

## Investor Attention: Can Google Search Volumes Predict Stock Returns?

Claudia Yoshinaga<sup>1</sup>

claudia.yoshinaga@fgv.br |  0000-0002-7756-6083

Fabio Rocco<sup>1</sup>

f.bouroc@gmail.com |  0000-0003-0151-7231

### ABSTRACT

This paper investigates the role of investor attention in predicting future stock market returns for Brazilian stocks using Google Search Volume (GSV). We tested whether lagged variations in GSV are followed by changes in excess returns by testing 57 stocks from the Ibovespa using weekly search data from Google Brazil from 2014 to 2018. Similar to previous research on the U.S. market, we found that increases in GSV are followed by lower excess returns. Additionally, we show that the more traded a stock is, the higher the effect. This is consistent with the hypothesis that higher individual investor attention leads to lower subsequent returns, suggesting that increasing popularity causes stock prices to deviate from their fundamental value.

### KEYWORDS

Investments, Abnormal Returns, Inefficient Markets, Behavioral Finance, Price Anomaly

<sup>1</sup>Fundação Getúlio Vargas (FGV EAESP),  
São Paulo, SP, Brasil

Received: 11/11/2019.  
Revised: 02/11/2020.  
Accepted: 02/29/2020.  
Published Online: 08/24/2020.  
DOI: <http://dx.doi.org/10.15728/bbr.2020.17.5.3>



## 1. INTRODUCTION

Many studies show that grabbing investor attention has an important impact on financial market characteristics, including liquidity, diversity of ownership (Grullon, Kanatas, & Weston, 2004), returns (Barber & Odean, 2007) and volatility (Andrei & Hasler, 2014). A quote commonly attributed to Joseph Kennedy from just before the Wall Street crash of 1929 links the popularity of stocks to predictable returns: “You know it’s time to sell when the shoeshine boys give you stock tips”. His point was that the increased popularity of stocks to the point where they are being traded by inexperienced investors supposedly indicates their imminent price decrease. The popularity of an investment may be related to the mutable variable known as investor attention, which has been studied by Barber and Odean (2007); Challet and Ayed (2013); Preis, Moat and Stanley (2013); Preis, Reith and Stanley (2010); and, more recently, Bijl, Kringhaug, Molnár and Sandvik (2016) and Mayer (2018).

Understanding the relationship between investor attention and future stock returns is fundamental to investment strategies. If one can predict the impact of an increase or decrease in investor attention on future performance, successful traders can explore this market anomaly. Our purpose in this paper is to evaluate whether GSV can predict stock returns in the Brazilian stock market. As in Bijl et al. (2016), we hypothesize that an increase in GSV for a stock during a certain week is followed by lower returns in the following weeks. To analyze this, we estimate a long-panel regression of the weekly market adjusted excess log-returns of 57 stocks on the Ibovespa against lags of GSV log-variation from January 2014 to December 2018, controlling for several market and firm factors. We estimated this using FGLS to account for serial correlation. We also included an interaction term in the model to check if investor attention could affect stocks differently according to the trading volume, a novel concept in this field’s literature.

In our main analysis, we find that a one percentage point increase in GSV in one week is followed by a decrease of 0.001 percentage points in returns in the following week. The next three lags are also negative and jointly statistically significant, accruing a decrease of 0.0035 p.p.. Next, we evaluate whether the results are sensitive for stocks that are more and less frequently traded. We find that more frequently traded stocks are more sensitive to shifts in investor attention. While the lags of GSV for stocks in the lower quartile of trading volume in the sample have an accrued negative effect for future returns of 0.0027 p.p., the accrued effect is 0.0045 for the stocks in the third quartile.

Our results contribute both to the existing debate about price anomalies, investor attention, and the predictive power of GSV and provide information to Brazilian investors who can exploit the pricing anomaly at play. The main motivation of this paper is to examine investor attention and its impact on stock returns in the Brazilian stock market; this information is currently missing in the literature. To do so, we investigated the predictability of stock returns in Brazil by analyzing the investor attention time series available for the Brazilian market using weekly data up to 2018. We contribute to academic research on behavioral finance by testing the link between investor attention and the return of the individual stock in the spotlight. We take GSV as a measure of investor attention as individuals usually collect information by via internet. Google has the lead position among all search engines available on the internet, with approximately 90% of market share. Specifically, regarding the GSV tool, this paper tests stock return predictability in a new environment, following methodology first proposed by Preis et al. (2010), and later improved by Bijl et al. (2016) and Challet and Ayed (2013), all of which study the American markets. That said, this paper is the first to analyze GSV and stock returns for Brazilian stocks.

Last, from an investor's perspective, this investigation offers powerful insight into the effectiveness of investor attention-based strategies such as one using GSV. Considering the peculiarity of Brazilian stock market trading compared to developed markets, this study offers local evidence for Brazilian investors in comparison to international studies with higher levels of individual investor trading. In 2019, B3 achieved an all-time record of 1.5 million individuals registered to trade stocks. This represents approximately 0.7% of the country's population; on the other hand, these investors are responsible for almost 20% of trading volume (B3, 2019). To illustrate this great difference, 52% of adults in the US market are invested in stocks (Gallup, 2016); this is an all-time low mark versus a peak of 65% in 2007, measured just before the subprime crisis. Considering that Brazilian markets currently have reached the lowest interest level in Brazil's economic history, it is expected that equity markets will get more attention and investment from individual investors.

## 2. LITERATURE REVIEW

It is important to mention that, across these studies, not only does the precise definition of investor attention change, but so do the tools used to quantify the attention variable itself. The attention of investors can be measured by different proxies, such as advertising expenses (Grullon, Kanatas and Weston, 2004), analysts' coverage (Liu, Wu & Chiang, 2014), 52-week historical high prices (Li & Yu, 2012), and media interest (Fang & Peress, 2009).

Notably, the increasing availability of big data has contributed by engaging more research on the topic and therefore providing more empirical evidence for many different markets. To illustrate the relevance of investor attention, one may turn to the recent Bitcoin phenomena and work in the vein of Kristoufek (2013). Throughout 2017, the correlation between Google searches worldwide for the word "bitcoin" and the price of this cryptocurrency quoted in US Dollars was approximately 0.95. By analyzing the recent dynamics of Bitcoin price evolution, Kristoufek (2013) identifies the dependence of cryptocurrency's price on investor sentiment, observing how price spikes generally occur in lockstep with overall investor interest in the asset. Apart from cryptocurrencies, Mayer (2018) demonstrates that nonfundamental attention-grabbing ads displayed at highly viewed college football tournament finals are also a powerful means of shifting stock prices. Thus, investor attention emerges as a key element for truly understanding price movements, momentum, and mean reversal dynamics in real markets.

New research tools have indeed helped deepen our understanding of the potential impacts of investor attention. Google's search analytics allows for time series analysis of GSV, where search volumes for specific words or expressions can be studied over time. This opens a new window for studying stock price predictability, since Google's activity functions are perhaps one of the best measurable proxies of investor attention to date. This information can be delimited in terms of location and period of time and can be collected over time to form a time series. As a result, an increasing number of studies have come to use this specific research tool (see, e.g., Bijl et al., 2016; Challet & Ayed, 2013; Da, Engelberg & Gao, 2011; Kristoufek, 2013; Mayer, 2018; Preis et al., 2013; Preis et al., 2010) to approximate investor attention, or sentiment, by GSV. Still, in spite of the wider array of possible studies based on GSV, such as its relationship with economic activity, unemployment, and general index fluctuation, as well as stock turnover and volume, any relationship between GSV and stock price returns presents a significant pricing anomaly that challenges Fama's (1970) random walk-oriented efficient market model.

Da, Engelberg and Gao (2011) first presented search frequency in Google as a measure of investor interest. Lou (2014) and Chemmanur & Yan (2019) explore the topic of investor attention by examining the impact of advertisement on stock prices. Mayer (2018) uses company advertising campaigns displayed in college football “bowls”, as these can be measured in terms of popularity by television viewer indexes, allowing for a sensitivity analysis of stock prices. Typically, one company acts as the major sponsor of a college bowl, and, in turn, gains television exposure through in-game advertisements and commercials. By using television indexes for the bowls as attention proxies and analyzing the impact of different levels of attention on the sponsoring companies’ stock prices, the author concludes that investor attention, specifically noise trader attention, causes abnormal returns in stocks of 1.51% in 1-5 days and, later, an almost complete reversal of 1.09% by the next week (Mayer, 2018).

Barber and Odean (2007) discuss that retail investors’ attention is a scarce resource, and given the thousands of options available to investors, there is a clear search problem at play. Such a problem is aggravated by the fact that individual investors tend to be net buyers, usually holding long positions and only selling stocks they already own (and short-selling much less frequently than institutional investors). The authors’ findings coincide with these conjectures, as they identify attention-driven buying in individual investors’ behavior as measured by the impact of attention-getting events such as news exposure, abnormal trading, and abnormal returns on stocks’ short-term trading volumes.

Inherently, GSV studies are relatively new to finance due to the nature of the tool used, which only became available and reliable post-2008. Although other related studies have been performed around the world, for instance linking trading volume to search volume, three key studies have looked specifically at the price-predicting power of GSV. First, Preis et al. (2010) attempted to define the relationship between price and GSV. While it was found that GSV had no statistically significant predictive power for stock returns, it appeared to have a strong correlation to trading volume. Later, Preis et al. (2013) tested whether a trading strategy based on buying and selling stocks according to certain Google search queries was able to outperform the market. The results indicated that Google queries were indeed able to predict stock returns, outperforming the market index by 310% in the seven-year period tested. Challet and Ayed (2013) worked in similar vein as Preis et al. (2013), albeit with less specific objectives, seeking to uncover whether GSV information is robust enough to predict market movements. However, one of the study’s most relevant contributions is in its methodological analysis and examination of potential biases, specifically those relating to backtesting.

The first type of bias identified is tool bias, where it is easy to find predictability in old data on financial markets utilizing modern tools (with strong computational power, especially when analyzing large data sets). For this reason, it is methodologically not advised to backtest old data using modern tool capabilities. The second bias type, data bias, relates to the actual availability of data and its relative significance at different points in time. For instance, Challet and Ayed (2013) point out that GSV data was only updated sporadically before August 2008, meaning that any study before this threshold automatically deals with different quality of data within the same data set. On another note, the authors affirm there is an inherent data revision problem, which, while it puts all available data on a comparable scale, may disrupt a time series. This impacts GSV significantly, given that, in 2012, the normalization method used by Google changed to what is currently a 100-point index system.

Finally, Google Trends – and its still experimental nature – signals the need for further measures to ensure biases do not significantly impact the experiment. Keyword query choice emerges as perhaps the most significant decision from an experimental point of view and carries a series of issues. When testing for GSV in a financial context, one would ideally want to distinguish (and isolate) the stock or company from the investor and consumer perspectives. Searching for the exchange ticker might be a reasonable solution, although it does remove significant queries by investors who were using the company’s actual name, for instance, to reach its investor relations website. Still, there is a problem with ticker changes, which must be accounted for in cases of mergers, listing segment changes, and a multitude of other reasons. In summary, Challet and Ayed (2013) contribute by adding a robust bias analysis to the literature while also defending the thesis that there are significant random (i.e. false-positive) conclusions to be drawn from GSV.

Last, Bijl et al. (2016) build upon these previously cited research papers in order to test for GSV’s price predictability using a slightly different methodology. Stock return predictability is tested for individual firms on a weekly basis using GSV for the American stock market (testing companies in the S&P500 index) from 2007 to 2013; they conclude that high GSV predicts subsequent low future returns. The conclusion is also tested in a practical context, using a long-short strategy that confirms the effect is not strong enough to be profitable after accounting for transaction costs. Methodologically, the authors admit to the unreliability of Google Trends data prior to 2008 but, nonetheless, use 2007 data to build 52-week rolling betas used in the CAPM model to test for excess return. Furthermore, they choose to restrain data regionally (US only) to reduce risk of non investor related searches, which intuitively should be more common outside of the US. This is judged to be less of an issue in Brazil, due to the relatively high international investor concentration in B3, representing approximately 27% of traded volume in 2018 as of April (B3, 2018). Nevertheless, Bijl et al. (2016) highlight that worldwide GSV queries are more likely to be “false positives” given that chances of overlapping words with different meanings increase exponentially. On keyword choice, the authors use company names instead of tickers, given many tickers are also company abbreviations, which still capture unwanted searches from retail clients and other non investors. For data processing, the authors added several other variables to the regression analysis for the sake of robustness, including the January effect, stock volatility and detrended log trading volume (Bijl et al., 2016).

### 3. DATA AND METHODOLOGY

The analysis covers the period between January 2014 and December 2018. We use the 57 stocks (54 companies) which composed the Ibovespa Index as of January 2018 and which had the necessary data available for the period. The choice of only including companies belonging to Ibovespa was made to guarantee that all stocks have relatively high daily traded volume and to assure prices dynamically adjust due to market information. In other words, it would do no good to include company stocks that do not trade frequently and, thus, do not react to market news, GSV, and other market variables, for days in a row. Moreover, considering that the Ibovespa, much like the S&P 500, concentrates the most relevant stocks of the B3 exchange (Castro et al. 2019), these are also the stocks most likely to have significant GSV data available, as suggested by Bijl et al. (2016).

The five trading years of 2014-2018 reflect different periods in Brazilian macroeconomic policy and economic performance. In 2014, the observed Gross Domestic Product (GDP) growth was aligned with Brazil’s emerging market peers; in the last two years, GDP has shrunk approximately 3%, with a slight upturn reflected in a mild 1% growth in 2017 (Bacen, 2018). This period also encompasses a wide range of fiscal and monetary policies. This, however, does not present a significant limitation to the study, as the study is not sector-based. Since we focus on Ibovespa firms, which are the most liquid and relevant stocks representing market movement, it should be approximately neutral to macroeconomic cycles. Interestingly, as demonstrated by Asness, Moskowitz and Pedersen’s (2013) and Daniel and Moskowitz’s (2016) momentum studies, anomaly tests are empirically sensitive to crises, making the selected time range also opportune for such reexamination.

To examine the predictive value of GSV on stock returns we estimate a panel regression model where each asset  $i$ ’s market adjusted excess log-return ( $r_i^e$ ) is explained by its own first five lags, the first to fifth lags of the log-variation in its GSV ( $gsv_i$ ), plus the first to fifth lags of trading volumes ( $vol_i$ ), as well as returns and GSV’s mid-term ( $\sigma_i^{mid}$  and  $gsv_i^{mid}$ ) and long-term ( $\sigma_i^{long}$  and  $gsv_i^{long}$ ) volatility, in addition to firms’ size and industry effects (10 industry GICS classifications), as written in Equation (1):

$$r_i^e = \beta_0 + \sum_{\ell=1}^5 \rho_{\ell} r_{i,t-\ell}^e + \sum_{\ell=1}^5 \theta_{\ell} gsv_{i,t-\ell} + \sum_{\ell=1}^5 \lambda_{\ell} vol_{i,t-\ell} + \beta_1 \sigma_{it}^{mid} + \beta_2 \sigma gsv_{it}^{mid} + \beta_3 \sigma_{it}^{long} + \beta_4 \sigma gsv_{it}^{long} + \beta_5 size_{it} + \sum_{j=1}^9 \alpha_j Industry_j + \varepsilon \tag{1}$$

The motivation for the addition of five lags of excess return, GSV, and trading volume is twofold. First, this allows the model to capture a wider range of data, potentially decreasing the impact of any data imprecision, such as if GSV for a given period actually represents volumes from the previous period. Second, the lags also allow the market to eliminate any pricing anomaly, such as momentum effects.

The panel is formed by 57 stock tickers across 261 weeks, creating a long-panel regression; this requires adding time-series asymptotic features to the estimation, in contrast to traditional short panels where the number of periods of time are assumed to be fixed and the number of individuals are assumed to grow indefinitely. Correlation across time for a given stock can be incorporated into the model using an Autoregressive Moving-Average (ARMA) model for the errors, allowing the parameters to differ among stocks (Cameron & Trivedi, 2005). If errors are heteroskedastic and the autocorrelation parameter also differs across stocks, we could estimate separate time series regressions for each stock using all the time periods to yield consistent ARMA parameters. Nevertheless, as explained by Cameron and Trivedi (2005), by using Feasible Generalized Least Squares (FGLS), we can also yield consistent estimates for these parameters using all NT observations allowing for heteroskedasticity and stock-specific correlation across time and correlation across stocks.

All variables are measured weekly. Market adjusted excess returns are measured as in Equation (2), where  $r_{it}$  are the weekly log-returns for each asset,  $r_{mt}$  is the log market return, and  $\beta$  is the one year rolling beta of each asset in the Bovespa Index (local index beta):

$$r_{it}^e = r_{it} + \beta r_{mt} \tag{2}$$

The trading volume series are detrended and transformed into a log function to attenuate peaks and valleys throughout the data set, since volume can be highly volatile. This methodology is based on Campbell, Grossman and Wang (1993) and Bijl et al. (2016):

$$vol_{it} = \log(volume)_{it} - \frac{1}{12} \sum_{s=t-11}^t \log(volume)_s \quad (3)$$

The volatility of returns is calculated daily using rolling windows of five and 20 trading days for mid- and long-term volatility, respectively referring to a one week and one-month trading period. Later, we select the measures at weekly intervals to match the other variables. For the volatility of GSV, since we only have original weekly data, mid- and long-term volatility is calculated using rolling windows of four and 12 weeks, respectively referring to one and three-month periods. We have:

$$\sigma r_{is}^{mid} = \sqrt{\sum_{s=1}^5 (r - \bar{r})_{is}^2} \text{ and } \sigma r_{is}^{long} = \sqrt{\sum_{s=1}^{20} (r - \bar{r})_{is}^2} \quad (4)$$

$$\sigma gsv_{it}^{mid} = \sqrt{\sum_{t=1}^5 (gsv - \overline{gsv})_{it}^2} \text{ and } \sigma gsv_{it}^{long} = \sqrt{\sum_{t=1}^{12} (gsv - \overline{gsv})_{it}^2} \quad (5)$$

Finally, size is measured as the log of the total market capitalization of each stock at each week, and the industry effects are captured by nine dummies for each of the GICS classifications minus one.

All variables except GSV are from Economatica and are quoted in Brazilian Reais (BRL) when applicable. The GSV variable is extracted directly from the Google Trends tool via the “gtrendsR” package for the R programming language, which includes several search options that must be carefully considered and parametrized. First, it is important to understand how Google presents its GSV data, as highlighted by Challet and Ayed (2013). GSV does not represent an absolute number of searches per keyword, but rather a standardized value ranging from 0-100. Under this methodology, 100 stands for the largest value of the series, which is dependent both on the time series chosen and the other keyword selected. Google allows for up to five simultaneous keyword searches, but the GSV data then standardizes each keyword’s series based on the other four selections. This can be a problem if we simultaneously search for the keywords “futebol” and “CMIG4” in Brazil, for example. Since the nominal search volume for “futebol” is much higher than “CMIG4”, the latter appears to be null compared to the former. Therefore, keywords with a higher search volume nullify any variation in keywords with a smaller volume when both are searched simultaneously to capture GSV; searching for more than one keyword can severely distort historical data.

Another relevant option in Google Trends is the geographical scope of GSV data. Given the relevant foreign investor participation in the Brazilian stock market, we could have opened the queries worldwide. However, considering Bijl et al. (2016) and the polluted preliminary data collected under the worldwide option, we chose to restrict GSV data to Brazilian queries so to

limit the risk of overlapping word meanings across the globe (Bijl et al., 2016). Finally, GSV data are published on Sundays and therefore are only available weekly; thus, Monday is the first trading day for markets to react when using GSV. This is the main reason we use weekly data for all the variables.

Next, perhaps the most crucial decision regarding Google Trends data collection comes in the form of the keywords we choose. While Bijl et al. (2016) chose to search for company names, arguing that tickers are very similar to company abbreviations, we searched for tickers in order to avoid noise in data caused by non-investors simply looking for company products or services. This is easy to imagine for retail companies, such as “Lojas Renner”, “Lojas Americanas”, and “Banco Itaú”, all of which have millions of clients in Brazil. Given that individual investors represent only a small fraction of the Brazilian population, searching for company names might be biased towards reflecting client attention rather than investor attention.

Therefore, we search the tickers of each stock separately for the whole period and restricted the results to Brazil. Since the data is standardized for each search, where the smallest value is zero and the highest value is 100, using the search volume (either raw or in the log form) could distort the data since 100 for one ticker does not have the same meaning as 100 for another ticker. This means that we cannot compare search volumes of different expressions (tickers, in this case), and the absolute number of queries is not available. That is the main reason why we adopted a log-GSV measure to capture the increase and decrease in searches from one week to the other. Therefore, we focused instead on the variation in search volumes from one week to another. In addition, search volumes tend to grow over time; using this variation avoids spurious regression results due to the nonstationary nature of the GSV series.

#### 4. RESULTS

Figure 1 shows the comparative evolution of market prices and GSV for a value-weighted portfolio comprised of the stocks in the sample. In Figure 1, the upward trend for the two samples is clear, although prices reach a minimum in the beginning of 2016, and the GSVs are very small during 2014.

Figure 2 shows a comparison between value-weighted portfolio market returns and GSV log-variation. Neither series seems to present trends, and the GSV is much more volatile than the returns. Finally, Figure 3 compares trading volume with GSV. Only recently do the two series seem to follow similar trends, showing a slight increase, and greater variation in the last weeks of 2018; this is especially true for the GSV series.

Table 1 shows the descriptive statistics for the regression variables. Although very close to zero, the weekly market excess returns are negative, the minimum is close to -50% (JBSS3 in May 2017) and the maximum is close to 50% (MGLU3 in December 2015). The most dramatic decrease (increase) in GSV, more than 400%, is from EQTL3 in September 2016 (CPFE3 in March 2018). In general, GSV is much more volatile than returns, as one can also see in Figure 2, and firms are generally very large.

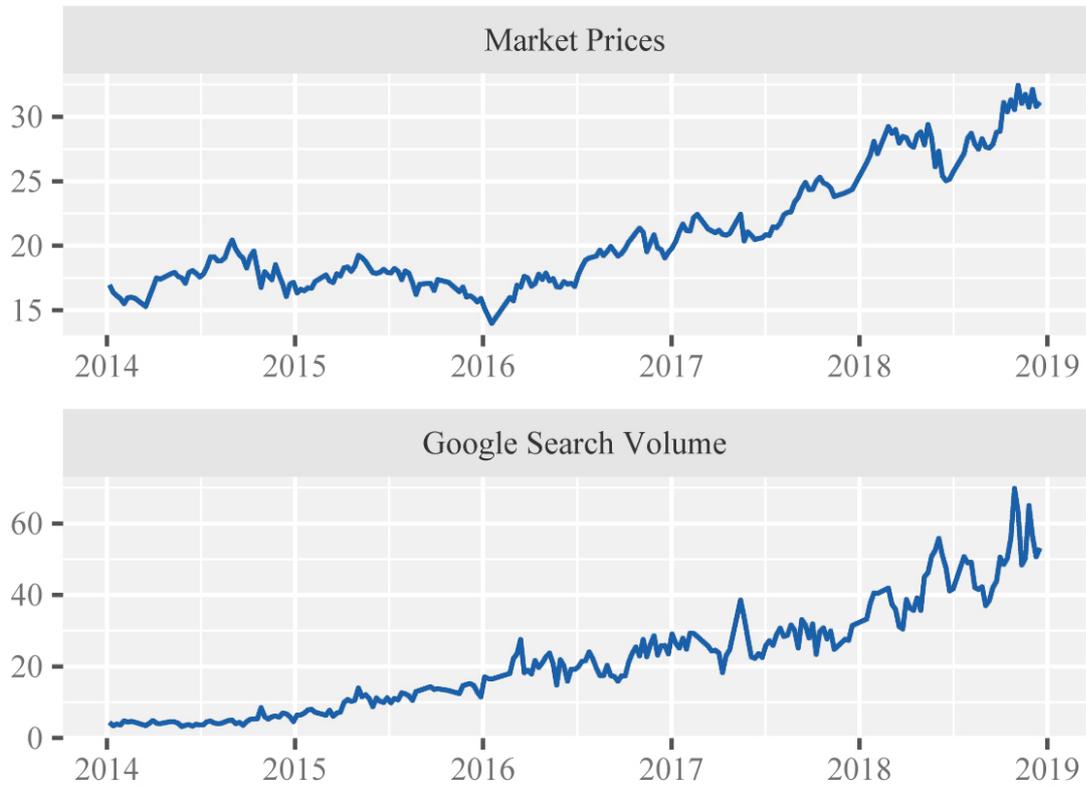


Figure 1. Market Prices Versus GSV  
Source: Research data (2019).

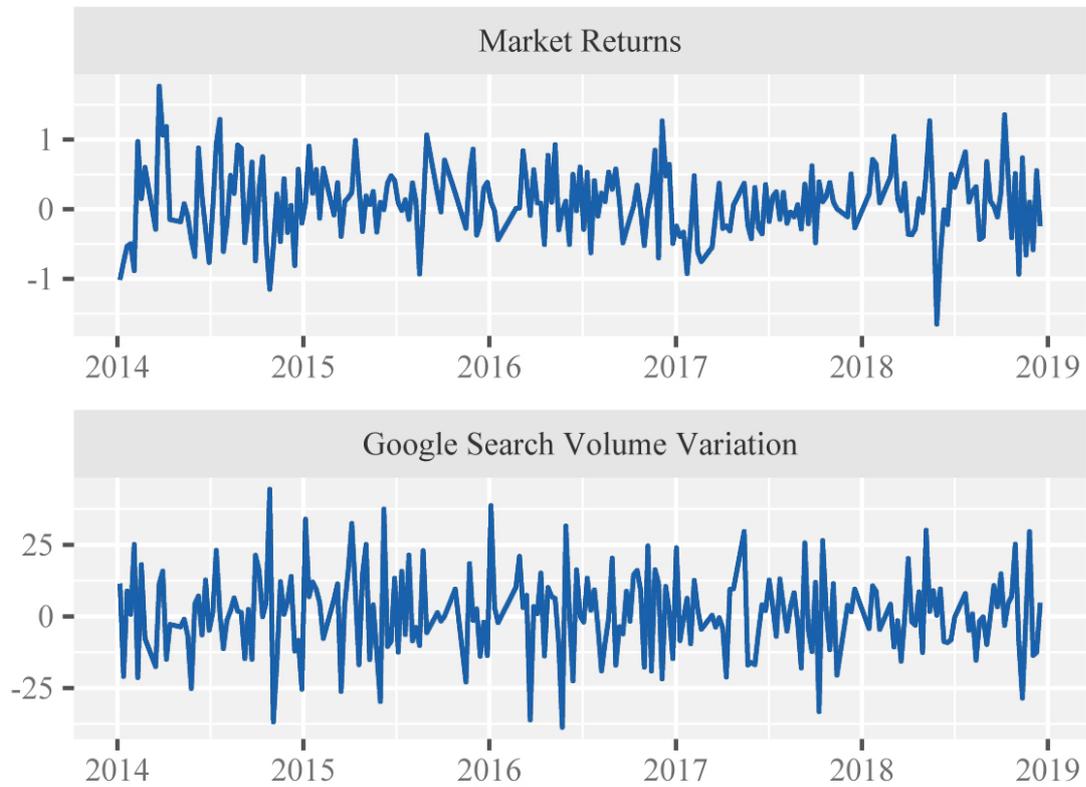
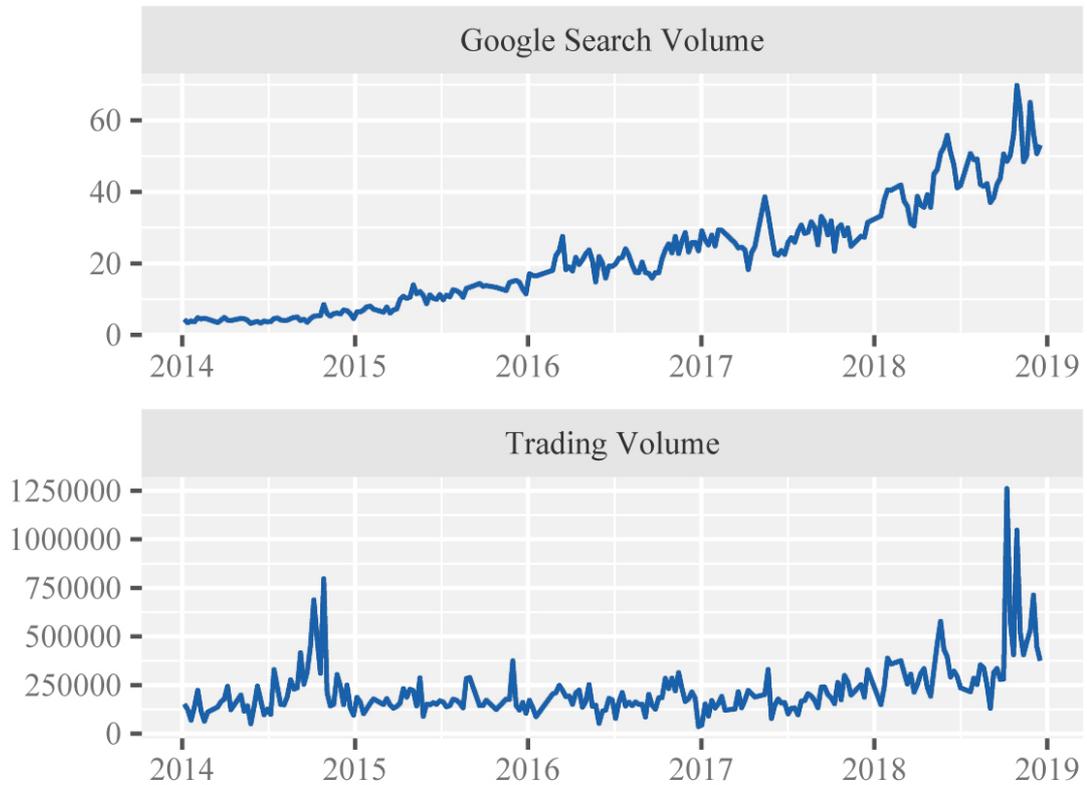


Figure 2. Market Returns Versus GSV Variation  
Source: Research data (2019).



*Figure 3.* Market Volume Versus GSV

*Source:* Research data (2019).

**Table 1**

*Summary Statistics*

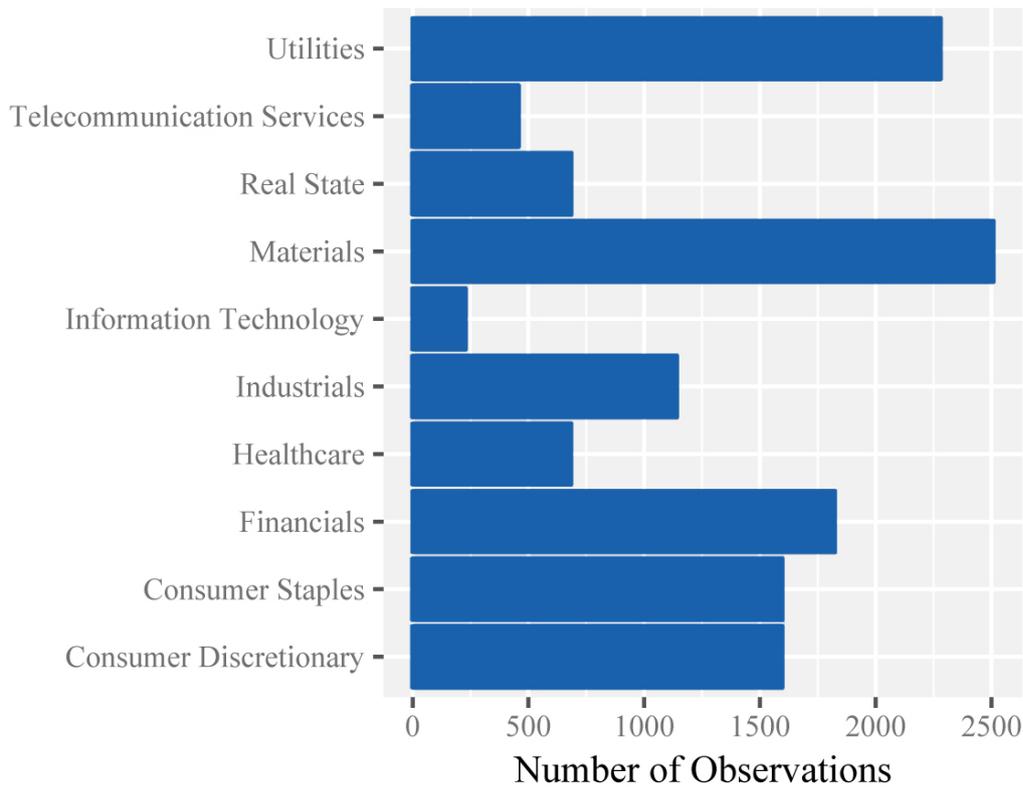
Variable	Stat	Value
Market Excess Returns	Min.	-48.834
	Mean	-0.032
	Max.	46.015
GSV Variation	Min.	-412.713
	Mean	0.926
	Max.	411.087
Trading Volume	Min.	-2.106
	Mean	0.028
	Max.	2.981
Returns Mid-term Volatility	Min.	0.005
	Mean	3.956
	Max.	39.537
Returns Long-term Volatility	Min.	0.296
	Mean	9.299
	Max.	47.868
GSV Mid-term Volatility	Min.	0
	Mean	109.770
	Max.	689.500

**Table 1**  
*Cont.*

Variable	Stat	Value
GSV Long-term Volatility	Min.	0
	Mean	216.600
	Max.	969.600
Size	Min.	12.060
	Mean	16.760
	Max.	19.770

*Source:* Research data (2019).

Figure 4 shows the distribution of the sample among the 10 industry classifications. Materials has the higher number of observations, followed by utilities. Next comes the financial sector, followed by the consumer discretionary and staples sectors. Information technology and telecommunications have the lowest number of observations.



**Figure 4.** Sample Distribution by Industry

*Source:* Research data (2019).

Finally, Table 2 shows the correlation matrix for the variables, including stock prices and the level of GSV. The contemporaneous correlation between market excess returns and GSV variation is close to zero, but prices and the level of GSV have a correlation of slightly more than 10% across the sample. The log-variation of GSV is mainly correlated with its level, trading volume and volatility. Larger firms have larger levels of GSV, but their variation is smaller. Larger firms also have lower returns volatility.

**Table 2**  
*Correlation Matrix*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Price	1	0.042	0.119	-0.004	0.034	-0.168	-0.224	-0.015	-0.022	0.225
(2) Excess Return		1	0.012	-0.002	0.092	0.008	0.011	0.007	0.004	0.027
(3) GSV			1	0.168	0.123	0.094	0.102	-0.297	-0.362	0.144
(4) GSV Var.				1	0.045	-0.034	-0.031	0.024	0.012	-0.002
(5) Volume					1	0.194	0.077	-0.006	-0.015	0.021
(6) Ret. MT Vol.						1	0.671	-0.088	-0.111	-0.111
(7) Ret. LT Vol.							1	-0.125	-0.164	-0.156
(8) GSV MT Vol.								1	0.816	-0.238
(9) GSV LT Vol.									1	-0.290
(10) Size										1

*Source:* Research data (2019).

Table 3 shows the FGLS estimation results for Equation (1). The first and the fifth autoregressive components of returns are statistically significant (the F test shows the five lags are jointly statistically significant at 1%). The first three lags of the log-variation of GSV are negative and statistically significant. All the lags except for the fifth one are negative. Jointly, the first four negative coefficients are also significant at the 1% level according to a F test. Although statistically significant, the effect of investor attention measured by GSV is very small. Considering only the first lag, one percentage point increase in GSV in one week is followed by a decrease of 0.001 percentage points in the returns the following week. Considering the four negative lags, the effect sums up to 0.0035 p.p.. Regarding the control variables, firm size is also a strong component in predicting returns (larger firms in the sample have higher returns). Volatility of returns and of GSV are not statistically significant, and trading volume is only significant for the first lag at the 10% level.

These results are aligned with Bijl et al. (2016), who also found negative effects of GSV on future excess returns. Therefore, the investor attention effect in the Brazilian market is similar to that in the U.S., despite the vast difference in development between the two markets. First, as previously reported, retail investors are much scarcer in Brazil, where historically high interest rates and negative market risk premiums drove individual investors to simple, fixed income products, such as government bonds. As a result, it is harder for individuals to cause significant price pressure on stocks, which tend to follow institutional investor demand (both nationally and internationally) according to the laws of supply and demand. This suggests a diminished predictive power from individual investor attentions as measured by internet queries, as its volumes are driven mostly by low-volume individual investors and not institutional investors who individually drive large trading volumes. Nevertheless, noise trading still seems to be able to play a role in explaining market returns.

As a robustness test, we decided to expand the analysis of GSV impact on Brazilian stocks. Our hypothesis was that investor attention may impact stock returns differently. To this purpose, we evaluated whether more frequently traded firms are more or less sensitive to variations in investor attention. To do that, we expanded Equation (1) to include interactions of GSV with trading volume. This interaction term is a novelty in the GSV and investor attention literature.

**Table 3**  
*Estimation Results*

	Dependent Variable: $r_i^e$
Intercept	-1.7349*** (0.5142)
$r_{t-1}^e$	-0.0272*** (0.0088)
$r_{t-2}^e$	-0.0112 (0.0088)
$r_{t-3}^e$	-0.0090 (0.0087)
$r_{t-4}^e$	0.0100 (0.0087)
$r_{t-5}^e$	0.0249*** (0.0087)
$gsv_{t-1}$	-0.0009*** (0.0003)
$gsv_{t-2}$	-0.0013*** (0.0004)
$gsv_{t-3}$	-0.0009*** (0.0004)
$gsv_{t-4}$	-0.0006 (0.0004)
$gsv_{t-5}$	0.0001 (0.0003)
$vol_{t-1}$	0.0656 (0.0416)
$vol_{t-2}$	0.0731* (0.0423)
$vol_{t-3}$	0.0329 (0.0424)
$vol_{t-4}$	-0.0090 (0.0421)
$vol_{t-5}$	-0.0262 (0.0412)
$\sigma r_{it}^{long}$	0.0112 (0.0082)
$\sigma r_{it}^{mid}$	0.0043 (0.0131)
$\sigma gsv_{it}^{long}$	0.0002 (0.0002)
$\sigma gsv_{it}^{mid}$	0.0002 (0.0003)
Size	0.0990*** (0.0283)
Industry effects	Yes
# stocks, # weeks, #observations	57, 228, 12,991
Wald statistics	66.4408***

**Note:** \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Source:** Research data (2019).

If the interactions are jointly negative (positive) and statistically significant, the negative effect of GSV is reinforced (minimized) the more frequently traded a stock is. Table 4 presents the FGLS estimations results for this model.

**Table 4**  
*Estimation Results: Interactions with Trading Volume*

	Dependent Variable: $r_i^e$
Intercept	-1.7530*** (0.5134)
$r_{t-1}^e$	-0.0274*** (0.0088)
$r_{t-2}^e$	-0.0118 (0.0088)
$r_{t-3}^e$	-0.0098 (0.0087)
$r_{t-4}^e$	0.0106 (0.0087)
$r_{t-5}^e$	0.0241*** (0.0087)
$gsv_{t-1}$	-0.0009*** (0.0003)
$gsv_{t-2}$	-0.0013*** (0.0004)
$gsv_{t-3}$	-0.0008*** (0.0004)
$gsv_{t-4}$	-0.0006 (0.0004)
$gsv_{t-5}$	0.0002 (0.0003)
$vol_{t-1}$	0.0653 (0.0418)
$vol_{t-2}$	0.0703* (0.0425)
$vol_{t-3}$	0.0349 (0.0425)
$vol_{t-4}$	-0.0129 (0.0423)
$vol_{t-5}$	-0.0201 (0.0414)
$\sigma r_{it}^{long}$	0.0122 (0.0083)
$\sigma r_{it}^{mid}$	0.0033 (0.0132)
$\sigma gsv_{it}^{long}$	0.0002 (0.0002)
$\sigma gsv_{it}^{mid}$	0.0001 (0.0003)
Size	0.1001*** (0.0283)

**Table 4**  
*Cont.*

	Dependent Variable: $r_i^e$
$gsv_{t-1} * vol_{t-1}$	0.0000 (0.0005)
$gsv_{t-1} * vol_{t-2}$	-0.0010* (0.0005)
$gsv_{t-1} * vol_{t-3}$	-0.0004 (0.0005)
$gsv_{t-1} * vol_{t-4}$	0.0001 (0.0005)
$gsv_{t-1} * vol_{t-5}$	-0.0016*** (0.0005)
Industry effects	Yes
# stocks, # weeks, #observations	57, 228, 12,991
Wald statistics	79.654***

**Note:** \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Source:** Research data (2019).

The basic results from Table 3 still hold. The GSV lags are negative and jointly statistically significant at 1%. However, while trading volume itself is not significant (the F test shows no joint significance for the five lags), the interactions with GSV are, and, although very small, they are negative, indicating that stocks more heavily traded tend to be more vulnerable to shifts in investor attention measured by internet queries. While we do not consider volume in the analysis, an increase in one p.p. in GSV accrues a decrease in returns of 0.0035 p.p., the effect is only 0.0027 for stocks in the first quartile of trading volume and of 0.0045 for stocks in the third quartile. Therefore, more heavily traded firms are more sensitive to changes in investor attention. This is consistent with the idea that investor attention is more concentrated on more heavily traded firms, implying that investors dedicate a larger part of their time researching stocks that are already under the spotlight of the market.

## 5. CONCLUDING REMARKS

The objective of this paper was to analyze the role of investor attention for predicting future market returns. Following previous research (see, e.g., Bijl et al., 2016; Challet & Ayed, 2013; Kristoufek, 2013; Mayer, 2018; Preis et al., 2013; Preis et al., 2010), we proxied individual investor attention by using internet queries in GSV. We estimated a regression of the returns from 57 stocks against the GSV variation of the previous first to fifth week and controlling for several market and firm effects.

Consistent with previous research in the U.S. market (Bijl et al., 2016), we found that higher investor attention, as measured by increases in GSV, are followed by lower returns and that this is reinforced by higher trading volume. This is consistent with the early ideas of J. Kennedy in the 1920s: that popularity may indicate that an asset has lost touch with its fundamental value, much before the scientific assessment of the impact of noise traders (e.g., De Long, Shleifer, & Summers, 1990).

This research adds to the debate on market prices anomalies and specifically contributes to the assessment of an emerging market, since the mainstream research focuses on developed markets, mainly the U.S. Additionally, our results offer insights on possible investment strategies based on following GSV variations. Future research could focus on this issue, testing for the profitability of trading strategies designed to benefit from given information about GSV, through a long-short portfolio backtesting exercise, for example.

Finally, a common thread woven throughout this paper is that technology should remain relevant in the field of investor attention and investor sentiment; this will impact the quality and availability of tools used to measure these complex yet abstract concepts. In this sense, further studies may very well be better positioned to confirm, expand or challenge the conclusions herein presented and contribute to a virtuous cycle where technology deepens our understanding of complex behavioral phenomena.

## 6. REFERENCES

- Andrei, D., & Hasler, M. (2014). Investor attention and stock market volatility. *The Review of Financial Studies*, 28(1), 33-72.
- Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3), 929-985.
- B3. (2019). Participação dos investidores. <http://www.bmfbovespa.com.br/pt-br/servicos/market-data/consultas/mercado-a-vista/participacao-dos-investidores/volume-total/>
- Bacen. (2018). Sistema de expectativas do mercado. <https://www3.bcb.gov.br/expectativas/publico/consulta/serieestatisticas>
- Barber, B. M., & Odean, T. (2007). 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.
- Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis*, 45, 150-156.
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.
- Campbell, J. Y., Grossman, S. J., & Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, 108(4), 905-939.
- Castro, F. H., Eid Jr, W., Santana, V. F., & Yoshinaga, C. E. (2019). Fifty-year History of the Ibovespa. *Revista Brasileira de Finanças*, 17(3), 47-65.
- Challet, D., & Ayed, A. B. H. (2013). Predicting financial markets with Google Trends and not so random keywords. *arXiv preprint*, 1307.4643. <https://arxiv.org/pdf/1307.4643.pdf>
- Chemmanur, T. J., & Yan, A. (2019). Advertising, attention, and stock returns. *Quarterly Journal of Finance*, 9(3), 1-51.
- Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5), 1461-1499.
- Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2), 221-247.
- De Long, B., Shleifer, A., & Summers, L. (1990). Noise trader risk in financial markets. *The Journal of Political Economy*, 98(4), 703-738.

- Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.
- Fang, L., & Peress, J. (2009). Media coverage and the cross section of stock returns. *The Journal of Finance*, 64(5), 2023–2052.
- Gallup (2016). Just Over Half of American Own Stocks, Matching a Record Low. <https://news.gallup.com/poll/190883/half-americans-own-stocks-matching-record-low.aspx>
- Grullon, G., Kanatas, G., & Weston, J. P. (2004). Advertising, breadth of ownership, and liquidity. *The Review of Financial Studies*, 17(2), 439–461.
- Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific reports*, 3, 3415.
- Li, J., & Yu, J. (2012). Investor attention, psychological anchors, and stock return predictability. *Journal of Financial Economics*, 104(2), 401–419.
- Lou, D. (2014). Attracting investor attention through advertising. *The Review of Financial Studies*, 27(6), 1797–1829.
- Mayer, E. J. (2018). Advertising, investor attention, and stock prices: Evidence from a natural experiment. <http://dx.doi.org/10.2139/ssrn.2468181>
- Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, 3, 1684.
- Preis, T., Reith, D., & Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: Insights from search engine query data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1933), 5707–5719.

#### AUTHOR'S CONTRIBUTION

Both authors contributed equally to this manuscript.

#### ACKNOWLEDGMENTS

The authors would like to thank Verônica Santana for her excellent research assistance and Nathália Gil for helping with the translation into Portuguese. We thank the Editor (Felipe Ramos), the Associate Editor (Luiz Felipe Girão) and two anonymous referees for helpful suggestions and comments. The authors, however, take full responsibility for all remaining errors. Claudia Yoshinaga acknowledges the financial support of FGV EAESP Pesquisa.

#### CONFLICTS OF INTEREST

The authors have no conflicts of interest.

#### COMPLIANCE WITH ETHICAL STANDARDS

This article does not contain any studies with human participants or animals performed by any of the authors.