Brazilian stock market performance and investor sentiment on Twitter

Dyliane Mouri Silva de Souza
*Programa de Pós-Graduação em Ciências Contábeis,
Universidade Federal da Paraíba, João Pessoa, Brazil, and*
Orleans Silva Martins
Departamento de Finanças e Contabilidade, UFPB, João Pessoa, Brazil

Abstract

Purpose – This study identified how investor sentiment on Twitter is associated with Brazilian stock market return and trading volume.

Design/methodology/approach – The study analyzes 314,864 tweets between January 1, 2017, to December 31, 2018, collected with the Tweepy library. The companies' financial data were obtained from Refinitiv Eikon. Using the netnographic method, a Twitter Investor Sentiment Index (ISI) was constructed based on terms associated with the stocks. This Twitter sentiment was attributed through machine learning using the Google Cloud Natural Language API. The associations between Twitter sentiment and market performance were performed using quantile regressions and vector auto-regression (VAR) models, because the variables of interest are heterogeneous and non-normal, even as relationships can be dynamic.

Findings – In the contemporary period, the ISI is positively correlated with stock market returns, but negatively correlated with trading volume. The autoregressive analysis did not confirm the expectation of a dynamic relationship between sentiment and market variables. The quantile analysis showed that the ISI explains the stock market return, however, only at times of lower returns. It is possible to state that this effect is due to the informational content of the tweets (sentiment), and not to the volume of tweets.

Originality/value – The study presents unprecedented evidence for the Brazilian market that investor sentiment can be identified on Twitter, and that this sentiment can be useful for the formation of an investment strategy, especially in times of lower returns. These findings are original and relevant to market agents, such as investors, managers and regulators, as they can be used to obtain abnormal returns.

Keywords Investor rationality, Investment strategy, Stock market

Paper type Research paper

1. Introduction

The efficient market hypothesis (EMH) supports that variations of stock prices for a long time were associated to a combination of historical prices with information environment (Fama, 1970, 1991). However, over the years, researchers began to question whether the investors' behavioral and rational elements would also be responsible for such variations (Menkhoff, 1998). Generally, these elements are called "investor sentiment" or "investor mood."

The finance literature reports a strong link between measures of investor sentiment (or mood) and stock market return (Lee, Jiang, & Indro, 2002; Brown & Cliff, 2005; Baker & Wurgler, 2006, 2007). For example, Kearney and Liu (2014) summarize important and influential findings of the association between investor sentiment and individual, firm and

market behaviors and performances, with different empirical models and forms of analysis, especially concerning textual sentiment. But the growing interaction among investors on the Internet has offered the opportunity to expand this literature, especially in emerging countries.

Exploring new ways of capturing investor sentiment or mood is important because most studies use only two sources to measure sentiment: opinion polls (Brown & Cliff, 2005) and market variables (Baker & Wurgler, 2006). In this sense, in the last few years, a series of researchers have turned their attention to capture the investor sentiment through online activities, since this type of approach instantly captures the sentiment, and this addresses criticism like that of Klibanoff, Lamont, and Wizman (1998), that a large part of the studies that examine investor behavior are retrospective and analyze only its determinants.

One of the fastest and most efficient ways to capture investors' sentiment or mood in online activities is through social network. The preceding literature has evidence that points to the power of social networks such as Twitter to identify sentiment, in addition to being useful for predicting stock market movements (Bollen, Mao, & Zeng, 2011; Mao, Wei, Wang, & Liu, 2012; Fan & Gordon, 2014; Wei, Mao, & Wang, 2016). However, Agarwal, Kumar and Goel (2019) highlight that most of this evidence is restricted to developed markets, with a scarcity of studies exploring emerging markets.

This gap is even greater in the Brazilian emerging market, the largest stock market in Latin America, which has seen the number of individual investors on the stock exchange increase from around 580,000 to four million in the last ten years (B3, 2021). Parallel to this growth, there was also the popularization of Twitter among investors, giving rise to the so-called "FinTwit," an acronym originated from finance and Twitter, which came to attract the attention of Brazilian financial media, due to its importance in representing the financial market mood, as in this community there is great interaction among investors, managers, analysts and other market participants (Tauhata, 2021).

Twitter is one of the key ways to obtain information about the market. This happens because, according to Wei et al. (2016), in this social network there are many investors, financial analysts, entities and news profiles, which usually post messages about what happens in the market. In this context, Mao, Counts and Bollen (2011) observed that the investor sentiment extracted from Twitter, as well as the number of tweets, can be used to forecast the daily return of the stock market. Despite the existence of some previous studies, little is known about these associations in emerging markets, especially in Brazil, which has experienced growth in the number of investors and their use of Twitter. This is the motivation of this study.

Agarwal et al. (2019) point that most studies on investor sentiment through online sources focus just on developed markets. For them, there is a research gap in the international literature when the subject involves investors' activities on the Internet and their impact on stock prices in emerging markets. And notably, emerging markets have more fragile information environments and bigger problems with information asymmetry, which leads to lower levels of market efficiency (Martins & Barros, 2021).

The objective of this study is to identify the investor sentiment on Twitter and its association with Brazilian stock market return and trading volume. Our findings demonstrate that investor sentiment is positively associated with the stock market movements, looking at stock market return and trading volume. This association is especially important in the contemporary period of these variables, and it varies over time. The positive association between volume of tweets and trading volume is especially important when the investor sentiment is negative.

Our findings also indicate that, although there is no dynamic relationship between Twitter sentiment and market performance, there is evidence of a significant association between this sentiment and stock market returns, especially in periods of lower returns. This association is

not limited to the volume of tweets published on the social network, but to the informational content of the messages, which reveals investor sentiment.

These findings are useful for market participants to define investment strategies and to monitor possible market manipulations. This is a relevant contribution because it suggests that the growth of investor interactions on the Internet can reduce information asymmetry and increase the market efficiency level, converging with Renault (2017) on the so-called more realistic model of market efficiency. This is especially important in emerging markets such as Brazil, which generally have lower information efficiency and legal protection for investors (Martins & Barros, 2021).

2. Development of hypotheses

The belief that markets predictability through online activities would be far from the efficient market hypothesis (EMH) has changed with the growth in the volume and importance of information sharing on the Internet and social networks (Baker & Wurgler, 2007; Bollen *et al.*, 2011; Agarwal *et al.*, 2019; Chen & Lo, 2019). For example, for Renault (2017), this would just be a more realistic model of market efficiency, like an EMH model in its semi-strong form.

The desire to understand the process of forming the asset pricing culminated in what is currently considered the EMH. This hypothesis brought with it the idea that prices are explained by information, so that the informational efficiency of a market is what will define the price of an asset, converging with Fama's (1970, 1991) view that predictability is possible in times of more online activity, even at low level, and ratified by Chen and Lo (2019). In this way, we consider that abnormal returns obtained in the market are rewards for the time and money spent by investors in monitoring a wide variety of information sources.

Over the years, market efficiency has been extensively tested (Malkiel, 2003). At the same time, studies that relate EMH to investor rationality have gained space, especially those that deal with behavioral finance and the investment decision process. Kahneman and Tversky (1979) show that human beings are endowed with limited rationality so that behavioral and psychological aspects influence the decision-making process. Thus, decisions made by market participants are not entirely based on reason – there is also emotion, and the social network is a meeting place for investors' emotions.

Since there are indications that the market is not entirely rational and that individual reactions may affect it, Brown and Cliff (2005) and Baker and Wurgler (2006, 2007) investigated how sentiment relates to stock returns, bringing evidence that they are indeed strongly associated, and that a high level of optimism in contemporary times causes current prices to be high, leading to low future returns. In this sense, Kim, Ruy and Seo (2014) demonstrate that investors should sell stocks that are under the focus of analysts, who "promise" high rates of earnings per share growth in the long run. For these authors, this strategy yields better returns than the buy-and-hold strategy.

We can suspect that investor sentiment is sensitive to stock market performance, as part of its decision-making process is guided by emotions, which can create a dynamic relationship between market performance and investor sentiment. In addition, little is known about this relationship in emerging markets such as Brazil, especially given the growth in the numbers of investors and the interactions of these investors on Twitter. Thus, our first hypothesis is:

H1. There is a positive and dynamic association between stock market performance and investor sentiment on Twitter in Brazil.

Investor sentiment is seen by literature as a determining factor in market movements. Faced with this and given criticisms such as that of Baker and Wurgler (2007), more traditional ways of capturing sentiments like opinion polls and market variables (Brown & Cliff, 2005;

Baker & Wurgler, 2006) have given way to capturing sentiment in online environments, especially through social networks, to get to know the market and capture investor sentiment directly and instantly (Fan & Gordon, 2014). A series of surveys have emerged intending to capture investor sentiment from the Internet, in which sentiment indices have been developed through websites, online newspapers, research tools and social networks.

Concerning the search for sentiment through websites, Kim and Kim (2014) investigated messages from Yahoo! Finance. Results show that sentiment does not predict future returns, volatility or volume, although present returns influence future sentiment. On the other hand, Silva (2017) and Galdi and Gonçalves (2018) found, through Valor Econômico newspaper website, a strong association between contemporary investor sentiment and contemporary return, with a trend of reversal over the days. These studies also found that Brazilian market rates negative and uncertain words with higher weight, which creates a stronger (negative) association between investor sentiment and market returns.

Mao et al. (2012) analyzed Twitter and identified that investor sentiment correlates significantly with various financial indicators of stocks, such as trading volume and performance indicators, and that tweets can be used to predict stock prices. In this line, Wei et al. (2016) found a strong association between volume of posted messages and stock return, in addition to attesting that using a strategy based on the perception of a market bullish helps to obtain greater returns.

Nisar and Yeung (2018) point out that investor sentiment on Twitter and stock prices are related. However, the volume of tweets does not appear to be associated with prices or turnover. Similarly, Oliveira, Cortez, and Areal (2017) found that the sentiment index extracted from Twitter was able to predict returns on S&P500 index, although it had low explanatory power in forecasting turnover and volatility. Thus, although for Nisar and Yeung (2018) the volume of tweets is not related to variables other than returns, for Mao et al. (2012), Wei et al. (2016) and Oliveira et al. (2017), the sentiment is useful in predicting the volume traded, making the second hypothesis emerge:

H2. Investor sentiment on Twitter is positively associated with stock market returns in Brazil, explaining part of such returns.

Regarding the analysis of sentiment by tweets, different studies used tweets from the social network StockTwits, demonstrating that forming portfolios based on a sentiment index is useful in gauging abnormal returns (Renault, 2017; Al-Nasseri & Ali, 2018). In this sense, Renault (2017) reveals that the exchange-traded fund (ETF) can be used to mirror S&P500 index and that the first half-hour in investor sentiment is useful to predict the return of the last half-hour of S&P500 ETF. Still, Al-Nasseri and Ali (2018) note that the higher the value of the sentiment index, the greater the trading volume.

A new way of capturing sentiment is through Google's search tool (Chen & Lo, 2019; Kim, Lučivjanská, Molnár, & Villa, 2019). Studies found no relationship between the volume of search on Google and the stock return; however, strong relationships were found with the volatility of the return, the trading volume and the turnover rate of the shares. So, in addition to the expectation of positive association between market performance and investor sentiment (Brown & Cliff, 2005; Baker & Wurgler, 2006, 2007), literature reports that the return on assets is affected by factors such as market risk, size, value, momentum and liquidity (Fama & French, 1993; Carhart, 1997; Keene & Peterson, 2007; Machado & Medeiros, 2011).

Lee et al. (2002) reveal that sentiment is a risk factor, having a linear relationship with market return. Some studies have found factors that could explain the average returns provided by market betas. These factors are company size, market value, leverage, book-to-market and earning/price ratios. Fama and French (1993) test these factors and find that the beta is not the only way to explain returns, as size and book-to-market also play a role, being a way to improve the explanatory power of the CAPM (capital asset pricing model).

Carhart (1997) added to Fama and French's (1993) three factors a further factor that captures the anomaly of the momentum effect. Also, Keene and Peterson (2007) added one more factor to the Fama–French model, liquidity, thus creating a five-factor model, which constitutes the model that best explains the stock return in the Brazilian context (Machado & Medeiros, 2011). In this sense, we believe that although sentiment is associated with returns, it is necessary to consider factors related to the market.

3. Method

To identify investor sentiment on Twitter, we used the model of Bollen *et al.* (2011) as the basis, using machine learning to classify the polarity of words through the "Google Cloud Natural Language API," where parameters vary from -1 to 1, with 1 being the extremely optimistic sentiment, -1 the extremely pessimistic sentiment and 0 neutral. We draw on Bollen *et al.* (2011) because they already have a sentiment index on Twitter that is robustly validated, in addition to being related to market movements in the US. Therefore, it is not our interest to analyze the efficiency of measuring sentiment, but rather to analyze the association between Twitter sentiment, market return and trading volume.

The investor sentiment on Twitter is calculated according to Equation (1), where, $Sent_{p,d}$ is the investor sentiment on Twitter for asset p on day d; $\sum \text{WPW}_d$ is the weight of positive words on day d; $\sum \text{WNW}_d$ is the weight of negative words on day d; $\sum \text{WPNW}_d$ is the sum of the weights of positive and negative words.

$$Sent_{p,d} = \frac{\sum \text{WPW}_d - \sum \text{WNW}_d}{\sum \text{WPNW}_d} \tag{1}$$

The tweets were collected through Tweepy library with the following parameters: (1) are in Portuguese and (2) contain terms referring to stocks or market indices. These terms were raised using the netnographic method, in which, for two months, there was participant observation aimed at identifying the behavior of individuals on Twitter, detecting patterns and how they usually refer to the market and the stocks that are part of Ibovespa. The period of the study was January 1, 2017, to December 31, 2018.

Based on this, the terms that represent each company were raised to capture the Twitter sentiment about the firms, for example, for Magazine Luiza S/A: MGLU3, \$MGLU, #MGLU and #MGLU3. We also looked for terms that refer to the Brazilian market as a whole, such as #BVMF3, \$BVMF, BVMF3, BVMF, B3SA3, BM&Fbovespa, B3, IBOV, IBOVESPA, Bolsa Brasileira, Brasil Bolsa Balcão and Bolsa de São Paulo. This search returned 1,195,575 tweets that were treated to remove links, tweets that mentioned YouTube (when they were liked to videos) and repeated tweets (they were not retweets or posted on different dates), reaching a final sample of 314,864 tweets.

These tweets were classified as positive ($+0.5$) and negative (-0.5) based on algorithm attribution, and then the daily Twitter Investor Sentiment Index ($ISI_{p,d}$) was constructed by accumulating the sentiments from up-to-date tweets. On days when there was no trading on the Brazilian stock exchange, tweets were added to the following trading day. So, the daily index is the average of the sentiment from the tweets for each trading day d, according to Equation (2).

$$ISI_{p,d} = \frac{1}{n} \sum_{i=1}^{n} Sent_{p,d} \tag{2}$$

Following the rationale of Renault (2017) in another market, we use the Ibovespa ETF (BOVA11), an asset that represents the main market index of the Brazilian stock exchange

and that has the largest trading volume in this market, to represent the Brazilian market movements. The BOVA11 return was calculated like in Equation (3), where $Ret_{p,d}$ is the observed return of portfolio p on day d, and $P_{p,d}$ and $P_{p,d-1}$ are the prices of BOVA11 on days d and $d-1$. The trading volume ($Vol_{p,d}$) is the total trades with BOVA11 on day d, as shown by Equation (4).

$$Ret_{p,d} = ln\left(\frac{P_{p,d}}{P_{p,d-1}}\right) \tag{3}$$

$$Vol_{p,d} = \sum_{i=1}^{d} Trades \tag{4}$$

All financial data were collected from Refinitiv Eikon.

3.1 Estimation models

To analyze the association between market performance and investor sentiment on Twitter, vector autoregression (VAR) models were used. The preceding literature provides evidence that the relationship between these phenomena is dynamic, especially when related to stock returns (Zhang, Deng, & Yang, 2010). Also, following Baker and Wurgler's (2006) rationale that macroeconomic factors affect investor sentiment, Equation (5) includes a set of macroeconomic variables.

$$r_d = \phi_0 + \Phi_1 r_{d-1} + \ldots \ \Phi_n r_{d-n} + \Psi_d + a_d \tag{5}$$

where r_d is a vector of k endogenous variables, ϕ_0 is a vector of constants, Φ_j is $k \times k$ matrices to be estimated for $d = 1, \ldots, n$, Ψ_d is a vector of macroeconomic variables, and a_d is white noise formed by a sequence of independent and identically distributed random vectors. Before these estimations, we conducted tests for the existence of a unit root in the variables, and the augmented Dickey-Fuller (ADF) test revealed its absence in all variables. The Jarque–Bera normality test revealed a non-normal distribution of residuals; however, the standard errors of the impulse response functions (IRF) are calculated via bootstrap.

Complementarily, to analyze the association of Twitter sentiment with the performance of the Brazilian stock market, we used quantile regression models. This type of model is adequate due to the absence of normality in the distribution of the series. So, we estimate quantile regressions in five different quantiles (0.10, 0.25, 0.50, 0.75 and 0.90) to explore this association in more depth, whether on days of high or low returns, especially considering that initial descriptive statistics point to an association that changes with the changing mood of the market.

Based on EMH semi-strong form (Fama, 1970, 1991), we analyzed the association of market performance with the Twitter sentiment using a 5-factor model, starting from the 4-factor model of Carhart (1997), plus the liquidity factor (IML – *illiquid-minus-liquid*) of Keene and Pedersen (2007). Since tweets are seen as a type of information, through Equation (6), we tested whether investor sentiment on Twitter has informational content to explain part of the Brazilian stock market performance.

$$\begin{aligned} Q_\tau(Ret_{p,d}) = \ &\beta_0(\tau) + \ \beta_1(\tau)ISI_{p,d} + \beta_2(\tau)MKT_{p,d} + \beta_3(\tau)SMB_{p,d} + \ \beta_4(\tau)HML_{p,d} \\ &+ \ \beta_5(\tau)WML_{p,d} + \ \beta_6(\tau)IML_{p,d} + \varepsilon_d \end{aligned} \tag{6}$$

where $Ret_{p,d}$ is the observed return of BOVA11 on day d, $ISI_{p,d}$ is the daily ISI related to BOVA11 on day d, $MKT_{p,d}$ is the market factor, $SMB_{p,d}$ is the size factor, $HML_{p,d}$ is the value

factor, $WML_{p,d}$ is the momentum factor and $IML_{p,d}$ is the liquidity factor. For Mao *et al.* (2012) and Oliveira *et al.* (2017), the volume of daily tweets is related to the market return. Therefore, we perform a robustness test in which we modified Equation (6) and changed the $ISI_{p,d}$ to the simple volume of tweets, which does not distinguish the sentiment or mood of the messages, as per Equation (7).

$$Q_\tau(Ret_{p,d}) = \beta_0(\tau) + \beta_1(\tau)TVT_{p,d} + \beta_2(\tau)MKT_{p,d} + \beta_3(\tau)SMB_{p,d} + \beta_4(\tau)HML_{p,d}$$
$$+ \beta_5(\tau)WML_{p,d} + \beta_6(\tau)IML_{p,d} + \varepsilon_d \qquad (7)$$

where $TVT_{p,d}$ is the total volume of tweets of portfolio p on day d.

4. Findings

Table 1 presents the descriptive statistics. Concerning trading volume (Vol), we note that the daily average was 3,672,363, with a strong heterogeneity among days. This trading volume is still low when compared to the daily average of trades (658,021,632) in a developed country like the United Kingdom (Nisar & Yeung, 2018), for example. Another relevant difference is the standard deviation of these volumes (average of 8% in the United Kingdom and 31% in Brazil).

The daily volumes of positive (VPT) and negative (VNT) tweets have similar behaviors since the average of positive messages was 243 (min. = 86 and max. = 1,136) and the average of negative tweets was 117 (min. = 32 and max. = 2,688). In total, we observed an average of 640 tweets per day, with a range from a minimum of 213 to a maximum of 5,281. In addition, the highest average of positive tweets reinforces our proxy's divergences from Silva (2017) and Galdi and Gonçalves (2018).

We note that the investor sentiment on Twitter (0.0339) and the daily return of BOVA11 (on average 0.0855%) are positive, which is associated with the positive performance of the Brazilian market during the analyzed period. When the sentiment is bigger, investors become more optimistic about investing in the market.

This finding is different from Silva (2017) and Galdi and Gonçalves (2018) when investigating the Brazilian market. This difference is explained by the distinction between the proxies used to capture the sentiment, and by the period of analysis, since the studies cited

Statistic	Mean	Median	SD	Min	Max
Vol	3,672,363	3,463,024	1,144,582	833,734	11,388,420
VPT	243.3096	224.0000	102.7091	86.0000	1,136.0000
VNT	117.3686	89.0000	175.3354	32.0000	2,688.0000
VTT	639.7495	575.0000	357.0173	213.0000	5,281.0000
IS	0.033866	0.036085	0.037105	−0.177825	0.189162
Ret	0.000855	0.000938	0.013029	−0.087995	0.045704
MKT	0.000470	0.000584	0.011827	−0.088402	0.044613
SMB	0.000446	0.000665	0.007399	−0.039959	0.027848
HML	0.000787	0.000762	0.007814	−0.056169	0.021858
WML	0.000563	0.000787	0.007337	−0.058128	0.023911
IML	0.000561	0.000718	0.007804	−0.030943	0.036575

Note(s): Vol is the trading volume, VPT is the volume of positive tweets, VNT is the volume of negative tweets, VTT is the total volume of tweets, ISI is the Twitter Investor Sentiment Index, Ret is the market return, MKT is the market factor, SMB is the size factor, HML is the value factor, WML is the momentum factor, IML is the liquidity factor

Table 1.
Descriptive statistics of the analyzed variables

only analyzed the market until June 2017, with previous years in which the Brazilian market had the worst performance, whereas this study encompasses a period known as the bull market after the impeachment of former President Dilma Rousseff.

The positive trend of the Brazilian market in this period can also be observed through the statistics of the traditional 4-factors of Carhart (1997) and the liquidity factor of Keene and Pedersen (2007). The averages and medians of portfolio returns for each of these five factors are positive. The market factor (MKT) had an average daily return of 0.0470%, the size factor (SMB) had an average daily return of 0.0446%, the value factor (HML) had an average of 0.0787%, the momentum factor had an average of 0.0563% and the liquidity factor had an average of 0.0561%. These findings converge with Machado and Medeiros (2011), which point out their importance to explain the pricing of assets in the Brazilian stock market.

Table 2 presents the correlation matrix between the variables of interest. The main finding points to a positive and significant correlation (0.1064, $p < 0.05$) between Twitter sentiment (ISI) and stock market return (Ret). This evidence converges with Bollen *et al.* (2011) and Nisar and Yeung (2018). However, unlike those studies, it reveals statistical significance.

The ISI also showed a positive and significant association (0.1057, $p < 0.05$) with the volume of positive tweets (VPT). On good days on the stock market, investors are happier on the social network. However, the association with the daily trading volume on the stock exchange was negative and significant (-0.2255, $p < 0.01$), which reveals that the sentiment was more positive on days of less trade.

4.1 Dynamic relationship between market performance and investor sentiment

Going beyond the initial evidence of a positive correlation between investor sentiment and the Brazilian market performance, we assume that this relationship is dynamic (Zhang *et al.*, 2010) and analyze the relationships between ISI, return and market trade volume through VAR models with up to 2 lags, based on Akaike and Schwarz criteria.

Table 3 shows the VAR models to test the dynamic relationship. Preceding literature documents that in major markets investor sentiment on Twitter can be a useful tool for forecasting market movements (Nisar & Yeung, 2018). However, in Brazil we realize that this evidence is observable only in the contemporary period, given that the correlation on day d was positive (according to Table 2), but the coefficient was low.

Furthermore, in plan A in Table 3, we can see that past ISI (with 1 or 2 lags) does not explain the current performance of the stock market ($p > 0.10$) through BOVA11. Likewise, in plan B, it was not possible to identify an association of past return (with 1 and 2 lags) with the current Twitter sentiment ($p > 0.10$). None of these associations were statistically significant. Thus, observing the findings of Nisar and Yeung (2018) that this association exists for a short window, we can derive these findings onto Brazil, adding that here we were able to identify statistical significance, but only for the association on the same day d.

Still, in plan B, we note that there is a relationship of persistence in Twitter sentiment since the Twitter SI lagged in 1 and 2 days was significant to explain the contemporary ISI ($p < 0.10$). Plan C in Table 3 presents the relationships between trading volume (Vol), ISI and return. We noticed that the return of the previous 2 days has a negative association with the current Vol ($p < 0.05$), as well as the past Vol (with 1 and 2 lags) has a positive association with the current Vol ($p < 0.01$).

Additionally, we could verify that inflation affected all variables of interest, being negatively related to Twitter sentiment ($p < 0.05$) and positively related to Vol ($p < 0.10$). GDP negatively affected investor sentiment ($p < 0.01$) and positively impacted Vol ($p < 0.05$). Exchange rate positively affected Vol ($p < 0.01$), and interest rate did not show significant relationship with the interest variables.

	Ret	Vol	ISI	VTT	VPT	VNT	MKT	SMB	HML	WML
Vol	−0.0352									
ISI	0.1064**	−0.2255***								
VTT	0.0432	0.1653***	−0.2726***							
VPT	0.0930*	0.0747*	0.1057**	0.8436***						
VNT	−0.0613	0.3130***	−0.6698***	0.7262***	0.4611***					
MKT	0.9862***	−0.0267	0.0980**	0.0504	0.0948**	−0.0576				
SMB	−0.1488***	−0.1752***	0.0837*	−0.0783*	−0.0334	−0.1094**	−0.1109**			
HML	0.4257***	−0.1542***	0.1516***	−0.0614	−0.0042	−0.1249***	0.4128***	0.3575***		
WML	0.2749***	−0.1473***	0.0682	−0.0795*	−0.0633	−0.1354***	0.2737***	0.0613	0.3279***	
IML	−0.2399***	−0.1465***	0.0854*	−0.0527	−0.0289	−0.0974**	−0.2009***	0.8975***	0.3113***	0.0393

Note(s): *** is significant at 1%, ** at 5% and * at 10%

Table 2.
Correlation matrix

Variable	Coefficient	Std. Error	Stat. z	$p > z$	[95% conf	Interval]
Panel A: return						
ISI_lag1	−0.0433	0.0288	−1.5100	0.1320	−0.0998	0.0131
ISI_lag2	−0.0027	0.0201	−0.1400	0.8920	−0.0420	0.0366
Ret_lag1	−0.0122	0.0572	−0.2100	0.8320	−0.1243	0.1000
Ret_lag2	−0.0449	0.0596	−0.7500	0.4510	−0.1618	0.0719
Vol_lag1	0.0004	0.0036	0.1100	0.9100	−0.0066	0.0074
Vol_lag2	−0.0074	0.0033	−2.2900	*0.0220*	−0.0138	−0.0011
Inflation	−0.0073	0.0034	−2.1100	*0.0350*	−0.0140	−0.0005
Interest rate	9.9258	0.7118	1.3000	0.1930	−0.4693	2.3209
Exchange rate	0.0032	0.0040	0.8000	0.4220	−0.0046	0.0111
GDP	−0.0006	0.0011	−0.5200	0.6000	−0.0028	0.0016
Constant	0.1152	0.0537	2.1400	*0.0320*	0.0099	0.2205
Panel B: investor sentiments on Twitter						
ISI_lag1	0.2251	0.0637	3.5300	*0.0000*	0.1002	0.3499
ISI_lag2	0.0860	0.0444	1.9400	*0.0530*	−0.0010	0.1731
Ret_lag1	0.0139	0.1267	0.1100	0.9130	−0.2344	0.2621
Ret_lag2	0.0176	0.1320	0.1300	0.8940	−0.2411	0.2762
Vol_lag1	0.0032	0.0079	0.4000	0.6890	−0.0124	0.0187
Vol_lag2	0.0005	0.0072	0.0700	0.9420	−0.0136	0.0146
Inflation	−0.0254	0.0076	−3.3300	*0.0010*	−0.0404	−0.0105
Interest rate	1.5853	1.5756	1.0100	0.3140	−1.5027	4.6733
Exchange rate	−0.0062	0.0089	−0.7000	0.4820	−0.0236	0.0111
GDP	−0.0085	0.0025	−3.3700	*0.0010*	−0.0134	−0.0035
Constant	0.0839	0.1190	0.7100	0.4810	−0.1493	0.3170
Panel C: trading volume						
ISI_lag1	0.1466	0.4811	0.3000	0.7610	−0.7964	1.0896
ISI_lag2	−0.0539	0.3354	−0.1600	0.8720	−0.7113	0.6035
Ret_lag1	−1.4280	0.9567	−1.4900	0.1360	−3.3030	0.4471
Ret_lag2	−2.3710	0.9966	−2.3800	*0.0170*	−4.3243	−0.4178
Vol_lag1	0.3392	0.0599	5.6600	*0.0000*	0.2218	0.4565
Vol_lag2	0.1419	0.0544	2.6100	*0.0090*	0.0353	0.2484
Inflation	0.1086	0.0576	1.8800	*0.0600*	−0.0044	0.2215
Interest rate	−7.2747	11.8989	−0.6100	0.5410	−30.5961	16.0466
Exchange rate	0.2270	0.0669	3.3900	*0.0010*	0.0959	0.3581
GDP	0.0408	0.0189	2.1600	*0.0310*	0.0037	0.0779
Constant	6.6815	0.8984	7.4400	*0.0000*	4.9207	8.4422
No. of Obs	270		*Equation*	*R-sq*	χ^2	$p > \chi^2$
Log-likelihood	1440.7250		IS	0.1878	62.4220	0.0000
AIC	−10.4276		Ret	0.0515	14.6714	0.1445
HQIC	−10.2510		Vol	0.3750	162.0310	0.0000
SBIC	−9.9878					

Note(s): ISI is the Twitter Investor Sentiment Index, Ret is the market return, Vol is the trading volume, GDP is the gross domestic product, $p < 0.01$ is significant at 1%, $p < 0.05$ is significant at 5% and $p < 0.10$ is significant at 10%

Table 3.
VAR models for the
variables of interest

To verify the relationship more accurately between the variables of interest, we analyzed the impulse-response using the impulse response function (IRF) and the forecast error variance decomposition (FEVD). Due to extent constraints, instead of presenting the graphical analysis we chose to present the accumulated impulse-response over 10 days. In each case, an impulse is defined as an 1% increase in ISI.

Table 4 shows that an 1% increase in current ISI has a positive response on future ISI, with bigger response on day $d+1$ (+0.0061%) and a decreasing effect over time. The return

Steps	ISI		Return		Volatility	
	IRF	FEVD	IRF	FEVD	IRF	FEVD
1	0.0061	1.0000	−0.0012	0.0192	−0.0132	0.0437
2	0.0037	0.9995	−0.0001	0.0287	−0.0136	0.0423
3	0.0013	0.9992	<0.0000	0.0283	−0.0032	0.0421
4	0.0006	0.9990	<0.0001	0.0283	−0.0028	0.0414
5	0.0002	0.9990	<0.0000	0.0283	−0.0014	0.0410
6	0.0001	0.9989	<0.0001	0.0283	−0.0009	0.0409
7	<0.0001	0.9989	<0.0001	0.0283	−0.0005	0.0409
8	<0.0001	0.9989	<0.0001	0.0283	−0.0003	0.0408
9	<0.0001	0.9989	<0.0001	0.0283	−0.0002	0.0408
10	<0.0001	0.9989	<0.0001	0.0283	−0.0001	0.0408

Note(s): ISI is the Twitter Investor Sentiment Index, IRF is the impulse response function, FEVD is forecast error variance decomposition

Table 4.
Impulse-response effect of the ISI on the variables of interest

response is negative (−0.0012%) on $d+1$, remaining close to zero from $d+2$ onward. The volatility response is also negative (−0.0132%) on $d+1$, reaching −0.0136% on $d+2$, however, with a tendency to zero over the following days. The FEVD shows the percentages of the forecast error variance for each variable, based on the VAR models presented in Table 3.

We also investigate whether there is Granger-causality between the main variables of our models, according to Table 5. A variable X Granger causes another variable Y if (given the past values of Y) past values of X are useful for predicting Y. For that, we regress Y on its own lagged values and lagged values of X. The null hypothesis of Granger-causality is equivalent to saying that the estimated coefficients on the lagged values of X are jointly zero.

Table 5 presents the chi-square statistics for our sample, showing that there is no causal relationship between ISI and return or trading volume. However, we can see that lagged values of trading volume cause the return of asset p on day d ($p < 0.05$). In the third plane of Table 5, we can also see that lagged values of return cause the trading volume of asset p on day d ($p < 0.05$). Therefore, we can verify that there is a bi-causal relationship between return and trading volume, as the null hypothesis of Granger's non-causality can be rejected at 5%.

Regarding the positive and dynamic relationship we assume between market performance and Twitter sentiment, our Hypothesis 1 was not confirmed. Despite the positive correlation identified, it was low (0.1064). In addition, we did not identify an association between these lagged variables, or even causality between them.

Equation	Excluded	χ^2	df	$p > \chi^2$
ISI	Ret	0.0269	2	0.9870
ISI	Vol	0.2258	2	0.8930
ISI	ALL	0.2419	4	0.9930
Ret	IS	2.6972	2	0.2600
Ret	Vol	6.0341	2	*0.0490*
Ret	ALL	8.2101	4	*0.0840*
Vol	ISI	0.0972	2	0.9530
Vol	Ret	7.2110	2	*0.0270*
Vol	ALL	7.2419	4	0.1240

Note(s): $p < 0.01$ is significant at 1%, $p < 0.05$ is significant at 5% and $p < 0.10$ is significant at 10%

Table 5.
Granger-causality between variables

4.2 Association between market return and investor sentiment
To analyze the association between stock market return and investor sentiment on Twitter, we estimated quantile regressions at five different quantiles. We know that market factors such as risk, size, value, momentum and liquidity explain stock returns in markets (Fama & French, 1993; Carhart, 1997; Keene & Peterson, 2007; Machado & Medeiros, 2011). Furthermore, Lee *et al.* (2002) state that investor sentiment is a risk factor that can explain abnormal returns. We then started from this evidence to test and present the association of investor sentiment on Twitter to market return, in different quantiles of returns over the analyzed period, in Table 6.

Table 6 presents the results for the 5-factor model estimated, with the addition of the Twitter ISI. The model estimated for the median (q.50) shows that ISI is not significant to explain stock market return (measured by the BOVA11 return). This finding diverges from our expectations. However, we can identify a positive and significant association between stock returns and Twitter sentiment (ISI) in periods of lower returns, particularly at the 10% and 25% quantiles (q.10 and q.25).

We can say that we have evidence that investor sentiment on Twitter is useful to explain stock returns in periods of lower returns in Brazil, which converges with previous literature (Mao *et al.*, 2012; Wei *et al.*, 2016; Oliveira *et al.*, 2017; Nisar & Yeung, 2018). These findings reinforce the assumptions of Brown and Cliff (2005) and Baker and Wurgler (2006, 2007), showing that there is a relationship between sentiment and return, which is also observed even in an emerging market, such as Brazilian. Based on this, we can say that Hypothesis 2 of this study, which suggests that the ISI explains part of stock returns, is confirmed only at times of lower returns (in the quantile of lower returns in our sample).

In those moments, the more positive the investor sentiment on Twitter, the greater the stock return (or, in periods of negative returns, they become less negative). In this sense, investors can use tweets in their investment strategies when the market operates with low returns, seeking to obtain above-average returns. This finding for the Brazilian emerging market converges with evidence also observed in the US market (Kim *et al.*, 2014; Wei *et al.*, 2016; Renault, 2017).

The findings in Table 6 also explain in a complementary way the positive and significant correlation that we identified in Table 2, between market return and Twitter sentiment

Variable	q.10	q.25	q.50	q.75	q.90
ISI	0.0173[***]	0.0080[***]	0.0032	−0.0025	−0.0004
	(0.0057)	(0.0030)	(0.0023)	(0.0033)	(0.0057)
MKT	1.0570[***]	1.0493[***]	1.0526[**]	1.0624[***]	1.0451[***]
	(0.0213)	(0.0111)	(0.0084)	(0.0123)	(0.0213)
SMB	0.0087	−0.0065	0.0040	−0.0630[*]	−0.0633
	(0.0653)	(0.0342)	(0.0258)	(0.0377)	(0.0653)
HML	0.1090[***]	0.0681[***]	0.0513[***]	0.0796[***]	0.1065[***]
	(0.0339)	(0.0178)	(0.0134)	(0.0196)	(0.0339)
WML	0.0253	0.0164	0.0271[**]	0.0094	−0.0311
	(0.0305)	(0.0160)	(0.0120)	(0.0176)	(0.0305)
IML	−0.0623	−0.0813[**]	−0.1089[***]	−0.0669[*]	−0.1028
	(0.0627)	(0.0328)	(0.0248)	(0.0362)	(0.0627)
Constant	−0.0026[***]	−0.0009[***]	0.0002[*]	0.0015[***]	0.0028[***]
	(0.0003)	(0.0001)	(0.0001)	(0.0002)	(0.0003)
No. of Obs	491	491	491	491	491
Pseudo R^2	0.8458	0.8502	0.8531	0.8497	0.8448

Note(s): *** is significant at 1%, ** at 5% and * at 10%

Table 6.
Association between
market return and
Twitter sentiment

(0.1064, $p < 0.05$). When we calculated this correlation in the first and last quantiles of the distribution of returns, we noticed that the correlation in the quantile of lower returns is stronger (0.4054, $p < 0.01$), while in the quantile of higher returns it becomes negative and nonsignificant (-0.0592, $p > 0.10$).

Concerning market factors, during the period analyzed, the most relevant factors for the Brazilian market were market and value factors, which had positive and significant effects in all analyzed quantiles ($p < 0.05$). This implies that companies with higher risk and higher book-to-market have higher returns, especially as these factors are associated with greater growth opportunities (Fama & French, 1993).

The liquidity factor had negative and significant effects in quantiles q.25, q.50 and q.75. The exceptions were the extremes (q.10 and q.90). This finding agrees with Machado and Medeiros (2011). Thus, the less liquid the stock, the greater the return. The size factor only showed negative and significant relationship at 10% in q.75, and the moment factor only had a positive and significant association in q.50 ($p < 0.05$).

Alternatively, we check whether the relevance of tweets in explaining stock market return lies in the informational content of the messages, which reflect investor sentiment, or in the total volume of tweets (TVT) that is published daily. Table 7 presents the estimated quantile regression models for different return quantiles. We can verify that in none of the estimated models the TVT was relevant to explain the market returns ($p > 0.10$). These findings are similar to those of Nisar and Yeung (2018), who found that the volume of tweets is not related to variables such as price, return and turnover.

In this sense, we can verify that the relevance of investor sentiment on Twitter lies in the informational content of the messages. The stock market return was not sensitive to the volume of tweets (TVT), either at times of higher or lower returns. Additionally, about market factors, the findings of this robustness analysis maintain the same direction as the previous one.

5. Conclusion

Our analyses suggest a positive and significant relationship between Brazilian market performance and investor sentiment on Twitter, especially at times of lower returns.

Variable	q.10	q.25	q.50	q.75	q.90
TVT	−0.0005	−0.0001	−0.0003	−0.0003	0.0000
	(0.0007)	(0.0003)	(0.0002)	(0.0003)	(0.0006)
MKT	1.0506***	1.0511***	1.0533***	1.0642***	1.0451***
	(0.0250)	(0.0119)	(0.0084)	(0.0121)	(0.0208)
SMB	0.0121	0.0239	−0.0054	−0.0650*	−0.0596
	(0.0769)	(0.0366)	(0.0259)	(0.0372)	(0.0639)
HML	0.1078***	0.0732***	0.0559***	0.0678***	0.1071***
	(0.0397)	(0.0189)	(0.0134)	(0.0192)	(0.0330)
WML	0.0362	0.0094	0.0202*	−0.0017	−0.0330
	(0.0359)	(0.0171)	(0.0121)	(0.0174)	(0.0299)
IML	−0.0947	−0.1050***	−0.1021***	−0.0635*	−0.1071*
	(0.0737)	(0.0351)	(0.0248)	(0.0357)	(0.0613)
Constant	0.0012	−0.0003	0.0025	0.0034	0.0030
	(0.0045)	(0.0022)	(0.0015)	(0.0022)	(0.0038)
No. of Obs	491	491	491	491	491
Pseudo R^2	0.8416	0.8481	0.8529	0.8497	0.8448

Note(s): *** is significant at 1%, ** at 5% and * at 10%

Table 7.
Association between market return and tweet volume

However, during the analyzed period it was not possible to identify a dynamic relationship between these two phenomena. We observed that the interaction among investors on Twitter increases especially in the worst moments of returns in this market: when trading volume increases, Twitter sentiment becomes more negative, and the association of this sentiment with market performance becomes stronger.

In stock markets, investors can have their investment decisions affected by emotional factors, because their rationality is limited. An example of this is the so-called herd effect. Thus, in pessimistic moments in the market, when there are lower returns, sharing information and sentiment on social networks like Twitter can help reducing information asymmetry, therefore making investment decisions more efficient. This can lead investors to think more rationally and analyze the expected future cash flow of the company, which determines its intrinsic value.

The daily total volume of tweets was also positively associated with the trading volume in the market when messages are endowed with negative sentiment. This finding is especially strong when we note that the first and second biggest spikes in volume took place during the Carne Fraca Operation and during the "Joesley Day," as they were accompanied by a series of negative news around scandals involving large companies and Brazil president. This confirms evidence from previous studies that Brazilian market places more emphasis on negative news, which converges with the heuristic of loss aversion. Based on Kahneman and Tversky (1979), perception related to loss is 2.25 times greater than that related to gain.

The findings of this study have relevant contributions to demonstrate that it is possible to use the monitoring of investor sentiment on Twitter as an aid in predicting Brazilian market fluctuations, or at least to understand their reasons. Thus, sentiment and volume of tweets have the potential to be used as auxiliary variables to investment strategies, as in the Brazilian market it seems that investors negotiate in the market while watching Twitter. Thus, we can follow the market's mood through this social network.

Our findings contribute to financial literature, converging on previous evidence, bringing results like those found in other markets and expanding them. The study demonstrated how the online activities of investors have effects in the Brazilian market, filling part of the gap in this literature. Additionally, we show in our evidence that despite the association between Twitter sentiment and market performance, we did not find a dynamic relationship or a clear cause-effect relationship between these phenomena.

This study is useful to demonstrate to market participants that investors can obtain abnormal returns with this kind of information, especially when the market performance is in its low side. There are practical implications for analysts and investors, as findings indicate that in some instances these agents may use Twitter in a complementary way to their analyses. Above all, it is necessary to be cautious when the sentiment is optimistic, or when the market is experiencing high returns. In these cases, we did not identify clear evidence of a positive association. The usefulness of these findings to regulatory and supervisory bodies of the Brazilian stock market is also noteworthy, since the monitoring of Twitter appears as a potential tool to identify possible market manipulations, especially with less liquid assets (more easily driven prices).

References

Agarwal, S., Kumar, S., & Goel, U. (2019). Stock market response to information diffusion through internet sources: A literature review. *International Journal of Information Management*, *45*, 118–131.

Al-Nasseri, A., & Ali, F. (2018). What does investors' online divergence of opinion tell us about stock returns and trading volume. *Journal of Business Research*, *86*, 166–178.

B3 – Brasil, Bolsa, Balcão. (2021). *Individual inverstor: An analysis of investor's evolution in B3*. São Paulo: B3. November.

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*, 1645–1680.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–151.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Brown, G., & Cliff, M. (2005). Investor sentiment and asset valuation. *The Journal of Business*, *78*(2), 405–440.

Carhart, M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57–82.

Chen, H., & Lo, T. (2019). Online search activities and investor attention on financial markets. *Asia Pacific Management Review*, *24*(1), 21–26.

Fama, E. (1970). Efficient markets: A review of theory and empirical work. *Journal of Finance*, *25*(2), 383–417.

Fama, E. F. (1991). Efficient capital markets: II. *Journal of Finance*, *46*(5), 1575–1617.

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

Fan, W., & Gordon, M. (2014). The power of social media analytics. *Communications of the ACM*, *57*(6), 74–81.

Galdi, F. C., & Gonçalves, A. M. (2018). Pessimismo e incerteza das notícias e o comportamento dos investidores no Brasil. *Revista de Administração de Empresas*, *58*(2), 130–148.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*, 263–291.

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, *33*, 171–185.

Keene, M., & Peterson, D. (2007). The importance of liquidity as a factor in asset pricing. *Journal of Financial Research*, *30*(1), 91–109.

Kim, S., & Kim, D. (2014). Investor sentiment from internet message postings and the predictability of stock returns. *Journal of Economic Behavior & Organization*, *107*, 708–729.

Kim, J., Ryu, D., & Seo, S. (2014). Investor sentiment and return predictability of disagreement. *Journal of Banking & Finance*, *42*, 166–178.

Kim, N., Lučivjanská, K., Molnár, P., & Villa, R. (2019). Google searches and stock market activity: Evidence from Norway. *Finance Research Letters*, *28*, 208–220.

Klibanoff, P., Lamont, O., & Wizman, T. (1998). Investor reaction to salient news in closed-end country funds. *Journal of Finance*, *53*(2), 673–699.

Lee, W., Jiang, C., & Indro, D. (2002). Stock market volatility, excess returns, and the role of investor sentiment. *Journal of Banking & Finance*, *26*(12), 2277–2299.

Machado, M., & Medeiros, O. (2011). Modelos de precificação de ativos e o efeito liquidez: Evidências empíricas no mercado acionário brasileiro. *Revista Brasileira de Finanças*, *9*, 383–412.

Malkiel, B. (2003). The Efficient markets hypothesis and its critics. *Journal of Economic Perspectives*, *17*(1), 59–82.

Mao, H., Counts, S., & Bollen, J. (2011). Predicting financial markets: Comparing survey, news, Twitter, and search engine data. *arXiv*, Preprint arXiv:1112.1051.

Mao, Y., Wei, W., Wang, B., & Liu, B. (2012). Correlating S&P 500 stocks with Twitter data. In *Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research* (pp. 69–72).

Martins, O. S., & Barros, L. A. B. C. (2021). Firm informativeness, information environment, and accounting quality in emerging countries. *The International Journal of Accounting*, 56(1), 1–50.

Menkhoff, L. (1998). The noise trading approach-questionnaire evidence from foreign exchange. *Journal of International Money and Finance*, 17(3), 547–564.

Nisar, T., & Yeung, M. (2018). Twitter as a tool for forecasting stock market movements: A short-window event study. *The Journal of Finance and Data Science*, 4, 101–119.

Oliveira, N., Cortez, P., & Areal, N. (2017). The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume, and survey sentiment indices. *Expert Systems with Applications*, 73, 125–144.

Renault, T. (2017). Intraday online investor sentiment and return patterns in the US stock market. *Journal of Banking & Finance*, 84, 25–40.

Silva, M. (2017). O efeito do sentimento das notícias sobre o comportamento dos preços no mercado acionário brasileiro. PhD thesis. Brasília: Universidade de Brasília.

Tauhata, S. (2021). *Na Fintwit, comunidade financeira no Twitter, pessimismo aumenta*. São Paulo: Valor Investe. available from: https://valorinveste.globo.com/mercados/noticia/2021/08/16/na-fintwit-comunidade-financeira-no-twitter-pessimismo-aumenta.ghtml.

Wei, W., Mao, Y., & Wang, B. (2016). Twitter volume spikes and stock options pricing. *Computer Communications*, 73, 271–281.

Zhang, Q., Deng, M., & Yang, S. (2010). Does investor sentiment and stock return affect each other: (S) VAR model approach. *International Journal of Management Science and Engineering Management*, 5(5), 334–340.

Corresponding author
Orleans Silva Martins can be contacted at: orleansmartins@yahoo.com.br

Associate Editor: Leandro Maciel