shown in Section 3.4. Finally, Section 3.5 provides a set of robustness and alternative specifications and Section 3.6 concludes.

3.2. 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[2] and Thomson Reuters News Analytics[3]). 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 et al. (2013)). The related literature also studies the effect of returns on media content Garcia (2014), 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 et al. (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 et al. (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

[2] `http://www.ravenpack.com/`
[3] `http://www.machinereadablenews.com/`

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[4]. Da et al. (2011) find that a higher SVI for stocks in Russell 3000 does forecast higher returns in the next two weeks, an effect which reverses within one year. Da et al. (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 et al. (2003) and Boyson et al. (2010) who examine the coincidence of extreme return shocks across groups of countries. A number of authors have used this methodology[5]. 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 Garcia (2014) as far as the analysis of News Flow is concerned and to Da et al. (2011) and Da et al. (2015) regarding 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 the crisis transmission from the Euro-periphery group to the other two groups (Euro-periphery vs. Euro-core, Euro-periphery vs. Non-euro).

3.3. The Data

The main area of study for this paper is the European Union, and 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,

[4]Google Trends can be found at: `http://www.google.com/trends/explore#cmpt=q`
[5]See, for example, Christiansen and Ranaldo (2009), Gropp et al. (2009), ?.

UK, Denmark)[6]. The Euro-periphery group consists of the five Eurozone countries that were most severely hit during the Euro-crisis. The Euro-core group consists of the biggest countries (in terms of market capitalization) 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 free-trading 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 until 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 into three subperiods (Pre-crisis, US-crisis and Euro-crisis) to be able to make comparisons between normal and abnormal times in the financial markets:

- the Pre-crisis period (from 1 January 2004 until 26 February 2007)
- the US-crisis period (from 27 February 2007 until 7 December 2009).[7]
- the Euro-crisis period (from 8 December 2009[8][9] until 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, Greek debt was downgraded by Fitch from A- to BBB+, with a negative outlook.

3.3.1. 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 analyze from Dow Jones Factiva[10] news articles covering the test period from January 1st, 2004 until 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

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

[7] 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`.

[8] "8 December 2009: Fitch downgrades Greece's credit rating The crisis escalates and shares fall around the world after ratings agency Fitch cuts Greece's long-term debt to BBB+, from A-. This is the first time in a decade that Greece does not have an A-rating, and pushes up the cost of borrowing." `http://www.theguardian.com/business/2010/may/05/greece-debt-crisis-timeline`

[9] "Financial markets tumble after Fitch downgrades Greece's credit rating" `http://www.theguardian.com/world/2009/dec/08/greece-credit-rating-lowest-eurozone`

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

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 includes the name of the country plus one of the following terms each time: crisis, debt, economy, deficit, default. For example, for Greece the news is 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 rose at various times points during the crisis. These five search terms were also found to be the most relevant key words used in Google searches.

3.3.1.1. Preprocessing the news data

These selection criteria result in a total of 110,800 news articles[11]. 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 items dominates the news sample, simply because they are 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 in 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

[11]We focus on news articles which are written in English. This is due to the fact that the finance literature has developed finance specific dictionaries in English (Loughran and McDonald (2011)). Important news may be written in local languages. Nevertheless, because of the dominant position of English in the finance world, these news would be rapidly transmitted through the international newswires and newspapers, which possess a global network of journalists and translators, who are able to disseminate information globally in a very rapid fashion.

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 opened 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.

3.3.1.2. 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[12]. 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 3.1, having a total of 24,402 news items:

Insert Table 3.1 here

We see from Table 3.1 that Pre-crisis most news are about Italy and Spain, but in the Euro-crisis period most news concerns Greece. Most news overall is 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.

3.3.1.3. Textual Analysis and Web Attention (SVI)

We use textual analysis, based on the Loughran and McDonald (2011) dictionaries[13]. These dicationaries have proved to be very well specified and robust for the finance domain, and have been used in a series of recent papers (Chen et al. (2013), Garcia (2014), Garcia (2013), Loughran and McDonald (2013), Jegadeesh and Wu (2013), Ahern and Sosyura (2014), Dougal et al. (2012), Kelley and Tetlock (2013)). We measure the positive media

[12] An alternative way might be to take the lead sentence(s) which could convey the central message of the article.

[13] These dictionaries can be found at `http://www3.nd.edu/~mcdonald/Word_Lists.html`

content as: $G_t = \sum_i \frac{g_{it}}{w_{it}}$, 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 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 (these metrics of "positive" and "negative" media content has also been used in Loughran and McDonald (2011), Chen et al. (2013), Garcia (2014), Garcia (2013), Loughran and McDonald (2013), Jegadeesh and Wu (2013), Ahern and Sosyura (2014), Dougal et al. (2012), Kelley and Tetlock (2013), Tetlock (2007), Tetlock et al. (2008)). Thus, we obtain the *Pessimism* (as in Dougal et al. (2012), Garcia (2014), and Garcia (2013)) on day t as the difference between the negative and the positive media measures:

$$Pessimism_t = B_t - G_t \tag{3.1}$$

The *Pessimism* measure is calculated for every Euro-periphery country, for every day, and gets one value for each of these countries, for every day. Then, the Euro-periphery *Pessimism* (P_t) is calculated as the average of the Euro-periphery pessimism factors on every day:

$$P_t = \frac{\sum_{j=1}^{5} Pessimism_{j,t}}{5} \tag{3.2}$$

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). The reason this metric is divided by 5, is that in the Euro-periphery there are five countries, so we need to divide the sum by 5 in order to get the average pessimism.

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:

$$WP_t = \frac{\sum_{j=1}^{5} Pessimism_{j,t} N_{j,t}}{\sum_{j=1}^{5} N_{j,t}} \tag{3.3}$$

where $N_{j,t}$ stands for the number of relevant news (on day t) for country j. In this case, we multiply each pessimism with the respective number of news per Euro-periphery country, and then we divide by the sum of articles, as the definition of the weighted average requests.

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.

In order to see how the algorithm works, some selected articles appear in Tables 3.2 and 3.3:

Insert Tables 3.2 and 3.3 here

Due to space constraints, we show a couple of news articles from two (2) Euro-periphery countries (Greece and Spain). For each country, 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 (Greek, Spain, Spanish), as well as one or more of the crisis keywords in its content (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 we selected are among the most pessimistic articles for every country. Thus, it is natural that most of the painted words are red. 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[14]. Thus, it is 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 provides a direct measure of Web Attention. Especially in crisis times, one could argue that the SVI can capture the uncertainty and the interest about topics and issues that trouble the markets and 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:

$$\textit{Web Search Volume Index: } SVI_{jt} = k, k = 0, ..., 100 \tag{3.4}$$

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

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

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[15]. 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[16,17].

Table 3.4 summarizes the News Flow and Web Attention (SVI) variable definitions:

Insert Table 3.4 here

The graphical illustrations for the SVI for Greece appears in Figure 3.1:

Insert Figure 3.1 here

One can clearly see that there are time periods where Web Attention spikes, in other words periods where people were searching a lot on 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 3.5:

Insert Table 3.5 here

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.

[15] 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.

[16] 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.

[17] 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.

3.3.2. 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 3.6 shows the summary statistics of the percentage (%) log-returns of the major stock market country indices[18].

Insert Table 3.6 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 have rather low values for all country groups as this was a period of relative calm 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 better 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 3.7.

Insert Table 3.7 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 the stock markets are negative during the Euro-crisis period, which indicates that higher news flow values during the Euro-crisis period

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

were associated with negative stock market returns for all three country groups[19]. Naturally, the information variables are quite significantly (positively) correlated among themselves.

3.3.2.1. Extreme Returns

We follow Bae et al. (2003) and define an extreme return as one that lies below (or above) a certain threshold of the marginal return distribution. This approach counts extreme returns within a group on a given day. The original study counts extreme returns for the entire period, using the 5th and the 95th percentiles as cutoffs. 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). Returns below the 10th percentile are defined as extreme bottom returns, while returns above the 90th percentile are defined as extreme top returns. This approach is applied to every country, for every group.

Bottom and top extreme returns counts for the entire period (1/1/2004-13/3/2013) are reported in Table 3.8. For each country we calculate the days for which it had an extreme bottom or top return separately. The extreme returns count for each group and day is calculated as the number of countries of the group that have extreme returns on that specific day[20].

Insert Table 3.8 here

Extreme bottom return counts are presented in the left side of Table 3.8, while top return counts are presented in the right side. A count equal to i for bottom returns denotes that i countries in the group had an extreme bottom return on this day.

Greece had the most days (106) on which it was the only Euro-periphery country having a bottom extreme return, followed by Ireland (56 days) and Portugal (37 days). All five Euro-periphery countries had an extreme bottom return in 54 days. This number was 109 days for the Euro-core and 55 days for the Non-euro groups. All five Euro-periphery countries had an extreme top return on 40 days (91 days for the Euro-core, 28 days for the Non-euro). As far as the Non-euro group is concerned, the Czech Republic was the country with the most days (84) being the only country with an extreme bottom return, and once more was the country with the most days (95) with the most extreme top returns days for the Non-euro group.

[19]Correlations with the extreme bottom returns (see Section 4.3) are significant and much higher than these of raw returns.

[20]Extreme return counts are arguably alternate measures of volatility. Crises are mainly reflected on extreme negative returns. We also take into account the extreme positive returns, for comparison purposes. The main emphasis is on extreme negative returns. One can calculate volatility using other methods as well, such as the standard deviation of high-frequency (intraday) returns, as in Chouliaras (2015c)

3.3.2.2. 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 3.2[21].

Insert Figure 3.2 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.

As in Bae et al. (2003) and Boyson et al. (2010), we create a polychotomous variable in order to capture the effects of the information variables on the probability of extreme returns of all three country-groups. In multinomial logistic regression models, if P_i is the probability of an event i out of m possible events, a multinomial distribution is defined as:

$$P_i = P(Y_t = i|x_j) = \frac{G(\beta_i' x_j)}{1 + \sum_{j=1}^{m-1} G(\beta_j' x_j)}, \tag{3.5}$$

where x is the vector of covariates and β_i the vector of coefficients associated with the covariates. The function $G(\beta_i' x)$ can take the form of an exponential function $exp(\beta_i' x)$, in which case Equation 3.5 represents a multinomial logistic (or multinomial logit) model.[22] To capture the range of possible outcomes 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. 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. The probability of an extreme return count of a specific level, P_i, is calculated by evaluating the covariates at their unconditional values,

$$P_{ij}^* = \frac{exp(\beta_i' x_j^*)}{1 + \sum_{j=1}^{m-1} exp(\beta_j' x_j^*)}, \tag{3.8}$$

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

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

[22] The estimation of such a model is done using maximum likelihood, defining a log-likelihood function for a sample of n observations as:

$$logL = \sum_{i=1}^{n} \sum_{j=1}^{m} I_{ij} logP_{ij}, \tag{3.6}$$

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:

$$pseudoR^2 = 1 - [logL_\omega / logL_\Omega]. \tag{3.7}$$

extreme return). As mentioned in Greene (2003), the coefficients of such a model are not easy to interpret.[23] The coefficients correspond to probabilities. The partial effects give us the marginal change in probability for a unit change in the independent covariate. We are interested in seeing whether these marginal effects are statistically significant or not.[24]

Y_t 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. In Equation 3.8, P_i is equal to $P(Y_t = i|x_t)$ where $Y_t = 0, 1, 2, ...k$ is the extreme returns count variable for the Non-euro, Euro-periphery and Euro-core respectively. We have k=5 for all three country groups, where x_t is the explanatory variable (covariate), on day t. In Equation 3.8, 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.

3.4. Empirical Findings

3.4.1. News Flow, Web Attention and Extreme Returns

In equation 3.5, 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 3.5 has the following form for $g_i(x_t)$:

$$g_i(x_t) = b_{0i} + b_{1i} X_{it} \tag{3.11}$$

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

[23] This is why in these models it is necessary to differentiate 3.5 in order to obtain the partial effects of the covariates on the probabilities

$$\delta_{ij} = \frac{\delta P_{ij}}{\delta \beta_i} = P_{ij}[x_j - \sum_{k=0}^{J} P_{ik}\beta_k] = P_{ij}[\beta_j - \bar{\beta}] \tag{3.9}$$

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

[24] 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}. 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 base outcome is 0 which stands for no extreme returns in the group. The coefficient for covariate x_{13} for outcome 3, which is β_{13} is the coefficient for the 1st covariate, calculated for the 3rd outcome, measures the probability of having the outcome 3 (3 extreme returns in the group), instead of the outcome 0 (no extreme returns in the group), for a unit change in the covariate x_{13}. In reality there is also the possibility of having the outcome 2 instead of 0 for a unit change in covariate x_{13}. Marginal effects calculate the probabilities associated with a covariate change of one in adjacent categories, not taking as an alternative only the base outcome (0 in our study). The coefficients of a multinomial logistic regression model exhibit what is known as the "log odds ratio" property:

$$\ln \frac{P_{ij}}{P_{i0}} = \beta_i' x_j \tag{3.10}$$

returns during the Euro-crisis appear in Table 3.9.

Insert Table 3.9 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)[25]. 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 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 Euro-periphery Weighted Pessimism increases the probability of this outcome by 1.4%. For the Non-euro group, an increase of one in 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 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 3.10.

Insert Table 3.10 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 3.10, 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.

[25] We are aware of the possible shortcomings of the $pseudo - R^2$ because of concerns raised in `http://www.ats.ucla.edu/stat/mult_pkg/faq/general/Psuedo_RSquareds.htm`. This is why the analysis that follows is based on the statistical significance and the magnitude of the marginal effects of the multinomial logistic regression models.

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. We provide evidence that financial crises in a region can affect other Eurozone and European Union equity markets. For diversification purposes, investors should diversify internationally and not only within the Eurozone and the European Union countries. Regulators must be aware that a crisis in a group of countries can have a serious impact on the Eurozone and the European Union. 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.

3.5. 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 an AR(1)-GARCH(1,1) model for each group's average returns:

$$Ret_t^{group} = c_0 + c_1 Ret_{t-1}^{group} + \epsilon_t \tag{3.12}$$

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

$$\sigma_t^2 = c_2 + c_3\sigma_{t-1}^2 + c_4\epsilon_{t-1}^2 \tag{3.13}$$

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.

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. Finally, a Granger causality analysis shows mainly a two-way relationship between News Flow variables and stock returns, during the Euro-crisis (consistent with the evidence of Tetlock (2007), Garcia (2013) and Garcia (2014)).

3.6. Conclusion

We use daily stock market data from January 2004 until 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 exposure. 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.

Table 3.1: 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 its 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 3.2: 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 color) and Positive (green color) word lists of Loughran and McDonald (2011).

Country	**Source**	**Date**	**Title**	**Text**
Greece	Dow Jones	20 July 2011	Fitch: Disorderly Greek Default Would Create Severe Volatility DJ Fitch: Material Threat Of Contagion From Disorderly Greek Default	Fitch: Disorderly Greek 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 Greek default "is material," Fitch Ratings said Wednesday. A disorderly Greek default 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 Greek default	ECB's Stark warns against Greek default 161 words 20 October 2011 11:16 Reuters News LBA English (c) 2011 Reuters Limited FRANKFURT, 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 3.3: Examples of selected stories for Spain. 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 Spain 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 Spanish economy 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-Spanish 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 Spanish economy emerged from recession in the third quarter of last year but state finances are still under scrutiny by the European Union .

Table 3.4: 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 3.5: 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 3.6: 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	0.030
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. 3.1. Web Attention (SVI) for Greece.

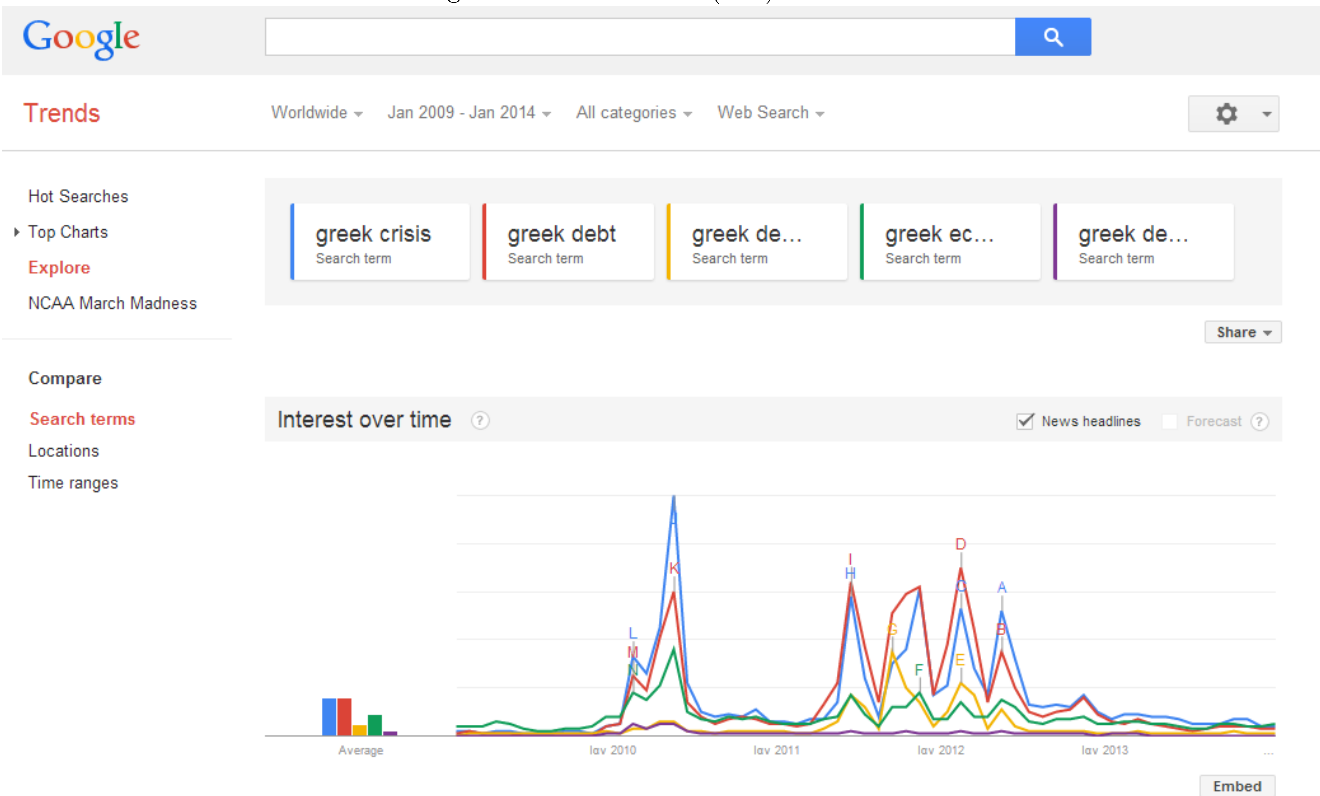


Fig. 3.2. Euro-periphery bottom extreme returns count, News Flow and Web Attention (SVI) factors

Table 3.7: 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 3.8: 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 3.9: 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 3.10: 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 3.9

Chapter 4

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

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 speed and breadth of coverage) provide a channel through which "bad" news is instantaneously circulated, and this stimulates worldwide "shocks" to stock prices, often 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.

4.1. Introduction

The Euro-crisis has been central theme for the financial press since it started, around the end of 2009. Tens of thousands of news articles were written about the Euro-periphery countries (Portugal, Ireland, Italy, Greece, Spain), with substantial debate on how bad their finances were and whether they should leave the common currency area (Eurozone). An issue of great concern has been whether bad news about them might affect the rest of the European countries, in case of some dramatical development regarding their finances.

The previous literature that dealt with news and events that were related to the European Financial Crisis (e.g. Arezki et al. (2011), Beetsma et al. (2013), Mink and De Haan (2013)) studied the impact of announcements such as rating changes, capturing events with the use of dummy variables, or classifying events as "good" or "bad". Such an approach can be problematic, not least because, it does not capture all available information because classifying events manually is a subjective process. Furthermore, it neglects the full spectrum of financial sentiment, in other words, how negatively or positively information is perceived. 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 at a very fast 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, this is the first paper to study the high-frequency effects of the Euro-crisis using data with are timestamped up to milliseconds. In addition, I use data from the two biggest international newswires (Thomson Reuters and Dow Jones) and I select articles related to the European Financial Crisis using a set of keywords that must be located in the content and the title of each selected article. I examine the impact of News Flow about the peripheral countries on international intraday stock returns, answering 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 center of financial discussions and policy making since the advent of the Euro-crisis. One step further, I answer the question about how fast these effects become visible in a large set of stock markets (European, North American, Asian). This paper also contributed to the policy decision making discussion since huge amounts of money had to be disbursed by European Union countries to bail out countries and banks on the brink of default. If news does not only affect the crisis countries, providing financial assistance is also an action that aims to control the risk of crisis propagation with a drop in worldwide stock prices (i.e. a drop in market capitalizations).

I find that intraday (high frequency) European Financial Crisis news affects international stock markets. The effect is not limited to the crisis-hit Euro-periphery countries, and not even just in the European Union countries, with overseas financial markets also affected (USA, Japan, China). More media pessimism 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 et al. (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 et al. (1992) who 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 et al. (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 crisis-hit countries affect the countries 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 quickly do these effects appear? Are multiple hours or days needed, or are small time frames such as 5 minutes and 30 minutes enough to start seeing the effects of these pessimistic articles? The theoretical models about noise traders, beliefs and sentiment provide an interesting framework through which to examine these questions.

Recent advances in the Data Mining and Sentiment Analysis techniques have led to the creation of companies which attempt to capture financial sentiment (for example, RavenPack[1] and Thomson Reuters News Analytics[2]). Recently, textual analysis techniques have been used to study the effect of media pessimism on stock prices: Tetlock (2007) is one of the first studies to use textual analysis in finance in order to analyze a Wall Street Journal column and it finds a significant effect of media pessimism on Dow Jones stock prices, using the Harvard IV-4 General Inquirer software. While Garcıa (2012) examined the impact of stock returns on media pessimism. Another area of application has been the study of corporate filings. Loughran and McDonald (2011) create finance-oriented word lists by selecting a part of words contained in the Harvard IV-4 dictionary, applying textual analysis in a set of US listed firms' annual report (SEC Form 10-K). Chouliaras (2015b) finds a different effect of 10-K pessimism for stocks with positive/negative returns in the previous period, while Chouliaras (2015a) finds that SEC Form 10-K pessimism negatively affects the number of institutions that hold the stock, after the filing. Researchers have recently analyzed the effect of textual pessimism on major online stock message boards (Antweiler and Frank (2004), Chen et al. (2013)). Garcia (2013) finds different effect of media pessimism on stock prices during different phases of the business cycle (expansions, recessions). Tetlock (2011) reports that a high level of textual similarity in firm-specific news leads to higher trading aggressiveness, for individual investors. Boudoukh et al. (2013) use a computational algorithm to classify and categorize news, and find that identified news have a higher impact on stock markets than news that can not be identified. Tetlock et al. (2008) find that more negative firm-specific news predict lower quarterly earnings. Loughran and McDonald (2013) find that higher uncertainty in IPO filings is associated with higher first-day returns and ex post volatility, while Jegadeesh and Wu (2013) use different weights for 10-K words based on the market reactions around the release of annual filings. Ahern and Sosyura (2014) use textual analysis on the field of Mergers and Acquisitions (M&As), and reports evidence that firms manipulate the

[1] http://www.ravenpack.com/
[2] http://www.machinereadablenews.com/

media in order to achieve better stock prices during the phase of M&As negotiation phase.

My study is related to Tetlock (2007), Garcia (2013), Garcıa (2012) and Chouliaras and Grammatikos (2013) since I also study the effect of media on the stock markets. I differentiate from the previous literature in a number of areas: First and foremost, I employ a database of 13145 articles from the two biggest international newswires (Dow Jones and Thomson Reuters), while previous papers typically analyze the effect of one financial column on a particular stock index. None of the previous studies deals with the high frequency effects of news during the Euro-crisis. The European Financial Crisis provides me with an interesting framework under which I study the effect of news on financial markets; going beyond firm specific news and the use of a single newspaper column. In my setting, the articles are released at any point during the 24 hours of a day. Given the fact that every article and every stock price is timestamped (at the level of minutes), I make sure that articles precede stock prices. Furthermore, the selected articles are related to multiple countries and I study their effects on eighteen (18) European, North American and Asian countries, while the previous literature typically studies the effects on just one stock index. Finally, this paper contributes to the policy debate since it studies the question of whether news about a group of crisis-hit countries can affect financially ”stronger” ones, and how fast this effect becomes visible in stock prices.

The rest of this paper is organized as follows. Section 4.2 presents the data. Section 4.3 presents the model specification and I explain how I study the effect of News Sentiment in a cross-country setting. Section 4.4 provides the empirical results. Section 4.5 concludes.

4.2. 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 6.1:

Insert Table 6.1 here

As far as the news articles data are concerned, I employ the Dow Jones Factiva news database[4]. The time period under examination starts from 1 December 2009, and it ends on 28 November 2012. Europe slipped into recession in 2009[5], while by December 2009 it was clear something was seriously wrong with the finances of Greece[6], which makes December 2009 a reasonable point to start our analysis. I end the analysis due to data availability

[3]Thomson Reuters Tick History can be found at: `http://thomsonreuters.com/tick-history/`

[4]Dow Jones Factiva can be found at `http://new.dowjones.com/products/factiva/`

[5]`http://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`

of 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.

4.3. The Methodology

4.3.1. 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. 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 was 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. I am then able to proceed studying the 5-minute (and 30-minute) effect of media content on high frequency stock returns. 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 record prices every 5 minutes (and 30 minutes). If between the 5-minutes (and 30-minutes) time intervals more than one story is 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" this 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.

4.3.2. News Flow

Can new information significantly affect stock markets? One can hypothesize that the release of information through media outlets, can be one of the factors that affect stock prices. The focus of this paper is to quantify the high-frequency financial sentiment of European Financial Crisis information flows, and relate it to stock prices. This way I contribute to the relatively new strand in the finance literature that examines directly how information is assimilated into financial markets.

In this context I analyze the content of news stories in the newswires. As far as the news stories are concerned, I extract and analyze news articles covering the time period from 1 December 2009 until 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 selected stories are those that contain the name of (at least) one Euro-periphery country, along with (at least one) of the following financial crisis terms: crisis, debt, economy, deficit, default. The retrieved articles were the ones that contained (at least) one of the terms that were mentioned in Section 4.3.2.

4.3.3. Preprocessing the news data

Initially, I exclude duplicate articles. It is very common that the same (or highly similar) pieces of information are distributed as frequently as ten times or more, especially in newswires such as the *Dow Jones Newswires* and the *Thomson Reuters* that release information in real time. A small number of observations might dominate the sample. 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. As a robustness check, I 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 in Section 4.4.2.

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

An article may refer to more than one topic, even though it may contain one or more of the keywords I defined previously. Every few lines, a new topic may be covered in the article. To attempt to achieve a higher degree of topic accuracy, on top of the previous keywords in the content in the article, I also search for keywords in the title of the articles. Very often the title is the main indicator of the content of the article. One can expect articles with one or more keywords in the title to be of higher precision to a topic, compared to selecting articles only via keywords in the content. For Greece, the title keywords are: "greece", "greece's", "greek", "greeks", "hellas", "hellenic". A similar set of keywords is used for each Euro-periphery country. After these filters are applied, the sources and the number of news items appear in Table 6.2, for a total of 13145 news articles passing the two steps of filters (content and title keywords):

Insert Table 6.2 here

One can see that most of the stories concern Greece: Out of 13145, 8128 concern the greek crisis. This is to be expected since Greece is the country that was the most badly hurt, and naturally the financial press took a great interest in covering events concerning Greek financial troubles. After Greece, the most articles were written about Spain (1822), followed by Ireland (1319), Italy (1268) and finally, Portugal (608). Quite interestingly, 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 of stories) the Euro-crisis, and that the keyword filters and not biased in terms of selecting stories from the one or other newswire.

The graphical plot of the number of articles during each time period is shown in Figure 6.1:

Insert Figure 6.1 here

Naturally, the number of stories is not constant as the Euro-crisis "escalates" and "calms down" at different periods. 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, with 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 on this day actually more than 500 crisis-related stories were released, a factor indicating the extremely high interest about the Euro-crisis 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 Greece[7]. Another spike occurs around March 2012, when the new bailout of Greece was finally granted[8]. 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 aid[9].

4.3.5. Textual Analysis

As a next step, we use textual analysis, based on the Loughran and McDonald (2011) dictionary[10]. We calculate the positive, the negative media content and the Pessimism as in Section 3.3.1.3, but in this case we calculate the media contents for each article separately. If more than one stories is 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. The Average Pessimism is calculated by taking the average values of pessimism, taking into account the articles that were released in every respective time interval[11]. Another metric I use is the *News Count* (N_t) which is the number of articles released between two financial prices (5 minute and 30 minute intervals) regarding any of the queries I am using.

[7]For example, an article covering the event from Guardian: `http://www.theguardian.com/business/2011/jul/21/european-debt-crisis-summit-euro`

[8]An article by BBC covering the event: `http://www.bbc.com/news/business-17338100`

[9]Another article by BBC covering the event: `http://www.bbc.com/news/business-13798000`

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

[11]I do not calculate the Average Positive and the Average Negative factors, since the Pessimism is found to clearly outperform both the Positive and the Negative factors (see the results of Table 6.8, with the Pessimism factor being significant for all countries, while the Positive factor is significant for only five countries, and the Negative factor is significant for eight countries. After all, the Pessimism factor is a product of the Positive and the Negative factor, since it linearly combines the two factors, in particular being the difference between the two factors, which happens to perform better, as also shown in other papers such as Garcia (2013).

The summary statistics for the media content (positive, negative, pessimism) and the stock markets appear in Table 6.3:

Insert Table 6.3 here

I see that stock markets have mean returns of 0 (normal since I am using 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 6.4, 6.5 and 6.6:

Insert Tables 6.4, 6.5 and 6.6 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 6.7:

Insert Table 6.7 here

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

4.4. Empirical Findings

4.4.1. 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:

$$R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i} \tag{4.1}$$

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 6.8:

Insert Table 6.8 here

Table 6.8 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 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 and affects American and Asian stock markets.

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.

"Average Pessimism" (4th row of Panel A) is significant for six (6) out of ten (10) countries, with a negative sign. Once more, "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 capitalizations 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. Once more, for the five minute intervals, the "Pessimism" factor is more significant than the individual "Positive" and "Negative" factors, 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 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.

4.4.2. The impact of repeat articles

In Section 4.3.3, 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

$$R_t = a_1 + b_1 Repeat_t + \sum_{i=1}^{5} c_i R_{t-i} \tag{4.2}$$

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/Obsershelp (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 6.10:

Insert Table 6.10 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 4.3.3, does not significantly affect our results.

4.4.3. Overnight Returns and News Similarity

In order to take into account the effect of media pessimism on overnight returns, along with news similarity, I add a dummy ($Overnight_t$) that takes the value of one (1) for overnight returns, and five (5) dummies respectively for the five (5) lagged values of returns ($Overnight_{j,t}$). On top of that, I use five dummies variables, one for each level of similarity ($D_{Similarity,i,t}$, taking into account five levels of news similarity: 50%, 60%, 70%, 80% and 90%). I use the following specification:

$$R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j} + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k} \tag{4.3}$$

The results appear in Tables 6.11, 6.12, 6.13, 6.14 and 6.15 for M_t being equal to the Pessimism, Average Pessimism, Positive, Negative and News Count factors respectively:

Insert Tables 6.11, 6.12, 6.13, 6.14 and 6.15 here

The vast majority of the coefficients of $Overnight \times Pessimism$ and $Overnight \times Average\,Pessimism$ are significant, as seen in Tables 6.11, 6.12, 6.14 and 6.15. It appears to be that overnight returns are not affected by media pessimism and negativity, even though the Pessimism, the Average Pessimism and the Negative coefficients are significant for the majority of countries. On the other hand, four coefficients are significant for the $Overnight \times Positive$, as seen in Table 6.13, three of which (Portugal, Germany and China) are positive, indicating higher stock returns when overnight media content is positive.

As far as the similarity dummies are concerned, the results are significant for Greece for all five (5) cases for the $D_{Similarity,i,t} Pessimism_t$ and the $D_{Similarity,i,t} Average\,Pessimism_t$ factors (Similarity greater than 50%, 60%, 70%, 80% and 90%), for Tables 6.11 and 6.12. The sign is negative, which indicates that an increase in pessimism is associated with lower stock prices, in the presence of highly similar news. Quite interestingly, the coefficient is positive for Germany (only for the $D_{Similarity,i,t} \times Pessimism_t$), which means that, even though the coefficients of Pessimism and Average Pessimism are negative for Germany, the coefficients of $D_{Similarity,i,t} \times Pessimism_t$, i.e. highly similar news does not lead to lower stock prices for Germany, even overall the media pessimism affects the German stock market in a negative way (i.e. a higher Pessimism and a higher Average Pessimism are associated with lower stock prices). As far as Tables 6.13, 6.14 and 6.15 are concerned, the coefficients are mostly insignificant for Table 6.15, while the coefficients are negative for Greece and Italy for the $D_{Similarity,i,t} \times Positive_t$ for Greece and Italy for Tables 6.13 and 6.14 (insignificant for Table 6.14), while the coefficients for Germany are once more positive for the Negative factor (Table 6.14).

4.4.4. 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. Volatility can be proxied by the squared return during the time interval. For this reason, I estimate the following model which takes studies the effect of media pessimism on squared returns, taking into account news similarity and overnight returns, using the dummies previously defined: $Overnight_t$, that takes the value of one (1) for overnight returns, and five (5) dummies respectively for the five (5) lagged values of returns ($Overnight_{j,t}$). On top of that, I use five dummies variables, one for each level of similarity ($D_{Similarity,i,t}$, taking into account five levels of news similarity: 50%, 60%, 70%, 80% and 90%). I use the following specification:

$$R_t^2 = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j}^2 + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}^2 \qquad (4.4)$$

The results appear in Tables 6.16, 6.17, 6.18, 6.19 and 6.20 for M_t being equal to the Pessimism, Average Pessimism, Positive, Negative and News Count respectively:

Insert Tables 6.16, 6.17, 6.18, 6.19 and 6.20 here

As far as the results of Table 6.16 are concerned (i.e. for M_t equal to the Pessimism factor) the results show that Pessimism positively and significantly affects volatility (proxied by squared returns) in the case of Germany and the US. Quite interestingly, the results are not significant for Portugal, Greece and Italy for the Pessimism factor. For Germany, the coefficient of pessimism on volatility (squared return) is equal to 0.000128, significant at the 1% level, with a t-stat equal to 2.68. For the US, the coefficient is 0.0002666, with a t-stat equal to 3.12. These positive coefficients indicate than an increase in media pessimism, is associated with higher volatility (as proxied by squared returns) for the case of Germany and the US. As far as the Overnight Pessimism is concerned, the coefficients are significant and positive for Portugal, Greece, Italy, Germany and Japan. These positive coefficients indicate that an increase in overnight pessimism, is associated with higher volatility. The coefficients for Portugal, Italy and Germany are significant at the 1% level, with coefficients equal to 0.000760, 0.00111 and 0.000515 respectively, with t-stats equal to 5.06, 2.99 and 3.05 respectively. For Greece and Japan, the coefficients are significant at the 10% level, with their magnitude being equal to 0.00139 and 0.00108 respectively, while their t-statistics are equal to 1.96 and 1.71 respectively. When the repeat dummies are taken into account, the effect on volatility is mostly negative, for the majority of countries, for the five cutoffs (50%, 60%, 70%, 80%, 90%) of textual similarity. For example, the coefficients of Portugal and Italy are equal to -0.0000778 and -0.000173, significant in the 1% level, with t-statistics equal to -2.63 and -2.79 respectively. As far as the results of Table 6.17 are concerned (i.e. for M_t equal to the Average Pessimism factor), the average pessimism coefficient is positive and significant for Italy, Germany, the US and Japan. In other words, a higher Average Pessimism is associated with a higher volatility, proxied by the squared returns. The coefficients for Germany and the US are equal to 0.000123 and 0.000246 respectively, significant at the 1% level, with t-statistics equal to 2.86 and 2.79 respectively, while the coefficients for Italy and Japan are equal to 0.000129 and 0.000348, with t-statistics equal to 1.68 and 1.65, significant at the 10% level. The effect of Overnight Average Pessimism is positive and significant for Portugal, Greece, Italy, Germany and Japan. The coefficients for Portugal, Italy and Germany are significant at the 1% level, with coefficients equal to 0.000716, 0.00158 and 0.00129, while their t-statistics are equal to 5.70, 2.67 and 3.87 respectively. The coefficients for Greece is equal to 0.00158, with a t-statistic equal to 2.15, significant at the 5% significance level. For Japan, the coefficient is equal to 0.00121, with a t-statistic equal to 1.65, significant at the 10% level. Finally, as was the case with the similarity dummies for pessimism, the effect of Average Pessimism when there are similar news articles is negative for the majority of countries (Portugal, Greece, Italy, Germany and the US). In particular, the effect is significant at the 1% level for Italy, with the coefficient being equal to -0.000187 and a t-statistic equal to -2.64. For Portugal, Greece and Germany, the coefficients are -.0000994, -0.000355 and -0.000129, significant at the 5% levels, with t-statistics equal to -2.43, -2.15 and -2.41 respectively. In Table 6.18, which stands for the Positive media factor, the majority of coefficients are negative for the Positive factor, while the majority of coefficients for $Overnight \times Positive$ are positive. The $D_{Similarity,i,t} \times Positive$ coefficients are negative for Germany and the US, while they are mostly positive for Japan and China. The results of Table 6.19 show a positive and statistically significant coefficient for Germany and the US for the Negative factor, which indicates that a more negative media content is associated with a higher stock market volatility for these two countries. In particular, the coefficient for Germany is 0.000144, statistically significant at the 1% level, with a t-statistic equal to 2.69. The coefficient for the US is equal to 0.000258, once more significant at the 1% level, with a t-statistic equal to 2.84. The results for volatility seem to be the most significant for the overnight returns, since we see significant and positive coefficients for five countries, with once more higher negativity in the media being associated with high overnight stock market volatility. The $D_{Similarity,i,t} \times Negative$ coefficients are mostly negative, with Japan and China being the exceptions with positive coefficients. Table 6.20 indicates the result for M_t being equal to the News Count fator. It seems like the number of news articles (News Count) is the best factor as far as the volatility is concerned: we find significant coefficients for five countries, with all coefficients being positive and statistically singificant at the 1% level. In particular the coefficients for Portugal, Greece, Italy, Germany and Japan are 0.00000323, 0.00000345, 0.00000295, 0.00000165 and 0.00000177 respectively, with t-statistics equal to 2.61, 4.32, 3.83, 5.36 and 7.09 respectively. The $Overnight \times$ News Count coefficients are significant for four countries, with three of the coefficients being positive. The $D_{Similarity,i,t} \times$ News Count coefficients are mostly negative, significant for Italy, while they are mostly positive for China.

4.5. 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 react very quickly when "absorbing" the shocks of media pessimism. Even time frames such as 5-minutes and 30-minutes can be enough for stock prices and volatility to be affected by increased 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 in "real time" by high frequency stories released through newswires. Repeated stories do not seem to influence stock markets, while they seem to decrease volatility.

Chapter 5

Conclusions and Future Work

5.1. Conclusions and Future Work

The European Financial Crisis has been one of the most significant and dramatic events of the past decades. The structure of the "European Union" and the creation of the common currency, under the "Eurozone", were under threat during the spread of the crisis[1]. It is a well established fact that before the crisis the authorities did not properly supervise countries with respect to their competitiveness and the macroeconomic imbalances. Subsequently, the enforcement of what was agreed has not been ideal in some cases. Another problem of the current setting has to do with the slow decision making capacity of the European Union, which has to do with institutional weaknesses, since the number of interacting countries was very large and heterogeneous. When the crisis started, there was no mechanism set in order to provide financial assistance to the euro area countries that needed it. Greece, and later on Ireland, Portugal, Spain and Cyprus were not able to borrow money at sustainable interest rate levels[2]. The EU had to create emergency crisis resolution mechanisms in order to deal with the crisis, namely the European Financial Stability Facility (EFSF) and the European Stability Mechanism (ESM)[3]. European governments used a total of 1.6 trillion euros (between 2008 and 2011) to support the banking system, an amount which is equivalent to 13% of the EU's annual Gross Domestic Product (GDP)[4].

This PhD Thesis consists of three papers which deal with the topic of the "European Financial Crisis". The first paper *("Extreme Returns in the European Financial Crisis")* deals with the dynamics of extreme stock returns and the transmission of financial shocks among fifteen European countries, split in three groups: the Euro-periphery (Portugal, Ireland, Italy, Greece and Spain), the Euro-core (Germany, France, the Netherlands, Finland, Belgium)

[1] http://ec.europa.eu/economy_finance/explained/the_financial_and_economic_crisis/why_did_the_crisis_spread/index_en.htm

[2] http://ec.europa.eu/economy_finance/explained/the_financial_and_economic_crisis/assisting_countries_in_trouble/index_en.htm

[3] http://ec.europa.eu/economy_finance/explained/the_financial_and_economic_crisis/responding_to_the_debt_crisis/index_en.htm

[4] http://ec.europa.eu/economy_finance/explained/the_financial_and_economic_crisis/responding_to_the_financial_crisis/index_en.htm

and the Non-euro (Poland, Sweden, Czech Republic, UK, Denmark). Financial data cover the period January 2004 to March 2013, and they are obtained from Thomson Reuters Datastream. The data concern stock indices prices, exchange rates, short term interest rates, conditional volatility estimated using an EGARCH(1,1) model. The paper finds the presence of transmission effects for the tails of the returns distributions for the Pre-crisis, the US-crisis and the Euro-crisis periods from the Euro-periphery group to the Non-euro and the Euro-core groups. Even before the crises periods there was a significant shock transmission from the Euro-periphery to the other two groups. The shocks transmitted during the crises were more substantial (in some cases, extreme bottom returns doubled). As extreme returns have become much more "extreme" during the financial crisis periods, the expected losses on extreme return days have increased significantly. Given the fact that stock market capitalisations in these country groups are trillions of Euros, a 1% or 2% increase in extreme bottom returns (in crisis periods) can lead to aggregate losses of tens of billions Euros in one single trading day. This paper uses multinomial logistic models, which, as is highlighted in Hosmer Jr et al. (2013), are extensively used in the research field of epidemiology, which studies the spreading of contagious diseases. Such models calculate the probability that a disease that has already affected N persons, will affect K persons.

The second paper *("News Flow, Web Attention and Extreme Returns in the European Financial Crisis")*, performs textual analysis on more than 24,000 news articles from seven leading international news providers, and finds 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. The News Flow data are obtained from the *Dow Jones Factiva* platform, while the Web Attention data are obtained from the *Google Trends* platform. The news articles come from seven news sources: *The Dow Jones Newswires, The Thomson Reuters Newswires, The Financial Times, The Wall Street Journal, The New York Times, The Telegraph and The Times.* The field of textual analysis has taken off lately, especially since the publications of Tetlock (2007) and Loughran and McDonald (2011). Researchers are attempting to quantify information that is embedded in the textual content of non-numerical sources, such as news and corporate filings. This paper contributes to the textual analysis literature, as well as the financial crisis literature. It is the first paper to use a direct measure of media pessimism, in order to answer the question of whether indeed news about the Euro-periphery countries are able to affect stock returns in two groups of European countries.

The third paper ("High Frequency Newswire Textual Sentiment Analysis: Evidence from International Stock Markets during the European Financial Crisis"), performs textual analysis on 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). The importance of high frequency news and the speed of price adjustments is very high, as highlighted in Foucault et al. (2015)'s recent theoretical model. Previous high frequency trading and "news" studies (Brogaard et al. (2014), Andersen et al. (2003), Hirschey (2013), Riordan et al. (2013)), typically consider macroeconomic announcements as their source of "news." The date and time of these announcements are well known in advance. In a crisis setting, especially one such as the Euro-crisis, a series of unprecedented events occurred, and the media served as a proxy for their content, depending on how positive or negative the tone was in the information they transmitted. To the best of my knowledge, this paper is the first to study the high-frequency effects of non-firm-specific news (Hirschey (2013), Riordan et al. (2013)) that is also unrelated to macroeconomic announcements (Brogaard et al. (2014), Andersen et al. (2003)).

In summary, this PhD Thesis contributes to a number of research fields: The first paper contributes to the field that studies the dynamics of extreme stock returns during a crisis period, and shows evidence that indeed extreme returns from one group in trouble (Euro-periphery) can affect other group of countries (Euro-core and Non-euro). The second paper contributes to the fields of "Web Attention" and "Textual Analysis" and shows that more Web searches and a higher media pessimism are associated with higher probabilities of extreme returns for all three groups, during the crisis period. Finally, the third paper, contributes in the fields of high frequency finance, textual analysis and information/price discovery. Indeed news can move stock markets in Europe, US and Asia, in a very rapid manner: 5-minutes and 30-minutes are enough for stock markets to (at least partially) assimilate the content of new information. The paper shows that news affect stock markets negatively and volatility positively (i.e. a higher media pessimism leads to lower stock prices and higher stock market volatility). Finally, this PhD Thesis contributes to the policy debate on whether indeed a group of countries that have financial problems can provoke negative equity returns in other countries. This debate is significant, especially given the attempts that the European countries make for a greater unification and integration.

An important issue that one should not neglect, has to do with the interdisciplinary responsibility and the inherent sensivitiy of the used/applied techniques. Natural Language Processing, Text Analysis, Machine Learning and others are not "warehouses", which offer tools because they are there. Instead, they are complex fields, which have found their place in research and which demand a 'scientific respect'. At this point, it is very important to explicitly highlight the interdisciplinary character of the thesis.

If we were to perform a SWOT analysis for the thesis, among the strengths we would likely include the fact that the tools and methods used in this research have stood out quite well in recently published (and ongoing working papers) research publications in finance (Bae et al. (2003) (RFS), Boyson et al. (2010) (JF), Loughran and McDonald (2011) (JF), Chen et al. (2013) (RFS), Garcia (2014), Garcia (2013) (JF), Loughran and McDonald (2013) (JFE), Jegadeesh and Wu (2013) (JFE), Ahern and Sosyura (2014) (JF), Dougal et al. (2012) (RFS), Kelley and Tetlock (2013), Tetlock (2007) (JF), Tetlock et al. (2008) (JF), Da et al. (2011) (JF), Da et al. (2015) (RFS), Bodnaruk et al. (2015) (JFQA)). On top of this literature, our contribution has to do with the application of these methods on a currently very important topic (the Euro-crisis) and on novel, unique and hand-collected data (news from Dow Jones Factiva and the Thomson Reuters/Dow Jones newswires). As far as the possible weaknesses are concerned, future researchers could try to relax the assumptions made in these research papers, and study alternative specifications in respect to the data, the methodologies and the econometric models. For example, one could study alternative dictionaries and/or different approaches to weighting words and/or the production of the financial sentiment time series.

At this point, it is rather important to argue what could be done next, in follow-up research: An excellent idea would be to study Country Stock Indices vs. Sectorial Stock Indices. As this research concentrates on financial crises, the effect of the examined breaking news would likely be different on financial stocks than non-financial stocks. Such distionction cannot be observed if one looks only at country stock indices, as in this thesis. Thus, future research could also study sectorial indices that represent specific types of economic activity (e.g. financials, industrials, high tech, pharmaceuticals, natural resources, energy etc.) or broader stock groups (e.g. cyclical vs. non-cyclical), especially when looking at the effects of news on the Euro-core or US stock markets, where there are many big non-financial listed companies. Such further research may likely reveal that there are more degrees of freedom for investors to diversify during financial crises than what is suggested by the current thesis.

Another very good application would be related to trading algorithms. Future research could look into algorithmic trading strategies based on news. For example, one could perform simulations where a position is opened after a news item is released, and subsequently closed some minutes later (or few hours later or next day, etc), and see whether after transaction costs, such strategies would have a positive return and an acceptable Sharpe ratio. This would also address the more fundamental question of whether markets are efficient in assimilating breaking news or, instead, an investor/trader who mechanically reacts to news can generate profits consistently after transaction costs.

Chapter 6

Appendix

Table 6.1: 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 6.2: 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. 6.1.

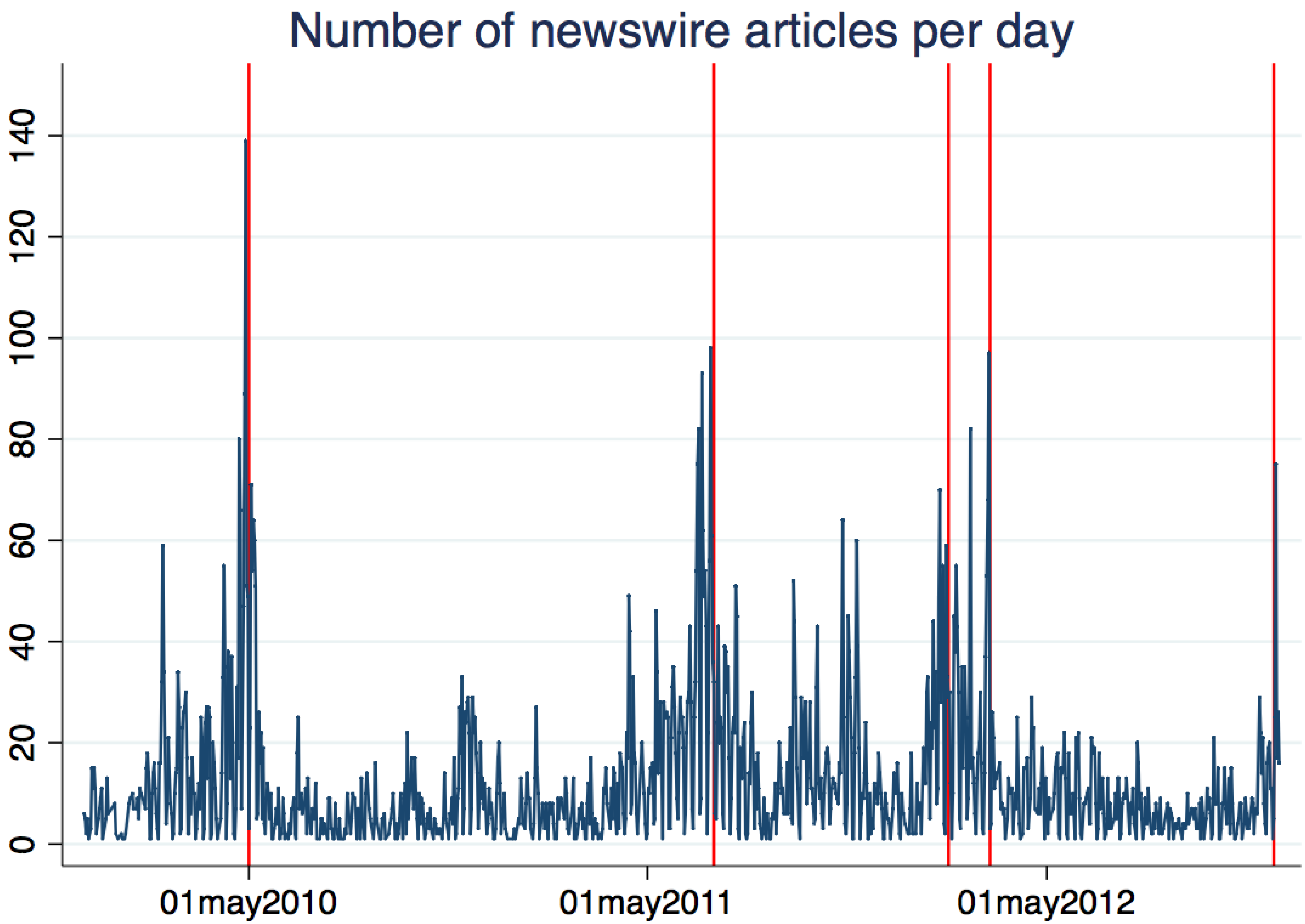

Number of newswire articles per day
0
20
40
60
80
100
120
140
01may2010
01may2011
01may2012

Table 6.3: Summary statistics for news articles and stock markets. The construction of the media content variables ("Negative", "Positive", "Pessimism") are defined in the Section 4.3.5

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 6.4: 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 color) and Positive (green color) 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 6.5: 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 color) and Positive (green color) word lists of Loughran and McDonald (2011).

Country	Source	Date	Title	Text
Ireland	Dow Jones		MARKET TALK: Fears Of Irish Debt Restructuring Rising	MARKET TALK: Fears Of Irish 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 Irish crisis just as worrying as Greece	Merkel says Irish 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 by Gernot Heller, Writing by Sarah Marsh)

Table 6.6: 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 color) and Positive (green color) word lists of Loughran and McDonald (2011).

Country	Source	Date	Title	Text
Italy	Dow Jones	20 June 2011	Moody's Analyst: Italy 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 Italian economy , 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 Italy 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 Italian economy '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 6.7: 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 color) word list of Loughran and McDonald (2011).

Country	**Source**	**Date**	**Title**	**Text**
Greece	Reuters	8 March 2011	IMF believes Greek 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 Dalgleish) IMF-GREECE/BORGES (UPDATE 1) Reuters Limited Document LBA0000020110310e737001eb

Table 6.8: 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 4.3.5. 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) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) France	(6) UK	(7) Switzerland	(8) US	(9) Japan	(10) China
Panel A: 30-minutes intervals										
Pessimism	-0.00812***	-0.0194***	-0.00886**	-0.00596*	-0.00757**	-0.00618**	-0.00536**	-0.0181***	-0.0767***	-0.0471**
	(-2.99)	(-2.99)	(-2.14)	(-1.70)	(-2.29)	(-2.51)	(-2.15)	(-2.75)	(-3.17)	(-2.31)
Positive	0.0131*	0.0296*	0.0169	0.0147*	0.0218***	0.00951	0.00927	0.0189	0.132**	0.0937
	(1.71)	(1.74)	(1.50)	(1.72)	(2.65)	(1.53)	(1.32)	(1.03)	(2.16)	(1.54)
Negative	-0.00768***	-0.0193***	-0.00784*	-0.00475	-0.00555	-0.00593**	-0.00499*	-0.0194***	-0.0736***	-0.0438*
	(-2.61)	(-2.73)	(-1.72)	(-1.23)	(-1.56)	(-2.17)	(-1.86)	(-2.82)	(-2.71)	(-1.93)
Average Pessimism	-0.00531**	-0.0139**	-0.0106**	-0.00609*	-0.00501	-0.00455*	-0.00332	-0.00978	-0.129***	-0.0537
	(-2.09)	(-2.17)	(-2.49)	(-1.82)	(-1.60)	(-1.81)	(-1.41)	(-1.33)	(-2.72)	(-1.42)
News Count	-0.0000111	0.0000579	-0.00000382	0.0000197	-0.0000228	0.0000130	-0.00000604	-0.00000460	-0.0000432	0.0000166
	(-0.17)	(1.54)	(-0.10)	(0.58)	(-0.56)	(0.53)	(-0.23)	(-0.49)	(-0.07)	(0.78)
Panel B: 5-minutes intervals										
Pessimism	-0.00374***	-0.00689***	-0.00353*	-0.00432**	-0.00330*	-0.000935	-0.00478***	-0.00728*	-0.0580***	-0.0501**
	(-2.84)	(-3.11)	(-1.93)	(-2.34)	(-1.95)	(-0.79)	(-3.45)	(-1.74)	(-3.18)	(-2.52)
Positive	0.00867**	0.0138**	0.00638	0.0122***	0.00658	0.00258	0.0102**	0.00868	0.0960*	0.0756
	(2.37)	(2.25)	(1.30)	(2.66)	(1.62)	(0.91)	(2.58)	(0.78)	(1.95)	(1.30)
Negative	-0.00311**	-0.00634**	-0.00322	-0.00322	-0.00293	-0.000705	-0.00416***	-0.00757*	-0.0567***	-0.0511**
	(-2.31)	(-2.55)	(-1.58)	(-1.62)	(-1.64)	(-0.52)	(-2.83)	(-1.74)	(-2.70)	(-2.24)
Average Pessimism	-0.00200**	-0.00522***	-0.00263*	-0.00272**	-0.00197	-0.000293	-0.00290***	-0.00104	-0.0683**	-0.0505*
	(-2.22)	(-2.58)	(-1.75)	(-2.00)	(-1.57)	(-0.32)	(-2.83)	(-0.26)	(-2.48)	(-1.83)
News Count	-0.0000473	0.0000274	-0.0000549	0.0000293	-0.0000223	-0.0000207	-0.0000552**	-0.000000619	0.00000551	0.0000147
	(-0.36)	(1.36)	(-1.54)	(0.75)	(-0.37)	(-1.01)	(-2.29)	(-0.09)	(0.23)	(0.66)

t statistics in parentheses

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

Table 6.9: 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 4.3.5. II 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***	0.00409*	0.00514***	0.00755***	0.00662***	0.00673***	0.00467***	0.00947**	0.0156***	0.00960**
	(2.81)	(1.69)	(2.83)	(5.30)	(5.16)	(6.58)	(4.46)	(2.22)	(2.90)	(2.34)
Positive	-0.00994***	-0.0137**	-0.0143***	-0.0147***	-0.0117***	-0.0118***	-0.00659**	-0.0143*	-0.0236*	-0.0187
	(-2.64)	(-2.15)	(-2.97)	(-3.94)	(-3.51)	(-4.45)	(-2.27)	(-1.91)	(-1.82)	(-1.48)
Negative	0.00249*	0.00285	0.00382*	0.00672***	0.00610***	0.00620***	0.00462***	0.00936**	0.0156***	0.00901**
	(1.93)	(1.05)	(1.90)	(4.19)	(4.31)	(5.50)	(3.99)	(2.06)	(2.65)	(1.98)
Average Pessimism	0.00504***	0.00706**	0.00828***	0.0104***	0.00778***	0.00871***	0.00604***	0.0142***	0.0141	0.0227**
	(3.75)	(1.95)	(3.07)	(3.46)	(1.75)	(3.26)	(5.06)	(1.35)	(0.57)	(2.02)
News Count	0.0000440***	0.0000272***	0.0000221**	0.0000121**	0.0000163**	0.00000554	0.0000123***	0.000000255	0.00000317	0.0000141***
	(3.54)	(3.80)	(2.53)	(2.01)	(2.12)	(1.30)	(2.73)	(0.07)	(0.72)	(3.11)
Panel B: 5-minutes intervals										
Pessimism	0.00126***	0.00172**	0.00280***	0.00306***	0.00250***	0.00299***	0.00210***	0.00297***	0.00451***	0.00402**
	(3.15)	(2.24)	(5.05)	(6.13)	(5.43)	(8.76)	(5.85)	(3.29)	(3.18)	(2.40)
Positive	-0.00291**	-0.00266	-0.00759***	-0.00726***	-0.00536***	-0.00546***	-0.00465***	-0.00314	-0.00656*	-0.00994**
	(-2.40)	(-1.22)	(-5.00)	(-5.61)	(-4.43)	(-6.42)	(-4.77)	(-1.19)	(-1.68)	(-2.03)
Negative	0.00104**	0.00170**	0.00213***	0.00251***	0.00215***	0.00272***	0.00180***	0.00315***	0.00456***	0.00344*
	(2.31)	(1.98)	(3.44)	(4.53)	(4.22)	(7.08)	(4.47)	(3.24)	(2.87)	(1.76)
Average Pessimism	0.00147***	0.00213**	0.00319***	0.00344***	0.00279***	0.00332***	0.00229***	0.00459***	0.00496	0.00983**
	(3.54)	(1.40)	(2.44)	(6.46)	(5.75)	(9.14)	(6.07)	(3.86)	(1.48)	(2.12)
News Count	0.0000117**	0.00000565**	0.00000656*	0.00000335	0.00000585*	0.00000184	0.00000286	-0.000000368	0.000000866	0.00000547***
	(2.01)	(2.06)	(1.93)	(1.32)	(1.84)	(1.05)	(1.56)	(-0.31)	(0.56)	(2.93)

t statistics in parentheses
* p<.1, ** p<.05, *** p<.01

Table 6.10: 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. 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, using the Ratcliff/Obsershelp (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.

	(1) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) US
$Similarity > 50\%$	-0.00000119	-0.000140**	-0.00000637	0.0000245	-0.00000736
	(-0.05)	(-2.19)	(-0.15)	(0.64)	(-0.17)
$Similarity > 60\%$	-0.00000955	-0.000129**	0.0000198	0.0000306	0.000000709
	(-0.43)	(-2.00)	(0.42)	(0.78)	(0.02)
$Similarity > 70\%$	0.00000165	-0.000119*	0.0000296	0.0000266	-0.00000120
	(0.07)	(-1.79)	(0.62)	(0.68)	(-0.03)
$Similarity > 80\%$	-0.00000505	-0.0000932	0.0000220	0.0000241	-0.00000421
	(-0.22)	(-1.37)	(0.47)	(0.58)	(-0.09)
$Similarity > 90\%$	-0.0000313	-0.0000234	-0.000238	0.0000735	-0.0000538
	(-0.37)	(-0.08)	(-1.39)	(0.62)	(-0.27)

t statistics in parentheses

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

Table 6.11: Effect of high frequency intraday news articles on stock returns - 30 minutes intervals. It could be the case that traders are triggered by repeated messages. Therefore, I provide a further check to study whether similar messages indeed affects stock prices and volatility. For this purpose, I construct a series of dummy variables that pick up the time intervals on which highly similar news articles exist, and add this to the regression. The model thus becomes $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j} + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}$ for the effect of overnight news and news similarity on stock returns. In this model, M_t takes the value of the Pessimism factor. The variables $D_{Similarity,i,t}$ are dummy variables that take the value one (1) least two similar news articles in the respective time interval. The news similarity cutoff is defined using five (5) different values: 50%, 60%, 70%, 80% and 90%. To calculate the similarity of articles, I compare the content of every story with articles released in the previous 30-minutes, using the Ratcliff/Obsershelp (Ratcliff and Metzener (1988)) pattern recognition algorithm which calculates the similarity of two strings. The dummy variable $Overnight_t$ is a dummy variable that takes the value one (1) if the return in the respective time interval is an overnight return.

	(1) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) US	(6) Japan	(7) China
Pessimism	-0.00688***	-0.0156***	-0.00453	-0.00974***	-0.0134***	-0.0160**	-0.0147*
	(-2.96)	(-2.79)	(-1.25)	(-3.27)	(-2.65)	(-2.19)	(-1.91)
$Overnight \times Pessimism$	0.00196	0.0152	0.0259	0.0137	-0.00134	-0.0263	0.0730**
	(0.22)	(0.88)	(1.52)	(1.23)	(-0.27)	(-0.47)	(2.01)
$D_{Similarity>50\%} \times Pessimism$	0.000712	-0.0166**	0.00111	0.00908**	0.00231	-0.0556	0.0111
	(0.26)	(-2.37)	(0.26)	(2.20)	(0.32)	(-1.33)	(0.24)
$D_{Similarity>60\%} \times Pessimism$	0.000643	-0.0154**	0.000851	0.00899**	0.00240	-0.0580	0.0174
	(0.23)	(-2.18)	(0.20)	(2.15)	(0.33)	(-1.40)	(0.38)
$D_{Similarity>70\%} \times Pessimism$	0.000806	-0.0160**	-0.0000896	0.00970**	0.00236	-0.0580	0.0265
	(0.29)	-2.23)	(-0.02)	(2.27)	(0.32)	(-1.40)	(0.55)
$D_{Similarity>80\%} \times Pessimism$	0.000835	-0.0155**	-0.00156	0.00963**	-0.0000113	-0.0564	0.0331
	(0.29)	(-2.15)	(-0.35)	(2.24)	(-0.00)	(-1.34)	(0.68)
$D_{Similarity>90\%} \times Pessimism$	0.00342	-0.0153**	-0.00152	0.00947**	-0.000754	-0.0553	0.0277
	(1.23)	(-2.07)	(-0.34)	(2.13)	(-0.10)	(-1.26)	(0.54)
$Overnight \times Return_{t-1}$	-0.0855	-0.153***	-0.00230	-0.0919**	0.00307	0.133	0.0195
	(-1.61)	(-2.83)	(-0.04)	(-2.46)	(0.03)	(1.24)	(0.22)
$Overnight \times Return_{t-2}$	-0.103**	-0.0439	-0.0351	-0.0742*	-0.0501	-0.00944	0.0695
	(-2.23)	(-1.16)	(-0.79)	(-1.68)	(-0.70)	(-0.11)	(0.66)
$Overnight \times Return_{t-3}$	-0.0168	-0.0218	-0.0157	0.0450	0.0813**	-0.160	0.0791
	(-0.46)	(-0.72)	(-0.39)	(1.41)	(2.42)	(-1.62)	(1.19)
$Overnight \times Return_{t-4}$	-0.0176	-0.00698	-0.00998	-0.0370	0.162***	0.0568	-0.0358
	(-0.33)	(-0.26)	(-0.25)	(-1.11)	(2.63)	(0.56)	(-0.54)
$Overnight \times Return_{t-5}$	-0.0701**	0.0257	-0.0414	-0.0402	0.237**	0.0539	0.0707
	(-2.08)	(1.07)	(-1.31)	(-1.32)	(2.53)	(0.34)	(0.75)
$Return_{t-1}$	0.0800***	0.0658**	0.0382	0.0223	-0.0543	-0.141	-0.0256
	(3.02)	(2.84)	(1.43)	(0.93)	(-0.88)	(-1.33)	(-0.30)
$Return_{t-2}$	0.0757**	0.0392*	-0.0253	0.0124	-0.0144	0.0494	-0.0698
	(2.45)	(1.85)	(-0.87)	(0.42)	(-0.39)	(0.57)	(-0.67)
$Return_{t-3}$	0.00515	0.00953	-0.000692	-0.0279	-0.0108	0.133	-0.0561
	(0.20)	(0.43)	(-0.03)	(-1.07)	(-0.36)	(1.34)	(-0.87)
$Return_{t-4}$	0.0162	-0.000813	0.00542	0.0219	-0.0667*	0.0352	0.0228
	(0.52)	(-0.04)	(0.25)	(0.81)	(-1.74)	(0.54)	(0.35)
$Return_{t-5}$	0.0270	0.00717	0.0240	0.0260	-0.106***	-0.0564	-0.0323
	(1.10)	(0.37)	(1.03)	(1.07)	(-2.77)	(-0.36)	(-0.35)

t statistics in parentheses
* p<.1, ** p<.05, *** p<.01

Table 6.12: Effect of high frequency intraday news articles on stock returns - 30 minutes intervals. It could be the case that traders are triggered by repeated messages. Therefore, I provide a further check to study whether similar messages indeed affects stock prices and volatility. For this purpose, I construct a series of dummy variables that pick up the time intervals on which highly similar news articles exist, and add this to the regression. The model thus becomes $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j} + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}$ for the effect of overnight news and news similarity on stock returns. In this model, M_t takes the value of the Average Pessimism factor. The variables $D_{Similarity,i,t}$ are dummy variables that take the value one (1) least two similar news articles in the respective time interval. The news similarity cutoff is defined using five (5) different values: 50%, 60%, 70%, 80% and 90%. To calculate the similarity of articles, I compare the content of every story with articles released in the previous 30-minutes, using the Ratcliff/Obsershelp (Ratcliff and Metzener (1988)) pattern recognition algorithm which calculates the similarity of two strings. The dummy variable $Overnight_t$ is a dummy variable that takes the value one (1) if the return in the respective time interval is an overnight return.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Portugal	Greece	Italy	Germany	US	Japan	China
Average Pessimism	-0.00545**	-0.00939	-0.00403	-0.00882***	-0.0116**	-0.0254***	-0.0277***
	(-2.36)	(-1.63)	(-1.04)	(-2.98)	(-2.06)	(-3.46)	(-4.32)
Overnight × *Average Pessimism*	0.00980	0.0185	0.0192	0.0156	-0.00218	-0.0148	0.0609
	(0.01)	(1.09)	(1.11)	(1.39)	(-0.43)	(-0.24)	(1.46)
$D_{Similarity>50\%}$ × *Average Pessimism*	0.0000315	-0.0182**	-0.000720	0.00537	0.000882	-0.0394	-0.00937
	(0.01)	(-2.42)	(-0.16)	(1.43)	(0.12)	(-0.89)	(-0.17)
$D_{Similarity>60\%}$ × *Average Pessimism*	-0.0000589	-0.0164**	-0.00137	0.00526	0.00121	-0.0470	0.000427
	(-0.02)	(-2.17)	(-0.30)	(1.38)	(0.16)	(-1.06)	(0.01)
$D_{Similarity>70\%}$ × *Average Pessimism*	0.000312	-0.0166**	-0.00217	0.00622	0.00114	-0.0470	0.0121
	(0.10)	(-2.16)	(-0.46)	(1.62)	(0.15)	(-1.06)	(0.21)
$D_{Similarity>80\%}$ × *Average Pessimism*	0.000346	-0.0164**	-0.00294	0.00592	-0.000810	-0.0449	0.0172
	(0.11)	(-2.11)	(-0.62)	(1.54)	(-0.10)	(-1.00)	(0.30)
$D_{Similarity>90\%}$ × *Average Pessimism*	0.00261	-0.0185**	-0.00297	0.00603	-0.00165	-0.0461	0.0175
	(0.86)	(-2.33)	(-0.62)	(1.54)	(-0.21)	(-0.98)	(0.29)
Overnight × $Return_{t-1}$	-0.0842	-0.150***	-0.00260	-0.0920**	0.00550	0.131	0.0198
	(-1.58)	(-2.75)	(-0.04)	(-2.47)	(0.06)	(1.22)	(0.23)
Overnight × $Return_{t-2}$	-0.102**	-0.0447	-0.0350	-0.0738*	-0.0504	-0.0108	0.0749
	(-2.21)	(-1.16)	(-0.79)	(-1.67)	(-0.71)	(-0.12)	(0.71)
Overnight × $Return_{t-3}$	-0.0169	-0.0243	-0.0155	0.0447	0.0800**	-0.160	0.0772
	(-0.46)	(-0.80)	(-0.38)	(1.40)	(2.38)	(-1.62)	(1.17)
Overnight × $Return_{t-4}$	-0.0185	-0.00590	-0.00995	-0.0367	0.163***	0.0568	-0.0361
	(-0.35)	(-0.22)	(-0.25)	(-1.10)	(2.66)	(0.56)	(-0.54)
Overnight × $Return_{t-5}$	-0.0709**	0.0269	-0.0413	-0.0409	0.239**	0.0544	0.0718
	(-2.10)	(1.11)	(-1.31)	(-1.34)	(2.55)	(0.34)	(0.76)
$Return_{t-1}$	0.0801***	0.0655***	0.380	0.0225	-0.0548	-0.140	-0.0261
	(3.02)	(2.83)	(1.42)	(0.94)	(-0.88)	(-1.32)	(-0.30)
$Return_{t-2}$	0.0752**	0.0397	-0.0250	0.0122	-0.0159	0.0481	-0.0744
	(2.43)	(1.86)	(-0.86)	(0.41)	(-0.43)	(0.55)	(-0.72)
$Return_{t-3}$	0.00518	0.0114*	-0.000547	-0.0281	-0.0106	0.134	-0.0572
	(0.21)	(0.51)	(-0.002)	(-1.07)	(-0.35)	(1.35)	(0.89)
$Return_{t-4}$	0.0167	-0.00131	0.00522	0.0222	-0.0675*	0.0348	0.0217
	(0.54)	(-0.07)	(0.25)	(0.82)	(-1.76)	(0.53)	(0.34)
$Return_{t-5}$	0.0277	0.00710	0.0237	0.0267	-0.106***	-0.0575	-0.0335
	(1.12)	(0.36)	(1.02)	(1.10)	(-2.78)	(-0.37)	(-0.36)

t statistics in parentheses

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

Table 6.13: Effect of high frequency intraday news articles on stock returns - 30 minutes intervals. It could be the case that traders are triggered by repeated messages. Therefore, I provide a further check to study whether similar messages indeed affects stock prices and volatility. For this purpose, I construct a series of dummy variables that pick up the time intervals on which highly similar news articles exist, and add this to the regression. The model thus becomes $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j} + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}$ for the effect of overnight news and news similarity on stock returns. In this model, M_t takes the value of the Positive factor. The variables $D_{Similarity,i,t}$ are dummy variables that take the value one (1) least two similar news articles in the respective time interval. The news similarity cutoff is defined using five (5) different values: 50%, 60%, 70%, 80% and 90%. To calculate the similarity of articles, I compare the content of every story with articles released in the previous 30-minutes, using the Ratcliff/Obsershelp (Ratcliff and Metzener (1988)) pattern recognition algorithm which calculates the similarity of two strings. The dummy variable $Overnight_t$ is a dummy variable that takes the value one (1) if the return in the respective time interval is an overnight return.

	(1) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) US	(6) Japan	(7) China
$Positive$	0.0146**	0.0520***	0.0244**	0.0169**	0.0127	0.0264*	0.0432**
	(2.15)	(3.50)	(2.47)	(2.41)	(1.18)	(1.66)	(2.10)
$Overnight \times Positive$	0.0650***	0.0409	0.0303	0.0877***	-0.0391**	0.250	0.258*
	(3.02)	(0.76)	(0.64)	(2.80)	(-2.17)	(1.17)	(1.72)
$D_{Similarity>50\%} \times Positive$	-0.00495	-0.0496**	-0.0252*	0.00463	0.0212	0.0632	0.0204
	(-0.49)	(-1.96)	(-1.83)	(0.46)	(1.34)	(0.59)	(0.22)
$D_{Similarity>60\%} \times Positive$	-0.00611	-0.0487*	-0.0228*	0.00531	0.0214	0.0438	0.0396
	(-0.60)	(-1.90)	(-1.66)	(0.52)	(1.35)	(0.41)	(0.42)
$D_{Similarity>70\%} \times Positive$	-0.00490	-0.0516**	-0.0255*	0.00505	0.0199	0.0438	0.0582
	(-0.47)	(-1.97)	(-1.83)	(0.49)	(1.23)	(0.41)	(0.63)
$D_{Similarity>80\%} \times Positive$	-0.00676	-0.0504*	-0.0257*	0.00450	0.00932	0.0466	0.0681
	(-0.64)	(-1.91)	(-1.77)	(0.43)	(0.58)	(0.44)	(0.72)
$D_{Similarity>90\%} \times Positive$	0.000940	-0.0633**	-0.0230	0.00659	0.0138	0.0416	0.0454
	(0.09)	(-2.46)	(-1.54)	(0.64)	(0.79)	(0.38)	(0.49)
$Overnight \times Return_{t-1}$	-0.0852	-0.152***	-0.00251	-0.0891**	-0.00148	0.134	0.0218
	(-1.60)	(-2.75)	(-0.04)	(-2.39)	(-0.02)	(1.24)	(0.25)
$Overnight \times Return_{t-2}$	-0.103**	-0.0436	-0.0367	-0.0741*	-0.0533	-0.00421	0.0688
	(-2.24)	(-1.14)	(-0.82)	(-1.68)	(-0.76)	(-0.05)	(0.65)
$Overnight \times Return_{t-3}$	-0.0172	-0.0250	-0.0160	0.0418	0.0846**	-0.153	0.0749
	(-0.47)	(-0.82)	(-0.40)	(1.30)	(2.48)	(-1.54)	(1.14)
$Overnight \times Return_{t-4}$	-0.0209	-0.00659	-0.0119	-0.0393	0.163***	0.0566	-0.0346
	(-0.39)	(-0.25)	(-0.29)	(-1.19)	(2.63)	(0.55)	(-0.52)
$Overnight \times Return_{t-5}$	-0.0684**	0.0270	-0.0418	-0.0401	0.240**	0.0586	0.0697
	(-2.02)	(1.13)	(-1.33)	(-1.32)	(2.56)	(0.37)	(0.73)
$Return_{t-1}$	0.0798***	0.0659***	0.0371	0.0211	-0.0544	-0.138	-0.0257
	(3.00)	(2.86)	(1.39)	(0.88)	(-0.87)	(-1.30)	(-0.30)
$Return_{t-2}$	0.0753**	0.0393*	-0.0253	0.0128	-0.0145	0.0458	-0.0718
	(2.44)	(1.85)	(-0.87)	(0.43)	(-0.39)	(0.53)	(-0.69)
$Return_{t-3}$	0.00486	0.0116	-0.000104	-0.0282	-0.00913	0.128	-0.0542
	(0.19)	(0.53)	(-0.00)	(-1.08)	(-0.30)	(1.29)	(-0.85)
$Return_{t-4}$	0.0179	-0.00128	0.00534	0.0228	-0.0650*	0.0360	0.0215
	(0.58)	(-0.07)	(0.25)	(0.84)	(-1.69)	(0.55)	(0.33)
$Return_{t-5}$	0.0263	0.00599	0.0240	0.0271	-0.104***	-0.0582	-0.0307
	(1.07)	(0.30)	(1.04)	(1.12)	(-2.71)	(-0.37)	(-0.33)

t statistics in parentheses

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

Table 6.14: Effect of high frequency intraday news articles on stock returns - 30 minutes intervals. It could be the case that traders are triggered by repeated messages. Therefore, I provide a further check to study whether similar messages indeed affects stock prices and volatility. For this purpose, I construct a series of dummy variables that pick up the time intervals on which highly similar news articles exist, and add this to the regression. The model thus becomes $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j} + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}$ for the effect of overnight news and news similarity on stock returns. In this model, M_t takes the value of the Negative factor. The variables $D_{Similarity,i,t}$ are dummy variables that take the value one (1) least two similar news articles in the respective time interval. The news similarity cutoff is defined using five (5) different values: 50%, 60%, 70%, 80% and 90%. To calculate the similarity of articles, I compare the content of every story with articles released in the previous 30-minutes, using the Ratcliff/Obsershelp (Ratcliff and Metzener (1988)) pattern recognition algorithm which calculates the similarity of two strings. The dummy variable $Overnight_t$ is a dummy variable that takes the value one (1) if the return in the respective time interval is an overnight return.

	(1) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) US	(6) Japan	(7) China
$Negative$	-0.00536**	-0.0120**	-0.00126	-0.00722**	-0.0144***	-0.0152*	-0.0101
	(-2.17)	(-2.04)	(-0.33)	(-2.20)	(-2.59)	(-1.89)	(-1.10)
$Overnight \times Negative$	0.00647	0.0134	0.0202	0.0158*	-0.00357	-0.00636	0.0627**
	(0.88)	(0.94)	(1.45)	(1.80)	(-0.85)	(-0.14)	(2.10)
$D_{Similarity>50\%} \times Negative$	0.000191	-0.0157**	-0.00110	0.00692**	0.00362	-0.0369	0.00892
	(0.08)	(-2.57)	(-0.29)	(2.05)	(0.62)	(-1.01)	(0.24)
$D_{Similarity>60\%} \times Negative$	0.0000529	-0.0148**	-0.00112	0.00690**	0.00372	-0.0407	0.0149
	(0.02)	(-2.39)	(-0.29)	(2.02)	(0.63)	(-1.12)	(0.40)
$D_{Similarity>70\%} \times Negative$	0.000271	-0.0154**	-0.00201	0.00741**	0.00353	-0.0407	0.0228
	(0.11)	(-2.45)	(-0.52)	(2.13)	(0.59)	(-1.12)	(0.59)
$D_{Similarity>80\%} \times Negative$	0.000132	-0.0150**	-0.00309	0.00732**	0.000868	-0.0392	0.0283
	(0.05)	(-2.37)	(-0.79)	(2.09)	(0.14)	(-1.07)	(0.71)
$D_{Similarity>90\%} \times Negative$	0.00262	-0.0157**	-0.00286	0.00738**	0.000640	-0.0393	0.0229
	(1.06)	(-2.41)	(-0.72)	(2.05)	(0.11)	(-1.02)	(0.55)
$Overnight \times Return_{t-1}$	-0.0846	-0.151***	-0.00280	-0.0910**	0.00470	0.132	0.0188
	(-1.59)	(-2.78)	(-0.05)	(-2.44)	(0.05)	(1.23)	(0.21)
$Overnight \times Return_{t-2}$	-0.102**	-0.0444	-0.0352	-0.0728*	-0.0507	-0.00760	0.0688
	(-2.22)	(-1.16)	(-0.79)	(-1.65)	(-0.71)	(-0.09)	(0.65)
$Overnight \times Return_{t-3}$	-0.0161	-0.0219	-0.0145	0.0450	0.0823**	-0.161	0.0809
	(-0.44)	(-0.72)	(-0.36)	(1.41)	(2.43)	(-1.62)	(1.22)
$Overnight \times Return_{t-4}$	-0.0186	-0.00622	-0.00971	-0.0373	0.163***	0.0602	-0.0347
	(-0.35)	(-0.23)	(-0.24)	(-1.12)	(2.60)	(0.59)	(-0.52)
$Overnight \times Return_{t-5}$	-0.0699**	0.0266	-0.0419	-0.0391	0.237**	0.0552	0.0712
	(-2.07)	(1.10)	(-1.33)	(-1.29)	(2.55)	(0.35)	(0.75)
$Return_{t-1}$	0.0803***	0.0663***	0.0382	0.0225	-0.0531	-0.141	-0.0244
	(3.03)	(2.86)	(1.43)	(0.94)	(-0.86)	(-1.32)	(-0.28)
$Return_{t-2}$	0.0756**	0.0397*	-0.0249	0.0123	-0.0137	0.0477	-0.0682
	(2.44)	(1.87)	(-0.85)	(0.42)	(-0.37)	(0.55)	(-0.66)
$Return_{t-3}$	0.00499	0.01000	-0.000349	-0.0275	-0.0106	0.135	-0.0561
	(0.20)	(0.45)	(-0.01)	(-1.05)	(-0.35)	(1.35)	(-0.87)
$Return_{t-4}$	0.0171	-0.000620	0.00544	0.0229	-0.0664*	0.0327	0.0223
	(0.55)	(-0.03)	(0.26)	(0.84)	(-1.73)	(0.50)	(0.35)
$Return_{t-5}$	0.0276	0.00708	0.0244	0.0267	-0.105***	-0.0565	-0.0317
	(1.12)	(0.36)	(1.05)	(1.10)	(-2.75)	(-0.36)	(-0.34)

t statistics in parentheses

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

Table 6.15: Effect of high frequency intraday news articles on stock returns - 30 minutes intervals. It could be the case that traders are triggered by repeated messages. Therefore, I provide a further check to study whether similar messages indeed affects stock prices and volatility. For this purpose, I construct a series of dummy variables that pick up the time intervals on which highly similar news articles exist, and add this to the regression. The model thus becomes $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j} + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}$ for the effect of overnight news and news similarity on stock returns. In this model, M_t takes the value of the News Count factor. The variables $D_{Similarity,i,t}$ are dummy variables that take the value one (1) least two similar news articles in the respective time interval. The news similarity cutoff is defined using five (5) different values: 50%, 60%, 70%, 80% and 90%. To calculate the similarity of articles, I compare the content of every story with articles released in the previous 30-minutes, using the Ratcliff/Obsershelp (Ratcliff and Metzener (1988)) pattern recognition algorithm which calculates the similarity of two strings. The dummy variable $Overnight_t$ is a dummy variable that takes the value one (1) if the return in the respective time interval is an overnight return.

	(1) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) US	(6) Japan	(7) China
$News\,Count$	-0.0000100	0.0000422**	-0.00000763	0.00000234	0.000000270***	0.00000487	-0.000000227
	(-0.35)	(2.05)	(-0.31)	(0.14)	(3.03)	(0.32)	(-0.89)
$Overnight \times News\,Count$	-0.00000578	0.0000121	0.0000934	0.0000537	-0.00000855	-0.000285	0.000863
	(-0.09)	(0.19)	(0.85)	(0.72)	(-0.44)	(-0.68)	(0.54)
$D_{Similarity>50\%} \times News\,Count$	0.0000335	-0.00000656	-0.00000510	0.0000465	-0.00000135	-0.0000192	0.00000780
	(0.91)	(-0.11)	(-0.15)	(0.90)	(-0.20)	(-0.79)	(0.45)
$D_{Similarity>60\%} \times News\,Count$	0.0000339	-0.00000228	-0.00000701	0.0000459	-0.00000135	-0.0000200	0.00000975
	(0.92)	(-0.04)	(-0.21)	(0.88)	(-0.20)	(-0.81)	(0.58)
$D_{Similarity>70\%} \times News\,Count$	0.0000336	-0.000000342	-0.00000657	0.0000411	-0.00000141	-0.0000200	0.0000131
	(0.90)	(-0.01)	(-0.20)	(0.78)	(-0.20)	(-0.81)	(0.81)
$D_{Similarity>80\%} \times News\,Count$	0.0000349	0.00000159	-0.0000233	0.0000430	-0.00000309	-0.0000190	0.0000121
	(0.93)	(0.03)	(-0.68)	(0.82)	(-0.44)	(-0.77)	(0.73)
$D_{Similarity>90\%} \times News\,Count$	0.0000570	-0.00000111	-0.00000782	0.0000439	-0.00000384	-0.0000196	0.0000133
	(1.59)	(-0.02)	(-0.24)	(0.82)	(-0.54)	(-0.79)	(0.80)
$Overnight \times Return_{t-1}$	-0.0847	-0.150***	-0.00164	-0.0892**	0.00167	0.135	0.0207
	(-1.58)	(-2.75)	(-0.03)	(-2.39)	(0.02)	(1.25)	(0.23)
$Overnight \times Return_{t-2}$	-0.101**	-0.0430	-0.0351	-0.0745*	-0.0544	-0.00288	0.0679
	(-2.21)	(-1.13)	(-0.79)	(-1.71)	(-0.78)	(-0.03)	(0.64)
$Overnight \times Return_{t-3}$	-0.0163	-0.0207	-0.0143	0.0449	0.0827**	-0.157	0.0809
	(-0.45)	(-0.68)	(-0.36)	(1.41)	(2.42)	(-1.58)	(1.22)
$Overnight \times Return_{t-4}$	-0.0175	-0.00520	-0.0102	-0.0360	0.165***	0.0621	-0.0336
	(-0.33)	(-0.20)	(-0.25)	(-1.08)	(2.66)	(0.60)	(-0.50)
$Overnight \times Return_{t-5}$	-0.0690**	0.0250	-0.0411	-0.0375	0.240***	0.0574	0.0725
	(-2.03)	(1.04)	(-1.30)	(-1.24)	(2.62)	(0.36)	(0.77)
$Return_{t-1}$	0.0803***	0.0655***	0.0382	0.0225	-0.0542	-0.141	-0.0248
	(3.04)	(2.83)	(1.43)	(0.94)	(-0.87)	(-1.32)	(-0.29)
$Return_{t-2}$	0.0756**	0.0413*	-0.0247	0.0127	-0.0136	0.0446	-0.0677
	(2.45)	(1.94)	(-0.85)	(0.44)	(-0.36)	(0.51)	(-0.65)
$Return_{t-3}$	0.00576	0.00990	0.000169	-0.0259	-0.00863	0.131	-0.0556
	(0.23)	(0.45)	(0.01)	(-0.99)	(-0.29)	(1.32)	(-0.87)
$Return_{t-4}$	0.0169	-0.000272	0.00569	0.0228	-0.0655*	0.0319	0.0217
	(0.55)	(-0.01)	(0.27)	(0.84)	(-1.70)	(0.48)	(0.34)
$Return_{t-5}$	0.0290	0.00925	0.0240	0.0278	-0.104***	-0.0557	-0.0315
	(1.17)	(0.47)	(1.03)	(1.16)	(-2.70)	(-0.35)	(-0.34)

t statistics in parentheses

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

Table 6.16: Effect of high frequency intraday news articles similarity on volatility (proxied by squared returns) - 30 minutes intervals. I also control for overnight returns and news similarity. The model thus becomes: $R_t^2 = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j}^2 + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}^2$. In this model, M_t takes the value of the Pessimism factor. 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/Obsershelp (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.

	(1) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) US	(6) Japan	(7) China
$Pessimism$	-0.0000472	-0.0000840	0.0000643	0.000128***	0.000266***	0.0000730	-0.000146
	(-0.83)	(-0.48)	(0.99)	(2.68)	(3.12)	(0.34)	(-0.96)
$Overnight \times Pessimism$	0.000760***	0.00139*	0.00111***	0.000515***	-0.000403***	0.00108*	-0.000476*
	(5.06)	(1.96)	(2.99)	(3.05)	(-5.75)	(1.71)	(-1.89)
$D_{Similarity>50\%} \times Pessimism$	-0.0000778***	-0.000299**	-0.000173***	-0.0000710	-0.000182*	0.00252***	0.00253**
	(-2.63)	(1.96)	(-2.75)	(-0.82)	(-1.89)	(4.26)	(2.17)
$D_{Similarity>60\%} \times Pessimism$	-0.0000736**	-0.000286**	-0.000177***	-0.0000685	-0.000184*	0.00250***	0.00247**
	(-2.51)	(-2.19)	(-2.79)	(-0.78)	(-1.89)	(4.20)	(2.10)
$D_{Similarity>70\%} \times Pessimism$	-0.0000713**	-0.000298**	-0.000182***	-0.0000671	-0.000178*	0.00250***	0.00254**
	(-2.43)	(-2.26)	(-2.86)	(-0.74)	(-1.82)	(4.20)	(2.07)
$D_{Similarity>80\%} \times Pessimism$	-0.0000622**	-0.000296**	-0.000185***	-0.0000772	-0.000178*	0.00251***	0.00256**
	(-2.14)	(-2.22)	(-2.89)	(-0.84)	(-1.81)	(4.20)	(2.01)
$D_{Similarity>90\%} \times Pessimism$	-0.0000874***	-0.000269*	-0.000167**	-0.0000740	-0.000221**	0.00260***	0.00267**
	(-3.16)	(-1.94)	(-2.47)	(-0.77)	(-2.28)	(4.19)	(2.02)
$Overnight \times Return_{t-1}^2$	-0.192*	-0.0504	-0.0718	-0.0640*	-0.320**	-3.026***	-0.368
	(-1.73)	(-0.95)	(-0.50)	(-1.70)	(-2.09)	(-4.48)	(-1.24)
$Overnight \times Return_{t-2}^2$	-0.335	-0.0494	-0.243*	-0.317**	0.284**	-0.144	-2.936*
	(-1.45)	(-1.54)	(-1.94)	(-2.56)	(1.97)	(-0.39)	(-1.82)
$Overnight \times Return_{t-3}^2$	-0.0939	-0.0773**	-0.243***	-0.138**	-0.0952	-1.221	-0.138
	(-0.93)	(-2.31)	(-3.12)	(-2.48)	(-1.35)	(-1.31)	(-0.62)
$Overnight \times Return_{t-4}^2$	-0.331	-0.0359**	0.00185	-0.239***	0.00324	0.177	0.121
	(-1.36)	(-2.27)	(0.03)	(-3.03)	(0.06)	(0.48)	(0.59)
$Overnight \times Return_{t-5}^2$	-0.0825	-0.0326*	-0.0980**	-0.0892	-0.0609	-0.601***	-1.270**
	(-1.05)	(-1.81)	(-2.16)	(-1.47)	(-1.16)	(-7.20)	(-2.45)
$Return_{t-1}^2$	0.224**	0.0665*	0.172*	0.0472	0.291*	3.014***	0.255
	(2.15)	(1.70)	(1.66)	(1.27)	(1.92)	(4.40)	(0.81)
$Return_{t-2}^2$	0.330	0.0437	0.242*	0.328***	-0.273*	0.0146	2.900*
	(1.45)	(1.48)	(1.93)	(2.66)	(-1.88)	(0.04)	(1.78)
$Return_{t-3}^2$	0.0798	0.0576*	0.225***	0.110**	0.0860	1.163	0.0613
	(0.76)	(1.80)	(3.14)	(1.97)	(1.24)	(1.22)	(0.27)
$Return_{t-4}^2$	0.351	0.0174	-0.0117	0.221***	-0.00233	0.219	-0.259
	(1.44)	(1.15)	(-0.21)	(2.78)	(-0.05)	(0.87)	(-1.18)
$Return_{t-5}^2$	0.0385	0.0203	0.0770*	0.0677	0.0381	0.475***	1.218**
	(0.40)	(1.14)	(1.81)	(1.11)	(1.32)	(6.46)	(2.33)

t statistics in parentheses

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

Table 6.17: Effect of high frequency intraday news articles similarity on volatility (proxied by squared returns) - 30 minutes intervals. I also control for overnight returns and news similarity. The model thus becomes: $R_t^2 = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j}^2 + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}^2$. In this model, M_t takes the value of the Average Pessimism factor. 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/Obsershelp (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.

	(1) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) US	(6) Japan	(7) China
$Average\,Pessimism$	0.0000110	0.000112	0.000129*	0.000123***	0.000246***	0.000348*	0.0000723
	(0.39)	(0.61)	(1.68)	(2.86)	(2.79)	(1.65)	(0.55)
$Overnight \times Average\,Pessimism$	0.000716***	0.00158**	0.00129***	0.000629***	-0.000397***	0.00121*	-0.000328
	(5.70)	(2.15)	(2.67)	(3.87)	(-5.83)	(1.86)	(-1.03)
$D_{Similarity>50\%} \times Average\,Pessimism$	-0.0000994**	-0.000355**	-0.000187***	-0.000129**	-0.000173*	0.00327***	0.00330**
	(-2.43)	(-2.15)	(-2.64)	(-2.41)	(-1.78)	(3.15)	(2.38)
$D_{Similarity>60\%} \times Average\,Pessimism$	-0.0000937**	-0.000343**	-0.000194***	-0.000127**	-0.000176*	0.00323***	0.00322**
	(-2.34)	(-2.06)	(-2.73)	(-2.36)	(-1.79)	(3.06)	(2.28)
$D_{Similarity>70\%} \times Average\,Pessimism$	-0.0000891**	-0.000348**	-0.000196***	-0.000133**	-0.000169*	0.00323***	0.00321**
	(-2.27)	(-2.07)	(-2.73)	(-2.45)	(-1.69)	(3.06)	(2.21)
$D_{Similarity>80\%} \times Average\,Pessimism$	-0.0000838**	-0.000339**	-0.000196***	-0.000142***	-0.000169*	0.00327***	0.00319**
	(-2.17)	(-1.98)	(-2.71)	(-2.65)	(-1.68)	(3.06)	(2.15)
$D_{Similarity>90\%} \times Average\,Pessimism$	-0.000102***	-0.000335*	-0.000181**	-0.000140**	-0.000215**	0.00339***	0.00333**
	(-2.66)	(-1.90)	(-2.43)	(-2.53)	(-2.19)	(3.03)	(2.16)
$Overnight \times Return_{t-1}^2$	-0.192*	-0.0493	-0.0722	-0.0647*	-0.316**	-3.029***	-0.357
	(-1.73)	(-0.93)	(-0.50)	(-1.72)	(-2.08)	(-4.49)	(-1.23)
$Overnight \times Return_{t-2}^2$	-0.332	-0.0475	-0.243*	-0.316**	0.284**	-0.132	-2.918*
	(-1.44)	(-1.51)	(-1.94)	(-2.56)	(1.98)	(-0.36)	(-1.81)
$Overnight \times Return_{t-3}^2$	-0.0951	-0.0743**	-0.242***	-0.138**	-0.0950	-1.215	-0.146
	(-0.93)	(-2.22)	(-3.11)	(-2.49)	(-1.35)	(-1.30)	(-0.66)
$Overnight \times Return_{t-4}^2$	-0.333	-0.0357**	0.00162	-0.238***	0.00358	0.158	0.128
	(-1.37)	(-2.25)	(0.03)	(-3.04)	(0.07)	(0.44)	(0.62)
$Overnight \times Return_{t-5}^2$	-0.0838	-0.0317*	-0.0971**	-0.0888	-0.0603	-0.596***	-1.259**
	(-1.08)	(-1.76)	(-2.15)	(-1.48)	(-1.15)	(-7.10)	(-2.45)
$Return_{t-1}^2$	0.224**	0.0668*	0.172*	0.0480	0.291*	3.019***	0.244
	(2.15)	(1.69)	(1.66)	(1.29)	(1.93)	(4.41)	(0.79)
$Return_{t-2}^2$	0.327	0.0423	0.242*	0.327***	-0.273*	0.00420	2.884*
	(1.44)	(1.46)	(1.93)	(2.65)	(-1.89)	(0.01)	(1.77)
$Return_{t-3}^2$	0.0809	0.0550*	0.225***	0.110**	0.0858	1.159	0.0701
	(0.77)	(1.72)	(3.14)	(1.96)	(1.24)	(1.22)	(0.31)
$Return_{t-4}^2$	0.352	0.0175	-0.0113	0.219***	-0.00281	0.238	-0.264
	(1.45)	(1.16)	(-0.20)	(2.77)	(-0.06)	(0.95)	(-1.21)
$Return_{t-5}^2$	0.0393	0.0202	0.0761*	0.0676	0.0369	0.476***	1.208**
	(0.41)	(1.14)	(1.79)	(1.12)	(1.28)	(6.50)	(2.33)

t statistics in parentheses

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

Table 6.18: Effect of high frequency intraday news articles similarity on volatility (proxied by squared returns) - 30 minutes intervals. I also control for overnight returns and news similarity. The model thus becomes: $R_t^2 = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j}^2 + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}^2$. In this model, M_t takes the value of the Positive factor. 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/Obsershelp (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.

	(1) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) US	(6) Japan	(7) China
$Positive$	0.000100	0.000616	-0.0000732	-0.000101	-0.000129	-0.000350	-0.000927**
	(0.44)	(1.31)	(-0.34)	(-1.03)	(-0.84)	(-1.10)	(-2.19)
$Overnight \times Positive$	0.00140***	0.00420*	0.00287***	0.00150***	-0.000598***	0.00222	-0.00175
	(4.68)	(1.82)	(2.77)	(3.43)	(-2.79)	(1.13)	(-1.27)
$D_{Similarity>50\%} \times Positive$	-0.000197	-0.000840	-0.000240	-0.000315***	-0.000388**	0.00483***	0.00334**
	(-0.93)	(-1.32)	(-1.16)	(-3.03)	(-2.24)	(2.72)	(2.34)
$D_{Similarity>60\%} \times Positive$	-0.000177	-0.000767	-0.000282	-0.000318***	-0.000385**	0.00460**	0.00311**
	(-0.85)	(-1.19)	(-1.36)	(-3.03)	(-2.21)	(2.57)	(2.22)
$D_{Similarity>70\%} \times Positive$	-0.000165	-0.000697	-0.000281	-0.000312***	-0.000382**	0.00460**	0.00288**
	(-0.81)	(-1.06)	(-1.34)	(-2.96)	(-2.17)	(2.57)	(2.10)
$D_{Similarity>80\%} \times Positive$	-0.000141	-0.000695	-0.000257	-0.000304***	-0.000447**	0.00461**	0.00267*
	(-0.69)	(-1.05)	(-1.19)	(-2.87)	(-2.53)	(2.57)	(1.82)
$D_{Similarity>90\%} \times Positive$	-0.000187	-0.000839	-0.000216	-0.000322***	-0.000441**	0.00463**	0.00248*
	(-0.93)	(-1.26)	(-0.94)	(-3.01)	(-2.35)	(2.51)	(1.84)
$Overnight \times Return_{t-1}^2$	-0.190*	-0.0519	-0.0722	-0.0630*	-0.320**	-3.032***	-0.358
	(-1.68)	(-0.97)	(-0.50)	(-1.67)	(-2.09)	(-4.47)	(-1.18)
$Overnight \times Return_{t-2}^2$	-0.326	-0.0515	-0.241*	-0.315**	0.284**	-0.152	-2.956*
	(-1.41)	(-1.61)	(-1.92)	(-2.54)	(1.97)	(-0.41)	(-1.84)
$Overnight \times Return_{t-3}^2$	-0.0984	-0.0761**	-0.243***	-0.139**	-0.0986	-1.222	-0.142
	(-0.96)	(-2.28)	(-3.11)	(-2.50)	(-1.40)	(-1.30)	(-0.64)
$Overnight \times Return_{t-4}^2$	-0.333	-0.0355**	0.00232	-0.240***	0.00347	0.130	0.133
	(-1.37)	(-2.30)	(0.04)	(-3.04)	(0.07)	(0.36)	(0.63)
$Overnight \times Return_{t-5}^2$	-0.0848	-0.0342*	-0.0997**	-0.0891	-0.0638	-0.604***	-1.303**
	(-1.08)	(-1.85)	(-2.20)	(-1.47)	(-1.19)	(-7.19)	(-2.49)
$Return_{t-1}^2$	0.222**	0.0684*	0.172*	0.0473	0.291*	3.019***	0.240
	(2.11)	(1.71)	(1.66)	(1.27)	(1.92)	(4.38)	(0.75)
$Return_{t-2}^2$	0.321	0.0439	0.241*	0.327***	-0.273*	0.0192	2.919*
	(1.41)	(1.50)	(1.92)	(2.65)	(-1.88)	(0.05)	(1.80)
$Return_{t-3}^2$	0.0841	0.0562*	0.226***	0.110**	0.0897	1.164	0.0609
	(0.79)	(1.77)	(3.14)	(1.97)	(1.31)	(1.21)	(0.27)
$Return_{t-4}^2$	0.353	0.0175	-0.0125	0.222***	-0.00363	0.265	-0.273
	(1.45)	(1.19)	(-0.22)	(2.79)	(-0.07)	(1.04)	(-1.20)
$Return_{t-5}^2$	0.0402	0.0227	0.0785*	0.0681	0.0397	0.474***	1.247**
	(0.41)	(1.25)	(1.84)	(1.11)	(1.36)	(6.43)	(2.37)

t statistics in parentheses

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

Table 6.19: Effect of high frequency intraday news articles similarity on volatility (proxied by squared returns) - 30 minutes intervals. I also control for overnight returns and news similarity. The model thus becomes: $R_t^2 = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j}^2 + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}^2$. In this model, M_t takes the value of the Negative factor. 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/Obsershelp (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.

	(1) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) US	(6) Japan	(7) China
$Negative$	-0.0000199	0.0000281	0.0000706	0.000144***	0.000258***	0.0000845	-0.000287
	(-0.47)	(0.18)	(1.09)	(2.69)	(2.84)	(0.37)	(-1.36)
$Overnight \times Negative$	0.000626***	0.00125**	0.000974***	0.000463***	-0.000335***	0.000863*	-0.000396*
	(5.43)	(2.10)	(3.08)	(3.47)	(-5.75)	(1.69)	(-1.65)
$D_{Similarity>50\%} \times Negative$	-0.0000705**	-0.000282**	-0.000144***	-0.0000756	-0.000164**	0.00230***	0.00208**
	(-2.30)	(-2.25)	(-2.62)	(-1.13)	(-2.10)	(4.75)	(2.31)
$D_{Similarity>60\%} \times Negative$	-0.0000660**	-0.000267**	-0.000150***	-0.0000739	-0.000165**	0.00227***	0.00202**
	(-2.20)	(-2.12)	(-2.71)	(-1.08)	(-2.10)	(4.66)	(2.23)
$D_{Similarity>70\%} \times Negative$	-0.0000634**	-0.000271**	-0.000154***	-0.0000727	-0.000160**	0.00227***	0.00205**
	(-2.15)	(-2.10)	(-2.77)	(-1.03)	(-2.03)	(4.66)	(2.17)
$D_{Similarity>80\%} \times Negative$	-0.0000550*	-0.000271**	-0.000154***	-0.0000795	-0.000164**	0.00229***	0.00207**
	(-1.87)	(-2.07)	(-2.73)	(-1.12)	(-2.05)	(4.65)	(2.08)
$D_{Similarity>90\%} \times Negative$	-0.0000760***	-0.000260*	-0.000137**	-0.0000791	-0.000192**	0.00237***	0.00215**
	(-2.65)	(-1.91)	(-2.27)	(-1.07)	(-2.44)	(4.64)	(2.09)
$Overnight \times Return_{t-1}^2$	-0.192*	-0.0502	-0.0724	-0.0649*	-0.324**	-3.029***	-0.371
	(-1.73)	(-0.95)	(-0.50)	(-1.73)	(-2.11)	(-4.48)	(-1.25)
$Overnight \times Return_{t-2}^2$	-0.335	-0.0493	-0.243*	-0.317**	0.284**	-0.148	-2.937*
	(-1.45)	(-1.55)	(-1.94)	(-2.57)	(1.97)	(-0.40)	(-1.82)
$Overnight \times Return_{t-3}^2$	-0.0953	-0.0762**	-0.242***	-0.138**	-0.0955	-1.222	-0.139
	(-0.94)	(-2.26)	(-3.11)	(-2.49)	(-1.34)	(-1.31)	(-0.63)
$Overnight \times Return_{t-4}^2$	-0.331	-0.0359**	0.00231	-0.239***	0.00386	0.174	0.116
	(-1.35)	(-2.27)	(0.04)	(-3.05)	(0.07)	(0.48)	(0.56)
$Overnight \times Return_{t-5}^2$	-0.0829	-0.0324*	-0.0971**	-0.0893	-0.0714	-0.599***	-1.269**
	(-1.06)	(-1.80)	(-2.15)	(-1.48)	(-1.37)	(-7.17)	(-2.45)
$Return_{t-1}^2$	0.224**	0.0669*	0.172*	0.0481	0.291*	3.018***	0.258
	(2.16)	(1.70)	(1.66)	(1.30)	(1.92)	(4.41)	(0.82)
$Return_{t-2}^2$	0.329	0.0433	0.242*	0.328***	-0.273*	0.0194	2.901*
	(1.45)	(1.47)	(1.93)	(2.66)	(-1.88)	(0.05)	(1.78)
$Return_{t-3}^2$	0.0813	0.0567*	0.225***	0.110**	0.0866	1.166	0.0627
	(0.78)	(1.76)	(3.13)	(1.96)	(1.25)	(1.22)	(0.28)
$Return_{t-4}^2$	0.350	0.0177	-0.0120	0.220***	-0.00264	0.223	-0.254
	(1.44)	(1.17)	(-0.21)	(2.79)	(-0.05)	(0.90)	(-1.15)
$Return_{t-5}^2$	0.0388	0.0207	0.0764*	0.0680	0.0390	0.475**	1.217**
	(0.40)	(1.16)	(1.80)	(1.12)	(1.34)	(6.46)	(2.33)

t statistics in parentheses

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

Table 6.20: Effect of high frequency intraday news articles similarity on volatility (proxied by squared returns) - 30 minutes intervals. I also control for overnight returns and news similarity. The model thus becomes: $R_t^2 = a_1 + b_1 M_t + \sum_{i=1}^{5} b_i D_{Similarity,i,t} M_t + b_3 Overnight_t M_t + \sum_{j=1}^{5} c_j R_{t-j}^2 + \sum_{k=1}^{5} c_k Overnight_{k,t} R_{t-k}^2$. In this model, M_t takes the value of the News Count factor. 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/Obsershelp (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.

	(1) Portugal	(2) Greece	(3) Italy	(4) Germany	(5) US	(6) Japan	(7) China
News Count	0.00000323***	0.00000345***	0.00000295***	0.00000164***	-5.72e-10	0.00000177***	1.22e-08
	(2.61)	(4.32)	(3.83)	(5.36)	(-0.37)	(7.09)	(0.50)
$Overnight \times$ News Count	0.00000153	0.00000383*	0.00000540*	0.00000244**	-0.00000117***	0.00000958	-0.000000473
	(1.28)	(1.73)	(1.70)	(2.00)	(-3.39)	(1.17)	(-0.04)
$D_{Similarity>50\%} \times$ News Count	-0.00000180	-5.75e-08	-0.00000177**	0.000000431	-4.18e-08	-0.000000436	0.000000892**
	(-1.61)	(-0.02)	(-2.22)	(0.35)	(-0.46)	(-1.11)	(2.01)
$D_{Similarity>60\%} \times$ News Count	-0.00000179	-2.08e-08	-0.00000179**	0.000000442	-4.22e-08	-0.000000427	0.000000871**
	(-1.60)	(-0.01)	(-2.24)	(0.36)	(-0.47)	(-1.08)	(1.97)
$D_{Similarity>70\%} \times$ News Count	-0.00000174	-4.36e-08	-0.00000180**	0.000000419	-3.87e-08	-0.000000427	0.000000845*
	(-1.58)	(-0.02)	(-2.25)	(0.33)	(-0.42)	(-1.08)	(1.91)
$D_{Similarity>80\%} \times$ News Count	-0.00000173	-3.28e-08	-0.00000180**	0.000000366	-4.38e-08	-0.000000417	0.000000850*
	(-1.58)	(-0.01)	(-2.27)	(0.29)	(-0.48)	(-1.05)	(1.89)
$D_{Similarity>90\%} \times$ News Count	-0.00000193*	7.15e-08	-0.00000196**	0.000000395	-4.76e-08	-0.000000392	0.000000871*
	(-1.75)	(0.03)	(-2.54)	(0.31)	(-0.51)	(-0.98)	(1.89)
$Overnight \times Return_{t-1}^2$	-0.214*	-0.0214	-0.0502	-0.0589	-0.320**	-3.039***	-0.372
	(-1.90)	(-0.42)	(-0.36)	(-1.58)	(-2.09)	(-4.51)	(-1.24)
$Overnight \times Return_{t-2}^2$	-0.327	-0.0207	-0.221*	-0.278**	0.284**	-0.207	-2.958*
	(-1.58)	(-0.82)	(-1.78)	(-2.47)	(1.97)	(-0.56)	(-1.83)
$Overnight \times Return_{t-3}^2$	-0.0809	-0.0613**	-0.226***	-0.121**	-0.0985	-1.258	-0.141
	(-0.80)	(-2.08)	(-2.94)	(-2.33)	(-1.40)	(-1.33)	(-0.64)
$Overnight \times Return_{t-4}^2$	-0.290	-0.0263**	0.000889	-0.208***	0.00353	0.165	0.137
	(-1.27)	(-1.97)	(0.02)	(-2.80)	(0.07)	(0.46)	(0.65)
$Overnight \times Return_{t-5}^2$	-0.0657	-0.0222	-0.101**	-0.0893	-0.0675	-0.552***	-1.276**
	(-0.84)	(-1.46)	(-2.27)	(-1.56)	(-1.24)	(-6.54)	(-2.45)
$Return_{t-1}^2$	0.251**	0.0589*	0.159*	0.0548	0.290*	3.049***	0.257
	(2.35)	(1.68)	(1.68)	(1.51)	(1.92)	(-4.51)	(0.81)
$Return_{t-2}^2$	0.325	0.0263	0.228*	0.298***	-0.272*	-0.207	2.922*
	(1.59)	(1.20)	(1.83)	(2.66)	(-1.87)	(-0.56)	(1.79)
$Return_{t-3}^2$	0.0698	0.0519*	0.214***	0.101*	0.0898	-1.258	0.0640
	(0.67)	(1.78)	(3.05)	(1.93)	(1.30)	(-1.33)	(0.28)
$Return_{t-4}^2$	0.311	0.0151	-0.00486	0.197***	-0.00393	0.165	-0.277
	(1.36)	(1.16)	(-0.09)	(2.63)	(-0.08)	(0.46)	(-1.23)
$Return_{t-5}^2$	0.0236	0.0147	0.0836**	0.0747	0.0410	-0.552***	1.222**
	(0.24)	(0.98)	(1.97)	(1.29)	(1.40)	(-6.54)	(2.33)

t statistics in parentheses
* p<.1, ** p<.05, *** p<.01

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

ONLINE APPENDIX. This appendix contains the full tables for the results of the paper, which are not included in the main body of the paper due to space constraints.

JEL classification: G01, G14, G15, D83.

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

NOTE : The following results concern the effect of ***media content on stock returns***, for a ***30-minute interval*** between stock prices.

Table 6.21: Portugal Stock Market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0532*	0.0523*	0.0527*	0.0526*	0.0523*
	(1.85)	(1.82)	(1.83)	(1.83)	(1.85)
$Return_{t-2}$	0.0148	0.0154	0.0146	0.0153	0.0160
	(0.56)	(0.59)	(0.56)	(0.58)	(0.61)
$Return_{t-3}$	0.0157	0.0152	0.0156	0.0153	0.0149
	(0.48)	(0.47)	(0.48)	(0.47)	(0.46)
$Return_{t-4}$	0.00767	0.00708	0.00660	0.00750	0.00854
	(0.26)	(0.24)	(0.22)	(0.25)	(0.29)
$Return_{t-5}$	0.0138	0.0140	0.0136	0.0136	0.0143
	(0.53)	(0.54)	(0.53)	(0.53)	(0.56)
Positive	0.0131*				
	(1.71)				
Negative		-0.00768***			
		(-2.61)			
Pessimism			-0.00812***		
			(-2.99)		
Average Pessimism				-0.00531**	
				(-2.09)	
News Count					-0.0000111
					(-0.17)
Constant	-0.0000719	0.000269***	0.000223***	0.000153**	0.0000500
	(-1.01)	(2.64)	(2.74)	(2.05)	(0.38)
Observations	5255	5255	5255	5255	5255
Adjusted R^2	0.003	0.004	0.004	0.003	0.003

t statistics in parentheses

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

Table 6.22: Ireland stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0580**	0.0580**	0.0575**	0.0576**	0.0585**
	(1.99)	(1.99)	(1.98)	(1.98)	(2.01)
$Return_{t-2}$	0.0348	0.0349	0.0346	0.0350	0.0351
	(1.18)	(1.19)	(1.18)	(1.19)	(1.20)
$Return_{t-3}$	-0.0420	-0.0415	-0.0419	-0.0417	-0.0415
	(-1.47)	(-1.46)	(-1.47)	(-1.46)	(-1.46)
$Return_{t-4}$	0.0179	0.0172	0.0176	0.0171	0.0175
	(0.75)	(0.72)	(0.74)	(0.72)	(0.73)
$Return_{t-5}$	0.0224	0.0212	0.0213	0.0217	0.0224
	(1.02)	(0.97)	(0.97)	(0.99)	(1.03)
Positive	0.00681				
	(0.99)				
Negative		-0.00664**			
		(-1.99)			
Pessimism			-0.00645**		
			(-2.18)		
Average Pessimism				-0.00443*	
				(-1.73)	
News Count					0.00000326
					(0.12)
Constant	0.0000293	0.000289***	0.000236***	0.000186***	0.0000716
	(0.42)	(2.61)	(2.89)	(2.59)	(1.10)
Observations	5080	5080	5080	5080	5080
Adjusted R^2	0.004	0.005	0.005	0.005	0.004

t statistics in parentheses

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

Table 6.23: Italy stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0304	0.0298	0.0298	0.0292	0.0302
	(0.92)	(0.90)	(0.90)	(0.89)	(0.92)
$Return_{t-2}$	0.0275	0.0287	0.0282	0.0286	0.0284
	(0.82)	(0.86)	(0.84)	(0.86)	(0.85)
$Return_{t-3}$	-0.00170	-0.00132	-0.00142	-0.00180	-0.00147
	(-0.05)	(-0.04)	(-0.04)	(-0.05)	(-0.04)
$Return_{t-4}$	-0.0410	-0.0416	-0.0420	-0.0421	-0.0406
	(-1.44)	(-1.46)	(-1.47)	(-1.48)	(-1.43)
$Return_{t-5}$	0.0131	0.0120	0.0117	0.0120	0.0132
	(0.49)	(0.45)	(0.44)	(0.45)	(0.50)
Positive	0.0169				
	(1.50)				
Negative		-0.00784*			
		(-1.72)			
Pessimism			-0.00886**		
			(-2.14)		
Average Pessimism				-0.0106**	
				(-2.49)	
News Count					-0.00000382
					(-0.10)
Constant	-0.0000551	0.000319**	0.000286**	0.000327***	0.0000801
	(-0.49)	(1.99)	(2.34)	(2.69)	(0.80)
Observations	4865	4865	4865	4865	4865
Adjusted R^2	0.002	0.003	0.003	0.003	0.002

t statistics in parentheses

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

Table 6.24: Greece stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0706**	0.0695**	0.0690**	0.0698**	0.0725**
	(2.24)	(2.22)	(2.20)	(2.22)	(2.30)
$Return_{t-2}$	-0.0231	-0.0225	-0.0223	-0.0228	-0.0253
	(-0.82)	(-0.80)	(-0.79)	(-0.81)	(-0.89)
$Return_{t-3}$	0.0294	0.0291	0.0290	0.0291	0.0297
	(1.15)	(1.14)	(1.13)	(1.13)	(1.15)
$Return_{t-4}$	-0.00105	-0.00229	-0.00270	-0.00156	0.00194
	(-0.04)	(-0.09)	(-0.10)	(-0.06)	(0.07)
$Return_{t-5}$	0.0193	0.0200	0.0200	0.0193	0.0184
	(1.01)	(1.04)	(1.04)	(1.01)	(0.96)
Positive	0.0296*				
	(1.74)				
Negative		-0.0193***			
		(-2.73)			
Pessimism			-0.0194***		
			(-2.99)		
Average Pessimism				-0.0139**	
				(-2.17)	
News Count					0.0000579
					(1.54)
Constant	-0.000160	0.000678***	0.000536***	0.000398**	-0.000128
	(-0.97)	(2.72)	(2.75)	(2.18)	(-0.98)
Observations	3653	3653	3653	3653	3653
Adjusted R^2	0.002	0.003	0.004	0.002	0.004

t statistics in parentheses

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

Table 6.25: Spain stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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) $Return_t$	(2) $Return_t$	(3) $Return_t$	(4) $Return_t$	(5) $Return_t$
$Return_{t-1}$	0.0218	0.0215	0.0216	0.0218	0.0226
	(0.68)	(0.67)	(0.67)	(0.68)	(0.70)
$Return_{t-2}$	-0.0219	-0.0214	-0.0213	-0.0219	-0.0216
	(-0.71)	(-0.70)	(-0.70)	(-0.71)	(-0.70)
$Return_{t-3}$	0.0171	0.0169	0.0165	0.0171	0.0176
	(0.58)	(0.57)	(0.56)	(0.58)	(0.59)
$Return_{t-4}$	0.0517*	0.0518*	0.0520*	0.0517*	0.0517*
	(1.86)	(1.86)	(1.87)	(1.85)	(1.85)
$Return_{t-5}$	-0.000743	-0.00125	-0.00129	-0.000965	-0.00155
	(-0.03)	(-0.05)	(-0.05)	(-0.04)	(-0.07)
Positive	0.0134				
	(1.15)				
Negative		-0.00759*			
		(-1.66)			
Pessimism			-0.00799*		
			(-1.84)		
Average Pessimism				-0.00359	
				(-0.77)	
News Count					-0.0000294
					(-0.91)
Constant	-0.0000930	0.000250	0.000203	0.0000947	0.0000851
	(-0.88)	(1.52)	(1.51)	(0.68)	(0.91)
Observations	4357	4357	4357	4357	4357
Adjusted R^2	0.002	0.002	0.003	0.002	0.003

t statistics in parentheses

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

Table 6.26: Germany stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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) $Return_t$	(2) $Return_t$	(3) $Return_t$	(4) $Return_t$	(5) $Return_t$
$Return_{t-1}$	0.0117	0.0108	0.0105	0.0106	0.0116
	(0.35)	(0.32)	(0.31)	(0.32)	(0.35)
$Return_{t-2}$	0.00141	0.00158	0.00169	0.00194	0.00136
	(0.04)	(0.05)	(0.05)	(0.06)	(0.04)
$Return_{t-3}$	-0.00622	-0.00591	-0.00606	-0.00618	-0.00558
	(-0.18)	(-0.17)	(-0.18)	(-0.18)	(-0.16)
$Return_{t-4}$	-0.0429	-0.0430	-0.0437	-0.0434	-0.0420
	(-1.51)	(-1.52)	(-1.55)	(-1.54)	(-1.48)
$Return_{t-5}$	-0.00817	-0.00739	-0.00741	-0.00740	-0.00847
	(-0.33)	(-0.30)	(-0.30)	(-0.30)	(-0.34)
Positive	0.0147*				
	(1.72)				
Negative		-0.00475			
		(-1.23)			
Pessimism			-0.00596*		
			(-1.70)		
Average Pessimism				-0.00609*	
				(-1.82)	
News Count					0.0000197
					(0.58)
Constant	0.0000151	0.000276**	0.000270***	0.000272***	0.0000745
	(0.17)	(2.17)	(2.84)	(2.99)	(0.92)
Observations	4864	4864	4864	4864	4864
Adjusted R^2	0.001	0.001	0.001	0.001	0.001

t statistics in parentheses

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

Table 6.27: France stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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) $Return_t$	(2) $Return_t$	(3) $Return_t$	(4) $Return_t$	(5) $Return_t$
$Return_{t-1}$	-0.00998	-0.0118	-0.0120	-0.0113	-0.0109
	(-0.47)	(-0.56)	(-0.57)	(-0.53)	(-0.52)
$Return_{t-2}$	-0.0196	-0.0195	-0.0199	-0.0196	-0.0187
	(-0.97)	(-0.97)	(-0.99)	(-0.97)	(-0.93)
$Return_{t-3}$	0.0299	0.0299	0.0295	0.0300	0.0303
	(1.11)	(1.11)	(1.09)	(1.11)	(1.12)
$Return_{t-4}$	0.00463	0.00561	0.00517	0.00536	0.00516
	(0.18)	(0.22)	(0.20)	(0.21)	(0.21)
$Return_{t-5}$	-0.0287	-0.0287	-0.0292	-0.0288	-0.0279
	(-1.38)	(-1.38)	(-1.40)	(-1.38)	(-1.34)
Positive	0.0218***				
	(2.65)				
Negative		-0.00555			
		(-1.56)			
Pessimism			-0.00757**		
			(-2.29)		
Average Pessimism				-0.00501	
				(-1.60)	
News Count					-0.0000228
					(-0.56)
Constant	-0.000125	0.000213*	0.000221**	0.000157*	0.0000901
	(-1.54)	(1.80)	(2.39)	(1.84)	(1.00)
Observations	5132	5132	5132	5132	5132
Adjusted R^2	0.002	0.001	0.002	0.001	0.001

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6.28: Austria stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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) $Return_t$	(2) $Return_t$	(3) $Return_t$	(4) $Return_t$	(5) $Return_t$
$Return_{t-1}$	0.103***	0.102***	0.102***	0.102***	0.103***
	(2.83)	(2.78)	(2.79)	(2.80)	(2.82)
$Return_{t-2}$	0.0229	0.0237	0.0237	0.0230	0.0238
	(0.61)	(0.63)	(0.63)	(0.61)	(0.64)
$Return_{t-3}$	0.0282	0.0268	0.0266	0.0279	0.0279
	(0.74)	(0.71)	(0.70)	(0.74)	(0.74)
$Return_{t-4}$	-0.0104	-0.00969	-0.0106	-0.0103	-0.00899
	(-0.28)	(-0.26)	(-0.29)	(-0.28)	(-0.24)
$Return_{t-5}$	0.0571**	0.0575**	0.0571**	0.0578**	0.0577**
	(2.54)	(2.55)	(2.53)	(2.57)	(2.57)
Positive	0.0190**				
	(2.06)				
Negative		-0.00984***			
		(-2.82)			
Pessimism			-0.0106***		
			(-3.21)		
Average Pessimism				-0.00593*	
				(-1.92)	
News Count					-0.0000252
					(-0.63)
Constant	-0.0000696	0.000385***	0.000330***	0.000216**	0.000139
	(-0.80)	(3.25)	(3.45)	(2.42)	(1.43)
Observations	4654	4654	4654	4654	4654
Adjusted R^2	0.012	0.012	0.013	0.011	0.012

t statistics in parentheses

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

Table 6.29: Belgium stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.00581	0.00402	0.00461	0.00462	0.00468
	(0.19)	(0.13)	(0.15)	(0.15)	(0.15)
$Return_{t-2}$	-0.0349	-0.0348	-0.0351	-0.0347	-0.0345
	(-1.22)	(-1.22)	(-1.23)	(-1.22)	(-1.22)
$Return_{t-3}$	0.0469*	0.0465*	0.0465*	0.0467*	0.0468*
	(1.75)	(1.73)	(1.73)	(1.74)	(1.75)
$Return_{t-4}$	0.0479	0.0488	0.0483	0.0485	0.0489*
	(1.59)	(1.61)	(1.60)	(1.60)	(1.65)
$Return_{t-5}$	-0.0132	-0.0129	-0.0135	-0.0129	-0.0123
	(-0.42)	(-0.41)	(-0.43)	(-0.41)	(-0.39)
Positive	0.0124*				
	(1.72)				
Negative		-0.00664**			
		(-2.24)			
Pessimism			-0.00714**		
			(-2.56)		
Average Pessimism				-0.00397	
				(-1.57)	
News Count					-0.00000134
					(-0.03)
Constant	-0.0000693	0.000233**	0.000196**	0.000119	0.0000259
	(-1.00)	(2.29)	(2.43)	(1.63)	(0.31)
Observations	5131	5131	5131	5131	5131
Adjusted R^2	0.004	0.005	0.005	0.004	0.004

t statistics in parentheses

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

Table 6.30: Finland stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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) $Return_t$	(2) $Return_t$	(3) $Return_t$	(4) $Return_t$	(5) $Return_t$
$Return_{t-1}$	0.0926**	0.0914**	0.0915**	0.0921**	0.0930**
	(2.23)	(2.20)	(2.21)	(2.21)	(2.26)
$Return_{t-2}$	0.0807*	0.0813*	0.0807*	0.0807*	0.0808*
	(1.89)	(1.91)	(1.89)	(1.89)	(1.92)
$Return_{t-3}$	0.0330	0.0325	0.0322	0.0328	0.0333
	(1.11)	(1.09)	(1.08)	(1.10)	(1.12)
$Return_{t-4}$	-0.0363	-0.0360	-0.0361	-0.0359	-0.0361
	(-1.24)	(-1.23)	(-1.23)	(-1.23)	(-1.23)
$Return_{t-5}$	0.00645	0.00627	0.00608	0.00617	0.00640
	(0.23)	(0.22)	(0.21)	(0.22)	(0.23)
Positive	0.00860				
	(0.96)				
Negative		-0.00780**			
		(-2.09)			
Pessimism			-0.00755**		
			(-2.24)		
Average Pessimism				-0.00527	
				(-1.61)	
News Count					-0.00000673
					(-0.22)
Constant	0.0000331	0.000346***	0.000281***	0.000225**	0.000117
	(0.35)	(2.65)	(2.87)	(2.38)	(1.47)
Observations	4387	4387	4387	4387	4387
Adjusted R^2	0.013	0.014	0.014	0.013	0.013

t statistics in parentheses

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

Table 6.31: UK stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	-0.0313	-0.0317	-0.0318	-0.0322	-0.0309
	(-1.02)	(-1.04)	(-1.04)	(-1.05)	(-1.00)
$Return_{t-2}$	0.0190	0.0191	0.0186	0.0195	0.0196
	(0.62)	(0.62)	(0.61)	(0.63)	(0.64)
$Return_{t-3}$	0.00116	0.000376	0.000434	0.000782	0.00134
	(0.04)	(0.01)	(0.01)	(0.02)	(0.04)
$Return_{t-4}$	0.0225	0.0230	0.0226	0.0223	0.0228
	(0.73)	(0.74)	(0.73)	(0.72)	(0.74)
$Return_{t-5}$	0.0166	0.0162	0.0162	0.0160	0.0169
	(0.59)	(0.57)	(0.57)	(0.57)	(0.60)
Positive	0.00951				
	(1.53)				
Negative		-0.00593**			
		(-2.17)			
Pessimism			-0.00618**		
			(-2.51)		
Average Pessimism				-0.00455*	
				(-1.81)	
News Count					0.0000130
					(0.53)
Constant	-0.0000423	0.000216**	0.000179***	0.000138**	-0.00000365
	(-0.63)	(2.33)	(2.58)	(2.00)	(-0.06)
Observations	5119	5119	5119	5119	5119
Adjusted R^2	0.001	0.001	0.001	0.001	0.000

t statistics in parentheses

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

Table 6.32: Switzerland stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.00918	0.00876	0.00906	0.00862	0.00851
	(0.22)	(0.21)	(0.22)	(0.21)	(0.20)
$Return_{t-2}$	-0.0277	-0.0275	-0.0280	-0.0273	-0.0270
	(-0.74)	(-0.74)	(-0.75)	(-0.73)	(-0.72)
$Return_{t-3}$	0.0533	0.0530	0.0528	0.0530	0.0536
	(1.48)	(1.47)	(1.46)	(1.47)	(1.48)
$Return_{t-4}$	-0.0783**	-0.0789**	-0.0790**	-0.0789**	-0.0784**
	(-2.02)	(-2.04)	(-2.04)	(-2.04)	(-2.03)
$Return_{t-5}$	0.00853	0.00812	0.00771	0.00872	0.00912
	(0.33)	(0.32)	(0.30)	(0.34)	(0.35)
Positive	0.00927				
	(1.32)				
Negative		-0.00499*			
		(-1.86)			
Pessimism			-0.00536**		
			(-2.15)		
Average Pessimism				-0.00332	
				(-1.41)	
News Count					-0.00000604
					(-0.23)
Constant	-0.0000396	0.000190**	0.000161**	0.000110*	0.0000462
	(-0.58)	(2.12)	(2.31)	(1.70)	(0.74)
Observations	4706	4706	4706	4706	4706
Adjusted R^2	0.003	0.003	0.003	0.003	0.002

t statistics in parentheses

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

Table 6.33: Norway stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0292	0.0280	0.0278	0.0279	0.0291
	(0.77)	(0.74)	(0.73)	(0.74)	(0.76)
$Return_{t-2}$	0.0396	0.0415	0.0415	0.0412	0.0402
	(1.01)	(1.05)	(1.05)	(1.05)	(1.02)
$Return_{t-3}$	0.000992	0.000776	0.000404	0.000759	0.00152
	(0.03)	(0.02)	(0.01)	(0.02)	(0.04)
$Return_{t-4}$	-0.0746**	-0.0745**	-0.0749**	-0.0748**	-0.0739**
	(-2.15)	(-2.15)	(-2.16)	(-2.16)	(-2.14)
$Return_{t-5}$	0.0182	0.0189	0.0182	0.0191	0.0196
	(0.64)	(0.67)	(0.64)	(0.67)	(0.69)
Positive	0.0202**				
	(2.03)				
Negative		-0.00669			
		(-1.64)			
Pessimism			-0.00827**		
			(-2.16)		
Average Pessimism				-0.00602*	
				(-1.69)	
News Count					0.00000664
					(0.22)
Constant	-0.0000528	0.000311**	0.000299***	0.000243**	0.0000795
	(-0.54)	(2.27)	(2.82)	(2.45)	(0.97)
Observations	4412	4412	4412	4412	4412
Adjusted R^2	0.006	0.005	0.006	0.005	0.005

t statistics in parentheses

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

Table 6.34: Brazil stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	-0.0182	-0.0185	-0.0193	-0.0187	-0.0176
	(-0.53)	(-0.53)	(-0.56)	(-0.54)	(-0.51)
$Return_{t-2}$	-0.00979	-0.00873	-0.00866	-0.00921	-0.00986
	(-0.26)	(-0.23)	(-0.23)	(-0.25)	(-0.26)
$Return_{t-3}$	0.0239	0.0234	0.0228	0.0233	0.0245
	(0.80)	(0.78)	(0.76)	(0.78)	(0.82)
$Return_{t-4}$	-0.0544*	-0.0527*	-0.0527*	-0.0535*	-0.0529*
	(-1.91)	(-1.84)	(-1.84)	(-1.87)	(-1.86)
$Return_{t-5}$	-0.00340	-0.00428	-0.00459	-0.00437	-0.00383
	(-0.16)	(-0.20)	(-0.22)	(-0.21)	(-0.18)
Positive	0.0193				
	(1.47)				
Negative		-0.00805			
		(-1.42)			
Pessimism			-0.00934*		
			(-1.84)		
Average Pessimism				-0.00901*	
				(-1.88)	
News Count					0.0000291
					(1.54)
Constant	-0.0000280	0.000369*	0.000341**	0.000337**	-0.00000343
	(-0.21)	(1.84)	(2.29)	(2.44)	(-0.04)
Observations	3031	3031	3031	3031	3031
Adjusted R^2	0.001	0.001	0.001	0.001	0.002

t statistics in parentheses

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

Table 6.35: Canada stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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) $Return_t$	(2) $Return_t$	(3) $Return_t$	(4) $Return_t$	(5) $Return_t$
$Return_{t-1}$	-0.0180	-0.0162	-0.0177	-0.0188	-0.0152
	(-0.47)	(-0.42)	(-0.46)	(-0.49)	(-0.40)
$Return_{t-2}$	-0.00375	-0.00453	-0.00391	-0.00325	-0.00518
	(-0.12)	(-0.14)	(-0.12)	(-0.10)	(-0.16)
$Return_{t-3}$	0.0124	0.0116	0.0112	0.0111	0.0131
	(0.75)	(0.71)	(0.68)	(0.67)	(0.80)
$Return_{t-4}$	-0.00568	-0.00473	-0.00488	-0.00621	-0.00536
	(-0.31)	(-0.26)	(-0.27)	(-0.34)	(-0.29)
$Return_{t-5}$	0.0151	0.0146	0.0144	0.0153	0.0153
	(0.60)	(0.59)	(0.58)	(0.62)	(0.62)
Positive	0.0222**				
	(2.52)				
Negative		-0.00591			
		(-1.64)			
Pessimism			-0.00785**		
			(-2.21)		
Average Pessimism				-0.0137***	
				(-2.66)	
News Count					0.00000249
					(0.97)
Constant	-0.000207**	0.000148	0.000152	0.000300**	-0.0000588
	(-2.29)	(1.22)	(1.54)	(2.25)	(-0.77)
Observations	1867	1867	1867	1867	1867
Adjusted R^2	0.001	-0.001	0.000	0.003	-0.002

t statistics in parentheses

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

Table 6.36: US Dow Jones stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	-0.161*	-0.163*	-0.163*	-0.162*	-0.163*
	(-1.68)	(-1.70)	(-1.71)	(-1.68)	(-1.70)
$Return_{t-2}$	0.0391	0.0432	0.0420	0.0394	0.0410
	(0.50)	(0.55)	(0.54)	(0.50)	(0.52)
$Return_{t-3}$	0.0132	0.00836	0.00978	0.0127	0.0127
	(0.22)	(0.14)	(0.16)	(0.21)	(0.21)
$Return_{t-4}$	-0.0254	-0.0236	-0.0240	-0.0253	-0.0271
	(-0.46)	(-0.43)	(-0.44)	(-0.46)	(-0.49)
$Return_{t-5}$	0.0548	0.0536	0.0542	0.0546	0.0567
	(1.40)	(1.37)	(1.39)	(1.40)	(1.46)
Positive	0.00853				
	(0.41)				
Negative		-0.0224***			
		(-2.60)			
Pessimism			-0.0190**		
			(-2.40)		
Average Pessimism				-0.00587	
				(-0.82)	
News Count					-0.0000119
					(-0.87)
Constant	-0.000101	0.000691**	0.000438*	0.000111	0.0000594
	(-0.49)	(2.23)	(1.83)	(0.51)	(0.46)
Observations	1529	1529	1529	1529	1529
Adjusted R^2	0.007	0.011	0.011	0.007	0.007

t statistics in parentheses

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

Table 6.37: US S&P 500 stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	-0.0605**	-0.0590**	-0.0596**	-0.0594**	-0.0600**
	(-2.03)	(-2.02)	(-2.03)	(-2.02)	(-2.01)
$Return_{t-2}$	-0.0142	-0.0132	-0.0127	-0.0151	-0.0151
	(-0.51)	(-0.47)	(-0.45)	(-0.54)	(-0.54)
$Return_{t-3}$	-0.0332	-0.0357	-0.0343	-0.0346	-0.0340
	(-0.84)	(-0.91)	(-0.87)	(-0.88)	(-0.87)
$Return_{t-4}$	-0.0224	-0.0228	-0.0219	-0.0220	-0.0232
	(-0.48)	(-0.49)	(-0.47)	(-0.47)	(-0.49)
$Return_{t-5}$	-0.0353	-0.0353	-0.0371	-0.0355	-0.0341
	(-0.79)	(-0.80)	(-0.84)	(-0.80)	(-0.76)
Positive	0.0189				
	(1.03)				
Negative		-0.0194***			
		(-2.82)			
Pessimism			-0.0181***		
			(-2.75)		
Average Pessimism				-0.00978	
				(-1.33)	
News Count					-0.00000460
					(-0.49)
Constant	-0.0000793	0.000695***	0.000517***	0.000308	0.000105
	(-0.47)	(2.87)	(2.65)	(1.46)	(0.82)
Observations	1418	1418	1418	1418	1418
Adjusted R^2	0.001	0.006	0.007	0.002	0.001

t statistics in parentheses

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

Table 6.38: Japan stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	-0.340	-0.340	-0.344	-0.337	-0.332
	(-1.08)	(-1.10)	(-1.13)	(-1.08)	(-1.04)
$Return_{t-2}$	0.179	0.205	0.204	0.183	0.185
	(0.69)	(0.79)	(0.79)	(0.71)	(0.71)
$Return_{t-3}$	0.323	0.313	0.315	0.335	0.318
	(1.28)	(1.25)	(1.26)	(1.34)	(1.27)
$Return_{t-4}$	0.0701	0.0772	0.0717	0.0722	0.0807
	(0.34)	(0.37)	(0.35)	(0.35)	(0.39)
$Return_{t-5}$	-0.0982	-0.102	-0.107	-0.101	-0.0932
	(-0.49)	(-0.50)	(-0.52)	(-0.50)	(-0.47)
Positive	0.132**				
	(2.16)				
Negative		-0.0736***			
		(-2.71)			
Pessimism			-0.0767***		
			(-3.17)		
Average Pessimism				-0.129***	
				(-2.72)	
News Count					0.00000837
					(0.31)
Constant	-0.000955	0.00245***	0.00194***	0.00317***	-0.0000432
	(-1.43)	(2.80)	(3.16)	(2.85)	(-0.07)
Observations	718	718	718	718	718
Adjusted R^2	0.013	0.017	0.022	0.017	0.006

t statistics in parentheses

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

Table 6.39: China stock market - 30 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 30-minute interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	-0.0618	-0.0855	-0.0795	-0.0747	-0.0740
	(-0.35)	(-0.49)	(-0.46)	(-0.43)	(-0.43)
$Return_{t-2}$	-0.204	-0.221	-0.220	-0.214	-0.211
	(-0.87)	(-0.94)	(-0.94)	(-0.91)	(-0.90)
$Return_{t-3}$	-0.126	-0.122	-0.128	-0.115	-0.110
	(-0.91)	(-0.89)	(-0.93)	(-0.84)	(-0.80)
$Return_{t-4}$	0.206	0.182	0.188	0.188	0.193
	(0.83)	(0.73)	(0.76)	(0.76)	(0.77)
$Return_{t-5}$	-0.0670	-0.0696	-0.0725	-0.0743	-0.0622
	(-0.32)	(-0.33)	(-0.34)	(-0.35)	(-0.30)
Positive	0.0937				
	(1.54)				
Negative		-0.0438*			
		(-1.93)			
Pessimism			-0.0471**		
			(-2.31)		
Average Pessimism				-0.0537	
				(-1.42)	
News Count					0.0000166
					(0.78)
Constant	-0.000329	0.00181**	0.00155***	0.00168*	0.000125
	(-0.49)	(2.35)	(2.74)	(1.92)	(0.26)
Observations	730	730	730	730	730
Adjusted R^2	0.001	0.001	0.003	-0.001	-0.002

t statistics in parentheses

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

NOTE : The following results concern the effect of

media content on stock returns,

for a ***5-minute interval*** between stock prices.

Table 6.40: Portugal Stock Market - 5 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 5-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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.106***	0.105***	0.104***	0.105***	0.107***
	(4.66)	(4.62)	(4.59)	(4.63)	(4.70)
$Return_{t-2}$	0.0287	0.0280	0.0282	0.0282	0.0282
	(1.23)	(1.20)	(1.21)	(1.21)	(1.20)
$Return_{t-3}$	0.0244	0.0230	0.0230	0.0235	0.0244
	(1.22)	(1.16)	(1.16)	(1.18)	(1.22)
$Return_{t-4}$	0.0453**	0.0456**	0.0457**	0.0456**	0.0454**
	(2.29)	(2.30)	(2.31)	(2.30)	(2.29)
$Return_{t-5}$	0.0122	0.0120	0.0117	0.0122	0.0128
	(0.60)	(0.59)	(0.57)	(0.59)	(0.63)
Positive	0.00867**				
	(2.37)				
Negative		-0.00311**			
		(-2.31)			
Pessimism			-0.00374***		
			(-2.84)		
Average Pessimism				-0.00200**	
				(-2.22)	
News Count					-0.0000473
					(-0.36)
Constant	-0.0000655**	0.0000961**	0.0000887**	0.0000461	0.0000654
	(-2.12)	(2.07)	(2.19)	(1.63)	(0.37)
Observations	6955	6955	6955	6955	6955
Adjusted R^2	0.004	0.003	0.004	0.003	0.005

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6.41: Greece Stock Market - 5 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 5-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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0108	0.0106	0.0103	0.0105	0.0109	0.0108
	(0.31)	(0.31)	(0.30)	(0.30)	(0.31)	(0.31)
$Return_{t-2}$	-0.0994***	-0.0999***	-0.0999***	-0.0997***	-0.0997***	-0.0999***
	(-3.99)	(-4.01)	(-4.01)	(-4.00)	(-4.00)	(-4.01)
$Return_{t-3}$	-0.0481**	-0.0475**	-0.0476**	-0.0477**	-0.0479**	-0.0484**
	(-2.02)	(-2.00)	(-2.00)	(-2.00)	(-2.01)	(-2.03)
$Return_{t-4}$	0.00109	-0.000274	-0.000585	-0.0000377	0.000402	-0.000596
	(0.05)	(-0.01)	(-0.02)	(-0.00)	(0.02)	(-0.02)
$Return_{t-5}$	0.0431**	0.0428**	0.0427**	0.0428**	0.0422**	0.0419**
	(2.31)	(2.30)	(2.29)	(2.30)	(2.28)	(2.26)
Positive	0.0138**					
	(2.25)					
Negative		-0.00634**				
		(-2.55)				
Pessimism			-0.00689***			
			(-3.11)			
Average Pessimism				-0.00522***		
				(-2.58)		
News Count					0.0000274	
					(1.36)	
Constant	-0.0000497	0.000253***	0.000220***	0.000179***	-0.00000910	0.000000499
	(-0.84)	(2.88)	(3.26)	(2.80)	(-0.19)	(0.01)
Observations	4791	4791	4791	4791	4791	4791
Adjusted R^2	0.009	0.009	0.010	0.009	0.010	0.010

t statistics in parentheses

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 6.42: Italy Stock Market - 5 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 5-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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0293	0.0296	0.0294	0.0295	0.0300	0.0305
	(1.23)	(1.25)	(1.24)	(1.24)	(1.28)	(1.29)
$Return_{t-2}$	0.00817	0.00845	0.00818	0.00832	0.00955	0.00929
	(0.24)	(0.25)	(0.24)	(0.24)	(0.28)	(0.27)
$Return_{t-3}$	0.0388	0.0386	0.0385	0.0384	0.0389	0.0390
	(1.56)	(1.56)	(1.55)	(1.55)	(1.57)	(1.57)
$Return_{t-4}$	-0.0152	-0.0153	-0.0152	-0.0153	-0.0118	-0.0135
	(-0.58)	(-0.59)	(-0.59)	(-0.59)	(-0.47)	(-0.52)
$Return_{t-5}$	0.0369	0.0364	0.0360	0.0364	0.0352	0.0368
	(1.22)	(1.21)	(1.19)	(1.20)	(1.18)	(1.23)
Positive	0.00638					
	(1.30)					
Negative		-0.00322				
		(-1.58)				
Pessimism			-0.00353*			
			(-1.93)			
Average Pessimism				-0.00263*		
				(-1.75)		
News Count					-0.0000549	
					(-1.54)	
Constant	-0.0000566	0.0000916	0.0000756	0.0000536	0.0000897	
	(-1.11)	(1.30)	(1.40)	(1.16)	(1.63)	
Observations	6899	6899	6899	6899	6899	
Adjusted R^2	0.001	0.001	0.001	0.001	0.005	

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6.43: Germany Stock Market - 5 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 5-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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	-0.0314	-0.0312	-0.0318	-0.0313	-0.0305	-0.0305
	(-1.60)	(-1.58)	(-1.62)	(-1.59)	(-1.54)	(-1.54)
$Return_{t-2}$	-0.00270	-0.00369	-0.00364	-0.00383	-0.00267	-0.00358
	(-0.08)	(-0.12)	(-0.11)	(-0.12)	(-0.08)	(-0.11)
$Return_{t-3}$	-0.0157	-0.0159	-0.0159	-0.0157	-0.0157	-0.0161
	(-0.64)	(-0.65)	(-0.65)	(-0.64)	(-0.64)	(-0.66)
$Return_{t-4}$	-0.0394	-0.0389	-0.0389	-0.0389	-0.0386	-0.0403
	(-1.06)	(-1.05)	(-1.05)	(-1.05)	(-1.05)	(-1.09)
$Return_{t-5}$	0.0385	0.0367	0.0368	0.0371	0.0386	0.0382
	(1.02)	(0.97)	(0.98)	(0.98)	(1.03)	(1.02)
Positive	0.0122***					
	(2.66)					
Negative		-0.00322				
		(-1.62)				
Pessimism			-0.00432**			
			(-2.34)			
Average Pessimism				-0.00272**		
				(-2.00)		
News Count					0.0000293	
					(0.75)	
Constant	-0.0000436	0.000148**	0.000151***	0.000112***	-0.00000417	-0.000209
	(-0.93)	(2.34)	(3.09)	(2.69)	(-0.07)	(-1.07)
Observations	6479	6479	6479	6479	6479	6479
Adjusted R^2	0.001	0.001	0.001	0.001	0.002	0.003

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6.44: France Stock Market - 5 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 5-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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0418	0.0415	0.0414	0.0415	0.0418
	(1.01)	(1.00)	(1.00)	(1.00)	(1.01)
$Return_{t-2}$	-0.0125	-0.0130	-0.0131	-0.0128	-0.0124
	(-0.53)	(-0.55)	(-0.56)	(-0.54)	(-0.53)
$Return_{t-3}$	0.0142	0.0144	0.0143	0.0144	0.0143
	(0.73)	(0.74)	(0.74)	(0.74)	(0.73)
$Return_{t-4}$	-0.0101	-0.0105	-0.0103	-0.0105	-0.0106
	(-0.45)	(-0.47)	(-0.46)	(-0.47)	(-0.47)
$Return_{t-5}$	0.0196	0.0197	0.0196	0.0198	0.0197
	(0.85)	(0.86)	(0.86)	(0.86)	(0.85)
Positive	0.00658				
	(1.62)				
Negative		-0.00293			
		(-1.64)			
Pessimism			-0.00330*		
			(-1.95)		
Average Pessimism				-0.00197	
				(-1.57)	
News Count					-0.0000223
					(-0.37)
Constant	-0.0000740*	0.0000663	0.0000540	0.0000215	0.00000658
	(-1.80)	(1.13)	(1.18)	(0.58)	(0.09)
Observations	6599	6599	6599	6599	6599
Adjusted R^2	0.000	0.000	0.001	0.000	0.000

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6.45: UK Stock Market - 5 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 5-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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0368	0.0370	0.0369	0.0369	0.0368
	(1.13)	(1.14)	(1.14)	(1.14)	(1.14)
$Return_{t-2}$	0.0119	0.0120	0.0117	0.0121	0.0121
	(0.56)	(0.56)	(0.55)	(0.56)	(0.56)
$Return_{t-3}$	-0.00225	-0.00227	-0.00239	-0.00217	-0.00225
	(-0.12)	(-0.12)	(-0.13)	(-0.11)	(-0.12)
$Return_{t-4}$	0.0216	0.0212	0.0213	0.0213	0.0211
	(1.03)	(1.01)	(1.02)	(1.02)	(1.01)
$Return_{t-5}$	-0.0295	-0.0298	-0.0300	-0.0296	-0.0297
	(-1.52)	(-1.53)	(-1.54)	(-1.52)	(-1.53)
Positive	0.00258				
	(0.91)				
Negative		-0.000705			
		(-0.52)			
Pessimism			-0.000935		
			(-0.79)		
Average Pessimism				-0.000293	
				(-0.32)	
News Count					-0.0000207
					(-1.01)
Constant	-0.0000142	0.0000270	0.0000274	0.0000119	0.0000400
	(-0.45)	(0.59)	(0.84)	(0.44)	(1.35)
Observations	7385	7385	7385	7385	7385
Adjusted R^2	0.000	0.000	0.000	0.000	0.002

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6.46: Switzerland Stock Market - 5 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 5-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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0687*	0.0679*	0.0676*	0.0685*	0.0623
	(1.76)	(1.74)	(1.74)	(1.75)	(1.61)
$Return_{t-2}$	-0.0477	-0.0494	-0.0494	-0.0485	-0.0472
	(-1.17)	(-1.21)	(-1.21)	(-1.19)	(-1.12)
$Return_{t-3}$	-0.00578	-0.00452	-0.00461	-0.00474	-0.00212
	(-0.16)	(-0.12)	(-0.12)	(-0.13)	(-0.06)
$Return_{t-4}$	0.0143	0.0149	0.0147	0.0145	0.0163
	(0.37)	(0.38)	(0.38)	(0.37)	(0.42)
$Return_{t-5}$	0.0639	0.0628	0.0628	0.0636	0.0607
	(1.60)	(1.58)	(1.58)	(1.59)	(1.54)
Positive	0.0102**				
	(2.58)				
Negative		-0.00416***			
		(-2.83)			
Pessimism			-0.00478***		
			(-3.45)		
Average Pessimism				-0.00290***	
				(-2.83)	
News Count					-0.0000552**
					(-2.29)
Constant	-0.0000771**	0.000131***	0.000116***	0.0000689**	0.0000951***
	(-2.04)	(2.80)	(3.16)	(2.26)	(2.69)
Observations	6468	6468	6468	6468	6468
Adjusted R^2	0.003	0.003	0.004	0.002	0.011

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6.47: US S&P 500 Stock Market - 5 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 5-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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.0555	0.0552	0.0549	0.0556	0.0558
	(1.41)	(1.40)	(1.40)	(1.41)	(1.42)
$Return_{t-2}$	0.00556	0.00129	0.00175	0.00464	0.00509
	(0.13)	(0.03)	(0.04)	(0.11)	(0.12)
$Return_{t-3}$	-0.0416	-0.0394	-0.0380	-0.0425	-0.0434
	(-0.86)	(-0.81)	(-0.78)	(-0.87)	(-0.90)
$Return_{t-4}$	-0.0875	-0.0866	-0.0877	-0.0863	-0.0863
	(-1.57)	(-1.54)	(-1.57)	(-1.54)	(-1.54)
$Return_{t-5}$	0.00335	0.00223	0.00338	0.00210	0.00196
	(0.10)	(0.07)	(0.10)	(0.06)	(0.06)
Positive	0.00868				
	(0.78)				
Negative		-0.00757*			
		(-1.74)			
Pessimism			-0.00728*		
			(-1.74)		
Average Pessimism				-0.00104	
				(-0.26)	
News Count					-0.000000619
					(-0.09)
Constant	0.0000273	0.000339**	0.000276**	0.000118	0.0000956
	(0.28)	(2.26)	(2.25)	(0.94)	(1.41)
Observations	1463	1463	1463	1463	1463
Adjusted R^2	-0.002	0.000	0.001	-0.002	-0.002

t statistics in parentheses

* $p<0.1$, ** $p<0.05$, *** $p<0.01$

Table 6.48: Japan Stock Market - 5 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 5-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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
$Return_{t-1}$	0.778*	0.755*	0.783*	0.706	0.734*
	(1.81)	(1.75)	(1.82)	(1.64)	(1.70)
$Return_{t-2}$	0.398	0.334	0.340	0.401	0.382
	(1.01)	(0.85)	(0.86)	(1.00)	(0.97)
$Return_{t-3}$	-0.567	-0.542	-0.562	-0.552	-0.534
	(-1.40)	(-1.34)	(-1.39)	(-1.35)	(-1.31)
$Return_{t-4}$	0.437	0.431	0.420	0.439	0.457
	(1.01)	(1.01)	(0.98)	(1.03)	(1.07)
$Return_{t-5}$	0.210	0.135	0.157	0.137	0.166
	(0.50)	(0.32)	(0.38)	(0.32)	(0.39)
Positive	0.0960*				
	(1.95)				
Negative		-0.0567***			
		(-2.70)			
Pessimism			-0.0580***		
			(-3.18)		
Average Pessimism				-0.0683**	
				(-2.48)	
News Count					0.00000551
					(0.23)
Constant	-0.000674	0.00192***	0.00149***	0.00174**	0.00000394
	(-1.26)	(2.64)	(2.86)	(2.41)	(0.01)
Observations	721	721	721	721	721
Adjusted R^2	0.007	0.012	0.015	0.009	0.002

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Table 6.49: China Stock Market - 5 minutes. The model I employ to study the effect of the content of news in high frequency stock returns is the following: $R_t = a_1 + b_1 M_t + \sum_{i=1}^{5} c_i R_{t-i}$ 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 interval, as defined in Section 3.5 of the paper. I control for five lags of returns (i.e. five 5-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)
	$Return_t$	$Return_t$	$Return_t$	$Return_t$	$Return_t$
Positive	0.0756				
	(1.30)				
$Return_{t-1}$	-0.376	-0.326	-0.327	-0.372	-0.368
	(-0.77)	(-0.67)	(-0.67)	(-0.77)	(-0.76)
$Return_{t-2}$	0.471	0.443	0.452	0.479	0.461
	(0.98)	(0.91)	(0.93)	(1.00)	(0.96)
$Return_{t-3}$	-0.662	-0.694	-0.673	-0.708	-0.681
	(-1.22)	(-1.27)	(-1.23)	(-1.30)	(-1.26)
$Return_{t-4}$	0.737*	0.748*	0.756*	0.727*	0.710
	(1.66)	(1.70)	(1.71)	(1.65)	(1.60)
$Return_{t-5}$	0.269	0.188	0.199	0.243	0.253
	(0.51)	(0.36)	(0.38)	(0.47)	(0.48)
Negative		-0.0511**			
		(-2.24)			
Pessimism			-0.0501**		
			(-2.52)		
Average Pessimism				-0.0505*	
				(-1.83)	
News Count					0.0000147
					(0.66)
ln_News Count					
Constant	-0.000292	0.00194**	0.00151***	0.00151**	0.0000601
	(-0.45)	(2.45)	(2.68)	(2.24)	(0.13)
Observations	759	759	759	759	759
Adjusted R^2	0.009	0.013	0.014	0.010	0.008

t statistics in parentheses

* p<0.1, ** p<0.05, *** p<0.01

Bibliography

Ahern, K. R. and Sosyura, D. (2014). Who writes the news? corporate press releases during merger negotiations. *The Journal of Finance*, 69(1):241–291.

Allen, F. and Gale, D. (2000). Financial contagion. *Journal of Political Economy*, 108(1):1–33.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Vega, C. (2003). Micro effects of macro announcements: Real-time price discovery in foreign exchange. *American Economic Review*, 93(1):38–62.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of Finance*, 59(3):1259–1294.

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

Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. J. (2010). Modeling financial contagion using mutually exciting jump processes. Working Paper 15850, National Bureau of Economic Research.

Bae, K.-H., Karolyi, G. A., and Stulz, R. M. (2003). A new approach to measuring financial contagion. *Review of Financial studies*, 16(3):717–763.

Beetsma, R., Giuliodori, M., de Jong, F., and 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.

Bekaert, G., Harvey, C. R., and Ng, A. (2005). Market integration and contagion. *The Journal of Business*, 78(1):39–69.

Bhanot, K., Burns, N., Hunter, D., and Williams, M. (2012). Was there contagion in eurozone sovereign bond markets during the greek debt crisis? Working Paper WP 006FIN-73-2012, The University of Texas at San Antonio, College of Business.

Bodnaruk, A., Loughran, T., and McDonald, B. (2015). Using 10-k text to gauge financial constraints. *Journal of Financial and Quantitative Analysis*, 50:623–646.

Boudoukh, J., Feldman, R., Kogan, S., and Richardson, M. (2013). Which news moves stock prices? a textual analysis. Technical report, National Bureau of Economic Research.

Boyson, N. M., Stahel, C. W., and Stulz, R. M. (2010). Hedge fund contagion and liquidity shocks. *The Journal of Finance*, 65(5):1789–1816.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *Review of Financial Studies*, 27(8):2267–2306.

Campbell, J. Y., Grossman, S. J., and Wang, J. (1992). Trading volume and serial correlation in stock returns. Technical report, National Bureau of Economic Research.

Caporin, M., Pelizzon, L., Ravazzolo, F., and Rigobon, R. (2013). Measuring sovereign contagion in europe. Technical report, National Bureau of Economic Research.

Chen, H., De, P., Hu, J., and 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. and Grammatikos, T. (2013). News flow, web attention and extreme returns in the european financial crisis.

Chouliaras, A. and Grammatikos, T. (2015). News flow, web attention and extreme returns in the european financial crisis.

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. and Ranaldo, A. (2009). Extreme coexceedances in new eu member states' stock markets. *Journal of banking & finance*, 33(6):1048–1057.

Corsetti, G., Pericoli, M., and Sbracia, M. (2005). 'some contagion, some interdependence': More pitfalls in tests of financial contagion. *Journal of International Money and Finance*, 24(8):1177 – 1199.

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

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

Daniel, K., Hirshleifer, D., and Subrahmanyam, A. (1998). Investor psychology and security market under-and overreactions. *the Journal of Finance*, 53(6):1839–1885.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, pages 703–738.

Dougal, C., Engelberg, J., Garcia, D., and Parsons, C. A. (2012). Journalists and the stock market. *Review of Financial Studies*, page hhr133.

Dungey, M., Fry, R., González-Hermosillo, B., and Martin, V. L. (2005). Empirical modelling of contagion: a review of methodologies. *Quantitative Finance*, 5(1):9–24.

Forbes, K. J. and Rigobon, R. (2001). No contagion, only interdependence: Measuring stock market co-movements. *Journal of Finance*, 57:2223–2261.

Foucault, T., Hombert, J., and Roşu, I. (2015). News trading and speed. *The Journal of Finance*.

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

Garcia, D. (2013). Sentiment during recessions. *The Journal of Finance*, 68(3):1267–1300.

Garcia, D. (2014). The kinks of financial journalism. *Available at SSRN 2517591*.

Greene, W. H. (2003). *Econometric analysis*. Pearson Education India.

Gropp, R., Duca, M. L., and Vesala, J. (2009). Cross-border bank contagion in europe. *International Journal of Central Banking*, 5(1):97–139.

Hirschey, N. (2013). Do high-frequency traders anticipate buying and selling pressure? *Available at SSRN 2238516*.

Hosmer Jr, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*, volume 398. John Wiley & Sons.

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*, volume 1, pages 221–233.

Jegadeesh, N. and Wu, D. (2013). Word power: A new approach for content analysis. *Journal of Financial Economics*, 110(3):712–729.

Kaminsky, G. L., Reinhart, C., and Vegh, C. A. (2003). The unholy trinity of financial contagion. Working Paper 10061, National Bureau of Economic Research.

Kelley, E. K. and Tetlock, P. C. (2013). How wise are crowds? insights from retail orders and stock returns. *The Journal of Finance*, 68(3):1229–1265.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 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(5):1049–1102.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

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

Lucey, B. and Sevic, A. (2010). Investigating the determinants of banking coexceedances in europe in the summer of 2008. *Journal of International Financial Markets, Institutions and Money*, 20(3):275–283.

Markwat, T., Kole, E., and van Dijk, D. (2009). Contagion as a domino effect in global stock markets. *Journal of Banking & Finance*, 33(11):1996 – 2012. Financial Globalisation, Risk Analysis and Risk Management.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of public economics*, 3(4):303–328.

Metiu, N. (2012). Sovereign risk contagion in the eurozone. *Economics Letters*, 117(1):35–38.

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

Missio, S. and Watzka, S. (2011). Financial contagion and the european debt crisis. CESifo Working Paper Series 3554, CESifo Group Munich.

Pericoli, M. and Sbracia, M. (2003). A primer on financial contagion. *Journal of Economic Surveys*, 17(4):571–608.

Ratcliff, J. W. and Metzener, D. E. (1988). Pattern-matching-the gestalt approach. *Dr Dobbs Journal*, 13(7):46.

Rigobon, R. (2002). *Contagion: How to Measure It?*, pages 269–334. University of Chicago Press.

Riordan, R., Storkenmaier, A., Wagener, M., and Zhang, S. S. (2013). Public information arrival: Price discovery and liquidity in electronic limit order markets. *Journal of Banking & Finance*, 37(4):1148–1159.

Shleifer, A. and Summers, L. H. (1990). The noise trader approach to finance. *The Journal of Economic Perspectives*, pages 19–33.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3):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(5):1481–1512.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467.

Thomadakis, A. (2012). Measuring financial contagion with extreme coexceedances. Technical report, School of Economics, University of Surrey.

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

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, pages 817–838.