WP 2020/4

Christophe Desagre and Catherine D’Hondt

Googlization and retail investors’ trading activity

LOUVAIN FINANCE

LFIN

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/lfin/working-papers.html

Googlization and retail investors' trading activity*

Christophe Desagre† Catherine D'Hondt‡

November 7, 2019

Abstract

Building on Barber and Odean (2008), a growing body of papers document a positive relationship between Google Search Volume Index (SVI) and equity market returns. Such findings suggest that increased attention is combined with a buying pressure that subsequently results in positive returns. This relationship has been established at the market level. In this paper, we focus on a sample of retail investors and use SVI to test whether their aggregate (signed) trading activity is related to attention as well. We find that the relationship between SVI and our retail investors' trading activity is positive, even when controlling for some socio-demographics or subjective investor characteristics. However, this relationship is not stronger for purchases than for sales, thereby providing no support for the buying pressure hypothesis. We also document a bi-directional causality between attention and trading activity, although the contemporaneous effects are economically stronger and predominate. Our results are robust to different measures of attention and trading activity.

Keywords: Investor attention ; Google SVI ; Retail investors

JEL Classification: D83, G11, G40

---

*We are grateful to the online brokerage house for providing the data. We also thank Gael Imad'Eddine, Alexander Kupfer, Werner De Bondt, Ivan Stetsyuk, Ruben Cox, Mikael Petitjean, Paolo Mazza, Yue Zhang, Rudy De Winne, and participants at 3L workshop (November 2017 - Brussels), 35^{th} AFFI Conference (May 2018 - Paris), 16^{th} Belgian Financial Research Forum (June 2018 - Brussels), Warsaw International Economic Meeting (July 2018 - Warsaw), and 3^{rd} European Capital Market Workshop (July 2019 - Dublin) for helpful comments. Any errors are the full responsibility of the authors.

†UCLouvain, Louvain School of Management & Louvain Finance (LIDAM), 151 Chaussée de Binche, 7000 Mons, Belgium, Email : christophe.desagre@uclouvain.be. Corresponding author.

‡UCLouvain, Louvain School of Management & Louvain Finance (LIDAM), 151 Chaussée de Binche, 7000 Mons, Belgium, Email : catherine.dhondt@uclouvain.be

1 Introduction

Most traditional valuation methods determine the value of a firm as the net present value of its expected future cash flows. However, a great deal of information may impact stock prices.[1] On the one hand, listed firms are forced to provide investors with information and, in turn, benefit from media coverage. On the other hand, investors can look for stock-specific or market-related information. On that aspect, Internet has dramatically facilitated access to all sources of information (e.g. newspapers, blogs, media coverage, analysts' reports, macroeconomic or political announcements, social networks, advertisements, etc.). Theoretically, more available information is valuable for investors, but only if they are able to make relevant analysis of it (Barber and Odean, 2001b). In practice, investors often cope with information overload, which is misleading and has non-trivial effects on trading behavior (Choi et al., 2002; Peress, 2014). Information overload is especially acute for retail investors because they have limited attention (Merton, 1987; Sims, 2003; Hirshleifer and Teoh, 2003; Peng and Xiong, 2006). Building on that feature, Barber and Odean (2008) bring empirical evidence that retail investors are net buyers of attention-grabbing stocks. Accordingly, an increase in retail investor attention results in temporary positive price pressure.

Measuring the level of investor attention is an empirical challenge. Several indirect measures are used in the literature such as news and headlines, abnormal trading volume, extreme returns, price limits, etc. News and media-related proxies (e.g. length of headlines, positive and negative tone of an article, analysts' recommendations and coverage, etc.) reflect information supply but not demand, since they do not allow to identify how many investors actually read the news.[2] As far as market-based measures are concerned, they allow to infer the reach and impact of news on the market. Significant news affect investors heterogeneously but result in more investors trading than usually. Relating abnormal volume to higher investor attention is nearly tautological (Barber and Odean, 2008). However, abnormal market volume or extreme returns can also be driven by liquidity-motivated large trades of institutional investors, which can bring some noise to market-based measures of investor attention.

[1] Thompson et al. (1987, p. 267) outline that *'the financial community appears to receive news from a variety of sources, and no single source contains a complete set of firm-specific news.'* This quote remains even true today. Moreover, individual investors are not sufficiently knowledgeable to use traditional valuation methods (Shiller et al., 1984; Nofsinger, 2005).

[2] In other words, front-page news deliver at best a measure of 'passive' attention. Ding and Hou (2015, p. 12) also describe these news as *'sporadic'*. As Da et al. (2011, p. 1462) note, *'a news article in the Wall Street Journal does not guarantee attention unless investors actually read it'*. This is a common criticism against the use of media-related proxies.

Alternatives to measure investor attention have recently emerged with the pervasive use of Internet, and, in particular, the growing availability of data related to online search queries. The best example is the so-called Google Search Volume Index (SVI hereafter).[3] In comparison with the aforementioned measures, SVI allows to directly capture the aggregate 'active' demand for information, either on a per-stock or per-index basis. In addition, SVI is not measured through financial market data. Vlastakis and Markellos (2012) report a correlation close to 15% between SVI and the number of headlines within the Thomson Reuters NewsScope Archive, thereby suggesting that both proxies cover different aspects of information.

Da et al. (2011) are among the first authors to use SVI as a direct proxy for investor attention. Based on their empirical findings, Da et al. (2011, p. 1497) conclude that *'SVI captures the attention of individual investors'* and that *'search volume is an objective way to reveal and quantify the interests of investors and therefore should have many other potential applications in finance.'* In the same vein, Bank et al. (2011, p. 239) note that *'search volume is indeed a powerful measure of investor recognition.'* These authors paved the way for a lot of empirical investigations using SVI to measure investor attention and the present work is part of them.

In this paper, we use SVI to address three research questions. Firstly, we focus on a sample of Belgian retail investors to investigate the relationship between SVI, which is then restricted to Belgium,[4] and retail trading activity. While previous work relies exclusively on market data (even when focusing on stocks that are traded more by retail investors), we directly analyze trades executed by retail investors. We are therefore able to investigate whether SVI helps explain trading activity for sub-samples of retail investors determined upon either socio-demographic characteristics that are common control variables (e.g. age, gender, education, and spoken language) or subjective characteristics that could affect trading behavior (e.g. financial literacy and risk aversion). Secondly, we are also able to determine whether attention is more related to purchases than sales, and thereby able to test the buying pressure hypothesis. Based on the latter, the relationship between attention and trading is expected to be stronger for purchases. Thirdly, we check whether we observe some causal relationships between attention and retail trading activity. Compared to the other empirical studies, we have two advantages as we are able to distinguish purchases from sales, and our results are not biased by any institutional trading.

Our main findings can be summarized as follows. We find that the relationship between SVI

[3] Google is by far the most popular search engine in the world, but other search engines data have also been used. For example, Ying et al. (2015) and Zhang et al. (2013) use Baidu data to analyze Chinese financial markets.

[4] The resulting SVI is based on all the queries sent from Belgium only.

and our retail investors' trading activity is positive, even when controlling for some socio-demographics or subjective investor characteristics. Such findings confirm that SVI is a reliable proxy for retail investor attention. However, our results do not bring evidence that this relationship is stronger for purchases than for sales, thereby providing no support for the buying pressure hypothesis. Finally, we document a bi-directional causality between attention and retail trading activity, although the contemporaneous effects are economically stronger and predominate in both cases. Our results are robust to several measures of attention and trading activity.

The remainder of this paper is organized as follows. In Section 2, we review the relevant literature. In Section 3, we present our data and sample. Section 4 reports our empirical results. We conduct some robustness checks in Section 5. Section 6 concludes.

2 Literature review

As mentioned earlier, Da et al. (2011) pioneered the use of SVI as a direct measure of investor attention. Their paper provides three key results. Focusing on all Russell 3000 stocks, these authors report that time-series correlations between the SVI based on stock tickers and usual indirect measures of attention (such as extreme returns, turnover, and news) are positive on average, although the level of correlation is low. Second, using trading volumes executed on different execution venues and the Dash-5 reports to infer monthly changes in orders and turnover from retail investors, Da et al. (2011) bring evidence that SVI rather captures the attention of retail investors. In particular, their findings suggest that SVI likely captures the attention of less sophisticated retail investors. Third, they document that shocks to investor attention predict higher stock prices in the short term and price reversals in the long run. Consistent with the buying pressure hypothesis and Barber and Odean (2008)'s framework, these results are stronger among stocks in which retail investor attention matters the most. This paper is the first of a growing list of empirical investigations based on SVI to measure investor attention. Depending on the data at hand, these empirical studies relate SVI to stock market returns, volatility, liquidity, and/or trading volume.[5] We review the main studies hereafter.

[5]Other variables have been investigated but to a lesser extent. For example, Ding and Hou (2015) focus on a sample of S&P500 stocks and show that SVI explains the breadth of ownership, i.e. firms that attract higher investor attention have more shareholders. Lately, SVI has been used to explain other assets' returns and characteristics: Goddard et al. (2015) analyze the effect of investor attention in the currency markets; Panagiotidis et al. (2019) use SVI to explain bitcoin returns; Li et al. (2015) analyze trader positions in commodity markets. We restrict our literature review to studies about stock markets.

In the niche devoted to variables that explain and/or predict returns, an increasing body of research finds that social media may be helpful. According to Bank et al. (2011), a portfolio long on the most-searched stocks and short on the least-searched stocks provides annual abnormal returns, even after controlling for well-known risk factors. Joseph et al. (2011) also provide evidence that a trading strategy using SVI is profitable. However, Bijl et al. (2016) find that this trading strategy is no longer profitable once transactions costs are considered. More recently, Heyman et al. (2019) suggest to short stocks that just experienced an increase in search volume as they will be subject to return reversals. Based on their analysis, these authors report that this investment strategy yields an annual interest of 9.1%, net of transaction costs. On the other hand, Takeda and Wakao (2014) find a weak relationship between investor attention and abnormal returns using the constituents of the NIKKEI225 index between 2008 and 2011. Their empirical results weakly support the buying pressure hypothesis, although their findings are stronger for small caps, which is consistent with Bank et al. (2011). Addressing the causality issue, Vozlyublennaia (2014) report that the relationship between market returns and SVI is bi-directional.

Since SVI is a proxy of information demand and investors are likely to need more information in periods of uncertainty, a bunch of papers relate SVI to market volatility. Assuming both that the presence of uninformed investors increases market volatility (Black, 1986; De Long et al., 1990) and that retail investors are likely to be uninformed, Dimpfl and Jank (2016) report that SVI is a relevant proxy of retail investors' attention since it is positively related to volatility. They even document that SVI Granger causes market volatility at the daily level. Vlastakis and Markellos (2012) show that the SVI for the query 'S&P500' is positively associated with stock idiosyncratic realized variance, despite conflicting results when they use the ticker. Aouadi et al. (2013) find that market-level attention increases volatility, while the relationship with stock-level attention varies across stocks. Vozlyublennaia (2014) brings evidence that volatility Granger causes attention, although the opposite is not true. Hamid and Heiden (2015) report that investor attention provides better prediction of volatility when volatility is higher. In a more theoretical approach, Andrei and Hasler (2014) provide a framework to link investor attention and stock market volatility. In their empirical application, they follow Da et al. (2011) and construct a proxy based on the SVI of several financial terms, called FEARS.[6]

With regard to the relationship between SVI and liquidity, the literature is scarcer. Aouadi et al. (2013) report conflicting results as an increase in attention at the stock-level leads to an

[6]FEARS stands for Financial and Economic Attitudes Revealed by Search. This index is constructed by aggregating the SVI of economic terms (a.o. 'recession', 'unemployment', 'bankruptcy', 'financial crisis', etc.).

improvement in liquidity while an increase in attention at the market-level has the opposite effect. Ding and Hou (2015) document an inverse relationship between relative spreads and attention, while they find no significant result when using the Amihud (2002)'s ratio as liquidity proxy. The former finding supports Bank et al. (2011)'s results, but the latter does not.

Our study is related to the research stream addressing the relationship between SVI and trading volume. Bank et al. (2011) document that stocks with higher (lower) changes in SVI exhibit positive (negative) changes in trading volume and turnover. Joseph et al. (2011) and Takeda and Wakao (2014) also report a positive correlation between investor attention and abnormal trading volume. Vlastakis and Markellos (2012) confirms a positive association between market trading volume and SVI, be it defined at the stock-level or at the market index-level. Aouadi et al. (2013) provide however evidence that the relationship between market trading volume and investor attention is stronger when they use the SVI on the market index name, i.e. CAC40 in their sample. As we are interested in explaining trading activity, and more specifically retail investors' trading activity, our work fits well in that strand of research.

The common ground in the aforementioned papers, summarized in Table A1, is that they build (implicitly or explicitly) on Barber and Odean (2008)'s framework of limited attention. Barber and Odean (2008) argue that retail investors are net buyers of attention-grabbing stocks. The rationale behind is that the decision to buy and to sell are fundamentally different, especially for retail investors. Retail buyers have to choose from a large set of available securities while sellers can only sell what they already own.[7] Their buying behavior is then more heavily influenced by attention than their selling behavior, which suggests that increased attention leads to a temporary buying pressure that subsequently results in positive returns.

Addressing the above buying pressure hypothesis, the present paper brings several contributions to the literature. In contrast with most authors who focus on market activity (or part of it) and can only infer retail trading volumes (or put some focus on stocks that are (supposed to be) the most traded by retail investors), we directly relate SVI to our retail investors' trading activity. Consequently, our results are not biased by any institutional trading. In addition, building on Da et al. (2011) who suggest that SVI likely captures the attention of less sophisticated retail investors, we check whether SVI helps explain trading activity for different sub-samples of retail investors. In particular, we distinguish sub-samples based upon either socio-demographic characteristics that are usual control variables (e.g. age, gender, education, and spoken language) or subjective characteristics that could affect trading behavior (e.g. financial literacy and risk

[7]Short-selling is often prohibited for retail investors.

aversion). Furthermore, since the empirical investigations have been conducted so far at the market level, distinguishing purchases from sales was not possible. We fill the gap in this paper since our data about retail investors enable us to sign trading volumes. We are therefore able to determine whether the relationship to attention differs between purchases and sales. Finally, we contribute to the ongoing debate about the causality issue by investigating the dynamics between our retail investors' trading activity and attention.

In a recent paper, Kostopoulos and Meyer (2018) also relate SVI to German retail investors' trading activity. There are two key differences between their empirical work and ours. Firstly, using the SVI based on each stock ticker,[8] our measure of attention is stock-specific while their measure is common across all stocks since they use the FEARS index (Da et al. (2015)). Secondly, by construction, the FEARS index provides a measure of pessimism (Kostopoulos and Meyer, 2018, p. 2), and therefore aims at measuring retail investors' sentiment.[9] By contrast, we conjecture that our SVI delivers a measure of attention, which is neutral by definition.

3 Data and sample

3.1 Sample of stocks

For the purpose of this paper, we use available trading accounts of Belgian retail investors (that will be described below). Hence, we need to identify a sample of stocks that meet two conditions: (1) each stock has to be traded by these retail investors and (2) the Google SVI must be available for each stock. Based on D'Hondt and Roger (2017) and Bellofatto et al. (2018) who use the same database as ours,[10] we know that Belgian retail investors tend to focus most of their trading activity on Belgian, US, French, and Dutch stocks.[11] We target therefore the constituents of the market indices representative of these four countries: BEL20, SBF120 (including the CAC40), AEX25, NASDAQ100 and S&P500.

Combining the targeted stocks and the available data about retail investors' trades, we end up with a sample of 455 stocks. For each of them, we collect historical monthly prices and market volume on Bloomberg for the period January 2004-March 2012. Previous empirical work is often

[8] We motivate this choice in Section 3.3.

[9] Da et al. (2015, p. 1) indicates that *'the FEARS index is a new measure of investor sentiment.'*

[10] These authors use the same database but focus on different samples, according to their research questions.

[11] This feature is mainly due to the small size of the Belgian stock market. About 150 stocks are listed on the Euronext Brussels Stock Exchange in comparison with more than 4,000 domestic companies listed in the US.

restricted to a few stocks, a single index, and/or a tiny time window. By contrast, we analyze both a large sample of stocks and a long time period. Our sample is made of 331 US stocks, 86 French stocks, 18 Dutch stocks and 10 Belgian stocks. The remaining 10 stocks were issued in various countries. Table A2 (in appendix) reports the five most traded stocks by our sample of retail investors, depending on the market index.

3.2 Sample of investors

Our data come from a Belgian brokerage house and cover a large sample of retail investors who traded online between January 2004 and March 2012. For each transaction, we have detailed information, i.e. the stock traded, the number of shares traded, the trade price, the trade direction, the trade currency, time-stamps, etc. As just mentioned, these available data allowed us to define our sample of 455 stocks. We then filter the investors who traded (at least once) one of these stocks, which results in 42,731 retail investors.

Over the 99-month period, our sample of retail investors executed 1,021,911 trades across the 455 stocks, among which 587,100 are purchases and 434,811 are sales (57% and 43%, respectively). Consistent with the literature, our retail investors are overall net buyers. In monetary volume, all their trades amount to 11,023,511,199 euros.[12] On a monthly basis, the typical investor makes 2.59 trades.[13] Regarding the universe of stocks, the average investor trades 6 stocks out of 455.

For each investor, we have additional data that we classify as either socio-demographic or subjective characteristics. Socio-demographics encompass age, gender, level of education, and spoken language. Our sample counts only 5,653 females (i.e. 13%) and the average (median) investor is 49 (48) years old in 2012.[14] For education, three distinct levels are available: 3,414 investors have no degree, 7,795 investors have secondary school/high school qualification, and 27,975 investors hold an university degree or equivalent.[15] The majority of our retail investors (i.e. 66%) have the highest level of education. As far as spoken language is concerned, Belgium has three

[12]When necessary, we use historical exchange rates to convert monetary volumes into euros.

[13]Due to our selection of stocks, this trading frequency is somewhat higher than what is observed on other samples used in the literature on retail trading activity. For example, Dhar and Zhu (2006) report an average of 60 trades on stocks per investor on a period from 1991 to 1995 (i.e. 1 trade per month per investor) and Barber and Odean (2000) find an average of 30 trades on stocks per investor over a 5-year period (i.e. 0.5 trade per month per investor).

[14]Age is determined in 2012 using the available year of birth.

[15]This information is missing for some investors.

official languages (French, Dutch and German), among which French and Dutch are spoken the most. Nevertheless, our retail investors had to choose from the three languages available on the online trading platform: French, Dutch or English. 52%, 43%, and 5% of them selected Dutch, French, and English, respectively.

Subjective investor characteristics are survey-based data collected by the brokerage house within the context of the MiFID regulation that came into force in November 2007 in the EU member states.[16] In short, this piece of regulation has made it compulsory for investment firms to collect specific information about their retail clients' needs and preferences. Accordingly, investment firms operating in the EU are obliged to submit questionnaires (that are then referred to as 'MiFID tests') to their clients in order to determine their level of knowledge and experience, their investment objectives as well as their financial capacity. Such items are usually covered in Investment Policy Statements (IPS) used in portfolio management delegation.[17]

In the present paper, we focus on two subjective characteristics: risk aversion and financial literacy. Since these data are survey-based, they are only available for the investors who completed the questionnaire, that is 20,119 investors (i.e. 47%). When considering risk aversion, investors had to self-report their attitude towards risk on a scale ranging from 1 (high risk aversion) to 5 (high risk tolerance). The majority of them seem to be risk tolerant since 65% declare a medium risk aversion and 28% even a high risk tolerance. Only 7% of the investors selected high risk aversion. As far as financial literacy is concerned, investors were required to self-assess their financial knowledge using the three available options: no knowledge, average knowledge, and good knowledge. 55% of the investors consider they have an average knowledge, while 30% declare to have a good knowledge. Only 15% self-report no knowledge.

3.3 Google SVI

When looking for stock-specific information on the Internet, investors can use several keywords. The query may be related to the firm name (e.g. 'Apple' such as in Bank et al. (2011), Jacobs and Weber (2011), Vlastakis and Markellos (2012), and Aouadi et al. (2013)) or the stock ticker

[16]MiFID stands for the Markets in Financial Instruments Directive. MiFID I (2004/39/EC) is known as the first version of this Directive while a review of it was recently implemented in January 2018 (known as MiFID II (2014/65/UE)). For more details, please visit the European Commission website (`http://ec.eurosopa.eu/internal_market/securities/isd/mifid2/index_en.htm`).

[17]MiFID tests can be viewed as a kind of regulated IPS that are required when any retail investor asks for financial advice and/or portfolio management services. For more details on the MiFID tests, please refer to Bellofatto et al. (2018).

(e.g. 'AAPL' such as in Da et al. (2011), Joseph et al. (2011), Drake et al. (2012), and Ding and Hou (2015)). The query may even combine the ticker with the word 'stock' (e.g. 'AAPL stock' as in Kristoufek (2013)). Using the ticker should provide a less noisy measure of demand for financial information because it does not include searches related to non-financial requests.[18] However, retail investors are often considered as unsophisticated investors, who may not be knowledgeable about the stock ticker.[19] We might also argue that (sophisticated) investors could prefer websites specialized in financial information over Google when looking for stock-specific information.[20] While the debate is still open, we opt for the less noisy proxy, i.e. the SVI defined on the stock ticker. Our retail investors are used to trade online and we focus on 455 stocks that are supposed to be quite familiar for them (since each one is part of a well-known market index). Therefore, we assume our investors are likely to know (at least) some of the tickers.

Using each stock ticker, we download the monthly SVI for the period from January 2004 to March 2012. Given our sample of retail investors, we restrict our data request to queries sent from Belgium, assuming that the corresponding SVI is a relevant proxy of attention for them.[21] We should stress that Google Trends provides the *relative* frequencies for a given keyword, meaning that we do not get the actual number of queries. Google Trends normalization process ensures that the SVI determined for a given keyword ranges from 0 to 100 over a specific period. The maximum value (100) corresponds then to the highest number of queries while all other values are scaled to that maximum.[22]

For our sample period that spans 99 months, Google Trends automatically provides SVI at a monthly frequency. By restricting the time period, it would be possible to get weekly or daily SVI but at the cost of losing comparability among sub-periods of time. Choosing an appropriate frequency is a key decision, but this choice is rarely documented.[23] Working with monthly SVI

[18] Da et al. (2011, p. 1466) note that *'searching for a stock using its ticker is less ambiguous.'* These authors report that the correlation between the SVI defined on the ticker and the SVI defined on the firm name is close to 10%.

[19] Aouadi et al. (2013, p. 675) indicate that *'French investors are more likely using the firm name to express their demand for stock-specific information on Google.'*

[20] During the sample period, the online brokerage house provides its retail clients with a free access to an investment advice tool on stocks through its web platform.

[21] Investors' nationality is not available but we can reasonably assume that most of the investors are Belgian or based in Belgium.

[22] We use the R package **gtrends** to download SVI. When SVI increases by one, it means that the actual number of search queries rises by 1% of the maximum number of queries submitted during the period (which is unknown).

[23] One exception is Hamid and Heiden (2015), who indicate that daily frequency is inappropriate for forecasting volatility.

seems appropriate given the trading frequency of our retail investors (see Subsection 3.2).

3.4 Measures of trading activity

To measure our retail investors' trading activity, we first aggregate the number of transactions ($N^T_{i,t}$), the number of shares traded ($Q^T_{i,t}$), and the monetary volume traded ($V^T_{i,t}$) for each stock i at each month t. We then adjust these trade-based variables to distinguish purchases from sales: $N^{B/S}_{i,t}$ refers to the number of purchases/sales, $Q^{B/S}_{i,t}$ to the number of shares bought/sold, and $V^{B/S}_{i,t}$ for the monetary volume bought/sold. These signed trade-based variables will allow us to directly test the buying pressure hypothesis, which is an important contribution of this paper.

Table 1 provides descriptive statistics about our market-based variables in Panel A, SVI in Panel B, and trade-based variables in Panel C. All statistics are computed across stocks and months. Panel A reveals an average stock price of 43.26 euros and a monthly return slightly positive (0.38%). Panel B reports an average SVI of 38.04, with a minimum of 0 and a maximum of 100. Panel C shows that the average number of trades is 23 per month while the corresponding median is only 2, suggesting a high heterogeneity in the sample. All the variables in Panel C exhibit means higher than the corresponding medians (and even higher than the upper quartile), indicating a high skewness in the sample. Only 72 stocks are traded every month during the entire sample period.

Table 1: Descriptive statistics

Variable	Minimum	1^{st} Quart.	Median	Mean	3^{rd} Quart.	Maximum	Std Dev.
Panel A: Monthly market-based variables							
$P_{i,t}$	-	19.56	32.25	43.26	49.54	2,323.00	88.98
$Vol_{i,t}$	-	17.34	45.87	116.97	110.74	11,186.42	314.64
$R_{i,t}$	-186.46%	-4.02%	0.90%	0.38%	5.49%	128.83%	0.10
Panel B: Monthly Search Volume Index							
SVI	-	17	35	38.04	58	100	26.49
Panel C: Monthly trade-based variables							
N^B	-	-	1.00	13.03	5.00	5,102.00	68.89
N^S	-	-	1.00	9.65	5.00	4,131.00	48.77
N^T	-	-	2.00	22.69	11.00	8,487.00	114.24
Q^B	-	-	50.00	6,727.17	1,543.00	3,142,160.00	43,077.31
Q^S	-	-	50.00	6,114.69	1,486.00	2,976,190.00	40,239.48
Q^T	-	-	220.00	12,841.86	3,244.00	6,118,350.00	82,178.44
V^B	-	-	1,568.12	127,948.61	39,531.48	46,406,656.63	700,181.97
V^S	-	-	1,623.92	116,773.59	37,845.37	53,306,451.65	666,114.56
V^T	-	-	6,380.66	244,722.19	80,672.42	93,766,776.84	1,343,932.11

This table reports descriptive cross-sectional statistics that include, for each variable, the minimum, 1^{st} quartile, median, mean, 3^{rd} quartile, maximum values as well as the standard deviation. Panel A provides market-based variables for our sample of stocks. Each stock monthly return is computed as: $R_{i,t} = ln(P_{i,t}/P_{i,t-1})$, with $P_{i,t}$ the stock i closing price on month t. $Vol_{i,t}$ is the monthly stock volume (in million of euros). Panel B refers to the Google Search Volume Index (SVI) for each stock ticker downloaded from https://trends.google.com/trends/. Panel C refers to trade-based variables that characterize our retail investors' aggregate trading activity across the 455 stocks in our sample. We provide the number of purchases, sales, and total trades (N^B, N^S, and N^T, respectively), the number of shares bought, sold, and traded (Q^B, Q^S, and Q^T, respectively), and the corresponding monetary volume in euros (V^B, V^S, and V^T, respectively).

To control for either some socio-demographic characteristics (that are age, gender, education, and spoken language) or other subjective characteristics that could affect trading behavior (such as risk aversion and financial literacy), we also build several sub-samples of retail investors and replicate on them the above aggregate trade-based variables. To the best of our knowledge, we are the first authors to combine such detailed measures of retail investors' trading activity and characteristics to investigate the impact of attention.

4 Empirical analysis

The extant literature documents a positive relationship between SVI and market trading volume (a.o. Da et al. (2011), Bank et al. (2011), Vlastakis and Markellos (2012), and Moussa et al. (2017)). As a preliminary step, we check whether this relationship holds in our sample of stocks with the following model:

$$Vol_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t} \tag{1}$$

in which $Vol_{i,t}$ is the market trading volume for stock i at month t, $SVI_{i,t}$ is the SVI for stock i at month t, $|R_{i,t}|$ is the absolute value of the return for stock i at month t, γ_i and δ_t are respectively stock- and time-fixed effects, and $\epsilon_{i,t}$ is the error term. We opt for a multivariate panel regression specification as in Bank et al. (2011). Given that SVI data are already normalized as explained in Subsection 3.3, we use SVI in levels. We show in Section 5 that this choice does not impact our results.

Table 2 reports the expected relationship between SVI and market trading activity: the higher the SVI for a stock at a given month, the higher is the corresponding market volume. Our results also show that the market trading volume increases with extreme returns. These findings are consistent with the literature and confirm that our SVI, based on queries sent from Belgium only, is a relevant proxy of investor attention at the market-level for our sample of stocks.

Table 2: Investor attention and market trading volume

Variable	$Vol_{i,t}$	
$SVI_{i,t}$	0.20	***
$\|R_{i,t}\|$	343.25	***
γ_i	YES	
δ_t	YES	
N	43,200	
R^2	68.64%	

This table reports the results for Equation 1: $Vol_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$, in which $Vol_{i,t}$ is the market trading volume for stock i at month t, $SVI_{i,t}$ is the SVI for stock i at month t, $|R_{i,t}|$ is the absolute value of the return for stock i at month t, γ_i and δ_t are respectively stock- and time-fixed effects, and $\epsilon_{i,t}$ is the error term. N is the number of observations and R^2 is the R-square. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity.

4.1 Analysis of retail investors' trading activity

To empirically test the relationship between attention and our retail investors' (signed) trading activity, we estimate the following regression model:

$$Y_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t} \tag{2}$$

wherein $Y_{i,t}$ is a (signed) trade-based variable defined in Subsection 3.4 to measure our retail investors' aggregate trading activity on stock i at month t. The set of explanatory variables is made of the SVI for stock i at month t ($SVI_{i,t}$), the market trading volume for stock i at month t ($Vol_{i,t}$), the (absolute)[24] return for stock i at month t ($R_{i,t}$), stock- and time-fixed effects (γ_i and δ_t, respectively),[25] and an error term ($\epsilon_{i,t}$).

We report the results in Table 3, wherein Panel A refers to unsigned trading activity while Panel B (C) refers to purchases (sales). Whatever the dependent variable, SVI displays a positive and significant coefficient at the 1% level. This positive relationship between attention and retail investors' trading activity is valid on both unsigned and signed trade-based measures. A higher attention is associated with a higher retail investors' trading activity on both market sides. When attention is higher, we observe more purchases but also more sales in our sample. Economically speaking, the number of trades increases by 0.27 on average when SVI increases by 1. This effect is relatively modest, although it should be put in perspective with the level of trading activity in our sample (documented in Table 1).

Table 3 shows that our retail investors' trades are also positively related to market volume.[26] More interestingly, we observe in Panels B and C the contrasting relationship between returns and signed trades, i.e. the number of purchases (sales) tend to decrease (increase) when returns increase. This relationship, which is however not always significant, is consistent with the disposition effect (i.e. tendency to sell stocks in bullish markets and reluctance to sell stocks in bearish markets).

[24] We take returns in absolute value when the dependent variable is not signed.

[25] As a robustness check, we control for observable time-effects by using the monthly market return computed as the log-return of the S&P500 index, and the CBOE index as a proxy for volatility. Results are qualitatively similar and are available upon request.

[26] Our results remain similar when we consider the market volume net of our retail trading activity.

Table 3: Investor attention and retail trading activity

Panel A: Unsigned trading activity	$N^T_{i,t}$		$Q^T_{i,t}$		$V^T_{i,t}$	
$SVI_{i,t}$	0.27	***	110.22	***	3,193.10	***
$Vol_{i,t}$	0.03	***	111.34	***	745.60	***
$\|R_{i,t}\|$	131.48	***	102,666.22	***	1,297,563.80	***
γ_i	YES		YES		YES	
δ_t	YES		YES		YES	
N	43,200		43,200		43,200	
R^2	43.95%		34.07%		43.81%	
Panel B: Purchases	$N^B_{i,t}$		$Q^B_{i,t}$		$V^B_{i,t}$	
$SVI_{i,t}$	0.15	***	59.97	***	1,637.58	***
$Vol_{i,t}$	0.02	***	60.92	***	420.70	***
$R_{i,t}$	(25.13)	***	1,364.61		(56,749.58)	
γ_i	YES		YES		YES	
δ_t	YES		YES		YES	
N	43,200		43,200		43,200	
R^2	40.10%		32.95%		42.86%	
Panel C: Sales	$N^S_{i,t}$		$Q^S_{i,t}$		$V^S_{i,t}$	
$SVI_{i,t}$	0.12	***	54.01	***	1,600.17	***
$Vol_{i,t}$	0.02	***	54.62	***	377.70	***
$R_{i,t}$	13.30	*	19,371.93	**	248,819.13	**
γ_i	YES		YES		YES	
δ_t	YES		YES		YES	
N	43,200		43,200		43,200	
R^2	42.66%		32.52%		41.45%	

This table reports the results of Equation 2: $Y_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3|R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$, wherein $Y_{i,t}$ is a (signed) trade-based variable defined in Subsection 3.4 to measure our retail investors' aggregate trading activity on stock i at month t. The set of explanatory variables is made of the SVI for stock i at month t ($SVI_{i,t}$), the market trading volume for stock i at month t ($Vol_{i,t}$), the (absolute) return for stock i at month t ($R_{i,t}$), stock- and time-fixed effects (γ_i and δ_t, respectively), and an error term ($\epsilon_{i,t}$). We use the absolute return in Panel A (unsigned trading activity) and the return in Panels B and C (purchases and sales, respectively). N is the number of observations and R^2 is the R-square. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity.

4.2 Socio-demographic characteristics

Socio-demographics such as gender, age, or education are usual control variables when investigating the behavior of retail investors. Both gender and age are recognized as major drivers of trading behavior (e.g. Barber and Odean (2001a); Goetzmann and Kumar (2008); Graham et al. (2009); Hoffmann et al. (2013); Hackethal et al. (2012); Bellofatto et al. (2018)). Similarly, the impact of education on investor behavior is established (a.o. Haliassos and Bertaut (1995); Campbell (2006); Van Rooij et al. (2011)).

Whether the relationship between investor attention and trading activity depends on such demographics is an empirical question that has not been addressed yet. To fill this gap, we estimate Equation 2 on six sub-samples, i.e. for each socio-demographic characteristic under scrutiny, we split our investors into two sub-samples. For gender, we simply separate men from women. Using the median age available in Subsection 3.2, we divide our investors into two equal groups, which allows us to flag any investor as 'Young' (below the median age) or 'Old' (above the median age). As far as education is concerned, we distinguish investors who hold an university degree ('High educ') from the others ('Low educ'). In addition, building on Grinblatt and Keloharju (2001) who focus on language (and culture effects) in Finland, we also estimate Equation 2 on three other sub-samples based on spoken language.

The results for gender, age, and education are provided in Table 4, in Panels A, B, and C, respectively.[27] Whatever the model, the coefficient of SVI is always positive and significant at the 1% level, which confirms our previous results. In Panel A, the marginal effect of $SVI_{i,t}$ appears stronger for men (0.40 versus 0.09). Panel B exhibits a stronger marginal effect of $SVI_{i,t}$ for younger investors (0.33 versus 0.19). As for Panel C, it reveals a stronger marginal effect of $SVI_{i,t}$ for low education investors (0.19 versus 0.04). The explanatory power slightly fluctuates across models. The highest R^2 is observed for 'Old' investors in Panel B.

[27]Results are reported for the dependent variable defined as the number of trades. We find similar results when considering the number of shares traded or the monetary volume. These unreported findings are available upon request.

Table 4: Investor attention and retail trading activity - socio-demographics

	Panel A: Gender				Panel B : Age				Panel C: Education			
	Men		Women		Young		Old		High educ		Low educ	
$SVI_{i,t}$	0.40	***	0.09	***	0.33	***	0.19	***	0.04	***	0.19	***
$Vol_{i,t}$	0.028	***	0.001	***	0.019	***	0.011	***	0.003	***	0.009	***
$\|R_{i,t}\|$	186.66	***	34.60	***	145.20	***	84.44	***	23.88	***	87.15	***
γ_i	YES		YES		YES		YES		YES		YES	
δ_t	YES		YES		YES		YES		YES		YES	
N	25,411		13,171		21,908		22,406		24,320		18,893	
R^2	43.31%		44.42%		38.62%		50.41%		44.17%		41.26%	

This Table reports the results of Equation 2: $Y_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$, wherein $Y_{i,t}$ is the number of trades. The set of explanatory variables is made of the SVI for stock i at month t ($SVI_{i,t}$), the market trading volume for stock i at month t ($Vol_{i,t}$), the absolute return for stock i at month t ($|R_{i,t}|$), stock- and time-fixed effects (γ_i and δ_t, respectively), and an error term ($\epsilon_{i,t}$). For each socio-demographic characteristic, we split our investors into two sub-samples. For gender (Panel A), we simply separate men from women. For age (Panel B), using the median age available in Subsection 3.2, we divide our investors into two equal groups, which allows us to flag any investor as 'Young' (below the median age) or 'Old' (above the median age). As far as education (Panel C) is concerned, we distinguish investors who hold an university degree ('High educ') from the others ('Low educ'). N is the number of observations and R^2 is the R-square. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity. Results are reported here for the dependent variable defined as the number of trades. Results are similar when considering the number of shares traded or the monetary volume. These unreported findings are available upon request.

Table 5 provides the results for the sub-samples based on spoken language. They are still consistent and outline that attention is positively related to trading activity, whatever the spoken language. The marginal effect of $SVI_{i,t}$ is similar for French-speaking and Dutch-speaking investors (0.25 and 0.26), while it appears much lower for English-speaking investors (0.04). The highest R^2 is observed when estimating our model on the sub-sample of French-speaking investors.

Table 5: Investor attention and retail trading activity - language

	Spoken language					
	FR		NL		EN	
$SVI_{i,t}$	0.25	***	0.26	***	0.04	***
$Vol_{i,t}$	0.016	***	0.010	***	0.003	***
$\|R_{i,t}\|$	110.53	***	121.96	***	11.09	***
γ_i	YES		YES		YES	
δ_t	YES		YES		YES	
N	22,038		20,775		9,552	
R^2	45.06%		41.17%		42.68%	

This Table reports the results of Equation 2: $Y_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$, wherein $Y_{i,t}$ is the number of trades. The set of explanatory variables is made of the SVI for stock i at month t ($SVI_{i,t}$), the market trading volume for stock i at month t ($Vol_{i,t}$), the absolute return for stock i at month t ($|R_{i,t}|$), stock- and time-fixed effects (γ_i and δ_t, respectively), and an error term ($\epsilon_{i,t}$). We divide our sample of investors according to the language chosen on the platform, i.e. French (FR), Dutch (NL), and English (EN). N is the number of observations and R^2 is the R-square. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity. Results are reported here for the dependent variable defined as the number of trades. Results are similar when considering the number of shares traded or the monetary volume. These unreported findings are available upon request.

4.3 Subjective investor characteristics

Some subjective individual attributes provide valuable insights into investor behavior (Dorn and Huberman (2005); Graham et al. (2009)). Focusing on financial literacy, Bellofatto et al. (2018) show that self-reported knowledge helps explain cross-sectional variations in retail investors' behavior.[28] Hence, it appears relevant to check whether the relationship between attention and trading activity is affected by such a self-declared literacy. For that purpose, we build two sub-samples to control for subjective financial literacy. Using the data available, we keep the extreme levels to create a sub-sample of low-literate investors (those who declare 'no knowledge' about financial markets) and high-literate investors (those who select 'good knowledge'), respectively.

Beugelsdijk and Frijns (2010, p. 2123) define uncertainty avoidance as *'the extent to which people feel (un)comfortable in situations with uncertain outcomes and their willingness to deal with*

[28] In the literature, a large body of papers show that financial literacy is related to different aspects of financial behavior (a.o. Kimball and Shumway (2006); Christelis et al. (2010); Van Rooij et al. (2011); Lusardi and Mitchell (2014);).

risk.' To examine the relationship between risk aversion and investor attention, Vlastakis and Markellos (2012) rely on a measure of variance risk premium. The latter is however computed at the market level and does not allow to check whether individual risk aversion affects the relationship between attention and trading activity. Using the level of risk aversion self-reported by our investors (described in Subsection 3.2), we are able to build two sub-samples aiming at distinguishing risk-averse investors (two lowest levels on the scale) and risk-tolerant investors (level 5).[29]

As in Subsection 4.2, we estimate Equation 2 on each of the four above sub-samples. We provide the results in Table 6, in Panel A for risk aversion and in Panel B for financial literacy. Whatever the model, our variable of interest, $SVI_{i,t}$, always exhibits a positive and highly significant coefficient. In Panel A, the marginal effect of $SVI_{i,t}$ appears stronger for risk-tolerant investors (0.13 versus 0.04). The R^2 is also higher when the model is estimated on this sub-sample of investors. Panel B reveals that the marginal effect of $SVI_{i,t}$ is somewhat stronger for high-literate investors (0.10 versus 0.08). The R^2 is also higher when the model is estimated on this sub-sample. We might hypothesize that risk-tolerant and/or high-literate investors tend to monitor the market more closely and are therefore more likely to follow market trends. Interestingly, the marginal effect of $Vol_{i,t}$ and $R_{i,t}$ on their trading activity is also stronger.

[29]This choice is mainly motivated by the question and the scale available in the questionnaire.

Table 6: Investor attention and retail trading activity - subjective characteristics

	Panel A: Risk aversion				Panel B: Financial literacy			
	Risk-averse		Risk-tolerant		Low-literate		High-literate	
$SVI_{i,t}$	0.04	***	0.13	***	0.08	***	0.10	***
$Vol_{i,t}$	(0.000)		0.007	***	(0.000)		0.005	***
$\|R_{i,t}\|$	13.94	***	52.95	***	35.52	***	43.63	***
γ_i	YES		YES		YES		YES	
δ_t	YES		YES		YES		YES	
N	13,428		17,069		8,812		18,932	
R^2	33.20%		40.68%		34.02%		46.71%	

This Table reports the results of Equation 2: $Y_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$, wherein $Y_{i,t}$ is the number of trades. The set of explanatory variables is made of the SVI for stock i at month t ($SVI_{i,t}$), the market trading volume for stock i at month t ($Vol_{i,t}$), the absolute return for stock i at month t ($|R_{i,t}|$), stock- and time-fixed effects (γ_i and δ_t, respectively), and an error term ($\epsilon_{i,t}$). Using the survey-data described in Subsection 3.2, we split our investors into two sub-samples aiming at distinguishing risk-averse investors (two lowest levels on the scale) and risk-tolerant investors (level 5). For financial literacy, we keep the extreme levels to create a sub-sample of low-literate investors (those who declare 'no knowledge' about financial markets) and high-literate investors (those who selected 'good knowledge'), respectively. N is the number of observations and R^2 is the R-square. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity. Results are reported here for the dependent variable defined as the number of trades. Results are similar when considering the number of shares traded or the monetary volume. These unreported findings are available upon request.

4.4 Does the relationship to attention differ between purchases and sales?

Table 3 displays coefficient estimates for SVI that are higher for purchases than for sales, whatever the dependent variable. This suggests a stronger marginal effect of $SVI_{i,t}$ on purchases. When investors want to buy shares, they can choose among a large set of stocks. By contrast, they mostly sell stocks that they already hold. This is especially true for retail investors who are often banned from short-selling, as it is the case in our sample. Such a discrepancy in opportunities could explain why the relationship between SVI and our trade-based variables is economically stronger for purchases than for sales. To test the statistical significance of this discrepancy, we estimate the following regression model:

$$Y_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 B + \alpha_3 (B^* SVI_{i,t}) + \alpha_4 Vol_{i,t} + \alpha_5 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t} \quad (3)$$

In contrast with Equation 2, $Y_{i,t}$ represents only either retail investors' purchases or sales. In the set of explanatory variables, we include a dummy variable equal to 1 for purchases and zero for sales (B) and an interaction variable ($B^{*}SVI_{i,t}$).[30] Using this specification allows us to test (i) whether retail investors' purchases are higher than their sales, and (ii) whether the relationship between trading activity and attention is stronger for purchases than for sales. The coefficient of this interaction variable will indicate whether the relationship between SVI and our investors' trading activity differs with respect to the trade direction. The other explanatory variables were defined previously and are made of the SVI for stock i at month t ($SVI_{i,t}$), the market trading volume for stock i at month t ($Vol_{i,t}$), the absolute return for stock i at month t ($|R_{i,t}|$), stock- and time-fixed effects (γ_i and δ_t, respectively), and an error term ($\epsilon_{i,t}$).

We report the results in Table 7. When the dependent variable is defined as the number of trades, the coefficient of the dummy variable (B) is positive and significant at the 1% level, indicating that there are more purchases than sales in our sample. Retail investors are generally net buyers of stocks, which is a phenomenon well-established in the literature. More importantly, the coefficient of the interaction variable ($B^{*}SVI_{i,t}$) is positive but not significant. Hence there is no significant difference in the relationship between SVI and our retail investors' purchases versus sales. This finding, which is still similar when the dependent variable is defined either in number of shares or in monetary value, brings no support for the buying pressure hypothesis.[31]

[30]This explain why the value of N in Table 7 is twice the value of the N in Table 3.

[31]We are not able to examine to what extent such a result is dependent upon the monthly frequency of our analysis, which leaves interesting avenues for future research.

Table 7: Investor attention and retail purchases versus sales

	$N^T_{i,t}$		$Q^T_{i,t}$		$V^T_{i,t}$	
$SVI_{i,t}$	0.13	***	52.54	***	1,590.98	***
B	3.00	***	416.18		10,709.76	**
$B * SVI_{i,t}$	0.01		5.15		11.14	
$Vol_{i,t}$	0.02	***	55.67	***	372.79	***
$\|R_{i,t}\|$	65.74	***	51,333.11	***	648,781.92	***
γ_i	YES		YES		YES	
δ_t	YES		YES		YES	
N	86,400		86,400		86,400	
R^2	40.33%		33.16%		42.38%	

This table reports the results of Equation 3: $Y_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 B + \alpha_3(B^* SVI_{i,t}) + \alpha_4 Vol_{i,t} + \alpha_5|R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$, wherein $Y_{i,t}$ is a signed trade-based variable defined in Subsection 3.4 to measure our retail investors' aggregate trading activity on stock i at month t. The set of explanatory variables is made of the SVI for stock i at month t ($SVI_{i,t}$), a dummy variable equal to 1 for purchases and zero for sales (B), an interaction variable ($B^* SVI_{i,t}$), the market trading volume for stock i at month t ($Vol_{i,t}$), the absolute return for stock i at month t ($|R_{i,t}|$), stock- and time-fixed effects (γ_i and δ_t, respectively), and an error term ($\epsilon_{i,t}$). N is the number of observations and R^2 is the R-square. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity.

4.5 Does attention cause trading?

All our analyses document a positive and significant relationship between attention and our retail investors' trading activity. A natural question that emerges is whether attention carries some predictive power for our retail trading activity. We address that issue by including a lagged term of $SVI_{i,t}$ (i.e. $SVI_{i,t-1}$) into Equation 2. More specifically, we compare an unrestricted model (UM - Equation 4a) that include both contemporaneous and lagged SVI, with three restricted models (RM) wherein $\beta_2 = 0$ (Equation 4b), $\beta_1 = 0$ (Equation 4c), and $\beta_1 = \beta_2 = 0$ (Equation 4d), respectively. These models are :

$$Y_{i,t} = \beta_1 SVI_{i,t} + \beta_2 SVI_{i,t-1} + a_1 Vol_{i,t} + a_2|R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t} \quad (4a)$$

$$Y_{i,t} = \beta_1 SVI_{i,t} + a_1 Vol_{i,t} + a_2|R_{i,t}| + \gamma_i + \delta_t + u_{i,t} \quad (4b)$$

$$Y_{i,t} = \beta_2 SVI_{i,t-1} + a_1 Vol_{i,t} + a_2|R_{i,t}| + \gamma_i + \delta_t + u_{i,t} \quad (4c)$$

$$Y_{i,t} = a_1 Vol_{i,t} + a_2|R_{i,t}| + \gamma_i + \delta_t + u_{i,t} \quad (4d)$$

All the variables were defined previously, but two points are worth noticing. First, Equation 4b corresponds to Equation 2, with only minor differences in the notation. Second, $\epsilon_{i,t}$ refers to the UM's residuals while $u_{i,t}$ refers to RM's residuals. As the unrestricted model contains more variables than the restricted model, the R^2 is *de facto* higher, but the question is to test whether this increase is statistically significant. We use an F-test, with the F-statistic being computed as follows:

$$F = \frac{(RSS_0 - RSS_1)/p}{RSS_1/(T-2p-1)} \sim F_{p,T-2p-1} \tag{5}$$

with p, the number of lags for the SVI variable (i.e. one). $RSS_0(=\sum_{i,t=1}^{N,T} u_{i,t}^2)$ and $RSS_1(=\sum_{i,t=1}^{N,T} \epsilon_{i,t}^2)$ correspond to the RM and UM sum of squared residuals, respectively. This approach slightly departs from a traditional Granger causality test for two reasons. First, we control for the contemporaneous effect in the regression. Second, the number of lags, i.e. one, is determined by economic intuition, rather than by information criteria (e.g. AIC or BIC).

Table 8 reports the results. When focusing on Panel A that refers to the dependent variable expressed in number of trades, both $SVI_{i,t}$ and $SVI_{i,t-1}$ in Equation 4a exhibit a positive and significant coefficient. Therefore, an increase in the past level of attention also leads to higher trading activity in the following month. However, the marginal effect of $SVI_{i,t-1}$ is lower, compared to the one of the contemporaneous SVI ($SVI_{i,t}$). When comparing Equations 4a and 4d, the R^2 increase is small, but highly significant. Comparing Equations 4b or 4c to Equation 4a, we conclude that including either $SVI_{i,t}$ or $SVI_{i,t-1}$ enhances the explanatory power of our model. These results are still consistent in Panel B (when the dependent variable is defined in number of shares) and in Panel C (when the dependent variable is defined in monetary value).

Table 8: Does attention cause trading?

Panel A: $N^T_{i,t}$	(4a)		(4b)		(4c)		(4d)	
$SVI_{i,t}$	0.22	***	0.27	***				
$SVI_{i,t-1}$	0.16	***			0.22	***		
$Vol_{i,t}$	0.03	***	0.03	***	0.03	***	0.03	***
$\|R_{i,t}\|$	131.68	***	131.48	***	132.17	***	132.04	***
γ_i	YES		YES		YES		YES	
δ_t	YES		YES		YES		YES	
N	43,199		43,200		43,199		43,20	
R^2	44.00%		43.95%		43.91%		43.80%	
F			39.74	***	72.85	***	156.42	***
Panel B: $Q^T_{i,t}$	**(4a)**		**(4b)**		**(4c)**		**(4d)**	
$SVI_{i,t}$	90.23	***	110.22	***				
$SVI_{i,t-1}$	68.34	***			93.82	***		
$Vol_{i,t}$	111.25	***	111.34	***	111.36	***	111.52	***
$\|R_{i,t}\|$	102,750.14	***	102,666.22	***	102,952.81	***	102,894.37	***
γ_i	YES		YES		YES		YES	
δ_t	YES		YES		YES		YES	
N	43,199		43,200		43,199		43,200	
R^2	34.09%		34.07%		34.06%		34.02%	
F			11.96	***	20.13	***	44.72	***
Panel C: $V^T_{i,t}$	**(4a)**		**(4b)**		**(4c)**		**(4d)**	
$SVI_{i,t}$	2,612.10	***	3,193.10	***				
$SVI_{i,t-1}$	1,987.40	***			2,725.00	***		
$Vol_{i,t}$	743.00	***	745.60	***	746.10	***	750.90	***
$\|R_{i,t}\|$	1,299,997.20	***	1,297,563.80	***	1,305,864.20	***	1,304,173.30	***
γ_i	YES		YES		YES		YES	
δ_t	YES		YES		YES		YES	
N	43,199		43,200		43,199		43,200	
R^2	43.87%		43.81%		43.78%		43.67%	
F			44.12	***	73.56	***	163.94	***

This Table reports the results of Equations 4a - 4d. Equation 4a is: $Y_{i,t} = \beta_1 SVI_{i,t} + \beta_2 SVI_{i,t-1} + a_1 Vol_{i,t} + a_2|R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$. Equation 4b (4c / 4d) is a restricted version of this Equation wherein $\beta_2 = 0$ ($\beta_1 = 0$ / $\beta_1 = \beta_2 = 0$). $Y_{i,t}$ is the number of trades, N^T (Panel A - upper part of the Table), the number of shares traded, Q^T (Panel B - middle part of the Table), or the monetary volume, V^T (Panel C - lower part of the Table). In the unrestricted model, the set of explanatory variables is made of the SVI for stock i at month t and at month $t-1$ ($SVI_{i,t}$ and $SVI_{i,t-1}$), the market trading volume for stock i at month t ($Vol_{i,t}$), the absolute return for stock i at month t ($|R_{i,t}|$), stock- and time-fixed effects (γ_i and δ_t, respectively), and an error term ($\epsilon_{i,t}$). N is the number of observations and R^2 is the R-square. F is the result of the F-test, which determines the statistical significance of the increase in the R^2 between the restricted and the unrestricted model. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity.

4.6 Does trading cause attention?

The previous subsection brings evidence that the past level of attention helps explain retail trading activity. We now test whether we observe any causality in the other direction, i.e. does trading cause attention? To do so, we replicate the same approach and estimate Equations 6a to 6d, which are defined as follows:

$$SVI_{i,t} = \beta_1 Y_{i,t} + \beta_2 Y_{i,t-1} + a_1 Vol_{i,t} + a_2 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t} \quad (6a)$$

$$SVI_{i,t} = \beta_1 Y_{i,t} + a_1 Vol_{i,t} + a_2 |R_{i,t}| + \gamma_i + \delta_t + u_{i,t} \quad (6b)$$

$$SVI_{i,t} = \beta_2 Y_{i,t-1} + a_1 Vol_{i,t} + a_2 |R_{i,t}| + \gamma_i + \delta_t + u_{i,t} \quad (6c)$$

$$SVI_{i,t} = a_1 Vol_{i,t} + a_2 |R_{i,t}| + \gamma_i + \delta_t + u_{i,t} \quad (6d)$$

All the variables were defined previously.[32] We report the results in Table 9. We find the expected relationship between any measure of trading activity (expressed either in number of trades ($N^T_{i,t}$), in number of shares ($Q^T_{i,t}$), or in monetary value ($V^T_{i,t}$)) and the contemporaneous SVI ($SVI_{i,t}$). This relationship is positive and significant at the 1% level. Adding the past level of trading activity to its contemporaneous level only improves slightly the explanatory power of the model, although the R^2 increase is almost always significant. Economically speaking, these findings indicate that trading activity tends to increase attention up to the next month. This might suggest that retail investors keep on monitoring the stocks that they have just bought.

[32] To obtain readable coefficients, $Q^T_{i,t}$ is expressed in millions of shares and $V^T_{i,t}$ is in millions of euros.

Table 9: Does trading cause attention?

Panel A: $SVI_{i,t}$	(6a)		(6b)		(6c)		(6d)	
$N^T_{i,t}$	0.0065	***	0.0097	***				
$N^T_{i,t-1}$	0.0049	***			0.0090	***		
$Vol_{i,t}$	0.0013	***	0.0013	***	0.0014	***	0.0017	***
$\lvert R_{i,t}\rvert$	0.9200		0.7928		1.5365		2.0699	
γ_i	YES		YES		YES		YES	
δ_t	YES		YES		YES		YES	
N	43,199		43,200		43,199		43,200	
R^2	87.21%		87.20%		87.20%		87.17%	
F			19.99	***	31.41	***	136.62	***
Panel B: $SVI_{i,t}$	(6a)		(6b)		(6c)		(6d)	
$Q^T_{i,t}$	5.5160	***	6.58979	***				
$Q^T_{i,t-1}$	1.6575				5.0544	***		
$Vol_{i,t}$	0.0009	*	0.00093	**	0.0012	***	0.0017	***
$\lvert R_{i,t}\rvert$	1.4369		1.39189		1.8690		2.0699	
γ_i	YES		YES		YES		YES	
δ_t	YES		YES		YES		YES	
N	43,199		43,200		43,199		43,200	
R^2	87.18%		87.18%		87.18%		87.17%	
F			2.63		13.79	***	35.38	***
Panel C: $SVI_{i,t}$	(6a)		(6b)		(6c)		(6d)	
$V^T_{i,t}$	0.6154	***	0.8301	***				
$V^T_{i,t-1}$	0.3326	**			0.7168	***		
$Vol_{i,t}$	0.0010	**	0.0010	**	0.0013	***	0.0017	***
$\lvert R_{i,t}\rvert$	1.1502		0.9874		1.8168		2.0699	
γ_i	YES		YES		YES		YES	
δ_t	YES		YES		YES		YES	
N	43,199		43,200		43,199		43,200	
R^2	87.21%		87.20%		87.20%		87.17%	
F			13.21	***	39.29	***	132.95	***

This Table reports the results of Equations 6a to 6d. Equation 6a is $SVI_{i,t} = \beta_1 Y_{i,t} + \beta_2 Y_{i,t-1} + a_1 Vol_{i,t} + a_2|R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$. Equation 6b (6c / 6d) is a restricted version of this Equation wherein β_2 =0 ($\beta_1 = 0$ / $\beta_1 = \beta_2 = 0$). In the unrestricted model, the set of explanatory variables is made of a measure of trading activity, i.e. either the number of trades ($N^T_{i,t}$), the number of shares traded ($Q^T_{i,t}$), or the monetary volume ($V^T_{i,t}$) for stock i at time t and at time $t-1$, the market trading volume for stock i at month t ($Vol_{i,t}$), the absolute return for stock i at month t ($|R_{i,t}|$), stock- and time-fixed effects (γ_i and δ_t, respectively), and an error term ($\epsilon_{i,t}$). N is the number of observations and R^2 is the R-square. F is the result of the F-test, which determines the statistical significance of the R^2 increase between the restricted and the unrestricted model. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity.

5 Robustness checks

We perform three robustness checks. Firstly, we restrict our analysis to the 372 stocks which remain in the sample during the whole period. By doing so, we obtain a *balanced* panel with which we estimate Equation 2 for the unsigned trade-based variables.[33] The results are reported in Table 10 and show that the relationship between retail trading and SVI is still positive and significant at the 1% level. The coefficients of SVI are very close to those observed in Table 3 (e.g., 0.28 versus 0.27; 104.41 versus 110.22; and 3,508.40 versus 3,193.10). R^2 are also in line with those previously obtained.

Table 10: Investor attention and retail trading activity - Balanced panel

	$N^T_{i,t}$		$Q^T_{i,t}$		$V^T_{i,t}$	
$SVI_{i,t}$	0.28	***	104.41	***	3,508.40	***
$Vol_{i,t}$	0.03	***	115.72	***	778.50	***
$\|R_{i,t}\|$	156.78	***	125,962.29	***	1,558,122.40	***
γ_i	YES		YES		YES	
δ_t	YES		YES		YES	
N	36,456		36,456		36,456	
R^2	42.48%		34.19%		43.15%	

This table reports the results of Equation 2: $Y_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$, wherein $Y_{i,t}$ is a unsigned trade-based variable defined in Subsection 3.4 to measure our retail investors' aggregate trading activity on stock i at month t. The set of explanatory variables is made of the SVI for stock i at month t ($SVI_{i,t}$), the market trading volume for stock i at month t ($Vol_{i,t}$), the (absolute) return for stock i at month t ($R_{i,t}$), stock- and time-fixed effects (γ_i and δ_t, respectively), and an error term ($\epsilon_{i,t}$). N is the number of observations and R^2 is the R-square. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity. The results are similar when considering the signed trade-based variables (purchases and sales). These unreported findings are available upon request.

As a second robustness check, we consider different specifications of SVI to ensure that the way SVI is defined does not influence our results. Some authors modify raw SVI either by taking the changes in SVI (Dzielinski, 2012), by using a standardized SVI (Bijl et al., 2016; Swamy and Dharani, 2019), or by taking the natural logarithm of SVI (Aouadi et al., 2013; Takeda and Wakao, 2014; Vozlyublennaia, 2014; Dimpfl and Jank, 2016). It is not straightforward to compare results across studies for that particular reason.[34] We then estimate the following

[33]The results are similar when considering the signed trade-based variables (purchases and sales). These unreported findings are available upon request.

[34]For example, Bank et al. (2011) and Ding and Hou (2015) find conflicting results for the relationship between

regression models:

$$\Delta Y_{i,t} = \alpha_1 \Delta SVI_{i,t} + \alpha_2 \Delta Vol_{i,t} + \alpha_3 |R_{i,t}| + \delta_t + \epsilon_{i,t} \tag{7a}$$

$$Y_{i,t} = \alpha_1 SSVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t} \tag{7b}$$

$$Y_{i,t} = \alpha_1 LN(SVI_{i,t}) + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t} \tag{7c}$$

with $\Delta Y_{i,t} = Y_{i,t} - Y_{i,t-1}$, $\Delta SVI_{i,t} = SVI_{i,t} - SVI_{i,t-1}$, and $\Delta Vol_{i,t} = Vol_{i,t} - Vol_{i,t-1}$. We standardize SVI at the stock-level, as follows:

$$SSVI_{i,t} = \frac{SVI_{i,t} - \mu_{SVI_i}}{\sigma_{SVI_i}} \tag{8}$$

where μ_{SVI_i} and σ_{SVI_i} are respectively the mean and the standard deviation of SVI for stock i over the sample period. All other variables were defined previously.[35]

The results are reported in Table 11, in Panel A for Equation 7a, in Panel B for Equation 7b, and in Panel C for Equation 7c. In Panel A, $\Delta SVI_{i,t}$ always displays a positive and significant coefficient. This means that any change in investor attention is positively related to a change in our retail investors' trading activity. In Panel B where we use a standardized SVI (SSVI) and in Panel C where we use the natural logarithm of SVI, we also find a positive and significant relationship between attention and retail trading activity. Consequently, our findings are robust to several specifications of SVI.

SVI and liquidity. Bank et al. (2011) use a lagged term of SVI while Ding and Hou (2015) take the change in SVI. It is not obvious whether their different results are linked to how SVI is specified.

[35]Given that SVI may have a value of zero, we take the natural logarithm of SVI+1 as it is done in Takeda and Wakao (2014).

Table 11: Alternative specifications of SVI

Panel A: Equation 7a - $\Delta SVI_{i,t}$						
	$\Delta N^T_{i,t}$		$\Delta Q^T_{i,t}$		$\Delta V^T_{i,t}$	
$\Delta SVI_{i,t}$	0.04	***	32.56	**	646.88	***
$\Delta Vol_{i,t}$	0.06	***	112.40	***	1,044.03	***
$\|R_{i,t}\|$	56.78	***	43,789.87	**	725,451.69	***
γ_i	NO		NO		NO	
δ_t	YES		YES		YES	
N	43,152		43,152		43,152	
R^2	2.09%		5.65%		3.03%	
Panel B: Equation 7b - Standardized SVI ($SSVI_{i,t}$)						
	$N^T_{i,t}$		$Q^T_{i,t}$		$V^T_{i,t}$	
$SSVI_{i,t}$	3.63	***	1,456.77	***	42,565.70	***
$Vol_{i,t}$	0.03	***	111.37	***	746.50	***
$\|R_{i,t}\|$	131.54	***	102,695.08	***	1,298,350.20	***
γ_i	YES		YES		YES	
δ_t	YES		YES		YES	
N	43,200		43,200		43,200	
R^2	46.01%		35.61%		45.56%	
Panel C: Equation 7c - Natural logarithm of SVI ($LN(SVI_{i,t})$)						
	$N^T_{i,t}$		$Q^T_{i,t}$		$V^T_{i,t}$	
$LN(SVI_{i,t})$	3.39	***	1,187.73	***	36,742.00	***
$Vol_{i,t}$	0.03	***	111.47	***	749.30	***
$\|R_{i,t}\|$	131.66	***	102,762.34	***	1,300,088.80	***
γ_i	YES		YES		YES	
δ_t	YES		YES		YES	
N	43,200		43,200		43,200	
R^2	45.99%		35.60%		45.53%	

This table reports the results of Equations 7a to 7c. Panel A reports the results of Equation (7a: $\Delta Y_{i,t} = \alpha_1 \Delta SVI_{i,t} + \alpha_2 |R_{i,t}| + \alpha_3 \Delta Vol_{i,t} + \gamma_t + \epsilon_{i,t}$, wherein $\Delta Y_{i,t} = Y_{i,t} - Y_{i,t-1}$. The set of explanatory variables is made of the SVI (in difference) for stock i at month t ($\Delta SVI_{i,t} = SVI_{i,t} - SVI_{i,t-1}$), the absolute return for stock i at month t ($R_{i,t}$), the differencied market trading volume for stock i at month t ($\Delta Vol_{i,t} = Vol_{i,t} - Vol_{i,t-1}$), time-fixed effects ($\delta_t$), and an error term ($\epsilon_{i,t}$). In Panel B, we report the results of Equation 7b: $Y_{i,t} = \alpha_1 SSVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$, wherein $SSVI_{i,t} = \frac{SVI_{i,t} - \mu_{SVI_i}}{\sigma_{SVI_i}}$, where μ_{SVI_i} and σ_{SVI_i} are respectively the mean and the standard deviation of SVI for stock i over the sample period. In Panel C, we report the results of Equation 7c: $Y_{i,t} = \alpha_1 LN(SVI_{i,t}) + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$, wherein $LN(SVI_{i,t})$ is the natural logarithm of SVI. N is the number of observations and R^2 is the R-square. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity. The results are similar when considering the signed trade-based variables (purchases and sales). These unreported findings are available upon request.

Finally, we construct a new trade-based measure that replicate the construction of SVI, i.e. we divide each observation by the maximum value observed over the sample period for stock i and multiply it by 100. We call this measure the Trade Volume Index (TVI), which is defined as follows:

$$TVI_{i,t}^{Y} = \frac{Y_{i,t}}{max(Y_{i,t})} * 100 \tag{9}$$

where $Y_{i,t}$ is one of the 9 monthly trade-based variables defined in Section 3.4. We then estimate the following model:

$$TVI_{i,t}^{Y} = \alpha_1 SVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t} \tag{10}$$

wherein the set of explanatory variables is identical to Equation 2. The results are reported in Table 12 and show that the relationship between our TVI and SVI is positive and significant. We conclude that our results are not affected by how we measure trading activity.

Table 12: Trade Volume Index

	$TVI_{i,t}^{N}$		$TVI_{i,t}^{Q}$		$TVI_{i,t}^{V}$	
$SVI_{i,t}$	0.0253	***	0.0119	**	0.0220	***
$Vol_{i,t}$	0.0096	***	0.0092	***	0.0075	***
$\|R_{i,t}\|$	29.6588	***	26.1575	***	22.2156	***
γ_i	YES		YES		YES	
δ_t	YES		YES		YES	
N	43,200		43,200		43,200	
R^2	49.53%		37.41%		39.55%	

This table reports the results of Equation 10: $TVI_{i,t}^{Y} = \alpha_1 SVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$. The set of explanatory variables is made of the SVI for stock i at month t, the market trading volume for stock i at month t, the absolute return for stock i at month t ($|R_{i,t}|$), stock- and time-fixed effects (γ_i and δ_t, respectively), and an error term ($\epsilon_{i,t}$). N is the number of observations and R^2 is the R-square. ***, **, * indicates significance at 1%, 5%, 10%, respectively. Standard errors are robust to heteroskedasticity. The results are similar when considering the signed trade-based variables (purchases and sales). These unreported findings are available upon request.

6 Conclusion

When investors desire to buy some stocks, they face an increasingly complex search problem, which is paradoxical since information has never been so easy to access. Their limited capacity of both attention and analysis leads them to focus on stocks that have recently caught their attention. In that context, a growing body of literature relies on search engines data, and most specifically the Google SVI, to measure investor attention. Most of the studies document a positive relationship between SVI and market indicators, such as market trading volume, stock returns, or volatility.

In this paper, we focus on a sample of Belgian retail investors to investigate the relationship between SVI (restricted to queries sent from Belgium only) and trading activity. We bring empirical evidence that SVI is positively related to our retail investors' trading activity. We estimate this relationship on several sub-samples of investors, identified upon either socio-demographics (age, gender, education, language) or individual subjective characteristics (risk aversion and financial literacy). The relationship between SVI and retail trading activity is positive and significant across all our models, which confirms that SVI is a relevant proxy of investor attention.

Using signed trades (to differentiate purchases from sales), we find no evidence of a stronger relationship between attention and purchases, thereby providing no support for the buying pressure hypothesis.

When addressing the causality issue between attention and trading, we find that the past level of attention (retail trading activity) helps explain our retail investors' trading activity (attention). However, the contemporaneous effects are economically stronger and predominate in both cases, pointing to a bivariate relationship between attention and retail trading activity.

We finally run several robustness checks to show that our results are robust to both various specifications of SVI and different measures of retail trading activity.

This paper brings several contributions to the literature since it combines SVI with retail trading activity. The data at hand allow us to get results that are not biased by any institutional trading. They also enable us to consider different sub-samples of retail investors. Furthermore, our retail data allow us to sign trading volumes to directly test the buying pressure hypothesis. Overall, this paper provides novel empirical evidence that investor attention is positively related to trading activity, beyond the market level. One shortcoming might be that our analyses are conducted at a monthly frequency. Given our retail investors' trading frequency, we could hardly use weakly or daily frequencies. This might potentially explain why we observe a bivariate causality between attention and retail trading activity.

References

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*(1), 31–56.

Andrei, D. and M. Hasler (2014). Investor attention and stock market volatility. *Review of Financial Studies 28*(1), 33–72.

Aouadi, A., M. Arouri, and F. Teulon (2013). Investor attention and stock market activity: evidence from France. *Economic Modelling 35*, 674–681.

Bank, M., M. Larch, and G. Peter (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management 25*(3), 239–264.

Barber, B. M. and T. Odean (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *Journal of Finance 55*(2), 773–806.

Barber, B. M. and T. Odean (2001a). Boys will be boys: Gender, overconfidence, and common stock investment. *Quarterly Journal of Economics 116*(1), 261–292.

Barber, B. M. and T. Odean (2001b). The internet and the investor. *Journal of Economic Perspectives 15*(1), 41–54.

Barber, B. M. and T. 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.

Bellofatto, A., C. D'Hondt, and R. De Winne (2018). Subjective financial literacy and retail investors' behavior. *Journal of Banking & Finance 92*, 168–181.

Beugelsdijk, S. and B. Frijns (2010). A cultural explanation of the foreign bias in international asset allocation. *Journal of Banking & Finance 34*(9), 2121–2131.

Bijl, L., G. Kringhaug, P. Molnár, and E. Sandvik (2016). Google searches and stock returns. *International Review of Financial Analysis 45*, 150–156.

Black, F. (1986). Noise. *Journal of Finance 41*(3), 528–543.

Campbell, J. Y. (2006). Household finance. *Journal of Finance 61*(4), 1553–1604.

Choi, J. J., D. Laibson, and A. Metrick (2002). How does the internet affect trading? Evidence from investor behavior in 401 (k) plans. *Journal of Financial Economics 64*(3), 397–421.

Christelis, D., T. Jappelli, and M. Padula (2010). Cognitive abilities and portfolio choice. *European Economic Review 54*(1), 18–38.

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

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

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). Noise trader risk in financial markets. *Journal of Political Economy 98*(4), 703–738.

Dhar, R. and N. Zhu (2006). Up close and personal: An individual level analysis of the disposition effect. *Management Science 52*(5), 726–740.

D'Hondt, C. and P. Roger (2017). Investor sentiment and stock return predictability: the power of ignorance. *Finance 38*(2), 7–37.

Dimpfl, T. and S. Jank (2016). Can internet search queries help to predict stock market volatility? *European Financial Management 22*(2), 171–192.

Ding, R. and W. Hou (2015). Retail investor attention and stock liquidity. *Journal of International Financial Markets, Institutions and Money 37*, 12–26.

Dorn, D. and G. Huberman (2005). Talk and action: What individual investors say and what they do. *Review of Finance 9*(4), 437–481.

Drake, M. S., D. T. Roulstone, and J. R. Thornock (2012). Investor information demand: evidence from Google searches around earnings announcements. *Journal of Accounting Research 50*(4), 1001–1040.

Dzielinski, M. (2012). Measuring economic uncertainty and its impact on the stock market. *Finance Research Letters 9*(3), 167–175.

Goddard, J., A. Kita, and Q. Wang (2015). Investor attention and FX market volatility. *Journal of International Financial Markets, Institutions and Money 38*, 79–96.

Goetzmann, W. N. and A. Kumar (2008). Equity portfolio diversification. *Review of Finance 12*(3), 433–463.

Graham, J. R., C. R. Harvey, and H. Huang (2009). Investor competence, trading frequency, and home bias. *Management Science 55*(7), 1094–1106.

Grinblatt, M. and M. Keloharju (2001). How distance, language, and culture influence stockholdings and trades. *Journal of Finance 56*(3), 1053–1073.

Hackethal, A., M. Haliassos, and T. Jappelli (2012). Financial advisors: A case of babysitters? *Journal of Banking & Finance 36*(2), 509–524.

Haliassos, M. and C. C. Bertaut (1995). Why do so few hold stocks? *Economic Journal 105*(432), 1110–29.

Hamid, A. and M. Heiden (2015). Forecasting volatility with empirical similarity and Google Trends. *Journal of Economic Behavior & Organization 117*, 62–81.

Heyman, D., M. Lescrauwaet, and H. Stieperaere (2019). Investor attention and short-term return reversals. *Finance Research Letters 29*, 1–6.

Hirshleifer, D. and S. H. Teoh (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics 36*(1), 337–386.

Hoffmann, A. O., T. Post, and J. M. Pennings (2013). Individual investor perceptions and behavior during the financial crisis. *Journal of Banking & Finance 37*(1), 60–74.

Jacobs, H. and M. Weber (2011). The trading volume impact of local bias: evidence from a natural experiment. *Review of Finance 16*(4), 867–901.

Joseph, K., M. B. Wintoki, and Z. Zhang (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: evidence from online search. *International Journal of Forecasting 27*(4), 1116–1127.

Kimball, M. and T. Shumway (2006). Investor sophistication, and the participation, home bias, diversification, and employer stock puzzles. *Ann Arbor: Unpublished Manuscript, University of Michigan*.

Kostopoulos, D. and S. Meyer (2018). Google search volume and individual investor trading. Working paper.

Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: quantifying the relationship between phenomena of the Internet era. *Scientific reports 3*(3415), 1–7.

Li, X., J. Ma, S. Wang, and X. Zhang (2015). How does Google search affect trader positions and crude oil prices? *Economic Modelling 49*, 162–171.

Lusardi, A. and O. S. Mitchell (2014). The economic importance of financial literacy: Theory and evidence. *Journal of Economic Literature 52*(1), 5–44.

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

Moussa, F., O. BenOuda, and E. Delhoumi (2017). The use of open source Internet to analysis and predict stock market trading volume. *Research in International Business and Finance 41*, 399–411.

Nofsinger, J. R. (2005). Social mood and financial economics. *Journal of Behavioral Finance 6*(3), 144–160.

Panagiotidis, T., T. Stengos, and O. Vravosinos (2019). The effects of markets, uncertainty and search intensity on bitcoin returns. *International Review of Financial Analysis 63*, 220–242.

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

Peress, J. (2014). The media and the diffusion of information in financial markets: evidence from newspaper strikes. *Journal of Finance 69*(5), 2007–2043.

Shiller, R. J., S. Fischer, and B. M. Friedman (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity 1984*(2), 457–510.

Sims, C. A. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*(3), 665–690.

Swamy, V. and M. Dharani (2019). Investor attention using the Google search volume index–impact on stock returns. *Review of Behavioral Finance 11*(1), 55–69.

Takeda, F. and T. Wakao (2014). Google search intensity and its relationship with returns and trading volume of Japanese stocks. *Pacific-Basin Finance Journal 27*, 1–18.

Thompson, R. B., C. Olsen, and J. R. Dietrich (1987). Attributes of news about firms: an analysis of firm-specific news reported in the Wall Street Journal index. *Journal of Accounting Research 25*(2), 245–274.

Van Rooij, M., A. Lusardi, and R. Alessie (2011). Financial literacy and stock market participation. *Journal of Financial Economics 101*(2), 449–472.

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

Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking & Finance 41*, 17–35.

Ying, Q., D. Kong, and D. Luo (2015). Investor attention, institutional ownership, and stock return: Empirical evidence from china. *Emerging Markets Finance and Trade 51*(3), 672–685.

Zhang, W., D. Shen, Y. Zhang, and X. Xiong (2013). Open source information, investor attention, and asset pricing. *Economic Modelling 33*, 613–619.

7 Appendix

Table A1: Literature review

Author	Time window	Frequency	Assets	Content of the query	SVI (modification)
Da et al. (2011)	Jan. 2004 – June 2008	W	Russell 3000 index	Ticker, Company name, Main product	SVI and ASVI
Bank et al. (2011)	Jan. 2004 – June 2010	W / M	XETRA-listed stocks	Name of the firm	Δ SVI
Joseph et al. (2011)	Jan. 2005 – Dec. 2008	W	S&P500 stocks	Ticker	/
Dzielinski (2012)	Jan. 2005 – June 2011	W	S&P500 index	"economy"	ΔSVI
Vlastakis and Markellos (2012)	Jan. 2004 – Oct. 2009	W	30 NYSE and NASDAQ stocks	Company name	Detrending procedure
Aouadi et al. (2013)	Jan. 2004 – June 2009	W	CAC40 stocks	Company name and "CAC40"	LN(SVI)
Takeda and Wakao (2014)	Jan. 2008 – Dec. 2011	W	NIKKEI225	Company name	
Vozlyublennaia (2014)	Jan. 2004 – Dec. 2012	W	6 asset indexes	6 keywords	LN(SVI)
Ding and Hou (2015)	Jan. 2004 – Dec. 2009	W	S&P500	Ticker	ASVI
Da et al. (2015)	Jan. 2004 – Dec. 2011	D	several indices	118 economic terms	ΔSVI
Goddard et al. (2015)	Jan. 2004 – Sep. 2011	W	Currencies	Currency (symbol and name)	Deseasonalized
Hamid and Heiden (2015)	Jan. 2004 – Oct. 2013	W	Dow Jones	"Dow"	
Bijl et al. (2016)	Jan. 2008 – Dec. 2013	W	S&P500	Company name	SSVI
Dimpfl and Jank (2016)	July 2006 – Dec. 2011	D / W	DJIA	"Dow"	LN(SVI)
Heyman et al. (2019)	Jan. 2004 – Dec. 2016	W	S&P500	Ticker	ASVI
Kostopoulos and Meyer (2018)	July 2005 – June 2015	D	retail investors' trades	198 economic terms	ΔSVI

Panagiotidis et al. (2019)	July 2010 – Aug. 2018	D	Bitcoin	"bitcoin"	trend adjustment
Swamy and Dharani (2019)	July 202 – June 2017	W	NIFTY50	Company name	SSVI
This study	Jan. 2004 - Mar. 2012	M	BEL20, CAC40, AEX25, NASDAQ100 and S&P500 stocks	Ticker	SVI, SSVI, and LN(SVI)

This Table summarizes some characteristics of the empirical studies mentioned in Section 2. M stands for monthly, W for weekly, and D for daily respectively. Readers who are interested in getting more details about these studies are invited to directly refer to the studies.

Table A2: Sample of stocks - Top 5 per market index

Number of trades	Company name
BEL20	
118,975	KBC
42,432	Bekaert
25,083	Proximus
24,108	Delhaize Group
23,944	Telenet Group
SBF120 (including CAC40)	
38,684	Engie
28,632	BNP Paribas
24,123	Total
20,129	AXA
17,779	Vallourec
AEX25	
19,148	Aegon
14,937	Tomtom
10,655	Sbm Offshore
7,885	Airbus
7,549	Koninklijke KPN
NASDAQ100 and S&P500	
22,863	Apple Computer
17,341	Pfizer
9,946	Bank of America
9,714	Microsoft Corporation
8,920	Intel Corporation

This table lists the five most traded stocks in our sample per market index.