A Work Project, presented as part of the requirements for the Award of a Masters
Degree in Finance from the NOVA – School of Business and Economics.

Exploring the Predictive Power of Google Searches over the US Stock Market

João Guilherme Martins Borges Sàágua; Student Number 588

A Project carried out on the Financial Markets theme, under the supervision of:

Professor Melissa Prado

6th January 2014

Exploring the Predictive Power of Google Searches over the US Stock Market

ABSTRACT

This paper takes search intensity for stock tickers in Google (SVI) as a direct measure of retail investor attention and assesses whether it holds predictive power over short-term market outcomes. In a sample of the most representative US stocks, during the period 2005 – 2008, I provide evidence that (1) surges of investor attention forecast higher stock liquidity and volatility; (2) depending severely on what is considered an abnormal level of SVI, retail investor attention can also be priced; and (3) SVI does not relate to firm-specific features, such as size and value. Furthermore, I extend the investigation to the aggregate market level, finding that investor attention to the market index predicts greater market liquidity, volatility and return.

Keywords: Investor attention; Search data; Stock market predictability Noise trading.

I. Introduction

Timing the major stock market outcomes, such as liquidity, volatility and returns has been a true challenge to investors and researchers over decades. Yet, a significant portion of short-term market phenomena remains unexplained.

Asset pricing models assume the market is efficient at least at a semi-strong level, meaning that all the information is immediately incorporated into security prices, after being publicly released. In order for this assumption to hold there is a fundamental requirement: investors must pay attention to the assets.

Information and attention form an interesting couple, the first is useless without the second, but an overabundance of information will most likely reduce the attention being paid to it. In short and quoting Herbert Simon, "a wealth of information creates a poverty of attention".

People's attention can be seen as a resource, and as any other resource in an economic environment, attention is also scarce. Hence, there is the need to be effectively attentive to the information available, which is mostly released in the form of news. What is more is that attention is time-varying, and following Andrei and Hasler (2012), investors' attention oscillates in accordance with the state of economy. Hence, attention fluctuations are usually between two extremes: either investors are very attentive to the financial and economic press, something that occurs mainly during downturns; or they almost disregard news, which is typically verified in a bullish phase.

After establishing some of investors' attention's features, such as being a limited resource whose usage varies over time, it is pertinent to bring back the asset pricing models into play, such as CAPM (1961-1966), Fama-French (1993) or Carhart (1994). These three key tools that aim at forecasting stock returns only look at market and firm-specific components, leaving out of consideration issues related to investors'

psychological behaviour. Naturally, this last element is particularly hard to measure, however recent studies carried out by Da, Engelberg and Gao (2011); Joseph, Wintoki and Zhang (2011) among others, present a timely effective measure of investors' attention, resorting to search frequency for companies' ticker symbols in Google, the search engine leading by far the landscape of internet search, possessing, in July 2013, a 67% share of all the searches executed by web users worldwide[1].

Since 2004, Google has been storing on its database the volume of searches for a certain word, expression or symbol, and making it available to public on a weekly basis, through a website called Google Trends (http://www.google.com/trends/).

A relevant link between investors' psychological behaviour, hereby translated by intensity of online searches, and market movements is of major importance, especially, if data collected on web platforms turns out to hold some predictive power.

To corroborate this idea and show its usefulness Choi and Varian (2009) argued that "harnessing the collective intelligence of millions of users, Google web search logs can provide one of the most timely, broad-reaching influenza monitoring systems available today", in other words, this data that allegedly mirrors "what we collectively think" represents an important piece of information about "what might happen in the future" (Rangaswamy, Giles and Seres, 2009, p.58).

In fact, when people search for something somewhere, they are explicitly being attentive to it, which makes a query for a stock in Google, or in any other search engine, an unequivocal manifestation of attention.

As it is, to test whether the Search Volume Index (SVI) for tickers, as a direct measure of investor attention, indeed holds predictive power over stock market activity I use a sample of stocks included in the S&P 500 over the period 2005-2008, and SVI impact is in the first place assessed on liquidity. As a matter of fact, I find that rises in

[1] Source: http://www.searchenginejournal.com

levels of investor attention verified in a certain week predict a higher share turnover ratio in the following week. Provided that investor attention has a significant impact on liquidity, I assess its relation with stock returns volatility, expecting it to be positive. Consistently, I encounter an also dynamic relationship between investor attention and volatility of returns, with a surge in the former leading to an increase in the latter.

Moving forward in the analysis, I evaluate if SVI also encloses some price impact. I disclose that only abnormal levels of attention registered during the present week are capable of predicting higher returns in the subsequent week, being consistent with the attention-induced price pressure theory of Barber and Odean (2008). However, this result does not appear to be as strong as other papers present it, with the statistically significance of this interaction being substantially affected by the measure of abnormal SVI employed.

After acknowledging the main effects of investor attention a question of major relevance is worth raising. What type of investors own the attention being captured by ticker searches in Google? The answer – individual/less sophisticated investors – can be found only by following intuition. Logically, institutional investors benefit from an easy access to specific platforms providing extensive financial information as well as vast databases, such as Bloomberg and Reuters, which makes them unlikely users of Google for ticker searching. In contrast, retail investors, lacking those professional tools, resort to Google or other universal search engines to seek financial information. Da et al. (2011) deliver the formal proof for this fact.

Once it is established that retail investors' attention is in charge, I look at what type of stocks they search in Google, with the objective of going a step further and substantiating some well-known anomalies with a possible explanation coming from investor attention. I put in place a naïve approach and a formal econometric analysis.

Both yield the same results: size anomaly, value anomaly and momentum anomaly do not relate to retail investor attention. In addition, firm-specific characteristics, such as market capitalisation and Book-to-Market ratio do not establish any relevant connection with SVI. In this part of the study, I only provide some evidence for a price reversal after the positive returns earned when investors are quite attentive.

The final exercise revisits the above mentioned market outcomes, from an aggregate market perspective, reaching the surprising conclusions that the search volume for the term S&P500 has explanatory power over market liquidity, volatility of returns and fluctuations of the index quote.

Besides contributing as a robustness test to the results documented by similar studies, using a sample of the most salient and economic meaningful US stocks, and the most appropriate method for each analysis, according to Petersen (2005), this paper originally approaches the theme of anomalies, finding no relevant results, and successfully extends the findings found for individual stocks to the overall market. Throughout this investigation of investor attention, noise trading theories were corroborated.

The rest of the paper is structured as follows. Section II reviews the relevant literature on this topic. Section III describes the specificities of the data used. Section IV summarises the tested hypotheses as well as the methodology followed. Section V presents and thoroughly discusses the empirical findings. Section VI concludes reinforcing the importance of the behavioural component in the financial markets universe.

II. Literature Review

Primarily, this paper establishes a link with the finance literature that approaches investors' psychological behaviours, particularly attention and sentiment. In the

background, it relates to studies that deal with the implications of noise trading on the market. The two previous phenomena are intimately related.

Shleifer and Summers (1990) extensively study the noise trading issue, suggesting that retail investors, who are often less informed, tend to trade impulsively, basing their investments on noisy signals which they perceive as good opportunities to participate in the stock market. The aggregation of those signals generates sentiment. In the same market, wise and informed investors who ponder risk – arbitrageurs – co-exist. Their role is to explore securities' mispricing, driving their price close to fundamentals. Yet, some misalignment between stock price and company's intrinsic value tend to subsist, indicating that noise traders not only influence the market, but also sometimes do it so intensely that they limit the action of risk-averse arbitrageurs. The reason why this is possible as well as its proof is provided by Barber, Odean and Zhu (2009a), who investigate sets of retail brokerage data and report high correlations between trades from individual investors, from which persistent psychological biases are identified. As a consequence, noise traders may or not be responsible for some well-known anomalies that have been alive over years, but, through investor sentiment, they are very likely in charge of temporary market events, such as extreme trading volumes, volatility shocks and significant security mispricing, as shown by Barber et al. (2009b).

Going back to the genesis of sentiment, to spot the signals, investors must have paid attention to financial assets.

The first research on investors' cognitive sphere was done by Merton (1987), who posits that each investor knows only a finite set of stocks, leading to a non-diversified portfolio. In his model equilibrium, stocks which have low visibility and thus are more neglected by investors achieve higher returns. For the stocks that are often on the investor's spotlight, Barber et al. (2008) theorise an attention-induced price

pressure, supporting that retail investors in the presence of attention-grabbing events typically become net buyers, and provided that the supply of stocks is inelastic (at least in the short-run), an upwards price pressure is verified, leading to positive returns. They back-up this hypothesis with evidence from three attention-grabbing events: extreme trading volume, abnormal returns and news coverage.

Besides the above three, other proxies of investor attention exist, however all of them are indirect, since they rely on the assumption that when these events took place investors paid attention to them, a statement that does not always hold. In these circumstances, Da et al. (2011) propose a direct measure of investor attention, Search Volume Index (SVI) in Google for stock tickers, which not only reveals attention but also allows gauging its level.

Furthermore, to avoid the risk of using a variable that is intended to have a true and unique explanatory power over stock returns, but is no more than a proxy for other external events, such as news coverage, Da et al. (2011) demonstrate in their exercise that 95% of the variation in web search queries, occurs independently from the amount and type of economic and financial news, which are sporadic while SVI is continuous.

Da et al. (2011) depart from the work of Barber et al. (2008) and assert that SVI points out a buy decision for the asset being searched, as when investors intend to purchase a stock, they can choose from an immense universe of stocks, so they pick from the ones which called their attention, whereas when they consider selling a stock, they usually have available only the stocks they own, on which they should already possess a considerable amount of knowledge.

Da et al. (2011) proceed to the clarification of whose attention is coming into play with Google searches for tickers. In order to demonstrate that retail investors are the ones responsible for this sort of online search, they analyse a vast set of data for

orders and trades across different market centres. Using Madoff as a representative centre for individual investors and NYSE as a typical centre for institutional investors, they reach the conclusion that abnormal levels of SVI consistently have a great impact on the first, while the second remains unaffected, confirming what intuition suggests: retail investors, less sophisticated and unlikely current users of professional finance platforms are the ones who utilise Google to acquire financial information about stocks.

Finally, using the sample of Russel 3000 stocks, they corroborate Barber et al. (2008) attention-induced price pressure, and also Barber et al. (2009) price reversal as a second consequence of noise trading. In this last case, Da et al. (2011) document that returns revert in approximately the same magnitude, within a year horizon.

In the same field, Joseph, Wintoki and Zhang (2011) triangulate these results within the S&P 500 sample, adding evidence for positive dynamic correlation between search intensity and trading volume. Taking a step forward they reach the striking result that SENT (their name for the SVI variable) is positively correlated with the market risk factor (Rm-Rf), following they find out that SENT correlations with Value (HML) and Momentum (UMD) are similar to the ones verified between these two variables and the market factor, which leads them to the conclusion that SENT can potentially mimic the market risk factor. Furthermore, when applying the Carhart model with the extra variable for search intensity to the empirical data, Joseph, Wintoki and Zhang (2011) realise that after controlling for all the four factors, SENT still possesses incremental information, suggesting that asset pricing models should take under consideration investors' behaviour variables in an attempt to forecast stock returns.

Finally, Vlastakis and Markellos (2012) evaluate the impact of attention paid to the 30 largest NASDAQ companies' names on the respective stock returns' volatility. Vlastakis et al. (2012) account for both, information demand and information supply, by

using SVI and news. They find a high correlation between SVI and historical volatility and a lower, still statistically significant, relation with idiosyncratic implied volatility.

Dimpfl and Jank (2012) carry out a different analysis on the grounds of volatility and they find a strong co-movement of Dow Jones index volatility and retail investor attention, measured by SVI. Reporting Granger causality between the two realities, Dimpfl and Jank (2012) identify the following pattern: "a fundamental shock in volatility triggers noise trading, which, in turn, leads to more volatility."

III. Data

To quantify the main variable of interest of this study – investor attention – I use the Search Volume Index in Google for the stocks included in the main US stock index, the S&P 500. Concretely, this dataset is the one from the paper "Investor Information Demand: Evidence from Google Searches around Earnings Announcements" (2012), which is freely provided by its authors Drake, Roulstone, and Thornock[2].

Originally, these data have daily frequency and cover the period under analysis, ranging from the 1st of January 2005 to the 31st of December 2008. As this dissertation targets the disclosure of more comprehensive relations between aggregate queries for stocks on a universal search engine and broad market outcomes, there was a need to convert the data frequency to weekly. That was rigorously executed on Excel, resorting to pivot tables, which allow the conversion of daily data into weekly data, by taking its sum.

Furthermore, Drake et al. (2012) downloaded this data using the fixed scaling option, a feature that scales each SVI value according to a specific observation, the first one. This feature does not harm the interpretation of regression coefficients,

---

[2] Source: http://byuaccounting.net/drake/ProgramsData1.php; The Search Volume Index data was not retrieved directly from Google Trends since nowadays Google only provides data normalised and scaled to the sample average, disabling comparisons among different stocks.

particularly, if they are in a logarithmical form the interpretation is done as usual, considering a 1% increase in the explanatory variable.

The companies' stocks were identified in Google through their ticker symbols, rather than companies' names or even companies' most famous products or services. The reasoning behind this choice relates to the fact that the relevant searches for the sake of this analysis are the ones that aim at investment purposes. Therefore, it was assumed that people would search for a stock ticker symbol in order to collect relevant financial information, which is quite useful when a limited amount of money can be allocated to a wide variety of stocks, diverging a lot in performance and fundamentals. Trying to capture investors' attention by looking at search data on companies' names or most famous products would turn out to be inaccurate, not only because a company may be known by more than one name or because those names may have other meanings (Ex: Apple, Amazon) but also because people may search them on Google, the worldwide search engine leader, for various reasons other than investing. Hence, tickers, uniquely assigned symbols were chosen as the best signal of investor's attention, raising no ambiguities, except for the tickers that may have other meanings or even be confused with other acronyms, (Ex: AA, ABC, ZION, COST). Companies with this sort of noisy tickers or companies for which the stock ticker coincides with the name of the company itself (Ex: IBM) were eliminated from the sample, being consistent with the aim of capturing only the attention of potential investors.

Google Trends data for stock tickers knows some upfront limitations. (1) Due to its nature it is likely to contain some noise; as described above and also by the application of a final filter, all the efforts to minimise its impact were done. (2) It may also have some degree of seasonality, in the aftermath of specific company events or correlate with news. Nevertheless, this fact does not affect the analyses' rigour, as

already explained, Da et al. (2011) prove that "SVI indeed captures investor attention and is different from existing proxies of attention".

Finally, as one of the goals of this paper is to provide a robustness test to previous studies, the variable accounting for abnormal levels of investor attention was constructed in multiple ways[3].

Regarding the sample of stocks, in order to avoid survivorship bias, before applying filters, it comprises all the stocks that were ever included in the S&P 500, during the time span under analysis. As a matter of fact, there was a need to acknowledge the historical constituents of the referred index, information that was possible to access through Wharton Research Data Services website[4].

After possessing all the company tickers, firm-specific characteristics and market data for individual stocks, the market index and the VIX index were retrieved from a Bloomberg terminal and Yahoo! Finance. Data on macroeconomic indicators was obtained from Trading Economics website[5]. Weekly Fama-French factors were obtained from Fama and French data library[6]. For the analysis at the aggregate market level, SVI for the term S&P500 was downloaded directly from Google Trends.

After all the previously mentioned necessary adjustments to guarantee the validity of the analyses, the two datasets (SVI and Market data) were merged and a final filter was applied: all the stocks that did not have both, a valid SVI (different from zero for more than 70% of the observations) and valid market data (complete time-series of return and trading volume), were excluded. Once this narrowing down process, which

---

[3] Abnormal SVI is the difference between the log-SVI of each observation and: the log-median over the past 4 weeks; the log-median over the past 6 weeks; the log-median over the past 8; the log-median of the whole sample. Abnormal SVI was also computed in the same way, but using log-averages instead of log-medians. Finally, for the returns analysis I also tried out variations in SVI. When nothing is specified, the Abnormal SVI variable being used corresponds to $logSVI_t$- $logmedian(SVI_{t-1},…SVI_{t-4})$).
[4] https://wrds-web.wharton.upenn.edu/wrds/
[5] http://www.tradingeconomics.com/
[5] http://mba.tuck.dartmouth.edu/pages/ faculty/ken.french/data_library.html

manually goes through each individual stock, was completed, I attained the final sample, which is composed by 215 stocks.

As the stocks included in the S&P 500 are mostly liquid, there was no reason to worry about bid-ask bounce problems derailing the analysis of returns.

A final remark concerning winsorization of the variables ought to be made. Winsorizing the variables is a double-edged sword in this case. On the one hand, decreases or cancels the impact of extreme observations, potentiating a higher correlation between different realities. On the other hand, it forces the loss of a relationship of utmost interest, the relationship between abnormal values of SVI and abnormal values of return or volume. Thus, I decided not to winsorize the variables.

IV. Hypotheses development and Methodology

Building upon the work of Barber, Odean and Zhu (2009), which comprehensively investigates the impact of noise trading on asset pricing, and Barber and Odean (2008), which aims at explaining the triggers of noise trading, by studying attention-grabbing events, I develop a robustness test to the results posteriorly found in the related literature about retail investors' attention, as measured by the intensity of web searches for stock tickers in the worldwide leader search engine Google.

In that sense, in order to evaluate whether Google SVI possesses predictive power over main stock market outcomes, such as liquidity, volatility and returns, the following hypotheses are tested:

H1: Seasholes and Wu (2007) document that several events to which retail investors pay attention lead to more active trading, usually with individual investors becoming net buyers. Joseph, Wintoki and Zhang (2012) find that surges in the volume of search queries for stock tickers precedes an increase in trading volume over the following week. In this paper, I test this effect of retail investor attention on stocks'

liquidity, using share turnover ratio, from a dynamic and contemporaneous perspective, assessing the impact of previous week SVI and current week SVI, respectively, on the current week share turnover ratio.

H2: Vlastakis and Marlellos (2012) present strong evidence that information demand, using web searches for companies' name, helps explaining contemporaneous stock returns volatility, both historical and implied. On the same field, Dimpfl and Jank (2012) report Granger causality between web search queries for tickers and realised volatility. I evaluate whether unequivocal declarations of attention to individual stocks throughout the current and previous weeks predict changes in volatility, using a sample of S&P 500 firms.

H3: Da, Engleberg and Gao (2011) support that a rise in individual investor attention forecasts a significant price impact, generating short-term positive abnormal returns in the subsequent week. I test if this result holds in a lagged relationship and if it is also present within the current week. First, if SVI for individual stocks has a truly significant effect on stock returns, and if so, in which direction, constituting a robustness test not only to Da et al. (2011), Joseph et al. (2012), but also to the attention-induced price pressure of Barber and Odean (2008).

H4: If the results presented by Da et al. (2011) turn out to be corroborated by the empirical data, in an attempt to go a step further, I investigate whether retail investors' attention may be related to some well-known anomalies, such as the Size anomaly, the Value anomaly and the Momentum anomaly.

H5: To finalise, I examine if the impact of SVI on individual stocks liquidity, volatility and returns is extended to the stock market at an aggregate level, gauging whether the search volume for the term S&P500 produces changes in the aforementioned market outcomes for the market index itself.

To conduct these analyses, whenever pertinent, I start by tabulating the data with the goal of providing a first insight to stock features and sample descriptive statistics. A formal econometrical analysis follows. To rigorously study the impacts of investors' attention on stocks, I resort to panel regressions with fixed-effects, an extension of the classical linear model, which enables obtaining unbiased estimators, by controlling for unobservable confounding factors that vary across units (stocks) but are constant over time, for instance industry or firm-specific effects. To ensure the validity of inference, the standard errors computed are robust to heteroskedasticity (White standard errors) and also clustered by units, accounting for a possible presence of correlation in the residuals within each firm. I also use Fama-Macbeth panel regressions[7], whose standard errors address the time possible time correlation in the residuals.

A remark concerning the decision rule for the usage of the above mentioned regressions ought to be made. When a firm-effect is present, fixed-effects panel regressions with robust clustered standard errors are implemented. Conversely, when a time-effect is encountered, Fama-Macbeth is the correct choice.[8]

Finally, to predict the variance of returns, whenever possible, GARCH (1,1) model – the standard tool to estimate stock returns volatility, introduced by Bollerslev (1986) – is used. Briefly, the GARCH (1,1) maximum-likelihood (or quasi-maximum likelihood) estimation suggests that the best forecast for the variance of stock returns is a weighted average of the most recent variance registered and the information contained in the most recent squared residual.

---

[7] Created by Fama and Macbeth (1973), this econometrical method, for panel data, determines the risk-premium of each risk factor in a two-step procedure: first, it carries out time-series regressions, regressing the returns of each stock against the proposed risk factor determining the Beta's of each asset in relation to that risk factor; second, it runs cross-sectional regressions having the returns of each stock as dependent variable and the previously calculated Beta's as the independent ones, to finally get the risk-premium for each risk-factor, originally called Gamma by Fama and Macbeth (1973).

[8] Petersen (2005) provides a useful discussion as it comes to which model is the most appropriate to be used in different contexts. Moreover, Petersen (2005) also clarifies that the identification of the form of dependence present in the data is made by a comparison between White standard errors and the standard errors yield by each model. If SE clustered by group are much larger than White SE, there is a firm-effect. If SE from Fama-Macbeth are much larger than White SE, time-effect is present.

V. Empirical Findings

V.1 Data Framing – Stock features and descriptive statistics

In this first subsection, I provide a broad overview of the general characteristics and descriptive statistics of the 215 S&P500 stocks included in the sample. In order to do so, the sample was sorted by date, and within this first arrangement, quintiles were formed in accordance to the attention level paid to the individual stocks (Table 1).

Table 2 contains the statistics for both market events and firm-specific features for the five weekly portfolios formed on the basis of SVI[9].

Starting off with liquidity, it is clearly visible that the portfolio registering the top levels of investors' attention comprises the mostly traded stocks. *H1* seems to be corroborated by this simple first approach, as even though for the other quintiles volume and turnover fluctuate around the same values, it is possible to notice a rising tendency as we move from lower to upper quintiles, suggesting a positive interaction between investors' attention and liquidity.

Along with increased liquidity, it is likely to come more volatility. In fact, this preliminary approach to the data suggests that. Similarly to liquidity, significantly greater volatility registered over the past 30 days is observable on the portfolio receiving the highest investor attention, whereas in the remaining quintiles a smooth upward trend can be identified, which may be related to retail investor attention.

Moving to stock returns, the existence of a pattern is not as obvious as before, yet there seems to be a rising tendency as we move from the lowest to the highest SVI quintiles. However, by looking at the portfolio receiving the largest attention, a sudden break in the tendency is detected, as returns drop sharply compared to the previous portfolio. Basically, this fact points out that the most searched stocks were the ones

---

[9] Table 1 provides both statistics, median and mean. The preliminary conclusions drawn are valid regardless the measure chosen, as they display essentially the same patterns.

performing worst, something that, at first sight, appears to contradict the positive correlation between abnormal levels of SVI and abnormal levels of returns documented by Da et al. (2011) and by Joseph et al. (2012).

Even so, it is interesting to perceive that the 4th quartile encloses the best performers and that, when the impact of outliers is mitigated, there appears to be a positive correlation between SVI and investors' attention up to the 4th quartile, including, which also disputes one main finding from Merton (1987), who supports that stocks enjoying low visibility usually perform slightly better than more visible peers.

Naturally, no rigorous conclusion can be drawn at this stage, however the empirical data may suggest that stocks benefit from visibility up to some point, after which too much attention leads to abnormal trading and, in turn, abnormal trading leads to volatility spikes, which can jeopardise returns. The addition of a higher risk likely prompts a sell-off decision for an average individual investor, risk-averse by nature, generating price falls. It is worth to stress that this simply theorised, and not yet formally proven, relation of retail investor attention, volatility and return, does not concern systematic volatility but instead some source of idiosyncratic volatility, as documented by Ang, Hodrick, Xing and Zhang (2006). In the present situation, this idiosyncratic volatility may be generated by a too large level of individual investors' attention, which ends up penalising stock ownership by reducing stock returns.

After getting a preliminary flavour of what general effects investor attention may have on market outcomes, and before focusing on the individual econometrical analysis of each one, I deepen the analysis on firm-specific characteristics as well as past returns to see if it is possible to establish a potential relation between retail investor attention and some well-known anomalies, such as size, value and momentum.

Table 2 indicates that, roughly speaking, the larger companies are, the more attention is paid to them. As for the Book-to-Market ratio, it is possible to observe that on the two extreme quintiles, the one with the lowest SVI and the one with the highest SVI, we find the most overvalued stocks, i.e. the stocks with the lowest BTM. This overvaluation can be a possible explanation for the most negative returns earned by these two portfolios. Taking the two together, it is possible to extrapolate a possible link with retail investor attention. This data points out that the lowest attention-grabbing firms are the smallest ones, while the mostly searched stocks are the largest. In this scenario, a potential behavioural bias of individual investors can be considered: investors pay greater attention to big capitalisation stocks, as they are more often mentioned in the press and noticing a considerable overvaluation, they tend to get rid of them, generating the negative returns. On the other side, retail investors are not that attentive to smaller firms. Especially, if they appear to be extremely overvalued compared to other public firms, hardly they will consider a purchase. The higher P/E ratio figures for these two quintiles corroborate the noticeable overvaluation.

By looking at the average past return earned in each portfolio, it is visible that the lowest past returns are registered on the highest SVI quintile, which may suggest that retail investors' common behaviour may lead to the persistency of negative returns, through the following mechanism: in the first place investors' attention is caught by extreme negative returns, as such they search them on Google, verifying their overvaluation and bad performance, preceding a sell-off decision that contributes to intensify the negativity of returns.

In sum, even though, through this simple approach, firm-specific characteristics appear to be related retail investor attention, there is no preliminary evidence for any kind of link between SVI and well-known market anomalies.

It is important to reinforce that all the hypotheses described in this subsection have no rigorous proof. They are exclusively grounded on the interpretation of descriptive tables using finance intuition. Throughout the next subsections, a thorough formal analysis will be conducted.

V.2 The effect of retail investor attention on stock's liquidity

In this subsection, I examine the potential link between the intensity of search queries on Google for stock tickers and stocks' liquidity by carrying out a set of analyses using panel regressions with fixed effects, and computing robust standard errors clustered by firm.

Before describing the procedure followed, it is relevant to establish the distinction between trading volume and share turnover ratio, which made me opt for the latter to conduct this investigation. In their essence, both are liquidity measures. However, while trading volume is a raw liquidity metric which simply counts the amount of shares of stock that were bought and sold over a certain time span; share turnover compares that same volume traded to the total number of shares outstanding during the same time period through a ratio. Its use, as opposed to trading volume, enables a more accurate interpretation of stock's liquidity, particularly when there is a large difference between the number of shares outstanding across different companies.

As a matter of fact, and since the results obtained were pretty much the same for both liquidity measures, the focus of this analysis will be on turnover, whose results are shown in Table 3.

To test whether volume of searches in Google correlates with stock turnover, the starting point was to set up a model that aims at explaining turnover through the usual explanatory variables, such as contemporaneous stock price, market capitalisation and

past returns up to the fourth lag. Naturally, transaction costs, such as brokerage commissions, would be a relevant variable to account for, however as the stocks in the sample are stocks from the S&P 500, they should all be liquid enough to consider the bid-ask spread negligible; and brokerage fees are impossible to control as there is no data reporting which broker made the deal.

All control variables turned out to be significant except for contemporaneous stock price, possibly because it corresponds only to the closing stock price of the week, and the R^2 of this model was 10.7%. Following, the contemporaneous SVI in logarithmical form was introduced as explanatory variable and its impact on turnover appears to be positive and statistically significant at 1%, presenting the first piece of evidence for the existence of a true interaction between investors' attention and stocks liquidity. In order to account for non-linearity in this relationship, a quadratic term – logarithmical SVI squared – was added to the equation. Apparently, the solely linear relation between SVI and turnover was overestimating the effect of the former on the latter, a fact demonstrated by the negative coefficient associated with the non-linear SVI term, which is statistically significant at 1%.

Finally, Joseph et al. (2011) document a slightly different relationship between search volume and stock liquidity, measured by trading volume. In their study, they find a dynamic interaction, through which abnormal SVI levels registered in a certain week predict a higher trading volume in the following week. Motivated by their findings, two lagged variables of SVI, both linear and quadratic, were included in the equation:

$$Turnover_{i,t} = \alpha_i + \beta \sum controls_{i,t} + \beta LogSVI_{i,t} + \beta LogSVI_{i,t-1} + \beta LogSVI^2_{i,t} + \beta LogSVI^2_{i,t-1} + e_{i,t} \quad \text{(1)}$$

The above specification, whose results are in Column (6) of Table 3, has an adjusted R^2 of 13.6% and it no longer provides strong evidence for the existence of a

contemporaneous relationship[10]. Instead, it supports that the interaction between retail investor attention and liquidity is dynamic corroborating what was found by Joseph et al. (2012). Furthermore, it is also non-linear and as such the impact of SVI on turnover can be interpreted as follows: SVI indeed leads an increase in turnover, however its impact is decreasing in magnitude. Numerically, *ceteris paribus*, on average, a 1% increase in SVI registered during the current week forecasts a 0.01532% rise in share turnover ratio during the following week.

Once the relationship between retail investor attention levels and liquidity, measured through turnover was established, it is relevant to assess a more specific connection between abnormal levels of attention and abnormal turnover, the only one reported by Joseph et al (2011). The determination of abnormal turnover requires abnormal trading volume, which was computed as in Joseph et al. (2012): the difference between the trading volume of each week and the average weekly volume over the whole sample period. Then, to calculate the abnormal share turnover ratio, it suffices to divide abnormal trading volume by the total amount of shares outstanding. The main variable of interest, abnormal SVI, is computed as described in the data section.

The methodology followed to investigate the existence of an interaction between abnormal SVI and abnormal turnover was the same as before. Firstly, I implemented a model to explain abnormal turnover with the usual control variables and an additional one, BTM, a dummy variable that takes the value 1 when the Book-to-market ratio is equal or below 0.35, a threshold arbitrarily established in accordance with the descriptive statistics presented in the previous subsection; and 0 otherwise. The rationale behind this binary variable is that when the Book-to-Market ratio drops to values that indicate an exaggerated overvaluation, investors are affected by an irrational

---

[10] Only the non-linear term remained significant, however only at a 10% level. Hence, I considered its impact negligible. However, the two contemporaneous variables were not suppressed from the model, because their inclusion seems relevant as the goodness of fit metric points out and also because removing the linear and keeping the quadratic alone does not make much sense. So, in this case, keeping the variables does not constitute a violation of the Principle of Parsimony.

behavioural bias, called disposition effect, extensively explained by Barberis and Xiong (2009), prompting the sale of winners and the holding of losers, related to the fact that investors prefer to recognise gains too early than to realise losses. In this case, I assume that the disposition effect leads to an abnormal trading volume. The coefficient associated to this variable is positive as expected and is statistically significant at 1%.

Once this model was set up, abnormal SVI variables were added to the equation similarly to what was done before. At a first stage, only the linear and contemporaneous abnormal SVI but then the non-linear term as well. Likewise the previous analysis, both turned out to be statistical significant. The results are available in Table 4.

Being consistent with the earlier finding about a lagged interaction between investors' attention and turnover, the 1-week lag of the abnormal SVI was introduced in the regression, yielding the final model:

$$Turnover_{i,t} = \alpha_i + \beta \sum controls_{i,t} + \beta AbnSVI_{i,t} + \beta AbnSVI_{i,t-1} + \beta AbnSVI_{i,t}^2 + \beta AbnSVI_{i,t-1}^2 + e_{i,t} \quad (2)$$

Just as before, this model, whose goodness of fit equals 17.4%, suggests a strong dynamic relationship between abnormal SVI levels and abnormal turnover[11], and no significant contemporaneous relationship. Specifically, a 1% increase in current week abnormal SVI predicts an average rise of 0.0257% in abnormal turnover during the subsequent week.

All in all, conversely to what was initially found in the first step of the analyses, only in a dynamic relationship, retail investor attention, on both normal and extreme levels, captured via volume of searches in Google for stock tickers exercises explanatory power over stock's normal and abnormal liquidity, hereby measured by share turnover ratio and abnormal share turnover ratio, respectively. The bottom line conclusion is that when individual investors search stocks intensely on Google, they are

---

[11] The results obtained for the interaction between abnormal levels of retail investor attention and abnormal turnover are robust across all different constructions of the abnormal SVI variable, described in the data section.

considering an investment decision which, in some cases, will indeed occur in the following week, affecting liquidity expressively. These results are in line with the work of Barber et al. (2009), Barber and Odean (2008) and Seasholes and Wu (2007) as it comes to a significant interaction between individual investors' attention and trading volume. They also triangulate the findings of Joseph et al. (2012), using a more accurate measure of liquidity and resorting to the most appropriate model, fixed-effects with robust clustered standard errors. Furthermore, I show that not only abnormal levels of attention interact with stock liquidity, but also normal levels, i.e. the weekly amount of Google searches *per se*. Still, the impact of the former is greater than the impact of the latter.

V.3 Retail investor attention and historical volatility of stock returns

In the previous subsection, I formally demonstrated that retail investor attention has a true impact stock liquidity. As previously argued, retail investors, who tend to be less informed and less sophisticated, in general underlie their investments decisions on an emotional component, prompted by noisy signals, which they perceive as good opportunities, rather than on a rational exercise weighing risk and firm fundamentals (see Shleifer and Summers (1990)). Therefore, as noise traders, retail investors can affect not only the value of individual securities they trade, but also the whole market (see Barber at al.(2009), Peng and Xiong (2006)). One of the common effects of noise trading corresponds to an increase in stock returns volatility, as posited by Dimpfl and Jank (2012). In line with this, and provided that liquidity also increases with peaks of investor attention, I investigate its impact on the historical volatility of stocks being searched in Google as well as the validity of noise trading theories within this data.

In order to do so, I started by implementing a GARCH (1,1) model, which aimed at explaining volatility using the two usual variables (past variance and past squared residual) as well as an exogenous variable: retail investor attention − SVI. Unluckily, this model yielded a non-stationary process for the variance, meaning that I am facing an explosive GARCH ($\alpha+\beta > 1$) from which nothing can be validly concluded, according to Francq and Zakoian (2008). Table 5 contains the model results.[12]

Therefore, I was forced to use a sub-optimal model to approach the estimation of volatility. I chose the panel regression with fixed-effects and robust clustered standard errors. This model is very likely to contain measurement error, so I recognise it as a limitation of this analysis, however as shown by Islam et al. (2012), who compare various volatility estimates from several models, reaching the conclusion that statistical significance of the coefficients as well as their direction tend to remain unchanged across most of them, including OLS and GARCH.

Nevertheless, care must be taken when linearly estimating volatility. Past volatility cannot be included in the sample, however a proxy for it can. Hence, as a control variable, I include in the model 1-week lagged VIX return. From a wide variety of factors, ranging from geopolitical and systemic issues to market and firm-specific events, which are said to affect volatility of returns; I picked macroeconomic aspects (monthly industrial production growth), interest rates (weekly return of the risk-free asset), market outcomes for each firm (volume traded and past return), firm-specific events (a Bloomberg estimate for 1-year dividend growth), to use as other control variables. The dependent variable is volatility over the last 30 days and the results are available in Table 6.

---

[12] Two reasons may explain the explosive GARCH (1,1). The first is that the variance process itself is non-stationary and the second is that the variables input in the model are non-stationary themselves. If the second turned out to be true most of the inference done so far would be invalid. Hence, to be as much diligent as possible, I submitted stock returns and SVI to a unit-root Augmented Dickey Fuller Test with 10 lags. These two tests did reject the null hypothesis of unit-root. (Returns are believed to be stationary, however when great shocks occur, they may become non-stationary and as the sample comprises the beginning of the financial crisis, it is pertinent to test their stationarity.) Auxiliary Tables 5.1 and 5.2 contain the results of the ADF test.

Keeping consistency with what was done before, the first step of the procedure consisted in setting up a model to explain volatility, containing the aforementioned control variables. This model achieved a R^2 of about 46%, with all variables being statistically significant, except for dividend growth[13]. Following, the contemporaneous logarithmic SVI variable, accounting for the effect of retail investors' attention, was included in the model. This variable as well as its quadratic term appeared to be statistical insignificant, indicating no relation between retail investor attention and volatility. Motivated by the work of Dimpfl and Jank (2012) who find Granger causality between investors' attention and stock returns volatility in a dynamic relationship, lags of the SVI variable were added to the model up to the lag of 4th order, covering approximately the previous 30 days for which volatility was registered. The resultant final equation is as follows:

$$Vol_{30d_{i,t}} = \alpha_i + \beta \sum controls_{i,t} + \beta \sum_{k=0}^{4} LogSVI_{i,t-k} + e_{i,t} \qquad (3)$$

Interestingly, as lags of SVI were added to the equation, the more recent SVI variables (except the contemporaneous) verified a gain in significance, with all lags being statistically significant at 1% confidence level in the final model, summarised in column (7) of Table 6. The correlation between SVI and volatility appears to be clearly positive, for instance when the SVI registered in previous week increases by 1%, there is, on average an increase of 0.0235% in volatility during the following week, holding all other things constant. Furthermore, the irrelevant contemporaneous relation indicates that there is a one week delay between the search in Google and the subsequent investment made by retail investors.

---

[13] As dividend growth is a variable for which a lot of observations are missing and it is also a highly insignificant variable, in accordance with the principle of parsimony it was removed from the model.

These results[14] provide strong evidence that a higher level of retail investor attention leads to increases in volatility, a fact that not only substantiates the preliminary hypothesis presented in the first subsection, but is also consistent with the augment of trading volume driven by a greater intensity of search queries for stock tickers in Google. In fact, the surge in trading volume verified when investors are more attentive to individual stocks appears as the main intermediary of this link between SVI and volatility, making it relevant to revisit the noise trading theme.

In accordance to Barber et al. (2009), retail investors trade impulsively upon noisy signals, disturbing security prices' stability, introducing a higher level of risk, hereby translated into rise in idiosyncratic volatility. Hence, the findings for volatility gathered with the results obtained for liquidity, using SVI as a timely attention measure, support the consequences of noise trading at the individual asset level.

This result suggests that the inclusion of a time-effective measure of investors' attention in models aiming to forecast stock returns volatility can improve their predictive power. Undoubtedly, an augmented estimating potential is always a good thing in itself, but the applicability of this result has its major value during bearish phases, periods not only characterised by high turbulence in the markets, when firms fundamental value is very often camouflaged, but also known as periods of especially high investor attention (see Andrei and Hasler (2012)). Consequently, the relation between retail investors' attention and volatility is believed to be strengthened during downturns, with SVI explanatory power being boosted.

Regarding the work of Vlastakis and Markellos (2012), who extensively study the relation between information demand and stock market volatility in a sample of the 30 largest NASDAQ stocks, this result corroborates their findings using a different

---

[14] The results obtained for abnormal levels of investor attention are somehow in line with these findings for normal levels; nonetheless they can be misleading as they differ in terms of lags significance and coefficient direction for different constructions of the abnormal SVI variable. Therefore, considering them inconclusive, I do not present them.

approach, a different and broader sample and also a different concept for the search terms of interest: stock tickers instead of companies' names.

In addition, Baker and Wurgler (2007) posit that a good proxy for the difficulty of arbitrage of a stock is volatility of returns. Matching this result with Baker and Wurgler (2007) evidence, it is plausible to argue that, since more retail attention leads to higher volatility of stock returns, stocks which are often on the spotlight of retail investors, typically noise traders, are harder to arbitrage than stocks whose visibility is lower.

V.4 Is retail investor attention a predictor of stock returns?

Once the influence of retail investor attention on liquidity and volatility is determined, it interests a lot to assess whether investors' attention, as measured by SVI, has some price impact on the individual stocks level.

While Drake, Roulstone and Thornock (2012) document surges in information demand around the specific event of earnings announcement, leading, on average to higher returns, Da et. al (2011), Joseph et al. (2012) and Zhang et al. (2013) among others, report a broader effect of investors' attention, not specifically related to a particular event: when the volume of searches for stock tickers in Google and other search engines is abnormal in a certain week, on average, in the subsequent week abnormal positive returns are verified. This is where the true robustness test takes form as this is, intuitively, the less obvious effect of SVI.

In that sense, the analysis starts by evaluating whether normal levels of investor attention should be taken into account in asset pricing. At this first stage, I initially implement the Carhart (1997) four-factor model through an inverse Fama-Macbeth regression, following the method of Da et al. (2011) and then it is extended with the

inclusion of the logarithmical SVI as a risk factor. Besides the latter, the independent variables used are the weekly Fama-French factors and the stock returns earned in the previous week, as a proxy for momentum[15], given that there is no weekly momentum factor available in the online research databases.

The replicated four-factor model, whose dependent variable is excess return over the risk-free was fairly successful at explaining the returns of the 215 S&P500 stocks included in the sample, achieving a R^2 of 45%. Following, the contemporaneous logarithmic form of SVI was included in the model as well as the respective quadratic term, but their coefficients turned out to be statistically insignificant. However, the documented interaction between investors' attention and returns is rather dynamic. Thus, the linear and non-linear variables were added to the equation, being both statistically insignificant. Table 7 presents the empirical results.

This first approach that intended to explore the predictive power of SVI on stock returns, from a broad asset pricing perspective, yielded no statistical evidence for a general relation between normal levels of SVI and risk adjusted returns.

Yet, the hypothesis that even though usual levels of SVI do not exercise any explanatory power over excess stock returns, abnormal levels of retail investor attention are capable of leading abnormal returns, ought to be tested. Actually, this is the relation for which strong evidence is found in the aforementioned literature.

To evaluate whether this interaction applies in this sample of stocks during the time period under analysis, again using an inverse Fama-Macbeth regression the four-factor model was set up and abnormal investor attention variables were added to the specification. Being consistent with previous approach, first contemporaneous abnormal

---

[15] Usually momentum takes under consideration a much larger time span, however, in this case, what interests is a very short-term effect on stock price, so I consider that it suffices to account for the past week return. In fact, this variable turned out to be always significant at a 1% confidence level.

SVI, which turned out to be all statistical insignificant and, then, one-week lagged variables reaching the following final model:

$$Return_i - Rf = \alpha + \gamma\beta_{i_{MktRf}} + \gamma\beta_{i_{SMB}} + \gamma\beta_{i_{HML}} + \gamma\beta_{i_{pastret}} + \gamma\beta_{i\,AbnSVI} + \gamma\beta_{i_{AbnSVI^2}} + e_i \quad (4)$$

The output of this model is available in Table 8. Indeed, I find that abnormal levels of retail investor attention in the current week have a statistically and economically significant stock price effect in following week, even after the four-factor risk adjustment. The empirical evidence points out an exclusively linear relationship, which can be interpreted as follows: *ceteris paribus*, a 1% increase in abnormal SVI during a certain week leads to an average 0.0279% increase in the following week stock return, equivalent to a 1.46% rise in annual return.

This finding does not completely corroborate the results obtained by Da et al. (2011), who also find a negative non-linear relationship, suggesting that even though there is a linear augment in price, it decreases in magnitude after a certain point. Nevertheless, it is absolutely in line with the fundamental theory of attention-induced price pressure documented by Barber and Odean (2008), through which retail investors become net buyers of stocks that catch their attention, leading to a temporary increase in stock price, which takes place in the subsequent week.

Yet, when I assess the robustness of the results obtained, by using different metrics of abnormal SVI, either the difference between the current observations to a median over a longer period or to an average, or even to the whole sample median or average, I observe that the statistical significance is altered substantially. Table 9 presents the results of this test, showing that the real existence of a positive and dynamic relation between abnormal levels of investor attention and abnormal returns depends on the way the abnormal SVI variable is constructed. Moreover, establishing a direct comparison to Da et al. (2011), who use the difference between each SVI

observation and the median over the past 8 weeks, in this case there is still a positive linear interaction, though weakly significant: only at a 10% confidence level.

Summing up, first of all it is worth emphasising that the fact that the dynamic relationship between abnormal retail investor attention and abnormal stock returns, does not hold for all the abnormal SVI measures does not mean that it does not exist at all in general terms, it suggests that it may not be as strong and certain as other papers, which use the same measure for investor attention, report. Specifically, regarding the work of Da et al. (2011), I use a similar approach, but a different sample. In my case, the stocks under analysis are included in the main US market index, the S&P 500, an index with which various exchanges worldwide are correlated. Therefore, I would expect to see at least the same results holding in this case, provided that these are salient stocks, the US largest caps and the most meaningful firms, from an economical perspective. Another difference should be highlighted: Da et al. (2011) use DGTW returns, while I use Carhart four-factor model risk-adjusted returns. In what concerns the work of Joseph et al. (2012), I use the same sample, but with significantly less stocks, imposing a much stricter filter in order to eliminate as much noise as possible as well as invalid SVI observations, a fact that narrows down my sample to 215 stocks, and likely, I also use a different method, the more appropriate one according to Petersen (2005). Actually, Joseph et al. (2011) do not specify exactly which type of regressions and standard errors they use, they simply mention "run regressions" and "consistent with heteroskedasticity standard errors".

Finally, Barber et al. (2005) and Barber and Odean (2008) theories about noise trading and investor attention are not refuted or weakened. First, as Seasholes and Wu (2007) argue not all "attention-grabbing events lead to predictable behaviour", so search intensity of stocks on Google may be one of those phenomena lacking a real predictive

power and on top of that, it is not exactly an attention-grabbing event *per se* but a possible, however not necessary, consequence of one. SVI is rather a declaration of attention.

V.5 Retail investor attention and anomalies: A short history of big failures

In this field of special attractiveness to investors, Da et al. (2011) show that the increase in return that occurs in the week succeeding an outbreak in SVI is merely temporary, observing a price reversal after the second week, which is prolonged thereafter until the positive effect is more or less annulated, holding all other things constant. Since under specific circumstances I also found a statistically significant relationship between retail investor attention and stock returns, in order to go a step further, I assess whether investor attention may constitute a possible explanation to some other well-known market anomalies, such as the size anomaly, the value anomaly and momentum anomaly, or if, at least, investor attention somehow relates to firm-specific features and firm-specific events.

This investigation faces some limitations. As it comes to size, the sample includes S&P 500 stocks, thus no small capitalisation firms can be found, however it is still possible to conduct the analysis in relative terms, determining within large caps which are bigger and which are smaller. Regarding value, as shown in the first subsection, the majority of the companies present an overvaluation, hence a markedly distinction between value and growth stocks is impossible. Nevertheless, it is possible to compare within overvalued BTM ratios, identifying which of them are more overvalued or and which are closer to equity's book value. Finally, in the momentum examination, I consider only the previous week return as, investor attention fluctuates rapidly (see Da et al.(2011)) having only a very short-term effect and also due to the

assumption that investors are timely attentive, therefore a declaration of attention, through a web search in Google, occurs soon after the attention-grabbing event.

To gauge a possible relation between retail investor attention, as measured by SVI, and firm-specific characteristics and events, two methods were put in place. The first, a simple intuitive analysis consisted in weekly sorting the stocks according to the aforementioned factors, forming quintiles and deciles portfolios.

This naïve approach shown in Table 10 aimed at identifying possible patterns of SVI and returns across portfolios based on size, value and past returns.

In size portfolios, an upward trend in returns and in SVI seems to be present, however both are not monotonic, registering some significant breaks when deciles are analysed. If this relationship between retail investor attention and market capitalisation would exist, larger caps would perform smaller caps, contradicting any link between retail investor attention and the size anomaly.

Looking at portfolios based on BTM, no pattern is identifiable in SVI, whereas a downward trend is present in returns, as stocks' market prices approach their book equity value. From this analysis, no strong connection can be established between retail investor attention and the value anomaly. Nevertheless, it is observable that the most overvalued stocks, registered both the highest returns and very high SVI levels.

Finally for momentum portfolios, there may be some extreme positive returns leading higher retail investor attention, consistent with Barber and Odean (2008) theory of attention-grabbing events; nonetheless, it appears that the best performers of one week become the worst performers in the following suggesting the presence of the price reversal in this sample, similar to what Da et al. (2011) find for the Russel 3000 stocks.

The second method to assess the existence of any relation between investor attention and firm-specific characteristics and events corresponds to a formal analysis,

using a set of fixed-effect panel regressions and robust standard errors clustered by company, as a firm-effect seems to be present. I study the impact of investor attention on two different dependent variables: Jensen's alpha from the Carhart model and Jensen's alpha risk-adjusted for the market and individual stocks' characteristics.[16] The explanatory variables are: one week lagged abnormal SVI, the factor under analysis and an interaction variable of these two variables.

Table 11 shows that there is no relation between abnormal levels of investor attention and firm's market capitalisation, given the fact that not even a single interaction variable turned out to be significant.

Table 12 reports the results for the value factor. Similarly to what happened with size, no link was found between retail investor attention and the Book-to-Market ratio.

Considering Table 13, despite weakly significant (only at a 10% confidence level) some interaction variables suggest a relation between retail investor attention and past returns. The negative coefficient associated to these variables indicates price reversals rather than persistency in returns after peaks of investor attention.

On the whole, this analysis whose goal was to find possible explanations for the size anomaly, the value anomaly and the momentum anomaly via retail investor attention, ended up being very conclusive: in this sample of stocks, SVI does not relate to any of the mentioned anomalies[17]. In contrast, there is some evidence supporting price reversals after the positive price pressure generated by boosts of search volume for stock tickers in Google.

---

[16] Jensen's alpha from the Carhart model was computed by calculating the difference between the realised return and the one estimated with the betas of the four-factor model from the previous subsection. Jensen's alpha adjusted was estimated by computing the difference between the realised return and the return estimated using betas from a fixed-effect panel regression model with robust and clustered standard errors, whose explanatory variables were: market risk premium, firm's market cap, firm's BTM and firm's past weekly return.

[17] A third method was also employed: weekly portfolios (quintiles) were formed on the basis of each factor under analysis, and for each of them I ran Fama-Macbeth regressions, risk-adjusting excess returns with the four-factor model plus an abnormal SVI variable. The results were in line with the ones reported, no significant relation between investor attention and firm-specific features was found.

V.6 Impact of investors' attention on the aggregate market level

Once the effects of retail investor attention on individual stocks are established, in this final subsection I investigate if any of the three previously studied market outcomes – liquidity, volatility and returns – correlates with retail investor attention at an aggregate market level.

In order to measure investors' attention to the US stock market, volume of web search queries in Google was used. Naturally, the term[18] which was assumed to reflect deliberate paying of attention by retail investors was S&P500, the true name of the market index. The period and the frequency under analysis remain the same: weekly frequency from January 2005 to December 2008.

An important remark ought to be made. Contrarily to what happens with individual stocks, investors do not own and trade the index itself, they can only trade either S&P500 index futures, an ETF replicating the index or invest in S&P500 index funds. Hence, there is no physical match between the security searched and the security traded, a fact that is likely to yield different results from the ones obtained thus far.

Maintaining the previous order, the first market outcome object of study was liquidity. Unlike the approach used for individual stocks, liquidity for S&P500 index is measured directly through volume traded, since there is no data for the total amount of shares outstanding for all the S&P500 stocks allowing the computation of share turnover ratio. Trading volume, in this case, corresponds to all the shares traded for all stocks included in the index, over a weekly time span.

The analysis of market trading volume was conducted using time-series robust regressions and the respective results are in Table 14. The first specification containing only the controls, closing price and previous weekly return, achieved a R^2 of 56%.

---

[18] The ticker for S&P 500 index – SPX – was not included to avoid ambiguities that may result from this search term, provided that there is a multinational company whose name is exactly SPX.

Subsequently, the model was extended with the inclusion of linear and non-linear SVI variables, first the contemporaneous one and after that the lagged effect. I consider the results found as misleading for three reasons. First, none of the coefficients' significance seems stable as more variables are added to the model; second, the significance of the linear variables is always inferior to 10%; and third, weirdly, not only the adjusted goodness of fit diminishes, penalising the inclusion of irrelevant variables, but also the normal R^2 decreases. Therefore, I conclude that there is no evidence supporting a true interaction between normal levels of investor attention and market index volume traded.

Anyway, keeping consistency with what was previously done I examine a possible relationship between those two realities, at an abnormal level. The procedure followed was the exact same and the results are in Table 15. In this case, there is significant evidence supporting linear, contemporaneous and dynamic, links between abnormal investor attention and abnormal market liquidity. Numerically, a 1% increase in the current abnormal SVI leads to an average 3.17% rise in current abnormal volume traded, *ceteris paribus*; while 1% surge in abnormal SVI in the previous week makes the current week abnormal trading volume grow 2.9%, on average, holding all other things equal. A possible explanation I find for this finding relates to investors' behaviour as follows: individual investors use Google to search the overall performance of the market, after acknowledging it, they look for the best performers, leading to the increase in individual stocks liquidity, which in turn, amplifies the overall liquidity of the market index. I estimated a quick panel regression with fixed-effect and robust clustered standard errors, without any controls, whose results are in line with this theory, displaying a positive and significant, though low, correlation between SVI for individual stock ticker and the S&P500 term (Auxiliary Table 15.1).

Provided that retail investor attention affects the overall market liquidity, I study the existence of a possible relationship between retail investor attention and the volatility of market index returns.

Table 16 contains the results of the GARCH (1,1) model, which accounts for the exogenous SVI variables. I find consistency with the results for liquidity, as the contemporaneous SVI seems to affect volatility, leading to its increase in the subsequent period[19]. This relation is statistically significant at a 5% level. The economic explanation stems from the fact that, as it was showed for individual stocks, augments in liquidity prompted by peaks of investor attention come along with a rise in volatility. In line with this, are noise trading theories already explained throughout the paper.

Accordingly to the evidence provided by Andrei and Hasler (2012) stating that investors are more attentive during bearish phases, times which are also characterised by high volatility shocks; to ensure that the significance of the SVI coefficients is not only due to the crisis period included in the sample, I re-estimate the model excluding all the observations that occurred in the post-Lehman Brothers bankruptcy, on the 15th of September 2008, considered by many as the starting point to the financial crisis. As it is known, the collapse of this financial giant not only prompted a sharp fall in the equity market approximately until March 2009, throwing the S&P500 to values well below 1000 points, but also increased dramatically the volatility of the US stock market. The results for this robustness re-estimation are available in column 11 and 12 of Table 16, and they show that, although slightly smaller, the influence of investor attention on market's volatility persists even after excluding the crisis period.

Another meaningful relation can be studied from an overall market perspective: the correspondence between retail investors' attention and implied volatility.

---

[19] As mentioned in the data section, by using the most recent information, in this case contemporaneous, GARCH models yield a forecast of the variance of returns in the subsequent period. This is the reason why lagged SVI variables are insignificant in this model.

For that purpose, the VIX index[20] is used as dependent variable, similarly to what was done by Vlastakis and Markellos (2012). To test this hypothesis, I implement a model that aims at explaining VIX weekly values via times-series robust regressions. Controlling for the only two possibly important factors (within this dataset) in the context of a forward looking volatility measure – interest rates and industrial production growth – SVI variables were included in the model. Table 17 provides strong evidence for the existence of a negative correlation between retail investors' attention and implied volatility. This counter-intuitive result, for which I find no explanation, only holds when the crisis period is included in the sample, a fact that is even odder, provided that both variables peaked after Lehman Brothers bankruptcy. Excluding the crisis period from the sample, no relevant relations are found between retail investor attention and forward-looking volatility, suggesting that finance professionals do not account for the effects of noise trading in their previsions of market volatility, probably because, under normal circumstances, it has no significant impact on options implied volatility.

As a matter of fact, this paper does not succeed in extending the work of Vlastakis and Markellos (2012), who find a truly significant positive relationship between retail investors' attention and individual stocks' implied volatility, to the aggregate market level.

Finally, moving to the subject of returns, I assess whether the intensity of searches for the term S&P500 has a changing impact on the market index quote.

A set of robust time-series regression were again employed. In this case, there are almost no control variables to explain S&P500 weekly returns, provided that the Fama-French factors apply only to individual firms or portfolios, but not to the index.

---

[20] The CBOE VIX index is a proxy for the following 30-days market volatility, grounded on the basis of market expectations, which in turn are derived mainly, but not exclusively, from supply and demand for call and put options on the S&P500 index. The VIX is a widely used measure of implied volatility which is also considered to be an "investor fear gauge".

Due to the same reason a further analysis of risk-adjusted (abnormal) returns could not be carried out. Nonetheless, it is still possible to control for the effect of momentum, by including a lagged variable of returns[21]. Interestingly, the final model in column 5 of Table 18 suggests that a 1% increase in SVI for S&P500 in a certain week predicts a higher index return in the following week, by an average value of 0.636% points, *ceteris paribus*. The same argument for liquidity holds as a possible reason for the price impact, allowing its polish: finding the best performers, retail investors take long positions on them (consistent with previous findings as well as Barber and Odean (2008) and Da et al. (2011)), generating the price pressure that drives up stock returns and consequently the market index value. Nevertheless, care must be taken when interpreting this result, since there are not many control variables in the model and this effect of SVI may be due to some other factors for which the model does not account.

In a nutshell, at the aggregate level, retail investor attention, as measured by SVI for S&P500, possesses some influence on market events for the market index. It interacts with liquidity, when abnormal levels of both dimensions are considered; in line with this, when investors are more attentive to the market its volatility tend to increase and, last but not least, higher levels of investor attention to the market index also result in a rise in market index quote, on average. Yet, it is relevant to look at these results with a critical eye and do not take them as absolutely certain, provided that few control variables were included in the models, due to the nature of the market index and also due to data limitations. Moreover, the relationships documented occur through indirect mechanisms as there is no physical match between the term searched and the asset traded, the impact on the market index occurs via individual stocks.

---

[21] For the sake of rigour, I also specified that model accounting for past volatility, which is not often priced, however establishes a well-known negative correlation with stock returns. The significance of the lagged SVI variables remained unchanged.

VI. Conclusion

Investing without research is like playing stud poker and never looking at the cards.
Peter Lynch

With the growing internet dominance over the gathering and subsequent spreading of information, financial markets have witnessed an amplified efficiency, while investors enjoy an abundance of information, which they explore in an attempt to become more knowledgeable about the widest variety of firms.

In this paper, I take search intensity for stock tickers in Google as a deliberate manifestation of attention from retail investors, the ones who use such tools to seek financial information, when an investment shows up in their short-term horizon.

Numerous recent studies defend that the search frequency in Google and in other web search engines exerts a significant influence on market outcomes for individual stocks. Having as target the development of a robustness test to see whether the documented results hold in the sample of the most salient and economic representative US stocks, I provide evidence that surges of individual investor attention during the current week predict higher stock liquidity and greater volatility in the upcoming week. As for the price impact, I disclose that its interaction with attention might not be as plain as it has been reported and I assert that a significant effect of the latter on the former depends heavily on what is considered an abnormal SVI level. Throughout these examinations, the conclusions reached are in line with theories of noise trading and investor attention.

Beyond, I assess whether the link of web search volume with returns may relate to firm-specific characteristics, such as market capitalisation and Book-to-Market ratio, or to past returns. I observe that no relationship exists, except for a possible price reversal after the positive returns earned in the week that follows an SVI outbreak.

I also extend the investigation to an aggregate market level, accomplishing the result that volume of searches in Google for the term S&P500 establishes a dynamic positive correlation with the overall market index liquidity, volatility and return.

Taken all together, these findings deliver a broad picture of how retail investor attention, triggering noise trading, may affect the stock market in the short-run, enhancing our understanding of it.

In essence, search frequency in Google appears as a component of behavioural finance possessing a truly relevant impact on short-term market activity. Yet, its possible asymmetric effects as well as the markets on which it exercises some action, or even the relations it establishes with some phenomena, such as news from concrete nature, some specific market events or short-selling, are still lacking clarification and therefore may constitute interesting themes for future research.

VII. Appendix

Table 1 – This table displays the SVI descriptive statistics for the five weekly portfolios formed on the basis of SVI.

SVI Quintiles	# SVI observations	Median SVI	Average SVI	Minimum SVI	Maximum SVI
0	8,926	4.5	3.51	0	6.04
1	8,692	6.67	6.61	6.05	6.90
2	9,099	7.07	7.07	6.91	7.27
3	8,991	7.72	7.80	7.28	8.64
4	8,973	10.59	1.64	8.65	377.50

Table 2 – This table contains the statistics for market events and firm-specific features for five weekly portfolios formed on the basis of SVI.

Median SVI quintiles	Trading Volume	Turnover	Volatility 30 days	Weekly Return$_t$ (%)	Market Capitalization (10^6)	Book-to-Market Ratio	Price-to-Earnings Ratio	Weekly Return$_{t-1}$ (%)
0	10,886,219	.0262	24.805	.06	9864.1553	.3038	17	.11
1	11,184,378	.0277	24.902	.09	1182.1329	.3547	15	.04
2	11,284,496	.0281	24.790	.12	1136.4693	.3661	15	.08
3	10,960,400	.0292	25.060	.12	1201.887	.3295	15	.18
4	19,471,100	.0360	28.216	.05	1883.8795	.3294	16	.07

Average SVI quintiles	Trading Volume	Turnover	Volatility 30 days	Weekly Return$_t$ (%)	Market Capitalization (10^6)	Book-to-Market Ratio	Price-to-Earnings Ratio	Weekly Return$_{t-1}$ (%)
0	21,667,394	.0412	31.068	-.07	21670.343	.3574	23.708	.02
1	22,169,613	.0406	31.615	-.06	26047.091	.4367	18.621	-.15
2	21,345,907	.0430	32.401	-.09	25710.733	.4552	18.266	-.11
3	23,179,164	.0435	31.955	-.04	24362.124	.4065	18.903	.01
4	50,522,672	.0565	36.054	-.32	51415.824	.3928	23.952	-.30

Table 3 – This table reports the results from fixed-effects panel regressions. The dependent variable is share turnover ratio (in decimal form). The independent variables are adjusted closing price and market capitalization (both in log form), logarithmical past returns over the past 4 weeks and linear and quadratic SVI variables in log form, both contemporaneous and lagged. Robust standard errors clustered by firm are presented in parenthesis. $*$, $**$, and$***$ represent significance at the 10%, 5%, and 1% level, respectively.

Dependent variable Explanatory variables	(1) Turnover	(2) Turnover	(3) Turnover	(4) Turnover	(5) Turnover	(6) Turnover
logStock_Price	0.0107	0.00927	0.00944	0.00905	0.00913	0.00877
	(0.00856)	(0.00862)	(0.00869)	(0.00855)	(0.00858)	(0.00864)
logMkt_Cap	-0.0223***	-0.0244***	-0.0245***	-0.0238***	-0.0238***	-0.0245***
	(0.00661)	(0.00662)	(0.00666)	(0.00654)	(0.00654)	(0.00661)
pastret1	-0.0965***	-0.0920***	-0.0920***	-0.0922***	-0.0922***	-0.0914***
	(0.00788)	(0.00792)	(0.00792)	(0.00791)	(0.00790)	(0.00792)
pastret2	-0.0831***	-0.0779***	-0.0779***	-0.0787***	-0.0787***	-0.0783***
	(0.00713)	(0.00736)	(0.00735)	(0.00731)	(0.00731)	(0.00729)
pastret3	-0.0809***	-0.0750***	-0.0749***	-0.0776***	-0.0776***	-0.0758***
	(0.00750)	(0.00776)	(0.00774)	(0.00754)	(0.00752)	(0.00759)
pastret4	-0.0929***	-0.0875***	-0.0874***	-0.0896***	-0.0896***	-0.0872***
	(0.00770)	(0.00803)	(0.00801)	(0.00784)	(0.00782)	(0.00808)
logSVI		0.0150***	0.0108			-0.00805
		(0.00309)	(0.00813)			(0.00945)
logSVI_squared			0.000944			0.00764*
			(0.00237)			(0.00509)
logSVIlag1				0.00887***	0.0698**	0.0230***
				(0.00221)	(0.00314)	(0.00598)
logSVIlag1_sq					-0.00429***	-0.00768***
					(0.00178)	(0.00196)
Constant	0.220***	0.215***	0.220***	0.222***	0.224***	0.217***
	(0.0433)	(0.0432)	(0.0436)	(0.0427)	(0.0424)	(0.0433)
Observations	43,606	41,532	41,532	41,489	41,489	41,108
R-squared	0.107	0.128	0.128	0.121	0.121	0.136
Number of cid	215	215	215	215	215	215
Adjusted R-squared	0.107	0.128	0.128	0.121	0.121	0.136

Table 4 – This table reports the results from fixed-effects panel regressions. The dependent variable is abnormal share turnover ratio (in decimal form). The independent variables are adjusted closing price and market capitalization (both in log form); logarithmical past returns over the past 4 weeks; a dummy variable that takes the value 1 when BTM is below 0.35 and 0 otherwise; and linear and quadratic abnormal SVI variables in log form, both contemporaneous and lagged. Robust standard errors clustered by firm are presented in parenthesis. ∗, ∗∗, and∗∗∗ represent significance at the 10%, 5%, and 1% level, respectively.

	(1) ABN_ Turnover	(2) ABN_ Turnover	(3) ABN_ Turnover	(4) ABN_ Turnover	(5) ABN_ Turnover
Dependent variable Explanatory variables					
logPX_LAST	-0.0157	-0.0159	-0.0161	-0.0163	-0.0166
	(0.0124)	(0.0124)	(0.0123)	(0.0123)	(0.0124)
logCUR_MKT_	0.0212**	0.0212**	0.0216**	0.0216**	0.0210**
	(0.00969)	(0.00965)	(0.00961)	(0.00956)	(0.00965)
pastret1	-0.0841***	-0.0840***	-0.0833***	-0.0832***	-0.0826***
	(0.00795)	(0.00793)	(0.00794)	(0.00793)	(0.00795)
pastret2	-0.0679***	-0.0679***	-0.0681***	-0.0679***	-0.0678***
	(0.00662)	(0.00663)	(0.00660)	(0.00661)	(0.00663)
pastret3	-0.0615***	-0.0614***	-0.0630***	-0.0630***	-0.0624***
	(0.00658)	(0.00658)	(0.00642)	(0.00642)	(0.00648)
pastret4	-0.0741***	-0.0739***	-0.0753***	-0.0752***	-0.0734***
	(0.00698)	(0.00697)	(0.00681)	(0.00681)	(0.00708)
BTM	0.0390***	0.0389***	0.0390***	0.0388***	0.0384***
	(0.00679)	(0.00677)	(0.00677)	(0.00676)	(0.00676)
Abn_SVI	0.0285***	0.0320***			0.00132
	(0.00617)	(0.00500)			(0.00216)
Abn_SVI_squared		-0.00998**			-0.0189*
		(0.00590)			(0.0131)
Abn_SVIlag			0.00472*	0.00929***	0.0445***
			(0.00245)	(0.00234)	(0.00770)
Abn_SVIlag_squared				-0.0128***	-0.0188***
				(0.00409)	(0.00385)
Constant	-0.163***	-0.162***	-0.166***	-0.165***	-0.157***
	(0.0580)	(0.0578)	(0.0575)	(0.0572)	(0.0577)
Observations	41,071	41,071	41,028	41,028	40,647
R-squared	0.167	0.167	0.165	0.166	0.174
Number of cid	213	213	213	213	213
Adjusted R-squared	0.166	0.167	0.165	0.166	0.174

Table 5 – This table presents the results of the estimation of a GARCH (1,1) model, extended by the addition of exogenous SVI variables, contemporaneous and lagged in log-form. The conditional variance estimated corresponds to the conditional variance of logarithmical weekly returns. L.arch and L.garch denote the coefficients for the most recent squared residual and the most recent variance, respectively. Standard errors are in parenthesis. *, **, and*** represent significance at the 10%, 5%, and 1% level, respectively.

The coefficients of Arch and Garch components add up to values higher than 1, thus the variance process is non-stationary.

Dependent Variable Explanatory Variables	(1) RETURN	(2) ARCH	(3) RETURN	(4) ARCH	(5) RETURN	(6) ARCH
L.arch		0.127*** (0.00130)		0.141*** (0.00142)		0.142*** (0.00142)
L.garch		0.894*** (0.00111)		0.881*** (0.00120)		0.881*** (0.00121)
logSVI			-0.000241 (0.000420)			
logSVIlag1					0.000527** (0.000301)	
Constant	0.00198*** (0.000158)	1.57e-05*** (8.12e-07)	0.00247*** (0.000868)	2.06e-05*** (9.69e-07)	0.000949 (0.000879)	2.08e-05*** (9.75e-07)
Observations	44,466	44,466	42,202	42,202	42,152	42,152

Auxiliary Table 5.1 – This table reports the results from an Augmented Dickey-Fuller test made to weekly log-returns. The null hypothesis of unit root and therefore non-stationarity is rejected when p-value is higher than 0.05.

Fisher-type unit-root test for RETURN
Based on augmented Dickey-Fuller tests

Ho: All panels contain unit roots		Number of panels = 215	
Ha: At least one panel is stationary		Avg. number of periods = 206.82	
		ADF regressions: 10 lags	

		Statistic	p-value
Inverse chi-squared(430)	P	3370.9130	0.0000
Inverse normal	Z	-47.7359	0.0000
Inverse logit t(1079)	L*	-63.3129	0.0000
Modified inv. chi-squared	Pm	100.2843	0.0000

Auxiliary Table 5.2 – This table reports the results from an Augmented Dickey-Fuller test made to logarithmical SVI. The null hypothesis of unit root and therefore non-stationarity is rejected when p-value is higher than 0.05.

Fisher-type unit-root test for logSVI
Based on augmented Dickey-Fuller tests

Ho: All panels contain unit roots		Number of panels = 215	
Ha: At least one panel is stationary		Avg. number of periods = 197.04	
		ADF regressions: 10 lags	

		Statistic	p-value
Inverse chi-squared(430)	P	39.311.170	0.0000
Inverse normal	Z	-403.568	0.0000
Inverse logit t(1074)	L*	-705.704	0.0000
Modified inv. chi-squared	Pm	1.193.871	0.0000

Table 6 – This table reports the results from a fixed-effect panel regression. The dependent variable is volatility of individual stock returns over the past 30 days (in decimal form). The independent variables are volume traded in log form; 1-week lag log stock return; 1-week lag log-return for the VIX index; quarterly industrial production growth; the return on the risk-free asset; a Bloomberg estimate of the 1-year dividend growth, and five SVI variables accounting for the search intensity registered over the current week and the previous four. Robust standard errors clustered by firm are presented in parenthesis. $*$, $**$, and $***$ represent significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable	(1) VOL_30d	(2) VOL_30d	(3) VOL_30d	(4) VOL_30d	(5) VOL_30d	(6) VOL_30d	(7) VOL_30d
Explanatory Variables							
logVolume_traded	0.190***	0.179***	0.179***	0.180***	0.179***	0.179***	0.178***
	(0.0160)	(0.0138)	(0.0138)	(0.0139)	(0.0140)	(0.0141)	(0.0141)
pastret1	-0.219***	-0.196***	-0.196***	-0.195***	-0.199***	-0.197***	-0.199***
	(0.0631)	(0.0478)	(0.0475)	(0.0474)	(0.0485)	(0.0492)	(0.0495)
lagVIX_Return	-0.0822***	-0.0880***	-0.0881***	-0.0863***	-0.0873***	-0.0847***	-0.0884***
	(0.00715)	(0.00641)	(0.00635)	(0.00657)	(0.00661)	(0.00669)	(0.00680)
RF	-3.115***	-3.229***	-3.226***	-3.232***	-3.234***	-3.265***	-3.311***
	(0.106)	(0.0949)	(0.0951)	(0.0956)	(0.0961)	(0.0974)	(0.0985)
INDUSTRI	-1.985***	-1.954***	-1.953***	-1.926***	-1.919***	-1.852***	-1.751***
	(0.206)	(0.179)	(0.179)	(0.179)	(0.179)	(0.178)	(0.177)
DIV_GR_1Y	9.29e-08						
	(1.38e-06)						
logSVI		0.0262***	-0.00610	-0.00205	-0.00297	-0.00386	-0.00444
		(0.0117)	(0.0202)	(0.00908)	(0.00890)	(0.00890)	(0.00908)
logSVI_squared			0.00505				
			(0.00611)				
logSVIlag1				0.0530***	0.0255***	0.0265***	0.0233***
				(0.00981)	(0.00715)	(0.00690)	(0.00678)
logSVIlag2					0.0489***	0.0303***	0.0334***
					(0.00801)	(0.00705)	(0.00715)
logSVIlag3						0.0341***	0.0164***
						(0.00572)	(0.00526)
logSVIlag4							0.0276***
							(0.00528)
Constant	-2.569***	-2.403***	-2.379***	-2.443***	-2.466***	-2.468***	-2.465***
	(0.260)	(0.228)	(0.223)	(0.234)	(0.238)	(0.242)	(0.244)
Observations	34,113	42,036	42,036	41,594	41,339	40,985	40,657
R-squared	0.465	0.450	0.450	0.453	0.454	0.456	0.459
Number of cid	173	215	215	215	215	215	215
Adjusted R-squared	0.464	0.450	0.450	0.453	0.454	0.456	0.458

Table 7 – This table contains the results from inverse Fama-Macbeth regressions. The dependent variable is weekly excess return over the risk-free (in decimal form). The independent variables are the weekly Fama-French factors; the 1-week lag log-return; and linear and quadratic variables of logarithmical SVI, both contemporaneous and lagged one week. Fama-Macbeth standard errors adjusted for a possible time-effect in the residuals are in parenthesis. $*$, $**$, and $***$ represent significance at the 10%, 5%, and 1% level, respectively. (Fama-Macbeth regression does not provide Adjusted R-squared.)

Dependent Variable Explanatory Variables	(1) ExReturn	(2) ExReturn	(3) ExReturn	(4) ExReturn	(5) ExReturn	(6) ExReturn
MktRF	1.117***	1.113***	1.111***	1.117***	1.118***	1.111***
	(0.0297)	(0.0298)	(0.0299)	(0.0298)	(0.0298)	(0.0299)
SMB	0.0924**	0.0997**	0.104**	0.0790*	0.0797*	0.0924*
	(0.0454)	(0.0470)	(0.0467)	(0.0465)	(0.0465)	(0.0477)
HML	0.146***	0.147***	0.149***	0.143***	0.143***	0.153***
	(0.0490)	(0.0502)	(0.0498)	(0.0502)	(0.0498)	(0.0503)
pastret1	0.0349***	0.0354***	0.0357***	0.0348***	0.0340***	0.0343***
	(0.0110)	(0.0114)	(0.0114)	(0.0112)	(0.0112)	(0.0117)
logSVI		-0.260	9.756			7.875
		(0.229)	(11.62)			(13.89)
logSVI_squared			-2.476			-2.183
			(2.987)			(3.493)
logSVIlag1				-0.103	-10.19	4.351
				(0.235)	(13.04)	(19.87)
logSVIlag1_squared					2.671	-1.100
					(3.339)	(5.106)
Constant	-0.0599***	0.532	-9.597	0.176	9.751	-11.27
	(0.0225)	(0.455)	(11.31)	(0.461)	(12.74)	(23.44)
Observations	44,220	41,956	41,956	41,956	41,956	40,166
R-squared	0.452	0.461	0.466	0.462	0.465	0.478
Number of groups	214	214	214	214	214	214

Table 8 – This table contains the results from inverse Fama-Macbeth regressions. The dependent variable is weekly excess return over the risk-free (in decimal form). The independent variables are the weekly Fama-French factors; the 1-week lag log-return; and linear and quadratic variables of logarithmical abnormal SVI, both contemporaneous and lagged one week. Fama-Macbeth standard errors adjusted for a possible time-effect in the residuals are in parenthesis. *, **, and*** represent significance at the 10%, 5%, and 1% level, respectively. (Fama-Macbeth regression does not provide Adjusted R-squared.)

	(1)	(2)	(3)	(4)
Dependent Variable	ExReturn	ExReturn	ExReturn	ExReturn
Explanatory Variables				
MktRF	1.118***	1.119***	1.118***	1.118***
	(0.0301)	(0.0301)	(0.0300)	(0.0300)
SMB	0.101**	0.101**	0.0771	0.0759
	(0.0467)	(0.0466)	(0.0468)	(0.0469)
HML	0.145***	0.146***	0.139***	0.135***
	(0.0501)	(0.0501)	(0.0503)	(0.0498)
pastret_1	0.0362***	0.0355***	0.0359***	0.0351***
	(0.0113)	(0.0113)	(0.0113)	(0.0113)
Abn_SVI	-0.00754	-0.00893		
	(0.00892)	(0.00881)		
Abn_SVI_squared		0.164		
		(0.251)		
Abn_SVIlag			0.0193**	0.0279***
			(0.00876)	(0.00979)
Abn_SVIlag_squared				-0.324
				(0.274)
Constant	-0.000664***	-0.000658***	-0.000760***	-0.000656**
	(0.000219)	(0.000219)	(0.000248)	(0.000272)
Observations	41,630	41,630	41,630	41,630
R-squared	0.465	0.468	0.463	0.466
Number of groups	214	214	214	214

Table 9 – This table contains the results from the robustness test made to the impact of abnormal SVI on individual stock returns. This test was performed through inverse Fama-Macbeth regressions. The dependent variable is weekly excess return over the risk-free (in decimal form). The independent variables are the weekly Fama-French factors; the 1-week lag log-return; and linear and quadratic variables of logarithmical abnormal SVI registered in the previous week. The abnormal levels of SVI were computed as the difference between each log-SVI observation and the log-median (m) or the log-average (a) of the SVI registered during the 4, 6 and 8 previous weeks. "variationsinsvi" represents the logarithmical weekly variations in SVI. Fama-Macbeth standard errors adjusted for a possible time-effect in the residuals are in parenthesis. *, **, and*** represent significance at the 10%, 5%, and 1% level, respectively. (Fama-Macbeth regression does not provide Adjusted R-squared.)

Dependent Variable Explanatory Variables	(1) ExReturn	(2) ExReturn	(3) ExReturn	(4) ExReturn	(5) ExReturn	(6) ExReturn	(7) ExReturn	(8) ExReturn
MktRF	1.118***	1.118***	1.117***	1.118***	1.117***	1.119***	1.118***	1.115***
	(0.0300)	(0.0301)	(0.0300)	(0.0301)	(0.0301)	(0.0299)	(0.0298)	(0.0298)
SMB	0.0752	0.0767	0.0794*	0.0781*	0.0803*	0.0823*	0.0797*	0.0741
	(0.0468)	(0.0469)	(0.0468)	(0.0469)	(0.0469)	(0.0466)	(0.0465)	(0.0478)
HML	0.136***	0.136***	0.137***	0.135***	0.137***	0.137***	0.143***	0.133***
	(0.0498)	(0.0500)	(0.0503)	(0.0500)	(0.0503)	(0.0497)	(0.0498)	(0.0504)
pastret_1	0.0347***	0.0359***	0.0372***	0.0362***	0.0375***	0.0336***	0.0340***	0.0367***
	(0.0113)	(0.0113)	(0.0114)	(0.0113)	(0.0114)	(0.0113)	(0.0112)	(0.0114)
lagabn_svi_a4	0.0250**							
	(0.0101)							
lagabn_svi_a4_sq	-0.475							
	(0.365)							
lagabn_svi_a6		0.0154*						
		(0.00863)						
lagabn_svi_a6_sq		-0.431						
		(0.304)						
lagabn_svi_a8			0.0115					
			(0.00861)					
lagabn_svi_a8_sq			-0.245					
			(0.255)					
lagabn_svi_m6				0.0211**				

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
				(0.00907)				
lagabn_svi_m6_sq				-0.357				
				(0.240)				
lagabn_svi_m8					0.0153*			
					(0.00876)			
lagabn_svi_m8_sq					-0.164			
					(0.207)			
lagabn_svi_median						-0.00145		
						(0.00604)		
lagabn_svi_median_sq						0.142		
						(0.178)		
lagabn_svi_average							0.000574	
							(0.00568)	
lagabn_svi_average_sq							0.142	
							(0.177)	
Lagvariationsinsvi								0.00523
								(0.00331)
lagvariationsinsvi_sq								-0.0372
								(0.0310)
Constant	-0.000642**	-0.000618**	-0.000636**	-0.000660**	-0.000706***	-0.000261	-0.000862***	-0.000686**
	(0.000276)	(0.000279)	(0.000276)	(0.000268)	(0.000265)	(0.000616)	(0.000306)	(0.000281)
Observations	41,630	41,294	40,953	41,294	40,953	41,857	41,956	41,506
R-squared	0.466	0.467	0.468	0.467	0.468	0.465	0.465	0.466
Number of groups	214	214	214	214	214	213	214	214

Table 10 – This table documents the naïve approach to the analysis of a relation between firm-specific characteristics (size and value) or events (past return), SVI and returns. This table contains the SVI levels and returns for each of the five and ten weekly portfolios formed on the basis of (1) market capitalization, (2) Book-to-Market ratio and (3) past returns.

Size Quintiles

Market Cap	Median weekly RETURN (%)	Average weekly RETURN (%)	Median SVI	Average SVI
0	-.09	-.63	6.97	7.09
1	.11	-.1	7.06	7.22
2	.11	-.07	7.02	7
3	.16	.07	7.1	7.54
4	.15	.05	7.32	12.59

Value Quintiles

BTM	Median weekly RETURN (%)	Average weekly RETURN (%)	Median SVI	Average SVI
0	.27	.4	7.07	9.02
1	.2	.16	7.09	7.68
2	.11	.04	7.08	7.99
3	.13	-.12	7.09	8.93
4	-.32	-1.10	7.05	7.94

Momentum Quintiles

Past Return	Median weekly RETURN (%)	Average weekly RETURN (%)	Median SVI	Average SVI
0	.28	-.31	7.1	9.13
1	.39	.26	7.06	7.8
2	.18	.09	7.04	7.8
3	-.01	-.09	7.06	8.13
4	-.4	-.65	7.11	8.73

Size Deciles

Market Cap	Median weekly RETURN (%)	Average weekly RETURN (%)	Median SVI	Average SVI
0	-.24	-.93	7.03	7.64
1	0	-.33	6.92	6.55
2	.09	-.1	7.09	7.09
3	.12	-.09	7.03	7.36
4	.08	-.06	6.95	6.87
5	.14	-.08	7.09	7.13
6	.13	.06	7.08	7.63
7	.17	.09	7.14	7.45
8	.23	.07	7.11	8.53
9	.1	.03	8.3	16.66

Value Deciles

BTM	Median weekly RETURN (%)	Average weekly RETURN (%)	Median SVI	Average SVI
0	.32	.41	7.04	9.42
1	.22	.3	7.09	8.62
2	.13	.09	7.15	7.77
3	.26	.22	7.04	7.59
4	.1	.03	7.07	8.05
5	.12	.04	7.09	7.92
6	.18	.02	7.1	8.6
7	.09	-.25	7.09	9.26
8	-.025	-.31	7.05	8.25
9	-.72	-1.89	7.05	7.62

Momentum Deciles

Past Return	Median weekly RETURN (%)	Average weekly RETURN (%)	Median SVI	Average SVI
0	.02	-.78	7.12	9.28
1	.43	.16	7.09	8.99
2	.43	.29	7.06	7.7
3	.35	.24	7.06	7.9
4	.18	.14	7.02	7.79
5	.18	.05	7.06	7.81
6	0	-.03	7.04	8.23
7	-.03	-.15	7.08	8.03
8	-.26	-.26	7.1	8.48
9	-.62	-1.05	7.12	8.99

Table 11 – This table presents the results from fixed-effect panel regressions. The dependent variables are Jensen's Alphas (in decimal form)calculated as the difference between the actual realized returns and the ones estimated by the Carhart model and a model that risk-adjusts stocks for the market and their specific characteristics (market cap, BTM and weekly past returns). The independent variables are market capitalization (in log form); different abnormal SVI variables, comparing each observation to the median over the 4, 6 and 8 weeks; and interaction variables between market capitalization and abnormal SVI. Robust standard errors are in parenthesis. ∗, ∗∗, and∗∗∗ represent significance at the 10%, 5%, and 1% level, respectively.

	(1) Alpha_ Carhart	(2) Alpha_ Carhart	(3) Alpha_ Carhart	(4) Alpha_StockFeatures	(5) Alpha_StockFeatures	(6) Alpha_StockFeatures
Dependent Variable Expalanatory Variables						
lnMKTcap	0.0159***	0.0162***	0.0164***	0.00223**	0.00217*	0.00216*
	(0.00128)	(0.00129)	(0.00129)	(0.00112)	(0.00113)	(0.00114)
lagabn_svi_m4	-0.0132			0.0681**		
	(0.0064)			(0.0429)		
SVIm4_SIZE	0.00321			-0.00656		
	(0.00790)			(0.00603)		
lagabn_svi_m6		-0.0138			0.0445	
		(0.0574)			(0.0439)	
SVIm6_SIZE		0.00306			-0.00427	
		(0.00682)			(0.00421)	
lagabn_svi_m8			-0.0185			0.0346
			(0.0504)			(0.0406)
SVIm8_SIZE			0.00352			-0.00325
			(0.00594)			(0.00388)
Constant	-0.151***	-0.154***	-0.156***	-0.0198*	-0.0191*	-0.0191*
	(0.0121)	(0.0122)	(0.0122)	(0.0107)	(0.0108)	(0.0109)
Observations	41,661	41,321	40,976	41,203	40,867	40,526
R-squared	0.011	0.011	0.012	0.001	0.000	0.000
Number of cid	215	215	215	215	215	215

Table 12 – This table presents the results from fixed-effect panel regressions. The dependent variables are Jensen's Alphas (in decimal form)calculated as the difference between the actual realized returns and the ones estimated by the Carhart model and a model that risk-adjusts stocks for the market and their specific characteristics (market cap, BTM and weekly past returns). The independent variables are Book-to-Market ratio; different abnormal SVI variables, comparing each observation to the median over the 4, 6 and 8 weeks; and interaction variables between Book-to-Market ratio and abnormal SVI. Robust standard errors are in parenthesis. $*$, $**$, and$***$ represent significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable	(1) Alpha_Carhart	(2) Alpha_Carhart	(3) Alpha_Carhart	(4) Alpha_StockFeatures	(5) Alpha_StockFeatures	(6) Alpha_StockFeatures
Expalanatory Variables						
BTM	-0.0239***	-0.0239***	-0.0239***	0.000785	0.000730	0.000746
	(0.00343)	(0.00344)	(0.00343)	(0.00180)	(0.00180)	(0.00180)
lagabn_svi_m4	0.0550*			0.00417		
	(0.0331)			(0.00469)		
SVI_VALUE1	0.0298			-0.000374		
	(0.0339)			(0.00507)		
lagabn_svi_m6		0.0487**			0.00621*	
		(0.0283)			(0.00352)	
SVI_VALUE2		0.0263			0.00242	
		(0.0292)			(0.00361)	
lagabn_svi_m8			0.0463			0.00625
			(0.0432)			(0.00382)
SVI_VALUE3			0.0245			0.00244
			(0.0260)			(0.00367)
Constant	0.00932***	0.00930***	0.00930***	0.00105	0.00115	0.00117
	(0.00144)	(0.00145)	(0.00145)	(0.000755)	(0.000755)	(0.000758)
Observations	41,077	40,741	40,400	41,072	40,736	40,395
R-squared	0.020	0.019	0.019	0.000	0.000	0.000
Number of cid	215	215	215	215	215	215

Table 13 – This table presents the results from fixed-effect panel regressions. The dependent variables are Jensen's Alphas (in decimal form)calculated as the difference between the actual realized returns and the ones estimated by the Carhart model and a model that risk-adjusts stocks for the market and their specific characteristics (market cap, BTM and weekly past returns). The independent variables are 1-week lagged log-returns; different abnormal SVI variables, comparing each observation to the median over the 4, 6 and 8 weeks; and interaction variables between 1-week lagged log-returns and abnormal SVI. Robust standard errors are in parenthesis. $*$, $**$, and$***$ represent significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Alpha_Carhart	Alpha_Carhart	Alpha_Carhart	Alpha_StockFeatures	Alpha_StockFeatures	Alpha_StockFeatures
Expalanatory Variables						
pastret1	-0.0172	-0.0166	-0.0164	0.0203**	0.0206**	0.0201**
	(0.0141)	(0.0148)	(0.0158)	(0.00908)	(0.00921)	(0.00924)
lagabn_svi_m4	0.0139**			0.00477		
	(0.00644)			(0.00547)		
SVI_MOM1	-1.458*			-0.00325		
	(0.805)			(0.0174)		
lagabn_svi_m6		0.0137**			0.00301	
		(0.00629)			(0.00378)	
SVI_MOM2		-1.409*			-0.0186	
		(0.773)			(0.0164)	
lagabn_svi_m8			0.0139**			0.00297
			(0.00597)			(0.00352)
SVI_MOM3			-1.344*			-0.00614
			(0.768)			(0.0142)
Constant	-0.000834***	-0.000864***	-0.000874***	0.00142***	0.00149***	0.00152***
	(6.33e-05)	(5.74e-05)	(6.19e-05)	(2.64e-05)	(2.59e-05)	(2.82e-05)
Observations	41,661	41,321	40,976	41,203	40,867	40,526
R-squared	0.074	0.075	0.071	0.001	0.001	0.001
Number of cid	215	215	215	215	215	215

Table 14 – This table documents the results for time-series robust regressions. The dependent variable is logarithmical weekly trading volume. The independent variables are weekly closing index quote (in log form); the log-return obtained by the S&P 500 market index in the previous week; and linear and quadratic log-SVI variables, both contemporaneous and 1-week lagged. Robust standard errors are in parenthesis. ∗, ∗∗, and∗∗∗ represent significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable Explanatory Variables	(1) logVolume	(2) logVolume	(3) logVolume	(4) logVolume	(5) logVolume
logSP_closePrice	2.178*** (0.148)	0.809*** (0.216)	2.108*** (0.175)	1.954*** (0.185)	2.007*** (0.197)
pastret1	-6.166*** (0.672)	-7.616*** (0.879)	-5.122*** (0.733)	-5.204*** (0.741)	-3.844*** (0.788)
logSVI		-0.656 (0.587)	-7.97* (4.570)	-15.11* (8.591)	-10.96* (6.175)
logSVI_sq			5.442** (3.410)	5.258 (3.414)	3.800 (2.625)
logSVIlag				-0.581** (0.275)	-7.36* (4.349)
logSVIlag_sq					7.20* (4.221)
Constant	6.022*** (1.058)	16.97*** (1.804)	21.32*** (3.594)	22.47*** (3.633)	44.49*** (4.248)
Observations	207	207	207	207	207
R-squared	0.560	0.311	0.546	0.554	0.497
Adjusted R-squared	0.556	0.301	0.537	0.543	0.482

Table 15 – This table documents the results for time-series robust regressions. The dependent variable is logarithmical weekly abnormal trading volume. The independent variables are weekly closing index quote (in log form); the log-return obtained by the S&P 500 market index in the previous week; and linear and quadratic abnormal SVI variables, both contemporaneous and 1-week lagged. Robust standard errors are in parenthesis. *, **, and*** represent significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable Explanatory Variables	(1) Log Abn_Volume	(2) Log Abn_Volume	(3) Log Abn_Volume	(4) Log Abn_Volume
logSP_closePrice	-3.511***	-3.733***	-3.653***	-3.825***
	(0.538)	(0.546)	(0.553)	(0.573)
pastret1	-2.456	-3.295	-3.244	-3.204
	(2.496)	(2.507)	(2.528)	(2.542)
Abn_SVI	2.998**	4.041***	3.338**	3.168**
	(1.292)	(1.381)	(1.492)	(1.500)
Abn_SVI_sq		-17.68*	-16.49	-16.03
		(10.49)	(10.58)	(10.65)
lagAbn_SVI			1.633	2.891**
			(1.348)	(1.470)
lagAbn_SVI_sq				-13.78
				(11.65)
Constant	45.98***	47.66***	47.06***	48.35***
	(3.869)	(3.938)	(3.993)	(4.147)
Observations	104	104	104	104
R-squared	0.357	0.379	0.382	0.392
Adjusted R-squared	0.337	0.354	0.350	0.354

Auxiliary Table 15.1 – This table presents the results for a fixed-effect panel regression. The dependent variable is the weekly SVI for each stock ticker included in the sample. The independent variable is the weekly SVI for the term S&P500. Robust standard errors are presented in parenthesis. *, **, and*** represent significance at the 10%, 5%, and 1% level, respectively.

Dependent Variable Explanatory Variables	(1) SVI_stocks
SVI_sp500	0.0104**
	(0.00436)
Constant	8.734***
	(0.185)
Observations	44,681
Number of cid	215
R-squared	0.002

Table 16 - This table presents the results of the estimation of a GARCH (1,1) model, extended by the addition of exogenous SVI variables, contemporaneous and lagged in log-form. The conditional variance estimated corresponds to the conditional variance of logarithmical weekly returns earned by the S&P 500 index. L.arch and L.garch denote the coefficients for the most recent squared residual and the most recent variance, respectively. Standard errors are in parenthesis. *, **, and*** represent significance at the 10%, 5%, and 1% level, respectively.

The signal # indicates the models that were estimated excluding from the sample the period after Lehman Brothers bankruptcy, therefore, in these cases, the last observation included occurred on 12/09/2008.

Dependent Variable Explanatory Variables	(1) SP_Return	(2) ARCH	(3) SP_Return	(4) ARCH	(5) SP_Return	(6) ARCH	(7) SP_Return	(8) ARCH	(9) SP_Return	(10) ARCH	(11)# SP_Return	(12)# ARCH
L.arch		0.339***		0.342***		0.354***		0.359***		0.353***		0.268***
		(0.0587)		(0.0575)		(0.0647)		(0.0739)		(0.0785)		(0.0834)
L.garch		0.581***		0.570***		0.570***		0.565***		0.570***		0.710***
		(0.115)		(0.119)		(0.114)		(0.118)		(0.121)		(0.127)
logSVI			0.0138		0.825**		0.842**		0.965**		0.712**	
			(0.0139)		(0.375)		(0.408)		(0.486)		(0.378)	
logSVI_sq					-0.251**		-0.256**		-0.294**		-0.232*	
					(0.117)		(0.126)		(0.150)		(0.166)	
logSVIlag							-0.00226		-0.355		-0.425	
							(0.0164)		(0.537)		(0.756)	
logSVIlag_sq									0.109		0.203	
									(0.166)		(0.381)	
Constant	0.00132	6.19e-05*	-0.0210	6.45e-05*	-0.674**	5.96e-05*	-0.685**	6.06e-05*	-0.499	5.95e-05*	-0.987	6.38e-05*
	(0.00138)	(3.50e-05)	(0.0224)	(3.67e-05)	(0.300)	(3.19e-05)	(0.319)	(3.26e-05)	(0.390)	(3.25e-05)	(0.677)	(3.64e-05)
Observations	208	208	208	208	208	208	207	207	207	207	207	207

Table 17 – This table presents the results from time-series robust regressions. The dependent variable is the weekly value of the VIX index (in log form). The independent variable are the return earned by the risk-free asset, quarterly industrial production growth; and linear and quadratic logarithmical SVI variables, both contemporaneous and1-week lagged. Robust standard errors are in parenthesis. ∗, ∗∗, and∗∗∗ represent significance at the 10%, 5%, and 1% level, respectively.

The signal # indicates the models that were estimated excluding from the sample the period after Lehman Brothers bankruptcy, therefore, in these cases, the last observation included occurred on 12/09/2008.

Dependent Variable Explanatory Variables	(1) Log VIX_closePrice	(2) Log VIX_closePrice	(3) Log VIX_closePrice	(4) Log VIX_closePrice	(5) Log VIX_closePrice	(6)# Log VIX_closePrice
RF	-5.874***	-5.846***	-6.701***	-6.873***	-6.565***	-6.045***
	(0.843)	(0.894)	(1.013)	(1.032)	(1.042)	(0.852)
INDUSTRI	-14.67***	-13.76***	-10.87***	-10.45***	-9.928***	-3.534
	(2.580)	(2.654)	(2.908)	(2.920)	(2.911)	(3.005)
logSVI		-0.471*	-22.57***	-21.71***	-14.88**	5.146
		(0.253)	(5.447)	(5.442)	(5.900)	(6.898)
logSVI_sq			6.992***	6.760***	4.611**	-1.770
			(1.671)	(1.666)	(1.808)	(2.145)
logSVIlag				-0.220	-21.04***	-3.150
				(0.338)	(5.932)	(6.892)
logSVIlag_sq					6.508***	0.700
					(1.818)	(2.145)
Constant	3.184***	3.934***	21.43***	21.01***	32.19***	2.720
	(0.0659)	(0.441)	(4.420)	(4.405)	(5.272)	(6.647)
Observations	208	208	208	207	207	192
R-squared	0.393	0.392	0.441	0.443	0.484	0.346
Adjusted R-squared	0.387	0.383	0.430	0.429	0.468	0.325

Table 18 – This table documents the results for time-series robust regressions. The dependent variable is logarithmical weekly returns (in decimal form) earned by the S&P 500 index. The independent variables are the log-return obtained by the S&P 500 market index in the previous week; and the linear and quadratic abnormal SVI variables, both contemporaneous and 1-week lagged. Robust standard errors are in parenthesis. *, **, and*** represent significance at the 10%, 5%, and 1% level, respectively.

	(1)	(2)	(3)	(4)	(5)
Dependent Variable	SP_Return	SP_Return	SP_Return	SP_Return	SP_Return
Explanatory Variables					
pastret1	-0.224***	-0.226***	-0.223***	-0.276***	-0.268***
	(0.0467)	(0.0473)	(0.0544)	(0.0503)	(0.0513)
logSVI		-0.00314	1.160***	0.513*	0.192
		(0.0136)	(0.358)	(0.297)	(0.341)
logSVI_sq			-0.364***	-0.164*	-0.0642
			(0.111)	(0.0914)	(0.105)
logSVIlag				0.0238	0.913***
				(0.0177)	(0.317)
logSVIlag_sq					-0.277***
					(0.0972)
Constant	0.00107	0.00611	-0.920***	-0.437*	-0.892***
	(0.00125)	(0.0221)	(0.290)	(0.243)	(0.275)
Observations	206	206	205	206	206
R-squared	0.101	0.102	0.123	0.133	0.171
Adjusted R-squared	0.0970	0.0928	0.110	0.115	0.151

VIII. References

a) Articles

Andrei, Daniel, and Michael Hasler, 2012. "Investor's attention and stock market volatility." *SSRN eLibrary.*

Ang, Andrew, Robert Hodrick, Yuhang Xing, and Xiaoyan Zhang, 2006. "The cross-section of volatility and expected returns." *Journal of Finance* 61, 259 – 299.

Baker, Malcom, and Jeffrey Wurgler, 2007. "Investor sentiment in the stock market." *The Journal of Economic Perspectives* 21 (2): 129–151.

Barber, Brad, and Terrance Odean, 2008. "All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors." *Review of Financial Studies* 21 (2): 785–818.

Barber, Brad, Terrance Odean, and Ning Zhu, 2009a. "Do retail trades move markets?" *Review of Financial Studies* 21: 151–186.

Barber, Brad, Terrance Odean, and Ning Zhu, 2009b. "Systematic noise." *Journal of Financial Markets* 12: 547–569.

Barberis, Nicholas, and Wei Xiong, 2009. "What drives the disposition effect? An analysis of a long-standing preference-based explanation." *The Journal of Finance* 64 (2): 751–784.

Carhart, Mark M., 1997. "On the persistence in mutual fund performance." *Journal of Finance* 52: 57–82.

Choi, Hyunyoung, and Hal R. Varian, 2009. "Predicting the present with Google trends." *Google Research.*

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011a. "In search of attention." *The Journal of Finance* 66 (5): 1461–1499.

Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011b. "The sum of all fears: investor sentiment and asset prices." *Working Paper, University of Notre Dame and University of North Carolina at Chapel Hill.*

Daniel, Kent D., Mark Grinblatt, Sheridan Titman, and Russ Wermers, 1997. "Measuring mutual fund performance with characteristic-based benchmarks." *Journal of Finance* 52: 1035–1058.

Dimpfl, Thomas, and Stephan Jank, 2011. "Can internet search queries help to predict stock market volatility?" *SSRN eLibrary.*

Drake, Michael, Darren Roulstone, and Jacob Thornock, 2012. "Investor Information Demand: Evidence from Google Searches Around Earnings Announcements."*Journal of Accounting Research*, Forthcoming.

Engle, Robert, 2001. "GARCH 101: The use of ARCH/GARCH Models in Applied Econometrics." Journal of Economic Perspectives 15 (4): 157-168.

Fama, Eugene F., and Kenneth French, 1993. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* 33: 3–56.

Fama, Eugene F., and James D. MacBeth, 1973. "Risk return and equilibrium: Empirical tests". *Journal of Political Economy* 81, 607–636.

Francq, Christian, and Jean-Michel Zakoian, 2008. "Can one really estimate nonstationary GARCH models?" Unpublished.

Islam, Masudul, Lasker E. Ali, and Nahida Afroz, 2012. "Forecasting volatility of Dhaka stock exchange: linear vs non-linear models." *Journal of Science and Engineering,* 3 (2): 4 – 8.

Joseph, Kissan, M. Babajide Wintoki, and Zelin Zhang, 2011. "Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search." *International Journal of Forecasting* 27 (4), 1116–1127.

Li., Jun, and Jianfeng Yu, 2012. "Investor Attention, Psychological Anchors, and Stock Return Predictability." *Journal of Financial Economics* 104: 401-419.

Lo, Andrew W., and Jiang Wang, 2009. "Stock market trading volume." *The Handbook of Financial Econometrics*, Forthcoming

Merton, Robert C., 1987. "A simple model of capital market equilibrium with incomplete information." *The Journal of Finance* 42 (3), 483–510.

Peng, Lin, and Wei Xiong, 2006. "Investor attention, overconfidence and category learning." *Journal of Financial Economics 80* (3), 563 – 602.

Petersen, Mitchell A., 2005. "Estimating standard errors in finance panel data sets: comparing approaches." Working Paper 11280. *National Bureau of Economic Research.*

Seasholes, Mark S., and Guojun Wu, 2007. "Predictable behavior, profits, and attention." *Journal of Empirical Finance* 14 (5): 590–610.

Shleifer, Andrei, and Lawrence Summers, 1990. "The Noise Trader Approach to Finance." *Journal of Economic Perspectives,* 4, 19 – 33.

Vlastakis, Nikolaos, and Raphael Markellos, 2012. "Information demand and stock market volatility."*Journal of Banking & Finance* 36 (6): 1808–1821.

Zhang, Wei, Dehua Shen, Yongjie Zhang, and Xiong Xiong, 2013. "Open source of information,investor attention, and asset pricing." *Journal of Economic Modelling* 33: 613 – 619.

b) Books

Wooldridge, Jeffrey M., 2009. *Introductory Econometrics: A Modern Approach,* fourth edition. Cincinnati, OH: South-Western College Publishing, 2009