MASTER OF SCIENCE IN
MONETARY AND FINANCIAL ECONOMICS

MASTERS FINAL WORK
DISSERTATION

CAN GOOGLE DATA MEASURE MARKET SENTIMENT

BRUNO MIGUEL AFONSO FERREIRA

MASTER OF SCIENCE IN
MONETARY AND FINANCIAL ECONOMICS

Master Final Work

DISSERTATION

CAN GOOGLE DATA MEASURE MARKET SENTIMENT

BRUNO MIGUEL AFONSO FERREIRA

SUPERVISOR:

MARGARIDA ABREU

OCTOBER - 2016

Abstract

The purpose of this paper, on the subject of Behavioral Finance, is to use data from Google's Online Search Query, the largest search engine in the world, and its product Google Trends, to create a variable which will serve as a measurement proxy for market sentiment. The paper will focus on studying the correlation of Google measured market sentiment with the returns of the Portuguese Stock Index, PSI-20. To test this, both linear OLS and VAR regressions will be implemented, using Google data as an explanatory variable for PSI-20 returns, while at the same time using data from other control variables to filtrate the fundamental financial analysis. Additionally, the created sentiment proxy will be compared with other known sentiment proxies in terms of accuracy and promptness in explaining market behavior.

The paper concludes that Google data is indeed capable of appropriately measuring sentiment's influence on the Portuguese market, and it shows more complete results than other proxies from previous research.

Keywords: Behavioral Finance, Market Sentiment, Google Trends, VAR model, proxy variable

Resumo

O propósito deste artigo, na temático de Finanças Comportamentais, é usar os dados da pesquisa on-line do Google, o maior motor de busca do mundo, e seu produto Google Trends, para criar uma variável que servirá como uma proxy do sentimento no mercado. O artigo irá concentrar-se em estudar a correlação do sentimento do mercado medido pelos dados providenciados pelo Google, com os retornos do Índice da Bolsa Portuguesa, o PSI-20. Para realizar este teste, irão ser aplicadas ambas regressões lineares de OLS e modelos VAR, usando dados do Google como uma variável independente dos retornos do PSI-20, enquanto que ao mesmo tempo, serão usados dados de outras variáveis de controlo para filtrar a análise financeira fundamental. Além disso, a proxy de sentimento criada será comparada com outras previamente utilizadas, no que toca a precisão, prontidão, e capacidade para explicar o comportamento do mercado em geral.

O documento conclui que os dados do Google são realmente capazes de medir adequadamente a influência do sentimento no mercado Português, e mostra resultados mais completos do que outras proxies previamente utilizadas noutros trabalhos.

Agradecimentos

Em primeiro lugar gostaria de agradecer à Professora Margarida Abreu. Pela sua preciosa ajuda, pelo seu constante apoio, orientação, e acima de tudo, pela sua inacabável paciência, sem os quais não concluiria este trabalho.

Gostaria também de agradecer ao Professor António Afonso por me ter proporcionado a oportunidade de participar neste programa de mestrado e pelos conselhos sempre pertinentes.

À minha família, e em particular aos meus avós, pelo seu contínuo apoio, ensinamentos e encorajamento. Agradeço-lhes a sua persistência e preocupação em todos os aspectos da minha vida, mesmo quando eu não notava. Sem eles não seria a pessoa que sou hoje.

A todos os professores e colegas de faculdade, sem os quais não teria chegado ao fim deste percurso. Sinto-me honestamente grato por ter realizado este caminho com vocês, ser influenciado por vocês e ter criado tantas memorias e laços que me seguirão na vida.

Um agradecimento especial aos meus amigos, que tanto fizeram por me apoiar em todos os momentos do último ano, nos bons e particularmente nos maus. Aos meus amigos do BNP Paribas pela sua companhia e por todas as horas de divertimento na aventura que foi o último ano. Um agradecimento em particular, ao Miguel, ao Gonçalo, ao Marco e ao Luís pela sua verdadeira amizade e a este último também por conselhos dados tendo em vista a finalização deste trabalho.

À Débora, por tudo o que me ensinou e de muitas formas, continua a ensinar.

A todos, o meu mais sincero *Obrigado*.

Table of Contents

List of Tables

List of Figures

Glossary

AbnTurnover – Abnormal Turnover;

AGMS – Sentiment measure by Google using search query with the aggregation of the most explanatory search terms related with PSI-20 returns;

OLS – Ordinary Least Squares Regression;

VAR – Vector Auto Regressive Regression;

Bubble – Trading value of assets strongly deviates from the fundamental value;

ESIER - Confidence indicators focusing on measurement of the European area;

ESIEU - Confidence indicators focusing on measurement of the Euro Zone area;

ESIPT - Confidence indicators focusing on measurement of Portugal;

FEARS – Sentiment measurement proxy based on data from online search query;

GDP - Quarterly data of GDP's homologous growth rate;

GMS – Google Market Sentiment – Proxy created with search query volume retrieved from Google Trends;

Google Trends – Tool of measurement for query search volume;

GSVI - Google Search Volume Index;

ICE - Indicador de clima Económico - Monthly time series regarding Portuguese investor confidence;

PSI20Ret – Returns of the PSI-20 Index;

SVI – Search Volume Index;

SVIW - Search Volume Index after the winsorization process;

SVIWB - Search Volume Index after winsorization and treatments for heterokedasticity;

VIX – Volatility Index of the American financial market – “Fear Gauge”;

VSTOXX - Volatility Index of the European financial market – “Fear Gauge”;

1 Introduction

This paper proposes the use of Google Trend's data as a means to boost previous works on the topic of market sentiment, creating a proxy variable composed by Google Trends' data and evaluate its ability to measure market sentiment in the Portuguese stock market. The proxy is then compared with other acknowledged proxies for market sentiment with the purpose of determining which one portrays more accurately, thoroughly and timely the aggregate investor behavior.

Market Sentiment, "Animal Spirits", "Irrational Exuberance", whatever the reader prefers to call it, has a considerable impact on the market and the Economy in general. The psychology of decision making can no longer be disregarded as a crucial feature in Economics and Finance.

We have witnessed bubbles and market crashes, from the asset prices and real-estates in Japan during the 80's, passing through the dotcom bubble of the late 90's and the real-estate bubbles all around the world during the 2000's. Speculative frenzies date back to the "Tulip Mania" in the Netherlands in the 17th century and still occur today. None of these events are fully explained by modern economic theory or financial notions. The idea that investor's sentiment, other than financial reasoning, can actually drive the market has faced many objections and critics, but also many supporters from as early as Keynes to more recent economists like Shiller, being a latent concept on the back of the minds of many scientists.

It can be seen as a reflection of information asymmetry, moral hazard and/or adverse selection, as a consequence of uncertainty in the financial markets, or simply as one of the costs of poor risk-taking attitudes by a large amount of investors who end up basing their action on incorrect sources. Whatever the way you look at it, the conclusion is always the same; the markets are affected by their participants, and systematic, financially unfounded behaviors, can have a large impact on the market.

Today, the difficulty lies mostly on the accurate method of measuring the effect of market sentiment, of pinpointing the extent of irrational investor's impact on price changes. It is on this aspect that search query data can provide valuable support. Nowadays, where all kinds of information are easily obtained through the internet, an indicator for the amount of search queries including specific words can reveal a great deal about consumer and investor behavior. This type of data can serve as a precise insight into people's motivations, interests, and desires, and help translate agent's true perceptions of the world.
Being the largest and most outright used search engine in use today, data regarding Google search query volume has the potential of being a strong indicator of peoples' preferences, through means of their online search habits.

With online search query data, it is possible to create a variable which adequately portrays agents' expectations and sentiment towards the market in a more transparent and timely fashion. To test this hypothesis an OLS regression is applied, focusing on the PSI-20 returns as the dependent variable. To precisely measure Market Sentiment's effects, other economic sources of information passing for control variables are introduced in the regression. At the same time, this new proxy is compared with other know proxies of sentiment measurement. From this, it will be possible to conclude that Google Search Volume indeed serves as a proxy for market sentiment, having recognizable explanatory power over the behavior of marker returns. To test for the predictability power of the created variable, the methodologic approach includes a series of VAR models. These serve as support for the conclusions present by Da et al (2014), congruent with previous research like De Long et al (1990), that sentiment's effect creates a contemporaneous push in prices and returns, which then suffer a decrease in the following periods (a reversal) has sentiment fades and fundamental values' are restored to their dignified importance. Also, has an extra robustness test towards the created proxy, a test concerning an alternative process of Google data treatment is conducted, and by means of direct comparison it is possible to conclude that the methodology implemented by Da et al (2014) provides better results.

The paper is structured as follows: firstly, the topic is generally introduced introduce providing important concepts and definitions; next, the data to be used is presented and the necessary calculations and data treatments are explained; the ensuing section explains the methodology approach used in to address the core questions of the paper; and the final two sections present results and form conclusion remarks.

2 Market Sentiment

2.1 Fundamentals of Market Sentiment

Going back to one of the fathers of modern economics, John Keynes (1936) advocated for the existence of what he would name "animal spirits", aggregate investor behaviors that affect the market, driving prices away from their fundamental values. Ever since, we have witnessed stock market crashes, bubbles, and other surprising and otherwise unexplained market behaviors by mainstream economic theory. It reached the point where economic thought begins to adapt, as it becomes logical to generate new economic and financial theories, or expand the existing ones, with imported concepts which were previously only studied by disciplines like psychology and sociology.

On this regard, Behavioral Economics/Finance has made some significant advances towards a better, fuller comprehension of the human behavior. These breakthroughs can be considered on an individual investor behavior perspective, dealing with biases and the role of decision making, or through investors' aggregate influence projected onto the market through less-than-optimal decision making. With the arising of new ways of interpreting uncertainty, the addition of auxiliary concepts like market sentiment and the development of attention theories, the formulated hypotheses to describe the decision making process proposed in the existing literature are increasingly more realistic and fit better with investors' actual actions in the market.

2.1.1 ***Uncertainty***

Uncertainty is a key issue in economic research. Agent's decision making is driven by the degree of their uncertainty regarding future events, their possible payoffs and correspondent likelihoods, which are reflected by agent's knowledge span in the present, ability to process information and availability of said information. Furthermore, the choices made are dependent on people's relation with risk, and, maybe just has importantly, how people *choose* to interpret said risk.

The Expected Utility Hypothesis is Economic theory's generally acknowledged approach to deal with uncertainty. Its conclusions are based in the computation of the expected value of the possible outcomes taking into consideration their respective probabilities and payoffs. Although the theory conveys the perception of the rational choice, the sufficient conditions for its application may not always hold for decision makers in real-life situations. On this line of thought, the existence of uncertainty and risk, imposing themselves into the market through asymmetry of information among the participants, can lead to cases of market inefficiency from which known problems such adverse selection or moral hazard transpire.

2.1.2 Heuristics, Biases and Prospect Theory

Behavioral Economics addresses the problem of uncertainty. Tversky and Kahneman (1974) and Kahneman (2011) contribution on the topic arose with the conception of the Prospect Theory, a Behavioral Economics challenger for the traditional von Neumann-Morgenstern utility theory. This theory's assumptions are based on the existence of biases, to which the human mind is prone to undergo, and heuristics or rules of thumb that the brain uses to facilitate decision making. At the same time, Prospect Theory incorporates innovative concepts from the field of psychology into economic theory, by considering relative gains or losses rather than final payoffs, and individual probability weightings. It challenges the basic Economic assumption that all agents are rational and always chose the best possible options. However, sometimes this comes at the cost of reaching rushed resolutions for problems, meaning that people, with haste of reaching a solution for a more complex problem tend to make interpretation or calculation based mistakes that ultimately lead to poor decision making. When applied to the fields of Finance and Economics these decisions (provided by less-than-optimal interpretations of new pieces of information for example) can mean that the subsequent aggregate actions of investors can affect the market in ways that do not reflect fundamental information. We can ascertain that when the effect of individual biases, prone to affect agents, is generalized to a sufficient number of (irrational) investors, prices and stock returns can show some unforeseen movements.

For the purposes of this paper, Behavioral biases can be categorized into 3 main classes: *Perception and information processing biases* such as availability bias, *Framing biases* such as accessibility or anchoring, and *Representativeness*.

In most economic textbooks it is stressed that one of the fundamental frameworks of economic thought is that of agents being rational. However, in practice, because of constraints related with time, memory and capacity to process all available information, that is not usually the case. **Perception and Processing Biases** focus on those limits, and how the investor is subjective to associations and unconscious preferences that dictate their financial actions.

Tversky and Kahneman (1974) had already stressed the importance of **Availability** of information, and the existence of a bias associated with the process of relying on information which is readily available, bypassing further research for newer and better data to justify conclusions. Applied to financial markets, this particular form of heuristics can take the form of investors focusing on opposing opinions regarding the future performance of a specific stock, which halts them from examining the real value of the stock, or alternately by focusing on a particular piece of news instead of the overall information regarding the firm and industry.

Framing and Mental Accounting is another major behavioral bias category and it states that, unlike what is defended by the rational theory of choice, the formulation in which problems and situations are presented does have an impact in the way agents address them, and can go as far as altering usual actions and preferences.

A prime example of this type of bias is what Behavioral Economists name as **Accessibility**, and its importance has been highlighted by authors like Kahneman, (1974). It consists in the subjective interpretation of the facts at hand, which differs from person to person, and as such serves to justify the different, sometimes broadly opposite actions played by market participants. Put in another way, when an agent processes information and makes a choice, context matters, as does the mental state or the ease with which a particular idea or feeling comes to mind. **Anchoring** is also a common bias, conveying people's tendency to rely too heavily on a set piece of information, named the "anchor". This serves as the basis of comparison, basically functioning like a subconsciously implied benchmark which instigates the ceasing of looking for further information and extrapolation of conclusions based on a piece of info which might even turn out to be irrelevant.

The third category, closely linked with uncertainty, **Representativeness Heuristics** consists in the way the brain can choose to mentally adapt the conditions of a difficult problem to solve, transforming it into a simpler problem with an easier solution. Many times this means extrapolating past results or incurring in stereotypical notions and errors.

Emotions, moods, reaction to news and other major events or even trivial everyday events, can direct investors' actions and attitudes toward risk, prompting them to make excessively risky moves, or in other words, directing them for situations where the possible payoffs do not fully incorporate the compensation necessary for the amount of risk taken. It is common for individual investors to manage their investment portfolio based on companies they like, or the ones they have good experiences with, regardless of their actual historical or current performance in the markets. For example, Kahneman (2011) writes a passage where he tells the story of an acquaintance that invested in Ford, merely because he liked the firm and was happy with its products, not basing his decision on any particular source of information or belief that the stock was undervalued at the time. (Abreu, Mendes (2012) also deals with this subject).

What mattered to his decision was merely the way he *felt* about the firm, about the confidence he was willing to entrust in it.

2.1.3 Collective Sentiment

All the effects indicated above are biases and heuristics applying to investor's on an individual level. However, major divergences with fundamental financial reasoning occur due to the aggregate behavior of individuals, and while it seems unreasonable to believe that a majority of investors might all be induced in error, the truth is that many times unsophisticated investors seem to incur in what is designated as Herd behavior (Chang et al (2000)). This collective phenomena consists in people's following large groups, not taking into account the validity of the decision being made, based on the fallacy that large groups of investors cannot be wrong. This behavior is most notably displayed during the occurrence of bubbles.

A sense of collective thinking can drive the market and as such, factors which affect the large collectives of investors will indirectly impact market forces.

Kamstra et al (2003) link sentiment with seasonal affected disorder. Focusing on previous research linking periods of depression to periods of low daylight, the authors' research finds positive correlations between periods of higher risk aversion and a seasonal variation of equity gains, verifying the idea that amount of day time can be linked with returns and portray agents overall sentiment and mood. Edmans et al (2007) link mood with the stock market through soccer results. They find positive correlations between national teams' soccer results (particularly negative result) and market returns for that country, resulting in a collective mood which is projected onto investors and affect financial actions. During this rush to catch the trend, it can be very costly to be left out, and that thought is also a driving factor for the promotion of herd behavior, and creation of financial bubbles.

2.2 Investor Sentiment

Sentiment's distinctive effect, and its toll on the overall market performance, has been the basis for a large amount of academic research. Authors develop distinct perspectives on these inaccuracies of human aptitude and undertake different approaches to capture these effects' magnitude, hence the multiple possible variables that can be used in order to quantify sentiment's effect in the market. In its core, market sentiment can be perceived as the aggregate sense of optimist or pessimism towards future financial performances, which translates in the amount of risk investors are willing to take (Baker and Wurgler (2007)). If investors are driven by an excessive wave of optimism, market deals should materialize an increase in prices above their fundamental value, as the peoples' shared euphoria induced behavior displays a sort of bandwagon effect. This way, phenomena like heuristics and biases, emotions and mood, reaction to news and individual interpretations, all caused by the effects of uncertainty and asymmetry of information existing in the market, can contribute to the formulation of a distorted image of the financial reality. This effect of investors' overall sentiment over the market can generate robust deviations from fundamental prices contributing to a moreover menacing implication of the existence of a "crowd mentality", which helps explain periods of unpredictable, random-walk like market performances.

These effects are particularly prominent on individual agents without access to major sources of information, the "unsophisticated investors" or like it is usually named in the literature, irrational investors. De Long et al (1990) explain the incidence of sentiment in agents by categorizing investors into two types, rational investors who act in the market based on fundamental information, and irrational investors, noise traders, whose actions can be affected by their emotions towards a certain stock, a firsthand piece of news, or a particular event of interest. These individual investors tend not to follow the traditional sense of economic and financial procedures in the evaluation of assets, they fail to diversify and hedge investments, and they conduct their own independent research on the market.

Shiller and Akerlof (2009) go as far as discussing several behavioral phenomena, or the so called “animal spirits”, like the effects of confidence and the creation of stories, both of which affect the economy and stock markets. Given their influence on expectations of future payoffs and investment appraisal, their importance on asset price movements should not be disregarded.

Barberis et al (1998) dealt with investor sentiment in market interactions and asset price formation through means of an investor sentiment model based on the interaction of biases such as over and under reaction to news and announcements. Their interpretation of sentiment is linked with the occurrence of news and its reception by investors a having a measurable effect in stock price movements. They provided guidance for future research alluring ideas of price shifts due to overreactions to news, and the behavior subsequent to stock market crashes.

The mentioned research comes to show the multifaceted range of effects that investor sentiment can have on the markets, and the importance of figuring out how to deal with such phenomena, capable of causing massive financial turmoil.

2.3 Measures of Sentiment

While the previous works deal with the theoretical effects and implications of the market's behavioral spectrum, others focus on measuring its effects in a more practical way. As Baker and Wurgler (2007) put it "Now, the question is no longer, as it was a few decades ago, whether investor sentiment affects stock prices, but rather how to measure investor sentiment and quantify its effects". In the referred paper, the authors themselves construct a composite sentiment index by aggregating several variables which were considered as indirect proxies for sentiment measurement in previous researches (trading volume, dividend premium, first-day returns on IPO's, amongst others). Their composite variable reflects the observed historic events like bubbles and other crashes on the stock market, succeeding in capturing sentiment. Robert Shiller (2010) on another hand looks at surveys with hopes they can provide some insight towards agents' sentiment flows around known periods of overconfidence and surge of financial bubbles.

Also resorting to surveys to face the problem of sentiment measurement, Brown and Cliff (2004, 2005) decided to use a time series based on newsletter inquiries with the purpose of comparing survey data with other known proxies for market sentiment, reaching the conclusion that they indeed serve as a viable option. The conducted tests, including VAR tests on a composite variable that measures market sentiment, provides them enough confidence to state that their method of measuring sentiment is able to capture asset price deviations from their fundamental values as well as the correspondent future return reversals expected from the guidelines of market sentiment models.

2.4 ***Google Data***

2.4.1 ***Google data and Attention***

A large set of researches focuses on another topic of Behavioral Finance, relating Google data with **attention,** and proposing a distinct explanation for investor's judgement regarding stock acquisition.

Google's data brought forth an innovative tool for researches, assembling support for attention grabbing theories like that fashioned by Barber and Odean (2008).This theory of attention, in the epicenter of this type of research, states that individual investors are net buying of attention-grabbing stocks. A typical investor can choose between thousands of stocks, so it is unreasonable to believe "irrational investors" maximize gains while considering all available options of investment. This happens because investors are faced with limitations of time and on their capacity of information processing. This last is a fundamental foundation of the field of Behavioral Economics, denominated Bounded Rationality, Kahneman (2003). There is simply too much to compute and too many investment options for a single person to make the very best decision every single time, even considering theoretical economic approaches where agents are unbiased. A more realistic hypothesis is that, if we consider that short selling its not common practice for individual investors, meaning they can only sell what they have, investors will trade more those stocks that are "newsworthy", that is, the stocks that get their attention somehow. Like previously stated, this is greatly interpolated with the Availability bias.

In Barber and Odean (2008), the authors test the possibility that surges of attention could promote a contemporaneous stock prices' rush, which would then be followed by a reversal in returns a few weeks after the fact. Their results indeed appear to support their proposed theory and make explicit that individual investors are affected by behavioral heuristics.

Based on this initial ground setting work, a considerable amount of researchers have directed their research towards the idea that the way individuals pay attention to firms and stocks, better explain unsophisticated investor's attitudes on the stock market. As such, several proxies for attention, like the occurrence of important pieces of news concerning the firms, volatility, abnormal returns and liquidity, have been tested in the studies conducted to attest the value of this theory.

Significant for our case, Da et al (2011) make good use of Google queries data and find correlations between it and other proxies for attention such as turnover, extreme returns and news. As an added plus, they find that this sort of data can capture investor attention (linking to Barber and Odean's theory) in a more timely fashion than other proxies, which take longer to exhibit how changes in attention affect the markets.

Bank et al (2011) focus on the case of German stocks, but arrive to similar conclusions regarding the capability of Google search volume to measure attention since their results show that increases in trading activity are related with investor recognition. They, in fact, suggest a positive correlation with stock liquidity, which they attribute to a reduction in asymmetric information.

On a less specific study, Latoeiro et al (2013) focus on stocks comprising the EURO STOXX, and find that Google queries' behavior precedes changes in trade volume and volatility, and reversals occur in the following weeks.

Aouadi et al (2013) focus on the French case, using Google data to research attention implications on stock market activity and volatility. Their findings are in line with other research, mainly, a positive correlation between liquidity and attention grabbing stocks. Their results are likewise robust to different ways of calculating volatility, and even the recent financial crisis, to which the effect of attention on stock performances and volatility does not subdue.

Although without mentioning Barber and Odean's theory, but still with the purpose of finding new information to support attention theories, Mondria et al (2010) sought to find how attention affects portfolio choices, creating another measure for attention based in data from Yahoo (and not Google, although Google greatly dominates the web search market), and study the determinants for attention allocation for US investors. Their key conclusion has implications on both behavioral economics and decision making, since they find two-way causality between attention allocation and home country bias.

2.4.2 ***Google data and Market Sentiment***

The idea of using data from Google Trends as an explanatory variable for financial phenomena has been gaining some track in the past years. Choi and Varian (2012) have used Google Trends' data in order to forecast near contemporaneous values for macroeconomic indicators such as unemployment, consumer confidence, and consumer patterns, particularly concerning automobile sales and travel destination planning.

For example, if a person is interested in purchasing a new car, the best way to gather information about possible models and brands and compare vehicles is to conduct online research. One individual will eventually find one particular car to his liking and might intensify the search for more detailed characteristics or financing methods. The aggregate search habits of the population can then mirror sales for that particular model. The same applies for the stock market.

They go as far as to state that Google Trends' data can be used to help predict incidence of some diseases like influenza and flu, mapping its progress and effect, based on the amount of queries of search terms related with those diseases, like symptoms and methods of treatment. After this work, academics understood the usefulness of Google data in "predicting the present" and favored its ability in explaining the behavior of consumers and investors, resulting in an observable increase of research regarding these topics in recent periods. Irrational investors, like previously defined, have access to considerably less materials than what are designated the "sophisticated" or "traditional" investors, and as such they resort to sources of information which might not be the most appropriate, might not be relevant or even contain incorrect or incomplete data. This leads to different expectation over stock performance and increases divergence on the overall attitude towards the market. Given the wonders made possible by modern technology and its worldwide generalization, it is reasonable to consider that investors acquire large chunks of information regarding the state of the economy and financial markets over the internet. This paper's objective is to focus on the particular activity of these agents and seize that information to get a quantifiable understanding on how investors perceive the market.

In the last few years, particularly after data from internet search queries was made available in a readily and outright fashion, a new topic of research began to sprang, with a focus on using web search queries as a quantifiable measure of otherwise unquantifiable occurrences. Subsequently, many authors have drifted towards this nature of data in hopes of capturing behavioral phenomena. The proposed premise of this paper is that web searches of particular search terms convey preferences and attitudes of irrational investors which use said method for financial information gathering. We test the implication that it is a more transparent manner and with faster contemporaneous adjustment in relation to other known measures of sentiment measurement.

Joseph, et al (2011) check online ticker data for its ability to forecast stock returns, trading volume, and volatility. They find reasons to believe that web data can be used as a forecasting tool, particularly for stocks that are considered harder to arbitrage, which are more striking to be affected by sentiment shocks, like some previous research has shown (Baker and Wurgler (2007)).

This paper will closely follow the methodology of Da et al (2014), in which the authors create an index with the purpose of measuring investor Sentiment. Their creation, the FEARS index, is based on data from online search query volume retrieved from the major search engine, Google, by mean of their product, Google Trends. The authors are able to show that this indicator of negative sentiment predicts aggregate market returns, particularly for those stocks considered as favored by unsophisticated investors (the effect is stronger for equities and small stocks, instead of treasury security returns). It shows an inverse correlation with contemporaneous returns, but also an increase in FEARS shows an increase in returns in the ensuing days, hence hinting the existence of reversal patterns, which is consistent with sentiment theory. Regarding volatility (one of the measures used is VIX), the results show a positive contemporaneous correlation between FEARS and VIX, again followed by a reversal in the succeeding days, supporting previous research stating that noise trading has a temporary effect on price volatility. Lastly, because mutual fund holders are usually individual investors, sentiment (negative sentiment in particular) is more likely to influence the behavior of these types of assets. The authors also find that FEARS predicts mutual fund flows out of equity funds towards bond funds, meaning that a spike in FEARS is followed by that particular shift, which embodies a flight to safety.

3 Data

3.1 Google Market Sentiment – GMS

Google Trends allows everyone access to data concerning internet queries through their website. An interested party can specify the search term of choice and constrain the search in terms of geographic location where the queries are posted, as well as time period limits. Figure 1 illustrates data retrieved from Google Trends concerning the amount of searches made for the term “crise”, the Portuguese word for crisis, in Portugal during the period of 2004-2015.

[Insert Figure I]

Notably, the massive increase in the amount of searches in August 2008 allows a glimpse into peoples’ sudden awareness regarding the county’s economic reality.

It is necessary to point out that the data Google provides is not an indicator of absolute search volume, but rather it is measured relatively to the highest overall amount of searches, ie, the time series is scaled by the time series' maximum. This means that the data provided is normalized, being presented in reference to the maximum number of hits the search term achieved during the specified period of time. This way, the values returned by Google Trends are ranged from 0 to 100, and if, for example for a particular search term, on a particular time, the observation is 60, that value is relative to the maximum (100) number of search queries for that term. The highest point in the graph will always have a value of 100, translating the period where the amount of searches was at its zenith. The value for a particular geographic location where the query is placed is also in relation with the overall number of searches in the region it is inserted. Google Trends actually gives us the likelihood of that term being searched in that location, relative to close countries, as to not beneficiate more populated countries with an immensely greater number of "hits".

Previous authors like Da, et al (2014), remark the definition of search terms as the first crucial step on this line of research, engraving the importance of using accurate, appropriate terms that are related with the research and leave little to no room for generic interpretations of what the person searched when he made that web-query. For example, ambiguous search terms like "receitas" which might mean financial revenues or meal recipes, should fall in the category of terms ignored by this approach. In order to broad the range of the research, a series of search terms that would encompass several aspects of economics and financial fields were compiled. Part of this research was based on the search terms covered on the word trees created by Ana Oliveira-Brochado (2014). Using the Harvard IV dictionary the research range was broadened, including additional search terms considered to be of the economic and finance jargon. This dictionary was previously used by Da, et al (2014) and Tetlock (2007) and comes with the aggregate advantage of allowing classification for each term into words of positive and negative connotation. Some other search terms were the result of general surveys, conducted with the purpose of finding which terms agents (and individual low frequency investors) tend to search for. Unlike previous works, this research focuses on search terms in the Portuguese language. The reason for this is that it is believed to better reflect the research methods of individual investors, that is, it would be expected that individual, non-sophisticated investors conduct searches in their official language, other than English. This could be interpreted as a form of home country bias, as an investor preference for his own national market is reflected by his preference in conducting relevant research using his native language instead of using a more global language where he might yield better and more diversified information. After searching for about three hundred search terms, data was retrieved after having identified the ones without major timeline breaks, as many of the chosen search terms did not present a sufficient number of queries as calculated by Google's method of scalar multiplication, meaning that people would not search that term enough for it to be relevant for the time series to be retrievable. Furthermore, although Google's data goes as far back as 2004, another large concern faced was that some terms did not possess data ranging that far, showing several null results for a large amount of terms for long periods of time. Because of this the timeline was shortened from 2007 until 2015, in order to keep the possibility of evaluating possible effects of the financial crisis. These major challenges reduced the sample significantly, to around 180 search terms.

Having downloaded time series' for each of the surviving terms, the variable was designated has Google Search Volume Index, or GSVI for short. Some series however were only treated on a monthly basis, contrasting with others, in weekly basis. This is probably due to the fact that through a particular period of time, some specific terms do not have enough "hits" to be considered significant after the scalar multiplication, and as such, their frequency turns out to be on a monthly basis. To advance the research the terms possessing a weekly basis were selected. Although the number of usable terms is reduced significantly, the final results prove to be more consistent and reliable.

Further, the SVI (Search Volume Index) was computed for each time series, ie, the natural logarithm of the GSV (Google Search Volume Index) series.

$$SVI_{j,t} = ln\left(GSVI_{j,t}\right)$$

Because some of these search terms also have some brief periods of GSV equaling zero, in order to deal with the logarithm, the choice fell on the replacement of all zero observations with 0.1 in order to reduce order alteration after applying the logs. Next the data is transformed further in order to take into consideration weekly variations of the variable instead of its level.

$$\Delta SVI_{j,t} = SVI_{j,t} - SVI_{j,t-1}$$

Afterwards great care was taken to identify possible outliers at the 10% level (5% each tail) allowing the undergoing of a process of winsorization for each search term, ie, replacing values above the 95th percentile and bellow the 5th percentile for the values correspondent to said percentiles. This way outliers and extreme values are eliminated, removing major problematic aspects of Google Trends' computation of data at the time of download. It is a necessary process, considering the nature of GSVI's computation by Google.

Next, an Analysis of Variance tests (ANOVA tests) are performed. The purpose of these tests is to check if there exists a statistical difference between the means of the variables. The null hypothesis is that the monthly means are equal. If the null hypothesis is rejected, meaning that there is significant change in data depending on the month, the data must be deseasonalized by regressing the time series of the variable ΔSVIW on monthly dummies, and keeping the residuals for further use. This process was applied to both the GSVI variable, the "raw" data, and to the ΔSVIW (ΔSVI after winsorization) and the results show that both transformations lead to different and opposite conclusions. The best results seemed to derive from applying the methodology to the raw data. For example, the search term "irs" should, in theory, present some seasonality, as more people search that term around the due date for dealing with taxes. The ANOVA tests on GSVI are in concordance with this idea, as for other search terms, so analysis was conducted considering the results for this variable. Finally, in order to deal with heterokedasticity, for each search term, a scalar multiplication of its standard deviation was applied to ΔSVIW, creating the variable ΔSVIWB. This concludes the treatment of Google Trends' raw data, which is now ready for further use.

The ensuing step thins the range of search terms even further by identifying those search terms which are better correlated with market returns. To examine this, backward-looking rolling regressions were conducted, with the purpose of identifying those terms possessing better historical relationships with contemporaneous market return. The regressions have a rolling window of 55 and a step size of 26 weeks. From the 180 terms that were tested, 25 proved to be statistically significant at least at the 90% level, showing t-stats around 1.60. From those, the 10 terms with best significance either with positive and negative correlation with market returns where chosen to create the variables GMSpos and GMSneg.[1] The images pictured in figures 2 and 3 are somewhat similar to one another and depict rather erratic behaviors. In spite this, particularly for the case of GMSneg, the illustrations does seem consistent with some periods in which worse market sentiment would theoretically be expected, showing a considerable increase in negative sentiment 2008 onward.

[Insert Figure III & **Figure IV**]

3.2 Google Market Sentiment – AGMS

Additionally, an alternate method of testing the robustness of the GMS variables consists in using a different technique to form them. Google Trends allows users to search words in an aggregate method, combining different terms in the same search query. So by typing the search terms which better correlate with returns onto Google Trends in such a fashion (as identified before), the Google Market Sentiment variables can be constructed anew as AGMS (Aggregate Google Market Sentiment).

[1] * GMSpos words: “Saldo”; “Procura emprego”; “Inovação”; “Abono”; “Investimento”; “Trabalhadores”; “Risco”; “Desenvolvimento”; “Oferta de Emprego”; “Benefícios”.
GMSneg: “Tarifa”; “Pensao”; “Preço ouro”; “Emprestimo”; “Dinheiro”; “Ordenado”; Comercio”; “Carreiras”; “Ordenado mínimo”; “Revenda”.

[Insert Figure II]

After the data referring to this new sentiment proxy's time series is retrieved, the same data transformations we applied to each individual search term are performed, that is, logarithmization, first differences, winsorization and heterokedasticity testing.

3.3 Data Treatment

In the following sections, the created proxy for market sentiment will be confronted with a series of tests in order to assess its efficiency in expressing the effects of sentiment, and robustness checks through comparisons with other proxies for market sentiment. To test Google's ability to quantify sentiments' influence on the financial market, further data was gathered, ranging from period 2007 until 2015. PSI-20 Index's weekly opening values from Bolsa de Lisboa were retrieved and went under the following procedure of return's calculation:

$$PSI20ret = \ln(P_t) - \ln(P_{t-1})$$

In order to accurately assess the worth of this new proxy, it is necessary to introduce some control variables into our model. As such two types of control variables were added to this analysis:

- Economic control variables: GDP, Inflation, Unemployment – Serving has a way of incorporating fundamental macroeconomic information into market returns;
    - *GDP* - Quarterly data of GDP's homologous growth rate gathered from INE
    - *Inflation* – Monthly data of homologous variation of the IHPC (índice harmonizado de preços no consumidor) gathered from EUROSTAT.

- *Unemployment* – Homologous rate on a monthly basis got from INE, calculated has a percentage of total population.

- Sentiment Measurement proxies: To be used as benchmarks, providing a term of comparison in order to provide insight in how well web search queries' data perform relative to other sentiment proxies.

  - *Economic Sentiment Indicator–ESI* – This variable is a composite of five sectoral confidence indicators with different weights. Based on surveys and calculated has an index by EUROSTAT, the gathered data serves as a sentiment proxy for the areas:
    - Europe - *ESIER*
    - Euro Zone - *ESIEU*
    - Portugal - *ESIPT*
  - *Indicador de clima Económico – ICE* - Monthly time series produced by INE based on surveys and regarding Portuguese investor confidence.
  - VSTOXX – Index designed to reflect expectations about market volatility, serves as a "Fear Gauge". The time-series is retrieved in a daily basis, but the data used is the one respecting the first day of the week.
  - Turnover – Turnover on PSI-20 Is acquired from Bolsa de Lisboa, as it was done for returns.
  - Abnormal Turnover – From the previous data, AbnTuronver is computed as:

$$AbnTurnover = \frac{\ln(Turnover_t)}{\sum_1^{52} \ln(Turnover_{t-1})}$$

It is noteworthy to point out that some of these variables were only available on a monthly or even quarterly basis, causing obvious concern since the study's main variables are created with a weekly frequency. Because there appears to be no ideal way of dealing with this issue, the decision tilted towards keeping the corresponding monthly values throughout the periods in question, that is between January 2007 and September 2015.

Table I – Control Variables

Variable	Source	Type
GDP	INE – Instituto Nacional de Estatística	Growth rate, Moving average **[2]
Inflation	INE – Instituto Nacional de Estatística	Harmonized Consumer Prices Index, Rate calculated has a Moving average *
Unemployment	INE – Instituto Nacional de Estatística	Moving average rate*
ESI - Economic Sentiment Index (Europe - ESIER, Euro Zone - ESIEU, and Portugal - ESIPT)	European Commission	The Economic sentiment indicator is a composite measure (average = 100) regarding surveys applied *
ICE - Índice Clima Económico	INE – Instituto Nacional de Estatística	Index based on surveys, calibrated as reference to GDP *
VSTOXX	STOXX	The VSTOXX® volatility index expresses the fluctuation range expected by the market, which is the implied volatility of the EURO STOXX 50® Index.
PSI 20 Index values	Bolsa de Lisboa	Opening prices
Turnover	Bolsa de Lisboa	Turnover is expected to rise as people become overoptimistic.
Abnormal Turnover	Bolsa de Lisboa	-

[2] *Monthly Data
**Quarterly Data

3.4 ***Correlation Matrix***

To develop an overall picture regarding how the sentiment measurement proxies perform in relation to each other, a correlation matrix was performed. The highest, statistically significant correlations are, as it was to be expected, among the different sentiment measuring proxies for Portugal, and the surrounding areas, ESIER, ESIEU, ESIPT, ICE, as they are computed in similar fashion (mostly resorting to surveys) and some from the same institutions. Because of the negative effects of using such correlated proxies as explanatory variables in the same regression, it was decided to relinquish the usage of some of these from the ensuing econometric models. The variables belonging to this array chosen to be inserted in the models will be ESIEU and ICE for their relevance and significance. VSTOXX also shows significant negative correlations, reaching tolerably high values with some sentiment proxies. This was to be expected, as an increase in Euro STOXX 50 Index's volatility should be interpreted as an increase in pessimist and negative sentiment towards the market.

Regarding the created proxies, the results only appear to be significant amongst each other, presenting a small and, most puzzling, positive correlation, as illustrated by the figures 3 and 4 where GMSpos mimics GMSneg behavior through some periods. With the exception of some periods where the behavior seems to go on the same direction, the GMS variables do not seem to fully replicate the behavior of the remaining known sentiment proxies. Possible explanations can be that, either Google data does not fully encompass the effects of sentiment in the market or, given this unlikely supposition given the solid theoretical reasoning behind the past application of this data in the explanation of sentiment, that it does a more accurate job in doing so, particularly on a more timely fashion, when taking in consideration the data frequency of the other proxies in comparison.

[Insert Table II]

3.5 Hypothesis Testing

The focus of this next section is to test the fundamental hypothesis of this paper - Does this created variable succeeds in capturing the sentiment impact on Index returns, and if so, how does it compare with other known proxies? The methodology proposed to deal with this issue consists in three layers of tests. Firstly we check basic correlations between sentiment proxies and apply causality tests. Main expectations involve the viewing of significant correlations between the behavior of GMS variables and other sentiment proxies. The causality tests will provide a better idea of causality direction, and to be consistent with preceding theories (particularly Delong et al. (1990)), it is expected that the created proxy for sentiment to be a cause of market returns.

Secondly, a series of simple OLS regressions is implemented, introducing the control variables individually and in sequence with the purpose of following GMS's impact has the model becomes more complex. The intuition behind the relation between the novel measurement proxy and returns is that increases in positive sentiment will result in pressure to increase prices, followed by a fall shortly after as the euphoria fades away, returning prices to the fundamental value, while a negative sentiment wave toward the market explains a decrease in returns, representing a negative correlation. Prior to any experimentation, the results are expected to show positive correlations between GDP and PSI-20 returns, given its representation of the economy's well-being. A positive correlation is also to be expected between inflation and PSI-20 returns. Unemployment is anticipated to show a negative correlation as higher rates of unemployment are characteristic of periods with low economic and financial performances. Additionally, all the remaining sentiment variables are expected to present positive correlation with the returns, with the exception of VSTOXX given that it measures fear, and not confidence.

Lastly, the same process as before is applied but this time resourcing to Vector Autoregressive Models with the purpose of obtaining some insight about GMS's capacity to predict Returns. Having recognized the contemporaneous explanatory power of this model, and most particularly, of GMS on measuring market sentiment, the focus will shift towards the recognition of lagged sentiment over returns and the predictability power of said variables. In order to test this, the chosen methodology comprises a series of VAR models in which the several control variables are systematically added to the regression. These tests serve several purposes. Firstly, the models will allow us to quantify GMS's effect on the Index's returns, effectively measuring the impact of sentiment on the market, up until a couple weeks' delay. Secondly, the methodology permits the testing of the consistency and robustness of these results. As control variables are added to the model we expect the findings regarding GMS to remain virtually unchanged. Lastly, the inclusion of lags, endorses backing of some broad conclusions regarding the predictability of returns.

As an additional robustness test on the results, data regarding AGMS variables is used, comparing the previously created proxy for market sentiment measurement with another way using Google Trends data which possibly consists in another way to apprehend the same effect. The same methodology and testing are implemented.

4 Results

4.1 ***Granger Causality***

The next step consists in finding out the type of relation present among the variables in study. As such, Granger causality tests were applied, considering 1, 2, and 4 lags. With the intention of reducing printing space, the test itself is not presented in the current paper. This test serves the purpose of finding the causality relation between variables. To be considered relevant, the sentiment proxies should Granger Cause PSI-20 returns, and not the other way around, in which sentiment is caused by returns.

In all models, the null hypothesis that GMSneg does not Granger Cause PSI 20 returns is rejected, stressing the practicality of Google Trends data in measuring the effect of sentiment on returns and showing that GMSneg causes and shows explanatory power over returns. However the same does not happen to the GMSpos. This might comprise evidence that the negative effect of sentiment, at least as measured by Google data, is more prone to affect investors, and therefore the market, than its positive counterpart proxy. In fact, Da, et al. (2014) create their sentiment proxy, FEARS, using only terms which are negatively correlated with returns, corroborating the notion that negative sentiment as a greater impact on forces that drive the market. As for the sentiment control variables ESIER, ESIEU, ICE and VSTOXX, they are all Granger Caused by returns. With the introduction of more lags, these control variables' causality relations become bilateral, has they become both Granger caused and the cause of returns. Turnover, Abnormal Turnover and Unemployment do not show any causality relation, while Inflation seems to be Granger caused by Returns, which in turn are caused by GDP.

4.2 OLS Estimation

Following this paper's premise that GMS conveys peoples' sentiment, the goal is to prove that these behavioral effect's influence the financial markets by quantifying the novel proxy's explanatory power on the PSI 20 Index returns. Initial focus will be on the simplest regression in order to gain a broader insight on these variables' behavior, designating Index returns has the dependent variable and the rest as contemporaneous explanatory variables, demonstrated as follows:

$$PSI20ret_t = c + \beta_1\, GMSpos_t + \beta_2 GMSneg_t + \alpha_i\, Control\ Variables_t + \varepsilon_t$$

As previously stated, because of the verified correlations between sentiment proxies, the model only includes ESIEU and ICE. The summarized findings (presented in Table III) seem to support previous suppositions. GMS show large statistical significance and the correlations are consistent with what was speculated, GMSpos has a contemporaneous positive correlation with returns, and GMSneg a negative one. Again, when comparing both proxies, athough both are statistically significant at 1%,negative sentiment seems to have stronger explanatory power over returns than positive sentiment, which might reflect its greater impact on market fluctuations, as anticipated before. VSTOXX, also shows significance and a contemporaneous negative relationship with returns. Because both serve as negative sentiment measurement variables, when directly comparing VSTOXX and GMSneg, this last one appears to be a more competent measure of negative sentiment, presenting higher coefficients and t-stats. The effects endure notwithstanding the introduction of the control variables.

[Insert Table II]

4.3 Vector Autoregressive Model – VAR

These models, shown in Table IV, start with the simplest version of the VAR model including only GMS variables lagged by two weeks which set the tone for the subsequently more complex models, to which the remaining control variables will be added one at a time in the same order as in the OLS testing. The simplest model's low R-squared value indicates that these sentiment proxies alone are not enough to explain returns' movement, also being noteworthy that only the one week lagged proxies are statistically significant. This, in association with the low coefficients presented by said variables leads to the belief that sentiment, at least measured by Google data, although possessing some ability in predicting returns, it is not considered a particularly determinant factor. Other findings support those previously evidenced by Da, et al. (2014), attesting the existence of reversals on the relations with returns. Although these authors deal with daily data, this new approach of weekly lags still yields signal changes, with GMSpos comprising a negative correlation, and GMSneg a positive one after the first week. The fact that lagged GMS relations' inverse in relation with the contemporaneous relation indicates a fast reversal of sentiment, which could mean that sentiment (with particular emphasis on negative market sentiment given the power of GMSneg) has a relatively strong, although short impact in the Portuguese market. Compared to GMSneg, VSTOXX also shows a signal reversal, but at the second week and not at the first. This can be interpreted has GMSneg being a faster predictor of sentiment reversals towards the market than VSTOXX . The addition of the control variables sees the model becoming slightly better at explaining returns and the GMS' outputs holding sound.

[Insert Table III]

Additionally, the same methodology was applied to a model with 4 lags (Table V) in order to investigate the same effects over the course of a larger period of time, anticipating the possibility of supplementary reversals. Here the results seem to appear more inconsistent, with variables like GMSpos' signs of correlation changing as more variables are introduced to the model. However, these peculiar findings are not statistically significant. To note that GMSneg shows significance at the 10% level in the last regressions, meaning that the reversion ends after about the 2 week mark. Otherwise, most findings are in line with the 2 week lagged model. GDP is statistically significant at the time of its introduction onto the model, but loses significance as more variables are introduced, while the same happens for Sentiment control variables. ESIEU loses all significance in this new model while although it is close to 90% significance in the first lag, something that remains true has other variables are introduced and even in the final model. ICE's sign and significance remain the same they were in the previous model. VSTOXX shows significance only at the first lag, somewhat contradicting the 2 lags model. ICE on the other hand shows significance throughout the 4 lags, but with sign variability.

[Insert Table IV]

4.4 AGMS

The same tests as before with are applied with the AGMS variables as the created sentiment measurement proxies' in order to test their ability to express market sentiment in Index Returns, and compare these results with the ones containing the GMS models, turning this into a supplementary test on its robustness.

The correlation matrix does not suffer any substantial changes since the variables considered statistically significant with the GMS variables are the same as the ones considering AGMS, also with the same correlation directions. Meanwhile, the only difference amongst both ways of calculating sentiment through Google as experienced by the Granger Causality tests is that no AGMS shows any type of causality with returns. The OLS regressions, behave similar to the previous models with GMS, with the same evolution has control variables are introduced. The VAR models go a step further in weakening these proxies since in none of the models they are significant at the minimal level of 90%. These results seem to be congruent with the idea that the GMS method of measuring sentiment is more adept at reflecting its effects on returns compared to the AGMS variables.

[Insert Table V & Table VI &Table VII]

5 Conclusion

This paper shows that it is possible to use Google Trend's data to create proxy variables encompassing investor's feeling of optimism and pessimism towards the market.

Through a series of statistical procedures a variable composed of data from the amount of queries performed in Google was created, inferring that it correctly represents individual investor's market perception. By means of a series of linear OLS regressions the hypothesis that this new proxy for market sentiment possesses explanatory power over Psi-20 Index Returns was proven, and in fact it is more precise when compared with other known market sentiment proxies, with particular emphasis on Google's reliability in explaining negative sentiment.

Furthermore, applying similar methodologies to a series of VAR regressions, the results show that when compared with other sentiment proxies, Google data has the advantage of showing predictability power over returns, justifying the belief that internet queries can foreshadow market performance, following the results recognized by Da et al (2014).

These findings can be useful to future research in various topics, from achieving more precise estimates on consumption, to a better understanding of stock performance or even to reach a more realistic characterization of investor behavior in economic models. This topic can also be supported with a more accurate categorization of appropriate queries and update of the search terms in the creation of an aggregate market sentiment time series. The study can also be enriched by increasing the range of the study behind the Portuguese and American stock markets, or even introducing other financial products such as bonds and derivatives.

6 Bibliography

Abreu, M., Mendes, V., (2012). Information, Overconfidence and Trading: Do the Sources of Information Matter?, Journal of Economic Psychology, Vol. 33, Issue 4, pp. 868–881

Aouadi, A., Arouri, M. and Teulon, F. (2013). Investor Attention and Stock Market Activity: Evidence from France, Economic Modelling, Vol. 35, pp.674-681

Baker, M and Stein, J. (2004). Market liquidity as a sentiment indicator, Journal of Financial Markets, Vol. 7, Issue 3, pp. 271-299

Baker, M. and Wurgler, J. (2007). Investor Sentiment in the Stock Market. Journal of Economic Perspectives, Vol.21, Issue 2, pp. 129–151

Bank, M., Larch, M. and Peter, G. (2011). Google Search Volume and Its Influence on Liquidty and Returns of German Stocks, Financial Markets and Portfolio Management, Vol. 25, Issue 3, pp. 239-264.

Barber, B. and Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors, Review of Financial Studies, Vol. 21, Issue 2, pp. 785-818

Barberis, N., Shleifer, A. and Vishny, R. (1998). A Model of Investor Sentiment, Journal of Financial Economics, Vol. 49, Issue 3, pp. 307-343

Brown, G. and Cliff, M. (2004). Investor Sentiment and the Near-term Stock Market, Journal of Empirical Finance, Vol. 11, Issue 1, pp 1-27

Brown, G. and Cliff, M. (2005). Investor Sentiment and Asset Valuation, The Journal of Business, Vol. 78, Issue 2, pp. 405-440

Chang, E., Cheng, J., Khorana, A. (2000),An examination of herd behavior in equity markets: An international Perspective, Journal of Banking and Finance, Vol.24, Issue 10, pp. 1651-1679

Choi, H. and Varian, H. (2012). Predicting the Present with Google Trends, The Economic Record, VOL. 88, Issue 1, pp 2–9

Coval, J. and T. Moskowitz (1999), Home bias at home: Local equity preference in domestic portfolios, Journal of Finance, Vol. 54, Issue 6, pp 2045-2073.

Da, Z., Engelberg, J. and Gao, P. (2011). In Search of Attention, Journal of Finance, Vol 66, Issue 5, pp. 1461-1499

--. (2014). The Sum of All FEARS: Investor Sentiment and Asset Prices, The Review of Financial Studies, Vol. 28, Issue 1, pp.1-32

De Long, J., Shleifer, A., Summers, L. and Waldmann, R. (1990). Noise Trader Risk in Financial Markets, Journal of Political Economy, Vol. 98, Issue 4, pp. 703-738

Edmans, A., García, D. and Norli, Ø. (2007). Sports Sentiment and Stock Returns, The Journal of Finance, Vol. 62, Issue 4, pp. 1967-1998

Frazzini, A., and Lamont, O. (2008). Dumb Money: Mutual Fund Flows and the Cross-section of Stock Returns, Journal of Financial Economics, Vol. 88, Issue 2, pp. 299-322

Joseph, K., Wintoki, M. and Zhang, Z. (2011). Forecasting Abnormal Stock Returns and Trading Volume Using Investor Sentiment: Evidence from Online Search, International Journal of Forecasting, Vol.27, Issue 4, pp.1116-1127

Kamstra, M., Kramer, L. and Devi, M. (2003). Winter Blues: A SAD Stock Market Cycle, The American Economic Review, Vol. 93, Issue 1, pp. 324-343

Kanheman, D. (2003). A Perspective on Judgment and Choice: Mapping Bounded Rationality, American Psychologist, Vol. 58, Issue 9, pp. 697–720

--------------. (2011). Thinking Fast and Slow, 1st Ed. New York: Farrar, Straus and Giroux

Keynes, J. M. (1936). The general theory of employment, interest and money. London: Macmillan.

Latoeiro, P., Ramos, S. and Veiga, H. (2013). Predictability of Stock Market Activity Using Google Search Queries, Statistics and Econometrics Working Paper No. ws130605, Universidad Carlos III de Madrid.

Lee, C., Shleifer, A. and Thaler, R. (1991). Investor Sentiment and the Closed-End Fund Puzzle, The Journal of Finance, Vol. 46, Issue 1, pp. 75-109

Lee, W., Jiang, C. and Indro, D. (2002). Stock Market Volatility, excess returns, and the role of investor sentiment, Journal of Banking & Finance, Vol.26, Issue 12, pp. 2277–2299

Mondria, J., Wu, T. and Zhang, Y. (2010). The Determinants of International Investment and Attention Allocation: Using Internet Search Query Data, Journal of International Economics, Vol.82, Issue 1, pp.85-95

Preis, T., Moat, H. S. and Stanley, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends, Scientific Reports, Vol. 3, pp.1684

Shiller, R. and Akerloff, A. (2009). Animal Spirits: How Human Psychology Drives the Economy, and Why It Matters for Global Capitalism, New Jersey: Princeton University Press

Shiller, R. (2010). Measuring Bubble Expectation and Investor Confidence, Journal of Psychology and Financial Markets, Vol.1, Issue 1, pp.49-60

Tetlock, P. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market, The Journal of Finance ,Vol. 62, pp. 1139-1168

Tversky, A. and Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. Science, New Series, Vol. 185, Issue 4157, pp. 1124-1131.

7 Appendix I – Figures

Figure I - Graph for Google Trends search term input “crise” (Print Screen)

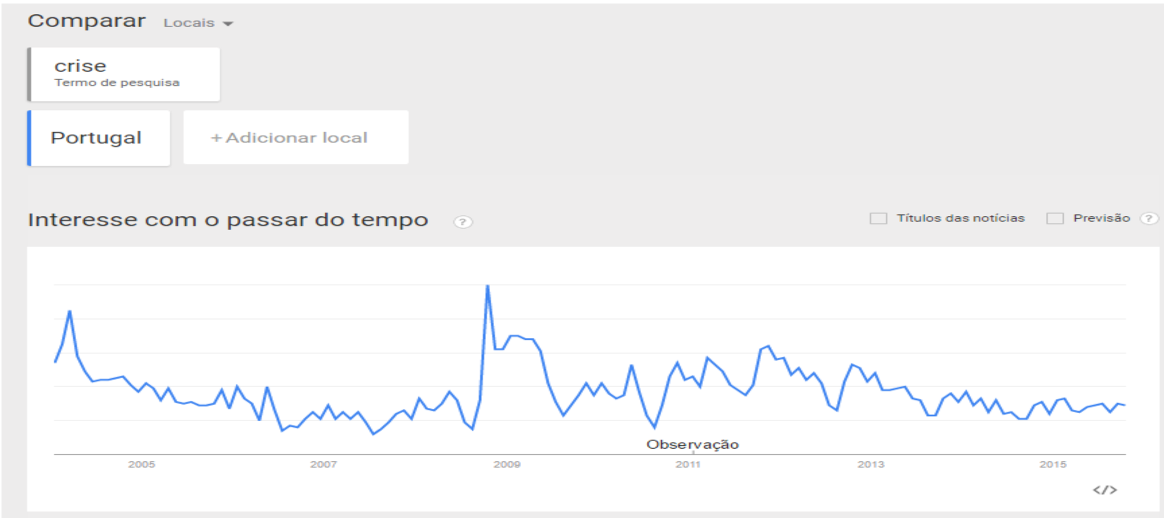


Figure II - Graph portraying the aggregate search term input “pensao+preço ouro+emprestimo+dinheiro+ordenado+comercio+ordenado minimo+revenda+tarifa+carreiras” from Google Trends. (Print Screen)

Figure III - GMSpos Graph - Graphical Representation of the generated variable GMSpos. It illustrates the combined evolution of the ΔSVWIB for the search terms: “Abono”, “Beneficios”, Desenvolvimento”, “Inovação”, “Investimento”, “Procura Emprego”, “Risco”, “Saldo”, “Trabalhadores” and “Oferta de Emprego”.

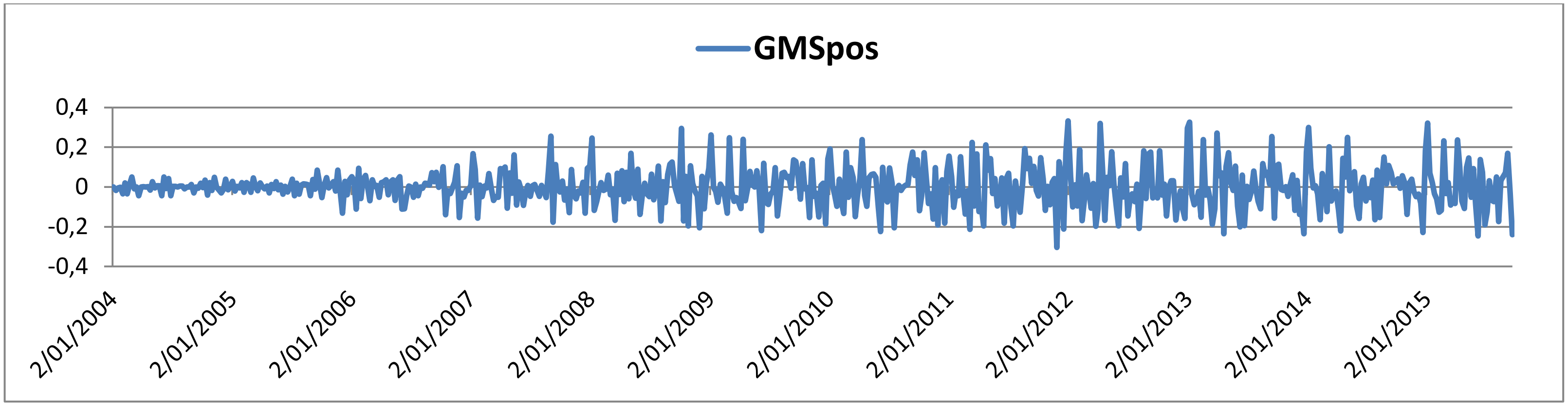
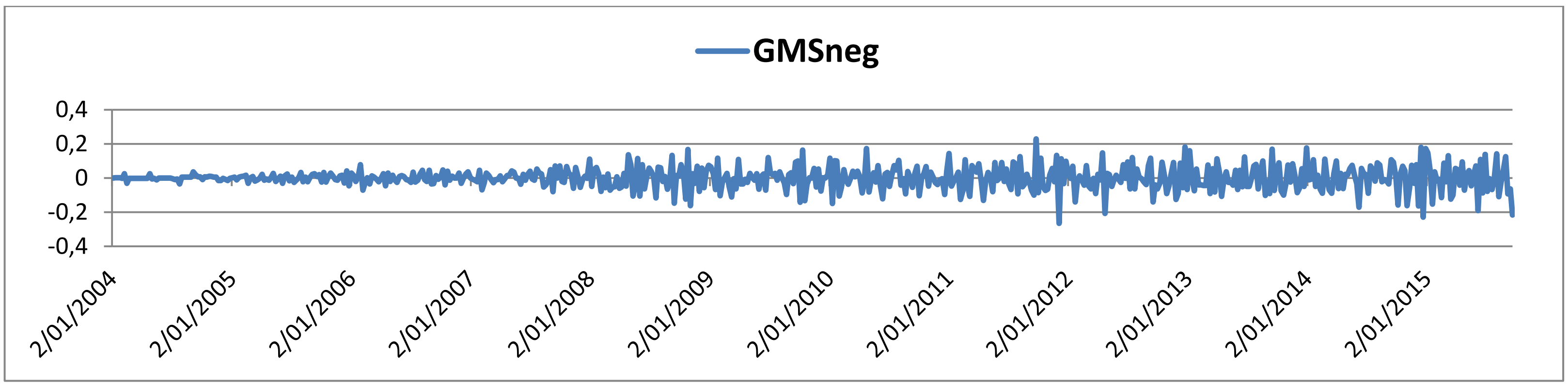


Figure IV – GMSneg Graph - Graphical Representation of the generated variable GMSneg. It illustrates the combined evolution of the ΔSVWIB for the search terms: “Tarifa”; “Pensao”; “Preço ouro”; “Emprestimo”; “Dinheiro”; “Ordenado”; Comercio”; “Carreiras”; “Ordenado mínimo”; “Revenda”.



8 Appendix II – Tables

Table II – Correlation Matrix – It portrays relations between several Sentiment Proxies: the created proxies and main focus of this paper GMSpos portraying positive sentiment, and GMSneg portraying negative sentiment; Economic Sentiment Indexes for different regions, namely Europe (ESIER), Euro-Zone (ESIEU), and Portugal (ESIPT); Indicador de Clima Económico (ICE); and the "fear Gauge" volatility index for Europe (VSTOXX). In brackets is shown the value of the T-Statistic of the correlation among variables.

Correlation [t-Statistic]	GMSPOS	GMSNEG	ESIER	ESIEU	ESIPT	ICE	VSTOXX
GMSPOS	1.000000 -----						
GMSNEG	0.175919 [3.807729]	1.000000 -----					
ESIER	-0.005466 [-0.116468]	0.010651 [0.226951]	1.000000 -----				
ESIEU	-0.007738 [-0.164874]	0.007972 [0.169869]	0.989296 [144.4545]	1.000000 -----			
ESIPT	0.004228 [0.090089]	0.009826 [0.209369]	0.793522 [27.78408]	0.765411 [25.34227]	1.000000 -----		
ICE	0.005519 [0.117597]	0.010787 [0.229860]	0.580197 [15.17837]	0.569411 [14.75890]	0.919569 [49.86544]	1.000000 -----	
VSTOXX	-0.036620 [-0.780799]	-0.000463 [-0.009874]	-0.588618 [-15.51419]	-0.536380 [-13.54161]	-0.405177 [-9.443068]	-0.142944 [-3.077340]	1.000000 -----

Table III – OLS Regressions - The simplest model lacks any control variable, resorting solely on GMS variables as explanatory instruments for the depend variable, PSI-20 Index Returns, encompassing the effect of sentiment on the Portuguese market. Each column represents a new regression in which a new control variable is added to the model individually. As the model becomes more complex it captures in a fuller fashion the different aspects that influence market returns and provides a basis for comparison with other known sentiment proxies. *,**,*** correspond to 90%, 95%, 99% level of statistical significance respectively.

	OLS 1	OLS 2	OLS 3	OLS 4	OLS 5	OLS 6	OLS 7	OLS 8	OLS 9
GMSPOS	0.07009***	0.067596***	0.065791***	0.065632***	0.065620***	0.064881***	0.064936***	0.064285***	0.063974***
GMSNEG	-0.13096***	-0.13033***	-0.12914***	-0.12916***	-0.12921***	-0.13707***	-0.13717***	-0.13711***	-0.13677***
VSTOXX		-0.00078***	-0.00117***	-0.00120***	-0.00118***	-0.00122***	-0.00122***	-0.00120***	-0.00125***
ESIEU			-0.00068***	-0.00075***	-0.00079***	-0.00080***	-0.00080***	-0.000662**	-0.000808**
ICE				0.000553	5.31E-05	-7.76E-05	-0.000116	0.001739	0.002064
GDP					0.000548	0.000564	0.000566	0.000277	0.000517
ABNTURNOVER						-0.521613	-0.555218	-0.423329	-0.326189
TURNOVER							1.71E-12	-9.83E-12	-1.22E-11
UNEMPLOYMENT								0.000362**	0.000368**
INFLATION									0.000874
C	-0.001735	0.018504	0.095277	0.102684	0.106531	0.118561	0.119281	0.102441	0.115495
R-squared	0.122389	0.175072	0.206639	0.207185	0.207414	0.215130	0.215147	0.224410	0.225168
Adjusted R-squared	0.118515	0.169597	0.199603	0.198376	0.196823	0.201220	0.199211	0.206648	0.205402
Prob(F-statistic)	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
Durbin-Watson stat	1.969083	1.959700	1.973330	1.971678	1.972907	1.984513	1.984914	2.009244	2.004787

Table III – VAR Models with 2 lags - As with the OLS regressions, the VAR models are also conducted with the individual sequential addition of the control variables to the simplest model as the farthest right columns in the table contain the more complex models. PSI-20 returns is the dependent variable in all the models. The simplest model, on the far left column contains, the dependent variable's own lags (up until 2 weeks), and the lagged versions of the GMS variables has explanatory variables. *,**,*** correspond to 90%, 95%, 99% level of statistical significance respectively.

	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5	VAR 6	VAR 7	VAR 8	VAR 9
LNPSI20RET(-1)	0,01015	-0,0733	-0,08619	-0,09012	-0,09241	-0,13313**	-0,1354**	-0,14271**	-0,14508**
LNPSI20RET(-2)	0,000517	0,027365	0,003539	0,006524	0,009947	0,049562	0,049829	0,043913	0,048172
GMSPOS(-1)	-0,0267*	-0,02671*	-0,02441*	-0,02419*	-0,02397*	-0,02418	-0,02332	-0,02333	-0,02402
GMSPOS(-2)	0,001977	0,004399	0,003459	0,003854	0,003386	-0,00013	0,000674	0,001052	0,000536
GMSNEG(-1)	0,087606***	0,088446***	0,088741***	0,090754***	0,089141***	0,096166***	0,095458***	0,094587***	0,09438***
GMSNEG(-2)	0,028409	0,034583	0,03638	0,037885*	0,038471*	0,054061**	0,053113**	0,051668**	0,052703**
VSTOXX(-1)		-0,00123**	-0,00127**	-0,00124**	-0,00119**	-0,0015***	-0,00152***	-0,00153***	-0,00149***
VSTOXX(-2)		0,001265**	0,001385***	0,001352***	0,001339***	0,001672***	0,001682***	0,001683***	0,0017***
ESIEU(-1)			0,008709***	0,009944***	0,009334***	0,009788***	0,009935***	0,009577***	0,009644***
ESIEU(-2)			-0,00878***	-0,00993***	-0,00941***	-0,00982***	-0,01002***	-0,00958***	-0,00949***
ICE(-1)				-0,01837	-0,02272	-0,03247**	-0,03407**	-0,03301**	-0,03319**
ICE(-2)				0,017525	0,020763	0,031501**	0,032582**	0,032634**	0,032477**
GDP(-1)					0,007549	0,009283*	0,008899	0,008835	0,008221
GDP(-2)					-0,00633	-0,00906*	-0,00865	-0,00873	-0,00829
ABNTURNOVER(-1)						1,615644	1,69127	1,681579	1,618267
ABNTURNOVER(-2)						0,192354	-0,20676	-0,17997	-0,24299
TURNOVER(-1)							2,42E-12	-2,18E-12	-1,74E-12
TURNOVER(-2)							2,08E-11	1,63E-11	1,76E-11
UNEMPLOYMENT(-1)								0,000264	0,000188
UNEMPLOYMENT(-2)								-5,53E-05	2,40E-05
INFLATION(-1)									-0,00489

	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5	VAR 6	VAR 7	VAR 8	VAR 9
INFLATION(-2)									0,003986
R-squared	0,035937	0,049226	0,107548	0,111942	0,11766	0,140395	0,14319	0,145938	0,14733
Adj. R-squared	0,022996	0,032133	0,087402	0,087777	0,089522	0,10439	0,102604	0,10075	0,097439
F-statistic	2,77708	2,879964	5,338514	4,632414	4,181475	3,89937	3,528082	3,229545	2,953074

Table IV – VAR models with 4 lags - same methodology as in Table IV, only considering up to 4 lags. *,**,*** correspond to 90%, 95%, 99% level of statistical significance respectively.

	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5	VAR 6	VAR 7	VAR 8	VAR 9
LNPSI20RET(-1)	0.011820	-0.087625	-0.092349	-0.098785	-0.102124*	-0.157546**	-0.172518***	-0.166036	-0.163931**
LNPSI20RET(-2)	-0.003514	-0.012263	-0.079226	-0.070078	-0.065444	-0.033347	-0.050762	-0.053583	-0.050706
LNPSI20RET(-3)	0.077564	0.075215	0.056409	0.057938	0.058648	0.068720	0.059111	0.072791	0.072274
LNPSI20RET(-4)	0.013248	0.034361	-0.009051	-0.007636	-0.007718	-0.030213	-0.029677	-0.027825	-0.028466
GMSPOS(-1)	-0.029144**	-0.027764*	-0.025873*	-0.026605*	-0.026496*	-0.025336	-0.022230	-0.022359	-0.022863
GMSPOS(-2)	-0.000108	0.002579	0.000123	0.001237	0.001362	-0.000671	0.002833	0.004070	0.003947
GMSPOS(-3)	-0.022523	-0.020855	-0.021876	-0.022158	-0.021065	-0.013259	-0.011316	-0.013921	-0.013312
GMSPOS(-4)	0.002298	0.001385	-0.001725	-0.001373	-0.000544	0.010899	0.011727	0.011003	0.010816
GMSNEG(-1)	0.089129***	0.088886***	0.084284***	0.085835***	0.083840***	0.088287***	0.085244***	0.085173***	0.085154***
GMSNEG(-2)	0.032756	0.040373	0.040846	0.040115	0.039151	0.052128*	0.048473*	0.048843*	0.049548*
GMSNEG(-3)	0.017988	0.021031	0.022535	0.019497	0.019073	0.016106	0.015044	0.015979	0.015428
GMSNEG(-4)	0.016446	0.020993	0.013669	0.010159	0.007154	-0.002217	-0.003771	0.000179	0.000130
VSTOXX(-1)		-0.001322**	-0.001104**	-0.001110**	-0.001068**	-0.001465***	-0.001485***	-0.001431**	-0.001407**
VSTOXX(-2)		0.000835	0.000847	0.000822	0.000841	0.001069	0.000958	0.000912	0.000891
VSTOXX(-3)		0.000396	0.000493	0.000641	0.000547	0.000806	0.000781	0.000857	0.000844
VSTOXX(-4)		0.000228	0.000224	0.000278	0.000338	0.000252	0.000378	0.000341	0.000322

	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5	VAR 6	VAR 7	VAR 8	VAR 9
ESIEU(-1)			0.009545***	0.010590***	0.010046***	0.010544***	0.010936***	0.011106***	0.011233***
ESIEU(-2)			-0.008380***	-0.011114***	-0.010780***	-0.011280***	-0.011216***	-0.011091***	-0.011196***
ESIEU(-3)			0.006099***	0.006624**	0.006891**	0.006803**	0.006609**	0.006308**	0.006279**
ESIEU(-4)			-0.007190***	-0.005807***	-0.005888***	-0.005776***	-0.006143***	-0.006170***	-0.006285***
ICE(-1)				-0.011187	-0.014673	-0.025961	-0.029543*	-0.029443*	-0.028199*
ICE(-2)				0.041195**	0.042469**	0.057767**	0.057696**	0.057731**	0.056913**
ICE(-3)				-0.010703	-0.007969	-0.000549	0.000528	-0.001382	-0.001099
ICE(-4)				-0.020756	-0.021717	-0.032601**	-0.031030*	-0.029674*	-0.030161*
GDP(-1)					0.006937	0.008965*	0.008850*	0.008622	0.008818
GDP(-2)					-0.003233	-0.006710	-0.007033	-0.006632	-0.006385
GDP(-3)					-0.004448	-0.005456	-0.005520	-0.006141	-0.006145
GDP(-4)					0.001237	0.003006	0.003534	0.003912	0.003617
ABNTURNOVER(-1)						1.417213	1.503386	1.456437	1.502515
ABNTURNOVER(-2)						0.202474	0.016394	-0.009993	0.042032
ABNTURNOVER(-3)						0.212950	0.076803	0.195348	0.244113
ABNTURNOVER(-4)						-0.364625	-0.774747	-0.805955	-0.762120
TURNOVER(-1)							-1.05E-11	-7.56E-12	-8.33E-12
TURNOVER(-2)							1.21E-11	1.35E-11	1.35E-11
TURNOVER(-3)							2.00E-11	1.89E-11	1.85E-11
TURNOVER(-4)							2.38E-11	2.54E-11	2.47E-11
UNEMPLOYMENT(-1)								-0.000310	-0.000380
UNEMPLOYMENT(-2)								0.000662	0.000750
UNEMPLOYMENT(-3)								-0.002249	-0.002245
UNEMPLOYMENT(-4)								0.001793	0.001758
INFLATION(-1)									-0.001010
INFLATION(-2)									0.003680
INFLATION(-3)									-3.19E-05
INFLATION(-4)									-0.001899

R-squared	0.046588	0.063710	0.155795	0.171914	0.177434	0.212038	0.220868	0.225063	0.225857
Adj. R-squared	0.020527	0.029272	0.116621	0.125370	0.122985	0.142384	0.142519	0.137500	0.128536
F-statistic	1.787640	1.849978	3.976988	3.693616	3.258724	3.044159	2.819040	2.570283	2.320747
Log likelihood	915.4745	919.5700	942.9679	947.3246	948.8361	826.6989	828.9244	829.9908	830.1933

Table V - Correlation Matrix involving AMGS variables – This table works the same as Table II, only with Sentiment as measured by Google computed through a different measure.

Correlation t-Statistic	AGMSPOS	AGMSNEG	ESIER	ESIEU	ESIPT	ICE	VSTOXX
AGMSPOS	1.000000 -----						
AGMSNEG	0.237974 5.220547	1.000000 -----					
ESIER	0.008257 0.175930	0.017414 0.371097	1.000000 -----				
ESIEU	0.008369 0.178327	0.016452 0.350591	0.989296 144.4545	1.000000 -----			
ESIPT	0.021421 0.456518	0.019601 0.417718	0.793522 27.78408	0.765411 25.34227	1.000000 -----		
ICE	0.021682 0.462086	0.014026 0.298881	0.580197 15.17837	0.569411 14.75890	0.919569 49.86544	1.000000 -----	
VSTOXX	-0.037605 -0.801821	0.011418 0.243312	-0.588618 -15.51419	-0.536380 -13.54161	-0.405177 -9.443068	-0.142944 -3.077340	1.000000 -----

Table VI - VAR Models with 2 lags for AGMS – This table depicts the same data treatment as Table IV but for the AGMS variables.

	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5	VAR 6	VAR 7	VAR 8	VAR 9
LNPSI20RET(-1)	-0,04269	-0,123**	-0,137**	-0,140**	-0,136**	-0,175***	-0,176***	-0,188***	-0,190***
LNPSI20RET(-2)	0,004915	0,026368	0,002995	0,004604	0,006347	0,037915	0,03874	0,032853	0,035131
AGMSPOS(-1)	-0,115742	-0,136985	-0,091195	-0,092978	-0,118152	-0,134876	-0,130676	-0,129814	-0,136589
AGMSPOS(-2)	0,044839	0,076257	0,063116	0,070564	0,08996	-0,004899	0,014468	0,017157	0,019419
AGMSNEG(-1)	0,189299	0,325415	0,191353	0,218856	0,289048	0,230199	0,232032	0,239661	0,240785
AGMSNEG(-2)	-0,02548	0,046544	0,185405	0,163827	0,131432	0,216539	0,184088	0,167367	0,138536
VSTOXX(-1)		-0,0012**	-0,0012**	-0,0012**	-0,0013**	-0,0015***	-0,0016***	-0,0016***	-0,0015***
VSTOXX(-2)		0,0012**	0,0013***	0,0013**	0,0013***	0,0016***	0,0016***	0,0017***	0,0017***
ESIEU(-1)			0,008***	0,009***	0,009***	0,009***	0,009***	0,009***	0,009***
ESIEU(-2)			-0,008***	-0,009***	-0,009***	-0,009***	-0,009***	-0,009***	-0,009***
ICE(-1)				-0,014823	-0,020963	-0,029*	-0,030*	-0,029*	-0,029*
ICE(-2)				0,014042	0,019402	0,028*	0,029*	0,029*	0,029*
GDP(-1)					0,008*	0,009*	0,009*	0,009*	0,008864
GDP(-2)					-0,007593	-0,009*	-0,009*	-0,009*	-0,009*
ABNTURNOVER(-1)						1,193493	1,307462	1,27315	1,205239
ABNTURNOVER(-2)						0,24915	-0,173262	-0,141038	-0,206425
TURNOVER(-1)							5,35E-13	-4,38E-12	-4,08E-12
TURNOVER(-2)							2,22E-11	1,68E-11	1,82E-11
UNEMPLOYMENT(-1)								0,000761	0,000714
UNEMPLOYMENT(-2)								-0,000511	-0,000463
INFLATION(-1)									-0,003807
INFLATION(-2)									2,92E-03
C	-0,001815	-0,001701	0,006335	-0,002171	0,005714	-0,028497	-0,020398	-0,028464	-0,041706
R-squared	0,003348	0,016184	0,074399	0,07752	0,087847	0,106956	0,109788	0,114139	0,115241
Adj, R-squared	-0,010029	-0,001503	0,053505	0,052418	0,05794	0,068338	0,066245	0,065731	0,061766
F-statistic	0,250295	0,915027	3,560801	3,088245	2,937357	2,769572	2,52137	2,35786	2,155064

Table VII - VAR Models with 4 lags for AGMS – This table depicts the same data treatment as Table V but for the AGMS variables.

	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5	VAR 6	VAR 7	VAR 8	VAR 9
LNPSI20RET(-1)	-0.042929	-0.143**	-0.148**	-0.154***	-0.156***	-0.211***	-0.224***	-0.223***	-0.220***
LNPSI20RET(-2)	0.008773	-0.017719	-0.083898	-0.076133	-0.073697	-0.048918	-0.061638	-0.060668	-0.058493
LNPSI20RET(-3)	0.072597	0.065987	0.040096	0.040559	0.034963	0.053417	0.045663	0.054526	0.054897
LNPSI20RET(-4)	0.021429	0.039385	-0.000715	0.003358	0.003515	-0.015889	-0.013779	-0.013943	-0.015210
AGMSPOS(-1)	-0.192744	-0.205447	-0.096166	-0.102062	-0.114617	-0.092931	-0.079094	-0.064648	-0.072471
AGMSPOS(-2)	-0.055030	-0.035943	-0.085746	-0.061438	-0.046615	-0.153565	-0.128073	-0.120644	-0.122584
AGMSPOS(-3)	-0.297031	-0.275938	-0.227910	-0.217401	-0.199303	-0.078370	-0.055858	-0.063835	-0.060004
AGMSPOS(-4)	-0.092953	-0.103632	-0.091890	-0.092943	-0.113369	0.033692	0.044641	0.045112	0.042080
AGMSNEG(-1)	0.287365	0.441092	0.105115	0.034088	0.070881	-0.043750	-0.033024	-0.073768	-0.046573
AGMSNEG(-2)	0.100268	0.303281	0.407924	0.303487	0.270173	0.440870	0.409108	0.425368	0.454535
AGMSNEG(-3)	0.516996	0.636949	0.340430	0.299039	0.210872	0.257644	0.272209	0.267785	0.284765
AGMSNEG(-4)	-0.014622	0.129077	0.029487	-0.022530	0.054453	-0.054352	-0.054104	-0.053441	-0.034152
VSTOXX(-1)		-0.0013**	-0.0011**	-0.0011**	-0.0012**	-0.0016***	-0.0016***	-0.0016***	-0.0016***
VSTOXX(-2)		0.000647	0.000634	0.000616	0.000682	0.000849	0.000757	0.000740	0.000719
VSTOXX(-3)		0.000500	0.000607	0.000740	0.000702	0.001075	0.001031	0.001088	0.001069
VSTOXX(-4)		0.000326	0.000315	0.000381	0.000471	0.000340	0.000470	0.000436	0.000426
ESIEU(-1)			0.009***	0.01***	0.009***	0.01***	0.01***	0.011***	0.011***
ESIEU(-2)			-0.008***	-0.011***	-0.010***	-0.011***	-0.011***	-0.011***	-0.011***
ESIEU(-3)			0.006***	0.007***	0.007***	0.0075**	0.0073**	0.0072**	0.0071**
ESIEU(-4)			-0.007***	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***	-0.006***
ICE(-1)				-0.006828	-0.012144	-0.021672	-0.025832	-0.025677	-0.024829
ICE(-2)				0.038*	0.041**	0.055**	0.056**	0.055**	0.054**
ICE(-3)				-0.014926	-0.011706	-0.002089	-0.000588	-0.001151	-0.001252
ICE(-4)				-0.017735	-0.019237	-0.032*	-0.031*	-0.031*	-0.031*
GDP(-1)					0.008*	0.009*	0.009*	0.009*	0.009*

	VAR 1	VAR 2	VAR 3	VAR 4	VAR 5	VAR 6	VAR 7	VAR 8	VAR 9
GDP(-2)					-0.005356	-0.007707	-0.008176	-0.008113	-0.007927
GDP(-3)					-0.004872	-0.005618	-0.005545	-0.005887	-0.005826
GDP(-4)					0.002465	0.003528	0.004137	0.004357	0.004035
ABNTURNOVER(-1)						1.092650	1.201296	1.151183	1.192159
ABNTURNOVER(-2)						0.287467	0.182986	0.182061	0.232264
ABNTURNOVER(-3)						0.340085	0.080996	0.145916	0.189769
ABNTURNOVER(-4)						-0.148756	-0.534672	-0.551834	-0.526541
TURNOVER(-1)							-1.32E-11	-1.07E-11	-1.11E-11
TURNOVER(-2)							9.05E-12	9.90E-12	1.01E-11
TURNOVER(-3)							2.65E-11	2.65E-11	2.66E-11
TURNOVER(-4)							2.49E-11	2.66E-11	2.62E-11
UNEMPLOYMENT(-1)								0.000146	8.44E-05
UNEMPLOYMENT(-2)								-7.58E-05	-1.32E-05
UNEMPLOYMENT(-3)								-0.001043	-0.001022
UNEMPLOYMENT(-4)								0.000879	0.000842
INFLATION(-1)									-0.000454
INFLATION(-2)									0.002955
INFLATION(-3)									0.000893
INFLATION(-4)									-0.002760
C	-0.001650	-0.004173	-0.019553	-0.045416	-0.042332	-0.079149	-0.060956	-0.059328	-0.051583
R-squared	0.012623	0.030966	0.124983	0.139517	0.151994	0.188660	0.198954	0.200324	0.201118
Adj. R-squared	-0.014367	-0.004677	0.084379	0.091153	0.094223	0.114480	0.115609	0.106795	0.097121
F-statistic	0.467693	0.868778	3.078101	2.884713	2.630950	2.543283	2.387096	2.141830	1.933891

Table IX – Summary of literature – This table contains a brief summary of major papers dealing with Market Sentiment, from interpretations of market Sentiment and its measurement to different methodologies applied and major conclusions and scientific contributions.

Authors	**Core Assumptions**	**Methodology**	**Conclusions**
Choi and Varian (2012)	Google Trends display real consumer preferences and behaviors.	Simple seasonal AR models with Google Trends data outperform those without data.	Search Engine data can serve has contemporaneous forecasters of economic indicators.
Tversky and Kahneman (1974)	People misjudge likelihoods and reliability of information concerning uncertain events. These effects have a particular emphasis on the Economic and Financial areas.	Surveys and practical experiments– Behavioral Economics	Identification of several heuristics that affect agents' decision making process under uncertainty, leading to systematic errors.
Coval and Moskowitz (1999)	The puzzling, well-documented existence of a strong preference for investing in domestic markets – Home-country bias. Sets to assess the importance of geographic proximity.	Measure the degree of preference for proximate equities using global coordinates and identifying the top holdings for each fund manager.	Information asymmetries drive geographic preferences and go as far as indicting a link between local equity preferences and cross-sectional asset pricing.
De Long et al (1990)	Existence of irrational investors, "noise traders". Its abundance affects price formulation through arbitrage limitations and increase in risk they themselves create.	Overlapping generation model with two types of agents and two assets, one riskless.	Financial market anomalies can be explained by the idea of noise trader risk. Rational Investors are forced to "respond" given the extent of irrational investors' influence.
Baker and Wurgler (2007)	Investor Sentiment is taken has exogenous. Harder to arbitrage stocks are susceptible to wider gaps between their current and fundamental prices during periods of higher or lower sentiment.	Create a variable to measure sentiment based on several known proxies (VIX,IPO's First Day Returns, Volume, etc…). Assess results against mutual fund flows, current returns, and test predictability among several groups of stocks depending on their difficulty of arbitrage.	Prove that harder to arbitrage stocks are indeed more affected by sentiment, which in turn has a measurable effect on the market.
Kamstra, et al (2003)	Using evidence from psychology that weather affects mood, the authors test the consequences	The authors model differences in seasonal variation in daylight as a proxy for sentiment and measure its influence on returns and risk	Even controlling for well-known environmental factors, the seasonal effect still has a significant impact

	that that change in mood can have on the markets through variations in risk aversion behavior.	tolerance.	on returns on a worldwide case.
Shiller, R. (2010)	Through surveys the author can get an idea regarding investors' outlook on the market performance.	Compare survey results with historical data and existing sentiment indicators related with the stock market.	Surveys serve has a form to measure investor confidence.
Barberis et al (1998)	People have idiosyncratic excessive reactions to news, which affect lead to distinct and subjective actions on the market regarding the statistical weight and importance of the news.	Parsimonious model with one investor and one asset with two possible states of nature and following a random walk.	The impact of news seems to be inconsistent with economic theory, with important news being taken lightly by investors, and vice-versa.
Brown and Cliff (2004)	Investigate the formation of expectations over the market (sentiment) using surveys.	Analysis focuses on evaluating survey ability to measure sentiment relative to other sentiment measures, and on survey's ability to predict returns.	Correlation with other sentiment proxies. Little predictability power with returns
Brown and Cliff (2005)	Check long run relation between investor sentiment and market returns.	Use survey time series to explain pricing errors, indicating market is overvalued during periods of optimism.	Surveys indeed predict market returns over the next 1-3 years and have the ability to explain market deviations from intrinsic value. Irrational sentiments affect price level.
Barber and Odean (2008)	Investors are net buyers of attention grabbing stick given limitations of time from investors when searching for investment alternatives.	Attention grabbing stocks are reported in news, hence they search news databases to categorize stocks depending on their attention seizing prospects. Check results against the volume and returns. In the model, agents are faced with many options, and it is tested if they pursue the one that caught more attention on the news.	Agents do in fact acquire more attention grabbing stocks than otherwise, and the effect is particularly felt in unsophisticated investors. Extreme returns are noted for said stock on the very short-turn.
Da et al (2014)	Construct an investor sentiment measure based on internet search behavior of households, which is an improvement in	FEARS is related with market returns, volatility (VIX), and mutual fund flows to test the "noise trading" hypothesis.	FEARS has a negative correlation with market returns but shows reverses in the future, consistent with sentiment induced mispricing,

	revealing preferences.		and the relative effects are consistent with the notion of flight-to-safety. They confirm a strong correlation between FEARS and VIX, as well as predicting reversals.
Lee et al (1991)	Indicate market sentiment as a possible solution for the closed-end fund puzzle, which supposedly reflects the expectations of individual investors. The proposed hypothesis indicates that discounts are high when investor sentiment is low/pessimistic about future returns.	The authors compare Closed-Fund Discount movements amongst with stock returns. They use the difference between the net asset value of a fund's holdings and it's the fund's market price as a proxy to market sentiment, stating that because those types of funds are mainly held by individual investors, the difference reflects better sentiment's effect	They find that changing investor sentiment makes funds riskier than the portfolios they hold and so causes average underpricing of funds relative to fundamentals, illustrating the effect of sentiment on these assets.
Lee et al (2002)	Shifts in perception of risk by noise traders is associated with sentiment	GARCH model to test the impact of noise trader risk with conditional volatility and returns serving as proxies for sentiment.	Investor Sentiment indeed affects returns through means of risk perception.
Preis et al (2013)	Online search queries constitute "early warning" signs for market movements, even anticipating future trends.	Analyze several search terms related with the stock market and implement a hypothetical investment strategy based on market movements.	Google Trends data can actually predict some economic behavior trends; Provide a quantifiable relation between search volume and stock market prices.
Edmans et al (2007)	International soccer results serve as a proxy for mood, taking into consideration the importance given to it by some countries. Its analysis can generate meaningful insights on mood's effect over the market, as mood swings will influence positive and negative perspective over the market.	After regressing market returns with its own lags, and dummies comprising working days in order to remove any of those effects, the authors regress the resulting residuals with dummies involving loses and gains in sports, in order to quantify each effect.	Soccer results, particularly important National Team matches, have an impact on mood which is conveyed to the market and most heavily felt by small stocks.
Joseph et al (2011)	Online financial ticker searches can forecast abnormal stock	Empirical strategy involves weekly classifications of each stock of the S&P500 into	Confirming previous studies, this paper finds that search intensity

	returns and trading volumes. The effect is predicted to be more prominent on stocks that are harder to arbitrage and more volatile.	quintiles and building a portfolio of long positions on those stocks which show highest search intensity.	indeed forecasts abnormal returns and trading activity during the previous week. Their findings regarding stock volatility indicate that search intensity serve as a valid proxy for investor sentiment.
Da et al (2011)	Given constraints of time and information processing, individual investors are prone to acquire those stocks that get their attention. The authors hypothesize that search volume serves as a measure of investor attention towards particular stocks.	They correlate search query (SVI) data with other proxies for attention and find positive but with low levels of correlation. A VAR shows that SVI is better and faster at predicting changes in attention. They use order execution reports to successfully find evidence that SVI illustrates the behaviors of retail investors. They also use IPO data in relation with SVI to prove attention hypothesis by Barber and Odean (2008).	Internet search volume constitutes a direct proxy for individual investor attention. As a test on Barber and Odean's (2008) theory, they find that SVI predicts increases in stock prices and a following reversal.
Bank et al (2011)	Google data allows perceiving the real attention firms receive through the number of internet queries, which is correlated with stock market performance.	Employ new data set for the German Stock Market, focusing on the firm's Stock Ticker, and employing google data as a variable. Univariate Analysis between average stock portfolios, comparing trading activity measured by Google and illiquidity. Panel Analysis considering lags of trading volume measured by google, control variables and different measures of illiquidity.	Significant correlation between Google searches and Trading Activity. Negative correlation with illiquidity, presumably due to asymmetric information. Positive Short-run correlation with future stock returns
Latoeiro et al (2013)	Web search queries provide insight on investment decisions via the volume of trading activity, as measured through liquidity	Use Yahoo Search Engine and focus on Market Indices; Explain market activity measures like trading abnormal volume and volatility with abnormal google searches (difference between verified search amount and previous four week average), also considering control variables like lagged versions and other proxies. Sort portfolios according to web search amounts	Web searches foresee a drop in index returns' and increase in volatility.
Aouadi et al	Investor's decision making	Use several measures for liquidity	Investor attention, measured by

(2013)	constraints generate mispricing. Attention theories can explain these mispricing, by using Google's Web queries (GSV) to measure attention.	Perform Correlation Coefficients between GSV and stock traded volume. Regress Amihud illiquidity ration with GSV and control for other known drivers of liquidity, including lags. Also regress GSV and trading volume with the standard deviations of returns to measure Google's effect on market volatility.	Google is a significant determinant of stock market volatility and a driver of stock market liquidity.
Frazzini and Lamont (2006)	The reallocation of money between funds can act as a proxy for sentiment by analyzing the stocks compromising said funds. Since funds usually go for safer stocks, stock with high ownership percentage by funds show market pessimism on said stocks.	Correlate stock returns and recorded trades of mutual fund flows, using the latter as a proxy for sentiment. Fund flows have positive contemporaneous correlations with stock returns.	If a fund holding a particular stock receives strong inflows, the performance of that stock will be inferior. Irrational Investors in fact lose money.
Mondria et al (2010)	Search query dataset serve as a measure for attention for stocks	Models how attention impacts market decision making. Measure attention allocated to national stock using instrumental variables.	The authors find two way causality between home bias and attention proxies.