

Investor sentiment analysis in the Brazilian market: empirical evidence through technical news

Nataniel Geraldo Mendes Peixoto¹ 

Pedro Reis^{2,3} 

Antônio Pedro Soares Pinto^{2,3} 

Abstract

Purpose – The proliferation of news disseminated through electronic media influences investor decision-making and stock prices. This study investigates the statistical association between investor sentiment, as measured using the Loughran and McDonald, VADER, and Harvard-IV lexical dictionaries, and the performance of the Brazilian stock market (Ibovespa).

Theoretical framework – The study is grounded in behavioral finance, emphasizing investor sentiment and cognitive biases as determinants of market behavior. This concept is supported by prior research linking financial news sentiment to stock market performance.

Design/methodology/approach – A dataset of 9,515 exclusive financial news articles from Valor +News (November 2021 to August 2024) was retrieved from *Valor Econômico* and analyzed using Python. The empirical strategy combines multiple linear regression with lagged specifications, vector autoregression (VAR), Granger causality tests, monetary policy subperiod analyses, and first-difference robustness models. Gold, Bitcoin, the exchange rate, country risk, and news volume were included as control variables.

Findings – VADER and Loughran and McDonald exhibit consistent positive associations with Ibovespa returns across the main specifications, whereas Harvard-IV is more sensitive to model specification. Additional analyses indicate that the evidence is predominantly contemporaneous and should be interpreted as statistical associations rather than causal effects. The exchange rate, Bitcoin, and country risk are also statistically significant, while news volume is not.

Practical & social implications of research – The findings highlight the usefulness of lexical sentiment measures for analyzing financial markets and emphasize the importance of interpreting sentiment-return relationships as contemporaneous associations.

Originality/value – This study provides new evidence for the Brazilian market by combining an exclusive financial news dataset with multiple lexical dictionaries and complementary robustness analyses.

Keywords: Investor sentiment, lexical approach, financial news.

1. Universidade Politécnica de Viseu, Escola Superior de Tecnologia e Gestão, Viseu, Portugal
2. Universidade Politécnica de Viseu, Centro de Investigação em Serviços Digitais, Viseu, Portugal
3. Universidade da Beira Interior, Núcleo de Estudos em Ciências Empresariais, Covilhã, Portugal

How to cite:

Peixoto, N. G. M., Reis, P., & Pinto, A. P. S. (2026). Investor sentiment analysis in the Brazilian market: empirical evidence through technical news. *Revista Brasileira de Gestão de Negócios*, 28(3), e20260062. <https://doi.org/10.7819/rbgn.v28i3.4357>

Received on:

Mar/26/2026

Approved on:

Sep/01/2026

Responsible editor:

Prof. Dr Vinicus Brunassi

Reviewers:

Edimilson Costa Lucas; Jefferson Colombo

Evaluation process:

Double Blind Review

This article is open data



Revista Brasileira de Gestão de Negócios

DOI: 10.7819/rbgn.v28i3.4357

I Introduction

Specialized media has become an increasingly important source of information for companies, particularly those listed on capital markets, where the principle of equal access to information prevails (Galdi & Gonçalves, 2018). Market participants rely on the large amount of information disseminated daily from various unpredictable sources.

According to the efficient market hypothesis formulated by Fama (1970), asset prices reflect all available information. This makes it impossible to achieve abnormal gains, or returns that exceed the risk-adjusted benchmark. However, subsequent studies identifying anomalies in this hypothesis contributed to the development of behavioral finance (Mussa et al., 2008).

The relationship between news and market behavior has been widely studied, leading to research on investor sentiment (Galdi & Gonçalves, 2018). Baker and Wurgler (2007) define investor sentiment as beliefs about future cash flows and investment risks that are not supported by market fundamentals. Researchers have studied this phenomenon since the 1990s, building on the work of Shleifer and Summers (1990) and Barberis et al. (1998).

Recently, studies on sentiment analysis in Brazil have used quantitative proxies, such as the Consumer Confidence Index (CCI) (Piccoli et al., 2018) and the Baker and Wurgler methodology (Mello, 2019; Duarte, 2023), to measure investor sentiment, instead of analyzing news media content. Other sources include Twitter (Carosia et al., 2019; Souza & Martins, 2020; Santos, 2022) and annual financial reports (Pagliarussi et al., 2016). Studies have also applied machine learning (Souza & Martins, 2020) and lexicon-based methods (Santos, 2022).

Recent evidence further reinforces the relevance of this research agenda. Terraza et al. (2024) document the dynamic interactions among Bitcoin, gold, and stock market volatility. Pereira Jr. and Arevalo (2024) examine the relationship between gold, exchange rates, and the Ibovespa using ARDL methods. Brito et al. (2025) show that crypto assets became more significant for retail investors during the period from 2020 to 2022. Finally, Jurafsky and Martin (2024) highlight the ongoing development of methods for processing language.

Within the lexical approach, studies employ various dictionary strategies. Some develop proprietary dictionaries (Silva, 2017; Pagliarussi et al., 2016; Silva & Machado, 2019), while others use existing ones. For example, Costa (2021) applied Silva's (2017) model. Consolidated dictionaries are also emerging, as seen in Valcareggi's (2022) use of the dictionary created by Shapiro et al. (2020). Castro (2023) used the VADER dictionary to analyze Twitter news and demonstrated its applicability to social media.

However, the empirical relevance of the Harvard-IV, Loughran and McDonald, and VADER dictionaries in sentiment analysis applied to specialized financial news remains underexplored. This study evaluates the statistical association of each dictionary with Ibovespa returns, avoiding a causal interpretation of same-day relationships. Although they are all lexical dictionaries, they were developed at different times, and lexicons evolve over time (Tahmasebi et al., 2021).

In this context, the study addresses the following research question: "Is investor sentiment in the Brazilian market, as measured by financial news and a lexical approach, statistically associated with Ibovespa returns?" The study tests the hypothesis that sentiment measures are positively related to the Brazilian stock market, as Silva (2017) suggested, while recognizing that contemporaneous regressions identify association rather than causal influence.

For the empirical investigation, news articles from the "Finance" category of the *Valor Econômico* "Valor +News" service are used. These articles were accessed under a prior research license. Investor sentiment was measured using the Harvard-IV, Loughran and McDonald, and VADER lexical dictionaries. The dataset was processed in Python and analyzed by estimating multiple linear regression models using time-series data in Stata Now 19.5 Basic Edition. Ibovespa returns were used as the dependent variable, and sentiment, exchange rate, gold price, Bitcoin price, country risk, and news frequency were used as explanatory variables. The sample period is from November 2021 to August 2024, reflecting the period for which Valor +News data were available.

News content may be associated with stock volatility and market indices (Mitra & Mitra, 2011). Investor optimism or pessimism may reflect behavioral biases and changing perceptions of risk. Therefore, the empirical interpretation in this study is framed

in terms of statistical association rather than direct causal impact.

The main findings are that VADER and Loughran and McDonald are positively and significantly associated with Ibovespa returns in the main specifications, whereas Harvard-IV is significant mainly in individual and first-difference specifications. Additional lagged and VAR/Granger tests indicate limited and -predictive content specific to the variables, with evidence of bidirectional feedback for VADER. These results refine prior Brazilian evidence by showing that specialized financial news contains relevant sentiment information and that the relationship should be interpreted as a contemporaneous association rather than a causal influence.

The present study builds upon and expands prior Brazilian research in several ways. Unlike the study by Silva (2017), which examined financial news from Valor Econômico (2024) using an author-developed Portuguese financial sentiment dictionary, this study used the subscription-based Valor +News service to compare three established lexical dictionaries within the same empirical framework. Compared to the work of Santos (2022), which analyzed economic and financial news disseminated through the *Folha de S.Paulo* Twitter account using VADER and LM-SA-2020 in a vector autoregressive framework, this study focuses on specialized, full-text financial news and its association with Ibovespa returns. The present study also complements Valcareggi (2022), who examined economic news using the Shapiro et al. (2020) sentiment dictionary, by directly comparing Loughran and McDonald, VADER, and Harvard-IV within the same empirical framework.

The distinctiveness of the Valor +News dataset arises not only from its restricted access but also from its specialized nature. The service provides curated financial content produced by professionals for a specialized financial audience. Unlike news disseminated through social media platforms, where content is subject to platform-specific formats and information environments, Valor +News represents a distinct layer of the Brazilian financial information ecosystem. Together, these features enable the study to evaluate the dependence of Brazilian financial news sentiment and market returns on the lexical measure employed.

The remainder of the paper is organized as follows: literature review, methodology, results, and conclusions.

2 Literature review

2.1 Evolution of finance studies

One of the central premises of modern finance is investor rationality. This premise is supported by the work of Modigliani and Miller (1958, 1961) with their capital structure irrelevance propositions, Markowitz (1952) with his portfolio theory, Sharpe (1964) with his asset pricing model, Black and Scholes (1972) with their option pricing framework, and Fama (1970) with his efficient market hypothesis (Famá et al., 2008). However, evidence of market anomalies and abnormal returns has challenged this view, raising questions about full rationality. In this context, behavioral finance emerged as an extension of modern finance, addressing its limitations (Mussa et al., 2008).

According to Shleifer and Summers (1990), behavioral finance essentially rests on two pillars. The first concerns the limits to arbitrage, as it is a risky and therefore constrained activity. The second pillar incorporates the psychological dimensions of investors, specifically their decision-making processes and cognitive biases.

Silva (2017) argues that market inefficiencies are associated with the realization that investors do not exhibit the level of rationality assumed by the efficient market hypothesis. Instead, they display bounded rationality, which is influenced by psychological factors known as cognitive biases, such as negativity bias.

Behavioral finance does not reject modern finance, but rather highlights its limitations in explaining observed market phenomena (Milanez, 2003). Several studies over the past few decades have documented market anomalies, including the day-of-the-week effect, the January effect, momentum, and behavioral biases such as excessive optimism, pessimism, and loss aversion (Banz, 1981; Jaffe et al., 1989; Haugen & Jorion, 1996; Mussa et al., 2008).

Studies such as that of García (2013) highlight the link between psychology and behavioral economics, emphasizing news as a key source in decision-making. The author identifies a correlation between market sentiment and news content, especially during periods of uncertainty when investor sensitivity increases. Therefore, recessionary periods may influence expectation formation differently than periods of economic expansion.

Investor heterogeneity is a key consideration, as individuals process news differently (Silva, 2017). Cognitive distortions reshape beliefs beyond traditional rationality. Thus, bounded rationality and psychological factors undermine expectation homogeneity and generate noise in price adjustment.

2.2 Technology and finance

Santos (2022) highlights the increasing reliance on technological and computational methods to understand economic phenomena, such as predicting market behavior and anticipating events. Advances in data mining and sentiment analysis have made it possible to extract implicit information from textual sources, such as news articles and reports, thereby enhancing the analysis of financial markets.

Medhat et al. (2014) group sentiment classification techniques into three main approaches: machine learning, lexicon-based, and hybrid methods. Machine learning uses algorithms and linguistic features to classify text, while lexicon-based approaches employ predefined dictionaries to identify sentiment polarity. Hybrid methods combine elements of both.

Sentiment lexicons are widely used in sentiment analysis (Hutto & Gilbert, 2014), and several have been developed and refined in recent years. Liu (2012) notes that sentiment is primarily captured through words with positive or negative connotations. Research in this field has primarily used three dictionaries: Harvard-IV, Loughran and McDonald, and VADER.

The Harvard-IV lexicon originated from the General Inquirer (Stone et al., 1966), which is one of the earliest manually constructed text analysis tools. It has been widely used in sentiment studies (Esuli & Sebastiani, 2005; Kamps et al., 2004; Turney & Littman, 2003), as noted by Hutto and Gilbert (2014), and it represents words in binary categories of positive and negative sentiment (Jurafsky & Martin, 2024).

However, its application in financial contexts is limited. Loughran and McDonald (2011) demonstrate that a significant portion of words classified as negative in the Harvard-IV dictionary are not perceived as negative in financial texts. To address this issue, Loughran and McDonald developed a domain-specific dictionary that more accurately captures financial sentiment. This dictionary has since become widely adopted in empirical studies (Silva, 2017).

Hutto and Gilbert (2014) developed VADER (Valence Aware Dictionary and Sentiment Reasoner), a tool that assigns values for positive, negative, and neutral sentiment. Because of its accuracy across different text types, it has become widely used in natural language processing studies. Santos (2022) used this dictionary in his research.

2.3 Investor sentiment

Contrary to classical theory, Baker and Wurgler (2006) show that investor sentiment influences the U.S. market. Using a sentiment index based on six proxies from 1962 to 2001, they found that higher sentiment inflates stock prices and makes valuation and arbitrage more difficult. Meanwhile, lower sentiment is associated with higher subsequent returns.

Silva (2017) constructed a sentiment index based on a proprietary dictionary of positive and negative words to analyze the impact of news on Brazilian stock market volatility under conditions of economic uncertainty. Using daily data from the Ibovespa, DJIA, VIX, and S&P 500 and 45,304 news items from *Valor Econômico* from July 2011 to June 2017, the study found that pessimism was associated with declining returns during periods of uncertainty. The results also indicate that textual sentiment predicts volatility and that news content influences investor perceptions of market and economic conditions.

Silva and Machado (2019) examined whether the systematic risk of Vale's preferred shares (VALE5) was influenced by the tone and volume of news related to the Samarco disaster. Using news from *Valor Econômico*, a proprietary lexicon, and ARCH-GARCH models to estimate dynamic beta, they found that tone and volume affected the stock's beta, particularly between the fifth and ninth deciles. This indicates a relationship between media coverage and systematic risk.

Souza and Martins (2020) examined the relationship between Twitter information and stock returns and trading volume in Brazil. Using 314,864 tweets from 2017 to 2018 and a machine learning-based sentiment index (Google Cloud Natural Language API), they found that optimism was positively associated with contemporaneous returns but predicted lower future returns. Tweet volume is also related to trading activity, with higher trading volume associated with

more negative messages, indicating a link between social media sentiment and market behavior.

Valcareggi (2022) examined the impact of economic news on the Bovespa Index by using a sentiment index derived from 104,290 news articles and the lexical dictionary of Shapiro et al. (2020). The study found that, in stable environments, sentiment and the index moved in the same direction. However, under favorable conditions, pessimistic news did not prevent increases; under unfavorable conditions, positive news did not stimulate the index.

Santos (2022) applied sentiment analysis to economic and financial news to assess its impact on macroeconomic dynamics. This study used Twitter data from *Folha de São Paulo* (2011–2022), vector

autoregressive models, and the LM-SA-2020 and VADER dictionaries. The results show that the LM-SA-2020 dictionary outperforms the VADER dictionary in explaining macroeconomic variables. This highlights the importance of dictionary selection and the relevance of sentiment indices, especially during expansion periods. The study also found that sentiment shocks affect Ibovespa performance. ISV (VADER) generated short-lived positive effects, followed by rapid reversion. In contrast, ISL (LM-SA-2020) produced more persistent negative impacts. These results indicate distinct dynamics depending on the index used.

Several other studies on sentiment analysis are included in the literature, such as those presented in Table 1.

Table 1
Summary of related studies

Author	Information Source	Dictionary/ Sentiment Proxy	Main Findings / Conclusions
Tetlock (2007)	Newspaper articles	Harvard-IV	* High pessimism in news is associated with stock price declines. * Strong pessimism or weak optimism in published news is related to higher trading volume.
Pagliarussi et al. (2016)	Corporate reports	Developed their own dictionary	* Weak association between tone metrics and trading volume or stock price volatility. * Positive word lists are positively associated with abnormal returns and negatively associated with return volatility.
Galdi and Gonçalves (2018)	Newspaper articles	Pagliarussi et al. (2016) dictionary	* Market places greater weight on uncertainty and negative words. * Terms such as “uncertainty” are negatively related to returns, while “negative” terms are linked to volatility.
Mello (2019)	Market data	Adaptation of Baker and Wurgler (2006) methodology	* Investor sentiment explains the returns of higher-volatility stocks. * More speculative stocks are harder to arbitrage and more exposed to irrational sentiment.
Igarashi et al. (2021)	Newspaper articles	Machine Learning - Amazon Comprehend	* The correlation between sentiment and stock price movement is predominantly moderate.
Reis and Pinho (2020a)	Market data	CCI, economic sentiment indicator, VSTOXX, gold bullion, yield spreads	* Strong co-movement between U.S. and European stock markets. * Optimism has a stronger impact on volatility than pessimism, except during the 2008–2009 period.
Reis and Pinho (2020b)	Market data	13 investor sentiment proxies	* Different measures of investor sentiment explain stock returns. * Some measures significantly predict irrational investor behavior and returns.
Costa (2021)	Newspaper articles	Silva (2017) dictionary	* Lower textual complexity enhances the transmission of information into prices. * Less complex news is associated with higher subsequent corporate investment levels.
Duarte (2023)	Market data	Adaptation of the methodology of Baker and Wurgler (2006)	* Sentiment is more pronounced in pre-crisis periods. * The proxy “net equity fund inflows” captures sentiment changes more quickly. * Findings demonstrate effective indicators for assessing investor sentiment.

3 Data and methodology

3.1 Model variables

The Ibovespa Index is used as a proxy for market returns (the dependent variable), and the explanatory variables include investor sentiment derived from news content, the USD/BRL exchange rate, gold and Bitcoin returns, Brazil's country risk (as measured by the CDS index), and daily news frequency. Previous studies have widely used the Ibovespa (e.g., Galdi & Gonçalves, 2018; Valcareggi, 2022; Igarashi et al., 2021; Marschner et al., 2021), as well as investor sentiment (e.g., Silva, 2017; Silva & Machado, 2019; Costa, 2021). Other variables examined in the literature include the exchange rate (Marschner et al., 2021; Paraboni et al., 2016), gold and Bitcoin (Karameshinova, 2023; Vartanian Jr. et al., 2022; Castro, 2023; Pereira Jr. & Arevalo, 2024), CDS-based country risk (Sing, 2019; Azevedo No., 2019), and news frequency (Silva & Machado, 2019; Ferguson et al., 2015; Faustino, 2013).

Ibovespa returns are computed as the logarithmic variation of closing prices (Santana, 2018). Data for Bitcoin, gold, the exchange rate, and CDS were obtained from Investing.com and cover the period from November 2021 to August 2024. Daily variations are calculated based on closing prices (Supplementary Data 1 – STATA Database).

Bitcoin is included as a control variable because crypto-assets became more mainstream during the sample period, potentially capturing risk appetite and portfolio reallocation across risky assets (Terraza et al., 2024; Brito et al., 2025). Bitcoin is measured as the growth rate of its price in U.S. dollars. USD/BRL is included separately to control for exchange rate dynamics relevant to the Brazilian market.

The analysis was conducted using multiple linear regression for time series, estimated in Stata Now 19.5 Basic Edition (Supplementary Data 2 – STATA Script), as follows:

$$\begin{aligned} \ln_ibov_n = & \beta_0 + \beta_1 qty_news_n + \\ & \beta_2 sent_VADER_n + \beta_3 sent_LM_n + \\ & \beta_4 sent_HIV_n + \beta_5 Bitcoin_n + \\ & \beta_6 Gold_n + \beta_7 USD_BRL_n + \\ & \beta_8 CDS_Brazil_n + \varepsilon_n \end{aligned} \quad (1)$$

where $\ln_ibov_n$ denotes the logarithmic return of the Ibovespa in period n ; β_0 is the intercept; qty_news_n is the daily number of news items; $sent_VADER_n$, $sent_LM_n$, and $sent_HIV_n$ represent investor sentiment measured by the VADER, Loughran and McDonald, and Harvard-IV

dictionaries, respectively; $Bitcoin_n$ and $Gold_n$ denote the returns of Bitcoin and gold; USD_BRL_n is the exchange rate variation; CDS_Brazil_n is the variation in Brazil's Credit Default Swap; and ε_n is the error term. Coefficients β_1 to β_8 indicate the magnitude and direction of the effects on $\ln_ibov_n$.

Standard diagnostic tests were applied, including the Dickey-Fuller test for stationarity (Dickey & Fuller, 1979), the variance inflation factor (VIF) for multicollinearity, the Breusch-Godfrey test for autocorrelation (Breusch, 1978; Godfrey, 1978), and the Breusch-Pagan test for heteroskedasticity (Breusch & Pagan, 1979) (Supplementary Data 2 – STATA Script, Supplementary Material)

The diagnostic tests support the adequacy of the baseline specification. The Dickey-Fuller tests indicate stationarity; the Breusch-Godfrey and Breusch-Pagan tests do not indicate residual autocorrelation or heteroskedasticity; the VIF results do not suggest severe multicollinearity; and the Ramsey RESET test does not reject the null hypothesis of no omitted variables ($F(3, 685) = 1.32$, $p = 0.267$). These results are detailed further in Section 4 of the manuscript. However, financial returns may still exhibit conditional heteroskedasticity, volatility clustering, heavy tails, and asymmetric responses to shocks. For this reason, lagged specifications, VAR/Granger causality tests, and monetary policy subperiod regressions were added as complementary robustness checks (Supplementary Data 2 – STATA Dofile). GARCH-type models are left for future research as this paper focuses on the conditional mean of Ibovespa returns rather than conditional volatility.

3.2 News processing and investor sentiment

When constructing the investor sentiment index, information from financial news headlines and full content written by specialists and disseminated via subscription were considered. The general public only had access to headlines, while the full content required a subscription. The dataset was obtained from *Valor Econômico* via the “Valor +News” service, a leading source of financial information in Brazil and abroad, with daily publications from November 2021 to August 2024. The original JSON file contained 9,583 news items.

The database was processed using NotePad++ to ensure the Python code developed for this study correctly interpreted the JSON structure. HTML tags were removed using regular expressions. The Google Translator

module from the Python deep-translator library was used to translate the news into English since the lexical dictionaries contain English entries. Reis et al. (2015) concluded that translation does not significantly affect the accuracy of sentiment analysis. Therefore, translation does not represent a methodological limitation when the target language matches the dictionary.

The news files were processed separately for each of the three Python scripts corresponding to the lexical dictionaries: Harvard-IV, Loughran and McDonald (2011), and VADER (Hutto & Gilbert, 2014). The library pysentiment2 0.1.1 (DeRobertis, 2024) was used for Harvard-IV and Loughran and McDonald, while vaderSentiment 3.3.2 (Hutto, 2024) was used for VADER. Both libraries provide functions for measuring sentiment through a lexical approach with implementations that calculate text polarity.

The sentiment (tone) of each news item was represented by its polarity. For the Loughran and McDonald and Harvard-IV dictionaries, polarity was calculated based on the relative frequency of positive and negative words. For VADER, however, it was derived from word valence. The -1 to +1 interval represents the theoretical scale of the sentiment measures, where -1 indicates the most negative sentiment and +1 indicates the most positive sentiment (Silva & Machado, 2019). The minimum and maximum values reported in Table 2 are empirical sample values and therefore do not need to reach the theoretical boundaries.

For days with more than one news item, the arithmetic mean of the sentiment values was computed following Silva and Machado (2019). The final dataset contains 9,515 news items, which are aggregated into 697 daily observations. Each observation represents the

average sentiment on days when the Ibovespa index varied (Supplementary Data 1 – STATA Database).

Although prior evidence supports multilingual sentiment analysis, automated translation of financial text may introduce semantic noise. To address the reliability of the sentiment measures, the revised study reports two complementary validation procedures. First, during the paper stage, an author-led manual sampling check evaluated whether the automated Portuguese-to-English translation preserved the financial meaning of the news and whether the dictionary outputs were plausible in context. Second, a Portuguese-language large language model (LLM)-assisted benchmark was constructed from a stratified sample of 100 news articles classified as positive, neutral, or negative according to their expected tone for the Ibovespa. Then agreement rates, Cohen's kappa, precision, recall, and F1-score were computed by comparing this benchmark with the VADER, Loughran and McDonald, and Harvard-IV classifications.

4 Results analysis

4.1 Descriptive statistics

The initial analysis examined the descriptive statistics of the model variables (Table 2), providing an overview of the dataset and a preliminary understanding of their distribution and behavior.

Regarding the Ibovespa ($\ln_{ibov}$), 50% of the variations were equal to or less than 0.04274%, and 95% were equal to or less than 1.84%, with a mean of 0.0424%. The standard deviation was approximately 1.1%, ranging from a minimum of -3.4% to a maximum of 5.3%. Silva (2017) reported a positive mean for Ibovespa returns,

Table 2
Descriptive statistics of the model variables

Variable	Obs	Mean	Percentiles		Std. Dev.	Min	Max
			50%	95%			
$\ln_{ibov}$	697	.000424	.0004274	.0184775	.0111302	-.034488	.053928
qty_news	697	13.65136	14	20	3.845486	2	31
sent_VADER	697	.5254248	.5506	.8323067	.2191688	-.57755	.96885
sent_LM	697	-.2838067	-.2888985	-.0352045	.1423894	-.7456858	.1867071
sent_HIV	697	.2433882	.2429599	.4057619	.1004818	-.0863323	.5678228
Bitcoin	697	.0000403	-.0013	.0497	.0319629	-.1563	.1194
Gold	697	.0004073	.0003	.0159	.0093175	-.0281	.0314
USD_BRL	697	.0000568	0	.0165	.0092238	-.0458	.0371
CDS_Brazil	697	-.0002139	-.0002	.0373	.0366329	-.2847	.4219

though it was deflated by the Broad National Consumer Price Index (IPCA). He also reported a negative median, suggesting the presence of outliers in the distribution. In contrast, Souza and Martins (2020) reported positive mean and median values using exchange-traded fund (ETF) returns.

Regarding the investor sentiment variables (sent_VADER, sent_LM, and sent_HIV), their empirical results differ across indicators, despite their theoretical range being -1 to +1. Figure 1 shows their behavior over the study period, with distinct value ranges but a broadly similar pattern over time.

For VADER (sent_VADER), 50% of the daily average sentiment values were ≤ 0.55 , and 95% were ≤ 0.83 , with a mean of 0.52. The standard deviation was approximately 22%, ranging from a minimum of -0.57 to a maximum of 0.96. The minimum indicates an average absence of strongly negative news, while the maximum value indicates highly optimistic days.

For Loughran and McDonald (sent_LM), 50% of the values were ≤ -0.288 and 95% were ≤ -0.035 , with a mean of -0.28. The standard deviation was approximately 14%, with a minimum of -0.74 and a maximum of 0.18. No values above 0.18 were observed, indicating a tendency toward negative classification.

For Harvard-IV (sent_HIV), 50% of the values were ≤ 0.242 , and 95% were ≤ 0.405 , with a mean of

0.243. The standard deviation was approximately 10%, ranging from -0.086 to 0.56. Overall, the results indicate a tendency toward positive classification, with at least half of the observations showing positive sentiment and a relatively low minimum value for negativity.

These differences reflect the context and purpose of each lexicon. Loughran and McDonald focuses on corporate reports and disclosures, which contain more negative terms related to risk, loss, and uncertainty. VADER, which is designed for social media, uses informal language and shows a more balanced distribution. Harvard-IV, which was developed for psychological studies, is also relatively balanced and broadly applicable.

4.2 Regression model assumptions

Dickey-Fuller confirmed the assumption of series stationarity for all the variables included in the model (Table 3). The p-value (0.000) is lower than the significance level and indicates that the null hypothesis should be rejected, thereby confirming the absence of a unit root.

The assumption of heteroskedasticity was tested using the Breusch-Pagan test (Table 4). Since the p-value was greater than the significance level (0.3785), the null hypothesis of homoskedasticity could not be rejected, indicating that the variance of the errors remained constant and did not vary as a function of the independent variables.

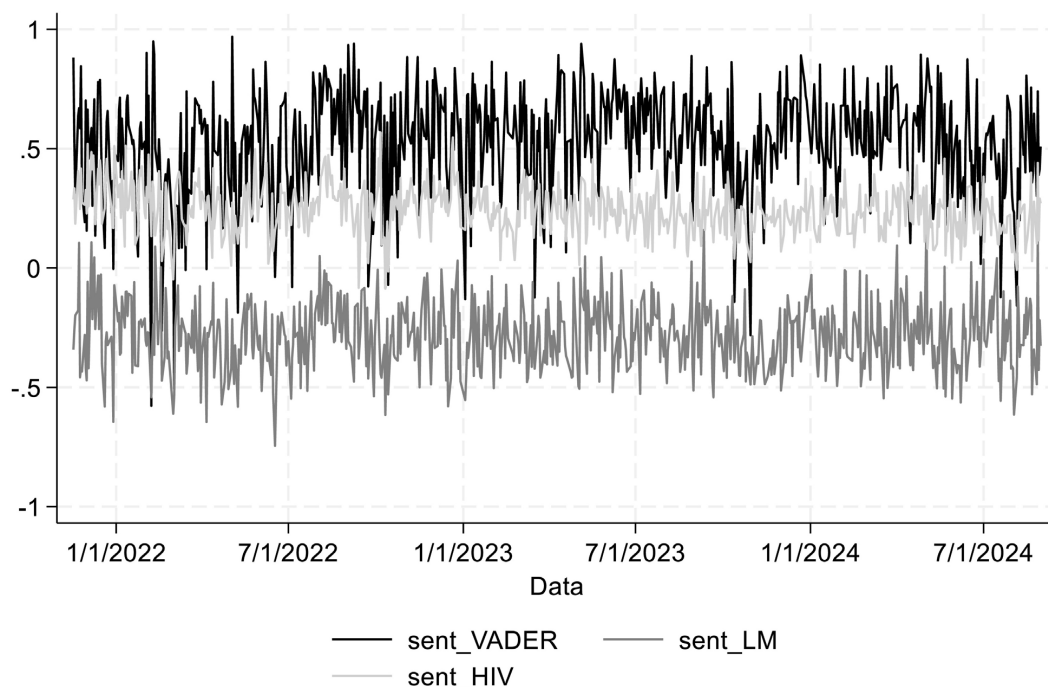


Figure 1. Investor sentiment dynamics according to different indicators

Table 3
Dickey-Fuller test results

Variable	Test	Critical Value			p-value Z(t)
	Statistic	1%	5%	10%	
ln_ibov	-23.228	-3.430	-2.860	-2.570	0.0000
qty_news	-13.349				0.0000
sent_VADER	-20.195				0.0000
sent_LM	-22.981				0.0000
sent_HIV	-17.173				0.0000
Bitcoin	-25.707				0.0000
Gold	-23.097				0.0000
USD_BRL	-23.999				0.0000
CDS_Brazil	-31.225				0.0000

Table 4
Breusch-Pagan test results

<i>Breusch-Pagan/Cook-Weisberg Test for Heteroskedasticity</i>	
<i>Assumption: Normal error terms</i>	
<i>Variable: Fitted values of ln_ibov</i>	
H0:	<i>Constant variance</i>
chi2(1) =	0.78
Prob > chi2 =	0.3785

The assumption of residual autocorrelation was assessed using the Breusch-Godfrey LM test (Table 5). Since the p-value exceeded the significance level (0.6422), the null hypothesis of no autocorrelation could not be rejected.

The assumption of no multicollinearity was assessed using the variance inflation factor (VIF) test (Table 6). The mean VIF was 1.21 and all individual values were well below the conventional threshold of 10, indicating the absence of severe correlations among the explanatory variables in the model.

4.3 Correlation matrix

Table 7 presents the correlation matrix results for the model variables, providing an overview of the pairwise relationships and offering preliminary evidence for regression analysis.

The correlation matrix reveals several statistically significant relationships at the 5% level. The sentiment variables (sent_VADER, sent_LM, and sent_HIV) are positively correlated with Ibovespa variations, indicating that optimistic sentiment is associated with higher index

returns, and that these variables tend to move in the same direction.

The correlations among the three sentiment dictionaries are moderate rather than redundant: VADER-LM = 0.4497, VADER-Harvard-IV = 0.4044, and LM-Harvard-IV = 0.5061. These correlations suggest that the dictionaries capture related yet distinct dimensions of news sentiment, which justifies both individual and joint specifications (“horse-race” model). However, the moderate correlations require cautious interpretation of joint coefficients because each dictionary only partially overlaps with the others.

Although not significant at the 5% level, the volume of news (qty_news) shows a negative relationship with Ibovespa returns, suggesting that higher news frequency is associated with lower stock performance. Faustino (2013) also found no statistically significant correlation between news volume and Ibovespa returns, though the coefficients were positive.

Bitcoin and gold exhibited a positive and statistically significant correlation with the Ibovespa, indicating that increases in their prices are associated with higher index

Table 5
Breusch-Godfrey LM test results

<i>Breusch-Godfrey for Autocorrelation</i>			
lags(p)	chi2	df	Prob > chi2
1	0.216	1	0.6422
<i>H0: no serial correlation</i>			

Table 6
VIF Test Results

<i>Variable</i>	<i>VIF</i>	<i>1/VIF</i>
sent_LM	1.50	0.668065
sent_HIV	1.43	0.697472
sent_VADER	1.42	0.705601
USD_BRL	1.18	0.850373
Gold	1.06	0.945840
Bitcoin	1.05	0.948244
qty_news	1.04	0.959235
CDS_Brazil	1.02	0.976322
Mean VIF	1.21	

Table 7
Correlation matrix of the model variables

	ln_ibov	qty_news	sent_VADER	sent_LM	sent_HIV	Bitcoin	Gold	USD_BRL	CDS_Brazil
ln_ibov	1.0000								
qty_news	-0.0353	1.0000							
sent_VADER	0.3212*	0.1533*	1.0000						
sent_LM	0.2876*	0.0285	0.4497*	1.0000					
sent_HIV	0.2455*	0.0738	0.4044*	0.5061*	1.0000				
Bitcoin	0.2246*	0.0422	0.1002*	0.0824*	0.0846*	1.0000			
Gold	0.0920*	-0.0327	0.0537	0.0727	0.0925*	0.0916*	1.0000		
USD_BRL	-0.6051*	0.0575	-0.2785*	-0.2026*	-0.2037*	-0.1955*	-0.1853*	1.0000	
CDS_Brazil	-0.0710	0.0126	-0.0621	-0.0327	-0.0436	-0.0815*	-0.1198*	0.0155	1.0000

Note: Coefficients marked with * are statistically significant at the 5% level ($p < 0.05$).

returns. Mello (2019) also identified a positive relationship between sentiment and commodities using an International Monetary Fund (IMF) index.

Karameshinova (2023) found a strong positive correlation between Bitcoin and the S&P 500, indicating co-movement; however, no causal effect was identified. Gold also showed a positive but weaker association. Terraza et al. (2024) reported that Bitcoin is positively correlated with major U.S. stock indices, especially during periods of instability. This finding reinforces its speculative nature and limited role as a safe-haven asset. In contrast, gold exhibits a weak or negative relationship, confirming its role as a hedge.

Contrary to studies suggesting stronger linkages, Klein et al. (2018) found no significant relationship between Bitcoin and traditional financial markets. They reported very low correlations with the S&P 500 and other indices. The authors concluded that, unlike gold, Bitcoin is not a useful hedge and is mainly characterized by high volatility and speculative behavior.

Bitcoin returns showed a positive and statistically significant correlation with investor sentiment variables (sent_VADER, sent_LM, and sent_HIV). This result is consistent with Castro (2023), who also observed this relationship using an autoregressive distributed lag (ARDL) model with tweet-based data and VADER sentiment.

Country risk (CDS_Brazil), which was not significant at the 5% level, and the exchange rate (USD_BRL), which was significant at the 5% level, both showed a negative relationship with Ibovespa returns. This indicates that increases in country risk and the U.S. dollar are associated with lower stock returns.

Mello (2019) also found a negative relationship between sentiment, measured using a methodology adapted from Baker and Wurgler (2006), and risk, proxied by the Emerging Markets Bond Index (EMBI). Similarly, Sing (2019) reported a negative correlation between CDS risk and the S&P 500, with significant relationships observed across pre- and post-crisis periods. Marschner et al. (2021) also identified a negative relationship between the exchange rate and sentiment measured by the Consumer Confidence Index (CCI) and Business Confidence Index (BCI).

4.4 Multiple linear regression

After verifying the assumptions that ensure the validity and robustness of the results, the estimated values presented in Table 8 could be obtained based on the model expressed in Equation 1.

The empirical interpretation proceeds in two steps. First, the individual contribution of each sentiment

dictionary was assessed through separate specifications in which VADER, Loughran and McDonald, and Harvard-IV enter the model separately (Models 2–4). The first-difference robustness specification (Model 5) is also discussed at this stage to assess whether the main relationships remain statistically significant when focusing on changes in the variables over time. Second, the joint specification (Model 1) includes the three sentiment measures simultaneously, allowing for a horse-race comparison of their incremental explanatory contribution after controlling for news volume, Bitcoin, gold, USD/BRL, and country risk.

In Models 2, 3, and 4, including or excluding investor sentiment variables did not substantially change the explanatory power compared to the baseline Model 1, with adjusted R^2 values ranging from 39.1% to 40.4%. The sentiment variables (sent_VADER, sent_LM, and sent_HIV) were all statistically significant in their respective specifications ($p < 0.001$), indicating that the optimism or pessimism captured by the different dictionaries contributes to explaining variations in the Ibovespa index.

Bitcoin consistently had a positive and statistically significant effect across all specifications, while the exchange rate (USD_BRL) had a negative and highly significant effect. In contrast, news volume (qty_news) and gold remained statistically insignificant, suggesting that only the

Table 8
Individual and joint dictionary specifications and first-difference robustness model

Variable	(1) ln_ibov	(2) ln_ibov	(3) ln_ibov	(4) ln_ibov	(5) D.ln_ibov
qty_news	-0.0000900 (-1.06)	-0.0000983 (-1.15)	-0.0000375 (-0.44)	-0.0000491 (-0.57)	—
sent_VADER	0.00525** (3.21)	0.00831*** (5.54)	—	—	—
sent_LM	0.00855** (3.11)	—	0.0131*** (5.45)	—	—
sent_HIV	0.00391 (1.04)	—	—	0.0138*** (3.97)	—
Bitcoin	0.0346** (3.21)	0.0358*** (3.33)	0.0354** (3.25)	0.0361** (3.23)	—
Gold	-0.0468 (-1.30)	-0.0401 (-1.12)	-0.0471 (-1.30)	-0.0485 (-1.34)	—
USD_BRL	-0.642*** (-14.81)	-0.655*** (-15.13)	-0.672*** (-15.63)	-0.682*** (-15.65)	—
CDS_Brazil	-0.0144* (-1.86)	-0.0145+ (-1.77)	-0.0162* (-2.04)	-0.0161* (-2.00)	—
D.qty_news	—	—	—	—	-0.000280+ (-1.95)
D.sent_VADER	—	—	—	—	0.00467* (2.19)
D.sent_LM	—	—	—	—	0.00746* (2.17)
D.sent_HIV	—	—	—	—	0.0107* (2.12)
D.Bitcoin	—	—	—	—	0.0453*** (3.57)
D.Gold	—	—	—	—	-0.0535 (-1.22)
D.USD_BRL	—	—	—	—	-0.599*** (-11.21)
D.CDS_Brazil	—	—	—	—	0.00178 (0.28)
Constant	0.000420 (0.22)	-0.00255* (-1.98)	0.00471** (3.15)	-0.00220+ (-1.65)	0.000267 (0.51)
N	697	697	697	697	538
R ²	0.419	0.406	0.409	0.396	0.391
Adjusted R ²	0.412	0.401	0.404	0.391	0.382

Notes: Robust t-statistics are reported in parentheses. Statistical significance is indicated by + $p < 0.10$; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

qualitative content of news and select financial indicators immediately impact the stock market.

Regarding risk variables, CDS_Brazil showed marginal to moderate significance in Models 2–4. Negative coefficients indicated that increases in risk were associated with declines in the Ibovespa. This suggests more conservative investor behavior under economic and political uncertainty. The consistency of the signs of the coefficients and the stability of their magnitudes across models further support the robustness of the findings, indicating that the effects are not highly sensitive to changes in model specifications.

Model 5, which incorporates first-differenced variables, shows that all three sentiment indicators (D.sent_VADER, D.sent_LM, and D.sent_HIV) are statistically significant. This reinforces the relevance of daily changes in news sentiment and the inclusion of Harvard-IV in first differences, while maintaining a cautious interpretation based on association rather than causal influence.

After examining the individual specifications and the first-difference robustness model, the discussion now turns to the full specification, which includes the three sentiment measures simultaneously. This joint model allows us to evaluate the incremental explanatory contribution of each dictionary while controlling for the presence of the other sentiment measures and the remaining explanatory variables.

According to the baseline Model 1 in Table 8, the R-squared indicates that 41.86% of the variation in the dependent variable ($\ln_{ibov}$) is explained; the remaining 58.14% is attributed to factors outside the model. Similar values are reported in the literature. For example, Marschner et al. (2021) found an explanatory power of 42% and 49% in models using the Consumer Confidence Index (CCI) and the Business Confidence Index (BCI), respectively.

When the three sentiment measures are estimated simultaneously, only VADER (sent_VADER) and Loughran and McDonald (sent_LM) remain statistically significant (p -values = 0.001 and $0.002 < 0.05$), indicating that these measures contribute to explaining variations in the Ibovespa. The positive coefficients suggest that optimistic news is associated with stock market appreciation.

These findings are consistent with prior studies. Silva (2017), for example, reported a significant positive relationship between media news and market performance. Valcareggi (2022) found that in stable periods, the sentiment

index moves proportionally with the Ibovespa; in other periods, however, the reaction is disproportionate. Galdi and Gonçalves (2018) identified associations between “uncertainty” and “negative” words with Ibovespa returns and volatility, suggesting market pessimism. Piccoli et al. (2018) found that pessimistic and optimistic sentiments both influence market behavior and asset pricing.

The results indicate a statistical association between sentiment and Ibovespa returns on the same day, since both are measured over the same daily window. While this pattern is consistent with rapid information incorporation, it should not be interpreted as evidence of a unidirectional causal effect, as market movements may also influence the tone of same-day news coverage.

The Harvard-IV sentiment variable was not statistically significant (p -value = $0.298 > 0.05$). This result is consistent with Loughran and McDonald (2011), who argue that non-domain-specific lexicons may yield less relevant classifications. This motivated the development of their own widely used dictionary. Conversely, Tetlock (2007) found a correlation between news content and market behavior using Harvard-IV, linking pessimism to stock price declines.

Unlike the sentiment measures, the volume of news (qty_news) was not statistically significant in explaining Ibovespa variations (p -value = $0.291 > 0.05$). This suggests that news quality may be more important than quantity. Faustino (2013) also found no significant effect of news volume or semantic content on Ibovespa returns.

The country risk variable (CDS_Brazil) was not significant at the 5% level (p -value = 0.063), although it may be considered marginally significant at the 10% level. The negative coefficient (-0.014) indicates that higher risk is associated with lower Ibovespa returns, which is consistent with more conservative investor behavior. Azevedo No. (2019) also reported a significant negative relationship between CDS and Ibovespa.

The exchange rate (USD_BRL) was statistically significant (p -value = $0.000 < 0.05$), indicating its relevance in explaining variations in the Ibovespa index. The negative coefficient (-0.64) suggests that dollar appreciation and real depreciation are associated with lower stock returns. This behavior reflects a shift toward safer assets during periods of uncertainty, which reduces exposure to emerging markets such as Brazil.

Pereira Jr. and Arevalo (2024) also found a statistically significant negative relationship between the Ibovespa and the exchange rate without lags. They argue

that increases in the exchange rate reflect capital outflows and currency shortages. Meanwhile, Marschner et al. (2021) and Paraboni et al. (2016) reported a negative and significant relationship between the exchange rate and sentiment measured by the Consumer Confidence Index (CCI), though Marschner et al. (2021) found no significance for the Business Confidence Index (BCI). Sing (2019) identified a negative, though not significant relationship between the exchange rate (USD/BRL) and both Brazilian (MSCI Brazil) and U.S. (S&P 500) indices.

Bitcoin was statistically significant ($p\text{-value} = 0.001 < 0.05$), indicating its contribution to explaining Ibovespa variation. The positive coefficient (0.034) suggests that Bitcoin appreciation is associated with higher stock returns, possibly reflecting investor demand for riskier assets in favorable economic contexts. However, Karameshinova (2023) found no significant relationship between Bitcoin and the S&P 500.

Gold was not statistically significant ($p\text{-value} = 0.195 > 0.05$), so its effect on the Ibovespa cannot be confirmed. Pereira Jr. and Arevalo (2024), however, found a significant positive relationship with gold (LnGold), arguing that it does not provide hedging or diversification benefits. Karameshinova (2023) identified gold as the second most influential variable on Bitcoin.

4.5 Additional robustness analyses

Lagged L1-L4 specifications were estimated to evaluate whether sentiment retains explanatory content beyond the contemporaneous daily window (Supplementary Data 3 – Appendix A). The positive contemporaneous coefficients do not persist uniformly across lags. The strongest lagged evidence for VADER appears at the third lag, which has a negative sign. This indicates that the baseline result is mainly contemporaneous and not evidence of a stable predictive channel.

VAR and Granger causality tests were also estimated with lags ranging from one to four observations (Supplementary Data 4 – Appendix B). VADER sentiment Granger-causes $\ln_ibov$; however, the reverse direction is also significant, indicating feedback between market returns and subsequent news sentiment. For Loughran and McDonald and Harvard-IV, however, there is no evidence that lagged sentiment Granger-causes $\ln_ibov$. These results reinforce the need to interpret the main findings as statistical associations rather than causal effects.

Finally, the sample was divided into three non-overlapping monetary policy subperiods based on the Selic/Copom cycle: tightening, restrictive plateau, and easing/pause (Supplementary Data 5 – Appendix C). The subperiod results show that the association between sentiment and return is strongest during the restrictive plateau, while the relevance of individual dictionaries varies across regimes. This suggests that the relationship between news sentiment and market returns is sensitive to macrofinancial conditions.

The validation exercise further indicates that Loughran and McDonald has the highest agreement with the Portuguese-language LLM-assisted benchmark, followed by Harvard-IV and VADER (Supplementary Data 6 – Appendix D). These results support using dictionary-based measures as aggregate proxies for financial news tone while indicating that they should not be interpreted as perfect article-level sentiment classifications.

5 Conclusion

This study examines the relationship between investor sentiment and the Brazilian stock market, as represented by the Ibovespa index, by analyzing specialized news from the *Valor Econômico* Valor+News service. A lexicon-based approach was applied with the Harvard-IV, Loughran and McDonald, and VADER dictionaries, combined with multiple linear regression and additional robustness checks based on first differences, lagged models, VAR/Granger tests, and monetary-policy subperiods.

The results show that the Loughran and McDonald and VADER dictionaries were consistently statistically significant across all models, while Harvard IV was only significant in Model 4 and the first-difference version of Model 5. These findings indicate that optimistic sentiment is associated with stock appreciation. News volume was not significant in any model, suggesting that content quality outweighs quantity.

Other variables showed consistent associations: Bitcoin was positively related to Ibovespa returns, while USD/BRL was negatively and strongly associated with returns. Including Bitcoin is motivated by evidence that crypto-assets have become more integrated into broader investment behavior, potentially capturing risk appetite and portfolio reallocation during the sample period (Terraza et al., 2024; Brito et al., 2025).

This study contributes to the Brazilian investor sentiment literature in three ways. First, it complements

the work of Silva (2017), Santos (2022), and Valcareggi (2022) by using a specialized, expert-produced, subscription-based news source rather than social media or broader news archives. Second, it compares three lexical dictionaries within the same empirical framework, showing that VADER and Loughran and McDonald are more stable in the main regressions, whereas Harvard-IV yields less robust results across model specifications. Third, the additional robustness tests show that the main evidence is predominantly contemporaneous and should be interpreted as statistical association rather than causal impact.

This study has several limitations. First, the relatively short time span may limit the generalizability of the results across market conditions. Second, while the manual sampling check and the LLM-assisted validation bolster confidence in the sentiment pipeline, the study lacks a substantial corpus labeled by multiple independent experts. Third, the analysis focuses on the conditional mean of Ibovespa returns and does not directly model conditional volatility through GARCH-type specifications. Fourth, the VAR/Granger evidence indicates possible feedback between returns and VADER sentiment, reinforcing the need for caution regarding reverse causality.

Future research could build on this framework by using larger corpora of Portuguese text labeled by experts, comparing dictionary-based measures with transformer models in Portuguese, and estimating GARCH, EGARCH, or GJR-GARCH models to evaluate whether sentiment also affects conditional volatility and asymmetric risk responses.

Furthermore, although artificial intelligence tools were employed to translate and refine the textual content, they did not influence the analytical framework or the interpretation of the results.

References

- Azevedo No., G. A. D. (2019). *Fundamentos macroeconômicos do credit default swaps nos mercados emergentes* [Dissertação de mestrado, Fundação Getúlio Vargas]. <https://hdl.handle.net/10438/28084>
- Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4), 1645-1680. <https://doi.org/10.1111/j.1540-6261.2006.00885.x>.
- Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *The Journal of Economic Perspectives*, 21(2), 129-151. <https://doi.org/10.1257/jep.21.2.129>.
- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1), 3-18. [https://doi.org/10.1016/0304-405X\(81\)90018-0](https://doi.org/10.1016/0304-405X(81)90018-0).
- Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, 49(3), 307-343. [https://doi.org/10.1016/S0304-405X\(98\)00027-0](https://doi.org/10.1016/S0304-405X(98)00027-0).
- Black, F., & Scholes, M. (1972). The valuation of option contracts and a test of market efficiency. *The Journal of Finance*, 27(2), 399-417. <https://doi.org/10.2307/2978484>.
- Breusch, T. S. (1978). Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17(31), 334-355. <https://doi.org/10.1111/j.1467-8454.1978.tb00635.x>.
- Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5), 1287-1294. <https://doi.org/10.2307/1911963>.
- Brito, M. C., Chague, F., & Colombo, J. (2025). Distracted by crypto. *SSRN*. <https://doi.org/10.2139/ssrn.5018192>
- Carosia, A. E. O., Coelho, G. P., & Silva, A. E. A. (2019). Analyzing the Brazilian financial market through portuguese sentiment analysis in social media. *Applied Artificial Intelligence*, 34(1), 1-19. <https://doi.org/10.1080/08839514.2019.1673037>.
- Castro, P. A. C. (2023). *Análise de sentimento dos tweets e sua influência sobre o preço do bitcoin* [Dissertação de mestrado, Universidade Federal do Ceará]. <http://repositorio.ufc.br/handle/riufc/75404>
- Costa, R. R. A. (2021). *Notícias financeiras e o custo da informação: A relação entre complexidade textual e a difusão de informação no mercado financeiro* [Tese de doutorado, Universidade Federal do Rio Grande do Norte]. <https://repositorio.ufrn.br/handle/123456789/48506>
- DeRobertis, N. (2024). *PySentiment*. <https://nickderobertis.github.io/pysentiment/index.html>

- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427-431. <https://doi.org/10.1080/01621459.1979.10482531>.
- Duarte, C. (2023). *Sentimento do investidor no mercado de ações brasileiro* [Dissertação de mestrado]. Universidade de São Paulo. <https://doi.org/10.11606/D.96.2023.tde-12072023-145002>.
- Esuli, A., & Sebastiani, F. (2005). Determining the semantic orientation of terms through gloss classification. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management* (pp. 617-624). Association for Computing Machinery. <https://doi.org/10.1145/1099554.1099713>.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417. <https://doi.org/10.2307/2325486>.
- Famá, R., Cioffi, P. L. de M., & Coelho, P. A. R. (2008). Contexto das finanças comportamentais: Anomalias e eficiência do mercado de capitais brasileiro. *Revista de Gestão USP*, 15(2), 65-78. https://revistas.usp.br/regel/pt_BR/article/view/36638.
- Faustino, C. C. R. (2013). *O impacto das notícias no mercado financeiro brasileiro* [Dissertação de mestrado, Fundação Getúlio Vargas]. <https://hdl.handle.net/10438/10532>
- Ferguson, N. J., Philip, D., Lam, H., & Guo, J. M. (2015). Media content and stock returns: The predictive power of press. *Multinational Finance Journal*, 19(1), 1-31. <https://doi.org/10.17578/19-1-1>.
- Galdi, F. C., & Gonçalves, A. M. (2018). Pessimismo e incerteza das notícias e o comportamento dos investidores no Brasil. *RAE*, 58(2), 130-148. <https://doi.org/10.1590/s0034-759020180203>.
- García, D. (2013). Sentiment during recessions. *The Journal of Finance*, 68(3), 1267-1300. <https://doi.org/10.1111/jofi.12027>.
- Godfrey, L. G. (1978). Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica*, 46(6), 1293-1301. <https://doi.org/10.2307/1913829>.
- Haugen, R. A., & Jorion, P. (1996). The January effect: Still there after all these years. *Financial Analysts Journal*, 52(1), 27-31. <https://doi.org/10.2469/faj.v52.n1.1963>.
- Hutto, C. J. (2024). *vaderSentiment*. <https://github.com/cjhutto/vaderSentiment>
- Hutto, C., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1), 216-225. <https://doi.org/10.1609/icwsm.v8i1.14550>.
- Igarashi, W., Valdevieso, G. S., & Igarashi, D. C. C. (2021). Análise de sentimentos e indicadores técnicos: Uma análise da correlação dos preços de ativos com a polaridade de notícias do mercado de ações. *Brazilian Journals of Business*, 3(1), 470-486. <https://doi.org/10.34140/bjbv3n1-029>.
- Jaffe, J., Keim, D. B., & Westerfield, R. (1989). Earnings yields, market values, and stock returns. *The Journal of Finance*, 44(1), 135-148. <https://doi.org/10.1111/j.1540-6261.1989.tb02408.x>.
- Jurafsky, D., & Martin, J. H. (2024). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition with language models* (3rd ed.). Stanford University. <https://web.stanford.edu/~jurafsky/slp3>
- Kamps, J., Marx, M., Mokken, R. J., & De Rijke, M. (2004). Using WordNet to measure semantic orientations of adjectives. *LREC*, 4, 1115-1118. <https://doi.org/10.63317/222z3zcs9izx>.
- Karameshinova, E. A. (2023). *Análise de sensibilidade: Fatores que influenciam o preço das Criptomoeas* [Dissertação de mestrado, Universidade de Lisboa]. <http://hdl.handle.net/10071/30911>
- Klein, T., Thu, H. P., & Walther, T. (2018). Bitcoin is not the New Gold – A comparison of volatility, correlation, and portfolio performance. *International Review of Financial Analysis*, 59, 105-116. <https://doi.org/10.1016/j.irfa.2018.07.010>.

- Liu, B. (2012). *Sentiment analysis and opinion mining*. Springer. <https://doi.org/10.1007/978-3-031-02145-9>.
- Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65. <https://doi.org/10.1111/j.1540-6261.2010.01625.x>.
- Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91. <https://doi.org/10.2307/2975974>.
- Marschner, P. F., Ceretta, P. S., & Ambrozini, M. A. (2021). Impactos dos Sentimentos Racionais e Irracionais dos Investidores sobre o Retorno e a Volatilidade do Mercado Brasileiro. In *21º USP International Conference in Accounting*. Universidade de São Paulo. <https://congressosp.fipecafi.org/anais/21UspInternational/ArtigosDownload/3323.pdf>
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4), 1093-1113. <https://doi.org/10.1016/j.asej.2014.04.011>.
- Mello, P. H. S. (2019). *O sentimento do investidor no mercado de ações brasileiro* [Dissertação de mestrado, Fundação Getúlio Vargas]. <https://hdl.handle.net/10438/28827>
- Milanez, D. Y. (2003). *Finanças comportamentais no Brasil* [Dissertação de mestrado]. Universidade de São Paulo. <https://doi.org/10.11606/D.12.2003.tde-09022004-130012>.
- Mitra, G., & Mitra, L. (2011). *The handbook of news analytics in finance*. Hoboken, USA: John Wiley & Sons. <https://doi.org/10.1002/9781118467411>.
- Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance and the theory of investment. *The American Economic Review*, 48(3), 261-297. <https://www.jstor.org/stable/1809766>
- Modigliani, F., & Miller, M. H. (1961). Dividend policy, growth, and the valuation of shares. *The Journal of Business*, 34(4), 411-433. <https://doi.org/10.1086/294442>.
- Mussa, A., Yang, E., Trovão, R., & Famá, R. (2008). Hipótese de mercados eficientes e finanças comportamentais: As discussões persistem. *FACEF Pesquisa*, 11(1), 5-17. <http://periodicos.unifacef.com.br/facefpesquisa/article/view/109/173>
- Pagliarussi, M. S., Aguiar, M. O., & Galdi, F. C. (2016). Sentiment analysis in annual reports from Brazilian companies listed at the BM&FBovespa. *BASE-Revista de Administração e Contabilidade da Unisinos*, 13(1), 53-64. <https://revistas.unisinos.br/index.php/base/article/view/base.2016.131.04>
- Paraboni, A. L., dos Santos Fraga, L., Bender Fo., R., Vieira, K. M., & Coronel, D. A. (2016). Impacto de variáveis macroeconômicas e corrupção na Petrobrás no Sentimento de Mercado. *Revista ESPACIOS*, 37(31). <https://www.revistaespacios.com/a16v37n31/16373120.html>
- Pereira Jr., J. R. G., & Arevalo, J. L. S. (2024). Relations between the price of gold, oil, and the exchange rate with the IBOVESPA: An application of the ARDL model. *Revista de Gestão Social e Ambiental*, 18(10), e08688. <https://doi.org/10.24857/rgsa.v18n10-011>.
- Piccoli, P., da Costa Jr., N. C. A., da Silva, W. V., & Cruz, J. A. W. (2018). Investor sentiment and the risk-return tradeoff in the Brazilian market. *Accounting and Finance*, 58(S1), 599-618. <https://doi.org/10.1111/acfi.12342>.
- Reis, J., Gonçalves, P., Araújo, M., Pereira, A., & Benevenuto, F. (2015). Uma abordagem multilíngue para análise de sentimentos. In *Anais do IV Brazilian Workshop on Social Network Analysis and Mining*. Sociedade Brasileira de Computação. <https://doi.org/10.5753/brasnam.2015.6767>.
- Reis, P. M. N., & Pinho, C. (2020a). A new European investor sentiment index (EURsent) and its return and volatility predictability. *Journal of Behavioral and Experimental Finance*, 27, 100373. <https://doi.org/10.1016/j.jbef.2020.100373>.
- Reis, P. M. N., & Pinho, C. (2020b). A reappraisal of the causal relationship between sentiment proxies and stock returns. *Journal of Behavioral Finance*, 22(4), 420-442. <https://doi.org/10.1080/15427560.2020.1792910>.
- Santana, H. N. (2018). *Eleições presidenciais brasileiras e a volatilidade do IBOVESPA: relações com variáveis conjunturais e risco político* [Dissertação de mestrado,

- Universidade Federal de Minas Gerais]. <http://hdl.handle.net/1843/30298>
- Santos, L. S. (2022). *Aplicações de técnicas de análise de sentimentos às notícias econômicas e sua relação com variáveis macroeconômicas* [Monografia de conclusão de curso, Universidade Federal Fluminense]. <http://app.uff.br/riuff/handle/1/32843>
- Shapiro, A. H., Sudhof, M., & Wilson, D. J. (2020). Measuring news sentiment. *Journal of Econometrics*, 228(2), 221-243. <https://doi.org/10.1016/j.jeconom.2020.07.053>.
- Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425-442. <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>.
- Shleifer, A., & Summers, L. (1990). The noise trader approach to finance. *The Journal of Economic Perspectives*, 4(2), 19-33. <https://doi.org/10.1257/jep.4.2.19>.
- Silva, M. D. O. P. (2017). *O efeito do sentimento das notícias sobre o comportamento dos preços no mercado acionário brasileiro* [Tese de doutorado, Universidade Federal da Paraíba]. <https://repositorio.ufpb.br/jspui/handle/123456789/12575>
- Silva, M. D. O. P., & Machado, M. A. V. (2019). Índice de sentimento textual: Uma análise empírica do impacto das notícias sobre risco sistemático. *Revista Contemporânea De Contabilidade*, 16(40), 24-42. <https://doi.org/10.5007/2175-8069.2019v16n40p24>.
- Sing, A. M. K. (2019). *Análise dos Credit Default Swap (CDS) de títulos soberanos* [Dissertação de mestrado, Insper Instituto de Ensino e Pesquisa]. <https://www.sidalc.net/search/Record/dig-insper-br-11224-2604/Description>
- Souza, D. M. S., & Martins, O. S. (2020). *Sentimento do investidor via twitter e desempenho do mercado acionário brasileiro*. Associação Nacional de Programas de Pós-Graduação em Ciências Contábeis.
- Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvie, D. M. (1966). *General inquirer*. MIT Press.
- Tahmasebi, N., Borin, L., Jatowt, A., Xu, Y., & Hengchen, S. (2021). *Computational approaches to semantic change*. Language Science Press. <https://doi.org/10.5281/zenodo.5040241>.
- Terraza, V., İpek, B., & Rounaghi, M. M. (2024). The nexus between the volatility of Bitcoin, gold, and American stock markets during the COVID-19 pandemic: Evidence from VAR-DCC-EGARCH and ANN models. *Financial Innovation*, 10(1), 1-29. <https://doi.org/10.1186/s40854-023-00520-3>.
- Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139-1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>.
- Turney, P. D., & Littman, M. L. (2003). Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4), 315-346. <https://doi.org/10.1145/944012.944013>.
- Valcareggi, S. M. (2022). *Análise de sentimentos para o mercado brasileiro* [Dissertação de mestrado]. Universidade de São Paulo. <https://doi.org/10.11606/D.96.2022.tde-24012023-165235>.
- Valor Econômico. (2024). *Venda de conteúdo*. <https://valor.globo.com/venda-de-conteudo/>
- Vartanian Jr., P. R., Moura, A. A., Racy, J. C., & Simioni, R. (2022). Preços dos non-fungible tokens (NFTs), criptomoedas, taxa de juros e ouro: uma análise econométrica (jan. de 2019 – ago. de 2022). In *Anais do Fórum Mackenzie de Liberdade Econômica*. Universidade Presbiteriana Mackenzie. [https://www.even3.com.br/anais/libecon2022/564426-precos-dos-non-fungibletokens-\(nfts\)-criptomoedas-taxa-de-juros-e-ouro--uma-analise-econometrica-\(jan-de-2019/\)](https://www.even3.com.br/anais/libecon2022/564426-precos-dos-non-fungibletokens-(nfts)-criptomoedas-taxa-de-juros-e-ouro--uma-analise-econometrica-(jan-de-2019/))

SUPPLEMENTARY MATERIAL

Supplementary material accompanies this paper.

Supplementary Data 1 – STATA Database

Supplementary Data 2 – STATA Script

Supplementary Data 3 – Appendix A_Lagged models L1-L4

Supplementary Data 4 – Appendix B_ VAR and Granger-causality tests

Supplementary Data 5 – Appendix C_Monetary-policy subperiod regressions_

Supplementary Data 6 – Appendix D_Manual and LLM-assisted validation of sentiment variables

Financial support:

This work was funded by the Foundation for Science and Technology (FCT), I.P., through Project Ref. UIDB/05583/2020 and Project Ref. UID/04630/2025. We would also like to thank the Research Centre in Digital Services (CISeD), the Polytechnic University of Viseu, and the Research Centre for Business Sciences (NECE) for their support. The DOI identifier is 10.54499/UID/04630/2025.

Research data availability statement - Open science:

The full data supporting the findings of this study has been made available at:

Peixoto, Nataniel; Reis, Pedro; Pinto, António, 2026, "Investor Sentiment Analysis in the Brazilian Market: Empirical Evidence Through Technical News", <https://doi.org/10.7910/DVN/OX5OAS>, Harvard Dataverse, V1, UNF:6:xBAOl2P+k93GXqOm1kMW6Q== [fileUNF]

Conflicts of interest:

The authors have no conflicts of interest to declare.

Copyrights:

RBGN owns the copyrights of this published content.

Plagiarism analysis:

RBGN uses the iThenticate tool to perform plagiarism analysis on all its articles at the time of submission and after manuscript approval.

Disclaimer on the Use of Artificial Intelligence

The authors declare that artificial intelligence (AI) platforms were utilized in various stages of this research to enhance accuracy, efficiency, and overall quality. The specific contributions of AI to this work include:

(X) Enhancing readability and proofreading - ChatGPT

(X) Coding assistance - ChatGPT

(X) Formatting and reference standardization - ChatGPT

The authors declare that, after using this tool/service, they reviewed and edited the content as needed and take full responsibility for the content of the publication.

Authors:

1. Nataniel Geraldo Mendes Peixoto, Master in Finance, Universidade Politécnica de Viseu, Escola Superior de Tecnologia e Gestão, Viseu, Portugal.

E-mail: nataniel.geraldo@yahoo.com.br

2. Pedro Reis, Coordinator Professor, Universidade Politécnica de Viseu, Centro de Investigação em Serviços Digitais, Viseu, Portugal. Núcleo de Estudos em Ciências Empresariais, Universidade da Beira Interior, Covilhã, Portugal.

E-mail: pedroreis@estgv.ipv.pt

3. António Pedro Soares Pinto, Coordinator Professor, Universidade Politécnica de Viseu, Centro de Investigação em Serviços Digitais, Viseu, Portugal. Núcleo de Estudos em Ciências Empresariais, Universidade da Beira Interior, Covilhã, Portugal.

E-mail: spinto@estgv.ipv.pt

Authors' contributions:

1st author: Definition of research problem; development of hypotheses or research questions (empirical studies); development of theoretical propositions (theoretical work); definition of methodological procedures; data collection; literature review; statistical analysis; analysis and interpretation of data; critical revision of the manuscript; manuscript writing.

2nd author: Development of hypotheses or research questions (empirical studies); definition of methodological procedures; literature review; statistical analysis; analysis and interpretation of data; critical revision of the manuscript; manuscript writing.

3rd author: Development of hypotheses or research questions (empirical studies); definition of methodological procedures; literature review; statistical analysis; analysis and interpretation of data; critical revision of the manuscript; manuscript writing.