PONTIFICAL CATHOLIC UNIVERSITY OF PARANÁ
SCHOOL OF BUSINESS
GRADUATE PROGRAM IN ADMINISTRATION

JESSICA DE CASTRO

ESSAYS ON THE RELATIONSHIP BETWEEN INVESTOR ATTENTION AND ASSET PRICING

CURITIBA
2023

JESSICA DE CASTRO

ESSAYS ON THE RELATIONSHIP BETWEEN INVESTOR ATTENTION AND ASSET PRICING

Doctoral thesis presented to the Graduate Program in Administration, research line in Digital Marketing of the Business School from Pontifical Catholic University of Paraná, as a partial requirement to obtain a Doctoral degree in Administration. Under the guidance of Prof. Ph.D. Pedro Guilherme Ribeiro Piccoli.

CURITIBA
2023

Dados da Catalogação na Publicação
Pontifícia Universidade Católica do Paraná
Sistema Integrado de Bibliotecas – SIBI/PUCPR
Biblioteca Central
Sônia Maria Magalhães da Silva – CRB 9/1191

C355e
2023

Castro, Jessica de
Essays on the relationship between investor attention and asset pricing/ Jessica de Castro ; orientador: Pedro Guilherme Ribeiro Piccoli. – 2023.
123 f. ; il. : 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Paraná, Curitiba, 2023
Bibliografia: 115-123

1. Investidores (Finanças). 2. Modelo de precificação de ativos.
3. Administração. I. Piccoli, Pedro Guilherme Ribeiro. II. Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Administração. III. Título.

CDD. 20. ed. – 658

TERMO DE APROVAÇÃO

ESSAYS ON THE RELATIONSHIP BETWEEN INVESTOR ATTENTION AND ASSET PRICING

Por

JESSICA DE CASTRO

Tese aprovada em **11 de setembro de 2023** como requisito parcial para obtenção do Título de Doutora no Programa de Pós-Graduação em Administração, Área de Concentração em Administração Estratégica, da Escola de Negócios da Pontifícia Universidade Católica do Paraná.

Profª. Dra. Angela Cristiane Santos Póvoa
Coordenadora do Programa de Pós-Graduação em Administração

Pedro Piccoli

Prof. Dr. Pedro Guilherme Ribeiro Piccoli
Orientador

Prof. Dr. Newton Carneiro Affonso da Costa Júnior
Examinador

Profª. Dra. Andrea Maria Accioly Fonseca Minardi
Examinadora

Profª. Dra. Cláudia Emiko Yoshinaga
Examinadora

Prof. Dr. Francisco Henrique Castro Júnior
Examinador

To my parents.

ACKNOWLEDGMENTS

I thank God, as there is nothing without Him.

I extend my gratitude to my most excellent advisor, who is, above all, humane. What an example of a teacher, a mentor, a person. I aspire to be even a fraction of what you are when “I grow up”; there's no escaping the superlatives here – a heartfelt thank you. I admire you, and I will carry your surname with immense pride throughout my academic journey.

To the committee examiners for their availability in reviewing and evaluating the research.

To my parents, who gave me so much with so little. Completing a doctoral thesis in finance and realizing that even though you were practically illiterate, it was you who taught me the fundamental pillars of finance for life – it's truly an honor.

To Dibas, for staying even when everything was chaos. One of the greatest challenges is finding someone willing to make sacrifices in the name of our goals. Here we are after the storm, I am ready to live out yours!

To my friends, who were my mental refuge. Your presence, company during my beloved coffee breaks, the always unfunny jokes, and the "hang in there, it'll be over soon" phrases provided the energy to keep going.

To my friend Carol, it was your encouragement that helped me overcome my fear of not being accepted into a doctoral program. It seems like everything worked out, didn't it? Undoubtedly, our lives have changed.

Thanks to UNICENTRO for granting me the license to complete my professional qualification.

Thanks to Professor Marcos Kuhl, as the department head, who promptly accommodated my requests to make travel for my doctoral activities feasible.

Thanks to PUCPR for the warm welcome, thank you to Professor Damião who in a distant past made himself absolutely accessible at a conference, showing the possibility of one day studying at this institution.

Finally, thank you to Capes for the financial support and facilitation of the development of this course.

"Living without reading is
dangerous, because it forces
you to believe in what they tell
you." (Mafalda)

ABSTRACT

Standard asset pricing models typically rely on the assumption that markets swiftly incorporate new available information, thus enabling the best possible estimation of asset values. In reality, this information supply and estimation require significant attention from investors to process and incorporate into their decisions. It's important to note that classical theory assumes that not all investors have access to accurate and timely information, reducing their ability to efficiently process available information, leading to noise-based trading. Investor attention can play a pivotal role in asset pricing, as it impacts the demand for different assets and subsequently their market prices. In this research, organized into three essays, we examine the effects of investor attention on asset price dynamics. The studies employ a quantitative data analysis approach using secondary data from the Brazilian stock market. Within these essays: (1) We ascertain whether online searches are a coherent proxy for the presence of retail investors in the Brazilian stock market, addressing a significant factor affecting research on individual investors – the lack of access to noise trader behavior; (2) We utilize the online search proxy to investigate the attention-return relationship in the gold spot market under different market conditions; and (3) We explore whether value dynamics are more crucial in explaining attention changes compared to noise variables, given the belief that individual investors trade on noise. Overall, the results (1) strongly support the view that online searches are a coherent proxy for the presence of retail investors in the stock market, as online searches exhibit a strong association with future retail investor trading, observed across various market conditions. Additionally, alternative approaches commonly used in literature to capture investor attention via online searches consistently prove to be reliable measures of future investor trading. (2) We further document that the attention-return relationship in the gold spot market is sensitive to market states, as we identify a positive (negative) association between returns and contemporary attention during bull (bear) markets. This sensitivity is influenced by the magnitude of returns, as the significance of the relationship, based on a quantile model, almost monotonically increases as return distribution diverges from the median. Moreover, we discover that investor attention is influenced by recent trends in gold spot prices in both market states. Overall, these findings suggest that the past performance of gold returns attracts the attention of individual investors who subsequently trade to profit from the trend, reinforcing price movements, consistent with the temporal momentum framework. Lastly, (3) the results also indicate that fundamental variables significantly affect investor attention. Sensitivity analysis conducted through univariate regression for each variable underscores this finding by revealing that fundamental dynamics remain more important for investment decisions than dynamics of noise variables. Therefore, these studies contribute important empirical insights to the finance field by exploring aspects related to the individual investor approach and investor attention.

FIGURE INDEX

THESIS STRUCTURE

The thesis is structured into three main parts: the first is the research characterization, the second encompasses the original articles (submitted/published), and finally, the concluding remarks of the studies.

PART 1	**CHARACTERIZATION OF THE RESEARCH**
	It covers the introductory section of the studies, contextualizes the research problem, identifies practical and theoretical justifications, general objectives, theoretical framework, and theories adopted in subsequent studies.
Section 1.1	Presents the research context.
Section 1.2	Introduces the guiding research question of the thesis.
Section 1.3	Identifies the main objectives of the study.
Section 1.4	Justifications.
Section 2	Identifies the structural concepts and theoretical context of the studies.
Section 3	Presents the developed studies and how they are grouped within a continuous conceptual and practical context.
PART 2	**DEVELOPED ARTICLES**
	Part 2 includes the three main articles that are part of this thesis. The arrangement of the articles is described in Section 3 of Part 1.
Article 1	*Do online searches actually measure future retail investor trades?*
Article 2	*Attention-return relation in the gold market and market states*
Article 3	*Where are retail investors looking?*
PART 3	**CONCLUSIONS**
	Presents the conclusion of the studies, identifies study limitations, and outlines future research for the academic development of the topic.

SUMMARY

PART 1

The first part of the document presents the study, the context, the justifications, research questions, and objectives. This section also demonstrates the contribution of each developed article to achieving the overall objectives of the thesis, explores the common theme among the research, and provides a holistic view of the studies conducted.

1 INTRODUCTION

1.1 Presentation of the theme.

As a contribution to the context of modern finance, studies on human behavior and irrationality have been evolving. These studies, aligned with concepts from other fields like Psychology and Sociology, aimed to explain individuals' financial decisions that deviated from the prevailing theory at the time and to bridge the gap between economic and financial concepts. This stream of finance considers the investor's limited rationality, allowing emotions and interpretive biases to influence their financial decisions. Consequently, this individual behavior can influence the market, contradicting the Efficient Market Hypothesis (EMH) (De Bondt, 1998). In this sense, old criticisms of the EMH have found a place in new research.

Investor attention is a central theme in behavioral finance literature. The specialized literature on investor attention operates under the assumption that attention is a scarce resource. In other words, as investors cannot keep up with all market actions or assimilate all relevant information, it's crucial for them to make a judicious selection of specific actions to monitor (Hirshleifer and Teoh, 2003). Various factors can affect investor attention, such as the volume of available information, the complexity of financial data, emotional burden in investment decisions, among others. Moreover, attention is a limited resource over time, meaning investors need to prioritize the information they want to absorb at any given moment. A growing body of empirical evidence suggests that investor attention fluctuates over time and that this affects asset prices (Da *et al.*, 2011). For instance, high levels of attention cause buying pressure and sudden price reactions (Barber and Odean, 2008; Barber *et al.*, 2009), while low levels lead to minimal reactions to announcements (Dellavigna and Pollet, 2009). In fact, prices only react to new information when investors pay attention to it.

To manage investor attention, it's important for individuals to adopt strategies for selecting relevant information and maintaining focus on actions they deem promising. These strategies can include using technical and fundamental analysis tools, keeping up with financial news, and analyzing past stock performance. Standard asset pricing models generally assume that markets quickly incorporate new information, enabling the best possible estimation of asset values. In reality, processing this information requires significant investor attention to process and incorporate into their decisions. It's worth noting that classical theory assumes that not all investors have access to accurate and timely information and that not all investors can efficiently process available information. These investors can be influenced by sensational news, unfounded investment tips, or irrational behaviors, leading to noise-based trading. Noise-based trading deviates from the principles of market efficiency that assert asset prices reflect all available information, and investors are rational and use correct information in making investment decisions.

Investor attention can play a significant role in asset pricing as it affects the demand for different assets and, consequently, their market prices. Investors who are attentive to relevant news and economic events can quickly react, buying or selling assets based on this information. This can lead to fluctuations in asset prices, even without significant changes in underlying economic fundamentals. Additionally, investor attention can be influenced by emotional factors such as fear, anxiety, and excessive optimism. This can lead to irrational behaviors, like buying or selling assets based on rumors or hearsay rather than solid economic fundamentals. As a result, asset prices can be distorted in relation to their economic fundamentals, especially in the short term. These distortions can be corrected over time as investor attention shifts and economic fundamentals adjust.

Several recent studies (Welch, 2022; Cheng *et al*., 2021; Desagre and D'Hondt, 2021; Chen, 2020; Baur *et al*., 2020) suggest that investor attention can play a significant role in determining asset prices. Important news or information isn't reflected in prices until investors pay attention to it. Despite the growing empirical evidence, we identify some unexplored gaps, such as the relationship between the Google proxy and retail trading, investor behavior in specific markets, and individual investor behavior characteristics. In this research, we study the effects of investor attention on asset price dynamics. We emphasize that attention is a scarce cognitive resource (Kahneman, 1973). Paying attention to a task inherently requires the reallocation of cognitive resources from other tasks. In the realm of investment decisions, given the abundance of available information and the inevitability of limited attention, investors must be selective in processing information.

Merton (1987) is one of the early studies that demonstrate the significance of investor attention in influencing security prices. Despite this, the exact role of information and investor attention in market equilibrium and market efficiency remains uncertain. The influential work of Grossman and Stiglitz (1980) suggests that more information (an increased number of informed investors in their model) leads to more informative prices, implying a more efficient market. If heightened attention results in more information being absorbed by the market, then attention should enhance market efficiency. However, some researchers, such as Da *et al*. (2011), who explore the attention of retail investors, argue that excessive attention might generate additional noise, thereby potentially reducing market efficiency. The challenge partly lies in the significant difficulty of determining the appropriate measure of attention.

Numerous studies have explored proxies for investor attention based on internet search mechanisms (Da *et al*., 2011, Da *et al*., 2015, Vlastakis and Markellos, 2012, Joseph *et al*., 2011, Mondria *et al*., 2010). The approach adopted by these studies involves the use of keywords in an attempt to capture investor interest in specific assets, such as stocks, cryptocurrencies, commodities, and more. While Da *et al*. (2011) assert that this approach is a direct and unequivocal measure of attention, it's important to recognize that keywords might not fully capture the researcher's trading intention. The chosen keywords could represent interest in other situations related to the research topic, rather than indicating investment interest. One argument supporting the use of such keywords as a measure of attention is the notion that professional investors wouldn't use company codes as search queries on platforms like Google, as they have access to more sophisticated search tools.

Within the limitations of each study, a significant challenge arises from the lack of access to noise trader behavior, as mentioned earlier. Various proxies are employed to measure investor attention, including trading volume, news announcements, media coverage, and even price limits. The main argument linking attention is that if an asset's value generates a higher return, investors will focus their attention on these movements. However, some research highlights that proxies such as volume, returns, and media news might not effectively fulfill their role in capturing attention (Brooks, 1998; Dimpfl and Jank 2012; Da *et al*., 2011). As an alternative, the volume of keyword searches on search engine sites has also been utilized as a proxy to capture non-sophisticated investors' intention to buy specific assets. Noise traders have less access to sophisticated tools, thus directing their attention only when they're interested in obtaining information about specifics assets.

1.2 Research Problem

Studies document the importance of investor recognition and attention for asset pricing and trading (Merton, 1987; Barber and Odean, 2008). Within the limitations of each study, one of the challenges faced lies precisely in the lack of access to noise trader behavior. To overcome this issue, most attention-related studies use indirect proxies for investor attention. The proxies used to measure investor attention have been represented as excessive trading volume and stock returns (Barber & Odean, 2008), new news announcements (Chan, 2003; Fang and Peress, 2009), advertising (Chemmanur and Yan, 2009; Grullon *et al*., 2004), and even price limit events (Seasholes and Wu, 2007). The main argument related to attention is that if the asset value yields higher returns, investors will pay attention to these movements; however, some research indicates that such proxies may not effectively capture attention (Brooks, 1998; Dimpfl and Jank, 2012; Da *et al*., 2011).

In the search for a preferred proxy that better represents investor attention, new alternatives have emerged. The most utilized channel in recent years has been the Internet, as this tool has democratized and facilitated access to information for non-sophisticated investors. However, this ease of information access represents not only advantages but also potential problems, given that an excessive volume of information is available and is not always processed rationally. This can contribute to the unpredictable behavior of the noise trader, who may act on false or untimely information, among other factors (Choi *et al*., 2020).

Arising from this growing expansion of online information, options for measuring investor attention have emerged with the widespread use of the internet. As an alternative, the volume of keyword searches on search engine websites has also been used as a proxy to capture individual investor attention. The noise trader has less access to sophisticated tools and therefore directs attention only when interested in obtaining information about specific assets. Da *et al*. (2011) argue that aggregated search frequency on Google is a direct and unequivocal measure of investor attention. Additionally, several recent studies also utilize the Google Trends tool as a proxy for investor attention (Choi and Varian, 2009; Da *et al*., 2011; Vozlyublennaia, 2014; Baur and Dimpfl, 2016; Swamy *et al*., 2019; Meshcheryakov and Winters, 2020).

Although there is a growing interest in research on the use of online search mechanisms as a proxy capable of capturing and measuring investor attention, there are still few studies that test the possibility of verifying whether search mechanisms indeed represent the non-sophisticated investor's trading intent for specific assets. This scenario highlights the issues that permeate the study of investor attention and its influence on capital market behavior, which this work aims to help address.

First, a natural obstacle to investigating the influence of retail investors on markets is related to the (un)availability of data, as retail transaction data is not publicly accessible. This lack of information hinders a complete and accurate understanding of retail investor behavior and their influence on financial markets. Given this limitation, research relies on proxies to overcome this deficiency and obtain information about noise trader behavior. However, the use of indirect measures can compromise the quality of results and introduce significant biases into such research. Our main contribution aims to provide greater reliability to the proxy proposed by Da *et al*. (2011), which has been used in numerous studies in recent years: the use of Google search volume as a way to operationalize individual investor attention (Choi; Varian, 2009; Da *et al*., 2011; Vozlyublennaia, 2014; Baur and Dimpfl, 2016; Swamy *et al*., 2019; Meshcheryakov and Winters, 2020). In this context, we seek to identify whether online searches are a coherent proxy for the presence of retail investors in the stock market, addressing one of the main challenges affecting research on individual investors, namely the lack of access to noise trader behavior. In this regard, in partnership with the Brazilian Stock Exchange, we gained access to data specifically related to the trading volume of retail investors in this market, thereby enhancing the reliability of results related to this category of investors' behavior.

Second, general evidence indicates that investor attention plays a significant role in understanding the behavior of various markets. Studies in this area (Jain and Biswal, 2019; Chen, 2020; Baur *et al.*, 2020) consistently demonstrate that investors are susceptible to cognitive and emotional biases that affect their investment decisions. These biases can be influenced by selective attention to certain assets or information. For instance, investors heavily focused on a particular sector might ignore important information about other sectors, leading to price distortions and investment opportunities. In light of this, we intend to investigate the attention-return relationship in the gold spot market using the proxy of online search activity.

Finally, individual investors are often characterized as traders who rely on less reliable information due to their limited experience and knowledge. This implies the possibility that they make investment decisions based on less precise information, referred to as noise. While this assertion may hold true in some cases, it is essential to acknowledge that generalizing all individual investors in this manner is simplistic and could lead to erroneous conclusions. It is common knowledge that, compared to institutional investors, individual investors generally have less access to privileged information and specialized resources. They may rely more on public sources of information, such as financial news, social media, and broker recommendations, which are not always reliable or accurate. Additionally, individual investors may have less knowledge of fundamental analysis and asset valuation techniques, which can impact their investment decisions. However, it is crucial to highlight that not all individual investors fit this profile. It is plausible that there are well-informed, experienced, and disciplined individual investors who conduct thorough analyses and make informed decisions based on reliable information. These investors may dedicate time and effort to research and obtain relevant knowledge about the markets they invest in. Therefore, we aim to identify whether value dynamics are more critical in explaining changes in attention than noise-related variables.

1.3 Research Objectives

1.3.1 Overall Objective

Analyzing the potential relationship between investor attention and various aspects of asset pricing

1.3.2 Specific Objectives:

- To examine whether online searches are indeed related to future trading activities of retail investors, as assumed by various studies on investor attention (First Study);

- To investigate the attention-return relationship in the gold spot market and its dependence on market states (Second Study);

- To identify whether value dynamics are more significant in explaining changes in attention compared to noise variables for individual investor attention (Third Study).

1.4 Theoretical and Practical Justification

This study presents three important contributions to the literature. Firstly, our study significantly contributes to the investor attention literature, particularly when conducting an empirical test on Da *et al*. (2011) proposition, in which the authors claim that the proxy based on keyword searches on search engine websites, specifically in this case, Google Trends, "is an accurate and unequivocal" measure of investor attention, as the authors did not conduct tests themselves to prove the effectiveness of the SVI (Search Volume Index) as a measure capable of capturing the real purchasing intention of retail investors. It is worth noting that Da *et al*. (2011) study, as well as several other related studies on the subject (Joseph *et al*., 2011; Vozlyublennaia, 2014; Goddard *et al*., 2015; Baur and Dimpfl, 2016; Yung and Nafar, 2017; Han *et al*., 2017; Jain and Biswal, 2018), correlate the established proxy with overall market returns, assuming that professional investors do not influence asset returns. Specifically, Da *et al*. (2011) found a positive and significant relationship between the Google Trends indicator and stock market indices. Therefore, among the study's limitations, there could be biases (stemming from institutional investors) that influence the obtained results.

Our research stands out primarily for segmenting market investor profiles and successfully relating the proxy proposal for investor attention to data related solely to individual investors in the Brazilian market. This allows for a more careful and objective analysis regarding the efficiency of using search mechanisms related to investor attention. Our results highlight the SVI as a tool capable of measuring investor attention. Thus, this initial study contributes to the literature by addressing one of the major limitations in investor attention research, which lies in the difficulty of measuring this variable.

Having assumed keyword search mechanisms as a secure proxy for investor attention, we further explored the behavior of individual investors in the market. Therefore, our second contribution stems from the theory of investor attention, which argues that unsophisticated investors are more likely to buy assets when sentiment is optimistic (Baker and Wurgler, 2006), thus influencing their prices. While research has been conducted (Da *et al*., 2011; Joseph *et al*., 2011; Vozlyublennaia, 2014; Goddard *et al*., 2015; Yung and Nafar, 2017; Dastgir *et al*., 2019) focusing on specific assets, and although the widespread use of search mechanisms is well established, the literature on the dynamics of internet search trends related to the commodity gold remains limited. Therefore, the second study in this thesis aims to contribute to gold and Google Trends literature by using SVI data regarding global searches for the price of this precious metal. Research on gold and Google Trends (Baur and Dimpfl, 2016; Han *et al*., 2017; Jain and Biswal, 2018) has been developed addressing price volatility, the relationship between price changes and news, expectations of future returns, and even pointing to causal relationships in specific markets like India. However, our study investigated the behavior of the price of this asset, considering different market conditions by separating it into moments of highs and lows. This examination helps discern the influence of noise traders on market movement, especially during moments of highs (positive returns) when investor sentiment tends to be more optimistic.

Lastly, investors, in general, have access to an ever-increasing array of information from various sources. However, literature often mentions the challenge people face in assimilating all available information due to their limited cognitive resources (Kahneman, 1973). Even with abundant information, investors might not be able to pay attention to and select truly important information that aids in making informed investment decisions and subsequent portfolio performance. From the wealth of information, investors with limited attention must allocate it efficiently across different assets and over time (Han *et al*., 2018). Prices react to new information only when investors pay attention to it. Attention allocation precedes portfolio allocation and can lead to infrequent portfolio decisions, affecting aspects of asset price dynamics, such as return predictability (Huberman and Regev, 2001).

Despite the extensive literature on investor attention effects, information about the characteristics that attract the attention of individual investors who choose to trade assets remains relatively undocumented. This leads to questioning what truly affects investors' attention allocation among different types of information they are exposed to. Hence, this aspect of the research is relevant to the literature, considering the empirical evidence reported here that supports the assumption that individual investors make efficient investment decisions (Kelley and Tetlock, 2013; Chen *et al*., 2016; Dahlquist *et al*., 2017). This goes against the mainstream theoretical view that regards retail investors as noise traders (Black, 1986; Shleifer and Summers, 1990; De Long *et al*., 1990). Additionally, by documenting the characteristics of assets that attract the most attention from investors, we support the proposition that retail investors are not irrational; they buy in bear markets and sell in bull markets (Chau *et al*., 2016).

In addition to theoretical contributions, this study also offers practical insights. It considers the chosen market's profile, which has been seeing a record influx of new investors year after year, along with a growing number of new companies going public on the Brazilian stock exchange. This characteristic has spurred interest in the research. The decision to study the Brazilian market was driven by various factors. Firstly, we obtained access to data from the Brazilian market as a whole, provided by the Brazilian Stock Exchange, which allowed us to access overall buying and selling figures, as well as the percentage of retail investor participation in trades. This provides substantial support for the analyses conducted.

Another interesting aspect presented by the Brazilian market is the significant increase in the number of individual investors entering the Brazilian Stock Exchange. The number of individual accounts grew from 560,000 in 2016 to 5,000,000 by December 2022, representing an over 8-fold increase during this period. This rise, combined with the theoretical perspective, is related to the representativeness that noise investors gain in the stock market. Considering the Herd Behavior Model, even though irrational investors exist, they aren't initially considered relevant enough to influence market movements. However, as the volume of trades by noise investors grows, the impacts of their cognitive and behavioral-based decisions might indeed start affecting market behavior. Additionally, this scenario demonstrates the increasing interest of Brazilian investors in stock market assets.

Another characteristic of the Brazilian market that contributes to the methodology adopted in this research is the specificity of the names of assets traded on the Brazilian Stock Exchange. Tickers with a specific standard are used to identify assets, making them a reliable source of information on investor interest. This characteristic has been discussed as a limitation in studies by Da *et al*. (2011) and Joseph et al. (2011). In their studies, searching for a keyword with the company's name can inadvertently bias the reason for the searches. An individual conducting a search might be interested in other aspects related to the company, such as online purchases. This could distort the demand for information. In this context, the Brazilian scenario is more congruent in data collection, as searches for specific tickers restrict interest solely to information about the asset itself. Challet and Ayed (2013) note that the choice of keywords is a crucial ingredient when using Google Trends for analysis.

2 BACKGROUND THEORETICAL

2.1 Efficient Market Hypothesis

The scope of Modern Finance Theory is primarily based on the Efficient Market Hypothesis (EMH). This hypothesis has its origins in the doctoral thesis of Louis Bachelier (1870-1947), titled "Theorie de la speculation." Bachelier (1900) describes that financial asset prices are comparable to a "random walk," which asserts that expected returns are independent of available information.

Although the EMH is quite old, it gained prominence only in 1970 with the publication of Eugene Fama's work. According to Fama (1970), a market is efficient when prices fully reflect all available information. In other words, it's the ability of financial markets to absorb available information about a particular asset and incorporate it into its trading price.

To test the EMH, Fama (1970) relied on a market equilibrium model that allows asset pricing, as presented by Sharpe in 1964. This model addresses the theory of market equilibrium of asset prices under conditions of risk, aiming to clarify the relationship between an asset's price and various components of its overall risk (Sharpe, 1964).

Considering the EMH's assumption that price incorporates relevant information about the asset, this theory suggests that an investor is unlikely to consistently outperform the market in the medium term, except through random movements (Shleifer, 2000). This is because market efficiency relies on the randomness of asset price fluctuations, making it impossible for an investor to consistently achieve returns above the market.

Fama (1970) explains that the conditions that determine an efficient market are primarily based on three aspects: first, there are no transaction costs in asset trading; second, all information is available at no cost to all market participants; and third, there is no information asymmetry, meaning everyone has the same interpretive ability, resulting in a consensus on asset pricing. In this context, it can be said that the current price of an asset fully reflects available information.

Therefore, Fama (1970) discusses price behavior in three aspects: the "Fair Game" model, the "submartingale" model, and the "random walk" model. The first is based on average returns behavior rather than the entire probability distribution. The fair game model doesn't imply positive asset returns; rather, it suggests unbiased expectations. Hence, in a relatively large sample, the expected return of an asset is, on average, equal to the current return. In the submartingale model, tomorrow's price is expected to be higher than today's price, indicating positive expected returns. Lastly, the random walk model posits that asset prices behave randomly, rendering trend projection and predictability infeasible.

Through the lens of the Efficient Market Hypothesis, the ideal market is one where prices accurately reflect all available information. Market efficiency has been classified into three forms. Initial tests define the first form, known as the weak form of market efficiency. In this form, the market fundamentally incorporates past information about assets. In this case, today's price is a function of yesterday's price plus the expected return and a random component derived from asset information.

The second form, the semi-strong form of market efficiency, occurs when the asset incorporates not only past behavior but all publicly available information about the asset's price, including financial statements and relevant news disclosures.

Lastly, the third form, the strong form of market efficiency, incorporates all types of available information, whether public or not. In this case, any information, even that known by an insider, is factored into the asset price (Jensen, 1978).

In these tests, Fama (1970) sought to identify the weak form by measuring how well past returns predict future returns. If patterns of behavior, investment schemes, or price correlations with variables can be found, then the market is considered weak form inefficient.

For the semi-strong form, tests seek to specify how quickly asset prices reflect public information, such as specific news and announcements of profit distribution. The faster prices adjust due to a certain event, the more efficient the market is considered, as it offers less opportunity to earn abnormal profits through exploiting that information. Finally, there are tests to evaluate efficiency in the strong form, where prices reflect not only public information but all obtainable information, including insider information. Tests for this form of efficiency aim to detect whether any insider or investor possesses privileged information that isn't fully reflected in prices and whether they could benefit from such information by earning abnormal profits (Fama, 1970). In general, EMH tests aim to identify inefficiencies and their nature in the studied markets. Much of the criticism of EMH is directed at its strong form, precisely due to the difficulty in measurement.

Jensen (1978) comments that the strong form of efficient market hypothesis is an extreme form, one that few people have treated as anything other than a logical conclusion from a set of possible hypotheses. The semi-strong form of Efficient Market Hypothesis, on the other hand, represents the accepted paradigm and is what is generally understood more broadly as the efficient market hypothesis. In semi-strong efficient markets, asset prices should rapidly adjust as new information emerges, and changes occur from period to period in the risk-free interest rate level and asset risk premiums.

In summary, there are three main arguments that underlie the EMH. First, the EMH assumes that investors are rational, leading to rational assessment and pricing of assets. The second argument suggests that the participation of irrational investors in the market is random, canceling out possible gains. The final argument posits that even if irrational investors are present in the market, rational arbitrageurs predominate and neutralize the effect of non-sophisticated investor trading (Schleifer, 2000).

Market efficiency doesn't occur naturally or autonomously; it's the actions of agents that make them more or less efficient. By seeking the greatest gains, each agent contributes to the market's efficiency as a whole because continuous trading nullifies the advantages that an investment scheme might provide. In this sense, given that agents' actions determine market efficiency, various studies (Schleifer; Summers, 1990; De Long et al., 1990; Grossman, 1975; Black, 1986; Barber; Odean,

2007) recognize the presence of non-rational investors in the market, unlike the argument defended by the EMH, which assumes all investors are rational and therefore their decisions will be too.

Next, we present an approach to the non-rational investor responsible for anomalies that contradict market efficiency. A possible explanation for such anomalies is provided by Behavioral Finance, which claims there's a non-rational pattern in individuals' decision-making process. This opens up the possibility that unsophisticated investors influence the market, as their decisions are not random but follow a pattern.

2.2 Noise Trader Approach

As previously mentioned, one of the assumptions of EMH states that even though irrational investors are present in the same market, rational arbitrageurs prevail and counteract the effects of non-sophisticated investors' trades. In other words, within the same market, two types of investors coexist, both having access to the same assets and trading opportunities (Barber and Odean, 2008). According to EMH, rational investors are responsible for making decisions that maintain the market in equilibrium, while retail investors engage in trades that are not significant enough to influence the market. Investors can be divided into those who trade based on information (professional and institutional investors) and those who trade based on noise (noise traders), who are essential for maintaining market liquidity. Since all transactions are assumed to process information rationally, the market could collapse if not for noise traders, as the possibility of gains would be null (Grossman, 1975; Black, 1986).

In this context, the research by Schleifer and Summers (1990) and De Long *et al*. (1990) plays a pivotal role in the formation of the theoretical framework called the "Noise Trader Approach," which challenges the assumption of the Efficient Market Hypothesis. The noise trader approach is based on two main assumptions: firstly, some investors are not entirely rational, and their demand for risky assets is influenced by beliefs or sentiments not entirely justified by fundamental news. Consequently, assets might not efficiently reflect available information, causing price deviations from their intrinsic values. The second assumption concerns arbitrage, where arbitrageurs, rational investors, tend to make rational decisions when there is a price deviation, contributing to restoring the asset's price to its fundamentals. Such investors are not influenced by beliefs or sentiments (Schleifer and Summers, 1990).

One distinguishing characteristic of these investors is that while institutional investors focus on specific sectors or rely on performance indicators and professional advice, individual investors tend to trade assets that capture their attention. This might be due to news in the media, unusual trading volume, or assets showing extraordinary returns (Barber and Odean, 2007).

Two perceptions exist regarding asset pricing discussions. One acknowledges the presence of a significant volume of noise traders operating in the market. On the other hand, defenders argue that only rational investors have the power to influence asset pricing. The defense that non-rational investors have no significance in the market is evident in EMH, stemming from previous arguments by Friedman (1953) that irrational investors do not affect the market. This assertion is grounded in the fact that the demands of irrational investors are offset in the market by rational arbitrageurs who trade against them, ultimately aligning prices with fundamental values. Moreover, non-sophisticated investors make uninformed decisions, leading to losses and their eventual disappearance from the market.

However, according to the Noise Trader Approach, the unpredictability of noise traders' beliefs creates a risk in asset prices that prevents rational arbitrageurs from aggressively betting against them. As a result, prices can significantly deviate from fundamental values, even in the absence of fundamental risk, due to the presence of both non-rational investors and implementation costs that limit arbitrageurs' actions in eliminating price differences between intrinsic and market values. This theory argues that it might be difficult for sophisticated investors to reverse deviations caused by less rational investors (Shleifer and Vishny, 1997). Furthermore, sustaining a disproportionate amount of risk that they themselves create enables noise traders to achieve higher expected returns than rational investors.

The model sheds light on various financial anomalies, including excess volatility in asset prices, mean reversion of stock returns, underestimation of closed-end mutual funds, and the Mehra-Prescott puzzle (De Long *et al*., 1990; Shiller, 1984). One reason for this behavior resulting in anomalies that don't follow a rational decision pattern is the so-called investor sentiment, which can be defined as optimism or pessimism and overall overreactions. According to Baker and Wurgler (2006), one possible definition of investor sentiment is a propensity for speculation. The authors argue that sentiment drives relative demand for speculative investments, thus causing cross-sectional effects. Even if arbitrageurs act, the effects remain on assets. The main effect of the weighting function is overweighting the distribution's tails it's applied to. This overweighting doesn't represent a bias in beliefs; it's a modeling device that captures the common preference for a wealth distribution resembling a lottery (Kahneman and Tversky, 1982).

In this context, Barber and Huang (2008) identify a positive asymmetry in investors' risk preference. The evidence lies in the human tendency to gamble. Even if the probability of gains is low, investors tend to take risks. This behavior in the market is likened to lottery games, where even though a player has an extremely low chance of winning the lottery, they compare it to the zero probability if they don't play, justifying their gambling action. This behavior distorts the small probability of actual gains that can be achieved and demonstrates the risk attraction characteristic of this investor profile. Other research also supports the findings of Barber and Huang (2008), including Kumar (2009), Bali *et al*. (2010), Albuquerque (2012), Amaya *et al*. (2012), Andrei and Hasler (2015), and Mitton and Vorkink (2016).

Recent research has sought to identify turbulence in the financial market, represented by investor attention, to better understand the behavioral factors influencing decisions (Barber and Odean, 2008; Huddart *et al*., 2012; Dimpfl and Jank, 2012). Besides behavioral biases, individual investor attention levels also contribute to the imbalance of stock market transactions. For noise traders, purchases attract more attention than sales, as buyers must choose from a vast array of available securities, while sellers can only sell what they already own. This can result in abnormal returns and temporary buying pressure (Barber and Odean, 2008).

As attention is not directly observable, proxies capturing part of its variations are adopted. Examples include excessive trading volume and stock returns (Barber and Odean, 2008), new news announcements (Chan, 2003; Fang and Peress, 2009), advertising (Chemmanur and Yan, 2009; Grullon *et al*., 2004), and price limit events (Seasholes and Wu, 2007). These studies have yielded results consistent with the investor attention hypothesis. However, as these measures indirectly capture attention, variables such as transaction volume and returns may not satisfactorily capture the degree of investor attention. Some studies, for instance, did not find significant evidence of the relationship between transaction volume and price volatility in their tests (Brooks, 1998; Dimpfl; Jank, 2012). Similarly, Da *et al*. (2011) noted that media news releases do not reliably proxy investor attention. According to Desagre and D'Hondt (2020), the primary deficiency is that these metrics reflect information supply, not demand.

The internet has significantly contributed to information access for investors, a fact that has sparked interest in the relationship between online information and investors. It's worth noting that the online environment, due to its information overload, can also contribute to the unpredictable behavior of noise traders (Choi *et al*., 2020). As a result of this growing expansion of online information, alternatives for measuring investor attention have emerged with the widespread use of the internet, such as Wikipedia, Twitter, Baidu, and the Google Search Volume Index. Compared to metrics mentioned as proxies for investor attention, the SVI (Search Volume Index) allows for the direct capture of demand for information in a more specific manner (Desagre; D'Hondt, 2020) and will be further discussed in the next topic.

2.3 Google Trends and Attention Trades

Google Trends has been addressed in various studies as a proxy capable of capturing investor attention (Choi and Varian, 2009; Da *et al*., 2011; Joseph *et al*., 2011; Vozlyublennaia, 2014; Bijl *et* *al*., 2016; Baur and Dimpfl, 2016). It is a tool that measures the volume of online searches for specific terms that individuals perform on the google.com search engine. Google, occupying the largest share of search engine popularity globally, makes Google Trends a tool to reveal information demand. This feature provides an index that portrays the search volume, which is called SVI (Search Volume Index). The first records of this platform date back to 2004, and the official launch occurred in 2006, being since then used in various research studies.

Recently, Google search volume has been widely used in financial literature to assess the attention of noise traders in different contexts. The work of Ettredge *et al*. (2005) was pioneering in the use of Google Trends in the economic context, focusing on analyzing unemployment series in the United States. They concluded that the Google Trends series is related to unemployment during the sample period, suggesting that this type of input information can help predict macroeconomic variables.

Choi and Varian (2009) described how to use Google Search Insights data to predict various economic metrics, including initial unemployment claims, demand for cars, and vacation destinations. More recently, Da *et al*. (2015) constructed an investor sentiment indicator called FEARS based on Google Trends data analysis. Their results allow for predicting short-term return reversals, temporary increases in volatility, and mutual fund flows in equity and debt funds.

Studies using this proxy are currently being applied to various asset classes. In the stock market, the seminal study by Da *et al*. (2011) proposed a new measure for investor attention, relating stocks belonging to the Russell 3000 index. Correlations between stocks and SVI were identified, particularly capturing individual investor attention. Furthermore, the results also showed a correlation between SVI and IPO. Along the same lines, Bank *et al*. (2011) state that Google search volume captures investor attention in the stock market, particularly that of uninformed investors. Joseph et al. (2011) examined the ability of online quote searches to predict abnormal stock returns and trading volumes, highlighting the potential of employing online search data for other forecasting applications.

In the stock market context, Vozlyublennaia (2014) investigated a link between the performance of various stock index categories and investor attention, as measured by Google search probability. They found a significant short-term change in index returns after an increase in attention, demonstrating that increased investor attention reduces return predictability and thus enhances market efficiency.

Bijl *et al*. (2016) investigated whether Google Trends data could be used to predict stock returns. Their results corroborate previous studies, showing that high levels of Google search volume predict high returns in the first two weeks, followed by price reversals. In the realm of currencies, Goddard *et al*. (2015) studied the relationship between active investor attention measured by the Google search volume index and the dynamics of prices for different currencies. Investor attention was found to be correlated with trading activities of major participants in the currency market. Their results suggest that investor attention is a priced risk source in the foreign exchange markets.

Han *et al*. (2018) empirically investigated whether investor attention is important for the exchange rate movements of nine countries using Google search volume as a proxy for attention. The results support the claim that investor attention contains information that influences exchange rate movements. Wu *et al*. (2019) examined financial contagion between currency markets through the new channel of investor attention measured by SVI. The findings support the attention reallocation channel as an important mechanism of contagion between currency markets and show that attention functions as a predictive variable.

Regarding cryptocurrencies, Dastgir *et al*. (2019) examined the causal relationship between Bitcoin attention, as measured by Google Trends queries, and Bitcoin returns. They observed a bidirectional causal relationship between Bitcoin attention and returns. Eom *et al*. (2019) empirically investigated the statistical and predictable characteristics of Bitcoin returns and volatility. The results suggest that Bitcoin seems to be an investment asset with high volatility and investor sentiment dependence, rather than a monetary asset. Ibikunle *et al*. (2020) investigated how increased attention affects the price discovery process of Bitcoin. The results show that the noise element of the Bitcoin price is driven by attention levels, implying that high attention levels are linked to increased individual investor trading activity in the Bitcoin market, while institutional investor trading activity is driven by arbitrage rather than attention.

Real Estate Investment Trusts (REITs) were studied by Yung and Nafar (2017), who investigated the effect of retail investor attention on expected REIT returns. Their results show that REITs generating high retail investor attention, measured by SVI, achieve higher returns compared to REITs that do not generate retail investor attention.

In the precious metals' scenario, Baur and Dimpfl (2016) studied investor attention in gold price movements by analyzing the relationship between them. The authors identified a positive relationship between gold price volatility and search term queries. Additionally, they found a strong asymmetrical effect of negative changes in the gold price on search queries, indicating that investors are more likely to search for bad news than good news about gold on Google. The authors also researched other precious metals and identified distinct behavior for gold.

Han *et al*. (2017) investigated relationships between investor attention and future gold returns. The results demonstrate that the past return of gold typically has a significant impact on investor attention. Furthermore, the authors also discovered that investor attention is closely associated with the futures market, indicating that investor attention incorporates significant information about expected future prices, providing an alternative explanation of the attention-return relationship and the presence of extreme economic conditions.

For commodities, Yao *et al*. (2017) aimed to empirically explore the impact of investor attention on crude oil prices using Google search volume index as a proxy for investor attention. The results indicate that investor attention has a significant negative impact on West Texas Intermediate (WTI) crude oil prices. The authors also identified that investor attention shocks contribute to long-term fluctuations in WTI crude oil prices and, when the business cycle remains in expansion, it positively influences investor attention and WTI oil prices.

In the Brazilian context, Ramos *et al*. (2017) analyze the predictive ability of Google searches on the Brazilian financial market. The results demonstrate a predictive effect between search levels and financial variables, primarily in the equity market. Furthermore, the authors found strong evidence of a causal relationship between the financial market and Google search levels.

Yoshinaga and Rocco (2020) investigate the role of investor attention in predicting future Brazilian stock market returns using Google Trends data. They tested whether lagged variations in GSV are followed by changes in excess returns using weekly Google Brazil search data from 2014 to 2018. The results show that increases in GSV are accompanied by lower excess returns, supporting the hypothesis that higher individual investor attention causes stock prices to deviate from their fundamental value. Still in the Brazilian context, Pereira et al. (2020) analyzed how Google searches influence the return, volatility, and trading volume of stocks composing the Ibovespa index. The results indicate that investor information demand is partially satisfied by search engine queries, which is also observed in increased trading volume after a historical search volume shock. For the authors, while higher returns attract investor attention, higher volatility tends to reduce investor attention.

Despite the extensive use of the Google Search Volume (SVI) as a proxy for investor attention, studies investigating whether Google search volume is actually associated with subsequent retail investor transactions are still scarce. In recent studies, Kostopoulos *et al*. (2020) relate SVI to trades of a sample of individual investors from Germany, using the FEARS index, and find a positive correlation between high FEARS index and the choice to trade risk assets. The authors contribute to the investor sentiment theory by noting that less sophisticated investors are more prone to sentiment.

Desagre and D'Hondt (2020) used trading accounts of a sample of retail investors collected from a Belgian online brokerage. The authors found that the relationship between SVI and retail trading activity is positive, but it is not stronger for purchases than for sales. They also identified evidence of a bidirectional causality between attention and individual investor trades.

It should be noted that despite the extensive literature documenting a positive relationship between the Google Search Volume Index (SVI) and returns or market volumes across various asset types, the studies have some limitations and provide gaps to be explored. First, most of the cited studies use market data for their respective tests, meaning that return movements can be influenced by institutional investors (Da *et al.*, 2011; Bank *et al*., 2011, Joseph *et al*., 2011; Vozlyublennaia, 2014; Swamy *et al*., 2019; Meshcheryakov and Winters, 2020). Second, selected keywords to capture investor attention may not necessarily be linked to interest in company-related information from Google Trends searches, which can also bias results (Da *et al*., 2011; Joseph *et al*., 2011; Challet and Ayed, 2013). Third, data on individual investor trades are limited to a specific segment within a market, which can lead to conclusions that cannot be extrapolated to the market as a whole (Desagre and D'hondt, 2020; Kostopoulos *et al*., 2020).

In this context, seeking to contribute to the field of noise trader approach studies, this work innovates by analyzing whether the pattern of online searches actually signifies retail investor intent to trade certain assets in the Brazilian scenario.

3 ARTICLE STRUCTURE

This section presents the three main studies proposed to address the general objectives of the thesis.

Article 1: Starting from one of the key factors affecting research on individual investors, namely the lack of access to noise trader behavior, the first study aims to test an alternative metric that has been used as a proxy for investor attention. To do so, we conducted an empirical study using the daily trading volume exclusively by retail investors in the Brazilian stock market from 2018 to 2020. We found that online searches show a strong association with future trades by retail investors and that this relationship is observable under different market conditions. Furthermore, we also documented that alternative approaches commonly used in literature to capture investor attention through online searches, such as tickers vs. company names, market index attention vs. aggregated attention to individual stocks, or log differences vs. abnormal log differences in search volumes, all provide consistent measures for capturing future trades by investors. Overall, our findings strongly support the view that online searches are a coherent proxy for the presence of retail investors in the stock market.

Article 2: We document that the attention-return relationship in the gold spot market is sensitive to market states, as we identify a positive (negative) association between returns and contemporaneous attention during bull (bear) markets. This sensitivity is influenced by the magnitude of returns, as the significance of the relationship, based on a quantile model, increases nearly monotonically as the return distribution deviates from the median. Additionally, we find that investor attention is influenced by recent trends in the spot gold price in both market states. Overall, these results suggest that the past performance of gold returns attracts the attention of individual investors who, in turn, trade to profit from the trend, reinforcing price movements, which is consistent with the momentum structure of the time series.

Article 3: Considering the Noise trader theory that posits retail investors trade based on noise and cannot impact market performance, the third study analyzes information on the characteristics that attract the attention of individual investors who choose to trade assets. Using Google search volume as a proxy for individual investor attention and adopting the tickers of companies composing the IBRX100 Index from 2014 to 2021 as search terms, variables to measure noise and value variables, which are supposedly not considered by noise traders, are analyzed. The results indicate that fundamental variables have significant effects on investor attention. Sensitivity analysis conducted through univariate regression for each of the relevant variables supports this finding by revealing that fundamental dynamics remain more important for their investment decisions than dynamics of noise variables.

PARTE 2

"Part II" adds the main original articles written for this thesis. The presentation of the articles was done in "Part 1," section 3.

ARTICLE 1

- Paper published in the International Review of Financial Analysis Volume 86, March 2023, 102552

Do online searches actually measure future retail investor trades?

ABSTRACT

Using the total daily amount traded exclusively by retail investors in the Brazilian stock market from 2018 to 2020, we find that online searches exhibit a strong association with the future trades of retail investors and that this relationship is observable under different market conditions. Moreover, we also document that the alternative approaches commonly used in the literature to capture investor attention with online searches, such as tickers *versus* company names, market index attention *versus* aggregated individual stock attention, or log-differences *versus* abnormal log-differences of search volumes, are all consistent measures to capture future investor trades. Overall, our findings strongly support the view that online searches are a coherent proxy for the presence of retail investors in the stock market.

Keywords: investor attention, online searches, retail investors, google trends

JEL Classification: G10, G12, G14

1. Introduction

The role of retail investors in financial markets has been attracting growing interest in asset pricing literature. Recent studies find that these investors can move stock prices (Barber *et al*., 2009; Han and Kumar, 2013), distort the mean-variance relationship (Yu and Yuan, 2011), and play a relevant role in well-documented puzzles, such as low beta (Bali *et al*., 2017) and low-risk anomalies (Liu *et al*., 2018). A natural barrier to investigating the influence of retail investors in markets relates to data availability since data on retail trades are not publicly available. Even though some studies access this private data from brokerage firms (e.g., Kumar, 2009), this limits the generalization of the results since this data naturally covers only a small part of the retail investors' universe (Kelley and Tetlock, 2013; Wang and Zhang, 2015). In this context, a group of studies (Wen *et al*., 2019; Da *et al*., 2020; Cheng *et al*., 2021) employ online search volume as a proxy for the role of individual investors in the market, relying on the belief that investors' searches and trades are related (Da *et al*., 2011). This assumption is based on the intuition that less sophisticated investors will search for information using search engines, such as Google, before trading the given security. Although this assumption is widely assumed in the corresponding literature, studies that directly test this assumption are extremely scarce. This is the primary contribution that this study aims to bring, since, to the best of our knowledge, this is the first to test this relation using the total amount traded exclusively by retail investors in a given market.

In this study, in collaboration with the Brazilian Stock Exchange (B3), we obtain access to the total daily amount traded by local retail investors from 2018 to 2020. We compare this measure with the Aggregated Retail Attention (ARA) index following Da *et al*. (2020). More precisely, the ARA index is composed of the Google Search Volumes (GSV) of the tickers of stocks belonging to the Brazilian market index during our sample period, weighted by their market capitalization. Hence, the evidence reported throughout the study is inferences for the aggregated level of the retail market on both sides: trades and attention, circumventing the limited generalization issue. There are at least two advantages to be gained from focusing on Brazilian market. First, the tickers of stocks in the Brazilian market are a conjunction of letters and numbers (e.g., PETR3 for "Petrobrás"), which completely circumvent the bias of using dubious terminologies (e.g., "F" for "Ford Motor Company" on the NYSE) that could distort the usage of on-line searches as a measure of attention on the corresponding company (Da *et al*., 2011). Second, during our sample period, the number of retail accounts in the Brazilian stock market exhibited an increase of more than 400%. This peculiar behavior is very important when running a robustness test on our central hypothesis, since it represents a natural experiment to investigate whether the attention-trade relation holds during periods when individual investors are less active in the market.

Our main results are as follows. The VAR analysis among ARA and trades indicates the influence of past online searches on future trades, whereas the relationship in the opposite direction is marginally observable (i.e., 10% significance level). We also run OLS models using retail trades as the dependent variable and find a strong association between the constructs. More interestingly, the coefficients for lagged online searches are always more significant than the contemporaneous in every OLS model examined, in line with the notion that the retail investor would search online before deciding to trade the given stock. This relation is also economically important since we document that an increase of one standard deviation in online searches provokes a future increase of 9.1% in the total amount traded by retail investors (t-stat = 7.26), which is almost eight times larger than the mean increase in trades of our sample.

Since our findings suggest that online searches and retail trades are related for the entire dataset, a natural question is whether this relationship is pervasive or restricted to specific market conditions when retail investors are more active in the market, given that the literature reports that individual investors exhibit a preference for skewness (Kumar and Lee, 2006) or to trade on recent trends (Barber and Odean, 2008). To investigate this, we examine the attention-trade relationship under different market states, namely, a) bear/bull market, b) high/low volatility periods, and c) accelerated increase in the number of retail accounts in the local market. For every market condition examined, we document a significant relationship between online searches and retail trades, which is always more acute for lagged searches, indicating that the attention-trade relationship holds conditionally since it is not driven by different market circumstances. This is the second contribution of this study since previous studies do not test for the unconditional hypothesis of the relation.

The literature employing online searches as a proxy to retail investor attention has a peculiar characteristic of adopting slightly different approaches to calculate the attention variable. First, while a group of studies employs the name of the stocks as the terminology of online searches (Aouadi *et al*., 2013; Bijl *et al*., 2016), claiming that less sophisticated investors would plausibly search for the company's under interest by its name, another large group of studies uses stock tickers (Joseph *et al*., 2011; Wen *et al*., 2019; Da *et al*., 2020; Cheng *et al*., 2021), advocating that this is a less ambiguous measure of investor attention since using a firm's name would incorporate searches that are not related to investments, such as looking for companies' products or job positions. Second, when focusing on market-level analysis, it is common to use the market index name as the search terminology (Vozlyublennaia, 2014; Hamid and Heiden, 2015; Baur and Dimpfl, 2016) since using the scaling procedure of Google Trends data, it is not possible to compare searches in large datasets, which is the case when focusing on the entire stock market. A reasonable question is whether focusing only on searches on the market index effectively captures the interest of retail investors towards the stock market since this procedure filters numerous searches on individual stocks. Finally, Google Trends data commonly exhibit spikes, time trends, or low-frequency issues (Da *et al*., 2011). To address this, many studies (Smith, 2012; Bijl *et al*., 2016; Da *et al*., 2020) calculate an abnormal attention measure that, in essence, is given by the log difference of the contemporaneous GSV and its moving window median. Meanwhile, another branch of studies calculates the attention measure as the log-difference of GSV on two consecutive periods (Vozlyublennaia, 2014; Baur and Dimpfl, 2016; Han *et al*., 2017; Wu *et al*., 2019). In this context, it is legitimate to conjecture whether this plethora of approaches can effectively capture retail investors' intention to trade in the market. This is the third contribution of the present study.

In this regard, we employ the methods mentioned above as alternative approaches to our ARA Index measure and investigate their relationship to retail trades. The results demonstrate a significant relationship between trades and contemporaneous and lagged online searches for each of these approaches. Moreover, the VAR analysis reports that online search influences future trade in all specifications examined, while the association in the opposite direction is virtually unobservable. The fact that the main findings remain almost untouched when employing different approaches provide a strong support to the assumption that online searches and future retail trades are actually connected. This pattern serves as a robustness check towards our central results and particularly interest the editors and reviewers in this field since it demonstrates that the distinct nuances in the methods to operationalize investor attention using online searches capture what they aim to: future trades on the stocks under scrutiny.

The study of the relation between online searches and future investor trades is not new to the literature. For example, using a dataset of 100,000 individual German brokerage accounts, Kostopoulos *et al*. (2020) find a negative association between individual trading activity and the FEARS index, the latter constructed using online searches on economic concern terminologies, as proposed by Da *et al*. (2015). Given that the FEARS index is a measure of individual sentiment, the aim of their study is to investigate how individual sentiment influences the trading behavior of individual investors, whereas our research question is of a methodological nature, focusing on whether online searches actually measure future investor trades, as assumed by a series of studies on investor attention. In a paper more closely related to ours, Desagre and D'Hondt (2021) use a sample of 42,731 Belgian retail investors, finding a positive association between online searches and individual trading activities. In light of their findings, we make four main contributions. First, the literature reports that sample size is an important issue when it comes to investigating individual trading behavior (Boehmer *et al*., 2021), because the results can be biased by intrinsic characteristics of the dataset that is used. Consequently, the findings from the literature so far (i.e., Desagre and D'Hondt, 2011) do not allow us to assume that on-line searches can be used as a proxy for retail trades. Since our findings are based on retail trades for the entire Brazilian market, we advocate that our study provides the first consistent evidence that retail trades can be proxied by on-line searches. Second, we test the attention-trades relation for different proxies for investor attention that use GSV (namely, using stocks' ticker, stocks' names, and names of market indices). This is an important contribution to the corresponding literature, since all of these specifications can be found in several previous studies. We also make a third contribution by investigating whether the strength of the relation can be driven by market states, which could limit the use of on-line searches as a proxy for future individual trades. Finally, our findings have important implications for the growing literature on the role of individual investors in financial markets. An important limitation in this field is related to data availability, given the difficulty involved in accessing individual trades. According to our findings, this difficulty can be easily circumvented by employing on-line searches as a proxy for individual trades, given the strong association between both constructs documented here, facilitating the dissemination of studies on retail investors.

The remainder of the paper is structured as follows. Section 2 describes the data and the methodology used to generate the main variables. Section 3 presents and discusses the results. Section 4 presents robustness tests, and Section 5 concludes the paper.

2. Data and Methodology

In cooperation with the Brazilian Stock Exchange (B3), we obtained a report containing the total daily amount traded exclusively by local retail investors between 2018 and 2020[1]. Given that our measure for investor attention is available on a weekly basis, as described below, we transform the daily trades into weekly trades by aggregating the daily trades (buys and sells) of retail investors during the corresponding weeks, as shown in Eq. (1), where the subscript d refers to the sells/buys on the given day. The focus on both sell and buy trades instead of focusing on the buy-side (as, for example, in the buy-sell imbalance measure) is justified as an effort to avoid the limited attention bias documented by Baber and Odean (2008)[2]. The log-difference of two consecutive weekly trades is our measure of the trades of retail investors (T), as shown in Eq. (2).

$$Agg\ Trades_t = \sum_{d=1}^{5} (Sells_d + Buys_d) \tag{1}$$

[1]The Brazilian Stock Exchange classifies foreign retail individual investors in the separate category of "Foreign Investors", which also includes foreign institutional investors. Since we have restricted the Google Searches region solely to Brazil, we advocate that our data for the retail trade amount is consistent with the data for investor attention.

[2]Desagre and D'Hont (2021) document that the relationship between GSV and trading activity is not stronger for buys than for sales, contradicting the limited attention hypothesis of Baber and Odean (2008). This evidence provides further support to include the sales in our trading measure.

$$T_t = ln\ (Agg\ Trades_t)\ - ln\ (Agg\ Trades_{t-1}) \quad (2)$$

Since our data on trades are at an aggregated level, we also construct an index of aggregated retail attention (ARA). In this regard, we downloaded GSV data using the tickers of the sampled firms as the search terminology. The GSV data is restricted to searches in Brazil, to be consistent with our retail trade amounts, which cover only Brazilian retail investors. Tickers have a particular format that combines letters and numbers on the Brazilian Stock Exchange and avoids confusion with ambiguous terminologies. For example, the ticker for Petrobrás' common stocks, one of the largest oil companies in the world, is PETR3, whereas for Vale do Rio Doce, the second-largest mining company in the world, it is VALE3. Meanwhile, it is plausible that many individual investors would employ a firm's name when searching for stock information. We address this issue through a robustness check and find that the main conclusions remain almost untouched by this alternative approach.

Given that Google Trends data for a given period is standardized with the largest number of searches peaking at 100, it forbids comparing GSVs of different terminologies for large datasets. To circumvent this, we follow Da *et al*. (2020) and aggregate the GSV for individual stocks (i) using the corresponding market capitalization[3] to weigh stock-specific GSVs on the ARA index[4], as indicated in Eq. (3).

$$ARA_t = \sum_{i=1}^{n} W_{i,t} GSV_{i,t} \quad (3)$$

where $W_{i,t}$ is the market capitalization of stock i on week t divided by the total market capitalization of the market index during the corresponding week, and $GSV_{i,t}$ is the GSV for firm i on week t. Finally, n refers to the number of companies listed in the market index on the given week t.

The ARA dataset is weekly because, for the entire period, Google Trends makes only weekly data available. Our measure of attention (S) is the log difference between ARA over two consecutive weeks.[5] We restrict our sample to the stocks that were part of the market index in any week of our sampled period, with the aim of focusing on more liquid stocks, which would lead to more representative results. Furthermore, given that individual investors are more likely to trade illiquid

[3] The data for the market capitalization are weekly and were obtained from Economatica®.

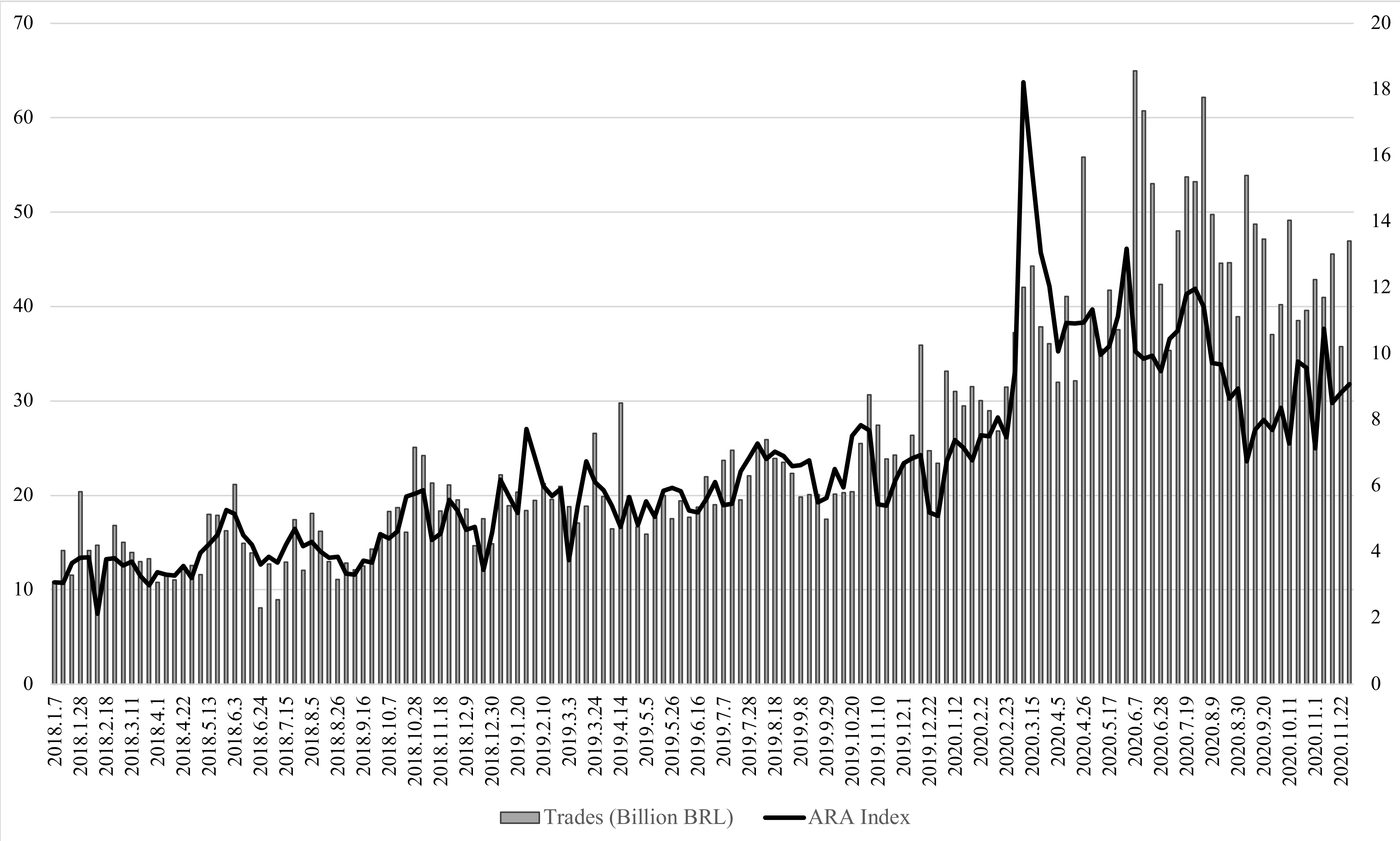
[4] We employ a simple average to generate an equally weighted ARA index in an alternative specification. The results are consistent and can be found in the robustness section.

[5] We also employ the approach suggested by Da *et al*. (2011) that uses the log difference between ARA on a given week and its past eight weeks median as a measure of attention. The results are virtually the same and are brought in the robustness section.

stocks (Kumar, 2009; Bali *et al*., 2017), this choice would avoid our results being biased by the behavior of less liquid securities that could drive the behavior of ARA. This filter results in a sample of 78 stocks.

Figure 1 shows the weekly trades ($Agg\ Trades_w$) in billion BRL (columns, right axis), together with the ARA index (solid line, left axis). The ARA index spikes in March 2020 due to the fear of the COVID-19 outbreak. Furthermore, there is a clear escalation in the amount traded by retail investors during the sample period. The total amount traded by these investors increased from 225 billion BRL (61 billion USD) in 2018 to 576 billion BRL in 2020 (112 billion USD), which can be partially explained by the concomitant decrease in local interest rates from 6.90% in January 2018 to 1.90% in December 2020. Most importantly, a clear association between both variables suggests that investor attention is related to trade.

Figure 1: Retail trades and attention.



The figure displays the total weekly amount traded (buy and sell) by Brazilian retail investors (gray columns, right axis) together with the ARA index (solid line, left axis), generated using the weekly GSV of the tickers of stocks belonging to the local market index. Trade is in billion BRL. The dataset is from 01.01.2018 to 11.30.2020.

To better illustrate some features of the main variables, Table 1 presents the weekly corresponding summary statistics. In a typical week, Brazilian retail investors trade 7.4 billion BRL (1.74 billion USD), with a growing trend captured by the positive asymmetry. Given that our ARA index is constructed with the weighted average of individual stocks, it peaks at 63.76 (and not at 100, as the raw Google Trends data do), which means that in a typical week, the online searches on Brazilian stocks represent almost one-third of the sample peak. The stationary distributions (i.e., S and T) have positive means, moderate skewness, and high variance. They also exhibit extreme observations with similar magnitudes albeit, as shown in Figure 2, these extreme swings are not coincident. In the following section, we examine whether similar patterns are associated.

Table 1: Descriptive statistics

	ARA Index	Agg. Trades	S	T
Mean	22.96	7.40	0.0035	0.0097
Median	20.53	6.01	0.0007	-0.0167
Std Dev	9.52	3.66	0.17	0.21
Asymmetry	1.15	1.01	0.26	0.40
Kurtosis	1.78	0.19	2.81	0.33
Min.	7.46	2.31	-0.59	-0.54
Max.	63.76	18.57	0.65	0.59

The table shows the summary statistics of the Aggregated Retail Attention (ARA) index for the total amount traded by retail investors in the Brazilian stock market (Agg. Trades) and for the corresponding stationary series (S and T, respectively), measured by the log difference of two consecutive periods. Aggregated trades is in billion BRL. The ARA index is the average Google search volumes of stock tickers belonging to the local market index, weighted by their market capitalization. The data is weekly, from 01.01.2009 to 11.30.2020.

3. Results

3.1 VAR Analysis

We begin the empirical analysis by investigating the Granger causality among trades and attention by employing a VAR approach. The intuition for testing a bidirectional relation is as follows. Although many studies assume that online searches are a consistent proxy for investors' intentions to trade on a given security (Joseph *et al.*, 2011; Vozlyublennaia, 2014), it is plausible to believe that the relation departs from the opposite direction. In the latter case, attention to security would be caused by variation in trades, in the sense that an abnormal volume transacted could grab investor attention, what is supported by the evidence that online searches and stocks' turnover are correlated (Da *et al.*, 2011). This investigation is addressed by the following models, where the number of lags is determined using the AIC:

$$T_{t,i} = \delta_0 + \delta_1 T_{t-1,i} + \cdots + \delta_n T_{t-n,i} + \gamma_1 S_{t-1,i} + \cdots + \gamma_n S_{t-n,i} + \varepsilon_t \quad (4)$$

$$S_{t,i} = \delta_0 + \delta_1 T_{t-1,i} + \cdots + \delta_n T_{t-n,i} + \gamma_1 S_{t-1,i} + \cdots + \gamma_n S_{t-n,i} + \varepsilon_t \quad (5)$$

The estimates for the parameters are presented in Table 2 together with the Granger causality test. The results show that changes in attention are positively related to future trades, mainly one week ahead. This association is also economically important, since an increase of one standard deviation in attention is associated with a 7.7%% increase in the total amount traded by retail investors in the following week (0.453 x 0.17 = 7.7%), which is almost eight times higher than the mean weekly growth of trades in the sampled period. The relationship in the opposite direction (i.e., trades Granger-causing attention) is marginally significant (Z = 0.136 / 0.073 = 1.87) and restricted to the first lag. These findings indicate that, even though past trades changes grab investors' attention at an aggregated level, the influence of past attention on future trades is more significant and persistent, providing support for the assumption that online searches are a consistent proxy for future trades from these investors. This pattern is supported by the Granger causality test, which rejects the non-causality between attention and trades at the 1% level (chi-squared = 26.303), whereas the Granger causality in the opposite direction is not significant at the 5% level.

These results are partially consistent with the findings of Desagre and D'Hondt (2021), who find bidirectional causation between attention and trades in the Belgian market. Despite the difference in periodicity (the authors employ monthly data in their study), our evidence of marginal causation in the trade-attention direction can be plausibly explained by the time-trend properties of both attention and trades datasets. Since these variables exhibit a clear time trend, employing stationary data in this investigation is recommended. For example, in our dataset, the ARA index (aggregated trades) exhibits significant autocorrelation for up to 12 lags (13 lags), whereas for the stationary data (i.e., S and T), the autocorrelation ceases after one lag. Given that the authors do not employ stationary data in their causation analysis, their findings could potentially be biased by time-trend endogeneity. In fact, in an unreported test, we run Eqs. (4) and (5) using the ARA index and aggregated trades and find a bidirectional Granger causation (albeit presumably endogenous).

Table 2: VAR analysis between online searches and retail trades

	S_{t-1}	S_{t-2}	T_{t-1}	T_{t-2}	Intercept	adj. R^2
T_t	0.453	0.175	-0.551	-0.240	0.011	30.4
	[0.089]***	[0.093]*	[0.080]***	[0.077]***	[0.770]	
S_t	-0.292	-0.272	0.136	-0.026	0.009	13.0
	[0.080]***	[0.084]***	[0.073]*	[0.070]	[0.710]	
T does not Granger cause S	0.070	[5.305]				
S does not Granger cause T	0.000	[26.303]				

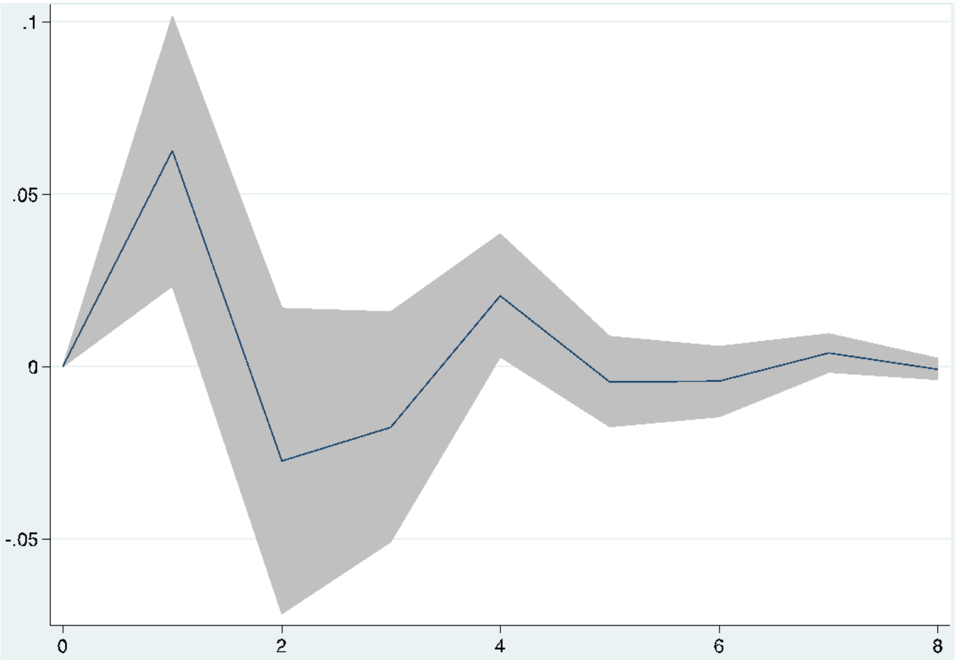
This table shows the estimates for the VAR models relating to online searches (S) and retail investors' trades (T). The number of lags is based on the AIC. See Section 2 for a detailed description of the construction of these variables. Standard errors are reported below the coefficients. The data were collected weekly from 01.01.2018 to 11.30.2020. In the lower section, the table reports the p-values of the Granger causality tests, in addition to the respective chi-squared between brackets. N = 149
***Significant at 1%, **5%, and *10% levels

To provide information on the timing and direction of the relationship documented above, Figure 2 shows an impulse-response chart. The chart on the left (right) shows the response of trades (attention) to one standard deviation of attention (trades). The orthogonalized impulse response functions are displayed for eight periods after a one standard deviation shock. The significance of the shock is given by confidence intervals that represent the plus/minus two standard deviations. Hence, when the confidence bands do not cross line zero, the impulse response is statistically different from zero at the 5% significance level. The charts indicate that the response of trades to changes in attention is stronger and more persistent than the relationship in the opposite direction, being observable until four weeks after the shock in attention.

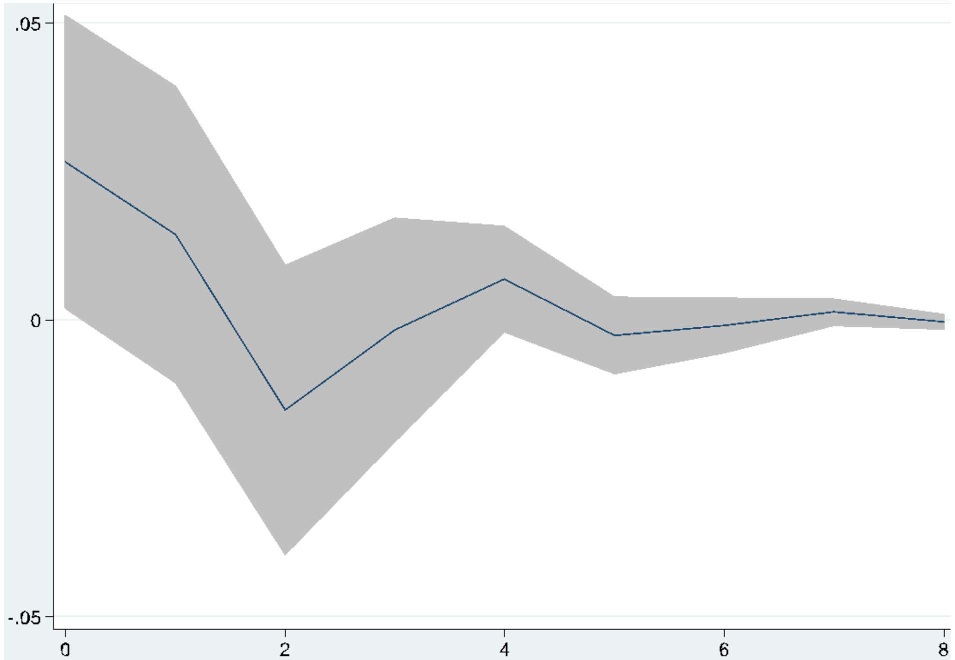
Figure 2: Impulse response function between retail trades and investor attention.

This figure plots the impulse response to one standard deviation of the VAR model between online searches (S) and retail investors' trades (T). The left (right) chart shows the responses of trades (searches) to search shocks (trades). See Section 2 for a detailed description of the construction of these variables. The sample period is from 01.01.2018 to 11.30.2020.

Response of trades to online searches



Response of online searches to trades



Overall, the findings indicate that increases in online searches are associated with future increases in retail investor trades. The following subsection provides an in-depth analysis of this relationship.

3.2 Relationship between attention and trades

This subsection investigates the relationship between attention and trade by employing two OLS models specified as follows:

$$T_t = a_0 + \sum_{n=0}^{2} b_n \cdot S_{t-n} + \varepsilon_t \tag{6}$$

$$T_t = a_0 + \sum_{n=0}^{2} b_n \cdot S_{t-n} + \sum_{n=1}^{2} c_n \cdot T_{t-n} + \sum_{n=0}^{2} d_n \cdot R_{t-n} + \varepsilon_t \tag{7}$$

The first equation analyzes the influence of contemporaneous and lagged attention on the trades of retail investors. In Eq. (7), we control for the autocorrelation of the dependent variable and for market return, measured by the log-difference of the market index in two consecutive weeks, since previous studies document that investor attention and market returns are related (e.g., Vozlyublennaia, 2014). The focus is on coefficients b's of each model, as well as between both models, to provide sensitivity analysis. The estimates are presented in Table 3 together with the *t*-statistics in parentheses. The results for the univariate model indicate a positive relationship between attention and trade in the following week, supporting the notion that retail investors search for online information before trading the corresponding stocks. In the extended approach (Model 7), the results are even more striking since we document a positive association between trades and contemporaneous as well as lagged attention. The fact that, for both models, the one-lag attention exhibits the highest significance indicates that retail investors search for online information regarding the stocks under interest before deciding to trade them. As previously demonstrated, the influence of attention on future trades is economically representative since, based on the coefficient of Model 7, an increase of one standard deviation in online searches provokes a rise in trades of 9.1% (0.54 x 0.17 = 9.1%) in the following week. The statistical significance of this relationship is also remarkable (t-stat = 7.26), given the moderate number of observations.

Table 3: Relationship between online searches and retail trades

	S_t	S_{t-1}	S_{t-2}	T_{t-1}	T_{t-2}	R_t	R_{t-1}	R_{t-2}	Intercept	adj. R^2
(6)	0.15	0.38	-0.03						0.01	7.9
	(1.60)	(4.91)***	(-0.30)						(0.59)	
(7)	0.23	0.54	0.24	-0.62	-0.26	0.26	0.41	1.14	0.01	36.8
	(2.74)***	(7.26)***	(2.45)**	(-6.66)***	(-3.07)***	(0.67)	(1.18)	(2.34)**	(0.40)	

This table shows the estimates for the regressions between online searches (S) and retail investors' trades (T) using the following OLS models:

$$T_t = a_0 + \sum_{n=0}^{2} b_n \cdot S_{t-n} + \varepsilon_t \tag{6}$$

$$T_t = a_0 + \sum_{n=0}^{2} b_n \cdot S_{t-n} + \sum_{n=1}^{2} c_n \cdot T_{t-n} + \sum_{n=0}^{2} d_n \cdot R_{t-n} + \varepsilon_t \tag{7}$$

See Section 2 for a detailed description of the construction of these variables. R is the weekly market return. The sample period is from 01.01.2018 to 11.30.2020. The numbers in parentheses are t-statistics from the Newey-West standard error estimator. The R^2 values are expressed as percentages. N=149.
***Significant at 1%, **5%, and *10% levels.

As mentioned in Section 2, we use the market capitalization of each stock to construct an ARA index. Consequently, the evolution of this index over time is influenced by the stock price dynamics. Since the price change is a direct function of trades (i.e., buy-sell balances), the results shown so far could be driven by an endogenous role from market value. To address this, we regress attention S on the contemporaneous and lagged market index changes and retain the residuals[6]. The standardized residuals (U_t) represent the portion of attention not explained by the evolution of market capitalization. We employ this orthogonal measure as our alternative proxy for attention in previous models, as shown in Eqs. (8) and (9):

$$T_t = a_0 + \sum_{n=0}^{2} b_n \cdot U_{t-n} + \varepsilon_t \tag{8}$$

$$T_t = a_0 + \sum_{n=0}^{2} b_n \cdot U_{t-n} + \sum_{n=1}^{2} c_n \cdot T_{t-n} + \sum_{n=0}^{2} d_n \cdot R_{t-n} + \varepsilon_t \tag{9}$$

The results are presented in Table 4. The findings remain virtually unchanged: there is a significant association between orthogonal attention and one-week-ahead trades in the univariate model, and a significant relationship between trades and contemporaneous as well as one and two-lag orthogonal attentions when controlling for autocorrelation and market return (Model 9). The significance of these relationships is almost untouched when this alternative approach is employed. Overall, the results indicate that the possible endogeneity arising from the market capitalization dynamics does not play a significant role in the relationship under scrutiny.

[6] More precisely, we run the following model: $S_t = a_0 + \sum_0^2 a_n \cdot R_{t-n} + u_t$, where R is the return of the market index defined as the log-difference of the index in two consecutive weeks. The only significant coefficient is the contemporaneous return (t-stat = 2.12).

Table 4: Relationship between orthogonalized attention and retail trades

	U_t	U_{t-1}	U_{t-2}	T_{t-1}	T_{t-2}	R_t	R_{t-1}	R_{t-2}	Intercept	adj. R^2
(8)	0.13	0.36	-0.09						0.01	8.2
	(1.48)	(4.80)***	(-1.02)						(0.64)	
(9)	0.22	0.51	0.21	-0.59	-0.27	0.28	0.68	1.06	0.01	38.0
	(2.52)**	(7.17)***	(2.14)**	(-6.26)***	(3.28)***	(0.73)	(1.87)*	(2.15)**	(0.44)	

This table shows the estimates for the regressions between orthogonalized attention (U) and retail investors' trades (T) using the following OLS models:

$$T_t = a_0 + \sum_{n=0}^{2} b_n \cdot U_{t-n} + \varepsilon_t \tag{8}$$

$$T_t = a_0 + \sum_{n=0}^{2} b_n \cdot U_{t-n} + \sum_{n=1}^{2} c_n \cdot T_{t-n} + \sum_{n=0}^{2} d_n \cdot R_{t-n} + \varepsilon_t \tag{9}$$

Orthogonalized attention (U) is obtained by regressing the online search variable (S) on the contemporaneous and lagged market index returns, aiming to capture the portion of attention that is not explained by the evolution of market capitalization. R is the weekly market return. The sample period is from 01.01.2018 to 11.30.2020. The numbers in parentheses are *t*-statistics from the Newey-West standard error estimator. The R^2 values are expressed as percentages. N=149.
***Significant at 1%, **5%, and *10% levels

The summary statistics of the dependent variable T reveals that its distribution exhibits a large dispersion relative to its mean (CV = 21.2) and that its median is relatively distant from the mean. This pattern suggests that the results documented so far could be biased by outliers, and that the mean value could not be a consistent proxy for the typical behavior of T on a given week. In both cases, the OLS model produces inefficient estimates of the trades–attention relationship. To investigate this hypothesis, we run a quantile regression using Model (7). To avoid heteroskedastic and autocorrelation issues, the standard errors of the coefficients were obtained using the block-bootstrap procedure. The results are summarized in Table 5, where the first row informs the quantile under analysis.

Table 5: Quantile regression of attention on trades

	0.05	0.10	0.25	0.50	0.75	0.90	0.95
S_t	0.369	0.415	0.243	0.228	0.221	0.240	0.081
	(3.29)***	(2.64)***	(1.88)*	(1.57)	(2.33)**	(1.56)	(0.74)
S_{t-1}	0.249	0.546	0.414	0.515	0.663	0.569	0.534
	(1.12)	(5.14)***	(4.57)***	(3.93)***	(4.38)***	(2.79)***	(4.76)***
S_{t-2}	0.106	0.184	0.115	0.276	0.356	0.343	0.239
	(0.39)	(1.45)	(0.87)	(1.96)**	(2.43)**	(1.13)	(2.48)**
T_{t-1}	-0.295	-0.432	-0.576	-0.686	-0.731	-0.677	-0.740
	(-1.88)*	(-4.30)***	(-5.59)***	(-4.64)***	(-4.62)***	(-4.61)***	(-4.83)***
T_{t-1}	0.020	-0.120	-0.275	-0.303	-0.382	-0.335	-0.442
	(0.15)	(-1.10)	(-2.48)**	(-2.68)***	(-3.01)***	(-2.24)***	(-1.93)*
R_t	-0.687	-0.035	-0.468	0.391	0.935	0.716	0.410
	(-0.75)	(-0.07)	(-1.04)	(0.63)	(1.87)*	(0.29)	(1.01)
R_{t-1}	0.035	-0.248	0.080	0.856	0.239	0.831	0.043
	(0.05)	(-0.60)	(0.18)	(0.93)	(0.59)	(0.55)	(0.23)
R_{t-2}	0.605	0.549	0.899	0.768	0.738	1.424	1.865
	(0.75)	(0.92)	(1.24)	(1.07)	(1.81)*	(0.83)	(7.66)***
Intercept	-0.245	-0.192	-0.096	0.006	0.116	0.200	0.278
	(-6.45)***	(-7.34)***	(-5.46)***	(0.33)	(5.97)***	(7.65)***	(13.77)***
R^2	21.8	31.7	36.4	38.7	38.0	36.4	36.4

The table shows the estimates for the quantile regressions between online searches (S) and retail investors' trades (T), using the following expression:

$$T_t = a_0 + \sum_{n=0}^{2} b_n \cdot S_{t-n} + \sum_{n=1}^{2} c_n \cdot T_{t-n} + \sum_{n=0}^{2} d_n \cdot R_{t-n} + \varepsilon_t \qquad (7)$$

See Section 2 for a detailed description of the construction of these variables. R is the weekly market return. The sample period is from 01.01.2018 to 11.30.2020. The corresponding quantiles are shown in the first row. The numbers in parentheses are blocked bootstrap *t*-statistics. The R^2 values are expressed as percentages. N=149.
***Significant at 1%, **5%, and *10% levels

The results indicate that in ordinary circumstances (e.g., the 50% quantile), there is a positive and significant relationship between trades and lagged attention, and that this relation is more pronounced for one-lag attention. Since this behavior is documented in the OLS model (Table 3), we can screen out the hypothesis that our main findings are biased by the presence of outliers. The fact that the attention coefficients are always positive in every quantile examined provides additional evidence that online searches are related to investor trades. It is also interesting to note that in almost all quantiles the one-lag attention is more significant than the contemporaneous, consistent with the notion that retail investors search the stock under interest online before deciding to trade it, as previously mentioned.

The findings thus far provide empirical support for the assumption that online searches are a consistent proxy for retail investors' intention to trade on a given security, as assumed in several studies (e.g., Da *et al*., 2011; Joseph *et al*., 2011; Vozlyublennaia, 2014; Bijl *et al*., 2016). In this context, a natural question is whether this relationship holds conditionally or is sensitive to market states. The following subsection addresses this issue.

3.3 Attention-trades relation and market states

It is well established in the literature that investors' attention is sensitive to market conditions, such as market volatility (Dimpfl and Jank, 2016), bull/bear markets (Piccoli and Castro, 2021), market sentiment (Kostopoulos *et al*., 2020; Cheng *et al*., 2021), or recent stock trends (Karlsson *et al*., 2009). Consequently, a natural conjecture is whether our previous findings are sensitive to market states, which could suggest that our results are driven by endogenous factors. To address this, we introduce a dummy variable in Eq. (7), aiming to capture the attention–trade relation under different market conditions, as shown in Eq. (10)

$$T_t = a_0 + \sum_{n=0}^{2} b_n D_{t-n} S_{t-n} + \sum_{n=0}^{2} c_n (1 - D_{t-n}) S_{,t-n} + \sum_{n=1}^{2} d_n T_{t-n} + \sum_{n=0}^{2} e_n \cdot R_{t-n} + \varepsilon_t \qquad (10)$$

where the dummy D_{t-n} equals 1 for the condition under scrutiny and 0 otherwise. In this study, the market states analyzed are as follows:

a) Bull/bear markets: In their classic study, Karlsson *et al*. (2009) find that individuals tend to pay more attention to their portfolios when recent returns are positive, and to avoid examining their stocks during negative markets, a behavior that the authors referred to as the Ostrich effect. Consequently, it would be reasonable to believe that our results could be biased by the propensity of individuals to be more attentive to the stocks they own during positive markets, leading to an increase in online searches under these circumstances. Therefore, the attention-trades relation documented above could be alternatively explained by the Ostrich effect. Moreover, Piccoli and Castro (2021) report that the attention-return relation is influenced by up/down regimes. Consequently, it is reasonable to conjecture that market performance affects the association between attention and trades. In this case, the dummy variable equals 1 if the contemporaneous market index return is positive (a bullish market).[7].

b) Volatility: the literature suggests that individual investors exhibit a preference for skewness (Barberis and Huang, 2008; Barberis and Xiong, 2012). Regarding investor attention and retail transactions, Kostopoulos *et al*. (2020) find that greater attention to economic concerns is related to individual investor sells of risky stocks, whereas Cheng *et al*. (2021) document that attention-grabbing stocks display a higher future crash risk due to the tendency of individual investors to buy risky stocks. Taken together, both studies suggest that volatility can play an important role in the attention-trades relation. Therefore, it would be reasonable to conjecture that our results are driven by high volatility periods, indicating that the attention-trades relation is actually explained by the preference for skewness. To examine whether the association holds during low volatility periods, we generate the conditional variance of the market by employing a GARCH (1,1) model, using weekly index returns. In this case, the dummy equals 1 (0) if the conditional market volatility in a given week is higher (lower) than the median

c) Increase in retail investors: As shown in Figure 1, the amount traded by Brazilian retail investors exhibited a sharp increase during the sampled period. This behavior can be explained by the scaling in retail investor accounts on the local stock exchange, which rose from 619 thousand at the end of 2017 to 3,200 thousand in 2020. This increment was clustered in the last two years of our dataset (i.e., 2019 and 2020), when the number of individual accounts more than tripled, presumably due to the decrease in local risk-free rates, as previously explained. Given this characteristic, it could plausibly be conjectured that the aforementioned attention–trade relationship might be driven by this particular increase, since retail investors are more prone to herding (Barber *et al*., 2009). To investigate this hypothesis, we used the dummy variable in Eq. 10 equals 1 during 2019 and 2020 and 0 during 2018, when the increase in the number of individual accounts was moderate.

[7] Even though the way we define bullish/bearish markets is supported by other studies (Vozlyublennaia, 2014; Han *et al*., 2017; Chen, 2020; Piccoli and Castro, 2021), this operationalization does not include neutral market conditions. To address this, in an additional unreported test, we defined a bullish (bearish) market as a week in which the market return is 0.5 standard deviation above (below) zero, while a week with market returns smaller than this threshold is classified as neutral. The results, which can be made available upon request, indicate that the positive attention-return relation is observable in every market condition. We thank an anonymous reviewer for raising this point.

The results for these three investigations are presented in Table 6, panels A, B, and C. In panel A, the estimates indicate similar patterns in the attention-trade relationship between the bullish (b's) and bearish (c's) for both regimes. It is also worth mentioning that lagged attention is the most significant parameter for both market states, providing additional evidence that retail investors actually search for online information before trading on the given stocks.

The pattern exhibited during high- and low-volatility periods is slightly different (see Panel B). Even though in both regimes the one-lag attention is positive and significant, indicating that individual investors' preference for skewness does not drive our main findings, it can be seen that the association is slightly more predominant when volatility is high, which is consistent with the view that volatility and individual trades are related (Kostopoulos *et al.*, 2020; Cheng *et al.*, 2021). This behavior suggests that investor attention provides a better explanation for contemporaneous and future trades of individual investors in high-volatility circumstances, which can be explained by the gambling preference of these investors (Barberis and Huang, 2008; Kumar, 2009; Barberis and Xiong, 2012).

Finally, the results in panel C demonstrate that the influence of online searches on future trades previously documented is not driven by the scalation of individual investors experienced by the Brazilian stock market during 2019-2020. Contrarily, the results indicate that during periods of "regular" growth of retail accounts in the local stock market (c's), this relation is not only significant but still more pronounced, indicating that our main findings are not biased by the peculiar growth of retail investors in the Brazilian market or by the COVID-19 outbreak in 2020, which significantly shook stock markets worldwide as well as investors' attention.

Table 6: Attention-trades relation and market states

$S_t.D_t$	$S_{t-1}.D_{t-1}$	$S_{t-2}.D_{t-2}$	$S_t.(1-D_t)$	$S_{t-1}.(1-D_{t-1})$	$S_{t-2}.(1-D_{t-2})$	T_{t-1}	T_{t-2}	R_t	R_{t-1}	R_{t-2}	Intercept	adj. R^2
Panel A: Bull/bear market												
0.285	0.573	0.142	0.187	0.561	0.358	-0.623	-0.263	0.377	0.501	1.129	0.006	36.0
(2.12)**	(4.45)***	(1.21)	(1.71)*	(5.60)***	(2.63)***	(-6.66)***	(-3.12)***	(0.86)	(1.36)	(2.19)**	(0.45)	
Panel B: High/low volatility												
0.255	0.442	0.225	0.221	0.663	0.211	-0.619	-0.255	0.218	0.469	1.126	0.006	35.9
(2.00)**	(4.46)***	(1.72)*	(1.85)*	(6.11)***	(1.66)*	(-6.42)***	(-2.96)***	(0.50)	(1.35)	(2.30)**	(0.44)	
Panel C: Increase of retail investors												
0.137	0.402	0.242	0.462	0.857	0.245	-0.641	-0.268	0.165	0.320	1.154	0.005	38.5
(1.34)	(4.65)***	(2.05)**	(2.80)***	(5.30)**	(1.71)*	(-6.49)***	(-3.06)***	(0.40)	(1.01)	(2.51)***	(0.45)	

The table shows the estimates for the regressions between online searches (S) and retail investors' trades (T) under different market states, captured by the dummy variable (D) using the following OLS model:

$$T_t = a_0 + \sum_{n=0}^{2} b_n D_{t-n} S_{t-n} + \sum_{n=0}^{2} c_n (1 - D_{t-n}) S_{,t-n} + \sum_{n=1}^{2} d_n T_{t-n} + \sum_{n=0}^{2} e_n R_{t-n} + \varepsilon_t \quad (10)$$

Where R is the weekly market return Panel A presents the results for bull/bear markets, where the dummy equals 1 if the contemporaneous market return is positive (bull markets). Panel B presents the results for high/low-volatility periods, where the dummy equals 1 if the contemporaneous conditional volatility, measured by a GARCH (1,1) model on market returns, is above the mean. Panel C shows the results when the dummy equals 1 for 2019 and 2020 when the Brazilian market experienced an acute increase in the number of individual investor accounts. The sample period is from 01.01.2018 to 11.30.2020. The numbers in parentheses are t-statistics from the Newey-West standard error estimator. The R^2 values are expressed as percentages. N=149.

***Significant at 1%, **5%, and *10% levels

Taken together, the findings in Table 6 suggest that the relationship between investor attention and trades is observable in different market states, providing strong support for the widespread assumption that online searches are a consistent proxy for retail investors' intention to trade on a given stock. However, the results so far have been obtained based on debatable methodological choices, especially regarding the construction of our aggregated attention proxy. To address this, the following section introduces different robustness checks that replicate the main analysis for the three different specifications for this proxy. First, employing the company's names instead of the ticker; second, employing alternative approaches to construct the aggregated attention index; and finally, calculating investor attention (S) using the detrended log difference of the ARA index, as proposed by Da *et al*. (2011).

4. Robustness checks

4.1 Searching on stock names

As mentioned in the Methodology section, our ARA index is constructed using stock tickers as the search terminology to avoid capturing individual attention unrelated to investment decisions. Since our focus is on retail investors' attention, it is legitimate to believe that this choice filters a good number of searches related to the investment decisions of more naïve investors that search using a company's name instead of its ticker. The disadvantage of using a firm's name as a proxy for attention is that it incorporates more searches unrelated to investment decisions, such as attention towards a company's products or job positions. Nevertheless, we advocate that this alternative approach is consistent with the nature of a robustness test that aims to verify whether our results are driven by our attention proxy. Furthermore, the results can also contribute to the debate on whether using a company's name in Google Trends actually captures investors' intention to trade on a given stock, as assumed by some studies (e.g., Aoaudi *et al*., 2013; Bijl *et al*., 2016). Table 7 presents the results of the VAR (Panel A) and OLS analyses (Panel B) employing this alternative approach to construct the ARA index.

Table 7: Attention-return relation using companies' names for online searches

	S_t	S_{t-1}	S_{t-2}	T_{t-1}	T_{t-2}	R_t	R_{t-1}	R_{t-2}	Intercept	adj. R^2
Panel A: VAR analysis										
T_t		0.21	0.05	-0.50	-0.16				0.01	26.7
		[4.04]***	[0.95]	[-6.04]***	[-2.09]***				[0.81]	
S_t		-0.46	-0.09	-0.02	0.03				0.01	19.2
		[0.082]	[0.086]***	[0.129]	[0.121]				[0.022]	
T does not Granger cause S			0.918		[0.17]					
S does not Granger cause T			0.000		[17.42]					
Panel B: OLS analysis										
(6)	0.13	0.21	-0.06						0.01	9.7
	(2.41)**	(4.19)***	(-1.06)						(0.66)	
(7)	0.12	0.27	0.04	-0.52	-0.19	0.11	0.33	1.18	0.00	34.9
	(2.37)**	(5.25)***	(0.85)	(-6.66)***	(-2.27)**	(0.26)	(0.90)	(2.49)**	(0.55)	

The table shows the estimates for the VAR (Panel A) and OLS models (Panel B) relating to online searches and retail trades, where the former is obtained using company names to construct the ARA index. The number of lags is based on the AIC. In panel A, the standard errors are reported below the coefficients, and in the lower section, the *p*-values of the Granger causality tests are reported in addition to the respective chi-squared between brackets. In panel B, the numbers in parentheses are *t*-statistics from the Newey-West standard error estimator. The sample period is from 01.01.2018 to 11.30.2020. The R^2 values are expressed as percentages. N=149.
***Significant at 1%, **5%, and *10% levels

The estimates for the VAR model indicate that the influence of attention on future trades is still present but is restricted to the first lag, disappearing after that, whereas no influence is found in the opposite direction. The fact that the attention-return relation is less pronounced in both directions in this alternative approach is consistent with the assumption that using a company's name instead of its ticker captures a larger proportion of searches that are not investment-related. Finally, it is worth mentioning that the Granger causality of attention on trades remains significant at the 1% level, whereas no causation is found in the opposite direction. In the OLS analysis (Panel B), the relationship between attention and trades remains almost the same. However, the parameter of the two-lag attention ceases to be significant, which can be plausibly explained by the notion that stock names are a less accurate measure of future trades on a given stock. Once more, the attention-trade relationship is more pronounced for one lag of attention.

Overall, the results support previous findings and the notion that online searches are linked with future trades from retail investors. Another important message from the results reported in Tables 3 and 7 is that the debate regarding using tickers or firms' names to capture investor attention is virtually innocuous since both approaches seem to satisfactorily capture investors' intention to trade on the given asset.

4.2 Aggregated attention index specification

The fact that we weight the stock's GSV using its market capitalization could bias our findings towards the attention-trade relation hypothesis since it is reasonable to expect that aggregated market values and aggregated retail trades are related. Moreover, this approach implicitly assumes that larger companies attract more attention than medium or small firms, which may not necessarily be the case, since the literature reports that retail investors in particular are attracted by smaller firms (Barber *et al*., 2009; Da *et al*., 2011; Cheng *et al*., 2021). We alternatively compute an equally weighted ARA index and replicate VAR and OLS analyses to circumvent this. Another possible specification for retail investors' attention is based on the GSV of the market index, an approach employed in several studies (e.g., Vozlyublennaia, 2014; Hamid and Heiden, 2015; Baur and Dimpfl, 2016). In this case, the approach captures investor attention at the market level, which could mean a less direct intention to trade on the stock market when compared to firm-level attention. The rationale behind this conjecture is provided by the limited attention framework (Barber and Odean, 2008), because the search for a specific stock would be more closely related to an intention to trade the security in question rather than a more generic search of the stock market index.

Given this context, our second alternative specification for aggregated attention is the GSV of the market index (i.e. "Ibovespa"). This second parameter serves as a robustness test for our main findings and allows us to test whether it properly captures the interest of retail investors in trading in the stock market, as assumed by these studies. The results of both approaches are presented in Table 8.

Table 8: Attention-return relation alternative specification for ARA

	S_t	S_{t-1}	S_{t-2}	T_{t-1}	T_{t-2}	R_t	R_{t-1}	R_{t-2}	Intercept	adj. R^2
Panel A: Equally weighted ARA Index										
VAR analysis										
T_t		0.35	0.15	-0.54	-0.23				0.01	
		[0.093]***	[0.096]	[0.082]***	[0.08]***				[0.014]	25.4
S_t		-0.36	-0.26	0.10	-0.07				0.01	
		[0.081]***	[0.083]***	[0.071]	[0.069]				[0.013]	16.1
T does not Granger cause S			0.073		[5.25]					
S does not Granger cause T			0.001		[14.44]					
OLS analysis										
(6)	0.20	0.29	0.01						0.01	4.0
	(1.94)*	(3.81)***	(0.08)						(0.62)	
(7)	0.29	0.46	0.19	-0.61	-0.23	0.03	0.51	1.19	0.00	36.2
	(3.20)***	(7.00)***	(1.97)**	-(6.69)***	(-2.62)**	(0.10)	(1.44)	(2.57)**	(0.47)	
Panel B: Attention on market index										
VAR analysis										
T_t		0.23	0.05	-0.55	-0.23				0.01	
		[0.047]***	[0.05]	[0.083]***	[0.079]***				[0.014]	30.1
S_t		-0.30	-0.15	0.08	-0.23				0.01	
		[0.082]	[0.087]***	[0.145]*	[0.138]				[0.025]	12.4
T does not Granger cause S			0.068		[5.38]					
S does not Granger cause T			0.000		[25.39]					
OLS analysis										
(6)	0.13	0.19	0.07						0.01	7.8
	(3.12)***	(3.86)***	(1.20)						(0.67)	
(7)	0.15	0.29	0.09	-0.62	-0.23	0.39	0.57	1.05	0.01	38.0
	(4.01)***	(6.36)***	(1.72)*	(-7.60)***	(-2.76)***	(1.11)	(1.31)	(1.97)**	(0.46)	

The table shows the estimates for the VAR and OLS models relating to online searches and retail trades, employing alternative specifications for ARA. We employ an equal average of stocks' GSV to generate the ARA index in panel A. In panel B, we use the GSV of the market index (i.e. "Ibovespa") as the proxy to ARA. For the VAR analysis, the standard errors are reported below the coefficients. The lower section reports the *p*-values of the Granger causality tests and the respective chi-squared between brackets. For the OLS model, the numbers in parentheses are *t*-statistics from the Newey-West standard error. The sample period is from 01.01.2018 to 11.30.2020. The R^2 values are expressed as percentages. N=149
***Significant at 1%, **5%, and *10% levels

Panel A presents the results for the equally weighted ARA index. The VAR model estimates demonstrate the influence of attention on trades one week ahead, while no relation is found in the opposite direction. Consistently, the Granger causation of attention on trades remains significant at the 1% level. The OLS analysis of Panel A reveals once again that the attention-trade relation is significant up to two lags and more pronounced for the one-lag parameter. Taken together: these findings demonstrate that the weighted average employed to construct the ARA index does not drive our main results. Finally, the R^2 values for both OLS models are slightly lower for the equally weighted ARA index, suggesting that the previous specification can better accommodate the data.

Panel B displays the estimates of attention using online searches on the market index. Once more, we document the influence of past attention on future trades, while the relationship in the opposite direction is marginally significant. This pattern is corroborated by the results of the Granger causality test. Finally, consistent with the results documented throughout the study, one lag of attention is the most significant parameter for forecasting retail investors' trades.

The findings in panel B indicate that employing online searches on the index is an adequate proxy for aggregated investor trades, as assumed by some studies (e.g., Vozlyublennaia, 2014; Hamid and Heiden, 2015; Baur and Dimpfl, 2016). Furthermore, a comparison between different specifications for aggregated retail investors' attention (i.e., value-weighted, equally weighted, and market index) demonstrate that, concerning retail trades, these measures share very similar patterns from both statistical and economic standpoints, supporting the view that they are capturing what they aim to achieve.

4.3 Abnormal attention

In their seminal study, Da *et al*. (2011) measure investor attention based on the abnormal level of online searches, defined as the log difference between the GSV in a given week and its prior eight weeks median. According to the authors, the intuition behind this is to avoid recent spikes in attention, time trends, or low-frequency issues. Following their seminal study, many studies have employed the abnormal attention method as a proxy for investor attention (Da *et al*., 2011; Smith, 2012; Bijl *et al*., 2016; Da *et al*., 2020). However, this approach is far from unanimous since another bulk of articles employs the most direct measure based on the log difference of GSV in two consecutive periods (Baur and Dimpfl, 2016; Han *et al*., 2017; Wu *et al*., 2019; Yoshinaga and Rocco, 2020). In this case, the attention measure could be understood as being noisier, since it would not control for attention spikes or time trends, leaving room for the hypothesis that this approach could lead to less consistent results. To address this, we calculated an alternative measure of attention (S) using the abnormal attention approach of Da *et al*. (2011), as expressed in Eq. (11):

$$S_t = log\ (ARA_t)\ - log\ [Med(ARA_{t-1}), \dots, (ARA_{t-8})] \tag{11}$$

Following the pattern in the previous subsections, we focus on the VAR and OLS analyses, whose estimates are presented in Table 9. Using this alternative approach, the VAR model confirms the influence of abnormal attention on future trades, whereas a less significant relationship is observable in the opposite direction, which is similar to the behavior documented by our main model. Moreover, the Granger causality test indicates that causation is significant only from attention towards trades. Curiously, the signal of the two lags attention is negative, suggesting that an increase in attention provokes a reversal in trades two weeks ahead, which could be related to the reversal pattern between online searches and future returns documented in the literature (Da *et al*., 2011) for the Brazilian market (Yoshinaga and Rocco, 2020).

Table 9: Attention-return relation using abnormal attention

	S_t	S_{t-1}	S_{t-2}	T_{t-1}	T_{t-2}	R_t	R_{t-1}	R_{t-2}	Intercept	adj. R^2
Panel A: VAR analysis										
T_t		0.45	-0.21	-0.53	-0.26				0.01	29.0
		[0.092]***	[0.087]**	[0.081]***	[0.081]***				[0.78]	
S_t		0.52	-0.24	0.14	0.08				0.00	27.1
		[0.082]***	[0.078]***	[0.072]*	[0.072]				[0.35]	
T does not Granger cause S			0.160		[3.66]					
S does not Granger cause T			0.000		[24.50]					
Panel B: OLS analysis										
(6)	0.15	0.22	-0.26						0.01	7.5
	(1.54)	(1.94)*	(-3.58)						(0.66)	
(7)	0.25	0.36	-0.06	-0.63	-0.33	0.31	0.70	1.16	0.00	35.0
	(2.67)***	(4.64)***	(-0.73)	(-6.75)***	(-3.69)***	(0.76)	(2.06)**	(2.40)**	(0.43)	

The table shows the estimates for the VAR (Panel A) and OLS models (Panel B) relating investor attention and retail trades, where the former is defined as the log difference between the GSV in a given week and its prior eight weeks median. The number of lags is based on the AIC. In panel A, the standard errors are reported below the coefficients, and in the lower section, the *p*-values of the Granger causality tests are reported in addition to the respective chi-squared between brackets. In panel B, the numbers in parentheses are *t*-statistics from the Newey-West standard error estimator. The sample period is from 01.01.2018 to 11.30.2020. The R^2 values are expressed as percentages. N=149

***Significant at 1%, **5%, and *10% levels

For the OLS analysis (Panel B), the association between attention and trades is similar to that observed when employing the equally weighted index. The main difference relies on the estimate for two lag parameters, whose negative coefficient indicates that an increase in abnormal online searches is associated with a decrease in trades two weeks ahead. This relation seems persistent, and observable in both OLS models and the VAR approach, albeit significant only in the latter. Overall, the results in Table 9 provide evidence that the attention-trades relationship documented is not driven by using log differences for two consecutive weeks to measure S. The findings also indicate that both approaches broadly used in the literature (abnormal attention and regular log differences), consistently capture future retail investor trades.

In summary, the evidence provided by the robustness checks does not indicate that methodological choices regarding the ARA index could bias our main findings. In addition, the results shown in the present section contribute to the ongoing debate on whether distinct approaches to attention based on online searches effectively capture retail investors' intention to trade on the given stock. Our findings indicate that the answer to this question is yes.

5. Conclusion

This study investigates whether online searches are actually related to retail investors' future trades, as assumed by several studies on investor attention. To address this, we generate an index of aggregated retail attention (ARA) composed of the Google Search Volumes of the tickers of stocks belonging to the local market index (Ibovespa) between 2018 and 2020, weighted by the corresponding market capitalization. We compare the weekly changes in the ARA index with a unique dataset containing the total amount traded by Brazilian retail investors made available by the local stock exchange (B3). The VAR analysis between both constructs indicates strong Granger causality of online searches on trades (significant at the 0.1% level), whereas the influence in the opposite direction is marginally significant (significant at the 10% level). Employing an OLS analysis, we also find a positive relationship between contemporaneous and lagged attention on trades and that this relationship is more pronounced in the case of one-lag attention, which is consistent with the view that retail investors search for stocks under interest in the web before deciding to trade them.

Our empirical analysis also shows that the attention-trades relationship is observable under different market states, such as bull/bear markets, high/low volatility regimes, and during distinct inflows of retail investors in the Brazilian stock market, suggesting that this relationship holds conditionally. Furthermore, we investigate whether the different specifications of online search measures commonly used in the study are accurate proxies for future retail investors' trades. Our results demonstrate that using stock names, stock tickers, or market index names as search terminologies in Google Trends are consistent approaches to capture attention. In the same line, we also report that employing log differences on two consecutive periods or using abnormal attention measures (i.e., the log-difference between GSV on a given week and its previous moving-window median) leads to similar results in terms of consistency. Overall, the results here documented indicate that online search measures are pertinent proxies for the presence of retail investors in the stock market. Fisher Black (1986, p. 530) predicted that "someday … [the] influence of noise traders will become apparent." Based on our findings, we conclude that Google Trends has uncovered this influence.

Acknowledgments

We are grateful to Christiane Bariquelli from B3 for supplying empirical data on retail investor trades. We also appreciate the comments from Claudia Yoshinaga, Newton C.A. da Costa Jr, Myrna Lima, and the participants at the 8th Brazilian Behavioral Economics and Finance Meeting.

References

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

Bali, T. G., Brown, S. J., Murray, S., Tang, Y., 2017. A lottery-demand-based explanation of the beta anomaly. *Journal of Financial and Quantitative Analysis*, 52(6), 2369-2397

Barber, B. M., 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(2), 785-818

Barber, B. M., Odean, T., Zhu, N., 2009. Do retail trades move markets? *Review of Financial Studies*, 22(1), 151-186

Barberis, N., Huang, M., 2008. Stock as lotteries: the implication of probability weighting for security prices. *American Economic Review*, 98 (5), 2066-2100.

Barberis, N., Xiong, W., 2012, Realization utility. *Journal of Financial Economics*, 104, 251-271

Baur, D. G., Dimpfl, T., 2016. Googling gold and mining bad news. *Resources Policy,* 50, 306-311

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

Bohemer, E., Jones, C., Zhang, X., Zhang, X., 2021. Tracking retail investor activity. *Journal of Finance,* 76(5), 2249-2305

Cheng, F., Wang, C., Chiao, C., Yao, S., Fang, Z., 2021. Retail attention, retail trades, and stock price crash risk. *Emerging Markets Review*, 49, 100821

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

Da, Z., Engelberg, J., Gao, P., 2015. The sum of all FEARS. Investor sentiment and asset prices. *Review of Financial Studies*, 28(1), 1-31

Da, Z., Hua, J., Hung, C., Peng, L., 2020. Market returns and a tale of two attentions. *SSRN Working Paper #3551662*. *SSRN Electronic Journal*

Desagre, C., D'Hondt, C., 2021. Googlization and retail trading activity. *Journal of Behavioral and Experimental Finance*, 29, 100453

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

Fisher, B., 1986. Noise. *Journal of Finance*, 41(3), 529-543

Hamid, A., Heiden, M., 2015. Forecasting volatility with empirical similarity and GoogleTrends. *Journal of Economic Behavior and Organization*, 117, 62-81

Han, B., Kumar, A., 2013. Speculative retail trading and asset prices. *Journal of Financial and Quantitative Analysis*, 48(2), 377-404

Han, L., Li, Z., Yin, L., 2017. The effects of investor attention on commodity futures markets. *Journal of Futures Markets*, 37(10), 1031-1049

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

Karlsson, N., Loewenstein, G., Seppi, D., 2009. The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty*, 38(2), 95-115

Kelley, E. K., Tetlock, P. C., 2013. How wise are crowds? Insights from retail orders and stock returns. *Journal of Finance*, 68(3), 1229-1265

Kostopoulos, D., Meyer, S., Uhr, C., 2020. Google search volume and individual investor trading. *Journal of Financial Markets*, 49, 100544

Kumar, A., 2009. Who gambles in the stock market? *Journal of Finance*, 64(4), 1889-1933

Kumar, A., Lee, C. M. C., 2006. Retail investor sentiment and return Comovements. *Journal of Finance*, 61(5), 2451-2486

Liu, J., Stambaugh, R. F., Yuan, Y., 2018. Absolving beta of volatility's effects. *Journal of Financial Economics*, 128(1), 1-15

Piccoli, P., de Castro, J., 2021. Attention-return relation in the gold market and market states. *Resources Policy*, 74, 102333

Smith, G. P., 2012. Google internet search activity and volatility prediction in the market for foreign currency. *Finance Research Letters*, 9(2), 103-110

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

Wang, Q., Zhang, J., 2015. Does individual investor trading impact firm valuation? *Journal of Corporate Finance*, 35, 120-135

Wen, F., Xu, L., Ouyang, G., Kou, G., 2019. Retail investor attention and stock price crash risk: Evidence from China. *International Review of Financial Analysis*, 65, 101376

Wu, Y., Han, L., Yin, L., 2019. Our currency, your attention: Contagion spillovers of investor attention on currency returns. *Economic Modelling,* 80, 49-61

Yoshinaga, C., Rocco, F., 2020. Investor attention: Can Google search volumes predict stock returns? *Brazilian Business Review*, 17(5), 523-539

Yu, J., Yuan, Y., 2011. Investor sentiment and the mean-variance relation. *Journal of Financial Economics*, 100(2), 367-381

ARTICLE 2

- Paper published in the Resources Policy Volume 74, December 2021, 102333

Attention-return relation in the gold market and market states

ABSTRACT

We document that the attention-return relation in the gold spot market is sensitive to market states since we identify a positive (negative) association between returns and contemporaneous attention during bull (bear) markets. This sensitivity is influenced by the magnitude of the returns, since the significance of the relation, based on a quantile model, increases almost monotonically as the return distribution moves away from the median. We further find that investor attention is influenced by recent gold spot price trends in both market states. Overall, these results suggest that gold return past performance attracts the attention of individual investors who, in turn, trade to profit from the trend, reinforcing the price movement, which is consistent with the time-series momentum framework.

Keywords: Gold, Market States, Investor Attention, Google Trends

JEL Classification: G14, G17, Q02, Q37

1. Introduction

There is a growing literature investigating the relation between asset returns and investor attention, the latter commonly proxied by the volume of on-line searches for the given asset. In this context, many studies document that investor attention can predict stock prices (Joseph *et al*., 2011; Vozlyublennaia, 2014; Bijl *et al*., 2016), exchange rates (Goddard *et al*., 2015; Han *et al*., 2018; Wu *et al*., 2019), cryptocurrency returns (Dastgir *et al*, 2019; Eom *et al*., 2019) and commodity prices (Han *et al*., 2017; Yao *et al*., 2017). In the specific case of the gold market, previous studies identify a bi-directional causality between attention and return (Jain and Biswal, 2019), and contend that attention predicts future prices (Han *et al*., 2017) or that on-line searches are related to volatility in spot prices (Baur and Dimpfl, 2016). Overall, the evidence indicates that investor attention plays an important role in understanding the behavior of the gold market. Since the attention-return relation seems to be well stablished in this market, a plausible question is how pervasive this relation may be; more specifically, whether this association is sensitive to market states, in terms of significance and sign. Furthermore, since individual investors exhibit a preference for skewness (Barberis and Huang, 2008), does this relation change conditional on the magnitude of the returns? If the answer to both questions is positive, what are the implications for gold pricing? These are the questions that the present study aims to address.

In this regard, we construct a weekly time-series of Google Search Volumes (GSV) for the term, "Gold price",[8] from 01.01.2007 to 12.31.2018, employing the log-difference as our proxy for investor attention, and analyze its relation to gold spot returns during bull (bear) markets, defined as those weeks when the return is positive (negative).[9] Our results indicate a positive (negative) relation between the two variables during up (down) markets, which is also asymmetric in terms of magnitude since it is more pronounced in bull markets. This association is also economically important, since a one-standard deviation change in investor attention corresponds to a contemporaneous change in price of 7.2% (-3.7%) in monthly terms for up (down) markets. This asymmetry in terms of significance and magnitude in the attention-return relation is the first contribution to the literature since, to our knowledge, this is the first study to conduct this investigation in the gold spot market.[10]

[8]In one of the robustness checks we employ alternative terminologies and obtain consistent results.
[9]In an alternative model, we define market states based on a four-week moving average and find virtually the same results.
[10]Bauer and Dimpfl (2016) analyze the asymmetric relation in the opposite direction (i.e., attention as the dependent variable) and document a stronger effect of negative returns on investor attention.

The study's second contribution is on how the attention-return relation varies with return distribution. More specifically, we run a quantile regression of gold returns and on-line searches and find that this association is absent only for returns around the median, while it is significant in all the remaining quantiles investigated (e.g., 0.05, 0.10, 0.25, 0.75, 0.90 and 0.95). Another interesting pattern is that the relation is always negative (positive) for the bottom (top) quartiles, and rises almost monotonically when the distribution moves from the mean to the tails. These features indicate that the relation is more pervasive during more turbulent circumstances, which is consistent with the preference for gambling by individual investors noted in some studies (Barberis and Huang, 2008; Kumar, 2009). Moreover, since gold is seen as a countercyclical investment, its behavior in different market states bears important implications for portfolio strategies, as well as for the literature on the influence of retail investors on financial markets.

Finally, we also investigate how investor attention is affected by previous returns based on the sign of the price change. Our results indicate that this influence is also asymmetric, being positive (negative) during bull (bear) markets, which indicates that recent price trends (increase or decrease) attract the attention of individual investors. Hence, a plausible conclusion from our aggregated results is as follows. Gold price past performance attracts the attention of retail traders who, aiming to profit from the trend, engage in trades that lead to a further increase or decrease in price, reinforcing the momentum effect. This explanation provides support to the time-series momentum anomaly documented by Moskowitz *et al*. (2012), which is also observable in the gold market since Baur *et al*. (2020) report that it generates abnormal returns for 12- and 24-month lagged price strategies. Consequently, the aggregated results reported in this study suggest that investor attention is a candidate explanation for the pervasive time-series momentum puzzle (Moskowitz *et al*., 2012; Papailias *et al*., 2021).

The remainder of the paper is structured as follows. Section 2 describes the data and the methodology. Section 3 presents and discusses the results, while Section 4 deals with the robustness tests. Section 5 concludes.

2. Data and Methodology

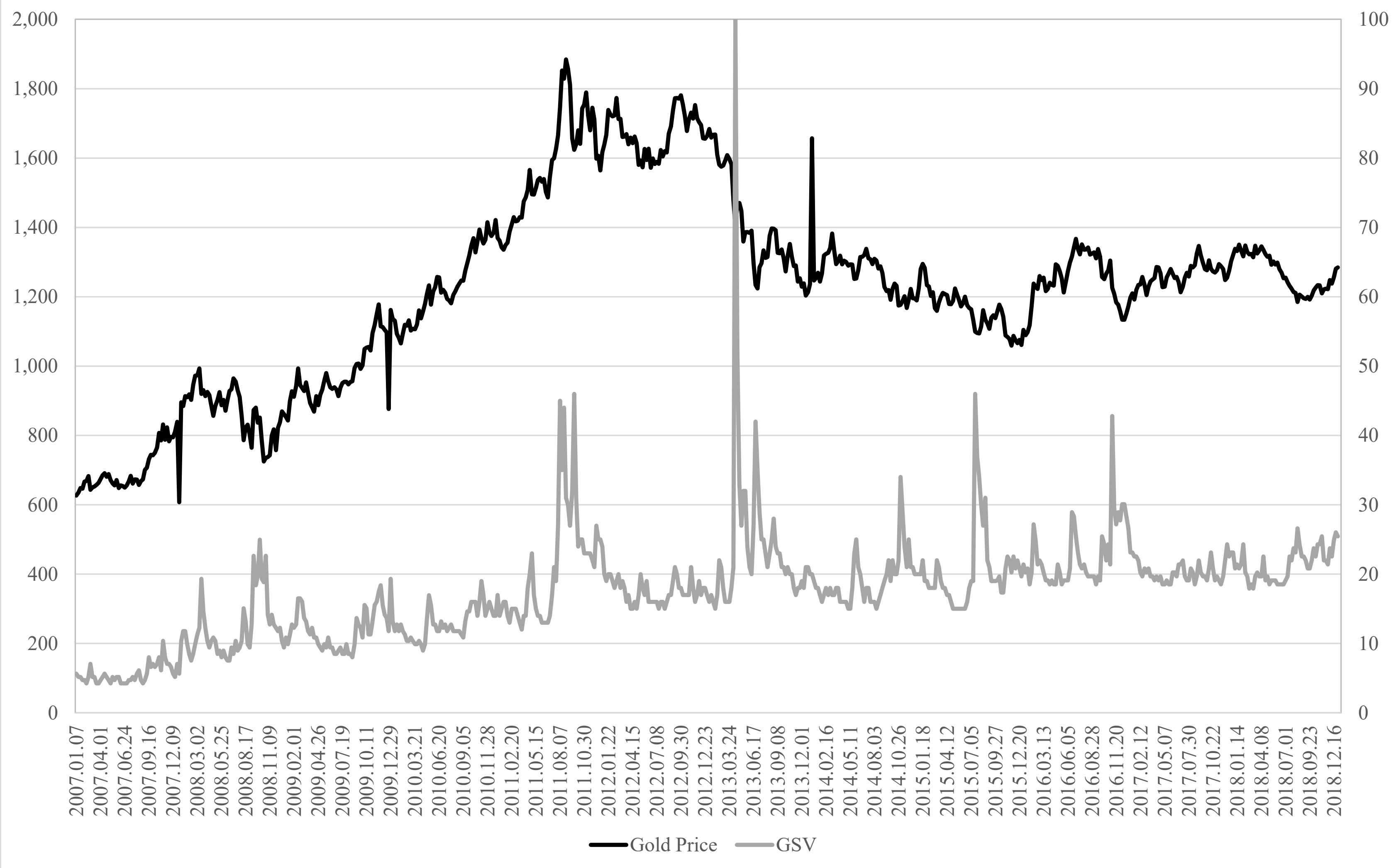
2.1 Data

The primary data for this study comprise weekly returns calculated based on the log-difference of the gold spot price in USD from 01.01.2007 to 12.31.2018, obtained from Investing.com. The data for Google Search were collected from the Google Trends platform (www.google.com/trends) using the term, “Gold Price”, for worldwide searches, with a filter for the option, “Finance”, available on this tool, as suggested by Baur and Dimpfl (2016).[11] For the entire sample, Google Trends makes available only monthly data, where the period with the largest number of searches receives an amount of 100. To obtain weekly data, we use 2 overlapping windows of 5 years (01.01.2007 to 12.31.2011 and 01.01.2011 to 12.31.2015) and 1 overlapping window of 4 years (01.01.2015 to 12.31.2018), since 5 years is the longest period for which Google displays GSV information on a weekly basis. The one-year overlaps are used to estimate the corresponding search volumes in relative terms. In this process, the window where the monthly GSV is the largest is the “reference window”, and the GSV of the remaining windows are calculated using the overlaps data to regress the GSV window (dependent variable) in the GSV reference window (independent variable). For our data, the monthly GSV peaks in April 2013. Hence, our second window assumes the role of the reference window, and the weekly GSV of the remaining two windows (1 and 3) are rescaled using the coefficients of the aforementioned regression. Appendix A exhibits the coefficients used in this process, as well as an illustrative example for additional clarification. Since the Google Trends time-series are non-stationary, we use the log-difference in GSV to examine the relations proposed in the study,[12] following several studies (Baur and Dimpfl, 2016; Dastgir *et al*., 2018; Zhang *et al*., 2018).

Figure 3 presents a chart associating gold price (left axis) and GSV (right axis). Several spikes and crashes in the gold price are associated with spikes in on-line searches, suggesting that the attention-return relation is more pronounced during extreme circumstances, a feature that will be discussed in more detail in a later section.

[11]In the robustness section, we include other possible key words that could be used in on-line searches for gold and extract the first principal component among them. The results are virtually the same.
[12]The time series of the changes in Google Trends measured by the augmented Dickey-Fuller test are stationary at the 1% level ($Z_{(t)}$= -28.165).

Figure 3: Gold spot prince and Google Search Volumes



Note: The figure plots the Gold spot price in USD (black line, left axis) together with the weekly GSV of "Gold price" searches.

To better illustrate some features of the main variables, Table 10 shows the summary statistics for gold returns (R) and changes in on-line searches (S), as well as the correlation among these variables (bottom line) for the entire dataset during both bull and bear markets, defined as the weeks with positive and negative returns, respectively.

Table 10: Summary Statistics

	Entire Sample		Bull Markets		Bear Markets	
	R	S	R	S	R	S
Mean	0.0011	0.0025	0.0208	-0.0034	-0.0212	0.0093
Median	0.0016	0.0000	0.0156	0.0000	-0.0141	0.0000
Std Dev	0.0383	0.1538	0.0327	0.1459	0.0314	0.1622
Skewness	0.9827	2.3629	7.7442	0.5102	-6.2015	3.8718
Kurtosis	39.7977	20.1201	73.4651	3.8869	50.4430	31.5778
Correlation	0.09		0.36		-0.11	

The table presents the weekly summary statistics for the entire sample (01.01.2007 to 12.31.2018) and for bull (bear) markets, defined as market states when the contemporaneous log-return (R) of gold is positive (negative). Attention (S) is defined as the log difference of Google Search Volumes for the term, “Gold Price”, filtering for the option, “Finance”, available on Google Trends.

There is a clear contrast between the distributions of the variables in the two market states: whereas in bull markets the mean attention is negative, with almost negligible skewness and low kurtosis, in bear markets the attention distribution has a positive mean, is positively asymmetric, and exhibits large tails. On the other hand, the return distribution during bull markets exhibits a more pronounced asymmetry and fatter tails than in bear states, which may appear to be puzzling at first glance. A plausible explanation for this pattern relies on the countercyclical nature of gold, whose price increases more sharply during downturns in other markets, leading to more volatile returns in gold bull markets. It is also interesting to note that the variables exhibit a positive (negative) correlation during bull (bear) markets that is more pronounced in magnitude than in the whole sample, suggesting that this relation is sensitive to market states. The next subsections describe how this association is quantitatively examined.

2.2 Econometric Models

We first analyze the return-attention relation for the entire dataset to contrast the results for the bull and bear markets. In this regard, we employ three models: (1) without controlling for lagged return; (2) controlling for autocorrelation by employing lagged returns; and (3) introducing an interactive term whose coefficient to measure the change in the coefficient of attention conditional on past returns, as suggested in previous studies (Vozlyublennaia, 2014;

Han *et al*., 2017; Han *et al*., 2018; Chen, 2020). The interaction term also aims to investigate whether investor attention improves market efficiency in the gold market. Given that the influence of past returns on current returns is an indication of market inefficiency, a significant coefficient of the interaction term would indicate that the effect of lagged returns is influenced by investor attention. The models are expressed as follows:

$$R_t = a_0 + \sum_{n=0}^{3} b_n S_{t-n} + \varepsilon_t \tag{1}$$

$$R_t = a_0 + \sum_{n=0}^{3} b_n S_{t-n} + \sum_{n=1}^{3} c_n R_{t-n} + \varepsilon_t \tag{2}$$

$$R_t = a_0 + \sum_{n=0}^{3} b_n S_{t-n} + \sum_{n=1}^{3} c_n R_{t-n} + \sum_{n=1}^{3} d_n R_{t-n} \times S_{t-n} + \varepsilon_t \tag{3}$$

where R_t is the return on the gold spot price and S_t is the log-difference of GSV in week t. The number of lags was determined using the AIC. For the market states models, (4) and (5), we add to Models (1) and (2) a dummy variable that equals 1 when the contemporaneous return is positive (bull market), and 0 otherwise (bear market).

$$R_t = a_0 + \sum_{n=0}^{3} b_n D_{t-n} S_{t-n} + \sum_{n=0}^{3} c_n (1 - D_{t-n}) S_{t-n} + \varepsilon_t \tag{4}$$

$$R_t = a_0 + \sum_{n=0}^{3} b_n D_{t-n} S_{t-n} + \sum_{n=0}^{3} c_n (1 - D_{t-n}) S_{,t-n} + \sum_{n=1}^{3} d_n R_{t-n} + \varepsilon_t \tag{5}$$

The focus here is on the coefficients, b_n and c_n, since distinct signs or magnitudes would indicate that the attention-return relation varied with market states. Finally, to avoid a potential bias related to our definition of bull and bear markets, for Models (6) and (7), we alternatively define a dummy variable using a four-week moving average: the dummy equals one if the average return during Weeks -3 to 0 is positive, and 0 otherwise.

3. Results

3.1 OLS Regressions

Table 11 presents the results of Models (1) to (3). The attention-return relation is mostly positive, albeit not significant. Noteworthily, the gold spot returns seem to follow an AR(1) model since the one-lag return coefficient is significant at the 1% level. In Model (3), the fact that, with one lag, the interaction term exhibits a coefficient with an opposite sign of the past return indicates that an increase in attention attenuates the relation between past and current returns, since the aggregated effect ($c_1 + d_1$) is close to zero, suggesting that investor attention improves market efficiency, as documented in previous studies (e.g.: Vozlyublennaia, 2014). Furthermore, the magnitude and significance of the interaction between attention and return with one lag suggests that the marginal effect of attention on future gold prices depends on past returns. More precisely, the positive coefficient indicates that an increase in attention during bull (bear) markets is associated with future positive (negative) returns on gold, providing support for the hypothesis that the attention-return relation is conditional on market states.

Table 11: OLS analysis between return and attention

	S_t	S_{t-1}	S_{t-2}	S_{t-3}	R_{t-1}	R_{t-2}	R_{t-3}	$S_{t-1}.R_{t-1}$	$S_{t-2}.R_{t-2}$	$S_{t-3}.R_{t-3}$	Intercept	adj. R^2
(1)	0.024	-0.005	0.016	0.008							0.001	0.7
	(0.84)	(-0.43)	(1.14)	(0.90)							(0.91)	
(2)	0.016	-0.001	0.013	0.012	-0.347	-0.124	-0.071				0.002	11.0
	(0.69)	(-0.13)	(1.04)	(1.16)	(-3.09)***	(-1.69)*	(-1.90)*				(1.10)	
(3)	0.008	-0.003	0.013	0.012	-0.441	-0.089	-0.075	0.701	-0.004	0.082	0.001	15.3
	(0.42)	(-0.25)	(1.17)	(1.28)	(-3.31)***	(-1.08)	(-1.79)	(2.94)***	(-0.03)	(0.90)	(0.93)	

The table presents the cross-sectional relation between returns (R) and attention (S) for the entire sample, using the following models:

$$R_t = a_0 + \sum_{n=0}^{3} b_n S_{t-n} + \varepsilon_t \tag{1}$$

$$R_t = a_0 + \sum_{n=0}^{3} b_n S_{t-n} + \sum_{n=1}^{3} c_n R_{t-n} + \varepsilon_t \tag{2}$$

$$R_t = a_0 + \sum_{n=0}^{3} b_n S_{t-n} + \sum_{n=1}^{3} c_n R_{t-n} + \sum_{n=1}^{3} d_n R_{t-n} \times S_{t-n} + \varepsilon_t \tag{3}$$

where the interaction variable in Model (3) aims to capture the effect of attention conditional on a unit increase in the past return. The sample period is from 01.01.2006 to 12.31.2018. The numbers in parentheses are the t-statistics from the Newey-West standard error estimator. The number of lags was determined using the AIC. The adjusted R^2 values are percentages. N = 624.
***Significant at 1%, **5% and *10% levels

To better explore this explanation, Table 12 presents the results of Models (4) to (7), which employ a two-regime analysis using a dummy variable during bull markets. There is a clear asymmetry in the signal for a contemporaneous relation between return and attention. While, in bull markets, an increase in attention is associated with a current rise in gold prices, there is a negative association between these variables in bear markets. This association is also economically important since, *ceteris paribus*, an increase of one standard deviation in on-line searches corresponds to a contemporaneous increase (decrease) in price of 6.0% (3.0%) in monthly terms during up (down) markets, using the results from Model (5). Furthermore, the two-regime models seem to better explain the variations in price since the adjusted R^2 is higher in these models than in the single-regime models (Models (1) to (3)). Although the asymmetry observed has already been reported in previous studies (Baur and Dimpfl, 2016; Jain and Biswal, 2019), ours is the first study, to the best of our knowledge, to address the magnitude of the contemporaneous relation during different market states.

Table 12: OLS analysis between return and attention conditional on market states

	$S_t.D_t$	$S_{t-1}.D_{t-1}$	$S_{t-2}.D_{t-2}$	$S_{t-3}.D_{t-3}$	$S_t.(1-D_t)$	$S_{t-1}.(1-D_{t-1})$	$S_{t-2}.(1-D_{t-2})$	$S_{t-3}.(1-D_{t-3})$	R_{t-1}	R_{t-2}	R_{t-3}	Intercept	adj. R^2
(4)	0.089	-0.014	0.031	-0.003	-0.030	0.010	0.023	0.022				0.001	5.7
	(2.29)**	(-0.72)	(1.13)	(-0.26)	(-2.07)**	(0.58)	(1.96)*	(1.26)				(1.09)	
(5)	0.081	0.013	0.032	0.010	-0.042	-0.009	0.015	0.024	-0.370	-0.167	-0.106	0.002	17.0
	(3.05)***	(0.71)	(1.28)	(0.72)	(-2.93)***	(-0.49)	(1.22)	(1.45)	(-3.53)***	(-2.12)**	(-2.70)***	(1.45)	
(6)	0.120	-0.006	0.022	0.018	-0.044	-0.008	0.005	-0.005				0.000	10.4
	(3.11)***	(-0.25)	(0.96)	(1.11)	(-3.30)***	(-0.64)	(0.46)	(-0.42)				(0.31)	
(7)	0.106	0.032	0.031	0.031	-0.051	-0.028	-0.007	-0.008	-0.369	-0.124	-0.059	0.001	20.7
	(4.54)***	(1.47)	(1.28)	(1.52)	(-3.66)***	(-1.83)*	(-0.61)	(-0.71)	(-3.58)***	(-1.57)	(-1.34)	(0.48)	

The table presents the cross-sectional relation between returns (R) and attention (S) for bull and bear markets using the following models:

$$R_t = a_0 + \sum_{n=0}^{3} b_n D_{t-n} S_{t-n} + \sum_{n=0}^{3} c_n (1 - D_{t-n}) S_{t-n} + \varepsilon_t \qquad (4)$$

$$R_t = a_0 + \sum_{n=0}^{3} b_n D_{t-n} S_{t-n} + \sum_{n=0}^{3} c_n (1 - D_{t-n}) S_{,t-n} + \sum_{n=1}^{3} d_n R_{t-n} + \varepsilon_t \qquad (5)$$

Where D is a dummy variable that equals (1) if the return is positive (bull market), and 0 otherwise (bear market), for Models (4) and (5). In Models (6) and (7), we employ a four-week moving average to define market states, where the dummy variable equals (1) if the moving average is positive (bull markets), and 0 otherwise. The sample period is from 01.01.2006 to 12.31.2018. The numbers in parentheses are the t-statistics from the Newey-West standard error estimator. The number of lags was determined using the AIC. The adjusted R^2 values are percentages. N = 624.

***Significant at 1%, **5% and *10% levels

When we employ a four-week moving average window to define the market states (Models (6) and (7)), the behavior previously described becomes both statistically and economically more significant. For example, the coefficients in Model (7) indicate that an increase in attention is associated with an increase (decrease) in price of 7.2% (3.7%) in the same week. Consistently, the R^2 scores for these models are higher than for Models (4) and (5). Overall, the coefficients in Models (6) and (7) suggest that recent gold price past performance better captures investor attention on this market and its corresponding influence on returns. If this is the case, we should then expect past returns on gold to be related to future changes in attention. This association is investigated in the next subsection.

3.2 Influence of past return on future attention

To examine the influence of past returns on future attention, we perform a VAR analysis employing current attention as the dependent variable. Since the focus of the study is on the attention-return relation conditional on market states, we create a pair of series for investor attention,[13] defined as $S_t^{BULL} = S_i D_t$ and $S_t^{BEAR} = S_t(1 - D_t)$, where D_t is the same dummy variable that captures a bull market based on the signal of the contemporaneous return.[14] We then run one VAR model for each market state attention (S_t^{BULL} on R_t and S_t^{BEAR} on R_t). The results are presented in Table 13. For brevity, we only report the results for the influence of returns on attention since this is our focus in this subsection.

Table 13: VAR analysis between attention and return

	S_{t-1}	S_{t-2}	S_{t-3}	R_{t-1}	R_{t-2}	R_{t-3}	Intercept	adj. R^2
Bull Markets								
S_t	0.082	-0.065	-0.056	-0.148	0.399	0.454	-0.003	4.7
	(2.03)**	(-1.61)	(-1.37)	(-1.26)	(3.23)***	(3.84)***	(-0.61)	
Bear Markets								
S_t	-0.053	-0.022	-0.043	-0.340	-0.180	-0.161	0.006	1.7
	(-1.33)	(-0.55)	(-1.09)	(-2.76)***	(-1.37)	(-1.30)	(1.24)	

The table presents vector autoregressive analysis between investor attention (S) and gold returns (R). To capture the difference in the relation between bull and bear markets, we created a pair of series for investor attention, defined as $S_t^{BULL} = S_i D_t$ and $S_t^{BEAR} = S_t(1 - D_t)$, where D_t is a dummy variable that equals (1) if the return is positive (bull market), and 0 otherwise (bear market). For brevity, we only report the results for the influence of

[13]This process is very similar to the one employed in the seminal paper by Kristoufek (2013) to investigate the return-attention relationship of Bitcoin after positive and negative weekly returns.

[14]The results employing the four-week moving average definition are virtually the same and are available upon request.

returns on attention. The sample period is from 01.01.2006 to 12.31.2018. The numbers in parentheses are the t-statistics. The R^2 values are percentages. N = 624
***Significant at 1% and **5% levels.

During bull markets, there is a positive influence of lagged return on attention, indicating that individual investors are attracted by recent price increases. Something similar happens during bear markets, since the negative sign of the one-lag return indicates that a decrease in price increases on-line searches in the following week. A comparison of the return coefficients in the two regimes indicates that the influence of past returns on attention is more intense in bull markets, which is consistent with the tendency of individual investors to pay more attention to assets with positive recent returns, as documented by Barber and Odean (2008).

As expected, these results confirm that gold price past performance draws the attention of individual investors, as evidenced by the coefficients in Model (7). Overall, the evidence from Tables 12 and 13 indicates an interesting behavior of the return-attention relation during up and down markets. Past gold price performance draws the attention of investors and induces them to take advantage of the momentum; in the process, the investors magnify the trending movement of the gold price, since we document a positive (negative) relation between contemporaneous attention and return during bull (bear) markets. This dynamic lends support to the time-series momentum described by Moskowitz *et al.* (2012), which was also documented for the gold market (Baur *et al.*, 2020; Papailias *et al.*, 2021). Overall, the results in Table 13 confirm an asymmetric behavior of the attention-return connection during up and down markets. The next subsection explores if this asymmetry depends on the magnitude of gold returns.

3.3 Quantile regression of return on attention

The evidence so far suggests that the attention-return relation in the gold spot market is sensitive to market states, and that the association can be explained by recent gold price behavior. If this is the case, it is plausible to expect that the association between on-line searches and gold prices is conditional on the magnitude of price changes. A natural way to investigate this hypothesis is to employ a quantile regression of returns on attention to examine if changes in the relation depend on the return distribution. More precisely, we expect a stronger association between the variables when the magnitude of the return is pronounced (e.g., extreme quantiles). Table 14 presents the results of the quantile regression using Model (2) specification.

To avoid heteroskedastic and autocorrelation issues, the standard errors of the coefficients were obtained using the block-bootstrap procedure.

Table 14: Quantile regression of returns on attention.

	0.05	0.10	0.25	0.50	0.75	0.90	0.95
S_t	-0.067	-0.059	-0.030	-0.005	0.041	0.054	0.086
	(-4.25)***	(-3.71)***	(-4.95)***	(-0.30)	(2.74)***	(3.08)**	(1.97)**
S_{t-1}	-0.014	-0.015	0.003	0.003	0.013	0.018	0.021
	(-0.87)	(-1.09)	(0.28)	(0.33)	(1.12)	(2.14)**	(1.82)*
S_{t-2}	0.003	0.001	0.009	0.000	0.010	0.006	0.016
	(0.37)	(0.08)	(1.30)	(-0.01)	(1.5)	(0.72)	(1.11)
S_{t-3}	0.006	-0.010	-0.001	0.000	0.009	0.010	-0.006
	(0.37)	(-0.48)	(-0.09)	(0.01)	(1.16)	(0.69)	(-0.60)
R_{t-1}	-0.291	-0.201	-0.098	-0.052	-0.104	-0.210	-0.270
	(-1.88)*	(-1.98)**	(-1.93)*	(-1.25)	(-1.74)*	(-3.58)***	(-2.32)**
R_{t-2}	-0.140	-0.090	-0.033	-0.020	-0.035	-0.067	-0.110
	(-1.36)	(-1.13)	(-0.83)	(-0.45)	(-0.74)	(-1.49)	(-1.67)*
R_{t-3}	-0.069	-0.029	-0.039	-0.063	-0.044	-0.065	-0.052
	(-1.36)	(-0.68)	(-1.08)	(-0.77)	(-1.87)*	(-2.93)***	(-2.19)**
Intercept	-0.039	-0.029	-0.013	0.002	0.017	0.031	0.041
	(-10.95)***	(-13.84)***	(-11.47)***	(1.43)	(12.46)***	(15.70)***	(8.92)***
pseudo-R^2	10.4	7.1	3.4	0.6	2.7	7.1	11.7
# obs	624						

The table presents the estimators from the quantile regression of gold returns (R) on attention (S) using the following expression:

$$R_t = a_0 + \sum_{n=0}^{3} b_n S_{t-n} + \sum_{n=1}^{3} c_n R_{t-n} + \varepsilon_t \qquad (2)$$

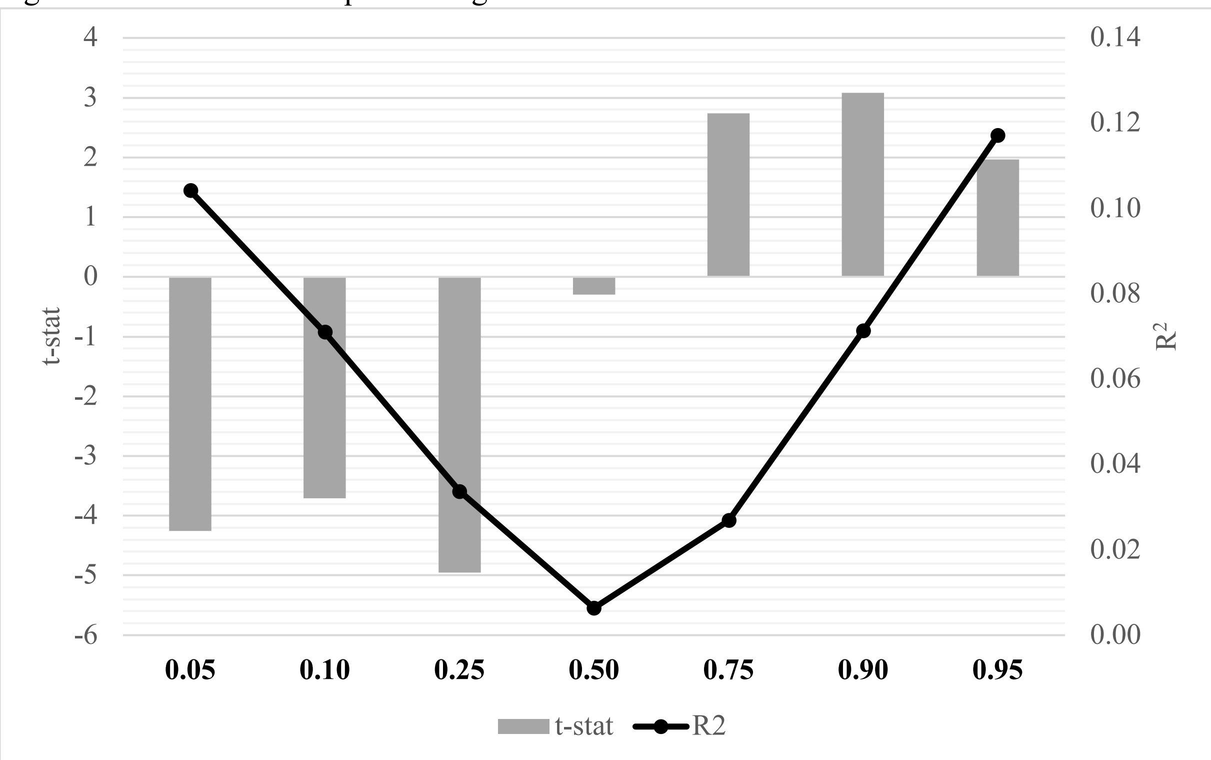
The corresponding quantile is identified in the first raw. For brevity, we report only the attention coefficients. The sample period is from 01.01.2006 to 12.31.2018. The numbers in parentheses are the block-bootstrap t-statistics. The pseudo-R^2 values are percentages. N = 624.
***Significant at 1% and **5% level.

The results indicate that the contemporaneous attention-return relation depends on the magnitude of the returns. While in regular circumstances (e.g., the 50% quantile) there is no association between the variables, during more volatile periods (lower and upper quantiles) we document a strong relation between searches and returns that is both economically and statistically significant. This pattern is consistent with the preference for skewness behavior (Barberis and Huang, 2008; Barberis and Xiong, 2012; Andrei and Hasler, 2015), since it indicates a stronger relationship between these constructs during more turbulent periods. Moreover, the sign of the coefficients indicates that the asymmetry of the return-attention relation documented in the previous subsections prevails, as it is always negative (positive) for the lower (higher) quantiles. Furthermore, it is worth noting that the explanatory power of attention on returns follows a U-shaped distribution, which is interesting since it indicates that investor attention is especially relevant in explaining prices precisely when this task is most challenging: in the tales.

It is also worth observing that the dependence of gold price on past returns is absent in the center quantile, and significant in remaining quantiles, especially during bull markets. This dependence behavior is consistent with the study of Baur *et al*. (2012), who document that stock returns exhibit an autocorrelation pattern in the tails, while no dependence is found in the central quantile[15].

To better illustrate the dependence of the attention-return relation on the return distribution, Figure 4 displays the t-statistics of the contemporaneous attention (columns) together with the R^2 of the corresponding quantile regression (right axis). The graph depicts a clear contrast in the relationship in bear and bull markets and its sensitivity to changes in the market states.

Figure 4: Estimates of the quantile regression of returns on attention.



Note: The figure plots the robust t-statistics (columns, left axis) of the contemporaneous attention from the quantile regression employing Model (2), together with the pseudo-R^2 (line, right axis) from the same model. The quantiles are represented on the horizontal axis.

[15] In their paper, Baur *et al*. (2012) report a negative dependence in upper quantiles, which is consistent to the pattern we document. However, the authors document a positive relation in lower quantiles, while we find a negative relation on the left tail, which indicates that the Gold market exhibits a peculiar behavior in terms of autocorrelation dependence during negative periods.

To that extent, the findings support the central hypothesis of the study, which posits that the association between on-line searches and gold spot prices depends on market states. To assess whether the results so far are driven by alternative explanations, the next section presents some robustness checks.

4. Robustness checks

4.1 On-line search terminologies

The previous findings were obtained using the term "Gold price" as a proxy for retail investor attention on the gold spot market. Notwithstanding this, it is reasonable to expect that some searches on the topic would employ alternative terminologies. Hence, to verify if the results previously reported could be driven by the choice of the proxy for investor attention, we downloaded the GSV for alternative specifications of on-line searches. More precisely, we employ two alternative terminologies, "Gold USD" and "Gold", with a filter to retain only the searches related to finance. We add to these alternative specifications the data from the terminology employed in Section 2 (i.e., Gold price); in the time series of these different terminologies, we run a Principal Component Analysis and extract the first component to capture the common factor among them.[16] The weekly log-difference in this factor is our alternative proxy for attention. Using this new proxy, we perform the same analysis as in Section 3. The results are presented in Table 15. For brevity, the table shows only the results for Model (5) in Panel A, and for the VAR analysis in Panel B.

[16]The factor analysis indicates only 1 factor, suggesting that these alternative terminologies are associated with one single factor that is closely related to the interest of retail investors in the gold spot market.

Table 15: Attention-return relation using alternative terminologies for on-line searches.

Panel A: OLS regression

$S_t.D_t$	$S_{t-1}.D_{t-1}$	$S_{t-2}.D_{t-2}$	$S_{t-3}.D_{t-3}$	$S_t.(1-D_t)$	$S_{t-1}.(1-D_{t-1})$	$S_{t-2}.(1-D_{t-2})$	$S_{t-3}.(1-D_{t-3})$	Intercept	Control	adj. R^2
0.091	0.036	0.008	0.015	-0.035	-0.020	-0.009	0.000	0.001	Yes	17.1
(3.03)***	(1.63)	(0.35)	(0.74)	(-1.67)*	(-1.12)	(-0.66)	(0.03)	(0.53)		

Panel B: VAR regression

	S_{t-1}	S_{t-2}	S_{t-3}	R_{t-1}	R_{t-2}	R_{t-3}	Intercept	adj. R^2
Bull Markets								
S_t	0.078	-0.038	-0.070	-0.221	0.378	0.408	0.002	4.5
	[0.040]	[0.040]	[0.040]*	[0.117]*	[0.124]***	[0.119]***	[0.004]*	
Bear Markets								
S_t	-0.089	-0.001	0.010	-0.348	-0.071	-0.214	0.001	2.5
	[0.040]**	[0.040]	[0.040]	[0.123]***	[0.131]	[0.123]*	[0.004]	

Panel A shows the results from the OLS regression of returns (R) on attention (S) for bull and bear markets. Attention is measured by the weekly log-difference of the first factor extracted from the Principal Component Analysis using three different terminologies of online searches: "Gold price", "Gold USD", and "Gold", with a filter to retain only the searches related to finance. The regression includes a dummy variable that equals 1 if the return is positive (bull market), and 0 otherwise (bear market). The control variables refer to the lagged returns (up to 3) to control for autocorrelation. The numbers in parentheses are the t-statistics from the Newey-West standard error estimator. Panel B shows the results from the VAR analysis between returns and attention during bull and bear markets. To capture the difference in the relation between the two market states, we created a pair of series for investor attention, defined as $S_t^{BULL} = S_i D_t$ and $S_t^{BEAR} = S_t(1 - D_t)$, where the dummy variable follows the previous definition. For brevity, we only report the results for the influence of returns on attention. The numbers in brackets are the standard errors. The sample period is from 01.01.2006 to 12.31.2018. The R^2 values are percentages. N = 624.

***Significant at 1%, **5% and *10% levels.

Overall, the results are consistent with the patterns previously documented: there is an asymmetric contemporaneous association between returns and attention during bear and bull markets, which is more pronounced in the latter market state. Furthermore, the VAR analysis confirms that the lagged return affects future attention in different ways depending on market states, showing a positive (negative) impact during bullish (bearish) conditions.

4.2 Subsample analysis

Subsample analysis is widely used in robustness checks to establish whether documented results are not driven by specific events. In the present case, this procedure is especially interesting since the split of the dataset into two periods creates two subsamples whose behavior coincides with that of a bull and a bear market, respectively. More precisely, the first subsample dates from 01.01.2007 to 12.31.2012, a wide bull market during which the gold spot price rose from 626 USD to 1,655 USD, an annual increase of 18%, on average; by contrast, in the second subsample (01.01.2013 to 12.31.2018), the price fell from 1,663 USD to 1,284 USD. Hence, this division creates a pair of subsamples that fits as natural experiments for the main purpose of the study: investigating the attention-return relation in the gold market conditional on market states. Specifically, employing the single-regime models (i.e., Models (1) and (2)) on each subsample, we would expect a positive (negative) relation between attention and contemporaneous return during the first (second) subsample. The results of this analysis are presented in Table 16.

Table 16: OLS analysis between return and attention on two subsamples

	S_t	S_{t-1}	S_{t-2}	S_{t-3}	R_{t-1}	R_{t-2}	R_{t-3}	Intercept	adj. R^2
Panel A: Subsample 2007-2012 (N = 312)									
(1)	0.079	-0.014	0.027	-0.001				0.003	7.2
	(1.82)*	(-0.65)	(1.11)	(-0.06)					
(2)	0.071	0.012	0.029	0.013	-0.385	-0.155	-0.090	0.005	18.5
	(2.12)**	(0.68)	(1.23)	(0.69)	(-3.04)***	(-1.44)	(-1.70)*	(1.80)*	
Panel B: Subsample 2013-2018 (N = 312)									
(1)	-0.033	0.002	0.002	0.011				-0.001	2.3
	(-4.18)***	(0.21)	(0.33)	(1.13)				(-0.62)	
(2)	-0.040	-0.013	-0.007	0.006	-0.299	-0.157	-0.080	-0.001	9.8
	(-4.43)***	(-0.95)	(-0.75)	(0.70)	(-1.49)	(-1.56)	(-1.06)	(-0.74)	

The table presents the cross-sectional relation between returns (R) and attention (S) for two subsamples, using the following models:

$$R_t = a_0 + \sum_{n=0}^{3} b_n S_{t-n} + \varepsilon_t \qquad (1)$$

$$R_t = a_0 + \sum_{n=0}^{3} b_n S_{t-n} + \sum_{n=1}^{3} c_n R_{t-n} + \varepsilon_t \qquad (2)$$

The first (second) subsample period is from 01.01.2006 to 12.31.2012 (01.01.2013 to 12.13.2018). The numbers in parentheses are the t-statistics from the Newey-West standard error estimator. The adjusted R^2 values are percentages.
***Significant at 1%, **5% and *10% levels

As expected, for both models, we find a contemporaneous positive attention-return relation for the first subsample (bull market) that becomes negative during the second (bear market). Additionally, we observe that the magnitude of the relation is consistent with the previous findings since it is economically stronger during the bull market (first subsample). Curiously, this association is statistically more pronounced in the second subsample. Although this fact may be counterintuitive at first glance, we argue that this behavior is rational, considering that gold is seen as a safe asset: a bull market for this commodity is usually associated with turbulent circumstances in the wider financial markets, which leads to higher volatility and, consequently, higher standard errors of the coefficient, undermining its statistical significance. Indeed, the variance of the returns in the first subsample is more than double that of the second,[17] which is expected since the former comprises both the subprime and the eurozone crises. Finally, it is notable that, in each subsample, there are returns whose signs contradict a granular definition of up and down markets (i.e., negative returns during the first subsample and positive in the second), lending bias to the results against our central hypothesis. That the results, despite the bias, still confirm our main findings, corroborates the view that market states play an important role in the attention-return relation.

5. Conclusion

This study investigates the attention-return relation in the gold spot market and its dependence on market states. Using GSV as a proxy for investor attention, we document a positive (negative) contemporaneous association between on-line searches and gold returns during bull (bear) markets, defined as the weeks during which the return is positive (negative). Additionally, we find that retail investors' attention is drawn by recent gold price trends, since we identify that lagged returns and future attention are proportionally related in both market states, suggesting that a price increase (decrease) attracts on-line searches. Overall, these findings indicate that gold price past performance stimulates retail investors' trading activity for profit, reinforcing the market state, consistent with the time-series momentum literature (Moskowitz *et al*., 2012; Papailias *et al*., 2021). Finally, we further report that this association is sensitive to the magnitude of the changes in price, since the quantile regression analysis shows that the attention-return relation is absent near the mean, monotonically rising when the return distribution approaches the tails. This pattern indicates that investor attention is more

[17] The variance for the first subsample is 2.01×10^{-3}, whereas it is 0.93×10^{-3} for the second.

intrinsically related to price changes during turbulent markets, which can be explained by a preference for skewness that has been observed in individual investors (Barberis and Huang, 2008; Barberis and Xiong, 2012; Andrei and Hasler, 2015).

References

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

Barber, B., 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(2), 785-818

Barberis, N., Huang, M., 2008. Stock as lotteries: The implication of probability weighting for security prices. *American Economic Review,* 98 (5), 2066-2100

Barberis, N., Xiong, W., 2012. Realization utility. *Journal of Financial Economics*, 104, 251-271

Baur, D., Dimpfl, T., Jung, R., 2012. Stock return autocorrelations revisited: A quantile regression approach. *Journal of Empirical Finance*, 19(2), 254-265

Baur, D., Dichtl, H., Drobetz, W., Wendt, V., 2020. Investing in gold – Market timing or buy-and-hold. *International Review of Financial Analysis*, 71,101281

Baur, D., Dimpfl, T., 2016. Googling gold and mining bad news. *Resources Policy*, 50, 306-311

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

Buchinsky, M., 1995. Estimating the asymptotic covariance matrix for quantile regression models a Monte Carlo study. *Journal of Econometrics*, 68, 303-338

Chen, Z., 2020. Investor attention and market correction. *Review of Behavioral Finance*. Vol. ahead-of-print No. ahead-of-print

Dastgir, S., Demir, E., Downing, G., Gozgor, G., Lau, C., 2019. The causal relationship between Bitcoin attention and Bitcoin returns: Evidence from the Copula-based Granger causality test. *Finance Research Letters,* 28, 160-164

Eom, C., Kaizoji, T., Kang, S., Pichl, L., 2019. Bitcoin and investor sentiment: Statistical characteristics and predictability. *Physica A*, 514, 511-521

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

Han, L., Li, Z., Yin, L., 2017. The effects of investor attention on commodity futures markets. *Journal of Futures Markets*, 37, 1031-1049

Han, L., Xu, Y., Yin, L., 2017. Does investor attention matter? The attention-return relation in gold futures market. *Economics Discussion Papers* # 2017-37

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

Jain, A., Biswal, P., 2019. Does internet search interest for gold move the gold spot, stock and exchange rate markets? A study from India. *Resources Policy,* 61, 501-507

Joseph, K., Wintoki, M., Zhang, Z., 2011. Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, 27, 1116-1127

Kristoufek, L., 2013. BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, 3, 3415

Kumar, A., 2009. Who gambles in the stock market? *Journal of Finance*, 64, 1889-1933

Moskowitz, T., Ooi, Y., Pedersen, L., 2012. Time series momentum. *Journal of Financial Economics*, 104, 228-250.

Papailias, F., Liu, J., Thomakos, D., 2021. Return signal momentum. *Journal of Banking & Finance*, 124, 106063

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

Wu, Y., Han, L., Yin, L., 2019. Our currency, your attention: Contagion spillovers of investor attention on currency returns. *Economic Modelling,* 80, 49-61

Yao, T., Zhang, Y., Ma, C., 2017. How does investor attention affect international crude oil prices? *Applied Energy*, 205, 336-344

Zhang, W., Wang, P., Li, X., Shen, D., 2018. Quantifying the cross-correlations between online searches and Bitcoin market. *Physica A*, 509, 657-672

Appendix A:

Table A: GSV of the window on the GSV of the reference window
The table displays the parameters of the regression of the weekly GSV of the dependent window in the weekly GSV of the reference window during the respective overlaps. Hence, the overlaps of Window 1 and Window 3 are regressed in the overlaps of Window 2 The b coefficients are used to generate the GSV of the given window relative to the GSV of the Reference Window. For instance, a GSV = 80 in Window 1 is resclaed to 37.72 in the Reference Window ($GSV_{Relative} = 80/2.121$). The datasets of the subsamples are as follows: Window 1: 01.01.2007 to 12.31.2011; Window 2: 01.01.2011 to 12.31.2015; Window 3: 01.01.2015 to 12.31.2018; The coefficients are followed by the t-statistic in parenthesis.

Window 1			Window 3		
a	b	R^2	a	b	R^2
0.218* (0.64)	2.121 (127.92)***	99.7	0.011 (0.03)	1.729 (110.91)***	99.6

***Significant at 1%, **5% and *10% levels

ARTICLE 3

Where are retail investors looking?

ABSTRACT: It is believed that individual investors trade based on noise. Based on this assumption, this article investigates whether noisy variables, such as the Relative Strength Index, Momentum, and Stochastic Index, attract more attention from these investors than value multiples, such as the price-earnings ratio, price-to-sales ratio and company value-EBITDA. The results suggest that value dynamics are more important when it comes to explaining changes in attention than noise variables, since fundamentalist indicators attract the attention of retail investors, which does not occur with indicators of a technical nature. Furthermore, when analyzing the characteristics of the assets that most attract investors' attention, we identified the following as significant: company size, the price-earnings ratio, the multiple EV-EBITDA and the company's beta. The positive sign displayed by the coefficients of these parameters suggests that larger, more expensive and riskier companies are more attractive to individual investors.

KEYWORDS: Investor Attention; Asset Characteristics; Technical Analysis; Capital Market.

1 Introduction

Investors in general have access to an increasingly wider range of information from a broad range of sources. Nevertheless, the literature has repeatedly shown that people experience a certain level of difficulty with regard to processing all the information that is available due to their limited cognitive resources (Kahneman, 1973). In addition, although a great deal of information is available, investors may not be able to pay attention and select really important information that enables them to make adequate investment decisions and consequently develop their portfolios.

Individual investors are often characterized as traders who depend on less reliable information due to having less experience and knowledge. This implies that they might make investment decisions based on less accurate information, which can be considered noise in the context of the financial market (Black, 1986; De Long *et al*., 1990, Shleifer and Summers, 1990). In this context, Sornette (2003) argues that the actions of noise traders can cause a drastic financial breakdown, as they act based on very little information.

However, it is important to note that this generalized description does not necessarily apply to all individual investors, as some may take a more sophisticated approach and are grounded in solid analytics. Nowadays, individual investors have access to a significant amount of financial information and analysis tools through online platforms, social media networks, specialist websites and brokerage services. This has improved the ability of individual investors to make more informed decisions and keep up with market trends (Farrel *et al*., 2022; Welch, 2022). In this respect, Chau *et al*. (2016), provide evidence that investors are better adapted to trades based on information garnered from research rather than market-based sentiment indicators. Choi *et al*. (2015) find that individual investors tend to act contrary to the market, while institutional investors follow the trends. Other studies, in turn, have demonstrated that individual investors improve market efficiency (Kelley and Tetlock, 2013; Bohemer *et al*., 2021).

Despite the extensive literature on the effects of investor attention, information regarding the characteristics that attract the attention of individual investors who choose to trade assets remains relatively undocumented. Thus, a gap remains concerning what actually attracts investors' attention to the different types of information that are available to them. As a considerable amount of information is at their disposal and investors do not have the capacity or time to absorb and analyze all the nuances at play (Barber and Odean, 2008), our intention is to attempt to identify, among the main available indicators, which of them attract the most attention from investors and define their options for trading assets.

The inclusion and use of the indicators selected for our analysis enables the identification of the characteristics that most attract the attention of individual investors. Each indicator provides specific information about different aspects of a company's performance and value, as well as market behavior. In our analysis, specifically, these indicators are used to identify whether noise information (i.e., from technical analysis) draws more attention from these investors than value variables (i.e., from fundamental analysis). As a proxy for individual investor attention, we use Google Search Volumes (GSV), adopting the tickers of the companies that make up the IBRX100 Index in the period spanning 2014 to 2021 as the search terms. The variables used to measure the noise are related to the technical analysis indicators, of which we used the Relative Strength Index, Moving Average Oscillator, high (low) momentum trends, Stochastic Index, and Bollinger Bands, all of which have been used in previous studies (Coalkley *et al*., 2016; Ni *et al*., 2020; Novy-Marx, R., 2012; Zhu and Zhou, 2009). The second set of variables is derived from the fundamental analysis and comprised the following indicators: Price-Earnings, Price-to-Sales Ratio, Enterprise Value-EBITDA (EV-EBITDA), Price-Book Value, Cash Flow, risk measure (beta), and Dividend Yield (Almanso *et al*.. 2022; Akhtar, 2021; Semenov, 2021; Baker *et al*., 2020; Igrejas *et al*., 2017). All the indicators were also obtained for the companies of the IBRX100 Index for the entire sample period.

In the first empirical analysis, we regress investor attention on the aforementioned variables. These results indicate that the fundamental variables of Price-Earnings, Price-to-Sales Ratio and EV-EBITDA have significant effects on investor attention. It is interesting to note that for Price-Earnings and EV-EBITDA the coefficients are positive, indicating that as a company's value rises, greater attention is paid by investors, suggesting a trend behavior on the part of these agents, as attention grows as assets become more expensive. Furthermore, the results of the regression analysis show that the other variables, including the three technical analysis variables, have no significant effects on investor attention. In other words, among the fundamental and technical variables analyzed, the Price-Earnings and EV-EBITDA indicators seem to be the most important with regard to attracting investors' attention. The sensitivity analysis conducted using univariate regression for each of the variables in question corroborates this finding by revealing that the dynamics of the fundamentals remain more important for investment decisions than the dynamics of the noise variables.

The second empirical analysis sought to identify the characteristics of stocks that most attract the attention of individual investors. For this purpose, we construct a dummy variable based on the accumulated attention over the last 6 months, using the accumulated GSV of individual assets. The stocks were then ordered based on the accumulated GSV, and those belonging to the upper (lower) quartile were classified as high (low) attention, being assigned a value of 1(0). The dummy variable was then regressed against the aforementioned set of variables. The results indicate that company size is important to investors. This is because the variable has a positive and significant coefficient, that is, larger companies tend to attract more investor attention. Regarding Price-Earnings and EV-EBITDA, it was observed that investors pay closer attention to assets whose Price-Earnings and EV-EBITDA ratio are higher. In addition, the analysis also reveals that investors have a preference for risk, as evidenced by the positive coefficient of the BETA variable. This result is in keeping with studies that support the idea that investors have risk asymmetry (Barberies and Huang, 2008; Kumar, 2009; Bali *et al.*, 2010, Albuquerque, 2012; Amaya *et al.*,2012; Andrei and Hasler, 2015; Mitton and Vorkink 2016).

Other characteristics of the group of assets that drew the most attention during the period under study are related to variables of a technical nature, such as the Relative Strength Index and Moving Averages. In this case, a negative coefficient was observed, indicating that assets with a low appreciation trend are more attractive to investors, as noted by Chau *et al.* (2016), for whom investors influenced by sentiment are not irrational, as they trade against consensus, buying stocks during bear markets and selling during bull markets.

This article is of relevance to the literature when we consider the counterintuitive nature of the empirical evidence reported here, supporting the assumption that individual investors make efficient investment decisions (Kelley and Tetlock, 2013; Chen *et al.*, 2016; Dahlquist *et al.*, 2017), moving away from the main theoretical current that assumes that retail investors are noise traders (Black, 1986; Shelifer and Summers, 1990, De Long *et al.*, 1990). Moreover, by documenting the characteristics of stocks that most attract the attention of investors, we corroborate the proposition that retail investors are not irrational, as they trade by buying in bear markets and selling in bull markets (Chau *et al*. 2016).

The remainder of this article is organized as follows. Section 2 describes the data and methodology used to answer the research question. The results are presented and discussed in Section 3, and the conclusions are found in Section 4.

2 Data and methodology

2.1 Data

The sample of this study is composed of those that make up the IBRX100 index in the Brazilian stock market, covering the period from 01.2014 to 12.2021. For these assets, the following indicators were obtained monthly through the Economática database:

a) Price-Earnings = Market price per share/Earnings per share. The result of the PE calculation indicates how many times the earnings per share are contained in the market price of the share.
b) Price-to-Sales Ratio = Share market price/Sales revenue. The PSR indicator, also known as the price/sales ratio, is a financial indicator that measures the relationship between a company’s market price and its sales revenue. It is used to assess the relative value of a company in relation to its revenue generation.
c) EV-EBITDA = Enterprise Value (EV)/EBITDA. EV/EBITDA is used to assess a company’s value in relation to its operating performance, regardless of capital structure and tax policies. It provides a view of the relationship between a company’s market value and its ability to generate earnings before financial and non-operating items.
d) Price-Book Value = Market price per share/Book value per share. The P/Book Value (PBV) indicator is a financial indicator that measures the relationship between the market price of a share and the book value per share of the company. It is used to assess whether a stock is trading at a premium or discount to its book value.
e) Cash Flow = Market price per share/Free cash flow per share. The P/Cash Flow (P/FCF) ratio is a financial ratio that measures the relationship between the market

price of a stock and the company's free cash flow per share. It is used to assess the relative value of a stock in relation to free cash flow generation.

f) Risk measure (beta) – The beta is calculated through statistical analyses, comparing the historical movement of prices of an asset in relation to the movement of the reference market, usually represented by a market index. Beta is used to help investors understand an asset's risk relative to the market.

g) Dividend Yield = (Dividends per share/Market price per share) * 100. The Dividend Yield is a financial indicator that measures the relationship between the dividends distributed by a company and the market price of its shares. It is used to assess the return provided to shareholders through dividend payments.

h) Relative Strength Index – The RSI is a technical indicator used to assess the speed and magnitude of changes in the price of a financial asset. The RSI is a momentum indicator that seeks to identify overbought and oversold conditions of an asset.

i) Moving Average Oscillator = Short-Term Moving Average - Long-Term Moving Average. It is a technical indicator used to identify the strength and direction of a price trend. It is calculated from a short-term Simple Moving Average (SMA) and a long-term Simple Moving Average.

j) Momentum = Current Price - Previous Price. Momentum is a financial indicator that measures the rate of change in the price of an asset over a given period. It is used in technical market analysis to identify the strength and direction of a price trend.

k) Stochastic Index – The stochastic indicator, or Stochastic Oscillator, is a technical indicator used in chart analysis and in predicting price movements in financial markets, developed by George Lane in the 1950s. Stochastic compares the current price of an asset with its price range over a given period of time. The purpose is to provide information on the strength and direction of a trend, as well as to identify overbought and oversold conditions.

l) Bollinger Bands - consist of three lines on the price chart: a central line and two outer bands. The central line is a simple moving average (usually of 20 periods) that represents the price trend over a given period. Outer bands are lines drawn above and below the central line, which are calculated based on the standard deviation of prices from the moving average.

m) Company size – The company's size is calculated using the natural logarithm of the Market Value.

n) Trading volume – The company's liquidity is calculated based on the natural logarithm of the average trading volume.

As a proxy for investor attention, we used the monthly Google Search Volume, collected from the Google Trends platform (www.google.com/googletrends), using the ticker of each company in the sample. As the time series of Google Trends are not stationary and are available on a relativized scale, with the highest search frequency represented by 100 and the other frequencies defined proportionally to this peak, when using the attention variable (S), we used the logarithmic difference of consecutive GSVs.

2.2 Econometric Models

In the first empirical analysis, we attempt to identify which variables related to fundamental and technical analysis have a significant impact on investor attention. To this end, we use panel data analysis with fixed effects, as shown in Equation (1):

$$S_{i,t} = a_{\mathrm{i}} + a_{\mathrm{t}} + b_n.PE_{it} + c_n.PSR_{it} + d_n.\mathrm{E}VEB_{it} + e_n.P\mathrm{BV}_{it} + f_n.\mathrm{CF}_{it} + g_n.RSI_{it} + h_n.MA_{it} + i_n.STO\mathrm{C}_{it} + \mathrm{j}_n.\mathrm{Dyear}_{it} + \varepsilon_{\mathrm{i}t} \quad (1)$$

Where $S_{i,t}$ refers to the variation of GSV of Google Trends data, obtained through the terms of each ticket of the stocks that comprise the IBRX100 index. The independent variables are Price-Earnings "PE_{it}", Price Sales Ratio "PSR_{it}", EV/EBITDA "$\mathrm{E}VEB_{it}$", Price/Book Value "PBV_{it}", Cash Flow "CF_{it}", Relative Strength Index "RSI_{it}", Moving Average Oscillator "MM_{it}", and Stochastic Index "STOC_{it}". "Dyear_{it}" *refers to a dummy variable for year, with 1 for each year and 0 otherwise. Finally, the parameter* a_{i} is the individual fixed-effect term used to control for idiosyncratic characteristics of companies, a_{t} *is the time fixed effect term,* while $\varepsilon_{\mathrm{i}t}$ is the random error term.

The second empirical test then sought to explain the characteristics of assets that most attract the attention of investors. For this purpose, logistic regression was used with fixed effects and panel data, as described in Equation (2):

$$
\begin{aligned}
D_{i,t} = a_{\mathrm{i}} + a_{\mathrm{t}} + b_n.SIZE_{it} + c_n.VOL_{it} + d_n.PE_{it} + e_n.PSR_{it} + f_n.EVEB_{it} \\
+ g_n.PBV_{it} + h_n.BETA_{it} + i_n.DY_{it} + j_n.RSI_{it} + k_n.MA_{it} \\
+ l_n.MOM_{it} + m_n.BB_{it} + n_n.STOC_{it} + \varepsilon_{it}
\end{aligned} \tag{2}
$$

Where $D_{i,t}$ is a dummy variable that assumes a value of 1 when assets attract attention and 0 otherwise. This variable was constructed based on attention over the last six months from the accumulation of GSV of each asset. Shares were then ordered based on the accumulated GSV, and those belonging to the highest (lowest) quartile were classified as high (low) attention, being assigned a value of 1(0). It should be noted that the data provided by the Google Trend platform are presented on a relativized scale, with the peak of attention equal to 100 and the other search frequencies proportional to this. For this reason, we chose to use the accumulated GSV to avoid possible bias resulting from news that could cause a spike, and not necessarily a company characteristic that might have drawn more attention from investors.

The independent variables refer to company size “$SIZE_{it}$”, volume of trades “VOL_{it}”, Price/Earnings “PE_{it}”, Price Sales Ratio “PSR_{it}”, Enterprise Value/EBITDA “$EVEB_{it}$”, Price-Book Value “PBV_{it}”, risk measure “$BETA_{it}$”, Dividend Yield “DY_{it}”, Relative Strength Index “RSI_{it}”, Moving Average Oscillator “MA_{it}”, high (low) momentum trend “MOM_{it}”, Stochastic Index “$STOc_{it}$”, and Bollinger Bands “BB_{it}”.

2.3 Descriptive Statistics

To illustrate some characteristics of the main variables better, the descriptive statistics for the data series are presented in Table 17:

Table 17: Descriptive statistics

	Mean	Standard Deviation	Quartile 1	Median	Quartile 3	Asymmetry	Kurtosis

PE	19.514	69.255	6.100	12.700	23.600	3.443	27.926
PSR	2.427	2.975	0.600	1.400	3.000	2.418	9.268
EVEB	10.413	13.367	6.000	8.600	12.600	1.504	14.671
PBV	2.775	2.990	1.000	1.800	3.300	2.463	10.612
CF	10.732	123.221	-3.000	9.000	22.000	0.413	21.226
BETA	0.949	0.522	0.580	0.870	1.220	0.690	3.195
DY	3.036	3.197	0.500	2.240	4.440	1.561	5.794
RSI	56.757	20.911	41.995	56.815	71.965	-0.062	2.372
MA	1.888	3.807	0.010	1.450	3.530	0.342	4.896
MOM	1.567	6.258	-1.230	1.105	4.160	0.189	5.830
STOC	54.244	29.955	28.260	56.480	81.780	-0.171	1.752
BB	15.739	10.695	8.030	13.330	20.730	1.252	4.666
SIZE	16.468	1.424	15.600	16.430	17.300	-0.093	3.324
VOL	13.806	1.787	13.142	14.076	14.878	-1.540	6.769

The table presents the results for the descriptive statistics of the variables Price/Earnings “PE”, Price Sales Ratio “PSR”, Enterpriseompany Value/EBITDA “EVEB”, Price/Book Value “PBV”, Cash Flow “CF”, risk measure “BETA”, Dividend Yield “DY”, Relative Strength Index “RSI”, Moving Average Oscillator “MA”, high (bearish) momentum trend “MOM”, Stochastic Index “STOC”, Bollinger Bands “BB ”, company size “SIZE”, Trading volume “VOL”. Information is displayed for the entire dataset (from 01.2014 to 12.2021).

We observe highly dispersed distributions for the variables, since the standard deviation is somewhat high in relation to the mean for most variables. This behavior is also shown in the kurtosis analysis, which presented high values, mainly for value variables.

As a way of adjusting the characteristics of the data observed in the initial tests, with the aim of inhibiting possible biases arising from extreme outliers, we use the winsorization technique, limiting the extreme values to the margins of 1 and 99%. Thus, the values shown in Table 17 are already presented in winsorized form.

Before proceeding to the regression models, we ran a model to identify possible variables with a high degree of correlation. Table 18 presents the correlation test for the data series (Price-Earnings “PE”, Price-to-Sales Ratio “PSR”, Enterprise Value-EBITDA “EVEB”, Price-Book Value “PBV”, Cash Flow Box “CF”, risk measure “BETA”, Dividend Yield “DY”, Relative Strength Index “RSI”, Moving Average Oscillator “MA”, high momentum trend “MOM”, Stochastic Index “STOC”, Bollinger Bands “BB”, company size “SIZE”, and Trading volume “VOL”) for the entire sample period (2014-2021).

Table 18: Correlation between the independent variables.

	PE	PSR	EVEB	PBV	CF	BETA	DY	RSI	MA	MOM	STOC	BB	SIZE	VOL
PE	1.000													
PSR	0.079	1.000												
EVEB	0.102	0.224	1.000											
PBV	0.152	0.159	0.279	1.000										
CF	0.049	-0.032	0.016	0.061	1.000									
BETA	-0.068	-0.113	-0.033	-0.228	-0.021	1.000								
DY	-0.049	0.109	-0.172	-0.146	-0.047	-0.213	1.000							
RSI	0.077	0.108	0.074	0.182	-0.012	-0.055	-0.095	1.000						
MA	0.066	0.125	0.218	0.224	0.027	-0.041	0.000	0.258	1.000					
MOM	0.041	0.105	0.140	0.224	0.007	-0.044	-0.072	**0.742**	0.491	1.000				
STOC	0.055	0.096	0.081	0.127	-0.017	-0.055	-0.083	**0.655**	0.047	0.471	1.000			
BB	-0.012	0.060	0.101	0.121	0.026	-0.190	0.078	-0.176	0.325	-0.023	-0.145	1.000		
SIZE	0.075	0.249	0.107	0.170	0.004	-0.063	0.107	0.097	0.246	0.175	0.092	0.366	1.000	
VOL	0.072	0.235	0.129	0.187	0.012	0.194	0.007	-0.016	0.196	0.082	-0.019	0.312	**0.654**	1

The Table presents the results referring to the correlation test for the variables Price/Earnings “PE”, Price Sales Ratio “PSR”, Company Value/EBITDA “CVEB”, Price/Book Value “PBV”, Cash Flow “CF” , risk measure “BETA”, Dividend Yield “DY”, Relative Strength Index “RSI”, Moving Average Oscillator “MA”, HIGH (bearish) momentum “MOM”, Stochastic Index “STOC”, Bollinger Bands “ BB”, company size “SIZE”, and Trading volume “VOL”). Information is displayed for the entire dataset (from 01.2014 to 12.2021).

Although we can identify in the above table a higher correlation between the variables of Momentum, Stochastic Index and Relative Strength Index, as well as for the Size and Volume variables (values in bold), for the VIF test, the highest value found was 3.68, not characterizing significant multicollinearity between the variables.

3 Analysis of The Results

3.1 Relationship between investor attention and fundamental and technical variables.

We begin the empirical analysis by identifying which variables related to fundamental and technical analysis have a significant impact on investor attention, as described in Equation (1). The results are summarized in Table 19, along with a presentation of the single-factor models in order to perform a sensitivity analysis that prevents the results from being biased by the way the econometric model was specified

Table 19: Relationship between investor attention and fundamental and technical variables

S	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
PE	**0.004**	**0.002**							
	(5.38)***	(4.32)***							
PSR	**-0.129**		-0.005						
	(-1.68)*		(-0.10)						
EVEB	**0.018**			**0.011**					
	(2.37)***			(2.20)***					
PBV	0.062				**0.062**				
	(1.44)				(2.39)***				
CF	0.000					0.000			
	(0.38)					(0.35)			
RSI	-0.002						0.000		
	(-0.77)						(-0.12)		
STOC	0.001							0.001	
	(0.54)							(0.85)	
MA	-0.055								-0.050
	(-1.35)								(-1.42)
Intercept	**0.140**	**0.147**	**0.152**	**0.151**	**0.148**	**0.151**	**0.153**	**0.149**	**0.133**
	(6.47)***	(9.90)***	(9.01)***	(9.06)***	(10.06)***	(9.02)***	(10.21)***	(10.26)***	(7.12)***
Firm FE	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	4151	6889	6003	6095	6997	6022	6487	6717	4901
R2	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

The table presents the results of the regression between the value and technical variables and the respective changes in Google Trends. Data from Google Trends for construction S refer to the variation of the GSV obtained through the terms of each ticket of the shares that make up the IBRX index. The independent variables refer to Price/Earnings “PE”, Price Sales Ratio “PSR”, EnterpriseValue/EBITDA “EVEB”, Price/Book Value “PBV”, Cash Flow “CF”, Relative Strength Index “IFR ”, Moving Average Oscillator “MA”, and Stochastic Index “STOC”. The results are displayed for the entire dataset (from 01.2014 to 12.2021). The t-statistics are below the coefficients. ***, **, * significant at 1%, 5% and 10% respectively.

Based on the results of the analysis (Table 19, Column 1), it was found that the fundamental variables Price-Earnings, Price-to-Sales Ratio and Enterprise Value-EBITDA have significant effects on investor attention. That is, these variables are important in the analysis of individual investors and can significantly affect their attention. It is interesting to note that for the Price-Earnings and Company Value-EBITDA variables, the coefficients are positive, indicating that as the company's value rises, more attention is paid by investors. This demonstrates that investors are optimistic with regard to the company's future performance, suggesting greater willingness to trade assets as they become more expensive, which indicates a behavior of following market trends. As for the PSR indicator, the coefficient is negative, suggesting that attention is greater when revenues fall. A lower PSR may indicate a possible undervaluation of the company, which could represent an investment opportunity. However, it is noteworthy that this relationship is marginally significant, not being observed in the corresponding single-factor model (Column 3), indicating that the significance of this association should be interpreted with caution.

Additionally, the results of the regression analysis show that the other variables, including the three technical analysis variables, do not have significant effects on investor attention. In other words, of the fundamental and technical variables analyzed, the Price-Earnings and Enterprise Value-EBITDA indicators seem to be the most important in terms of attracting the attention of investors.

Columns (2) to (9) contain the results of the regressions of each variable tested individually in order to assess whether the results obtained in the previous test (Column 1) are not biased by a possible relationship of the set of variables. We observe that the parameters of the Price-Earnings and Company Value-EBITDA variables remain significant, reinforcing the assumption that value measures draw more attention than technical measures, that is, that the dynamics of fundamentals are more important for investment decisions than noise (Choi *et al.*, 2015). Likewise, corroborating the previous findings, none of the technical variables proved to be significant when analyzed individually (Columns 7 to 9).

The findings reported here are in keeping with a recent stream of financial economic reports that retail traders are not a consistent proxy for naive investors (Choi *et al.*, 2015; Chau *et al.*, 2016; Kelley and Tetlock, 2017; Herve *et al.*, 2019).

3.2 Characteristics of the assets that most attract the attention of individual investors

In this subsection, we seek to identify the characteristics of the assets that most attract the attention of investors, in accordance with the process detailed in Subsection 2.2. To this end, we teste our objective by analyzing the logistic regression results that explain the characteristics of the assets that most attract investors' attention.

Table 20: Technical and fundamental variables that determine investor attention

D	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
SIZE	**0.889**	**0.750**												
	(2.81)***	(8.62)***												
VOL	**1.460**		**0.000**											
	(7.72)***		(5.01)***											
PE	**0.003**			**0.002**										
	(2)**			(2.63)***										
PSR	-0.056				**0.183**									
	(-0.54)				(3.44)***									
EVEB	**0.020**					**0.015**								
	(2.33)**					(3.52)***								
PBV	0.01						-0.022							
	(0.22)						(-0.86)							
BETA	**1.698**							**1.458**						
	(3.68)***							(6.53)***						
DY	0.013								-0.007					
	(0.35)								(-0.33)					
RSI	**-0.015**									**-0.014**				
	(-1.91)**									(-5.06)***				
MA	**-0.148**										**-0.083**			
	(-4.18)***										(-3.92)***			
MOM	-0.009											**-0.053**		
	(-0.4)											(-5.46)***		
BB	**0.069**												**0.090**	
	(3.2)***												(8.62)***	
STOC	**0.008**													**-0.005**

	(2.29)**													(-2.97)***
Firm FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Year FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Sector FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
N	1514	2870	2736	2557	2311	2470	2673	2338	2370	2520	1804	2428	2107	2618

The table presents the results of the logistic regression between the presented variables and the attention of the set of companies that most attracted investors' attention, "d" is a dummy variable that assumes a value of 1 when an asset attracts attention and 0 otherwise. The independent variables are company size "SIZE", Trading volume "VOL", Price/Earnings "PE", Price Sales Ratio "PSR", Enterprise Value/EBITDA "EVEB", Price/Book Value "PBV", measure "BETA", Dividend Yield "DY", Relative Strength Index "RSI", Moving Average Oscillator "MA", high (bearish) momentum trend "MOM", Stochastic Index "STOC", and Bollinger Bands "BB" . Results are displayed for the entire dataset (from 01.2014 to 12.2021). The t-statistics are below the coefficients. ***, **, * significant at 1%, 5% and 10%, respectively.

An analysis of Table 20 clearly shows that diverse variables are significant with regard to determining investor attention. Among these variables, we can highlight company size, volume, Price-Earnings, Enterprise Value-EBITDA, Beta, the Relative Strength Index, Moving Averages and Bollinger Bands. The size variable has a positive and significant coefficient, indicating that larger companies tend to attract greater investor attention. Volume is also highly significant and positive, indicating that investors have preferences for companies with greater liquidity. In relation to Price-Earnings and Enterprise Value-EBITDA, it was observed that investors pay more attention to assets with a higher Price-Earnings ratio and Enterprise Value-EBITDA, suggesting that they are willing to pay a premium for the company's profit or profit potential. In addition, the analysis also reveals that investors have a preference for risk, as evidenced by the positive coefficient of the Beta variable. This result is in line with several studies supporting the idea that investors prefer riskier assets. (Barberies and Huang, 2008; Kumar, 2009; Andrei and Hasler, 2015).

Other variables that attract investors' attention are those of a technical nature, such as the RSI and MA. In this case, a negative coefficient was observed, which indicates that assets with a tendency to depreciate are more attractive to investors.

In short, an analysis of Table 4 allows us to conclude that investors pay attention to several variables when making their investment decisions. Some factors such as Company Size, Price-Earnings and Company Value-EBITDA, in addition to preferences for risk and appreciation trends, are among the elements that influence investors' choices.

The results in Table 4, Columns (2) to (14), show evidence that most asset characteristics are significant at a 1% confidence level, and the single-factor analysis also revealed that only the Price-Book Value indicators and Dividend Yield are not statistically significant. However, when running the complete model (Column (1)), we observe that some variables that had been shown to be individually significant were no longer significant when analyzed together with the other indicators. Among them, the Price-to-Sales Ratio and Momentum may be mentioned. This characteristic can be attributed to an endogenous bias implicit in these variables.

On the other hand, company size, volume, Beta, the Relative Strength Index, Momentum, Bollinger Bands and the Stochastic Index remain highly significant in the analysis of the complete model. This suggests that these variables are important when it comes to explaining investors' behavior in relation to assets. Furthermore, the Price-Earnings and Company Value-EBITDA variables remain significant at confidence levels of 10% and 5%, respectively, indicating that they play an important role in explaining the model. Analysis of Table 4 reveals that several asset characteristics are significant in explaining investor behavior. This behavior highlights not only the characteristic of the retail investor as an investor who does not operate only based on noise, but also through paying attention to fundamental indicators, as well as the rationality for decision making in the face of asset trading (Choi *et al*., 2015; Chau *et al*., 2016; Wang and Zhang, 2015).

5 Final Considerations

In this study we investigate information about the characteristics that attract the attention of individual investors who choose to trade assets. We sought to identify the specific characteristics that attract most attention from investors and which of the different types of information available to them actually affect their attention.

We begin our analysis by identifying which fundamental and technical analysis variables have a significant impact on investor attention. The results show that fundamental analysis variables seem to be more important than technical analysis variables with regard to individual investor attention. Specifically, the Price-Earnings and Enterprise Value-EBITDA indicators seem to be the most important in this respect, while none of the technical variables proved to be significantly important in the analysis.

We then conducted a logistical regression test to identify the characteristics of assets that are more likely to attract investors' attention. We observed that the variables of Size, Price-Earnings, Company Value-EBITDA, Beta, Relative Strength Index, Moving Averages and Bollinger Bands are significant. The positive coefficient of the size variable indicates the preference of these investors for large companies. We also observed that investors pay more attention to assets whose Price-Earnings and Company Value-EBTIDA ratio are higher, indicating that investors are willing to pay a premium for the company's profit or profit potential. In addition, when we observed the Beta variable, we noted the positive coefficient indicating a preference for risk, corroborating several studies that indicate that retail investors have a preference for riskier assets (Barberies and Huang, 2008; Kumar, 2009; Andrei and Hasler, 2015). Finally, with regard to variables of a technical nature, the Relative Strength Index and Moving Averages presented a negative coefficient, indicating that assets with a tendency towards devaluation attract investor attention.

Given the results presented in this study, we can conclude that individual investors have different criteria and characteristics that lead to their paying attention to certain assets.

Fundamental analysis appears to carry more weight when choosing investments than technical analysis, indicating that the dynamics of fundamentals are more important for investment decisions than variables associated with noise. Moreover, company size and risk preference are also factors to be considered. It is important to point out that, although this study has limitations, such as the analysis of an emerging market or the selection of indicators used in the research process, it offers an initial understanding of what influences the attention of individual investors. This understanding can be useful both for investors and for companies that wish to attract the attention of the market.

Bibliography and References

Aharon, David Y. & Qadan, Mahmoud, 2020. When do retail investors pay attention to their trading platforms?, *The North American Journal of Economics and Finance*, 53, 101209.

Akhtar, T. 2021. Market multiples and stock returns among emerging and developed financial markets. *Borsa Istanbul Review*. 21(1), 44-56

Almansour, A. Y.; Hasan, E. F.; Matar, G. A. A.; Almansour, Y. M.; Haddad, H. 2022. Investigating the Influence of Financial Indicators on Stock Returns in the Presence of the COVID-19 Pandemic. *Asian Economic and Financial Review*. 12(10).

Andrei, D., Hasler, M., 2015. Investor attention and stock market volatility. *Review of Financial Studies*, 28(1), 33-71

Baker, H. K.; Ridder, A.; Råsbrant, 2020. J. Investors and dividend yields. *The Quarterly Review of Economics and Finance.* 76. 386-395,

Barber, B., Odean, T. 2001. The Internet and the Investor. *Journal of Economic Perspectives*. 15(1), 41-54

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

Barber, B., Odean, T., Zhu, N., 2009. Do Retail Trades Move Markets? *Review of Financial Studies*, 22(1), 151-186

Barberis, N., Huang, M., 2008. Stock as lotteries: the implication of probability weighting for security prices, *American Economic Review,* 98 (5), 2066-2100

Baule, R; Muenchhalfen, P. 2022. What is your desire? Retail investor preferences in structured products. *Review of Behavioral Finance* v. 14 ed 2 p. 197-222 DOI:10.1108/RBF-10-2020-0254

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

Black, F., 1986. Noise. *Journal of Finance*, 41(3), 529-543

Bohemer, E., Jones, C., Zhang, X., Zhang, X., 2021. Tracking retail investor activity. *Journal of Finance*, 76(5), 2249-2305

Chau F., Deesomsak, R., Koutmos, D, 2016. Does investor sentiment really matter? *International Review of Financial Analysis*, 48, 221-232

Choi, J., Kedar-Levy, H., Yoo, S., 2015. Are individual or institutional investors the agents of bubbles? *Journal of International Money and Finance*, 59, 1-22

Chen, H., Cow, E., Shiu, C., 2016. The informational role of individual investors in stock pricing: Evidence from large individual and small retail investors. *Pacific-Basin Finance Journal*, 31, 36-56.

Coakley, J.; Marzano, M.; Nankervis, J. 2016. How profitable are FX technical trading rules? *International Review of Financial Analysis*. 45. 273–282

Cookson, J. A., Krueger, P. M., & Noel, D. P. 2016. The Price-Sales Ratio and Stock Returns. *Journal of Financial and Quantitative Analysis*, 51(4), 1391-1421.

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

Dahlquist, M., Martinez, J., Söderlind, P., 2017. Individual investor activity and performance. *Review of Financial Studies*, 30(3), 866-899

Davydov, D.; Khrashchevskyi I.; Peltomäki J. 2021. Investor attention and portfolio performance: what information does it pay to pay attention to?, *The European Journal of Finance*, 27:17, 1740-1764, DOI: 10.1080/1351847X.2021.1911823

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

Devaney, J. (2014). Bollinger Bands and Forecasting the Financial Markets. *International Journal of Economics, Commerce and Management*, 2(11), 1-9.

Dimpfl, T., Jank, S., 2016. Can Internet Search Queries Help to Predict Stock Market Volatility? *European Financial Management*, 22(2), 171-192

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

Drake, M. S.; Jennings, J.; Roulstone, D. T.; Thornock, J. R. 2016. The Comovement of Investor Attention. *Management Science*. 63(9)

Elsayed, M. A., Daly, K. M., & Eldomiaty, T. I. 2017. Does the Price/Earnings Ratio Predict Market Returns? The Evidence from an Emerging Market. *Research in International Business and Finance*, 41, 445-458.

Farrel, M.; Green, T. C.; Jame, R.; Markov, S. 2022. The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics 145. 616–641*

Gargano, A; Rossi, A. G. Does It Pay to Pay Attention?, *The Review of Financial Studies*, Volume 31, Issue 12, December 2018, Pages 4595–4649,

Goddard, J., Kita, A., Wang, Q., 2015. Investor attention and FX market volatility. *Journal of International Financial Markets, Institution & Money*, 38, 79-96

Gounopoulos, D., Kallias, A., Newton, D., & Tzeremes, N. G. (2018). The Use of the Price-to-Cash Flow Ratio in Investment Decisions: Evidence from the UK Stock Market. *European Journal of Finance*, 24(9), 767-787.

Graham-Marr, A. (2018). The Use of Price Earnings Ratio in Investment Decisions. *International Journal of Economics, Commerce and Management*, 6(4), 197-206.

Griffioen, G. A. W. 2003. Technical Analysis in Financial Markets. *SSRN Electronic Journal*. doi:10.2139/ssrn.566882

Han, L., Xu, Y., Yin, L., 2018. Does investor attention matter? The attention-return relationship in FX markets. *Economic Modelling*, 68, 614-660

Han, L.; Wu, Y.; Yin L. 2018. Investor attention and currency performance: international evidence, *Applied Economics*, 50:23, 2525-2551, DOI: 10.1080/00036846.2017.1403556

Hoffmann, A. O. I.; Shefrinc, H. 2014. Technical analysis and individual investors, *Journal of Economic Behavior & Organization.* 107. 487–511

Huberman, G., and T. Regev. 2001. Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar. *The Journal of Finance.* 56. 387–396.

Igrejas, R. Silva, R. B.; Klotzle, M. C.; Pinto, A. C. F.; Silva P.V. J. G. 2017. Enterprise multiple and future returns of the Brazilian stock market. *REBRAE: Revista Brasileira de Estratégia*, 10 (3). 431-443

Joseph, K., Wintoki, M., Zhang, Z., 2011. Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, 27, 1116-1127

Kahneman, D. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Karlsson, N.; Loewenstein, G.; Seppi, D. J. 2009. The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty* 38:95–115.

Kelley, E., Tetlock, P., 2017. Retail short selling and stock prices. *Review of Financial Studies*, 30(3), 801-834

Kumar, A., 2009, Who gambles in stock market? *Journal of Finance*, 64, 1889-1933

Neely, C. J. (1997). Technical Analysis in the Foreign Exchange Market: A Layman's Guide. *Federal Reserve Bank of St. Louis Review,* 79(6), 31-46.

Ni, Y.; Day, M. Y. Huang, P.; Yu, S. R. 2020. The profitability of Bollinger Bands: Evidence from the constituent stocks of Taiwan 50, *Physica A: Statistical Mechanics and its Applications*, 551, 124-144

Novy-Marx, R., 2012. Is momentum really momentum? *Journal of Financial Economics*. 103(3), 429–453.

Shleifer, A., Summers, L., 1990. The Noise Trader Approach to Finance. *The Journal of Economic Perspectives*, 4(2), 19-33

Semenov, A. 2021. Measuring the stock's factor beta and identifying risk factors under market inefficiency. *The Quarterly Review of Economics and Finance.* 80. 635-649,

Sicherman, N.; Loewenstein, G.; Seppi, D. J.; Utkus, S. P. 2016. Financial Attention. *The Review of Financial Studies*. 29(4)

Sornette, D. 2003. "Critical market crashes", *Physics Reports*. 378(1). 1-98.

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, Q., Zhang, J., 2015. Does individual investor trading impact firm valuation? *Journal of Corporate Finance*, 35, 120-135

Welch, I. 2022. The Wisdom of the Robinhood Crowd. *The Journal of Finance*. 27(3). 1489-1527.

Willinger, M. 1989. Risk aversion and the value of information. *Journal of Risk and Insurance,* 104–112.

Wu, Y., Han, L., Yin, L., 2019. Our currency, your attention: contagion spillovers of investor attention on currency returns. *Economic Modelling,* 80, 49-61

Zhou, Z., Harvey, C. R., & Zhong, L. X. (2018). Dividend Yields and Expected Stock Returns: Evidence from Emerging and Developed Markets. *Journal of Financial Economics*, 129(2), 360-381.

Zhu, Y.; Zhou, G. 2009. Technical analysis: An asset allocation perspective on the use of moving averages. *Journal of Financial Economics*. 92(3). 519-544.

PART 3

"Part 3" presents the main conclusions of this thesis, as well as the limitations and directions for future work. Additionally, it encompasses the references used in this document.

4. CONCLUSION

The studies presented addressed aspects related to investor attention and its relationship with asset pricing. The first study aimed to identify whether online searches are a coherent proxy for the presence of retail investors in the Brazilian stock market, addressing one of the key factors affecting research on individual investors, namely the lack of access to noise trader behavior. Delving further into this knowledge, the second study employed the use of online search proxy to investigate the attention-return relationship in the gold spot market under different market conditions. Lastly, we sought to identify whether value dynamics are more important in explaining changes in attention than noise variables, as it's believed that individual investors trade based on noise.

A natural barrier to investigating the influence of retail investors on markets is related to data availability, as data on retail transactions is not publicly accessible. In this context, a group of studies (Wen *et al*., 2019; Da *et al*., 2020; Cheng *et al*., 2021) employs online search volume as a proxy for the role of individual investors in the market, operating under the belief that searches and investor trades are related (Da *et al*., 2011). Our main contribution in this study was to directly test this assumption using the total value traded exclusively by retail investors in the Brazilian market. In this regard, we compared this measure with the Aggregated Retail Attention (ARA) index following Da *et al*. (2020). Our main findings are as follows: VAR analysis between ARA and retail investor trades indicates the influence of past online searches on future trades, while the opposite relationship is marginally observable. We also conducted OLS models using retail investor trades as the dependent variable and found a strong association between the constructs. Furthermore, we identified that coefficients for lagged online searches are consistently more significant than contemporaneous ones in all examined OLS models, in line with the notion that retail investors would conduct online searches before deciding to trade a particular stock. We also examined the attention-trading relationship under different market states: a) bull/bear market, b) high/low volatility periods, and c) rapid increase in the number of retail accounts in the local market. For each examined market condition, we documented a significant relationship between online searches and individual investor trades, which is always more pronounced for lagged searches, indicating that the attention-return relationship remains conditional, not driven by different market circumstances. A significant limitation in this field is related to data availability, given the difficulty in accessing individual trades. According to our findings, this challenge can be easily overcome by using online searches as a proxy for individual trades, given the strong association between both constructs documented here, facilitating the publication of studies on retail investors.

General evidence also indicates that investor attention plays an important role in understanding the behavior of various markets. In this context, the second study specifically addresses the gold market. While an attention-return relationship has been previously documented, we deemed it plausible to analyze how widespread this relationship can be under different market circumstances. In this regard, we constructed a weekly time series of Google Search Volumes (GSV) for the term "gold price" from 01/01/2007 to 12/31/2018, using log differences as our proxy for investor attention, and analyzed its relationship with spot gold returns during bull (bear) markets, defined as weeks where the return is positive (negative). Our results indicate a positive (negative) relationship between the two variables during bull (bear) markets, which is also asymmetric in terms of magnitude, being more pronounced in bull markets. To further deepen this analysis, we examined how the attention-return relationship varies with the return distribution. Thus, we performed a quantile regression of gold returns and online searches and found that this association is absent only for returns around the median. Another interesting pattern is that the relationship is always negative (positive) for lower (upper) quartiles and increases almost monotonically as the distribution moves from the mean to the tails. Finally, we investigated how investor attention is affected by previous returns based on the direction of price change, and our results indicate that this influence is also asymmetric, being positive (negative) during bull (bear) markets, indicating that recent price trends (increase or decrease) attract the attention of individual investors. Therefore, a plausible conclusion from our aggregated results is that the past performance of the gold price attracts the attention of retail traders who, aiming to profit from the trend, lead to further price increase or decrease, reinforcing the momentum effect. This explanation provides support for the momentum anomaly documented by Moskowitz *et al*. (2012), which is also observable in the gold market, as Baur et al. (2020) report abnormal returns for 12- and 24-month lagged price strategies. Consequently, the aggregated results reported in this study suggest that investor attention is a plausible explanation for the time series puzzle (Moskowitz *et al*., 2012; Papailias *et al*., 2021).

Lastly, although investors, in general, have access to an increasingly wide range of information, literature often mentions that individuals struggle to assimilate all available information due to their limited cognitive resources (Kahneman, 1973). Individual investors are often characterized as traders who rely on less reliable information due to their lesser experience and knowledge. This implies the possibility that they make investment decisions based on less accurate information, which can be considered noise in the financial market context (Black, 1986; De Long *et al*., 1990, Shleifer and Summers, 1990). The third study considers information on the characteristics that attract the attention of individual investors who choose to trade assets. Using Google search volume as a proxy for individual investor attention and adopting the tickers of companies composing the IBRX100 Index from 2014 to 2021 as search terms, we analyzed variables to measure noise and value variables, which are supposedly not considered by noise traders. In an initial empirical analysis, we regressed investor attention on the aforementioned variables. These results indicate that fundamental variables have significant effects on investor attention, indicating that as a company's value rises, more attention is given by investors, suggesting trend behavior by these agents, as attention increases as assets become more expensive. The sensitivity analysis conducted through univariate regression for each of the relevant variables supports this finding by revealing that the dynamics of fundamentals remain more important for their investment decisions than the dynamics of noise variables. The second empirical analysis in this work aimed to identify the characteristics of stocks that attract the most attention from individual investors. The results suggest that larger companies tend to attract more investor attention. In relation to the Price-Earnings and Enterprise Value-EBITDA variables, investors pay more attention to assets with higher multiples. Additionally, the analysis also reveals that investors show a preference for risk, as evidenced by the positive coefficient of the BETA variable. Regarding technical variables, a negative coefficient is observed, indicating that assets with a declining trend are more attractive to investors, following the line of Chau *et al*. (2016), which suggests that sentiment-influenced investors are not irrational as they trade against consensus, buying stocks during bear markets and selling during bull markets.

4.1 Practical and Theorical Contributions

The financial literature has provided significant contributions regarding the theme of investor attention. The studies conducted in this thesis contribute to a better understanding of noise trader behavior in the asset pricing environment. Building upon initial theories whose concepts are based on 1) the difficulty of obtaining secure information for a proxy; 2) the important role of the retail investor in understanding the behavior of various markets; and 3) the individual investor operates based on noise, we can see that the process of more specific investigations incorporated new elements to encompass its own complexity and conjecture a new bias for the theory.

This research provides a contribution to understanding the behavior of retail investors in the market. With its quantitative approach, the current research offers a guarantee of security in using a method capable of capturing and measuring investor attention, as well as providing specific empirical analyses that can contribute to shaping the current behavior of the noise trader.

Thus, throughout the studies, we developed a greater understanding of the dimension of the investor's role and relevance in their market activity. The main research gap that was addressed was methodologically corroborating a secure proxy for the relevant theme. With our findings, this challenge can be easily overcome by using online searches as a proxy for individual trades, given the strong association between both constructs documented here, facilitating the dissemination of studies on retail investors.

Furthermore, this research brings theoretical advancements by explaining intrinsic aspects of the investor attention theme and how they can impact market behavior, both in the specific gold market and more broadly when referring to the investigation of characteristics in the Brazilian market. In this way, our studies contribute to the theories of noise trader behavior and investor sentiment.

4.2 Limitations

This research presents some limitations. Firstly, the study focuses on analyzing the Brazilian market, which, while presenting several favorable and contributory aspects for research, also has characteristics inherent to an emerging and relatively small market. Future studies could consider conducting tests in other markets and, in addition to the discussed issues, comparing across different markets.

Secondly, the focus of this study is on investigating individual investors. Future research could explore investigations using institutional investors to form the sample and allow for a better understanding of similarities and differences between these profiles, as institutional investors are maturing in an environment that is becoming more accessible to information for the noise trader.

Thirdly, in our second study, we only investigated the scenario of the gold market. As suggestions for further studies, it could be considered to extend the analysis to other commodities, enabling a broader panorama investigation, aimed at understanding the behavior of individual investors in that market more comprehensively.

Lastly, our third study initiates a discussion that allows for a counterargument to classic theories of noise trader behavior, which assert that individual investors operate based on noise. For our investigations, we employed a range of variables that enabled us to identify that noise trader behavior follows a sophisticated trend, based on value variables. Future studies could build upon our findings by exploring additional variables beyond those used in our investigations, thus corroborating with this new theoretical perspective and providing more support for further investigations.

BIBLIOGRAPHY AND REFERENCES

Aharon, D. Y.; Qadan, M. 2020. When do retail investors pay attention to their trading platforms?, *The North American Journal of Economics and Finance*, 53, 101209.

Akhtar, T. 2021. Market multiples and stock returns among emerging and developed financial markets. *Borsa Istanbul Review*. 21(1), 44-56

Albuquerque, R. 2012. "Skewness in stock returns: Reconciling the evidence on firm versus aggregate returns", *Review of Financial Studies*, v. 25, 1630-1673.

Almansour, A. Y.; Hasan, E. F.; Matar, G. A. A.; Almansour, Y. M.; Haddad, H. 2022. Investigating the Influence of Financial Indicators on Stock Returns in the Presence of the COVID-19 Pandemic. *Asian Economic and Financial Review*. 12(10).

Amaya, D.;Christoffersen, P.;Jacobs, K.; Vasquez, A. 2015. Does Realized Skewness Predict the Cross-Section of Equity Returns? *Journal of Financial Economics*, 118(1), 135-167.

Andrei, D., Hasler, M., 2015. Investor attention and stock market volatility. *Review of Financial Studies*, 28(1), 33-71.

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

Bachalier, L. 1900. Théorie de la speculation. (ThèseenMathématique). *Annales* Scientifiques de l'École Normale Supérieure, (17)3. Paris: Gauthier-Villars. 21-86.

Bali, T. G., Brown, S. J., Murray, S., Tang, Y., 2017. A lottery-demand-based explanation of the beta anomaly. *Joural of Financial and Quantitative Analysis*, 52(6), 2369-2397

Bali, T;Cakici, N;Whitelaw, R.2010. "Maxing out: stock as lotteries and the cross-section of expected returns," *Journal of Financial Economics.*99, 427-446

Ball, R. 1978. "Anomalies in Relationships Between Securities' Yields and Yield-surrogates." *Journal of Financial Economics* 6,103-126.

Baker, M.;Wulgler, J. A. 2006. Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, Vol 61 (4), 1645-1680.

Baker, H. K.; Ridder, A.; Råsbrant, 2020. J. Investors and dividend yields. *The Quarterly Review of Economics and Finance.* 76. 386-395,

Barber, B., Odean, T. 2001. The Internet and the Investor. *Journal of Economic Perspectives*. 15(1), 41-54

Barber, B.; 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(2), 785-818.

Barber, B.; Odean, T; Zhu, N. 2009, Do retail trades move markets? *Review of Financial Studies*, 22(1), 151-186.

Barberis, N.;Huang, M. 2008. "Stocks As Lotteries: The Implications Of Probability Weighting For Security Prices." *American Economic Review* 98(5): 2066- 2100.

Barberis, N., Xiong, W., 2012, Realization utility. *Journal of Financial Economics*, 104, 251-271

Baur, D., Dimpfl, T., Jung, R., 2012. Stock return autocorrelations revisited: A quantile regression approach. *Journal of Empirical Finance*, 19(2), 254-265

Baur, D.;Dimpfl, T. 2016. Googling gold and mining bad news. *Resources Policy*, 50, 306-311

Baur, D., Dichtl, H., Drobetz, W., Wendt, V., 2020. Investing in gold – Market timing or buy-and-hold. *International Review of Financial Analysis*, 71,101281

Behrendt, S.; Schmidt, A. 2018. The Twitter myth revisited: Intraday investor sentiment, Twitter activity and individual-level stock return volatility. *Journal of Banking & Finance* 96, 96(C). 355-367.

Behrendt, S.; Prange, P. 2019. What Are You Searching For? On the Equivalence of Proxies for Online Investor Attention. *Finance Research Letters*. 38, 101401.

Bijl, L.;Kringhuag, 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*, 41:529-543.

Bohemer, E., Jones, C., Zhang, X., Zhang, X., 2021. Tracking retail investor activity. *Journal of Finance*, 76(5), 2249-2305

Brooks, C. 1998. Predicting Stock Index Volatility: Can Market Volume Help? *Journal of Forecasting*, 17(1), 59-80.

Buchinsky, M., 1995. Estimating the asymptotic covariance matrix for quantile regression models a Monte Carlo study. *Journal of Econometrics*, 68, 303-338

Challet, D.,; Ayed, A. B. H. 2013. Predicting financial markets with Google Trends and not so random keywords. arXiv preprint, 1307.4643.

Chan, W. S. 2003. Stock price reaction to news and no-news: drift and reversal after headlines. *Journal of Financial Economics.*

Chau F., Deesomsak, R., Koutmos, D, 2016. Does investor sentiment really matter? *International Review of Financial Analysis*, 48, 221-232

Chemmanur, T.; Yan, A. 2009. "Product market advertising and new equity issues," *Journal of Financial Economics*, Elsevier, vol. 92(1), pages 40-65.

Chen, H., Cow, E., Shiu, C., 2016. The informational role of individual investors in stock pricing: Evidence from large individual and small retail investors. *Pacific-Basin Finance Journal*, 31, 36-56.

Cheng, F., Wang, C., Chiao, C., Yao, S., Fang, Z., 2021. Retail attention, retail trades, and stock price crash risk. *Emerging Markets Review*, 49, 100821.

Choi, P. M. S.; Choi, J. H.; Chung, C. Y. 2020. Do individual traders undermine firm valuation? *Finance Research Letters*. 36, 101567.

Choi, H.;Varian, H. 2009. Predicting Initial Claims for Unemployment Benefits, *Relatório Técnico*. Google Inc.

Choi, J., Kedar-Levy, H., Yoo, S., 2015. Are individual or institutional investors the agents of bubbles? *Journal of International Money and Finance*, 59, 1-22

Coakley, J.; Marzano, M.; Nankervis, J. 2016. How profitable are FX technical trading rules? *International Review of Financial Analysis*. 45. 273–282

Cookson, J. A., Krueger, P. M., & Noel, D. P. 2016. The Price-Sales Ratio and Stock Returns. *Journal of Financial and Quantitative Analysis*, 51(4), 1391-1421.

Da, Z.; Engelberg, J.; Gao, P. 2011. Search of Attention. *The Journal of Finance*, 66, 1461-1499.

Da, Z.; Engelberg, J; Gao, P. 2015, The Sum of All FEARS Investor Sentiment and Asset Prices? *The Review of Financial Studies*, 28(1), 1-32

Da, Z., Hua, J., Hung, C., Peng, L., 2020. Market returns and a tale of two attentions. *SSRN Working Paper #3551662*. *SSRN Electronic Journal*

Dahlquist, M., Martinez, J., Söderlind, P., 2017. Individual investor activity and performance. *Review of Financial Studies*, 30(3), 866-899

Dastgir, S; Demir, E.;Downing, G.;Gozgor, G.;Lau, C. 2019. The causal relationship between Bitcoin attention and Bitcoin returns: Evidence from the Copula-based Granger causality test. *Finance Research Letters*, 28, 160.164.

Davydov, D.; Khrashchevskyi I.; Peltomäki J. 2021. Investor attention and portfolio performance: what information does it pay to pay attention to?, *The European Journal of Finance*, 27:17, 1740-1764, DOI: 10.1080/1351847X.2021.1911823

De Bondt, W. 1998. A Portrait of the Individual Investor. European Economic Review, 42, 831-844.

De Bondt, W.; Thaler, R. 1985. Does the Stock Market Overreact? *Journal of Finance*, (40)3, 793-805.

De Long, J. B.;Shleifer, A.;Summers, L. H.; Waldmann, R. J. 1990. Noise Trader Risk in Financial Markets. Source: *The Journal of Political Economy*, 98(4), 703-738.

Dellavigna, S.; Pollet, J. M. 2009. "Investor Inattention and Friday Earnings Announcements," Journal of Finance, American Finance Association. 64(2), pages 709-749.

Desagre, C., D'hondt, C., 2021. Googlization and retail trading activity. *Journal of Behavioral and Experimental Finance*, 29, 100453.

Devaney, J. (2014). Bollinger Bands and Forecasting the Financial Markets. *International Journal of Economics, Commerce and Management*, 2(11), 1-9.

Dimpfl, T.; Jank, S. 2016. Can Internet Search Queries Help to Predict Stock Market Volatility? *European Financial Management*, 22, 2, 171-192.

Drake, M. S.; Jennings, J.; Roulstone, D. T.; Thornock, J. R. 2016. The Comovement of Investor Attention. *Management Science*. 63(9)

Elsayed, M. A., Daly, K. M., & Eldomiaty, T. I. 2017. Does the Price/Earnings Ratio Predict Market Returns? The Evidence from an Emerging Market. *Research in International Business and Finance*, 41, 445-458.

Eom, C.;Kaizoji, T.;Kang, S.;Pichl, L.2019. Bitcoin and investor sentiment: Statistical characteristics and predictability. *Physica A*, 514, 511-521

Fama, E.1970. Efficient capital markets: a review of theory and empirical work. The *Journal of Finance*, (25)2, 383-417.

Fama, E.1991. Efficient capital markets: II. The Journal of Finance, (46) 5, 1.575-1.618,

Fang, L. H.; Peress, J. 2009. Media Coverage and the Cross-Section of Stock Returns. *The Journal of Finance*, 64(5), p. 2023-2052.

Farrel, M.; Green, T. C.; Jame, R.; Markov, S. 2022. The democratization of investment research and the informativeness of retail investor trading. *Journal of Financial Economics 145. 616–641*

Friedman, M. 1953 "The methodology of positive economics". *Essays in Positive Economics*. Chicago: University of Chicago Press.
Fisher, B., 1986. Noise. *Journal of Finance*, 41(3), 529-543

Gargano, A; Rossi, A. G. Does It Pay to Pay Attention?, *The Review of Financial Studies*, Volume 31, Issue 12, December 2018, Pages 4595–4649, Goddard, J.;Kita, A.;Wang, Q. 2015. Investor attention and FX market volatility. *Journal of International Financial Markets, Institution & Money*, 38, 79-96

Gounopoulos, D., Kallias, A., Newton, D., & Tzeremes, N. G. (2018). The Use of the Price-to-Cash Flow Ratio in Investment Decisions: Evidence from the UK Stock Market. *European Journal of Finance*, 24(9), 767-787.

Graham-Marr, A. (2018). The Use of Price Earnings Ratio in Investment Decisions. *International Journal of Economics, Commerce and Management*, 6(4), 197-206.

Griffioen, G. A. W. 2003. Technical Analysis in Financial Markets. *SSRN Electronic Journal*. doi:10.2139/ssrn.566882

Grossman, S.J. 1975, On the efficiency of competitive stock markets where traders have diverse information, *Journal of Finance.*

Grossman, S; Stiglitz, J. 1980. "On the Impossibility of Informationally Efficient Markets." *American Economic Review.* 70(3): 393-408.

Grullon, B.; Kanatas, G.; Weston, J. P. 2004. Advertising, Breadth of Ownership, and Liquidity. *Review of Financial Studies*, 17(2). 439-461.

Hamid, A., Heiden, M., 2015. Forecasting volatility with empirical similarity and GoogleTrends. *Journal of Economic Behavior and Organization*, 117, 62-81

Han, L.;Xu, Y.;Yin, L.2018. Does investor attention matter? The attention-return relationship in FX markets. *Economic Modelling*, 68, 614-660

Han, L.; Wu, Y.; Yin L. 2018. Investor attention and currency performance: international evidence, *Applied Economics*, 50:23, 2525-2551, DOI: 10.1080/00036846.2017.1403556

Han, L.; Xu, Y.; Yin, L.2017. Does investor attention matter?The attention-return relation in gold futures market, *Economics Discussion Papers*, (37).

Han, B.; Kumar, A. 2013, Speculative retail trading and asset prices, *Journal of Financial and Quantitative Analysis* 48, 377–404.

Harvey, C.R.2001. The specification of conditional expectations. *Journal of Empirical Finance*, 8(5), 573-638.

HIRSHLEIFER, D.; TEOH, S. H., 2003. "Limited attention, information disclosure, and financial reporting," Journal of Accounting and Economics, Elsevier, vol. 36(1-3), pages 337-386.

Hoffmann, A. O. I.; Shefrinc, H. 2014. Technical analysis and individual investors, *Journal of Economic Behavior & Organization.* 107. 487–511

Huberman, G.,; Regev. T. 2001. Contagious Speculation and a Cure for Cancer: A Nonevent that Made Stock Prices Soar. *The Journal of Finance.* 56. 387–396.

Huddart, S. J.; Lang, M. H.; Yetman, M.2012. Volume and Price Patterns Around a Stock's 52-Week Highs and Lows: Theory and Evidence. *Management Science*, 55(1), 16-31.

Ibikunle, G.; McGroarty, F.; Rzayev, K. 2020. More heat than light: Investor attention and bitcoin price discovery. International Review of Financial Analysis, 69.

Igrejas, R. Silva, R. B.; Klotzle, M. C.; Pinto, A. C. F.; Silva P.V. J. G. 2017. Enterprise multiple and future returns of the Brazilian stock market. *REBRAE: Revista Brasileira de Estratégia*, 10 (3). 431-443

Jain, A., Biswal, P., 2019. Does internet search interest for gold move the gold spot, stock and exchange rate markets? A study from India. *Resources Policy,* 61, 501-507

Jensen. M.1978. Some anomalous evidence regarding market efficiency. *Journal of financial Economics,*(6), 95-101.

Joseph, K.;Wintoki, M.; Zhang, Z., 2011. Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting*, 27, 1116-1127.

Kahneman, D. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D.; Tversky, A.1982. The psychology of preferences. *Scientific American* 246:160-73.

Karlsson, N.; Loewenstein, G.; Seppi, D. J. 2009. The ostrich effect: Selective attention to information. *Journal of Risk and Uncertainty* 38:95–115.

Kelley, E., Tetlock, P., 2017. Retail short selling and stock prices. *Review of Financial Studies*, 30(3), 801-834

Kostopoulos, D., Meyer, S., Uhr, C., 2020. Google search volume and individual investor trading. *Journal of Financial Markets*, 49, 100544

Kumar, A., 2009. Who gambles in the stock market? *Journal of Finance*, 64(4), 1889-1933

Kristoufek, L., 2013. BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific Reports*, 3, 3415

Kumar, A.2009, Who gambles in stock market? *Journal of Finance*, 64, 1889-1933

Liu, J., Stambaugh, R. F., Yuan, Y., 2018. Absolving beta of volatility's effects. *Journal of Financial Economics*, 128(1), 1-15

Merton, R.C. 1987. A Simple Model of Capital Market Equilibrium with Incomplete Information. Journal of Finance, 42, 483-510

Meshcheryakov, A.; Winters, D. B. 2020. Retail investor attention and the limit order book: Intraday analysis of attention-based trading. International Review of Financial Analysis 81 (2022) 101627

Mitton, T.;Vorkink, K.2007. Equilibrium underdiversification and the preference for skewness. *Review of Financial Studies.* 20, 1255–1288.

Mondria, J. 2010. Portfolio choice, attention allocation, and price comovement. *Journal of Economic Theory*, 2010, 145(5) 1837-1864

Moskowitz, T., Ooi, Y., Pedersen, L., 2012. Time series momentum. *Journal of Financial Economics*, 104, 228-250.

Neely, C. J. (1997). Technical Analysis in the Foreign Exchange Market: A Layman's Guide. *Federal Reserve Bank of St. Louis Review,* 79(6), 31-46.

Ni, Y.; Day, M. Y. Huang, P.; Yu, S. R. 2020. The profitability of Bollinger Bands: Evidence from the constituent stocks of Taiwan 50, *Physica A: Statistical Mechanics and its Applications*, 551, 124-144

Novy-Marx, R., 2012. Is momentum really momentum? *Journal of Financial Economics*. 103(3), 429–453.

Pástor, L.; Sinha, M.; Swaminathan, B. 2008. Estimating the intertemporal risk-return tradeoffusing the implied cost of capital. *Journal of Finance* 63 (6), 2859–2897.

Papailias, F., Liu, J., Thomakos, D., 2021. Return signal momentum. *Journal of Banking & Finance*, 124, 106063.

Matheus Machado de Pereira, M. M.; Rosa, T. G.; Bender Filho, R. 2020. Influência Do Google Trends Em Ações Listadas Na Bolsa De Valores Brasileira: Evidências A Partir Da Modelagem Pvar. Artigo. *Revista eletrônica de administração.* 26 (3)

Piccoli, P., de Castro, J., 2021. Attention-return relation in the gold market and market states. *Resources Policy*, 74, 102333

Rabelo Junior, T. S.; Ikeda, R. H.2004. Mercados Eficientes e Arbitragem: um estudo sob o enfoque das finanças comportamentais. *Revista Contabilidade e Finanças*, (34).

Ramos, H. P.; Ribeiro.K. K. M.; Perlin, M. S. 2017. The forecasting power of internet search queries in the Brazilian financial Market. *Mackenzie Management Review*, 19(2). 184-210.

Seasholes, M. S.; Wu, G. 2007. Predictable behavior, profits, and attention, *Journal of Empirical Finance*, 14(5), 590-610.

Semenov, A. 2021. Measuring the stock's factor beta and identifying risk factors under market inefficiency. *The Quarterly Review of Economics and Finance.* 80. 635-649,

Sharpe, W.1964, Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, (19), 425-442.

Shefrin,H.; Statman, M. 2000. Behavioral portfolio theory. *Journal of Financial and Quantitative Analysis*, 35 (2), 127-151

Shleifer, A.2000. A. Inefficient Markets: An Introduction to Behavioral Finance. New York. Oxford University Press.

Shleifer, A.; Summers, L. H. 1990. The Noise Trader Approach to Finance. *Journal of Economic Perspectives*, 4 (2), 19-33.

Shleifer, A.; Vishny, R. W**.** 1997. A Survey of Corporate Governance. The Journal of Finance. 52(2). 737-783.

Shiller, R. J. 1984. "Stock Prices and Social Dynamics, "Brookings Papers on Economic Activity, 457-97.

Sicherman, N.; Loewenstein, G.; Seppi, D. J.; Utkus, S. P. 2016. Financial Attention. *The Review of Financial Studies*. 29(4)

Smith, G. P., 2012. Google internet search activity and volatility prediction in the market for foreign currency. *Finance Research Letters*, 9(2), 103-110

Sornette, D. 2003. "Critical market crashes", *Physics Reports*. 378(1). 1-98.

Swamy, S.; Munusamy, D.; Takeda, F. 2019. Investor Attention and Google Search Volume Index: Evidence from an Emerging Market using Quantile Regression Analysis. *Research in International Business and Finance*. 50. 1-17

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, W., 2018 Investor sentiment and the mean-variance relationship: European evidence. *Research in International Business and Finance*, 46, 227-239.

Wang, Q., Zhang, J., 2015. Does individual investor trading impact firm valuation? *Journal of Corporate Finance*, 35, 120-135

Welch, I. 2022. The Wisdom of the Robinhood Crowd. *The Journal of Finance*. 27(3). 1489-1527.

Willinger, M. 1989. Risk aversion and the value of information. *Journal of Risk and Insurance,* 104–112.

Wu, Y.; Han, L.; Yin, L.2019. Our currency, your attention: contagion spillovers of investor attention on currency returns. *Economic Modelling,* 80, 49-61.

Yao, T.; Zhang, Y.; Ma, C.2017. How does investor attention affect international crude oil prices? *Applied Energy*, 205, 336-344.

Yoshinaga, C. E.; Rocco, F. B. 2020. Investor Attention: Can Google Search Volumes Predict

Stock Returns? *Brazilian Business Review*. 17(5). 523-539.

Yu J.; Yuan, Y.2011, Investor sentiment and the mean–variance relation, *Journal of Financial Economics* 100, 367–381.

Yung, K.; Nafar, N. 2017. Investor attention and the expected return of reits. *International review of economics and finance*, 48, 423-439.

Zhang, W., Wang, P., Li, X., Shen, D., 2018. Quantifying the cross-correlations between online searches and Bitcoin market. *Physica A*, 509, 657-672

Zhou, Z., Harvey, C. R., & Zhong, L. X. (2018). Dividend Yields and Expected Stock Returns: Evidence from Emerging and Developed Markets. *Journal of Financial Economics*, 129(2), 360-381.

Zhu, Y.; Zhou, G. 2009. Technical analysis: An asset allocation perspective on the use of moving averages. *Journal of Financial Economics*. 92(3). 519-544,