Collective Intelligence in Financial Markets: Does consumer sentiment influence valuation of financial products?

By

Sergey Chekanskiy

M.S. International Finance
Institute of Business Studies, Moscow, 2010

SUBMITTED TO THE MIT SLOAN SCHOOL OF MANAGEMENT IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN MANAGEMENT STUDIES
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2012

Signature of Author: ______________________

May 11, 2012
MIT Sloan School of Management

Certified by: ______________________

Peter Gloor
Research Scientist
Thesis Supervisor

Accepted by: ______________________

Michael A. Cusumano
SMR Distinguished Professor of Management
Program Director, M.S. in Management Studies Program
MIT Sloan School of Management

Collective Intelligence in Financial Markets: Does consumer sentiment influence valuation of financial products?

By

Sergey Chekanskiy

Submitted to the MIT Sloan School of Management on May 11, 2012 in Partial
Fulfillment of the Requirements for the Degree of
Master of Science in Management Studies

ABSTRACT

This paper examines the relationship between the public mood associated with the economies of Italy, Spain and Greece, and prices of Credit Default Swaps on sovereign bonds of aforementioned countries. The effect of the changes in the public mood was measured by Granger causality tests and linear regression models. A price change prediction model was built based on the CART technique.

Results of the Granger tests suggest that constructed mood indices convey new and meaningful information about changes in CDS prices. Moreover, the extent to which this is true varies between countries. In the analyzed timeframe, public mood is a much better predictor for Spain than for Italy. Investigation of this difference revealed that there is a strong relationship between the mood associated with Spanish CDS and changes in the Italian CDS prices. This empirical evidence illustrates the spillover effect that troubles in one economy might have on another economy.

Thesis Supervisor: Peter Gloor

Title: Research Scientist

TABLE OF CONTENTS

ABSTRACT 1

LIST OF TABLES 3

1. INTRODUCTION 4

2. LITERATURE REVIEW 6

 2.1. Trading Philosophies 6

 2.2. Stock Market Predictions 7

 2.3. Macro-Economic Influence 8

 2.4. Early Indicators 8

 2.5. Mood tracking 9

3. OBJECTIVES, RESEARCH QUESTIONS AND HYPOTHESIS 12

4. METHODOLOGY AND DATA DESCRIPTION 14

 4.1. Data Collection 14

 4.2. Sentiment Extraction 16

 4.3. Sentiment Effect 18

5. RESULTS AND DISCUSSION 22

 5.1 Granger Test 22

 5.2 Linear Regression 23

 5.3 Cross Country Sentiment Effects 32

6. CONCLUSION AND RECOMENDATIONS 34

REFERENCES 36

APPENDICES 38

LIST OF TABLES AND FIGURES

1. Key Words..14

2. Distribution of Tweets per day..15

3. Percent of Tweets by Country Keyword..15

4. Summary of Lagged and Averaged Variables...17

5. Descriptive Statistics Sentiment..18

6. Granger Test Variable Pairs..19

7. Twitter Count and Sentiment-based Prediction Model...............................20

8. Sentiment Variable Groups...20

1. INTRODUCTION

The rapid growth of social media has opened numerous opportunities to study human behavior and social interactions. Social media has grown from community specific web-sites (like an early version of Facebook) into a global network of social interactions. Nowadays, almost every event anywhere in the world has a reflection in the social media, be it an article in the Wikipedia, a post on Twitter, a status update on Facebook or a simple Google search query. However, the biggest value of social media lies in the fact that millions of people freely discuss events or news, express their opinions, and give predictions or speculations online. When aggregated, this information may often result in decisions that are better that those made by every individual of the network (Surowiecki, 2004).

Stock market predictions in various forms have attracted both business and academia. Financial market is a fertile ground for research, primarily due to data availability and a high possible return on investment. The most intriguing quality of financial markets is that it is a result of social interactions. People make decisions to invest or divest and people provide this opportunity for others. This leads to an idea of behavioral economics and behavioral finance that even the most pragmatic decisions are affected by social, cognitive and emotional factors.

Even though every human investment decision is influenced by subconscious factors, the one specific decision that might be affected to a greater extent is the decision to buy insurance. Especially, insurance from the event is highly dependent on the behavior of the majority. A good example of this type of insurance is a credit default swap on a sovereign bond, which is a derivative security that insures the buyer from a case of a default of an entity. The difference between corporate and sovereign default events lies primarily in the amount of time those entities can postpone a default. Even with bad financials, a government can postpone a default for much longer periods of time, especially, if it manages to support investors' confidence. Sovereign default is a much more political event rather than an accounting one. In the highly globalized world a default of one country may have severe spillover effects on other countries. Moreover, no politician that plans to be reelected will make a decision to default if she/he has options to borrow. This leads to the fact that despite there may be many reasons for a sovereign default, it will not happen unless everyone loses confidence. This raises the importance of the insurance against a sovereign default and makes it a good potential speculative instrument.

The price for a CDS comes from the perceived probability of default of the underlying entity. The true default probability, obviously, depends on various objective factors. However, as the event of sovereign default is affected by the subjective perception of the true default probability by investors and the general public (investor confidence), the price of the insurance from this event should be affected as well.

This study answers the question whether the social media mood associated with particular country's economic situation has an influence on the change in the price of a CDS on this country's bond. In other words, can the social media be used to understand the pricing on the insurance from the sovereign default?

2. LITERATURE REVIEW

The pricing of financial assets has been a hot research topic since the establishment of financial markets. Numerous studies have been conducted to understand whether the stock market can be predicted. The topic is interesting not only for practical applications, but also from a theoretical standpoint. The predictability of the stock market would contradict the Efficient Market Hypothesis (EMH). There are three forms of effective market. First is weak efficiency, when only historical information is imbedded in the price of securities. Second is semi-strong efficiency, when news and all public information is reflected in the stock price. If market efficiency is strong then all information (including private) is instantaneously reflected in the price of a security. Subsequently, according to EMH stock market prices are for the most part affected by the new information, which is unpredictable.

Early papers on stock market predictions were based on a theory different from the EMH, but with a similar conclusion. The random walk theory concept was first used by K. Pearson in his study "The problem of the Random Walk" in 1905 and was applied to the stock market in 1973 by Burton Malkiel. The theory of random walks applied to stock prices says that the future price level of a security can be predicted with the same accuracy as a next number in the random number series (Fama, 1965)

2.1. Trading Philosophies

Both these theories have built the foundation for two distinct trading philosophies: a fundamental and a technical trading philosophy. Fundamentalists rely on analysis of nuts and bolts of the financial system. Fundamental analysis is the process of looking into numbers, derived from the economy, a particular sector or a company itself. The idea is that each stock has an intrinsic value that can be derived analyzing different ratios: liquidity, earnings, turnover etc. Comparing those ratios through time and across firms gives the insight into what determines the stock price and whether a particular instrument is fairly priced. The fundamental strategy looks more into the medium to long time frame. It compares the current stock price with the future stock price, when the market corrects itself and "the fair" price establishes. However, proponents on the EMH argue that intrinsic value of stock is equal to the current price.

Technicians, on the other hand, rely on a short term trading strategy and believe that market timing is crucial. Technical analysis looks at a historical and a time series data. It utilizes the idea that market timing opportunities may be found by comparing averaged historical prices and volume movements to the current price levels. The technical analysis uses terms like support and resistance levels to indicate price barriers where opportunities may exist. This trading philosophy is built around analysis of the market itself rather than a particular company. Technical analysis has three main assumptions. First, is that the market price and the volume reflect everything. Second, is that history repeats itself and the third is that prices move in trends. Technical analysis or the market timing strategy has one common assumption with the Effective Market Hypothesis; however, conclusions made in both cases are a bit different. Technicians believe that the stock price reacts to news slowly and since the market driving force is mostly human psychology - market prices show long-term trends that tend to repeat themselves and thus, can be predicted.

2.2. Stock Market Predictions

First studies on the topic tried to find an autocorrelation between past and present stock returns. Fama and French (1988) have studied returns on diversified portfolios of NYSE stocks from 1926 to 1985. The have found a strong negative autocorrelations on time intervals from 3 to 5 years. However, with exclusion of the 1926–1940 period from the study the autocorrelation disappears. Finally, Fama and French (1988) have come to a conclusion that irrational bubbles in stock prices cannot be distinguished from rational time-varying expected returns and, thus, stock prices likely to follow the random walk pattern.

More recent studies find much less support for the random walk theory. In the review of his earlier work Fama (1991) states that Efficient Market Hypothesis must be wrong. In their study of the Athens Stock Exchange Kavussanos and Dockery (2001) reject the EMH. Their findings are consistent with findings of Butler and Malaikah (1992), which as well reject the Efficient Market Hypothesis.

More recently Qian and Rasheed (2006) achieved 60% prediction accuracy for the Dow Jones Industrial Average Index. In their study they used using a combination of machine-learning classifiers—artificial neural network, decision tree and k-nearest neighbor.

2.3. Macro-Economic Influence

Another component for the understanding of pricing of the financial instruments is current macro-economic condition. In the past decade many studies were conducted on the topic of the links between stock market and macro-economic conditions. Gallagher and Taylor (2002) have found differences between aggregate demand and supply shocks, where former has only temporary effect and later have a more permanent effect on stock prices.

2.4. Early Indicators

Other research in the area shows that even if news is unpredictable, there are early indicators that can be extracted from the social media. Most studies were conducted for economic and commercial indicators; however the idea is true for other things as well. Gruhl, Guha, Kumar, Novak & Tomkins (2005) find that online chatter in the form of blog posts can be an early predictor for spikes in book sales at online retailer Amazon. Mishne & Glance (2006) study consumer sentiment towards movies and find that positive sentiment is a better predictor for a movie success than the volume of discussion alone. Their conclusion suggests that positive sentiment, included into a traditional predictive model for movie success improves its predictive power.

Schumaker and Chen (2009) examine news articles with different techniques for certain textual representations: bag of words, noun phrases, and named entities. The idea of this study starts from the fact that textual information from financial reports or breaking news can have dramatic effect on the price of security. This study utilizes more of the linguistic approach to the textual analysis, which is the main difference from most other studies in this area. Majority of previous studies in the area combined key words search with machine learning algorithms to assign certain securities' price movements to predetermined phrases and words. Their findings are that the inclusion of more precise textual representations into a predictive model yields better predictive power, meaning that there is meaning between the lines that are not being captured with a simple key words analysis.

There is no doubt that news affects security prices, but the public mood or a consumer sentiment might play no lesser role. Emotions play a big role in human decision-making. Emotion driven behavior may lead to poor choices and irrational decisions. As an example, George Ainslie (1975) shows that the impulsiveness makes people chose a suboptimal or an outright the worst of two alternatives. Wang (2011) illustrates that sense of complexity as a function of one's knowledge inversely affects the perception of risk.

If emotions affect individual decisions then the general level of a social mood has an effect on all kinds of business activities that involve individuals. Stock market can be taken as a metric for the social mood as well, but for the most part the stock market follows the public mood (Nofsinger, 2005). In his study Nofsinger identifies four stages of public mood: the rising mood, the peak positive mood, the declining mood, the peak negative mood. Each stage then is associated with certain emotional characteristics. For example, peak positive mood is associated with overconfidence, euphoria and trust. Peak positive mood leads to overconfidence making investors overstate growth opportunities causing securities to become overvalued leading to bubbles. The peak negative mood affects prices in the totally opposite way. Nofsinger finds that all business activities, become affected, but for some it takes longer to reflect changes in the public mood. For example, it takes time for the level of M&A and IPO to increase. However, it doesn't take as long for the business activity to decrease in case of the declining public mood and the stock market.

This idea is supported by Gilbert and Karahalios (2010) who derive emotional index from Live Journal posts and compare it to the S&P 500 Index. Using twenty thousand blog posts they build three indices of anxiety, worry and fear and test whether they convey information about the future stock price. Gilbert and Karahalios find that Anxiety Index has novel information about S&P 500 Index in 70% of analyzed days and one standard deviation in the Anxiety Index corresponds to about 0.4% of S&P returns.

2.5. Mood tracking

A reliable and scalable tool for assessment of public mood is required for the purpose of this paper. Historically public mood and public opinion were assessed using large scale surveys involving a representative sample from the total population. These kinds of surveys are extremely expensive and time consuming. But with the development of the internet and social media it became possible to extract the social mood directly from online media sources like Twitter, LJ etc. In the past years significant progress was made in the area of social media content analysis and particularly in the area of analyzing short text messages also known as Twitter Tweets.

Many automated tools for sentiment extraction of online texts are set up to identify just the polarity of the text, e.g. positive, negative or neutral. While even a simple metric like this can be successfully used in this study, the measurement of the scale of the emotion definitely conveys more information than just the polarity. For example, to identify misbehavior by a user, the algorithm has to be sensitive to the strength of the expressed emotions (e.g., Huang, Goh, & Liew, 2007).

One of the factors complicating the sentiment detection online is that in many cases people ignore certain grammar and spelling rules or use shortened versions of words and even sentences (Grinter & Eldridge, 2003). This is especially true for Twitter messages as their length is limited to 140 characters. Abbreviations can contain sentiment indicators ("lol", "rofl", etc) or strengthen sentiment ("omg" can be used as both bad and good sense).

The algorithm used for the purposes of this paper is called SentiStrength and it utilizes several new methods that capture both positive and negative sentiments simultaneously. This algorithm was originally build on the data sample of 2,600comments from MySpace and verified on about 1,040 comments. The main novel contributions of this algorithm are: a machine learning approach to optimize sentiment term weightings; methods for extracting sentiment from repeated letter non-standard spelling in informal text; and a related spelling correction method (Thelwall M., Buckley K., et al., 2010).

Applied algorithm shows about 61% accuracy in identifying positive and 73% accuracy in identifying negative sentiment, both based on strength scales from one to five and about 95% accuracy when measuring within plus or minus one class. The relative success of this algorithm is for the most part attributed to abilities to decode non-standard spellings and methods for boosting the strength of certain words.

While this algorithm is very promising there are still ways for improvement. Wilson et al., 2009 shows that sentiment extraction algorithms may be improved through linguistic processing, particularly with the dependency trees technique.

3. OBJECTIVES, RESEARCH QUESTIONS AND HYPOTHESIS

In a broad sense this paper aims to investigate whether public mood in relation to certain countries, extracted from social media has an influence on the valuation of certain financial products associated with those countries.

The recent European crisis and speculations on the topic of default probabilities of various European countries have motivated me to look into the relationship between the cumulative mood extracted from tweets associated with economic situations in three European countries: Italy, Spain and Greece and the Credit Default Swaps on these countries' sovereign bonds respectively.

On the one hand, CDS is a simple derivative contract. The buyer of the protection transfers the credit risk associated with an entity to the seller of the protection for a stream of premium payments. If a credit event occurs the seller of a CDS transfers the par value of the underlying bond to the buyer. A credit event is usually classified as a bankruptcy, a failure-to-pay or a restructuring. The payment for the protection is usually made quarterly and called the premium leg. The size of which is calculated from the quoted default swap spread, which is paid on the face value of the protection. Payments are made until the maturity of the underlying asset or until a credit event occurs, whichever happens first.

On the other hand, the gain or loss from a CDS position cannot be computed simply by taking the difference between current market quoted price plus received coupons and the purchase price. To value a CDS we need to use a term structure of default swap spreads, a recovery rate assumption and a model. The calculation of the value of a CDS requires a model because the riskiness of each premium payment has to be taken into account. It can be done by calculating the probability of the reference entity surviving to each premium payment date. These survival probabilities must be the arbitrage-free survival probabilities. These are the survival probabilities that are implied by the market default swap spreads (O'Kane and Turnbull, 2003). According to O'Kane and Turnbull such model must:

- Capture the risk of default of the reference entity;
- Model payment of the recovery rate as a percentage of the face value;

- Be able to model the timing of the default (especially important as the value of a default swap is the present value - all payments must be discounted to today),
- Be flexible enough to refit the term structure of quoted default swap spreads – the model should not generate any arbitrages;

The key component of this model should be the ability to capture the default probability of an underlying entity. There are two main approaches to the credit default modeling: the structural approach and the reduced form approach. The structural approach utilizes the ideal that a credit default is a result of some structural problems in the underlying entity, for example, shortage of the liquidity to cover immediate obligations. These models are usually extensions of Merton's 1974 firm-value model (O'Kane and Schloegl, 2001). Structural models say at what spread the bond should be trading, based on the internal characteristics of an entity. Thus, these models require information about the entity, be it a company or a country. This information doesn't come in the real time and it puts a hard limit of the usability of the structural approach.

In the reduced-form approach, the credit event process is modeled directly by modeling the probability of the credit event itself (O'Kane and Turnbull, 2003). Based on this approach, this probability of a default can be extracted from market prices.

This leads to the following idea: if the probability of a default of an entity is extracted from current market prices, which are affected by the public mood, as was shown in the previous chapter, then the public mood may be an early indicator of the default itself or of the change in the price of the protection against the default. Thus, the main research question of this paper is whether public mood convey any meaningful information about the valuation of CDS instruments on sovereign bonds of Italy, Spain and Greece.

4. METHODOLOGY AND DATA DESCRIPTION

The research consists of two main parts. First part is the data acquisition and the sentiment analysis. The second part is the analysis of the effects of the public mood on the CDS quotes.

4.1. Data Collection

Twitter has become very popular and it is being used to discuss a great variety of topics. This creates a problem when the result may get skewed by an effect that is far away from the economy or financial markets. Theoretically, taking every single Twitter message for the day and analyzing the total sentiment will produce the result that is biased towards a certain event that happened to attract the largest internet audience at that particular time, like a flash mob or news about a pop star. To minimize this effect I filter the data to refine only Tweets that are relevant to the topic i.e. those that talk about the economy, finance and markets of the three selected European countries.

The filtering was done on two dimensions. First, the content of the tweet should be relevant to the discussion about the state of the economy or financial market. Second, the tweet should be related to at least one of the selected countries: Italy, Spain or Greece or Europe in general. The filtering was done with the predetermined set of key words.

The first step for selecting key words was to identify all possible English words that are related to the state of economy or financial markets. The second step was to refine the list, leaving only most used words on Twitter. Using Trendistic[1] online service, that show how frequently a word is mentioned on Twitter, I have selected words that are being mentioned in more than 0.1% tweets for the day. The final list of key words is summarized in the following table:

Table 1 Key Words

Sentiment	Key Words
Neutral	economy, credit, employment, market, exchange, currency, FX, spread, yield, return
Negative	debt, default, bailout, crisis, unemployment, recession, spending
Positive	recovery, growth, expansion, improvement, regaining

Using the filtering approach described above I have obtained 195 604 Twitter tweets for the period from the 1st of January 2012 to 10th of March 2012. Sorting out tweets that contain non English or

[1] http://trendistic.indextank.com/

unreadable characters gives the final amount of 195 599 tweets. The distribution of tweets through the selected time period is shown in the next graph. To save space only weekends are marked on the X axis with "S" for Sunday and Saturday.

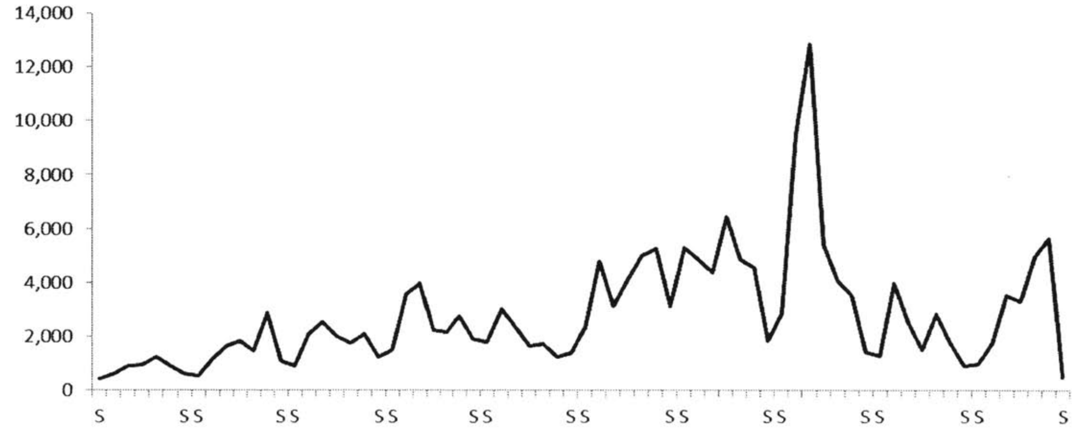

Figure 1. Distribution of Tweets per day

As expected, there is much less chatter about the market and the economy during weekends and much more during work days. At a later stage weekends were excluded from final models, because there is much less activity over the weekend and there are no trades happening over the weekend period.

Figure 2 illustrates the relative amount of tweets that contain country key words.

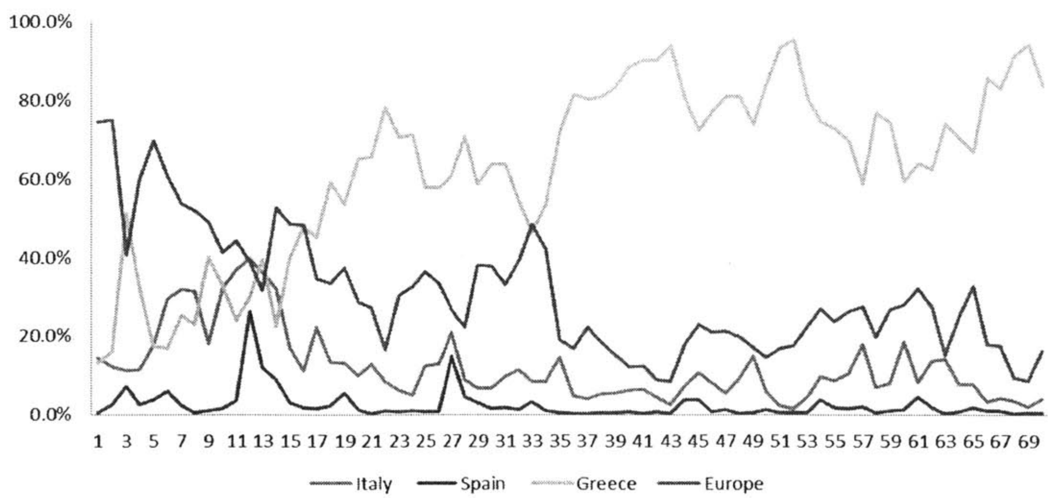

Figure 2. Percent of Tweets by Country Keyword

4.2. Sentiment Extraction

To extract the sentiment data from Twitter texts, I use the software called SentiStrength. It is developed by Mike Thelwall, Kevan Buckley, and Georgios Paltoglou from Statistical Cybermetrics Research Group in School of Computing and Information Technology, University of Wolverhampton. This software allows the computation of different sentiment metrics, from a simple one-dimensional positivity/negativity index to three-dimensional scale and trinary indices, which are discussed below.

It seems reasonable to include a separate metric for both the positive and the negative sentiment in the model. The research done by Fox (2008) shows that the positive and the negative sentiment can coexist and be relatively independent from each other. Same results were confirmed by Huppert and Whittington (2003) when levels of sentiment are not extreme and over long periods of time.

In this paper I use five different sentiment metrics, the first four are obtained with the help of the SentiStrenght software, and the fifth is manually constructed:

- Negativity Index – Captures only the negativity sentiment of the sentence. It changes from -1 to -5 depending on how overly negative the text is. The value of minus one means that it is non-negative and minus five means the text has very negative sentiment.
- Positivity Index – Similarly to the previous index, it captures only the positivity sentiment of the sentence. It changes from 1 to 5 depending on how overly positive the text is.
- Trinary Index – This index reflects both the positive and the negative sentiment. The name comes from the fact that unlike a binary index would do it includes a third – the neutral sentiment. This variable has three possible states: one, zero and negative one. One means that the text is more positive than negative; zero means it is neutral and negative one means that the overall sentiment is more negative than anything else.
- Scale Index – In essence this variable is similar to the previous one, but its value is allowed to change from -4 to +4, meaning it reflects the magnitude of the sentiment as well as its overall sign.
- Daily Mentions – This variable reflects how many times a particular country was mentioned on Twitter on a particular day. Every Tweet in the dataset that contains

country's name or its derivative (Greece – Greek) counts as 1 for this country on that day. Tweets containing mentions for more than one country count as 1 for every country they mention.

One more thing that affects the effectiveness of sentiment variables is the time lag. According to the EMH, strong efficiency means that the effect of the public information on the market price should be immediate, however, this idea was challenged numerous times and there is enough evidence to believe that there are certain inefficiencies that vary from market to market. The possible presence of market inefficiencies makes it reasonable to test lagged and averaged versions of created sentiment variables. I have chosen five working days as the maximum lag period and the maximum averaging period. There are two main reasons for that. The first is a relatively small data sample (49 days). The second is that a week (five working days) is a long time for the highly speculative market like the sovereign CDS market and in this environment any public information older than a week is very likely to be already priced in.

At this point it is unclear whether the nominal value or the percent change of the constructed variables is the best predictor of CDS prices. Thus, I test it for every combination of Index type/Lag amount/Change vs. Nominal value. Those combinations are summarized in the table below:

Table 2 Summary of Lagged and Averaged Variables

	Lag 1 to 5 Days	Percent Change for Lagged Variables Values	1 - 5 Days Moving Average	Percent Change of Moving Average Values	**Total**
Mentions	5	5	x	x	**10**
Positive	5	5	5	5	**20**
Negative	5	5	5	5	**20**
Trinary	5	5	5	5	**20**
Scale	5	5	5	5	**20**
Total	**25**	**25**	**20**	**20**	**90**

Same variables are constructed for each of the three selected countries for the total of 270 sentiment variables.

The table below presents descriptive statistics for four nominal non-lagged sentiment metrics constructed with SentiStrength software for each country.

Table 3 Descriptive Statistics Sentiment

Variable	N	Minimum	Maximum	Mean	Std. Deviation
Greece Positivity	49	1.1076	1.4431	1.235906	.0731188
Greece Negativity	49	-2.4909	-1.4503	-1.891210	.2113236
Greece Trinary	49	-.8790	.0200	-.541341	.1857300
Greece Scale	49	-1.2661	-.0072	-.655316	.2479055
Spain Positivity	49	1.0152	1.7128	1.217269	.1567355
Spain Negativity	49	-3.0702	-1.2360	-1.913771	.3952563
Spain Trinary	49	-.9787	.0957	-.502186	.2422994
Spain Scale	49	-1.8772	.3511	-.696508	.4193671
Italy Positivity	49	1.1410	1.6751	1.359284	.1089265
Italy negativity	49	-2.4686	-1.2958	-1.712653	.2514504
Italy Trinary	49	-.7124	.0962	-.287765	.1826928
Italy Scale	49	-1.2076	.0865	-.353373	.2791707

During the analyzed time period Spain received the biggest number of negative comments and, thus, has the lowest minimum negative sentiment value, followed by Greece and Italy. Not surprisingly, countries with the lower negativity score have the higher positivity index.

Looking at the Scale Index we see that all three countries have overall negative sentiment, however, Italy has the significantly higher score than both Spain and Greece.

Another interesting observation is that on the Scale basis Spain has much more volatile sentiment, with standard deviation of .419 versus .279 and .247 for Italy and Greece.

4.3. Sentiment Effect

I have used two approaches to find out whether the filtered twitter mood has a cause-effect relationship with prices of CDS on bonds of the selected European countries.

1. The first approach is the bivariate Granger causality test. If the variable X Granger-causes the variable Y then past values of X help better predict Y than past values of variable Y itself. Its mathematic calculations are based on linear regression modeling of stochastic processes (Granger 1969). To test the Granger-cause effect for k amount of lags we first model:

$$Y_t = \beta 0 + \beta_1 Y_{t-1} + \cdots + \beta_k Y_{t-k} + e_t$$

Then we ask if adding similar information about the variable X improves our prediction of Y. The new equation is the following:

$$Y_t = \beta 0 + \beta_1 Y_{t-1} + \cdots + \beta_k Y_{t-k} + \gamma_1 X_{t-1} + \cdots + \gamma_k X_{t-k} + e_t$$

In the latter model β coefficients do not provide any information, but if the γ coefficients are jointly significant, then we have an established cause.

The hypothesis that is tested here is whether sentiment indices provide useful information for projection of the future CDS price. The bivariate model is specified in a way where the Y variable is the daily change in the five year price of CDS protection and X variables are each of the individual sentiment indices, detailed in the previous section. The test is performed in the R software with the means of the free package "vars". The nature of the test requires that we build pairs of variables to be used in the Granger-cause test. For that purpose every sentiment variable is paired with the daily percent change of CDS prices. Each of the ten variable pairs is tested for each country. Below is the summary of tested variable pairs:

Table 4 Granger Test Variable Pairs

Pair 1	Pair 2	Pair 3	Pair 4	Pair 5
5yr CDS Percent Change	5yr CDS Percent Change	5yr CDS Percent Change	5yr CDS Percent Change	5yr CDS Percent Change
Absolute Value of Daily mentions	Percent Change of Daily mentions	Absolute Value of Negativity Index	Percent Change of Negativity Index	Absolute Value of Positivity Index

Pair 6	Pair 7	Pair 8	Pair 9	Pair 10
5yr CDS Percent Change	5yr CDS Percent Change	5yr CDS Percent Change	5yr CDS Percent Change	5yr CDS Percent Change
Percent Change of Positivity Index	Absolute Value of Trinary Index	Percent Change of Trinary Index	Absolute Value of Scale Index	Percent Change of Scale Index

Lag parameters selected in the model are from one to five days, which is in line with previous studies on a similar subject (Gilbert and Karahalios, 2009).

2. The second approach is a linear regression summarized in the following equation:

$$CDS5_{i,t} = Intercept + \delta Count_{i,t} + \beta Sent_{i,t} + \beta Fin_{i,t-1} \qquad (1)$$

Where:

$CDS5_{it}$ The change in the price of the 5 year CDS on the bond of the country i at time t

$Count_{i,t}$ Mentions count variables for the country i in the day t

$Fin_{i,\,t-1}$ Changes in the Dow Jones index for country i at time $t-1$ and changes in the CDS prices for country i at time $t-1$

$Sent_{i,t}$		Sentiment variables, described in the section "4.2 Sentiment Extraction"

In the Linear Regression analysis I am comparing regressions that include sentiment analysis versus "baseline models" - regressions that do not include sentiment analysis. Figure 3 gives a representation of this idea:

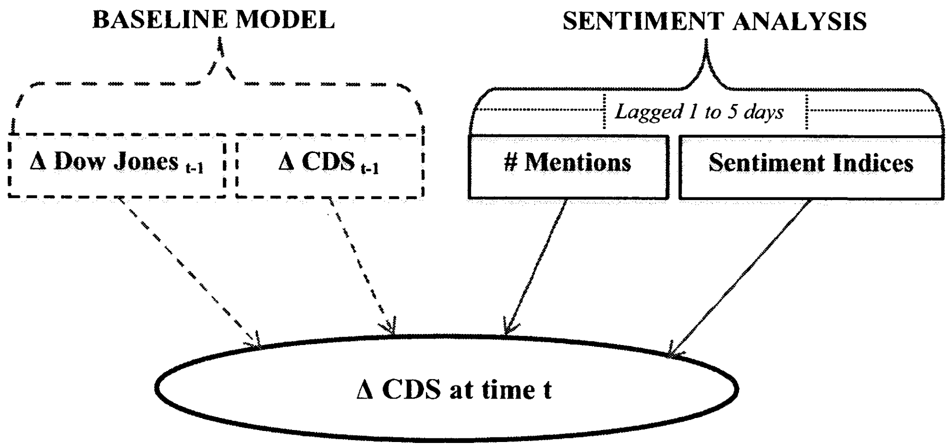

Figure 3. Twitter count and sentiment-based prediction model

Based on results of the correlation analysis of sentiment variables (see Appendix) these variables were divided into twelve groups to be tested if they can improve the baseline model.

Table 5 Sentiment Variable Groups

Group 1	Group 2	Group 3	Group 4
Lagged Nominal Amount of Daily Mentions	Lagged Changes in Daily Mentions	Lagged Changes in Daily Mentions	Lagged Changes in Daily Mentions
Lagged Nominal Negativity Index Level	Lagged Changes in Negativity Index	Moving Averages of Negativity Index	Changes in Moving Averages of Negativity Index
Lagged Nominal Positivity Index Level	Lagged Changes in Positivity Index	Moving Averages of Positivity Index	Changes in Moving Averages of Positivity Index

Group 5	Group 6	Group 7	Group 8
Lagged Nominal Amount of Daily Mentions	Lagged Changes in Daily Mentions	Lagged Changes in Daily Mentions	Lagged Changes in Daily Mentions
Lagged Nominal Full Scale Index Level	Lagged Changes in Full Scale Index	Moving Averages of Full Scale Index	Changes in Moving Averages of Full Scale Index

Group 9	Group 10	Group 11	Group 12
Lagged Nominal Amount of Daily Mentions	Lagged Changes in Daily Mentions	Lagged Changes in Daily Mentions	Lagged Changes in Daily Mentions
Lagged Nominal Trinary Index Level	Lagged Changes in Trinary Index	Moving Averages Trinary Index	Changes in Moving Averages of Trinary Index

Each lagged variable has five lag variations starting from one day lag to five day lag. Moving Averages start with two days averages and end with five day averages.

I used the classification regression trees technique (CART) as an extension to the linear regression approach to test whether variables identified in the regression model can be used to predict the movement of the CDS price. The CART model is a flexible method for specifying a conditional distribution of a dependent variable based on a given set of predictor variables x_1, x_2, etc. The model finds certain rules to divide the data into subsets where the distribution of the dependent variable is more homogeneous. All data points are assigned to a specific terminal node and every terminal node is defined by a set of rules. The model requires one additional input to figure the rules to divide the data and this input is the minimum number of observation that has to be assigned into one terminal node. Since the total amount of observations in the dataset is 49 this number should be relatively small to allow for some flexibility in the model. On the other hand it cannot be too small or otherwise there is a risk of over fitting the model.

For the CART model I have constructed a new binary dependent variable that is equal to 0 if the change in the CDS price for the day is negative and is equal to 1 if the change is positive. I have divided the dataset (49 observations) into the training part (28 observations) and the testing part (21 observations). I have used "rpart" package in the R software to build a CART model using the training set and checked the obtained model on the testing set to determine the accuracy of its predictions.

5. RESULTS AND DISCUSSION

5.1 Granger Test

The Granger causality test models relationships between variables linearly, which is a limitation of the model. However, it serves the main purpose of this paper, which is to capture the relationship between the sentiment and the security price and understand which sentiment measurement technique gives the most accurate result.

The major result of the Granger test is that sentiment variables do convey new information about changes in the CDS price. Moreover, almost universally across tested variables, lags that are longer than three days were insignificant. Best results were obtained with lags from one to three days long. This is intuitive given the fast paced nature of markets and their ability to absorb information very quickly. Another interesting finding is that nominal values of sentiment indices generally better explain variations in CDS prices than percent changes in the same variables.

Tests for Spanish Data

In the Spanish example the best predictor variable is the Scale Index lagged one day ($p = 0.007$ and $F = 7.6$), which remains significant at 5% level even with the increase in the lag parameter to three days. The second best predictor variable is the nominal Negativity Index ($p = 0.0048$ and $F = 8.3$), which remains significant at 5% level with lags extending to five days. The third best variable is the Mentions Count variable. But, this variable becomes significant only when lagged one day. Interestingly, it is the only variable which remains significant when taking its changes instead of nominal values. The last best variable for Spain is the Trinary Index, which is by its nature simpler version of the Scale Index. Trinary Index variable is significant at 10% level with lags of one and two days. The last observation is that changes in Scale Index are significant at 10% level only when taking its four days lag, which is interesting, but most likely, can be attributed to the small size of the sample. Full results of the Granger tests are presented in the Appendix.

Tests for Italian Data

Results of the Granger tests for Italian data are surprisingly different from the Spanish results. The price of Italian CDS is much less sensitive to the Twitter mood towards Italian economy. Only two variables convey new information about changes in the dependent variable. Negativity Index is significant when lagged one day ($p = 0.222$, $F = 5.4$). Positivity Index, which was insignificant for Spanish data, is now significant when lagged three days, ($p = 0.05$ and $F = 2.71$). All the other sentiment variables remain not significant with any of the selected lag parameters.

The nature and possible causes of this difference between two countries are investigated in the last part of this section.

5.2 Linear Regression

This approach requires an establishment of a baseline model to be able to capture any improvements that sentiment variables might add. The equation of a baseline[2] model used for the comparison is given below:

$$CDS5_{i,t} = Intercept + \beta Fin_{i,t-1} \qquad (2)$$

Where

$CDS5_{i,t}$	The change in the price of the 5 year CDS on the bond of the country i at time t
$Fin_{i,\,t-1}$	Changes in the Dow Jones index for country *i* at time *t-1* and changes in the CDS prices for country *i* at time *t-1*

Interestingly, Spain is the only country where lagged changes in Dow Jones (Spain) index and lagged changes in the CDS price are good explanatory variables for the current changes in CDS prices. The lagged change in the Dow Jones Index variable is significant at 10% level and the lagged change in the CDS price is significant at 1%, where the overall model has an R square of .33, explaining about 33% of the changes in the dependent variable. Italian and Greek baseline models are far less successful,. The possible explanation of this difference is discussed in the last section of this chapter.

[2] Detailed results for the baseline models for each country are presented in the Appendix.

Once baseline models are set up, the next step is to include sentiment related variables and identify if the quality of the new model improves. Below is a discussion of results for all 216 linear regressions identified above.

Variables in the later discussion are coded to conserve space. Each variable starts with a three letter country code, where "Spa" = Spain, "Ita" = Italy, "Cee" = Greece. The number after the "_" sign identifies how many days the variable was averaged in case of averaged variables and the amount of lag in all other cases. After the number goes the code for the variable type and if it is lagged or averaged, where "DM" = Daily Mentions, "DLP" = Lagged Positivity Index, "DALP" = Averaged Positivity Index. Same logic applies for "DLN", "DALN", being Lagged Negativity Index and Averaged Negativity Index respectively. Similarly, "DLT", "DALT", "DLS" and "DALS" being Lagged Trinary, Averaged Trinary, Lagged Scale and Averaged Scale Indices. If the variable has "_CH" at the end it is expressed as a percent change over the previous day $[(Var_t - Var_{t-1})/Var_{t-1}]$ and it is a nominal value if otherwise.

Spain Group 1

With the inclusion of sentiment variables from the first group the overall R square of the model has improved. The highest increase of R square was achieved with two day lagged variables: R^2 of .395, which is an improvement of .061 over the baseline model. However, individual sentiment variables are not significant even at the 10% level.

Coefficients Spain Group 1

Model ($R^2 = 0.395$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.036	.035		1.020	.313
	Spa_DJ_CH_Lag1	-1.103	.304	-.468	-3.627	.001
	Spa_CDS5_CH_Lag1	.239	.132	.252	1.812	.077
	Spa_2DM	-.001	.001	-.166	-1.379	.175
	Spa_2DLP	-.007	.026	-.036	-.270	.788
	Spa_2DLN	.014	.009	.180	1.478	.147

Spain Group 2

The second group of variables has showed the same dynamic, however, the biggest increase in the R square was achieved with three days of lagged variables: R^2 of .436, which is an improvement of .102 over the baseline model. Even though, individual sentiment variables are still not significant at the 10% level, the overall significance of sentiment variables has increased.

Coefficients Spain Group 2

Model (R^2 = 0.436)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.002	.004		.608	.546
	Spa_DJ_CH_Lag1	-1.309	.295	-.555	-4.434	.000
	Spa_CDS5_CH_Lag1	.172	.117	.181	1.468	.150
	Spa_3DM_CH	-.005	.003	-.199	-1.679	.101
	Spa_3DALP_CH	-.039	.024	-.199	-1.671	.102
	Spa_3DALN_CH	-.016	.013	-.143	-1.195	.239

Spain Group 3

Results for the third group are very similar to results of the previous one, with the exception that Mentions variable lagged 3 days is significant at the 10% level (almost 5% level in this model). The overall R^2 of .431 is almost similar to .436 in the previous model.

Coefficients Spain Group 3

Model (R^2 = 0.431)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.092	.048		1.943	.059
	Spa_DJ_CH_Lag1	-1.192	.288	-.505	-4.142	.000
	Spa_CDS5_CH_Lag1	.147	.131	.154	1.118	.270
	Spa_3DM_CH	-.007	.003	-.242	-2.003	.052
	Spa_3DAP	-.040	.039	-.138	-1.013	.317
	Spa_3DAN	.022	.017	.166	1.310	.197

Spain Group 4

The best model in the fourth group has R^2 of .414 and it includes variables lagged three days and three days moving averages of positivity and negativity indices. The only significant variable is still daily mentions change lagged three days.

Coefficients Spain Group 4

Model ($R^2 = 0.414$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.003	.004		.690	.494
	Spa_DJ_CH_Lag1	-1.251	.302	-.530	-4.139	.000
	Spa_CDS5_CH_Lag1	.190	.123	.200	1.544	.130
	Spa_3DM_CH	-.006	.003	-.239	-1.889	.066
	Spa_3DAP_CH	-.084	.059	-.176	-1.423	.162
	Spa_3DAN_CH	-.022	.040	-.078	-.562	.577

Spain Groups 5, 6, 7, 8

Next four groups introduce the new sentiment variable that includes effects of both negativity and positivity indices. The Scale variable changes between -4 and +4 capturing the overall mood and its magnitude.

The fifth group did not show any significant improvements.

The sixth group showed significant improvements in the overall fit of the model: the best model R^2 is .449 and individual sentiment variables are significant at 10% and 1% levels.

Coefficients Spain Group 6

Model ($R^2 = 0.449$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	.003		-.139	.890
	Spa_DJ_CH_Lag1	-1.086	.265	-.483	-4.090	.000
	Spa_CDS5_CH_Lag1	.152	.111	.164	1.368	.179
	Spa_4DM_CH	-.007	.004	-.231	-1.958	.057
	Spa_4DALS_CH	.008	.002	.363	3.160	.003

Notably, these results were achieved with sentiment variables lagged four days, unlike the 3 days lag in the previous models. However, taking longer lags did not produce any better results for this model. Substituting four days lagged Mention Change variable with three days lag of the same variable, which was significant in previous models has improved the result. In the model below the R^2 equals .453 which is 3 points higher than in the previous model and the 3 days lagged mention change variable is significant at 5% level.

Coefficients Group 6 Custom

Model ($R^2 = 0.453$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.000	.003		.056	.956
	Spa_DJ_CH_Lag1	-1.063	.263	-.473	-4.040	.000
	Spa_CDS5_CH_Lag1	.152	.111	.164	1.374	.177
	Spa_3DM_CH	-.006	.003	-.238	-2.036	.048
	Spa_4DALS_CH	.008	.002	.358	3.127	.003

Groups number seven and eight have produced much worse results than the group six and thus are not discussed in detail here.

Spain Groups 9, 10, 11, 12

The next set of groups of variables introduces trinary scale index, which has three possible states: -1/0/1 for negative/neutral/positive sentiment respectively.

Group nine started showing improvements over the baseline model with inclusion of two and three day lagged sentiment variables. Respective R squares for those models are .417 and .420. In both models the trinary variable is significant at 5% level.

Coefficients Spain Group 9

Model ($R^2 = 0.417$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.011	.010		1.163	.251
	Spa_DJ_CH_Lag1	-1.017	.287	-.431	-3.546	.001
	Spa_CDS5_CH_Lag1	.188	.124	.198	1.524	.135
	Spa_3DM	8.297E-005	.000	.127	1.048	.300
	Spa_3DLT	.031	.014	.265	2.148	.037

Group ten did not show any significant improvements over the baseline model.

Group eleven has started showing improvements over the baseline model with inclusion of three day lagged sentiment variables. Its R square is .437. And all variables are significant at least 10% level. The main difference from the results of the group nine is that with three days moving averages of Trinary index the Daily Mentions index becomes significant at 10% level.

Coefficients Spain Group 11

Model ($R^2 = 0.437$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	.027	.012		2.217	.032
	Spa_DJ_CH_Lag1	-1.143	.281	-.485	-4.069	.000
1	Spa_CDS5_CH_Lag1	.224	.116	.235	1.930	.060
	Spa_3DM_CH	-.006	.003	-.218	-1.865	.069
	Spa_3DAT	.050	.023	.255	2.201	.033

Group twelve did not show any improvements over the baseline model and is not discussed in details.

To sum up the results for Spain the best results were achieved with three to four day lagged variables. The positivity and the negativity indices were not significant in any of models. Best results were achieved with Scale and Trinary variables that incorporate both the negativity and the positivity sentiment in the one variable. However, there is a difference between the two. The nominal level of the Scale variable is insignificant in all models, where Changes in the Scale Index become significant at a certain lag point and overall improve the model. The opposite is true for the Trinary index. Which makes sense, because Trinary index does not capture the magnitude of the sentiment change and it only indicates the direction. Interestingly, three days moving average of the Trinary index yields better results than the three days lag of the same variable.

Taking into account the results of the individual regression groups I have identified three variables that explain the change in dependent variable the best and I used them together in the last regression (results are presented below). The inclusion of the three most successful variables from individual regressions yielded an R^2 of .521, which is the biggest improvement over the baseline model so far.

Additionally, in the resulting regression every variable is significant at least at the 10% level and most are significant at 1% or 5% levels.

Coefficients Spain Final Regression

Model (R^2 = 0.521)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.025	.011		2.312	.026
	Spa_DJ_CH_Lag1	-1.065	.249	-.474	-4.270	.000
	Spa_CDS5_CH_Lag1	.180	.106	.193	1.701	.096
	Spa_3DM_CH	-.006	.003	-.252	-2.279	.028
	Spa_4DALS_CH	.007	.002	.343	3.162	.003
	Spa_3DAT	.048	.020	.263	2.410	.021

Pearson correlation between 4DALS_CH, 3DM_CH and 3DAT variables is less than 0.08

Notably, results of linear regressions for Spain are in line with results from Granger tests. Same variable types work better than others. The major difference is that in linear regressions the "change" variables play a significant role, whether in the Granger causality test those were not significant.

CART Model

To test whether identified variables can be used to predict the movement of the CDS prices I have constructed a binary variable that is equal to 0 if the change is negative and equals to 1 if the change is positive for the given day. Then I have divided the dataset (49 observations) into the training part (28 observations) and the testing part (21 observations). I have used "rpart" package in the R software to build a CART model using the training set and have checked the obtained model on the testing set to determine the accuracy of its predictions.

The model was able to correctly predict 71% of the data points in the training set. Meaning it correctly predicted the direction of the change in CDS prices on Spanish bonds 71% of the time.

Italy Group 1

Group one of sentiment variables did not yield any significant improvements over the baseline model and thus is not discussed in detail.

The second group of sentiment variables showed no improvement until five day lagged variables were included. R^2 has increased to .139, however only the 5 days lagged negativity index variable was significant in the model.

Coefficients Italy Group 2

Model ($R^2 = 0.139$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.008	.005		-1.621	.113
	Ita_CDS5_CH_Lag1	.021	.216	.021	.098	.923
	Ita_DJ_CH_Lag1	-.077	.356	-.047	-.215	.831
	Ita_5DM_CH	.008	.028	.041	.272	.787
	Ita_5DALP_CH	-.052	.041	-.192	-1.243	.221
	Ita_5DALN_CH	-.057	.028	-.312	-2.073	.045

The third group showed exact same dynamic as the second one. There were no improvements until five days lagged variables were included and only Moving Average of the Negativity index was significant in these models.

Coefficients Italy Group 3

Model ($R^2 = 0.134$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.277	.168		1.652	.106
	Ita_CDS5_CH_Lag1	-.011	.214	-.012	-.054	.957
	Ita_DJ_CH_Lag1	-.074	.359	-.045	-.206	.838
	Ita_5DM_CH	.042	.030	.226	1.389	.173
	Ita_5DAP	-.073	.082	-.159	-.894	.377
	Ita_5DAN	.110	.048	.445	2.310	.026

The fourth group of sentiment variables did not yield any significant improvements over the baseline model and thus is not discussed in detail.

Group five of sentiment variables did not yield any significant improvements over the baseline model; however, it's worth mentioning that in these models only variables lagged 5 days back were significant.

Group six did not significantly improve over the baseline model, but interestingly, the Scale variable became significant at 10% level only when lagged three days back, as opposed to five day lags that worked better so far for Italy.

Coefficients Italy Group 6

Model ($R^2 = 0.096$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	-.006	.005		-1.238	.222
	Ita_CDS5_CH_Lag1	-.058	.223	-.059	-.257	.798
1	Ita_DJ_CH_Lag1	-.312	.357	-.194	-.874	.387
	Ita_3DM_CH	.002	.019	.016	.103	.918
	Ita_3DALS_CH	-.001	.001	-.283	-1.838	.073

Groups seven, eight, nine, ten and eleven of sentiment variables did not yield any significant improvements over the baseline model and thus are not discussed in detail.

Coefficients Italy Group 12

Model ($R^2 = 0.132$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
	(Constant)	-.008	.005		-1.550	.129
	Ita_CDS5_CH_Lag1	-.025	.218	-.025	-.116	.909
1	Ita_DJ_CH_Lag1	-.248	.344	-.157	-.719	.476
	Ita_2DM_CH	.007	.015	.067	.447	.657
	Ita_2DAT_CH	.002	.001	.359	2.358	.023

Summing up results for Italy, four variables were found significant when explaining variation in changes of CDS prices for Italian bonds. Those variables are the Change in Three Days Moving Average of the Scale Index (3DALS_CH), the Change in Two Days Moving Average of the Trinary Index (2DAT_CH), the Five Days Moving Average of the Negativity Index (5DAN) and the Change in Five Days Lagged Negativity Index (5DALN_CH). Including all four of those variables in one regression significantly improves the overall fit of the model ($R^2 = .250$) and all variables, except the Five Days Moving Average of the Negativity Index (5DAN), remain significant at the 10% level.

Coefficients Italy Custom Group

Model ($R^2 = 0.250$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.041	.061		.666	.509
	Ita_3DALS_CH	-.001	.000	-.249	-1.803	.079
	Ita_2DAT_CH	.004	.002	.242	1.741	.089
	Ita_5DAN	.030	.036	.120	.823	.415
	Ita_5DALN_CH	-.048	.026	-.260	-1.869	.069

Pearson correlations between those variables are less than 0.1

Results of linear regressions for Italy are perfectly in line with Granger test's results. We see the same pattern where only Negativity index is significant in most models.

5.3 Cross Country Sentiment Effects

Clearly, in the case of Italy, the sentiment variables work worse than in the case of Spain. One of the explanations might be that prices on Italian CDS are affected by the situation in other, more troubled European countries, like Spain, for example. To test this hypothesis I have included the variables from Spanish model with the best fit into the Italian model with the best fit. And by eliminating insignificant variables the model I got was the following:

Coefficients Italy Spanish Influence Model

Model ($R^2 = 0.591$)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.002	.003		-.695	.491
	Spa_DJ_CH_Lag1	-1.403	.251	-.576	-5.596	.000
	Spa_CDS5_CH_Lag1	.159	.103	.163	1.553	.128
	Spa_3DM_CH	-.007	.003	-.247	-2.373	.022
	Spa_3DALS_CH	-.005	.003	-.189	-1.855	.071
	Ita_5DALN_CH	-.037	.019	-.195	-1.928	.061

The final model for Italy that includes Spanish public mood variables has the R^2 more than double of the previous one. Moreover, Spanish variables are significant at 5% and 10% levels and only one Italian variable has remained significant: it is the Change in the Five Days Lagged Negativity Index (5DALN_CH). This result supports the notion that troubles in Spain can become a trigger to the escalation of the European crisis, driving Italy very close to bankruptcy. Spanish news is more important for the European market, thus, making Spanish sentiment variables better predictors of the Italian CDS prices. Moreover, including lagged the Dow Jones Index changes for Greece, Italy and Spain in the above model shows that changes in the Italian Index are not significant predictors of the Italian CDS prices, whether changes in Spanish and Greek indices are significant at 1% level. The overall fit of the model improves to the R^2 of .615.

The next table presents the results of a test whether Greek related mood variables have an effect on Italian CDS prices.

Coefficients Greece / Italy

Model (R^2 = 0.209)		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.064	.098		-.652	.518
	Cee_2DM	6.622E-007	.000	.044	.325	.747
	Cee_2DLP	-.044	.061	-.104	-.725	.472
	Cee_2DLN	-.059	.022	-.398	-2.731	.009

a. Dependent Variable: Ita_CDS5_CH

The overall fit of this model is not impressive to say the least R^2 = .209. However, one interesting insight is that the two days lagged negativity Index for Greece is a significant variable, meaning it partially explains changes in the Italian CDS price.

6. CONCLUSION AND RECOMENDATIONS

In this paper I have examined the relationship between the public mood associated with a set of European countries and prices on the CDS for the respective sovereign bonds. The timeframe of the study consists of 49 workdays starting from January 2^{nd}, 2012. The dataset is comprised of 195,599 Twitter Tweets. There were constructed five different mood indices, using nominal, percent changes and moving average values of those for the total of fourteen unique sentiment variables.

The relationship between public mood or consumer sentiment associated with a given economy and prices of the protection from the default of this country exists. The results of this study show that the public mood has a Granger-cause effect on CDS prices. This is further supported by the results of regression models including one to five days lagged sentiment variables. Comparison between results of models with different days of lagged variables shows that sentiment variables lagged from one to three days generally produce the best result. Variables with longer lags (four and five days) do not improve the model and are usually insignificant. This is not counterintuitive as financial markets are fast-paced by the nature and are known to absorb the new information very quickly. On the other hand, this study shows that the European market for CDS is not strongly efficient in terms of the Efficient Market Hypothesis. If the market was strongly efficient there would be no Granger-cause effect of the public mood and securities prices, as the effect of any news or the new information would be realized instantaneously and reflected in the current market price.

Not all employed sentiment indices proved to be good predictors of the CDS prices. The best results were achieved with the Scale Index that changes from -4 to +4 depending on how happy or unhappy the mood is. This allows capturing both the negative and the positive mood effects in the same variable and shows the magnitude of the sentiment. The Scale variable proved to be more effective than Trinary Index that has only three values: negative/neutral/positive. Another finding is that the simple negativity index conveys more information about changes in CDS prices than a simple positivity index. This supports the observation that the market quickly reacts to the bad news, but takes longer time to rebound on the better ones. The measure of the amount of chatter about a certain economy (Daily Mentions variable) was found to be important for prediction of CDS prices. In the case of Spanish regression the Changes in Daily mentions had negative coefficient, meaning that the less chatter more uncertainty and the higher the price for the protection, which is intuitive.

Another interesting finding is that there are cross country links and so called spillover effect may be seen in the effect that the mood associated with one country has on another country's CDS prices. In particular, there is a strong relationship between the mood towards Spanish economy and the price of CDS on Italian bonds. The model including Spanish sentiment variables achieved the highest R^2 (0.615) of all models for Italy.

The inclusion of the public mood variables in prediction models for prices of various market securities has a great potential. A simple CART model, build to predict the direction of the change in price for Spanish CDS, which included sentiment variables, had an accuracy of 71%, meaning that in seventy one percent of cases it guessed the increase or the decrease in the price of the CDS correctly.

The main limitation of the paper is the small data sample and the scope of the study limited to a small subset of countries. In the future study I would like to look at the same relationship at a longer time frame and across more countries. This will allow looking closer at cross-country links and possible cross-country spillover effects. Results of this paper may be used to build a prediction model for CDS prices on sovereign bonds of Spain and Italy. Or it may be used to improve existing models to better reflect market inefficiencies in relation to public mood or consumer sentiment expressed on social networks.

REFERENCES

Ainslie G., (1975). "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control". *Psychological Bulletin*, VOL. 82, No. 4

Bo Qian and Khaled Rasheed, (2007). "Stock Market Prediction with Multiple Classifiers", *Applied Intelligence: The International Journal of Artificial Intelligence*, Neural Networks and Complex Problem-Solving Technologies, 26(1), pp. 25 - 33.

Fama, E. F. and and Kenneth R. French, (1988). "Permanent and temporary components of stock prices", *Journal of Political Economy* 96, 246-273.

Fama, E.F., (1965). "The behavior of stock market prices", *The Journal of Business* 38, 34–105.

Fama E.F., (1991). "Efficient capital markets: II", *Journal of Finance* 46(5):1575–1617

Fox, E. (2008). "The Emotional Approach", *Emotion Science.* Basingstoke: Palgrave Macmillan.

Gallagher L, Taylor M, (2002)."Permanent and temporary components of stock prices: Evidence from assessing macroeconomic stocks, *Southern Economics* J 69:245–262

Granger, C., (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". *Econometrica* 37: 424–438.

Gruhl, D, Guha, R, Kumar, R, Novak, J, & Tomkins, A. (2005), "The predictive power of online chatter." *ACM*, New York, NY, USA, pp. 78–87.

Grinter, R. E., & Eldridge, M. (2003). "Wan2tlk? Everyday text messaging". CHI 2003, 441-448.

Gilbert, E & Karahalios, K., (2009). "Widespread worry and the stock market". Department of Computer Science University of Illinois at Urbana-Champaign

Huang, Y.-P., Goh, T., & Liew, C. L. (2007). "Hunting suicide notes in web 2.0 - Preliminary findings." *In Ninth Ieee International Symposium On Multimedia - Workshops, Proceedings* (pp. 517-521). Los Alamitos: IEEE.

Huppert, F. A., & Whittington, J. E. (2003). "Evidence for the independence of positive and negative well-being: Implications for quality of life assessment". *British Journal of Health Psychology*, 8(1), 107-122.

Jarrow and Turnbull, (1995), "Pricing Derivatives on Financial Securities Subject to Credit Risk", *Journal of Finance*, Vol. 50 (1995), 53-85.

Kavussanos MG, Dockery E, (2001). "A multivariate test for stock market efficiency: The case of ASE", *Applied Financial Economics* 11(5):573–579(7)

Kirt CB, Malaikah SJ, (1992). "Efficiency and inefficiency in thinly traded stock markets: Kuwait and Saudi Arabia". *Journal of Banking and Finance* 16(1):197–210

Lo AW, MacKinlay AC, (1997). "Stock market prices do not follow random walks." Market Efficiency: Stock Market Behavior in Theory and Practice 1:363–389

Liu, Y, Huang, X, An, A, & Yu, X., (2007). "ARSA: a sentiment-aware model for predicting sales performance using blogs." ACM, New York, NY, USA, pp. 607–614.

Malkiel, B. G., (1973). "A Random Walk Down Wall Street." *W.W. Norton*, New York.

Nofsinger, J., (2005). "Social Mood and Financial Economics", *Journal of Behavior Finance*. 6, 144–160.

O'Kane D., Turnbull S., (2003). "Valuation of Credit Default Swaps", Lehman Brothers QCR Quarterly, vol. 2003-Q1/Q2

O'Kane D., Schloegl S., (2001). "Modeling Credit: Theory and Practice", *Lehman Brothers International.*

Pearson, K., (1905). "The problem of the Random Walk." *Nature*. 72, 294.

Surowiecki, J., (2004). "The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations" Little, Brown.

Schumaker, R. P & Chen, H., (2009). "Textual Analysis of Stock Market Prediction Using Breaking Financial News: The AZFinText System" ACM *Trans. Inf. Syst.* 27, 12:1–12:19.

Thelwall M., Buckley K., Paltoglou G., Cai D., (2010). "Sentiment Strength Detection in Short Informal Text" Statistical Cybermetrics Research Group, School of Computing and Information Technology, University of Wolverhampton.

Wang M., (2011). "The Less You Know, the More You Are Afraid of—A Survey on Risk Perceptions of Investment Products", The Journal of behavioral Finance, 12: 9–19.

Wilson, T., Wiebe, J., & Hoffman, P. (2009). "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis." Computational linguistics, 35(3), 399-433.

APPENDICES

Table 2 Correlations Greece

Variables	Cee_1 DM	Cee_1 DM_CH	Cee_1 DLP	Cee_1 DLN	Cee_1 DLT	Cee_1 DLS	Cee_1 DALP_CH	Cee_1 DALN_CH	Cee_1 DALT_CH	Cee_1 DALS_CH
Cee_1DM	1	.121	-.080	-.197	-.357*	-.192	.042	.085	-.184	.179
Cee_1DM_CH	.121	1	-.206	-.117	-.185	-.161	-.284*	.060	.073	-.019
Cee_1DLP	-.080	-.206	1	.371**	.571**	.611**	.657**	-.271	-.333*	.213
Cee_1DLN	-.197	-.117	.371**	1	.858**	.962**	.145	-.576**	-.252	.138
Cee_1DLT	-.357*	-.185	.571**	.858**	1	.900**	.274	-.440**	-.247	.133
Cee_1DLS	-.192	-.161	.611**	.962**	.900**	1	.317*	-.571**	-.313*	.180
Cee_1DALP_CH	.042	-.284*	.657**	.145	.274	.317*	1	-.329*	.021	-.172
Cee_1DALN_CH	.085	.060	-.271	-.576**	-.440**	-.571**	-.329*	1	-.025	.234
Cee_1DALT_CH	-.184	.073	-.333*	-.252	-.247	-.313*	.021	-.025	1	-.970**
Cee_1DALS_CH	.179	-.019	.213	.138	.133	.180	-.172	.234	-.970**	1

Cee – Greece, Spa – Spain, Ita – Italy. Last letter of the variable indicates the type of the index: N – Negative, P – Positive, S – -4/+4 Scale, T – -1/0/1 Scale.

Table 3 Correlations Spain

Variables	Spa_1 DM	Spa_1 DM_CH	Spa_1 DLP	Spa_1 DLN	Spa_1 DLT	Spa_1 DLS	Spa_1 DALP_CH	Spa_1 DALN_CH	Spa_1 DALT_CH	Spa_1 DALS_CH
Spa_1DM	1	.656**	-.020	.428**	.400**	.396**	-.111	-.169	-.317*	-.319*
Spa_1DM_CH	.656**	1	-.204	.351*	.249	.254	-.149	-.252	.030	-.077
Spa_1DLP	-.020	-.204	1	-.040	.327*	.336*	.658**	-.125	-.139	-.250
Spa_1DLN	.428**	.351*	-.040	1	.761**	.928**	.012	-.658**	-.220	-.373**
Spa_1DLT	.400**	.249	.327*	.761**	1	.840**	.281	-.558**	-.390**	-.478**
Spa_1DLS	.396**	.254	.336*	.928**	.840**	1	.257	-.667**	-.259	-.445**
Spa_1DALP_CH	-.111	-.149	.658**	.012	.281	.257	1	-.078	.079	-.030
Spa_1DALN_CH	-.169	-.252	-.125	-.658**	-.558**	-.667**	-.078	1	.166	.421**
Spa_1DALT_CH	-.317*	.030	-.139	-.220	-.390**	-.259	.079	.166	1	.791**
Spa_1DALS_CH	-.319*	-.077	-.250	-.373**	-.478**	-.445**	-.030	.421**	.791**	1

Cee – Greece, Spa – Spain, Ita – Italy. Last letter of the variable indicates the type of the index: N – Negative, P – Positive, S – -4/+4 Scale, T – -1/0/1 Scale.

Table 4 Correlations Italy

Variables	Ita_1D M	Ita_1D M_CH	Ita_1D LP	Ita_1D LN	Ita_1D LT	Ita_1D LS	Ita_1D ALP_ CH	Ita_1D ALN_ CH	Ita_1D ALT_ CH	Ita_1D ALS_ CH
Ita_1DM	1	.403**	-.421**	-.243	-.386**	-.384**	-.217	.102	.260	.241
Ita_1DM_CH	.403**	1	-.289*	-.162	-.239	-.259	-.437**	.360*	.165	.304*
Ita_1DLP	-.421**	-.289*	1	.052	.346*	.437**	.643**	-.121	-.042	-.121
Ita_1DLN	-.243	-.162	.052	1	.878**	.921**	-.182	-.670**	-.214	-.220
Ita_1DLT	-.386**	-.239	.346*	.878**	1	.926**	.076	-.547**	-.222	-.236
Ita_1DLS	-.384**	-.259	.437**	.921**	.926**	1	.087	-.651**	-.210	-.245
Ita_1DALP_CH	-.217	-.437**	.643**	-.182	.076	.087	1	-.087	.010	-.114
Ita_1DALN_CH	.102	.360*	-.121	-.670**	-.547**	-.651**	-.087	1	.228	.373**
Ita_1DALT_CH	.260	.165	-.042	-.214	-.222	-.210	.010	.228	1	.709**
Ita_1DALS_CH	.241	.304*	-.121	-.220	-.236	-.245	-.114	.373**	.709**	1

Cee – Greece, Spa – Spain, Ita – Italy. Last letter of the variable indicates the type of the index: N – Negative, P – Positive, S – -4/+4 Scale, T – -1/0/1 Scale.

Baseline Linear Regressions

Model Summary for Spain

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.578[a]	.334	.304	.0254224

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-.001	.004		-.216	.830
	Spa_CDS5_CH_Lag1	.232	.120	.244	1.937	.059
	Spa_DJ_CH_Lag1	-1.093	.297	-.463	-3.675	.001

Model Summary for Italy

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.106[a]	.011	-.033	.0319806

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-.005	.005		-.998	.324
	Ita_CDS5_CH_Lag1	.123	.214	.123	.575	.568
	Ita_DJ_CH_Lag1	.041	.337	.026	.123	.903

Model Summary for Greece

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.239[a]	.057	.016	.0507053

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	.006	.007		.867	.390
	Cee_DJ_CH_Lag1	-.112	.216	-.075	-.518	.607
	Cee_5YR_CH_Lag1	-.239	.145	-.239	-1.650	.106

Granger Causality Spain

	Data Mentions				Data Mentions Change			
Lag	F-Test	df1	df2	p-value	F-Test	df1	df2	p-value
1	4.2657	1	90	0.0418	3.7526	1	90	0.0559
2	2.2667	2	84	0.1100	2.1746	2	84	0.1200
3	1.6441	3	78	0.1860	1.8487	3	78	0.1453
4	1.8853	4	72	0.1223	1.5823	4	72	0.1883
5	2.8814	5	66	0.0206	2.9028	5	66	0.0198

	Negativity				Negativity Change			
Lag	F-Test	df1	df2	p-value	F-Test	df1	df2	p-value
1	8.3791	1	90	0.0048	1.8523	1	90	0.1769
2	4.0952	2	84	0.0201	2.1246	2	84	0.1259
3	2.9601	3	78	0.0374	2.2595	3	78	0.0881
4	2.6777	4	72	0.0384	2.1143	4	72	0.0878
5	2.3668	5	66	0.0489	2.1282	5	66	0.0729

	Positivity				Positivity Change			
Lag	F-Test	df1	df2	p-value	F-Test	df1	df2	p-value
1	0.0013	1	90	0.9710	0.2098	1	90	0.6480
2	0.0509	2	84	0.9504	0.3257	2	84	0.7229
3	0.35	3	78	0.7892	0.2019	3	78	0.8948
4	0.2605	4	72	0.9023	0.1632	4	72	0.9563
5	0.2368	5	66	0.9449	0.1702	5	66	0.9727

	Trinary				Trinary Change			
Lag	F-Test	df1	df2	p-value	F-Test	df1	df2	p-value
1	3.7973	1	90	0.0545	2.713	1	90	0.1030
2	2.9564	2	84	0.0574	1.0128	2	84	0.3676
3	1.7304	3	78	0.1676	0.5652	3	78	0.6396
4	1.3285	4	72	0.2677	1.6148	4	72	0.1799
5	1.2082	5	66	0.3151	1.2285	5	66	0.3058

	Scale				Scale Change			
Lag	F-Test	df1	df2	p-value	F-Test	df1	df2	p-value
1	7.6152	1	90	0.0070	0.1351	1	90	0.7141
2	4.0676	2	84	0.0206	1.9597	2	84	0.1473
3	2.8092	3	78	0.0449	1.3966	3	78	0.2501
4	2.2325	4	72	0.0739	2.1934	4	72	0.0782
5	1.8211	5	66	0.1208	1.8107	5	66	0.1228

Granger Causality Italy

	Data Mentions				Data Mentions Change			
Lag	**F-Test**	**df1**	**df2**	**p-value**	**F-Test**	**df1**	**df2**	**p-value**
1	0.5934	1	90	0.4431	0.772	1	90	0.3819
2	0.9206	2	84	0.4022	1.0372	2	84	0.359
3	0.9094	3	78	0.4405	0.7695	3	78	0.5146

	Negativity				Negativity Change			
Lag	**F-Test**	**df1**	**df2**	**p-value**	**F-Test**	**df1**	**df2**	**p-value**
1	5.4158	1	90	0.0222	0.0358	1	90	0.8503
2	2.3456	2	84	0.102	0.8856	2	84	0.4163
3	1.7425	3	78	0.1652	0.6463	3	78	0.5876

	Positivity				Positivity Change			
Lag	**F-Test**	**df1**	**df2**	**p-value**	**F-Test**	**df1**	**df2**	**p-value**
1	1.2698	1	90	0.2628	2.4379	1	90	0.1219
2	1.3579	2	84	0.2628	2.4503	2	84	0.09241
3	2.7131	3	78	0.05054	1.6845	3	78	0.1772
4	1.8782	4	72	0.1236				

	Trinary				Trinary Change			
Lag	**F-Test**	**df1**	**df2**	**p-value**	**F-Test**	**df1**	**df2**	**p-value**
1	2.1515	1	90	0.1459	0.0222	1	90	0.8818
2	0.7823	2	84	0.4607	0.1505	2	84	0.8605
3	0.5745	3	78	0.6335	0.584	3	78	0.6273

	Scale				Scale Change			
Lag	**F-Test**	**df1**	**df2**	**p-value**	**F-Test**	**df1**	**df2**	**p-value**
1	2.5749	1	90	0.1121	0.0996	1	90	0.7531
2	1.2028	2	84	0.3055	1.6501	2	84	0.1982
3	0.7506	3	78	0.5253	1.4132	3	78	0.2452