HAL
open science

Noise traders and smart money: Evidence from online searches

Fabrice Hervé, Mohamed Zouaoui, Bertrand Belvaux

To cite this version:

Fabrice Hervé, Mohamed Zouaoui, Bertrand Belvaux. Noise traders and smart money: Evidence from online searches. Economic Modelling, 2019, 83, pp.141-149. 10.1016/j.econmod.2019.02.005 . hal-02065042

HAL Id: hal-02065042

https://u-bourgogne.hal.science/hal-02065042v1

Submitted on 26 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Noise Traders and Smart Money: Evidence from Online Searches

Fabrice HERVE
Université de Bourgogne, CREGO, IAE DIJON

Mohamed ZOUAOUI
Université de Bourgogne, CREGO, IAE DIJON

Bertrand BELVAUX
Université de Bourgogne, CREGO, IAE DIJON

Economic Modelling, forthcoming

ACCEPTED ARTICLE

The first two authors contributed equally and are listed alphabetically.
Contact address of authors: Fabrice Hervé, Université de Bourgogne, UBFC, IAE DIJON, CREGO, 2 Bd Gabriel, BP 26611, 21066 DIJON CEDEX (France), Email: fabrice.herve@u-bourgogne.fr; Mohamed Zouaoui, Université de Bourgogne, UBFC, IAE DIJON, CREGO, 2 Bd Gabriel, BP 26611, 21066 DIJON CEDEX (France), Email: mohamed.zouaoui@u-bourgogne.fr; Bertrand Belvaux, Université de Bourgogne, UBFC, IAE DIJON, CREGO, 2 Bd Gabriel, BP 26611, 21066 DIJON CEDEX (France), Email: bertrand.belvaux@u-bourgogne.fr. We are grateful for helpful comments and suggestions from Sushanta Mallick (the Editor) and two anonymous referees, and seminar participants at Université de Bourgogne (France).

Noise Traders and Smart Money: Evidence from Online Searches

Abstract

Traditional finance theory considers that the impact of noise traders' attention on asset prices is offset by attention from smart investors. This paper uses online search data to study the influence of noise traders and smart investors on stock returns and volatility. Adopting an original approach, we construct a proxy for smart investor attention based on investors' online search behavior provided by Wikipedia Page Traffic. We combine this new measure with a standard measure of noise traders' attention as proxied by Google Search Volume Index. We show for a sample of 87 French firms over the period 2008–2018 that only noise traders' attention influences stock returns. Noise traders' attention increases volatility by creating an extra risk that is priced into the market. Conversely, smart investors' attention decreases volatility because their presence stabilizes stock prices by reducing uncertainty. Our empirical results support a behavioral explanation of stock prices.

Keywords: attention measures, smart investors, noise traders, price pressure hypothesis, behavioral finance

JEL Classifications: G12, G4

ACCEPTED ARTICLE

1. Introduction

Classical finance theory assumes that investors are by and large Bayesian, forming fully rational expectations about future cash flows and investment risks. The classical theory concedes that some investors – so-called noise traders[1] – cannot be rational but argues that their positions are offset by smart investors' actions.[2] As a result, asset prices fully incorporate and perfectly reflect all available information. Proponents of behavioral finance challenge this view by invoking the implications of investors' cognitive biases for the price formation process. Many studies show that investors have limited attention and limited aptitude for processing information (Simon, 1955). Attention consumes cognitive energy, and since a very large amount of information is accessible, investors must be selective (Kahneman, 1973). Several theoretical studies provide a framework in which noise traders' attention can lastingly affect asset pricing (Black, 1986; De Long et al., 1990).

Generally, a measurement problem arises when it comes to determining which category of investors is influencing the market. The attention of any given investor is not directly observable because it is subjective and individual by nature. As suggested by Black (1986, p. 532), "[t]here will always be a lot of ambiguity about who is an information trader and who is a noise trader". In the empirical literature, the variables used to gauge investors' attention fall into two categories. The first comprises direct measurements of investors' actions, for instance, the number, type, and size of the orders performed (e.g. Avramov et al., 2006; Foucault et al., 2011). The second category is made up of indirect measurements, e.g. abnormal returns or abnormal volumes of stock transactions (Barber and Odean, 2008). These indirect measurements appear to be measurements of noise traders' attention. These proxy variables may be subject to major flaws. In some cases, these variables may not reflect investor behavior: a very large number of stock market orders may be performed because of management constraints of an individual investor for example.

[1] Such, not completely rational, investors are also called noise, liquidity, uninformed, irrational, and positive-feedback traders in the literature or individual or retail investors (cf. Shiller, 2003 or Avramov et al., 2006). De Long et al. (1990, p. 704) specify that "Black (1986) believes that such investors, with no access to inside information, irrationally act on noise as if it were information that would give them an edge". Following Kyle (1985), Black calls such investors "noise traders". In the present article, we call them "noise traders".

[2] The literature calls this category smart, professional investors, rational speculators, or informed traders (cf. Shiller, 2003). In what follows, we use the term "smart investors".

ACCEPTED ARTICLE

The purpose of this article is to study both the impact of the attention of noise traders and smart investors on the French stock market using aggregate online search data. The Internet and its development have had major effects on society and on individual behavior (Rubin and Rubin, 2010). Ever more people have access to the Internet in France: the World Bank's World Development Indicators show that the number of Internet users in France has grown substantially from 14.3% of the population in 2000 to 39.2% in 2004, 70.7% in 2008, and 84.7% in 2015. Looney et al. (2006) claim that one of the areas most impacted by the development of the Internet is that of personal finance, i.e. investment decisions by non-professional investors. In France, ACSEL reports that the number of stock-exchange orders executed by on-line brokers rose from 6.6 million to 11.6 million between 2004 and 2011. This evidence points to increased participation by households in the stock market whether directly or through intermediaries. Bogan (2008) shows that access to the Internet does indeed promote household investment on the stock market. This phenomenon can be tied in with the findings by Ahmed et al. (2003) that online trading leads to growth in the proportion of naive investors on the stock market. In this vein, the paper by Da et al. (2011) attests to the effect of noise traders on the stock market using an Internet-based measure of noise traders' attention. As already evoked, besides these noise investors, another category of so-called smart, rational investors – market professionals – is assumed to intervene on the markets. In this paper, we develop a new and direct measure of smart investors' attention and we simultaneously investigate the impact of the attention of noise traders and smart investors on stock returns and volatility.

Huberman and Regev (2001) tell an anecdote illustrating how asset prices may change depending on the way these two categories of investors allocate their attention to a given asset. Shares in Entremed, a small biotechnology firm, underwent a series of changes. On 28 November 1997, further to a publication in *Nature* and to short press reports on the development potential of the company's anti-cancer therapy, the stock price rose from $10 to $13 on average in the following days. On 4 May 1998, the same information was highlighted on the front page of the New-York Times and the price surged from $12 to $85 when markets opened and stood at $52 at closing. This anecdote shows the differentiated effects of both categories of investors on share prices and on market activity as a whole. Hong and Stein (2007) interpret those price changes based on the assumption there were two types of investors: a small group of scientific investors specializing in biotechnologies and a larger

ACCEPTED ARTICLE

group of general investors. There was a small group of informed smart investors and a group of naive, less informed investors.

Andrei and Hasler (2015) report that "[…] as Huberman and Regev (2001) point out, prices react to new information only when investors pay attention to it". If we can identify which category of investors pays attention to a given firm, then it seems possible to measure the influence of categories of investors on the stock market. In a recent study, investor attention has been measured using Google's Search Volume Index. Da et al. (2011) claim it provides a good measure of noise traders' attention. In the present article, we follow the same methodology to construct a proxy of smart investor attention based on investors 'online search behavior provided by Wikipedia Page Traffic. We propose and confirm the idea that smart investors' attention can be proxied using traffic or edits on a firm's Wikipedia page. Furthermore, we rely on two Internet-based measures – Google Search Volume Index and Wikipedia Page Traffic – to simultaneously investigate the effects of attention from two categories of investors (noise versus smart) on financial markets in France.

This research is original in that it uses a new and direct measure of smart investors' attention and, accordingly, combines the use of two separate measures of investor attention. We study the stocks traded on the SBF 120 between February 2008 and September 2018. The paper directly examines the relationships between measures of attention of the two categories of investors and stock returns and volatility. Thus, we propose an empirical test of the De Long et al. (1990) model. Our results confirm our research hypotheses. In accordance with the price pressure hypothesis, we find that noise traders' attention temporarily affects stock returns. In contrast, smart investors realize that the information may already be impounded into price. Consequently, they do not significantly influence long-term asset prices. Moreover, we hypothesize and validate empirically that noise (smart) traders' attention exerts a positive (negative) influence on the volatility of a stock's return. This evidence is in line with our story: smart investors incorporate information into prices and in doing so they reduce signal dispersion and asset volatility; whereas noise traders, trading on the basis of noisy information, increase volatility.

This paper contributes to the existing literature in several ways. First, we propose and validate a new Internet-based measure of smart investors' attention using Wikipedia pages about firms. In particular, the Wikipedia-based measure is new and easy to obtain and process. Second, to the best of our knowledge, this is the first paper to use two separate measures of investor attention based on online search data, one for each category of investors, noise

ACCEPTED ARTICLE

traders and smart investors. Indeed, the introduction of our new measure allows us to disentangle the effects of noise and smart investors on stock prices. This allows us to test the model of De Long et al. (1990). Third, we provide evidence for a behavioral explanation of stock prices. Our results show that the two categories of investors have separate influences on stock returns and volatility. Noise and smart investors behave in a way that is consistent with the proposals by De Long et al. (1990). Ultimately, the effect of noise traders fades away because informed investors do return prices to equilibrium.

The paper is organized as follows. In section 2, we review the literature on measures of investor attention and we state our research hypotheses. In section 3, we describe our data and present our variables. In particular, we detail our proxies of investor attention and provide some descriptive statistics. In section 4, we present the empirical methodology and discuss our results. Finally, section 5 concludes.

2. Literature review and development of hypotheses

We begin with a presentation of the literature on investor attention. This helps us to identify two separate measures of attention, one reflecting noise traders' attention, the other proxying for smart investors' attention. We then develop the research hypotheses.

2.1. Investor attention measures: Disentangling smart investors and noise traders

Nowadays financial markets produce huge volumes of information. Information about firms and more broadly about their environment is continuously disclosed to the general public via many different media (newspapers, radio, television, the Internet). Lazer et al. (2009) note that "the Internet offers an entirely different channel for understanding what people are saying, and how they are connecting". In this way, the Internet makes it possible to ascertain how people collect information, in other words, it throws light on what Simon (1955) calls the information gathering phase in a decision-making process.

Search engine query data have been widely studied. The articles by Moat et al. (2014) or Hervé and Zouaoui (2014) include literature reviews on their use in various areas (medicine, economics, marketing, etc.). Moat et al. (2014) come up with the idea that such data "provide a view of the beginning of the information gathering phase in real decision-making processes, which could not previously be observed" and that they may have a predictive character. In this vein, Bijl et al. (2016) use Google Trends data and show that they can be exploited to forecast stock returns. Han et al. (2018) find similar results for currency returns. Li et al. (2015) study crude oil prices and obtain improved predictions of crude oil

ACCEPTED ARTICLE

prices using the Google Search Volume Index as a proxy for investor attention. As Vozlyublennaia (2014) observes, the point of such measures is that they make it possible to fully appreciate the interest an investor expresses. Da et al. (2011) notice that data from search engines provide a direct measure of attention because it is a revealed measure. If people make a query about a specific firm it is because they are interested in it. Given that Google accounted for more than 90% of queries in France over the period 2008–2018, it seems logical that the volume of searches made via Google is an indicator of investors' attention in France.[3]

A number of papers, the most representative certainly being that by Da et al. (2011), examine the link between the volume of Internet search queries and the financial markets. Generally, Google search queries represent the attention of one category of investor in particular: individual investors (Da et al. 2011; Bank et al. 2011; Aouadi et al., 2013). Vozlyublennaia (2014) emphasizes that "The general consensus is that these measures capture the attention of retail investors". Likewise, Da et al. (2011) observe that such measurements reflect the attention of less sophisticated investors. If the volumes of Google searches illustrate the level of attention from noise traders and if a measure can be identified to represent the attention from smart investors, it then becomes possible to evaluate the different effects of the two categories of investors on financial markets.

Moat et al. (2013) suggest that "Wikipedia records may provide a proxy measurement of the information gathering process of a subset of investors". In the *Financial Times*, Bradshaw (2008) reports a study by a consultancy firm (Hallvarsson & Halvarsson) of journalists, financial analysts, and investors and asserts that 91% of respondents gathered information on firms from sources other than corporate websites and that more than 75% of them used Wikipedia.

Xu and Zhang (2013) argue that Wikipedia provides an information compilation service. The informational value produced via this service is therefore high. Wikipedia information gathering produces more relevant results than Google queries. The compilation of information means that corporate leaders no longer have an informational edge enabling them to manipulate private information about their businesses. With Wikipedia, it is more difficult

[3] Hoopes et al. (2015) underline that a distinction is made in psychology between two types of attention: exogenous attention for bottom-up information processing and endogenous attention for top-down information processing. Endogenous attention is goal-directed (e.g. investing one's wealth in shares) whereas exogenous attention is conditioned by events outside the process and requires an exogenous stimulus (e.g. announcement of an essential discovery for a firm). This research therefore investigates endogenous attention.

ACCEPTED ARTICLE

for them to control information because this online encyclopedia provides unbiased or bias-resistant information, unlike that conveyed by newspapers or financial analysts. The lack of bias results from a decentralized process of compilation of information by Wikipedia. Wikipedia permits investors to obtain information cheaply (Rubin and Rubin, 2010). More precisely, Rubin and Rubin (2010) argue that Wikipedia can be used to measure the proportion of smart investors for a given company. Their idea is that Wikipedia reflects the information processing pertaining to a company. This discussion suggests that a Wikipedia-based measure is a good proxy for smart investors' attention.

This reasoning depends on the editorial quality of Wikipedia and on its capacity to produce relevant information. An article in *Nature* (Giles, 2005) shows that Wikipedia seems to be as reliable as the Encyclopedia Britannica, which is a guarantee of the editorial quality of this online encyclopedia. Xu and Zhang (2013) explain that Wikipedia is able to disseminate relevant information rapidly. In 2009, it was almost the only information source to say that Dell was selling off a call center to Teleperformance (Xu and Zhang 2013, p. 1050).

Given this evidence, we argue that the intensity of searches via Google provides a good measure of the attention of individual unsophisticated investors – noise traders – whereas information gathering via Wikipedia provides an accurate measure of professional, smart investors' attention. We now turn to the expected effect of the attention of the two categories of investors on stock prices.

2.2. Development of research hypotheses

First, we discuss the expected effects of the two separate attention measures on stock prices. Then, we explain the expected effects of these attention measures on stock volatility.

2.2.1. Price pressure effect

Barber and Odean (2008) study the relationship between individual investors' attention and changes in stock prices. They assume that individual investors are net purchasers of stocks that attract their attention. Faced with the huge number of stocks available on the market, individual investors decide to invest in stocks that grab their attention, i.e. shares that are traded in abnormal volumes, that have extremely abnormal returns, or that come up in the news. Conversely, individual investors can only sell the securities they already hold since individual investors seldom sell short. Following this reasoning, individual investors are net buyers of attention-grabbing stocks. This causes a temporary rise in stock

ACCEPTED ARTICLE

prices followed by a drop resulting from the adjustment of prices. Da et al. (2011) reason along similar lines. This leads to the following hypothesis:

H1a (price pressure hypothesis): Ceteris paribus, noise traders' attention induces higher stock prices in the short run followed by a price correction in the long run.

Barber and Odean (2008) assert that "fully rational investors will recognize the limitations of buying predominantly stocks that catch their attention. They will realize that the information associated with an attention-grabbing event may already be impounded into price (since the event has undoubtedly been noticed by others)." The point is that rational, informed investors incorporate information into prices. Increased attention from smart investors should not have any particular effect on stock prices in the long term because, under hypothesis H1a, noise traders influence share prices significantly. If we follow Rubin and Rubin (2010), we might think that the attention of smart investors and the proportion of noise traders in the market are connected. In this case, if we refer to De Long et al. (1990), we should be confronted with a phenomenon of limits to arbitrage and noise trader risk. For those authors (p.713), "if sophisticated investors' horizons are long relative to the duration of noise traders' optimism or pessimism toward risky assets, then they can buy low, confident that they will be able to sell high when prices revert to the mean". Thus, if this condition holds here, we should see a positive short-term influence of smart investors' attention on stock returns and no significant long-term effect. This leads to the next hypothesis:

H1b (rational behavior of smart investors hypothesis): ceteris paribus, smart investors' attention induces higher stock prices in the short run and no significant effect (no price correction) in the long run.

There is a strong link between risk and return in finance. Accordingly, it can be assumed that if investor attention has a significant influence on stock returns, it should also affect the volatility of stock returns. This is the subject we now turn to.

2.2.2. Volatility and investor attention

Over the last decade, many financial economists have debated the effect of noise trading on volatility. In the theoretical models, the economy is characterized by two types of investors: smart investors who rationally anticipate asset prices and noise traders whose expectations lead to periods of over- or undervaluation of financial assets. Both types of investors are risk adverse and the equilibrium price reflects everyone's expectations.

ACCEPTED ARTICLE

Traditional financial theory (e.g. Fama, 1970) postulates that market prices fully and perfectly reflect all available information and argues that the positions of noise traders are offset by arbitrageurs bringing prices back to their fundamental value. As a result, noise trading has no effect on volatility. However, behavioral financial theory (e.g. Black, 1986; De Long et al., 1990) expresses a different view about the effect of noise trading on volatility. Indeed, the theoretical framework assumes that noise traders' sentiment is stochastic and cannot be perfectly forecast by smart investors. Because arbitrage has practical limits, smart investors fail to fully offset the effects of noise traders' trading. It follows that noise traders' trading influences stock price volatility. Several empirical studies identify a link between investor attention and volatility. They all use an attention measure based on Google search query. Generally, they find a positive relationship between volatility and noise traders' attention measured by the intensity of Google searches (Vlastakis and Markellos, 2012; Smith, 2012; Takeda and Wakao, 2014; Aouadi et al., 2013; Vozlyublennaia, 2014). The most basic implication of noise trader theory is that the existence of noise traders introduces an extra risk in financial markets and increases the volatility of market prices of securities. This risk is known in the literature as "noise trader" risk or investor sentiment risk. De Long et al. (1990) show formally that when the proportion of noise traders rises compared with the proportion of smart investors, asset volatility increases.

These arguments lead to the following hypothesis:

H2a (noise trading hypothesis): ceteris paribus, noise investors' attention has a positive effect on volatility.

A second strand of literature using rational expectations models studies the impact of smart investors on volatility. Hellwig (1980), Wang (1994), and Odean (1998) propose formal theoretical models with rational expectations. Their models build on the idea that the trades of rational, informed, smart investors provide a signal about their information and so stabilize prices, thereby contributing to reduced volatility. As the number of informed traders grows, the signal is improved and the effect of noise traders on prices is mitigated. This reduces the deviation of price from the fundamental value and as a result volatility decreases. Avramov et al. (2006) study the volatility of stock returns and its relation with trades empirically. They report that smart investors' transactions reduce the volatility of stock returns, while trades by less rational noise traders raise volatility.

Accordingly, we can formulate the following hypothesis:

ACCEPTED ARTICLE

H2b (smart investors and volatility): ceteris paribus, smart investors' attention exerts a negative influence on volatility

3. Data and measures of investor attention

We begin this section with a brief presentation of the data. We discuss in particular the variables used to measure the attention of the two categories of investors. We then outline the other variables used in the study and a few descriptive statistics.

3.1. Data

We use weekly stock returns data from the Thomson Reuters One database. In order to mitigate liquidity concerns, we focus on firms in the SBF 120 in February 2008.[4] Our sampling period runs from 10 February 2008 to 02 September 2018, i.e. 552 weeks. We use this period, although Google Trends data are available from as early as 2004, since until late 2007, it seems that Google Trends data were rarely updated and of doubtful reliability.[5]

The data from Google Trends were downloaded directly from the Google Trends site on 02 September 2018.[6] Like Aouadi et al. (2013), we define France as the geographical zone and exclude stocks for which fewer than 20 weeks of data are available and those for which Google search volumes are zero for eight consecutive weeks. We use company names and not stock tickers. In the French context, the use of tickers seems to be less suitable and leads to major data problems due to massive missing values (Aoudi et al., 2013).[7] The Wikipedia data are extracted from the revision history available for any page.[8] Due to missing values, 87 stocks remain in our sample. We use a historical measure of volatility over the previous 24 weeks.[9] This volatility measure is used by many authors (Foucault et al., 2011; Vlastakis and Markellos, 2012; Aouadi et al., 2013).

3.2. Investor attention measures

In this subsection, we detail the construction of our two measures of investor attention. Following Da et al. (2011), we use Google search queries to capture the attention of noise

[4] The SBF 120 index is composed of the 120 largest capitalizations on the French market.
[5] https://en.wikipedia.org/wiki/Google_Trends#
[6] http://www.google.fr/trends/
[7] This setting varies with the paper considered. Vozlyublennaia (2014) and Da et al. (2011) use ticker symbols, whereas Vlastakis and Markellos (2012), Aouadi et al. (2013), or Takeda and Wakao (2014) use firms' names. In the French context, we conducted search tests and obtained inconclusive Google searches and/or missing values.
[8] For example, for the French version of Wikipedia, access to the history of the "Sanofi" article is located at: http://fr.wikipedia.org/w/index.php?title=Sanofi&action=history
[9] As a robustness check, we compute 36 months' volatility and replicate our analysis. The results are qualitatively similar.

ACCEPTED ARTICLE

traders. This appears to be a good proxy for capturing investor attention essentially because of its capacity to replicate sudden surges in attention. The measure of noise traders' attention is similar to Da et al. (2011) and is as follows:

$$ASVI_{i,t} = \text{Log}(SVI_{i,t}) - \text{Log}\left[Med\left(SVI_{i,t-1}, \dots, SVI_{i,t-8}\right)\right] \qquad (1)$$

Where $\text{Log}(SVI_{i,t})$ is the logarithm of SVI for a firm *i* during week *t* and $\text{Log}[\text{Med}(SVI_{i,t-1}, \dots, SVI_{i,t-8})]$ is the logarithm of the median value of SVI during the previous 8 weeks. A positive (negative) ASVI reflects a sudden rise (fall) in noise traders' attention.

Rubin and Rubin (2010), Xu and Zhang (2013), and Moat et al. (2013) attempt to identify the usefulness of Wikipedia in informing investors. Building on these papers, we retain two measures of smart investors' attention in the form of two Wikipedia-based measures. The first one involves counting the number of consultations of a Wikipedia page on a given business and the second one counts the number of edits of contents of a Wikipedia page. With regard to the research results, no clear consensus arises in favor of either measure. Accordingly, it is not easy to choose between page traffic and page edits.[10]

Measuring traffic on Wikipedia pages is a good way to gauge smart investors' attention. Westen (2000) claims that online consultation of articles enables individuals to gather information as and when they wish, i.e. without waiting for publication in newspapers and independently of editorial choices. In addition, Liu (2005) indicates that – in an increasingly digital environment – individuals read ever more electronic documents and that reading such documents differs from reading printed documents. Online reading is non-linear and selective, in other words it is behavior that is more akin to navigation, to scanning and searching for keywords. This suggests that readers of Wikipedia pages on a given firm gather relevant information. More visits to a Wikipedia page would therefore reveal increased interest of readers who read smartly, which argues in favor of measuring smart investors' interest by using traffic on firms' Wikipedia pages. The results in Moat et al. (2013) confirm the value of such measures.

Measuring the attraction of smart investors for a firm via the volume of edits on firms' Wikipedia pages can be explained as follows. Editing is done by better informed individuals

[10] The same prevails in other areas. For instance, in computer science, Georgescu et al. (2013) notice that previous research identifies Wikipedia page traffic as a good way of detecting events. They show that Wikipedia page edits are valuable for detecting important events. These results are an indication of the capacity of editing on a Wikipedia page to provide investors with information. In marketing, Mestyan et al. (2013) rely on the number of pages consulted and the number of edits to predict how successful a film will be. It appears that the joint use of the two measures has significant predictive power. One explanation given by those authors is that Wikipedia users are informed users.

ACCEPTED ARTICLE

(Rubin and Rubin 2010). Editing, combining differing points of view about a firm, produces relevant and unbiased information (Xu and Zhang 2013). Accordingly, the more changes that are made to a Wikipedia page on a given firm, the lower information search costs and the better informed investors are. Rubin and Rubin (2010) assume that these low costs should increase the proportion of smart investors. They empirically confirm the idea that the number of revisions on a Wikipedia page measures the intensity of information searches and that the increased search for information (the rise in the number of edits on a page) is related to a higher proportion of informed investors.

At this point, a question may be raised about these measures. Someone who consults or edits a Wikipedia page will not necessarily be an investor in the market. Our argument is that it is because individuals consult or edit Wikipedia pages that information is disseminated and then attracts informed investors' attention. This approach is consistent with work in behavioral finance on measuring investor sentiment. For instance, the Michigan Consumer Confidence Index is used as a proxy for US investor sentiment (Lemmon and Portniaguina, 2006; Baker and Wurgler, 2007).

Two measures of smart investors' attention can be used:

$$AWIKI_{i,t} = \text{Log}(WIKITRAF_{i,t}) - \text{Log}\left[Med\left(WIKITRAF_{i,t-1}, \dots, WIKITRAF_{i,t-8}\right)\right] \quad (2)$$

Where $WIKITRAF_{i,t}$ is the average number of consultations of a Wikipedia page for a given firm *i* during a week *t* and $\text{Log}[\text{Med}(WIKITRAF_{i,t-1}, \dots, WIKITRAF_{i,t-8})]$ is the logarithm of the median value of traffic on the Wikipedia page of firm i in the previous 8 weeks. This measure is defined in a similar way to that for noise traders' attention (ASVI). The argument is that reasoning in terms of deviation from a trend better captures any spike in attention than measuring the level of traffic.

As the second measure relies on edits of pages being by nature a differential measure, we use the measure of Xu and Zhang (2013):

$$WIKIEDIT_{i,t} = Ln(1 + \text{number of modifications for firm } i \text{ during week } t) \quad (3)$$

Identifying which measure is best suited for our purpose is an empirical matter. We will discuss this point in detail in the following section. We use other variables taken from the literature (cf. Da et al., 2011). They come from the Thomson Reuters One database. Stock prices and trading volumes data are obtained from Datastream. Table 1 details the variables used in this article.

ACCEPTED ARTICLE

[Table 1 about here]

3.3. Summary statistics

Table 2 presents the descriptive statistics of our main variables, the correlation matrix between investor attention variables and the results of Granger causality tests. These descriptive statistics confirm the idea that the proxies we use do correspond to two separate categories of investors.

[Table 2 about here]

All the correlation coefficients display the expected positive sign. The strongest correlation is between WIKITRAF and WIKIEDIT (32.6%), which suggests that the revision of a Wikipedia page prompts Internet users to consult the Wikipedia page more often, thereby increasing traffic on the page. The variables traditionally used in the literature for measuring smart investors' attention (analyst attention, media attention) are weakly correlated with the SVI measure. Only the correlation coefficient between SVI and WIKITRAF is positive and significant at the 5% level. These results are in line with those of Da et al. (2011) who report a weak correlation between SVI and other usual investor attention measures.

Conversely, correlation coefficients between the WIKITRAF and WIKIEDIT measures and the other proxies traditionally used for measuring smart investors' attention are all positive and highly significant. These contemporaneous relations are significant at the 1% level. These findings suggest that the Wikipedia-based measures are relevant proxies for measuring smart investors' attention.

The strong contemporaneous correlations between the WIKITRAF variable and the other measures of smart investors' attention raise a question of causality. Is the WIKITRAF measure a leading indicator of smart investors' attention or vice-versa? To address this question, we conduct Granger causality tests on panel data. The results of these tests appear in Panel C of Table 2. They show that the WIKITRAF variable influences the traditional measures of smart investors' attention at the 5% confidence level (p-values less than 0.05). Moreover, the other traditional measures of smart investors' attention do not influence the WIKITRAF variable at the conventional significance level (all p-values are greater than 0.1). This result is not surprising because the traditional measures are market data. They incorporate information into prices once the investors have traded. The WIKITRAF measure is able to provide advanced information because it is the crowd that integrates information

ACCEPTED ARTICLE

into Wikipedia pages and the information is processed very rapidly (Xu and Zhang, 2013). Further, we conduct other Granger causality tests (not reported here for the sake of brevity, available upon request). Following Da et al. (2011), we examine the link between SVI and standard measures of attention and find no significant causal links. Finally, we study the causal relationship between the Google-based investor attention measure (SVI) and the Wikipedia-based investor attention measures (WIKITRAFF and WIKIEDIT). Wikipedia webpages tend to rank high in Google searches.[11] In looking for firms' names on the Google search engine, web users may well consult and use Wikipedia pages because the search engine tends to show them among the top webpage proposals, which may therefore bias both our measures of investor attention. The results of the Granger-causality tests show that we do not observe any causal relationship between the two categories of measures of investors' attention. This empirical evidence proves that we are dealing with two well-separated populations of investors, a smart population using Wikipedia and a noise trader population using Google.

4. Empirical modeling and findings

Here we present the method used for testing our two series of hypotheses and discuss our results.

4.1. The effect of investor attention on stock returns

To assess the effect of (smart and/or noise) investor attention on stock returns, we run the regression below (4). The specification builds on Da et al. (2011).

$$\frac{1}{K}\sum_{k=1}^{K} R_{i,t+k} = \alpha + \gamma_1 ATTENTION_{i,t} + \gamma_2 \operatorname{Log}(MARKETCAP)_{i,t} + \gamma_3 ATTENTION_{i,t} \times \operatorname{Log}(MARKETCAP)_{i,t} + \gamma_4 ABRETURN_{i,t} + \gamma_5 \operatorname{Log}(1 + ANALYSTS_{i,t}) + \gamma_6 NEWS_{i,t} + \gamma_7 ABTURNOVER_{i,t} + \varepsilon_{i,t} \quad (4)$$

Where $\frac{1}{K}\sum_{k=1}^{K} R_{i,t+k}$ is the dependent variable measuring average stock returns over K-periods of firm *i*. ATTENTION is the selected investor attention measure. It is measured in two ways: it represents the attention of either noise traders (ASVI) or smart investors (AWIKITRAF or WIKIEDIT). MARKETCAP is the stock market capitalization of firm i at date t. ABRETURN is the absolute value of characteristic-adjusted return as in Daniel et al.

[11] We are indebted to one of the anonymous reviewers for this suggestion.

ACCEPTED ARTICLE

(1997). ANALYSTS is the number of financial analysts who follow firm i at date t. NEWS is measured using the number of weekly mentions of each firm in a reference French newspaper, *Les Echos*. We use the occurrences of the firm's name in the headlines. ABTURNOVER is the abnormal turnover as in Da et al. (2011). K is the forecast horizon.

To assess the impact of the attention of the different categories of investors on stock returns, we use cross-sectional Fama-MacBeth regressions to estimate Model (4). For each week, the mean return calculated for the different horizons is regressed on the measure of (smart and/or noise) investor attention and the control variables at time *t.* All the explanatory variables are standardized and have a comparable scale of values. The reported estimates are the time-series averages of the weekly estimated coefficients. The p-value is based on the standard deviation of the coefficient estimate using the Newey-West (1987) formula with eight lags.

[Table 3 about here]

The results are presented in Table 3. The regression coefficient on a variable can be interpreted as the effect on the dependent variable of a one-standard deviation change in that variable. For panel A, we observe a positive significant relationship between the mean return and the variable measuring noise traders' attention for investment horizons k=1, 2, and 3 weeks. The phenomenon is reversed when the forecast horizon exceeds 4 weeks; a negative relation then arises between the noise traders' attention at time *t* and future stock market returns. An increase in the standard deviation of the ASVI variable is followed by a significant positive rise in market returns of 11.871 points over the following week but a significant fall of -2.098 points over the next 4 weeks. We also notice that, in the long-term regression (5–52 weeks), we obtain an estimated coefficient on the ASVI variable that is negative and significant (-19.827) at the 5% level. This amplitude of the fall in stock prices is comparable with the initial hike in stock prices over the first 3 weeks, which proves that the initial positive pressure on stock prices is fully reversed over the horizon of one year.

Moreover, it is observed that our variable for measuring noise traders' attention is significant even in the presence of investor attention measures traditionally used in the literature (ABRETRUN, ANALYSTS, ABTURNOVER). This finding shows that our Internet-based measure captures part of the individual investor attention that is not detected by the measures traditionally used in the literature.

ACCEPTED ARTICLE

Noise traders' attention has a greater impact on stocks that are difficult to value or to arbitrage (Baker and Wurgler, 2007; Da et al., 2011). Small cap stocks in particular come into this category. To refine our results and following Da et al. (2011), we introduce an interaction term into our regression. The interaction variable between the measure of individual investor attention (ASVI) and stock market capitalization (MARKETCAP) is negative and significant at the 1% level. This result suggests that the impact of noise traders' attention on stock returns is greater for small cap stocks. This finding is fully consistent with the findings in the field of behavioral finance. These results support our hypothesis H1a about price pressure. The values of the signs of these coefficients corroborate our central hypothesis about price pressure. Periods characterized by a high level of investor attention are followed by low levels of future returns (De Long et al., 1990). It also turns out that the price-pressure effect is more intense for small caps.

For panels B and C, a positive significant relationship is also observed between mean returns and the measure of smart investors' attention for forecast horizons of 1 and 2 weeks (AWIKITRAF or WIKIEDIT). Conversely, the measure of smart investors' attention does not influence the performance of stocks for forecast horizons exceeding 2 weeks, which corroborates hypothesis H1b. Smart investors' attention causes a rise in prices over the short term but has no impact on long-term prices.

Panel D of table 3 shows the results of the regression including both measures of noise and smart investors' attention. The results show that the coefficients of the noise traders' attention variable remain significant and display the expected signs. The influence of the smart investors' attention variable is non-significant at the conventional significance level. This finding supports the idea that the two investor attention measures are independent and correspond to two separate concepts. This is consistent with the previous results of the Granger causality tests.

Overall, whatever the panel under consideration, we obtain low R-squared varying in a range of 2.0% to 2.6%. This low level of power of our regressions results from their predictive nature. Studies similar to ours obtain similar coefficients: Da et al. (2011) report R-squared inferior to 2%, Vozlyublennaia (2014) provides slightly better coefficients but she is working with stock index returns, which are generally easier to predict than individual stock returns (Fama and French, 1993).

4.2. The effect of investor attention on stock volatility

ACCEPTED ARTICLE

To evaluate the effect of (smart and/or noise) investor attention on volatility, we use panel data regression. A long line of literature identifies a positive link between volatility and volume (e.g. Girard and Biswas, 2007; Giot et al., 2010; Karpoff, 1987; Chen et al., 2001). We therefore introduce the trading volume as a control variable into our regression (5). In addition, we use the market ASVI to control for market attention of noise traders (Vlastakis and Markellos 2012). We extract the Google search volume for the keyword "CAC 40". The specification of our model is in line with the literature (Aouadi et al., 2013; Vlastakis and Markellos, 2012). It is as follows:[12]

$$VOLATILITY_{i,t} = \alpha + \gamma_1 ATTENTION_{i,t} + \gamma_2 \text{Log}(MARKETCAP)_{i,t} + \gamma_3 ATTENTION_{i,t} \times \text{Log}(MARKETCAP)_{i,t} + \gamma_4 MARKETRETURN_{i,t} + \gamma_5 NEWS_{i,t} + \gamma_6 \text{ABTURNOVER}_{i,t} + \gamma_7 VOLATILITY_{i,t-1} + \gamma_8 ASVICAC40_{i,t} + TIME_{effects} + SECTOR_{effects} + \varepsilon_{i,t} \quad (5)$$

Where $\text{VOLATILITY}_{i,t}$ is the historical volatility computed over 24 weeks of firm i at date t. ATTENTION is the selected measure of investor attention. Investor attention is measured in two ways: it represents the attention of either noise traders (ASVI) or smart investors (AWIKITRAF or WIKIEDIT). MARKETCAP is stock market capitalization of the stock i at time t. MARKETRETURN is the log return of the market as a whole (CAC 40). NEWS is measured using the number of weekly mentions of each firm in a reference French financial newspaper, *Les Echos*. We use the occurrence of the firm's name in the headlines. ABTURNOVER is abnormal turnover as in Da et al. (2011). ASVICAC40 corresponds to the ASVI measure for the term "CAC 40". TIME and SECTOR are dummy variables that capture the temporal and sector of activity effects. All of the independent variables are standardized.

Before making the estimation, we conduct empirical tests to evaluate the relevance of the specification of our model in panel data. We first perform the Fisher test on the nullity of all the parameters. The value of the statistical test enables us to reject at the 1% level the null hypothesis of homogeneity of individual effects. Accordingly, the fixed-effects estimation method is more suitable than the simple pooled model. We also compare the fixed individual effects model with the random effects model using the Hausman test. The fixed-effects method of estimation proves systematically preferable to the random effects method of estimation. The statistical performance of the Hausman test implies rejection of the null hypothesis at the 1% level, confirming the occurrence of fixed effects.

[12] To be consistent with the literature we have not included in the regressions the other measures traditionally used in estimating returns (cf. equation 4). Their inclusion in the regressions does not affect the results and the estimated coefficients are not significant.

ACCEPTED ARTICLE

[Table 4 about here]

The results of the estimation in panel data for the coefficient associated with the investor attention variable, with the introduction of control variables, appear in table 4. Column (1) of the table presents the results using the ASVI variable as a measure of investor attention. The R-squared of the regression is relatively high (around 30%), confirming the quality of the model's overall fit. Overall fit is better for stock's volatility than for stock returns. This result is in line with Vozlyublennaia (2014). The coefficient on the ASVI variable is positive and significant at the 5% level even in the presence of control variables. In other words, noise traders' attention causes a rise in volatility. This finding is consistent with Vlastakis and Markellos (2012) and Aouadi et al. (2013). We also observe that the coefficient displayed on the interaction variable (ASVI× Log(MARKETCAP)) is negative and significant, showing that the impact of noise traders' attention on stock volatility is higher for small cap stocks. This result confirms hypothesis H2a that attention increases volatility and this is particularly so for small caps.

Columns (2) and (3) of the table display the results including smart investors' attention measures. We notice that only the Wikipedia page traffic has a significant influence on volatility. Edits of Wikipedia pages have no significant effect on stock volatility. The impact of Wikipedia page traffic is negative and significant at the 5% level. Put differently, smart investors' attention reduces stock volatility.

Following Aouadi et al. (2013), we introduce stock market capitalization as an interaction term. Firms with large capitalizations are most closely monitored by analysts; being highly visible, they are exposed to less uncertainty. Due to the constant close monitoring of these firms, we can expect that the focus of smart investors' attention on large capitalization firms contributes to reducing the volatility of their returns, because their increased attention is rapidly observed and taken into consideration by the market. Accordingly, we find that the coefficient of the interaction variable (AWIKI × Log(MARKETCAP)) is positive and significant at the 5% risk level. The negative impact of smart investors' attention on stock volatility is greater for large cap stocks. Our results confirm hypothesis H2b that greater attention from smart investors leads to a decrease in volatility.

Column (4) of the table displays the results combining measures of attention of both smart and noise investors. Overall our findings remain significant and show that noise traders' attention contributes to a rise in stock volatility especially for small stocks whereas the attention of smart investors reduces volatility in particular for large capitalization stocks.

ACCEPTED ARTICLE

5. Conclusion

Lazer et al. (2009) contend that digital traces, especially on the Internet, offer great research opportunities. In the present paper, we follow this view and argue that Internet-based measures and more specifically those from Wikipedia are relevant proxies for smart investors' attention. We confirm empirically the usefulness of such Wikipedia-based measures. Google-based measures prove to be representative of noise traders' attention. These two sets of measures are of particular value because they are ex-ante measures of investor attention – whether informed or not – to a given stock. This supplements the ex-post and indirect measures used in the literature (e.g. abnormal returns or abnormal transaction volumes on stock). Obvious advantages associated with the use of such measures are their low cost and the ease in obtaining them. They are publicly available and there are no restrictions on their use.

Modern finance theory considers that investor attention is not a topic which has to be discussed because investors are fully rational. They are able to process all the available information and to incorporate it into stock prices. Usually, two categories of investors are identified, noise traders and smart investors (De Long et al., 1990). The direct measure of smart investors' attention we propose here offers the possibility of empirically testing whether investor attention – and in particular attention from two separate categories of investors – matters. The combined use of two attention measures enables us to validate the predictions of the model of De Long et al. (1990) empirically. To the best of our knowledge, this has not been done before, with such aggregate online search data. This is an important contribution by our paper to the literature on finance and, in particular, on behavioral finance. All in all, our results point to a behavioral explanation for stock prices. Noise traders negatively influence prices but, in the end, their influence is offset by the behavior of informed investors. This suggest that periods of high investor attention should be followed by high volatility and low future stock returns.

Our results are consistent with the remark by Fisher Black quoted at the beginning of this article. We can argue that, with the development of the Internet, it is now possible to measure the proportion of attention devoted to a particular stock by both smart investors and noise

ACCEPTED ARTICLE

traders. So, there is no longer any ambiguity about this point. Ambiguity remains when we have to precisely identify who is a noise trader and who is not.

One question remains unresolved at the end of this article: Can these measures be used to put in place a profitable trading strategy? We have seen that they are predictive, which suggests they can be used in portfolio management. This points to an opening for future research.

ACCEPTED ARTICLE

References

ACSEL, http://www.acsel.asso.fr

Ahmed, A. S., Schneible, R. A., & Stevens, D. E. (2003). An empirical analysis of the effects of online trading on stock price and trading volume reactions to earnings announcements. *Contemporary Accounting Research*, *20*(3), 413-439.

Andrei, D. & Hasler, M. (2015). Investor attention and stock market volatility. *The Review of Financial Studies*, *28*(1), 33-72.

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

Avramov, D., Chordia, T., & Goyal, A. (2006). The impact of trades on daily volatility. *The Review of Financial Studies*, *19*(4), 1241-1277.

Baker, M. & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129-152.

Bank, M., Larch, M., & Peter, G. (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. and Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies,* 21, 785–818.

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

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

Bogan, V. (2008). Stock market participation and the internet. *Journal of Financial and Quantitative Analysis*, *43*(1), 191-211.

Bradshaw, T. (2008). Companies woo investors via social websites. *Financial Times.*

Chen, G. M., Firth, M., & Rui, O. M. (2001). The dynamic relation between stock returns, trading volume, and volatility. *Financial Review*, *36*(3), 153-174.

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

Daniel, K., Grinblatt, M., Titman, S., & Wermers, R. (1997). Measuring mutual fund performance with characteristic-based benchmarks. *The Journal of Finance*, *52*(3), 1035-1058.

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

Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work, *Journal of Finance*, 25 (2), 383-417.

Fama, E. F. & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56.

Fama, E. F. & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*(3), 607-636.

Foucault, T., Sraer, D., & Thesmar, D. J. (2011). Individual investors and volatility. *The Journal of Finance*, *66*(4), 1369-1406.

Georgescu, M., Kanhabua, N., Krause, D., Nejdl, W., & Siersdorfer, S. (2013, March). Extracting event-related information from article updates in wikipedia. In *European Conference on Information Retrieval* (pp. 254-266). Springer, Berlin, Heidelberg.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature, 438(15),* 900-901.

Giot, P., Laurent, S., & Petitjean, M. (2010). Trading activity, realized volatility and jumps. *Journal of Empirical Finance*, *17*(1), 168-175.

Girard, E. & Biswas, R. (2007). Trading volume and market volatility: Developed versus emerging stock markets. *Financial Review*, *42*(3), 429-459.

Han, L., Xu, Y., & Yin, L. (2018). Does investor attention matter? The attention-return relationships in FX markets. *Economic Modelling*, *68*, 644-660.

Hellwig, M.F. (1980). On the Aggregation of Information in Competitive Markets. *Journal of Economic Theory*, 22, 477–498.

Hervé, F. & Zouaoui, M. (2014). Quand la psychologie et la linguistique rencontrent la finance : le cas de la France. *Finance Contrôle Stratégie*, (17-1).
Hong, H. & Stein, J. C. (2007). Disagreement and the stock market. *The Journal of Economic Perspectives*, *21*(2), 109-128.
Hoopes, J. L., Reck, D. H., & Slemrod, J. (2015). Taxpayer search for information: Implications for rational attention. *American Economic Journal: Economic Policy*, *7*(3), 177-208.
Huberman, G. & Regev, T. (2001). Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *The Journal of Finance*, *56*(1), 387-396.
Kahneman, D. (1973). Attention and Effort, Englewood Cliffs: *Prentice-Hall*., 1-246
Karpoff, J. M. (1987). The relation between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, *22*(1), 109-126.
Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315-1335.
Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ..., & Jebara, T. (2009). Computational Social Science. *Science*, *323*(5915), 721-723.
Lemmon, M. & Portniaguina, E. (2006). Consumer confidence and asset prices: Some empirical evidence. *The Review of Financial Studies*, *19*(4), 1499-1529.
Li, X., Ma, J., Wang, S., & Zhang, X. (2015). How does Google search affect trader positions and crude oil prices? *Economic Modelling*, *49*, 162-171.
Liu, Z. (2005). Reading behavior in the digital environment: Changes in reading behavior over the past ten years. *Journal of Documentation*, *61*(6), 700-712.
Looney, C. A., Valacich, J. S., Todd, P. A., & Morris, M. G. (2006). Paradoxes of online investing: Testing the influence of technology on user expectancies. *Decision Sciences*, *37*(2), 205-246.
Mestyán, M., Yasseri, T., & Kertész, J. (2013). Early prediction of movie box office success based on Wikipedia activity big data. *PloS One*, *8*(8), e71226.
Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T. (2013). Quantifying Wikipedia usage patterns before stock market moves. *Scientific Reports*, *3*, 1801.
Newey, W. K. & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*(3), 703-708.
Odean, T. (1998). Volume, Volatility, Price, and Profit When All Traders Are Above Average. *Journal of Finance*, 53(6), 1887-1934.
Rubin, A. & Rubin, E. (2010). Informed investors and the internet. *Journal of Business Finance & Accounting*, *37*(7-8), 841-865.
Shiller, R. J. (2003). From efficient markets theory to behavioral finance. *The Journal of Economic Perspectives*, *17*(1), 83-104.
Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99-118.
Smith, G. P. (2012). Google Internet search activity and volatility prediction in the market for foreign currency. *Finance Research Letters*, *9*(2), 103-110.
Takeda, F. & Wakao, T. (2014). Google search intensity and its relationship with returns and trading volume of Japanese stocks. *Pacific-Basin Finance Journal*, *27*, 1-18.
Vlastakis, N. & Markellos, R. N. (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.
Wang, J. (1994). A Model of Competitive Stock Trading Volume. *Journal of Political Economy*, 102, 127–168.
Westen, T. (2000). E-Democracy: Ready or not, here it comes. *National Civic Review*, *89*(3), 217-228.
Xu, S. X. & Zhang, X. M. (2013). Impact of Wikipedia on market information environment: Evidence on management disclosure and investor reaction. *MIS Quarterly*, *37*(4), 1043-1068.

Table 1

Variable definitions.

Variable	Definition
Measures of investor attention	
SVI	Aggregate search frequency from Google Trends based on stock name
ASVI	The logarithm of SVI during the week minus the logarithm of median SVI during the previous 8 weeks
ASVICAC40	The logarithm of SVI for the term "CAC40" during the week minus the logarithm of median SVI during the previous 8 weeks
WIKITRAF	Average number of views for a given Wikipedia page in French
AWIKITRAF	The logarithm of WIKITRAF during the week minus the logarithm of median WIKITRAF during the previous 8 weeks
WIKIEDIT	Number of edits for a given Wikipedia page in French
Control variable	
NEWS	Number of news stories in the financial newspaper *Les Echos*
ANALYSTS	Number of analysts in I/B/E/S
TURNOVER	Trading volume
ABRETURN	Absolute value of characteristic-adjusted return as in Daniel et al. (1997)
ABTURNOVER	Abnormal turnover as in Da et al. (2011)
VOLATILITY	Historical volatility computed over the 24 weeks
MARKETCAP	Market capitalization

ACCEPTED ARTICLE

Table 2
Summary statistics, correlation matrix and Granger causality tests.

Panel A: Summary Statistics

	Mean	Std. Dev.	Min.	Max.
SVI	51.579	23.333	0	320
WIKITRAF	1237.81	1744.00	0	22814.12
WIKIEDIT	1.704	2.445	0	167
NEWS	2.651	4.892	0	67
ANALYSTS	11.044	4.918	0	17

Panel B: Correlation Matrix

	SVI	WIKITRAF	WIKIEDIT	NEWS	ANALYSTS	ABRETRUN	ABTURNOVER
SVI	1						
WIKITRAF	0.145**	1					
WIKIEDIT	0.090*	0.326***	1				
NEWS	0.093*	0.319***	0.236***	1			
ANALYSTS	0.067	0.218***	0.201***	0.022***	1		
ABRETRUN	0.039	0.151**	0.123**	0.198***	0.133**	1	
ABTURNOVER	0.024	0.222**	0.116*	0.167***	0.121**	0.215**	1

Panel C: Granger Causality Tests

Test 1: H_0: the Wikipedia measures fail to Granger predict the traditional measures of informed investor attention.
Test 2: H_0: the traditional measures of informed investor attention fail to Granger predict the Wikipedia measures.

Test 1: Causality from WIKIEDIT to	
ABRETURN	0.031
ANALYSTS	0.044
NEWS	0.032
ABTURNOVER	0.029
Test 1: Causality from WIKITRAF to	
ABRETURN	0.027
ANALYSTS	0.038
NEWS	0.024
ABTURNOVER	0.041
Test 2: Causality from ABRETURN to	
WIKIEDIT	0.288
WIKITRAF	0.382
Test 2: Causality from ANALYSTS to	
WIKIEDIT	0.173
WIKITRAF	0.301
Test 2: Causality from NEWS to	
WIKIEDIT	0.182
WIKITRAF	0.287
Test 2: Causality from ABTURNOVER to	
WIKIEDIT	0.335
WIKITRAF	0.287

Notes: This table presents the basic descriptive statistics (Panel A) and the Pearson correlation coefficients (Panel B) of the main variables used in the study for the period 2008–2018. The table also details the results of the Granger causality tests (Panel C) between Wikipedia measures and measures of smart investors' attention traditionally used in the literature. All variables are defined in Table 1. The tabulated statistics are the critical probabilities of the Granger test. The number of lags is determined by minimizing the Akaike and Schwarz information criteria (p=2).

ACCEPTED ARTICLE

Table 3
Investor attention and stock returns.

	Week 1	Week 2	Week 3	Week 4	Weeks 5-52
Panel A: The effect of noise traders' attention on stock returns					
ASVI	11.871***	10.012***	5.996***	-2.098*	-19.827**
Log(MARKETCAP)	3.009***	3.202***	2.182***	2.498***	1.253***
Log(MARKETCAP) × ASVI	-26.291***	-19.186***	-9.192**	5.133*	4.113*
ABRETURN	1.309***	-1.110	-1.106	-1.001	-0.528
Log(1+ANALYSTS)	-2.009**	-2.087**	-1.782*	-1.324	-1.244
NEWS	1.281*	1.104	1.102	1.065	1.104
ABTURNOVER	1.922**	2.100**	1.823*	1.725	1.702
Adjusted R^2	0.026	0.024	0.022	0.021	0.024
Panel B: The effect of smart investors' attention on stock returns (AWIKITRAF)					
AWIKITRAF	5.726**	4.0182*	3.087*	1.887	1.779
Log(MARKETCAP)	3.221***	3.114**	2.204***	2.376	1.088
Log(MARKETCAP) × AWIKITRAF	5.004**	3.302*	2.777*	2.263	2.197
ABRETURN	1.209***	-1.033	-1.014	-0.831	-0.398
Log(1+ANALYSTS)	-2.002**	-2.005**	-1.933*	-1.094	-1.332
NEWS	1.223*	1.234*	1.091	1.070	1.049
ABTURNOVER	1.933**	2.001**	1.902*	1.874	1.526
Adjusted R^2	0.023	0.022	0.022	0.020	0.022
Panel C: The effect of smart investors' attention on stock returns (WIKIEDIT)					
WIKIEDIT	4.032**	3.106*	2.044	1.997	1.802
Log(MARKETCAP)	3.010***	2.989**	2.202***	2.298	1.092
Log(MARKETCAP) × WIKIEDIT	5.066**	3.132*	2.898*	2.323	2.192
ABRETURN	1.267***	-1.021	-1.008	-0.816	-0.431
Log(1+ANALYSTS)	-1.832**	-2.002**	-1.924*	-1.077	-1.328
NEWS	1.210*	1.204*	1.074	1.055	1.044
ABTURNOVER	1.931**	1.991**	1.877*	1.882	1.643
Adjusted R^2	0.020	0.021	0.022	0.020	0.021
Panel D: The effect of both categories of investor attention on stock returns					
ASVI	11.651***	9.981***	6.267***	-2.244*	-18.788**
AWIKITRAF	5.422**	3.111*	2.117	1.956	1.772
Log(MARKETCAP)	3.088***	3.098***	2.126***	2.567***	1.119***
Log(MARKETCAP) × ASVI	-27.942***	-19.245***	-9.066**	5.233*	3.421*
Log(MARKETCAP) × AWIKITRAF	4.127**	3.088*	2.591*	2.293	2.008
ABRETURN	1.342***	-1.059	-1.069	-0.934	-0.535
Log(1+ANALYSTS)	-2.221**	-2.119**	-1.692*	-1.245	-1.249
NEWS	1.003	1.003	1.004	0.997	0.991
ABTURNOVER	1.892**	2.112**	1.768*	1.894	1.677
Adjusted R^2	0.027	0.026	0.024	0.023	0.024

Notes: This table reports the results from Fama-MacBeth (1973) cross-sectional regressions of equation (4). The dependent variable is the stock return during the first 4 weeks and during weeks 5 to 52. All variables are defined in Table 1. All independent variables are standardized. Standard errors are computed using the Newey-West (1987) formula with eight lags. *, ** and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from February 2008 to September 2018.

ACCEPTED ARTICLE

Table 4
Investor attention and historical volatility.

	(1)	(2)	(3)	(4)
INTERCEPT	-0.125***	-0.110***	-0.005	-0.126***
ASVI	0.051**			0.046**
AWIKITRAF		-0.024**		-0.021**
WIKIEDIT			-0.008	
Log(MARKETCAP)	-1.123**	-0.016**	-0.111**	-1.019**
Log(MARKETCAP) × ASVI	-0.155**			-0.151**
Log(MARKETCAP) × AWIKITRAF		0.129**		0.119**
Log(MARKETCAP) × WIKIEDIT			0.006	
MARKETRETURN	0.012*	0.014*	0.019*	0.016*
NEWS	0.008*	0.012*	0.013*	0.012*
ABTURNOVER	0.015**	0.012**	0.012**	0.011**
LAGGED VOLATILITY	0.022***	0.023***	0.025***	0.023***
ASVICAC40	0.019**	0.025**	0.029***	0.029***
Time fixed effects	YES	YES	YES	YES
Sector fixed effects	YES	YES	YES	YES
Adjusted R^2	0.315	0.304	0.298	0.324

Notes: This table reports the results from panel regressions as defined in equation (5). The dependent variable is the historical stock return volatility computed over 24 weeks. All variables are defined in Table 1. In each regression, we control for time and sector dummies. All independent variables are standardized. Standard errors are computed using the Newey-West (1987) formula with eight lags. *, ** and *** represent significance at the 10%, 5%, and 1% level, respectively. The sample period is from February 2008 to September 2018.

ACCEPTED ARTICLE