Journal of Behavioral and Experimental Finance 29 (2021) 100453

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Finance

journal homepage: www.elsevier.com/locate/jbef

Full length article

Googlization and retail trading activity☆

Christophe Desagre *, Catherine D'Hondt

UCLouvain, Louvain School of Management & Louvain Finance (LIDAM), 151 Chaussée de Binche, 7000 Mons, Belgium

ARTICLE INFO

Article history:
Received 29 June 2020
Accepted 17 December 2020
Available online 25 December 2020

JEL classification:
D83
G11
G40

Keywords:
Investor attention
Google SVI
Retail investors

ABSTRACT

A large body of literature documents a positive relationship between the Google Search Volume Index (SVI) and market returns or volumes. Such findings are consistent with a buying pressure due to increased attention. Unlike most of the studies that use market data, we use the trading accounts for a sample of retail investors. The advantage is twofold; we are able to disentangle purchases from sales, and our results are not biased by any institutional trading. We find that the relationship between the SVI and retail trading activity is positive but not stronger for purchases than for sales. We also demonstrate evidence of a bidirectional causality between attention and trading activity, though contemporaneous effects predominate. Our results are robust to controls based on sociodemographics or subjective investor characteristics, as well as various specifications of the SVI and different measures of trading activity.

1. Introduction

The Internet has dramatically facilitated access to information for investors. One drawback is information overload,[1] which may have nontrivial 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) demonstrate evidence that retail investors are net buyers of attention-grabbing stocks. The rationale is that their decision to buy and to sell are fundamentally different, i.e., buyers have to choose from a large set of available securities, while sellers can only sell what they already own.[2] Hence, increased attention is assumed to lead to a temporary buying pressure that subsequently results in positive returns.

Measuring investor attention is an empirical challenge. Several indirect measures based on news or media-related proxies (e.g., length of headlines, positive and negative tone of an article, analysts' recommendations and coverage, etc.) are used in the literature. Their main shortcoming is that they reflect information supply, not demand. Such measures allow instead inference of the reach and impact of news on the market but relating abnormal volume to higher investor attention is then nearly tautological (Barber and Odean, 2008). Moreover, abnormal market volume or extreme returns can be driven by liquidity-motivated large trades of institutional investors, which makes such market-based measures noisy.

Alternatives to measure investor attention have recently emerged with the pervasive use of the 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, the SVI allows direct capture of the aggregate 'active' demand for information, either on a per-stock or per-market index basis. In addition, the SVI does not depend on market data.

Da et al. (2011) pioneered the use of the SVI as a direct measure of investor attention. According to these authors, the *'SVI captures the attention of individual investors'* and *'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) claim that *'search*

☆ 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, Gunther Wuyts, and participants at the 3L workshop (November 2017 - Brussels), 35th AFFI Conference (May 2018 - Paris), 16th Belgian Financial Research Forum (June 2018 - Brussels), Warsaw International Economic Meeting (July 2018 - Warsaw), and 3rd European Capital Market Workshop (July 2019 - Dublin) for helpful comments. Any errors are the full responsibility of the authors.

* Corresponding author.
E-mail addresses: christophe.desagre@uclouvain.be (C. Desagre), catherine.dhondt@uclouvain.be (C. D'Hondt).

[1] Theoretically, more available information is valuable for investors but only if they are able to make relevant analysis of it (Barber and Odean, 2001b).

[2] Most of the time, retail investors are banned from short-selling.

[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.

https://doi.org/10.1016/j.jbef.2020.100453

volume is indeed a powerful measure of investor recognition.' These two papers paved the way for many empirical investigations based on the SVI as an investor attention proxy.[4] Most of these investigations relate the SVI to stock market returns (Bank et al., 2011; Joseph et al., 2011; Takeda and Wakao, 2014; Bijl et al., 2016; Heyman et al., 2019), volatility (Vlastakis and Markellos, 2012; Vozlyublennaia, 2014; Hamid and Heiden, 2015; Kim et al., 2019), liquidity (Aouadi et al., 2013; Ding and Hou, 2015), and/or trading volume (Bank et al., 2011; Joseph et al., 2011; Takeda and Wakao, 2014; Vlastakis and Markellos, 2012; Aouadi et al., 2013).[5]

Among the above literature, the relationship between the SVI and stock trading volume is of particular interest for the present paper. Empirical evidence clearly points to a positive relationship between market trading volume and the SVI (be it defined at the stock level or at the market index level), which provides support to the buying pressure hypothesis. All of these findings however rely on market activity (or part of it), from which retail trading volume can only be inferred (e.g., with a focus on stocks that are (supposed to be) the most traded by retail investors). In contrast with this approach, we directly relate the SVI to retail trading activity. Specifically, we use the trading accounts of a sample of Belgian retail investors to investigate the relationship between their trading activity and the SVI, which is restricted to all of the queries sent from Belgium only.[6] Our sample is made of 455 large cap stocks (mainly US, French, Dutch, and Belgian stocks) traded by a set of 42,731 retail investors over the period January 2004–March 2012. Hence, we contribute to the literature with a twofold advantage over previous work. First, our retail data enable us to sign trading volumes. Distinguishing purchases from sales, we are able to test the buying pressure hypothesis. Based on the latter, the relationship between the SVI and trading activity is expected to be stronger for purchases. Second, since we use retail data, our results may not be biased by any institutional trading. We should add that we analyze a large sample of stocks over a long time window, while previous empirical findings are often restricted to a few stocks, a single market index, and/or a short time period.

To further contribute to the literature, we investigate whether the SVI helps explain the trading activity of subsamples of retail investors determined by individual characteristics. Building on Da et al. (2011), who suggest that the SVI likely captures the attention of less sophisticated retail investors, we consider either sociodemographic 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). Finally, we also investigate the dynamics between the SVI and retail trading, i.e., whether attention does cause trading volume and vice-versa.

In a recent paper, Kostopoulos et al. (2020) use retail data to relate the SVI to the trading activity of a sample of German investors. Nevertheless, there are two key differences between their work and the present paper. First, using the SVI based on each stock ticker,[7] 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). Second, by construction, the FEARS index provides a measure of pessimism (Kostopoulos et al., 2020, p. 2) and therefore aims at measuring retail investor sentiment.[8] In this paper, the SVI based on each stock ticker delivers a measure of attention, which is neutral by definition.

Our main findings can be summarized as follows. Consistent with the extant literature, we find that the relationship between the SVI and our retail trading activity is positive. This further confirms that the SVI is a reliable proxy for retail investor attention. However, our results do not provide evidence that this relationship is stronger for purchases than for sales, thereby providing no support for the buying pressure hypothesis. In our sample, increased attention is associated with higher retail trading volume on both market sides. The most plausible explanation is information-based portfolio monitoring: retail investors search for information when they wish to buy some shares (i.e., when looking for investment opportunities) but also when they monitor their stock portfolios (i.e., when rebalancing their portfolios). Next, our findings still hold when controlling for some sociodemographics or subjective investor characteristics. Whatever the subsample of investors, the positive relationship between attention and trading activity still holds, although both the value of the SVI coefficient and the explanatory power of the model fluctuate. The marginal effect of the SVI looks stronger for men, for younger investors, for low education investors, for risk-tolerant investors, or for high-literate investors. For the dynamics between the SVI and trading volume, we document a bidirectional causality, although the contemporaneous effects are economically stronger and predominate in both cases. Our results are robust to various specifications of the SVI as well as to different measures of trading activity.

The remainder of this paper is organized as follows. Section 2 presents our data and sample. Section 3 reports our empirical results. Several robustness checks are provided in Section 4. Section 5 concludes.

2. Data and sample

2.1. Sample of stocks

For the purpose of this paper, we use available trading accounts of Belgian retail investors (which will be described in Section 2.2). Therefore, 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,[9] we know that Belgian retail investors tend to focus most of their trading activity on Belgian, US, French, and Dutch stocks.[10] Hence, we target 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 with the available retail trade data, we end up with a sample of 455 stocks. We count 331 US stocks, 86 French stocks, 18 Dutch stocks, and 10 Belgian stocks. The remaining 10 stocks were issued in various countries.

4 Table A.1 available in the appendix provides a big picture of the main empirical studies.

5 Other variables have been investigated, but to a lesser extent (e.g., the breadth of ownership Ding and Hou, 2015). Lately, the SVI has also been related to returns on currencies (Goddard et al., 2015), crypto-currencies (Panagiotidis et al., 2019), and commodities (Li et al., 2015).

6 This choice is made because of the availability of the retail data. The online brokerage house that provided us with these data was the leader on the market at the time of the sample period (2004–2012). There is no country-specific expectation.

7 We motivate this choice in Section 2.3.

8 Da et al. (2015, p. 1) indicate that *'the FEARS index is a new measure of investor sentiment.'* FEARS stands for Financial and Economic Attitudes Revealed by Search. This index is constructed by aggregating the SVI of economic terms (e.g., 'recession,' 'unemployment,' 'bankruptcy,' 'financial crisis,' etc.). In the same vein, Dimpfl and Kleiman (2019) extend the work of Da et al. (2015) to compute a pessimism index.

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

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

Table A.2 (in Appendix) reports the five most traded stocks in our retail data, depending on the market index. For each stock in the sample, we get historical monthly prices and market volume from Bloomberg for the period under scrutiny, namely, January 2004–March 2012. Previous empirical work is often restricted to a few stocks, a single market index, and/or a tiny time window. By contrast, we analyze both a large sample of stocks and a long time period.

2.2. Sample of investors

Our retail data come from an online Belgian brokerage house. They cover a large set of trading accounts over the 99-month period (i.e., January 2004–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 mentioned above, 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 a large set of 42,731 retail investors.

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

For each investor, we have matched additional data that we classify as either sociodemographic or subjective characteristics. Sociodemographics 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.[12] 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 a university degree or equivalent.[13] 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 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. The presence of several spoken languages does not lead to any country-specific expectation in this paper.[14]

The 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.[15] 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 and their investment objectives as well as their financial capacity.[16] Since these MiFID data are survey-based, they are only available for the investors who completed the questionnaire, that is, 20,119 investors (i.e., 47%).

Building on Da et al. (2011) who suggest that the SVI likely captures the attention of less sophisticated retail investors, we first focus on financial literacy. For the latter, investors were required to self-assess their financial knowledge using three available options: no knowledge, average knowledge, and good knowledge. 55% of the investors believe they have average knowledge, while 30% declare that they have good knowledge. Only 15% self-report no knowledge. Subjective financial literacy helps explain cross-sectional variations in retail investor behavior (Bellofatto et al., 2018). Thanks to the data at hand, we are able to check whether the relationship between attention and trading activity is affected by the level of self-declared literacy. In addition, we also consider subjective risk aversion. Vlastakis and Markellos (2012) examine the relationship between risk aversion and investor attention. However, these authors rely on a measure of variance risk premium computed at the market level, which does not allow one to check for individual differences. By contrast, we have an individual risk aversion measure since our 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.

2.3. 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 (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 provides a less noisy measure of the demand for financial information because the measure is less ambiguous.[17] On the other hand, one could argue that retail investors are likely unsophisticated investors who might not know the stock ticker.[18] In addition, when investors are sophisticated, they might prefer websites specialized in financial data when looking for stock-specific information.[19] Despite the lack of consensus, we opt for the less noisy proxy, i.e., the SVI defined on the stock ticker. Our retail investors are used to trading online, and we focus on 455 stocks that are supposed to be quite familiar to them (since each stock is part of a well-known market index). We can reasonably assume our investors are likely to know (at least some of) the tickers.

11 When necessary, we use historical exchange rates to convert monetary volumes into euros.

12 Age is determined in 2012 using the available year of birth.

13 This information is missing for some investors.

14 As argued by Bukovina (2016, p. 25), *'In the field of behavioral finance, social media big data is currently predominantly employed for modeling of sentiment in the US market. Therefore, it would be interesting to see the broader discussion about the presence of investor demand and sentiment also on other stock markets'*. From that point of view, we complement previous work dealing with markets outside the US (e.g., Taiwan Chen and Lo, 2019, Portugal Oliveira-Brochado, 2019, India Swamy and Dharani, 2019).

15 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 implemented in January 2018 (known as MiFID II (2014/65/UE)). For more details, please visit the European Commission website (https://ec.europa.eu/info/law/markets-financial-instruments-mifid-ii-directive-2014-65-eu_en).

16 Such items are usually covered in Investment Policy Statements (IPS) used in portfolio management delegation. 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).

17 Da et al. (2011) report that the correlation between the SVI defined on the ticker and the SVI defined on the firm name is close to 10%.

18 Aouadi et al. (2013) indicate that French investors are more likely to use the stock name when searching for stock-specific information on Google.

19 During the sample period, the online brokerage house provided its retail clients with free access to an investment advice tool on stocks through its web platform.

Table 1
Descriptive statistics.

Variable	Minimum	1st Quart.	Median	Mean	3rd 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, 1st quartile, median, mean, 3rd quartile, and maximum values as well as the standard deviation. Panel A provides market-based variables for our sample of 455 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 millions 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 aggregate trading activity across the 455 stocks. 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).

Using each stock ticker, we download the monthly SVI for the period from January 2004 to March 2012. Given our retail data, we restrict our Google data request to queries sent from Belgium, assuming that the corresponding SVI is a relevant proxy of attention for our sample of investors.[20] 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 to the highest number of queries, while all other values are scaled to that maximum.[21]

For our sample period, which spans 99 months, Google Trends automatically provides the SVI at a monthly frequency. By restricting the time period, it would be possible to get the weekly or daily SVI but at the cost of losing comparability among subperiods of time. Choosing an appropriate frequency is a key decision, but this choice is rarely documented.[22] In our case, working with the monthly SVI seems appropriate given the trading frequency of our retail investors (see Section 2.2).

2.4. Measures of trading activity

To measure our retail 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. Next, we 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 $V^{B/S}_{i,t}$ to the monetary volume bought/sold. These signed trade-based variables will allow us to 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, the SVI in Panel B, and trade-based variables in Panel C. All of the statistics are computed across stocks and months. Panel A reveals an average stock price of 43.26 euros and a monthly return that is 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. In fact, all of the variables in Panel C exhibit positively skewed distributions since means are higher than the corresponding medians (and even higher than the upper quartile). Only 72 stocks are traded every month during the entire sample period.

To control for either some sociodemographic characteristics (which 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 subsamples of retail investors and replicate on them the above aggregate trade-based variables.

3. Empirical analysis

A large body of papers document a positive relationship between the SVI and stock market trading volume (e.g., 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:

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

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 stock i return 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 as in Bank et al. (2011). Given that the SVI data are already normalized (see Section 2.3), we use the SVI in levels. We show in Section 4 that this choice does not impact our results.

Table 2 reports the expected relationship between the SVI and market trading activity: the higher the SVI for a stock at a given month, the higher is the corresponding market volume. The market trading volume also 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.

20 Investor nationality is not available, but we can reasonably assume that most of the investors are Belgian or based in Belgium.

21 We use the R package **gtrends** to download the SVI. When the 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).

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

Table 2
Investor attention and market trading volume.

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

This table reports the results for Eq. (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%, and 10%, respectively. Standard errors are robust to heteroskedasticity.

3.1. Analysis of retail trading activity

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

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

wherein $Y_{i,t}$ is a (signed) trade-based variable defined in Section 2.4 to measure the aggregate retail trading activity on stock i at month t. The set of explanatory variables is made using 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)[23] return for stock i at month t ($R_{i,t}$), stock- and time-fixed effects (γ_i and δ_t, respectively),[24] and an error term ($\epsilon_{i,t}$).

Table 3 reports the results, where Panel A refers to unsigned trading activity, and Panel B (C) refers to purchases (sales). Whatever the dependent variable, the SVI displays a positive and significant coefficient at the 1% level. This positive relationship between attention and retail trading activity is valid on both unsigned and signed trade-based measures. A higher attention is associated with a higher trading activity from retail investors on both market sides. This means that when attention is higher, retail investors execute more purchases but also more sales. For example, when the SVI increases by 1, the number of trades increases on average by 0.27, and the number of purchases and sales increase on average by 0.15 and 0.12, respectively. Such effects are relatively modest, though they should be put in perspective with the level of trading activity in our sample (documented in Table 1).

Table 3 shows that retail trades are also positively related to market volume.[25] More interestingly, Panels B and C reveal the contrasting relationship between returns and signed trades, i.e., the number of purchases (sales) tends 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).

3.2. Does the relationship to attention truly differ between purchases and sales?

Table 3 displays coefficient estimates for SVI that are higher for purchases than for sales. This indicates a stronger marginal effect of the SVI 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 is the case in our sample. Based on that fundamental difference, increased attention is assumed to lead to temporary buying pressure (Barber and Odean, 2008). To address this hypothesis, 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 Eq. (3), $Y_{i,t}$ represents either aggregate retail purchases or sales. The set of explanatory variables is similar to the one of Eq. (2), except that we add a dummy variable (B) equal to 1 for purchases and an interaction variable ($B^* SVI_{i,t}$).[26] These two variables allow us to check (i) whether retail purchases are higher than sales, and (ii) whether the relationship between trading activity and attention is statistically stronger for purchases compared to sales.

The results are provided in Table 4. When the dependent variable is defined in number of trades or monetary volume, the coefficient of the dummy variable (B) is positive and significant. This confirms that there are more purchases than sales in our sample. Retail investors are indeed net buyers of stocks, which is well-established in the literature. More importantly, the coefficient of the interaction variable ($B^* SVI_{i,t}$) is positive but never significant. This finding provides no support for the buying pressure hypothesis. In our sample, increased attention is associated with higher trading volume on both market sides. The most plausible explanation is that retail investors search for information when they wish to buy some shares (i.e., when looking for investment opportunities) but also when they monitor their stock portfolios (i.e., when considering any portfolio rebalancing). Information-based monitoring could explain the positive relationship between attention and sales. Despite being less likely, our findings might be due to the monthly frequency of our analysis. For investors who execute round trip-trades on the same stock within the same month, the link to attention would deserve an analysis at a higher frequency (e.g., with the weekly SVI on shorter time windows). Further research based on higher frequencies would provide additional insights for such active retail investors.

3.3. Sociodemographic characteristics

Sociodemographic characteristics such as gender, age, and education are typical 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 (e.g., Haliassos and Bertaut, 1995, Campbell, 2006, Van Rooij et al., 2011). Whether the relationship between attention and trading activity depends on such individual characteristics has however never been addressed. To fill the gap, we estimate Eq. (2) on six different subsamples of retail investors determined upon the above sociodemographic characteristics. Specifically, for each characteristic under scrutiny, we split our sample of retail investors into two subsamples. For gender, we separate men from women. Using the median age (available in Section 2.2), we divide our investors into two equal groups, which allows us to flag any investor as 'Young' (below the median

23 We take returns in absolute value when the dependent variable is not signed.
24 We also 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. The results are qualitatively similar and are available upon request.
25 Our results remain similar when we consider the market volume net of our retail trading activity.
26 This explain why the value of N in Table 4 is twice the value of the N in Table 3.

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 for Eq. (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 Section 2.4 to measure the aggregate retail 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 absolute returns in Panel A (unsigned trading activity) and returns 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.

Table 4
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 for Eq. (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 Section 2.4 to measure the aggregate retail 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.

age) or 'Old' (above the median age). As far as education is concerned, we distinguish investors who hold a 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 consider three other subsamples based on spoken language.

The results for gender, age, and education are provided in Table 5, in Panels A, B, and C, respectively.[27] Whatever the subsample of investors, the coefficient estimate of the SVI is positive and significant at the 1% level. The positive relationship between attention and trading activity still holds, although both the value of the SVI coefficient and the R^2 fluctuate across models. In Panel A, the marginal effect of the SVI appears much stronger for men (0.40 versus 0.09). Panel B exhibits a stronger marginal effect of the SVI for younger investors (0.33 versus 0.19). For Panel C, it reveals a stronger marginal effect of the SVI for low education investors (0.19 versus 0.04). If education is correlated with sophistication, this result provides some support to Da et al. (2011), who report that the SVI is a relevant proxy of attention especially for less sophisticated investors. The highest R^2 is observed for 'Old' investors in Panel B.

Table 6 provides the results for the three subsamples based on spoken language. The results show that attention is positively related to retail trading activity, whatever the spoken language. The marginal effect of the SVI 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 for the subsample of French-speaking investors.

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 5
Investor attention and retail trading activity — sociodemographics.

	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 for Eq. (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 sociodemographic characteristic, we split our sample of retail investors into two subsamples. For gender (Panel A), we separate men from women. For age (Panel B), using the median age (available in Section 2.2), we divide our investors into two equal groups 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 a 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%, and 10%, respectively. Standard errors are robust to heteroskedasticity. The results are reported here for the dependent variable defined as the number of trades. The results are similar when considering the number of shares traded or the monetary volume. These unreported findings are available upon request.

Table 6
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 for Eq. (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 retail investors according to their spoken language, 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%, and 10%, respectively. Standard errors are robust to heteroskedasticity. The results are reported here for the dependent variable defined as the number of trades. The results are similar when considering the number of shares traded or the monetary volume. These unreported findings are available upon request.

3.4. 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 the trading behavior.[28] On the other hand, Da et al. (2011) report that the SVI is likely to capture the attention of the less sophisticated retail investors.[29] Hence, it appears relevant to check whether the relationship between attention and retail trading activity is affected by financial literacy. For that purpose, we build two subsamples of investors depending on their level of financial literacy. Specifically, we focus on the extreme levels to get a subsample of low-literate investors (those who declare 'no knowledge' about financial markets) and another subsample of high-literate investors (those who select 'good knowledge').

We adopt a similar approach to consider investor subjective risk aversion. To examine the relationship between risk aversion and investor attention, Vlastakis and Markellos (2012) rely on a measure of variance risk premium.[30] The latter is however computed at the market level and does not allow one 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 (see Section 2.2), we are able to build two subsamples aiming at distinguishing risk-averse investors (those who selected one of the two lowest levels on the scale) and risk-tolerant investors (those who chose the highest level).[31]

As in Section 3.3, we estimate Eq. (2) on each of the above subsamples. We provide the results in Table 7, in Panel A for risk aversion and in Panel B for financial literacy. Whatever the subsample of investors, the SVI always exhibits a positive and significant coefficient. In Panel A, the marginal effect of the SVI appears stronger for risk-tolerant investors (0.13 versus 0.04). The

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

29 These authors use the trading volumes executed on different execution venues and the Dash-5 reports to infer monthly changes in orders and turnover from retail investors on all of the Russell 3000 stocks.

30 These authors 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.

31 This choice is mainly motivated by the question and the scale available in the MiFID questionnaire.

R^2 is also higher when the model is estimated on this subsample of investors (40.68% versus 33.20%). Panel B reveals that the marginal effect of the SVI 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 subsample (46.71% versus 34.02%). Although financial literacy is a reliable proxy of sophistication, our findings are not consistent with Da et al. (2011). We may argue that if risk-tolerant and/or high-literate investors tend to monitor closely their portfolios and market trends in general, they are more likely to look for information when considering any rebalancing. In accordance with information-based portfolio monitoring, the marginal effect of both market volume and returns on their trading activity is also stronger.

3.5. Does attention cause trading?

All of our findings show a positive and significant relationship between attention and retail trading activity. This obviously raises the question of whether the SVI carries some predictive power for retail trading activity. We address that issue by adding a lagged term for SVI ($SVI_{i,t-1}$) into Eq. (2). More specifically, we compare an unrestricted model (UM - Eq. (4a)) that includes both the contemporaneous and lagged SVI, with three restricted models (RM), wherein $\beta_2 = 0$ (Eq. (4b)), $\beta_1 = 0$ (Eq. (4c)), and $\beta_1 = \beta_2 = 0$ (Eq. (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 of the variables are similar to those previously used in Eq. (2), but it is worth noticing that $\epsilon_{i,t}$ refers to the UM residuals, while $u_{i,t}$ refers to RM residuals. As the UM contains more variables than the RM, the R^2 is *de facto* higher. The point is to test whether this increase is statistically significant. To do so, we use an F-test, with the F-statistic computed as follows:

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

where 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.[32]

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

3.6. Does trading cause attention?

Since the past level of attention helps explain retail trading activity, we now check for any causality in the other direction, i.e., does trading cause attention? To do so, we replicate the same approach and estimate Eqs. (6a) to (6d), which are defined as follows:

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

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

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

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

The results are provided in Table 9.[33] Whatever the measure of trading activity (number of trades ($N_{i,t}^T$), number of shares ($Q_{i,t}^T$), or monetary value ($V_{i,t}^T$)), we find a positive and significant relationship with the SVI. 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 in a given month tends to increase attention up to the next month. This suggests that retail investors keep on looking for information about the stocks that they have just traded, which is consistent with information-based portfolio monitoring. This might also provide some support to attention utility, i.e., retail investors continue to pay attention to information already known (Quispe-Torreblanca et al., 2020).

4. Robustness checks

We perform four robustness checks. First, we restrict our analysis to the 372 stocks present in the sample during the whole period. By doing so, we obtain a *balanced* panel, with which we estimate Eq. (2) for the unsigned trade-based variables.[34] The results are reported in Table 10 and show that the relationship between retail trading and the SVI is still positive and significant at the 1% level. The coefficient estimates of the SVI are very close to those displayed in Table 3 (e.g., 0.28 versus 0.27; 104.41 versus 110.22; and 3,508.40 versus 3,193.10). The R^2 are also in line.

As a second robustness check, we consider different specifications of the SVI to ensure that the way the SVI is defined does not influence our results. Some authors modify the raw SVI either by taking the changes in the SVI (Dzielinski, 2012), by using a standardized SVI (Bijl et al., 2016; Swamy and Dharani, 2019), or by taking the natural logarithm of the 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.[35] We then estimate the following regression models:

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

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

32 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).

33 All of the variables were defined previously. To obtain readable coefficients, $Q_{i,t}^T$ is expressed in millions of shares and $V_{i,t}^T$ is in millions of euros.

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

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

Table 7
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 for Eq. (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 Section 2.2, we split our investors into two subsamples 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 on the scale to create a subsample 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%, and 10%, respectively. Standard errors are robust to heteroskedasticity. The results are reported here for the dependent variable defined as the number of trades. The results are similar when considering the number of shares traded or the monetary volume. These unreported findings are available upon request.

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 for Eqs. (4a)–(4d). Eq. (4a) is: $Y_{i,t} = \beta_1 SVI_{i,t} + \beta_2 SVI_{i,t-1} + a_1 Vol_{i,t} + a_2 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$. Eq. (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%, and 10%, respectively. Standard errors are robust to heteroskedasticity.

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

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 the SVI at the stock level, as follows:

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

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	***
$\|R_{i,t}\|$	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	***
$\|R_{i,t}\|$	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	***
$\|R_{i,t}\|$	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 for Eqs. (6a) to (6d). Eq. (6a) is $SVI_{i,t} = \beta_1 Y_{i,t} + \beta_2 Y_{i,t-1} + a_1 Vol_{i,t} + a_2|R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$. Eq. (6b) ((6c)/(6d)) is a restricted version of this Equation wherein β_2 =0 (β_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%, and 10%, respectively. Standard errors are robust to heteroskedasticity.

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 for Eq. (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 Section 2.4 to measure aggregate retail 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%, and 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.

where μ_{SVI_i} and σ_{SVI_i} are respectively the mean and the standard deviation of the SVI for stock i over the sample period. All of the other variables were defined previously.[36]

The results are reported in Table 11, in Panel A for Eq. (7a), in Panel B for Eq. (7b), and in Panel C for Eq. (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 aggregate retail trading activity. In Panel B, where we use a standardized SVI (SSVI), and in Panel C, where we use the natural logarithm of the SVI, we also find a positive and significant relationship between attention and retail trading activity. Consequently, our findings are robust to several specifications of the SVI.

Third, we construct a new trade-based measure that replicates the construction of the SVI, i.e., we divide each observation by the maximum value over the sample period for stock i and multiply it by 100. We call this measure the Trade Volume Index (TVI),

36 Given that the SVI may have a value of zero, we take the natural logarithm of SVI+1 as in Takeda and Wakao (2014).

Table 11
Alternative specifications of the SVI.

Panel A: Eq. (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: Eq. (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: Eq. (7c) - Natural logarithm of the 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 for Eqs. (7a) to (7c). Panel A reports the results of Eq. (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 difference in 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 Eq. (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 Eq. (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 the SVI. N is the number of observations and R^2 is the R-square. ***, **, * indicates significance at 1%, 5%, and 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.

which is defined as follows:

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

where $Y_{i,t}$ is one of the 9 monthly trade-based variables defined in Section 2.4. We then estimate the following model:

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

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

Finally, we use an instrumental variable approach. In the main model, we investigate the relationship between trading activity and investor attention, as proxied by the SVI. This variable is obtained from Google Trends. As queries are restricted to those sent from Belgium, it can be argued that investor attention is endogenous to trading activity because investors could pay more attention once they have traded. In that case, the coefficient that measures the relationship between investor attention and trading activity is not consistent and our inferences are not valid. We adopt an instrumental variable approach to counter this argument. We choose the SVI without restricting the queries from Belgium. We call this measure the 'world SVI' (WSVI). We motivate the choice of this instrument as follows: (1) there is likely a positive link between the SVI and the WSVI, and (2) trading activity in Belgium is unlikely to affect the WSVI. We estimate the following two-stage least squares regressions:

$$SVI_{i,t} = \beta_0 + \beta_1 WSVI_{i,t} + \mu_{i,t} \tag{11a}$$

$$Y_{i,t} = \alpha_1 \hat{SVI}_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t} \tag{11b}$$

The results are reported in Table 13. First, we observe that β_0, the relationship between search volume index in Belgium and search volume index in the world, is positive. Then, using the predicted values of the SVI, $\hat{SVI}$, as regressors in the second-stage model, we find that the relationship between this variable and trading activity always exhibits a positive and statistically significant coefficient. Furthermore, the coefficients are higher than those in Table 3, suggesting that the impact of attention on trading activity is even higher than previously documented.

5. Conclusion

Over the last decade, a growing literature has shown that the Google Search Volume Index (SVI) is a reliable proxy for investor attention. Empirical evidence clearly indicates a positive relationship between stock market trading volume and the SVI (be it defined at the stock level or at the market index level). Such results provide support to the buying pressure hypothesis, i.e., increased attention is assumed to lead to a temporary buying pressure. The rationale behind this hypothesis is the fundamental difference between buying and selling decisions; i.e., buyers have to choose from a large set of available securities, while sellers

Table 12
Trade Volume Index.

	$TVI^N_{i,t}$		$TVI^Q_{i,t}$		$TVI^V_{i,t}$	
$SVI_{i,t}$	0.0253	***	0.0119	**	0.0220	***
$Vol_{i,t}$	0.0096	***	0.0092	***	0.0075	***
$\lvert R_{i,t}\rvert$	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 for Eq. (10): $TVI^Y_{i,t} = \alpha_1 SVI_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$. 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%, and 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.

Table 13
Instrumental variable.

Variable dependent: SVI - first stage			
	$SVI_{i,t}$		
β_0	17.40***		
$WSVI_{i,t}$	0.68***		
N	43,004		
R^2	0.49		
Variable dependent: Trading activity - second stage			
	$N^T_{i,t}$	$Q^T_{i,t}$	$V^T_{i,t}$
$\hat{SVI}_{i,t}$	0.34***	151.87***	4,074.50***
$Vol_{i,t}$	0.03***	111.13***	747.50***
$\lvert R_{i,t}\rvert$	132.78***	101,161.97***	1,311,265.00***
γ_i	Yes	Yes	Yes
δ_t	Yes	Yes	Yes
N	43,004	43,004	43,004
R^2	46.05%	35.63%	45.60%

This table reports the results of the first-stage Equation, $SVI_{i,t} = \beta_0 + \beta_1 WSVI_{i,t} + \mu_{i,t}$, and the second stage Equation, $Y_{i,t} = \alpha_1 \hat{SVI}_{i,t} + \alpha_2 Vol_{i,t} + \alpha_3 |R_{i,t}| + \gamma_i + \delta_t + \epsilon_{i,t}$, wherein $Y_{i,t}$ is a trade-based variable defined in Section 2.4 to measure our retail investors' aggregate trading activity on stock i at month t. The set of explanatory variables is made of the predicted SVI for stock i at month t ($\hat{SVI}_{i,t}$), the world SVI for stock i at month t ($WSVI_{i,t}$), 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%, and 10%, respectively. Standard errors are robust to heteroskedasticity.

can only sell what they already own. This difference is especially relevant for retail investors, who are most of the time banned from short-selling.

Unlike most of the studies that use market data, we directly relate the SVI to retail trading activity. In particular, we use the trading accounts of a large set of Belgian retail investors to investigate the relationship between their trading activity and the SVI, which is restricted to all of the queries sent from Belgium only. Our sample is made of 455 large cap stocks (mainly US, French, Dutch, and Belgian stocks) traded by a set of 42,731 retail investors over a 99-month period, i.e., from January 2004 to March 2012. Hence, we have a twofold advantage over previous work. First, we are able to disentangle purchases from sales to directly test the buying pressure hypothesis. Based on the latter, the relationship between the SVI and retail trading activity is expected to be stronger for purchases. Second, since we use retail data, our results are not biased by any institutional trading.

Consistent with the extant literature, we find that the relationship between the SVI and our retail trading activity is positive. This further confirms that the SVI is a reliable proxy for retail investor attention. However, our results do not provide evidence that this relationship is stronger for purchases than for sales, thereby providing no support for the buying pressure hypothesis. In our sample, increased attention is associated with higher retail trading volume on both market sides. We relate this result to information-based portfolio monitoring: retail investors search for information when they wish to buy some shares (i.e., when looking for investment opportunities) but also when they monitor their stock portfolios (i.e., when rebalancing their portfolios).

In addition, we investigate whether the SVI helps explain the trading activity of subsamples of retail investors determined upon individual characteristics. Our findings still hold when controlling for some sociodemographics that are common control variables (e.g., age, gender, education, and spoken language) or some subjective characteristics that could affect trading behavior (e.g., financial literacy and risk aversion). Whatever the subsample of investors, the positive relationship between attention and trading activity still holds, although both the value of the SVI coefficient and the explanatory power of the model fluctuate. The marginal effect of the SVI looks stronger for men, for younger investors, for low education investors, for risk-tolerant investors, or for high-literate investors.

We also examine the dynamics between the SVI and retail trading, i.e., whether attention causes trading volume and vice-versa. Our results show a bidirectional causality, although the contemporaneous effects are economically stronger and predominate in both cases. Such findings suggest that retail investors keep on looking for information about the stocks that they have

Table A.1
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
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
Kostopoulos et al. (2020)	July 2005 – June 2015	D	retail investors' trades	198 economic terms	ΔSVI
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. 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 A.2
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.

just traded, which is consistent with information-based portfolio monitoring.

We finally perform several robustness checks in order to show that our results are robust to various specifications of the SVI as well as to different measures of trading activity.

Appendix

See Tables A.1 and A.2.

References

Aouadi, A., Arouri, M., Teulon, F., 2013. Investor attention and stock market activity: evidence from France. Econ. Model. 35, 674–681.

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

Barber, B.M., Odean, T., 2001a. Boys will be boys: Gender, overconfidence, and common stock investment. Q. J. Econ. 116 (1), 261–292.

Barber, B.M., Odean, T., 2001b. The internet and the investor. J. Econ. Perspect. 15 (1), 41–54.

Barber, B.M., Odean, T., 2008. All that glitters: the effect of attention and news on the buying behavior of individual and institutional investors. Rev. Financ. Stud. 21 (2), 785–818.

Bellofatto, A., D'Hondt, C., De Winne, R., 2018. Subjective financial literacy and retail investors' behavior. J. Bank. Financ. 92, 168–181.

Bijl, L., Kringhaug, G., Molnár, P., Sandvik, E., 2016. Google searches and stock returns. Int. Rev. Financ. Anal. 45, 150–156.

Bukovina, J., 2016. Social media big data and capital markets—An overview. J. Behav. Exp. Financ. 11, 18–26.

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

Chen, H.-Y., Lo, T.-C., 2019. Online search activities and investor attention on financial markets. Asia Pac. Manage. Rev. 24 (1), 21–26.

Choi, J.J., Laibson, D., Metrick, A., 2002. How does the internet affect trading? Evidence from investor behavior in 401 (k) plans. J. Financ. Econ. 64 (3), 397–421.

Christelis, D., Jappelli, T., Padula, M., 2010. Cognitive abilities and portfolio choice. Eur. Econ. Rev. 54 (1), 18–38.

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

Da, Z., Engelberg, J., Gao, P., 2015. The sum of all FEARS investor sentiment and asset prices. Rev. Financ. Stud. 28 (1), 1–32.

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

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

Dimpfl, T., Kleiman, V., 2019. Investor pessimism and the German stock market: Exploring google search queries. Ger. Econ. Rev. 20 (1), 1–28.

Ding, R., Hou, W., 2015. Retail investor attention and stock liquidity. J. Int. Financ. Mark. Inst. Money 37, 12–26.

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

Drake, M.S., Roulstone, D.T., Thornock, J.R., 2012. Investor information demand: evidence from Google searches around earnings announcements. J. Account. Res. 50 (4), 1001–1040.

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

Goddard, J., Kita, A., Wang, Q., 2015. Investor attention and FX market volatility. J. Int. Financ. Mark. Inst. Money 38, 79–96.

Goetzmann, W.N., Kumar, A., 2008. Equity portfolio diversification. Rev. Financ. 12 (3), 433–463.

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

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

Hackethal, A., Haliassos, M., Jappelli, T., 2012. Financial advisors: A case of babysitters? J. Bank. Financ. 36 (2), 509–524.

Haliassos, M., Bertaut, C.C., 1995. Why do so few hold stocks? Econom. J. 105 (432), 1110–1129.

Hamid, A., Heiden, M., 2015. Forecasting volatility with empirical similarity and Google Trends. J. Econ. Behav. Organ. 117, 62–81.

Heyman, D., Lescrauwaet, M., Stieperaere, H., 2019. Investor attention and short-term return reversals. Finance Res. Lett. 29, 1–6.

Hirshleifer, D., Teoh, S.H., 2003. Limited attention, information disclosure, and financial reporting. J. Account. Econ. 36 (1), 337–386.

Hoffmann, A.O., Post, T., Pennings, J.M., 2013. Individual investor perceptions and behavior during the financial crisis. J. Bank. Financ. 37 (1), 60–74.

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

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

Kim, N., Lučivjanská, K., Molnár, P., Villa, R., 2019. Google searches and stock market activity: Evidence from Norway. Finance Res. Lett. 28, 208–220.

Kimball, M., Shumway, T., 2006. Investor Sophistication, and the Participation, Home Bias, Diversification, and Employer Stock Puzzles. University of Michigan, Ann Arbor, Unpublished Manuscript.

Kostopoulos, D., Meyer, S., Uhr, C., 2020. Google Search Volume and Individual Investor Trading. Elsevier,

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

Li, X., Ma, J., Wang, S., Zhang, X., 2015. How does Google search affect trader positions and crude oil prices? Econ. Model. 49, 162–171.

Lusardi, A., Mitchell, O.S., 2014. The economic importance of financial literacy: Theory and evidence. J. Econ. Lit. 52 (1), 5–44.

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

Moussa, F., BenOuda, O., Delhoumi, E., 2017. The use of open source Internet to analysis and predict stock market trading volume. Res. Int. Bus. Finance 41, 399–411.

Oliveira-Brochado, A., 2019. Google search-based sentiment indexes. IIMB Manage. Rev..

Panagiotidis, T., Stengos, T., Vravosinos, O., 2019. The effects of markets, uncertainty and search intensity on bitcoin returns. Int. Rev. Financ. Anal. 63, 220–242.

Peng, L., Xiong, W., 2006. Investor attention, overconfidence and category learning. J. Financ. Econ. 80 (3), 563–602.

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

Quispe-Torreblanca, E., Gathergood, J., Loewenstein, G., Stewart, N., 2020. Attention utility: evidence from individual investors. Available at SSRN 3527082.

Sims, C.A., 2003. Implications of rational inattention. J. Monetary Econ. 50 (3), 665–690.

Swamy, V., Dharani, M., 2019. Investor attention using the Google search volume index–impact on stock returns. Rev. Behav. Financ. 11 (1), 55–69.

Takeda, F., Wakao, T., 2014. Google search intensity and its relationship with returns and trading volume of Japanese stocks. Pac.-Basin Finance J. 27, 1–18.

Van Rooij, M., Lusardi, A., Alessie, R., 2011. Financial literacy and stock market participation. J. Financ. Econ. 101 (2), 449–472.

Vlastakis, N., Markellos, R.N., 2012. Information demand and stock market volatility. J. Bank. Financ. 36 (6), 1808–1821.

Vozlyublennaia, N., 2014. Investor attention, index performance, and return predictability. J. Bank. Financ. 41, 17–35.

Ying, Q., Kong, D., Luo, D., 2015. Investor attention, institutional ownership, and stock return: Empirical evidence from China. Emerg. Mark. Financ. Trade 51 (3), 672–685.

Zhang, W., Shen, D., Zhang, Y., Xiong, X., 2013. Open source information, investor attention, and asset pricing. Econ. Model. 33, 613–619.