repec

Revista de Educação e Pesquisa em Contabilidade
Journal of Education and Research in Accounting

ABRACICON
ACADEMIA BRASILEIRA DE CIÊNCIAS CONTÁBEIS

Periódico Trimestral, digital e gratuito publicado pela Academia Brasileira de Ciências Contábeis | Available online at www.repec.org.br

REPeC, Brasília, v. 17, n. 2, art. 2, p. 140-160, Apr./Jun. 2023 | DOI: http://dx.doi.org/10.17524/repec.v17i2.3111 | ISSN 1981-8610

Stock Return Predictability based on Textual Sentiment Analysis: a review

Renan Silva de Carvalho
https://orcid.org/0000-0001-5475-3233

Joyce Menezes da Fonseca Tonin
https://orcid.org/0000-0003-1371-1314

Simone Leticia Raimundini Sanches
https://orcid.org/0000-0002-7363-2573

Abstract

Objective: To analyze how research addressing textual analysis as a tool to predict the return on stocks in the capital market has evolved.
Method: Systematic literature review supported by the VosViewer software; 78 empirical studies written in English and indexed in the Web of Science were analyzed.
Results: The results show evidence that textual sentiment can predict the return on stocks and be captured from varied sources of information. Four categories emerged from the analysis corresponding to sources of information for textual analysis: financial news (31), corporate disclosures (29), social media (16), and other documents (8).
Contributions: This paper contributes to the academic milieu by showing the main findings of studies on the topic and suggesting topics for future research. In a practical context, it presents to investors evidence that textual information provided by companies may also cause reactions in the market.

Keywords: Textual Analysis; Textual Sentiment; Return on Shares; Systematic Literature Review.

Published in Portuguese and English. Original Version in Portuguese.

Round 1: Received in 5/25/2022. Review requested on 11/22/2022. Round 2: Resubmited on 12/21/2022. Review requested on 2/3/2023. Round 3: Resubmited on 3/4/2023. Accepted on 5/1st/2023 by Vinícius Gomes Martins, PhD (Editor assistant) and by Gerlando Augusto Sampaio Franco de Lima, PhD (Editor). Published on 6/30/2023. Organization responsible for the journal: Abracicon

1. Introduction

The Efficient Market Hypothesis (EMH) predicts that the prices of financial assets reflect all the content of disclosed relevant information (Fama, 1970). Despite existing criticisms of this theory, such as Kahneman (1994) and Lo (2004), it is assumed that the market is, at least, partially efficient (Markiel, 2003). One of the potential reasons concerns the large volume of information disclosed in corporate reports, press releases, and the news, making it difficult to manually and thoroughly analyze such information (Xing, 2017).

In this sense, it is in the interest of investors to seek ways to automate this information analysis process, considering that they are not able to absorb all relevant information as rationality is limited (Kahneman, 1994); automating this process would enable investors to acquire some level of competitive advantage. However, much of the information disclosed is qualitative, which, at first, would make it difficult to analyze without individual attention. In this context, textual analysis emerges. As noted by Loughran and McDonald (2016), it allows extracting the sentiment of information disclosed, transforming it into quantitative data, which can be treated to, among other things, predict changes in the prices of shares.

Studies seeking to trace the relationship between textual sentiment or tone and stock returns found evidence that a change in the textual tone of annual reports significantly influences variations in stock prices over a two-day window after the disclosure (Feldman, Govindaraj, Livnat, & Segal, 2010); the linguistic tone of conference calls on results is a significant indicator of abnormal returns and subsequent trading volume (Price, Doran, Peterson, & Bliss, 2012); and high levels of uncertainty in the texts of forms filed with the Securities and Exchange Commission (SEC) before an Initial Public Offering (IPO) process, have higher returns on the first day of trading on the stock exchange (Loughran & MacDonald, 2013).

In addition to the companies' financial disclosures, the literature presents textual analyses to capture news and social media sentiment. Tetlock's (2007) findings reveal that high levels of pessimistic words in the Wall Street Journal lead to lower returns in the market the next day. Likewise, Chen, De, Hu, and Hwang (2014) show that a textual analysis of investors' opinions transmitted through social media can predict the return of companies' shares, as well as profit surprises, which occur when a company reports earnings that are significantly higher or lower than expected.

Given the ample possibility of applying textual analysis in finance and accounting, changing how texts are analyzed with advanced technology shows a need to revisit the literature and organize the main contributions and gaps left, which can be explored in future research. In this sense, Kearney and Liu (2014), Xing (2017), and Loughran and McDonald (2020) conducted reviews on the theme; the latter focused on the use of textual analysis in social media, political bias, and fraud detection.

However, unlike the previous studies, this study specifically explores textual analysis to predict stock returns. That said, the objective here is to analyze how research using the textual analysis approach to predict stock returns in the capital market progressed. A systematic literature review was performed using the VosViewer software, in which 78 empirical papers published in English and indexed in the Web of Science database were analyzed. A starting date was not established for the search, and papers published until the end of 2021 were included.

This review enabled identifying four primary sources of information for textual analysis: financial news (31), corporate disclosures (29), social media (16), and a fourth source, which represents other documents (8). This study contributes to the literature by compiling the main evidence in each category, supporting new research, and indicating potential future directions. Additionally, there are practical contributions by showing stakeholders how qualitative information can affect prices in the market.

In addition to this introduction, this paper presents four other sections. The second section presents the literature review, discussing the efficient market hypothesis, its criticisms, and the emergence of textual analysis in finance; the methodological aspects are presented in the third section; the research results are evidenced in the fourth section; and the fifth and last section contains the final considerations.

2. Theoretical framework

2.1 The Efficient Market Hypothesis, Adaptive Market Hypothesis, and Textual Analysis

Market efficiency is key in the finance literature, as the (un)acceptance of such efficiency determines the development of financial models in the capital market field. The discussion about market efficiency arises with Fama (1970), who presents the Efficient Market Hypothesis (EMH), which assumes that the prices of financial assets reflect all the content of relevant disclosed information.

Fama (1970) separates market efficiency into three types: weak efficiency, semi-strong efficiency, and strong efficiency. Later, Fama (1991) replaces their nomenclatures for the tests: I) return predictability test; II) test for event studies; and III) private information test, respectively. The change in nomenclature did not change the essence of each test; there was only a change in the weak form, which now incorporates the prediction of returns.

An efficient market in its weak form predicts that the stock price completely reflects all past information, especially those referring to asset prices, and according to Fama (1991), the dividend yield and the price/profit ratio. A semi-strong efficient market assumes that prices instantly adjust to other public information, such as annual reports and stock splits, for example. Finally, tests for strong efficiency concern whether individual or group investors can access privileged information relevant to price formation and how this information is incorporated into prices.

After Fama's (1970) study, researchers started challenging the EMH and presented evidence contrary to his theory, showing some market anomalies and questioning the existence of an efficient market. Contrary evidence shows the January effect, in which returns in January are higher than those obtained in other months of the year (Rozeff & Kinney, 1976); the Price/Earnings (P/E) effect, in which stocks with lower P/E would bring higher returns than the market (Ball, 1992); and the obtaining of abnormal returns through the application of moving averages (Brock, Lakonishok, & Lebaron, 1992), among others.

In addition to the anomalies, other theories emerged to explain how the capital market functions. Contrary to one of the main assumptions of the EMH, which is investors' rationality, behavioral finance emerges with Kahneman and Tversky (1979) as precursors. They argue that investors are irrational and that emotions interfere with their decision-making. In this sense, the Adaptive Market Hypothesis (AMH) also opposes the EMH, stating that the market works in cycles and, at times, institutional factors and behavioral aspects change market conditions (Lo, 2004).

Accordingly, Lo (2004) explains that the AMH is based on an approach intended to reconcile the neoclassical structure of the EMH economic interactions with investors' psychological factors, which sometimes result in irrational behavior, consequently affecting the market. From the AMH perspective, there are trends in the market that can generate bubbles, panic, and bullish or bearish cycles, which are phenomena routinely witnessed in natural market ecologies (Lo, 2004).

A practical implication of the AMH is that using investment strategies to obtain above-market returns may work during certain times. However, once the market cycle changes, the strategy may not work as expected and return when conditions become more favorable (Lo, 2004).

Thus, even in a competitive capital market, arbitrage opportunities arise from time to time in an "Adaptive Market." The reason, as Grossman and Stiglitz (1980) point out, is that without these opportunities, there would be no incentive to collect and analyze information and no opportunities in the financial markets, thus causing them to collapse.

One of the assumptions of AMH is that market participants absorb information at different paces (Xing, 2017). Such differences cause certain efficiency deviations that may persist for short periods, allowing additional gains if identified in advance (Markiel, 2003).

Considering the AMH, and given so much information available about companies and the market, investors are interested in streamlining the process of information analysis to take the lead in decision-making and, consequently, obtain returns above the market. Thus, textual analysis emerges, allowing empirical studies that seek, among other possibilities, to verify whether the textual sentiment of public information about the capital market can predict stock price changes (Loughran & McDonald, 2016).

Financial news transmitted by traditional media (journalistic texts) is one of the main objects of textual analysis in the finance literature, which has the seminal work of Tetlock (2007) as one of the precursors. In this same perspective, the growing popularity of social media has created opportunities for investors to share their opinions about publicly traded companies and their market views. Thus, capturing the investors' sentiment through their publications and relating it to stock returns has become a promising field.

Corporate disclosures may also contain information that allows for identifying a company's prospects, and these expectations about future performance are of great interest to investors, who constantly seek to anticipate events. Disclosures that may have their textual content analyzed range from documents provided to regulatory bodies, even before an IPO, to other mandatory disclosures, such as annual reports, or voluntary disclosures, such as earnings conference calls.

3. Methodological procedures

A systematic literature review was chosen because it is a research method that uses the literature as a data source, seeking to synthesize and report evidence on a topic, using search techniques, and systematically organizing data to reduce research bias. Therefore, this research method is in accordance with the objective proposed here.

This study has a qualitative and descriptive approach, as it describes the sample's characteristics and theoretical basis. Additionally, it is classified as a documentary, as data were collected from the literature based on papers selected from a database (Creswell, 2010).

Data were collected through a search on Google Scholar to identify terminologies commonly used in studies on this topic and establish the search terms. The keywords from Price et al. (2012) and Jiang, Lee, Martin, and Zhou (2019) were used in this preliminary search.

Based on the results obtained from Google Scholar on the keywords previously mentioned, those with the highest number of occurrences were selected – one associated with textual analysis and the other associated with market reactions. Hence, the following terms were established: *Textual Analysis* and *Stock Return*, so the papers containing both terms in their titles, abstracts, or keywords were selected.

With the terms in hand, the Web of Science database was chosen to certify that only peer-reviewed papers would be selected (Ahmad & Omar, 2016), besides being a multidisciplinary database considered relevant (Wang & Waltman, 2016).

The search on the Web of Science returned 114 results, including papers published up to the end of 2021; a starting date was not determined for the search. The abstracts of all 114 papers were read, and those presenting an empirical association between textual analysis and the stock return were selected. At this point, 36 papers were excluded for not meeting this criterion. The phases of the systematic review are summarized in Figure 1:

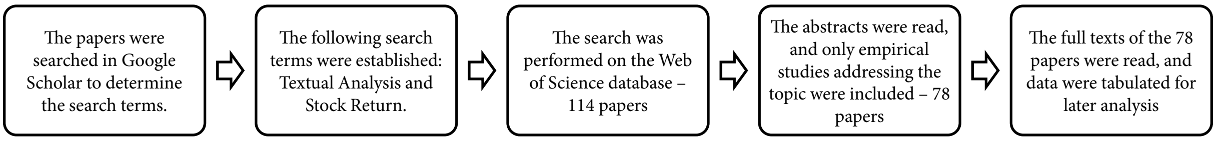

Figure 1. Systematic Review Phases

Source: developed by the author, 2022.

After this filtering process, the final sample comprised 78 papers. Figure 2 shows the progression in the number of publications on the subject over the years.

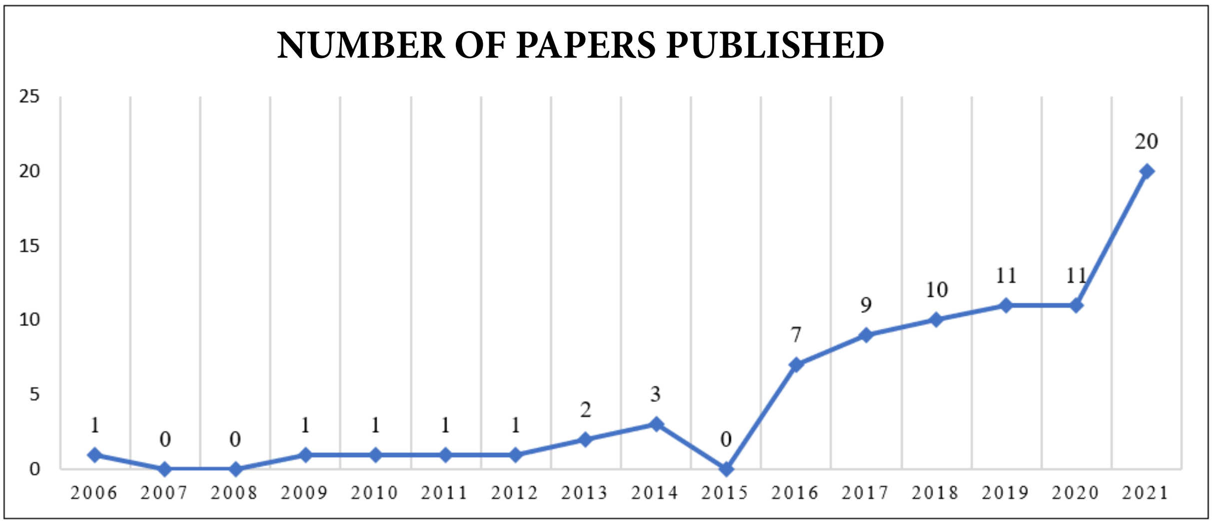

Figure 2. Number of papers published per year

Source: study's data, 2022.

Even before starting the analysis of papers, the tabulation showed a significant growth in the number of studies relating textual sentiment with the return of the shares in 2016; a continuous growth is verified from that year on up to 2020. Another leap in the number of papers on the subject is observed in 2021.

The papers were initially analyzed using the VosViewer® software. This software enabled visualizing a map with the most frequently used keywords and how they are connected, which helped to divide them into categories that represent the sources of information for the textual analysis. The full texts of papers were read to identify the documents in which the textual analysis was performed, i.e., source of information, and allocate the papers into categories.

4. Results and discussions

This section is divided into two main topics. The first presents data such as an overview of the papers, the journals where the papers were published, which countries were investigated, how research evolved over the years, and the methods adopted. The following topic deals with the studies' main contributions and research gaps. Each subtopic represents the documents that were the object of textual analysis in the sample.

4.1 Overview of the Papers Analyzed

The first part of this investigation identified the journals in which the papers were published and the impact factor of each. Table 1 shows a substantial variety of publications on the subject; the 78 papers in the sample were published in 50 different journals in the fields of accounting, finance, administration, and information technology.

Table 1
Periodicals where the papers were published

Periodical	Impact factor	Number of papers	Periodical	Impact factor	Number of papers
Journal of Banking and Finance	3.539	5	ACM Transactions on Information Systems	4.657	1
Journal of Behavioral Finance	1.798	5	Decision Sciences	4.551	1
Review of Financial Studies	8.414	4	Journal of Financial and Quantitative Analysis	4.337	1
Pacific-Basin Finance Journal	3.239	4	The Quarterly Review of Economics and Finance	4.324	1
Management Science	6.172	3	Journal of Corporate Finance	4.249	1
Review of Accounting Studies	4.011	3	Accounting. organizations and society	4.114	1
Applied Economics	1.916	3	Journal of Financial Markets	3.095	1
Journal of Financial Economics	8.238	2	Journal of Financial Research	2.811	1
International Review of Financial Analysis	8.235	2	Journal of Business Finance & Accounting	2.709	1
Knowledge-Based Systems	8.139	2	Multimedia Tools and Applications	2.577	1
Decision Support Systems	6.969	2	Accounting & Finance	2.473	1
International Review of Economics & Finance	3.399	2	Kybernetes	2.352	1
European Financial Management	2.295	2	Financial Analysts Journal	2.345	1
Computational Economics	1.741	2	Amfiteatru Economic	2.304	1

Periodical	Impact factor	Number of papers	Periodical	Impact factor	Number of papers
Journal of Intelligent & Fuzzy Systems	1.737	2	A Journal of Accounting, Finance, and Business Studies	2.060	1
The Quarterly Journal of Economics	15.563	1	Discrete Dynamics in Nature and Society	1.457	1
International Journal of Production Economics	11.251	1	Applied Economics Letters	1.287	1
Finance Research Letters	9.846	1	Revista de Administração de Empresas	1.100	1
Expert Systems with Applications	8.665	1	Lecture Notes in Computer Science	0.402	1
The Journal of Finance	7.870	1	Review of Behavioral Finance	N/A	1
Information Processing & Management	7.466	1	Intelligent Systems in Accounting, Finance and Management	N/A	1
Journal of Business Ethics	6.331	1	Global Business Review	N/A	1
Neural Computing and Applications	5.102	1	Journal of Advanced Computational Intelligence and Intelligent Informatics	N/A	1
Review of Finance	5.059	1	Qualitative Research in Financial Markets	N/A	1
Emerging Markets Finance and Trade	4.859	1	Review of Quantitative Finance and Accounting	N/A	1

Source: study's data, 2022.

Table 2 identifies the environments and markets where the studies were developed.

Table 2
Markets addressed in the papers

Market	No. of papers	Percentage (%)
United States of America	45	57.7
China	13	16.7
World	8	10.3
Taiwan	3	3.8
Germany	2	2.6
Brazil	2	2.6
India	2	2.6
England	1	1.3
Iran	1	1.3
Japan	1	1.3
Total	78	100

Source: study's data. 2022.

Note that the United States is the market most frequently investigated, representing 58% of the papers in the sample, followed by China, with 17%. Additionally, the studies involving emerging markets represent a small portion of the papers, with the possibility of studies conducted in the US being replicated in emerging markets for comparison purposes.

Eight studies were classified under "World," as they refer to studies addressing the market of several countries under different criteria. For example, Glasserman and Mamayski (2019) investigated the top 50 global banks, insurance companies, and real estate by market capitalization, thus covering several countries worldwide. Maragoudakis and Serpanos (2016) analyzed the leading European, Asian, and American stock markets. Furthermore, Anand, Basu, Pathak, and Thampy (2021) focused on the stock indexes of European Union countries.

In the next stage, a map was created in the VosViewer® software with the keywords most frequently used in the 78 studies in the sample (Figure 3) to support the preliminary analysis and identify how the subjects are divided within the topic and how matters have evolved over the years. Hence, the keywords appearing in more than ten studies were included.

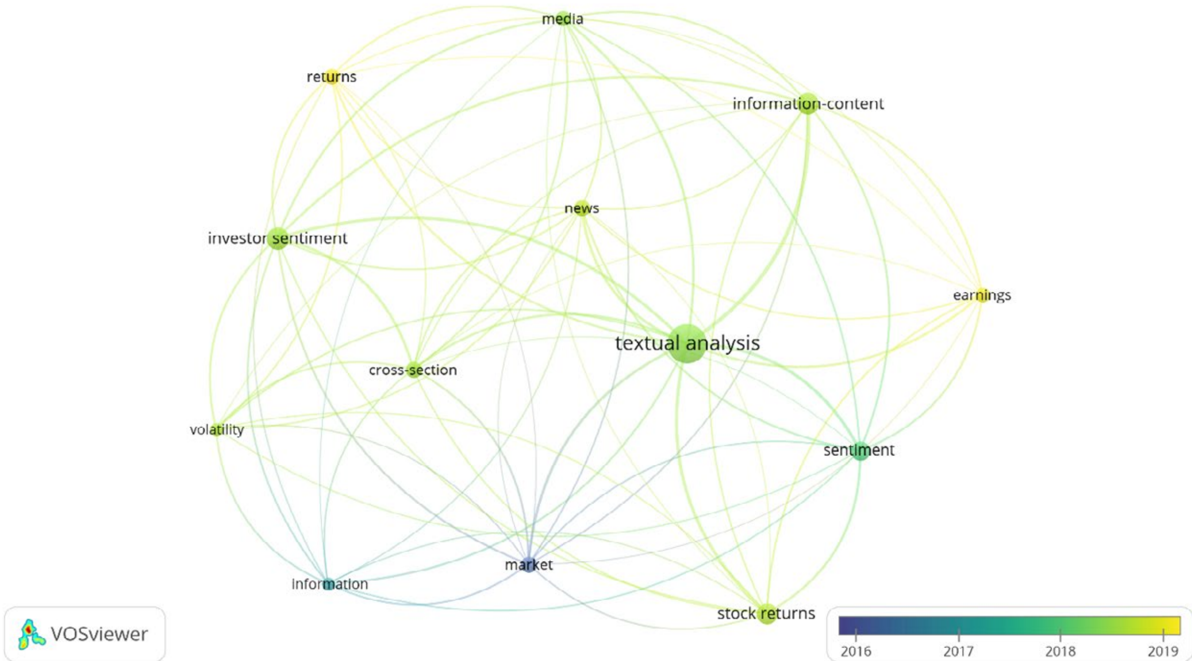

Figure 3. The progression of the papers' keywords

Source: study's data, 2022.

Figure 3 shows that the term "Textual analysis" is near the center of the map and is the keyword most frequently mentioned, with 45 occurrences, in addition to having a connection with all the other words. Note other keywords that also focus on capturing the sentiment of textual information, such as "Sentiment," "Information-content," and "Investor sentiment," in addition to the presence of "Returns," which refers to the return of shares.

There are also some keywords related to the object of textual analysis: "Media," "News," and "Earnings"; the last one refers to the sentiment in the companies' disclosure of the results. However, these words show the different possibilities of applying textual analysis so that the forecast of stock returns is not restricted to corporate documents disclosed by the company but also includes financial news; therefore, investor sentiment on social media might have predictive value.

As for how keywords evolved, most words are in shades of green, meaning they have been well distributed over the years. Among the words that refer to the most recent studies, the word "Earnings" stands out, which suggests that studies published from 2018 onwards focused on analyzing sentiment in the disclosure of results. It is worth mentioning that the software automatically created the caption with the description of the years, so it shows that research on this topic began to take shape from 2016 onwards.

Later, the papers were read to identify which documents were covered by the textual analysis. In addition to those in Figure 3, earnings conference calls stand out as the object of analysis in 10 papers. Table 3 presents the documents used for textual analysis and the frequency with which they were used.

Table 3
Objects of textual analysis

Object of analysis	Number of papers
Financial news	31
Corporate disclosures	29
Social media	16
Other documents	8

Source: study's data, 2022.

Note that the sum of the numbers in Table 3 surpasses the 78 articles that made up the sample. The reason is that some studies consulted multiple data sources, analyzing financial news in conjunction with social media (Sun, Najand & Shen, 2016; Maragoudakis, 2016; Li, Wang, Wang, Li, Liu, & Chen, 2017; Griffith, Najand, & Shen, 2019; Gan, Alexeev, Bird, & Yeung, 2020), or financial news together with corporate disclosures (Eachempati & Srivastava, 2021).

Most studies used a dictionary, classifying the words under a positive or negative tone or under other pre-established categories to capture the documents' textual sentiment. The dictionary most frequently used was that of Loughran and McDonald (2011), applied in 36 of the 78 articles analyzed. The reason is that a dictionary specifically created to capture sentiments in the financial field is preferable (Myšková, Hájek, & Olej, 2018).

In this sense, studies classifying the tone using the Loughran and McDonald dictionary (2011) and the Harvard dictionary showed that the dictionary created to be specifically used in finance is more accurate than the Harvard publication (Price et al., 2012; Ferris, Hao, & Liao, 2013; Li, Wu, & Wang, 2020).

Some researchers who chose not to use the dictionary by Loughran and McDonald (2011) created their list of words and classified the tone or sentiment of each to adapt it for their purposes. For example, Brau, Cicon, and McQueen (2016) argue that they created their word list due to the specificity of their objective, which was to analyze the strategic tone in IPO registration documents.

Other researchers did not use wordlists. It is the case of Ibriyamova, Kogan, Salganik-Shoshan, and Stolin (2017), who used semantic fingerprinting to measure the similarity between companies based on their descriptions and show that the degree of similarity helps to predict the correlation of returns on the shares of these companies. More sophisticated methods were also found to analyze textual sentiment, such as Schumaker and Chen (2009), who built a system to predict market variations by applying three textual analysis techniques through machine learning: I) Bag of words, II) noun phrases, and III) named entities. Also noteworthy is the study by Zhang, Wang, Zhu, Wang, and Ghei (2019), which proposes a hybrid neural network model for classifying and computing feelings.

Regarding the proxy for measuring stock returns, gross return for "company x" in the year/month/day/hour *t* after a particular event is predominantly used. Also important is the use of other measures, such as the Cumulative Abnormal Return (CAR), considering the difference between the gross return on the stock and the return on a market index (Kiesel, 2021), the Buy-and-Hold Return (Jiang, Lee, Martin & Zhou, 2019), and the abnormal return estimated by the market model (Campbell, Cecchini, Cianci, Ehinger, & Werner, 2019; Wu & Lin, 2017).

Studies using the excess return obtained through the risk factor model of Carhart (1997) were also found, such as Hillert, Jacobs, and Müller (2018), or the three risk factors by Fama and French (1993), and five risk factors by Fama and French (2015), adopted by Houlihan and Creamer (2017) and Buehlmaier and Whited (2018), respectively. The approach Doukas, Guo, Lam, and Xiao (2016) adopted was the market-adjusted abnormal return for short-term analysis and the Buy-and-Hold Abnormal Return (BHAR) for the long term; one can use BHAR to calculate the abnormal return of stocks with different volatility and lower bias in the long term than CAR (Yan, Xiong, Meng, & Zou, 2019).

4.2 Main Contributions and Research Possibilities

4.2.1 Financial News

The financial news was the studies' primary data source for textual analysis. The oldest paper in the sample is that of Schumaker and Chen (2009), which analyzed the news through experimental research and used a system based on machine learning. They show the correlation between the future prices of companies' shares and the news concerning these companies.

Other studies used a forecasting algorithm, with the aid of neural networks, to connect the textual sentiment of the news obtained by textual analysis to a predominantly numerical model based on technical analysis indicators to predict the intraday return of stocks (Geva & Zahavi, 2014; Li et al., 2020). The results showed that integrating market data with textual data increases the models' predictive power, according to tests conducted with 72 companies from the S&P 500 (Geva & Zahavi, 2014) and the Hong Kong Stock Exchange (Li et al., 2020).

Ahmad, Han, Hutson, Kearney, and Liu (2016) analyzed the news of 20 large companies in the United States of America (USA) published over ten years, using the dictionary of Loughran and McDonald (2011) with the words that are typically negative in finance contexts. They estimated a model with autoregressive vectors and found that the negative tone extracted from the news has a counterproductive effect on the return – which, when significant, tends to last for long periods instead of reversing in the short term.

In more recent papers, Glasserman and Mamayski (2019) show that news considered "unusual" and with negative sentiment increases volatility in the stock market, whereas "unusual" news with positive sentiment predicts lower volatility. From the same perspective, Gan et al. (2020) revealed that the link between volatility and news sentiment is more persistent than between returns and sentiment. The authors also found evidence that, from 2016 onwards, textual sentiment from social networks became a better predictor of market variations than news.

Debata, Ghate, and Renganathan (2021) analyzed sentiment in the context of the Covid-19 pandemic, specifically in 2020. They found that an index based on the intensity of Google's search volume is a more representative proxy for public sentiment than an index based on newspaper headlines. However, that paper was the only one in the sample to address the pandemic scenario, and its main limitation is that it addressed the Indian market only - thus leaving significant research gaps regarding the relationship between pandemic sentiment and the stock market.

In general, the studies that capture the textual sentiment of financial news vary considerably not only in the methods adopted, such as dictionaries, neural networks, and machine learning, but also in terms of focus. It is a consensus in the literature that news affect the prices practiced in the stock market and constitutes the source of information to which textual analysis is the most frequently applied in finance.

4.2.2 Corporate disclosures

Corporate disclosures correspond to textual documents that companies prepare and disclose. In this topic, corporate disclosures are divided into three categories: 10-K and 8-K Reports (14), Earnings Conference Calls (11), and Pre-IPO Documents (6). Again, the sum of the papers surpasses the 29 papers dealing with corporate disclosures, as some investigated more than one type of document.

By individually analyzing the return proxies used when the textual analysis was employed in these documents, CAR was found to be the most predominant measure, followed by Buy-and-Hold Return, which comprises 59% of the papers. A potential explanation is that several papers that focused on financial news and social media were published in journals in the field of information technology (IT), specifically capturing the textual sentiment with the use of technology; studies using corporate disclosures are concentrated in the fields of accounting and finance, where there is greater familiarity with more sophisticated return measures.

8-K documents deal with specific events, provide details and reports about what has been happening with a company, and are filed with the SEC. Feuerriegel & Prölloch (2021) and Filip, Lobo, Paugam, and Stolowy (2021) investigated how stock prices vary in response to the disclosure of mergers and acquisitions – the first adopted a data mining approach, more specifically a Latent Dirichlet allocation (LDA); and the second used its own list of words. Empirical evidence found in both studies showed a negative and statistically significant effect on abnormal returns in response to the disclosure of mergers and acquisitions at intervals of 1 and 3 days after the disclosure.

The effects resulting from the release of reports are not limited to short-term reactions. Cohen, Malloy, and Nguyen (2020) showed that changes in the language and construction of financial reports have substantial implications for the companies' future returns and operations. Such changes are gradually incorporated into asset prices over the 12 to 18-month period after the reporting change and have no short-term effects after the announcement.

Regarding the result of conference calls, Price et al. (2012) conducted a study to capture the tone of these conferences, separating the presentation part conducted by the administration from the questions and answers. The findings showed that the Q&A portion of these calls had significant predictive capacity for the CAR initial reaction and explanatory power during the 60 trading days after the conference call. Additional results also showed that the tone becomes more relevant in companies that do not pay dividends.

Manager sentiment analyzed in earnings conference calls is considered complementary to investor sentiment in forecasting stock returns, implying that manager sentiment has a different impact on the assessment than investor sentiment (Jiang et al., 2019). This finding allows the interaction of both feelings, obtained through different textual sources, to increase the predictive power of abnormal returns based on sentiments, which was not found in the studies analyzed in this review.

Praise made by analysts during the calls also proved to have explanatory power for stock returns. Milian and Smith (2017) examined 16,609 conference calls from S&P 500 companies and found that the number of accolades, such as "great year" and "good quarter," predicted abnormal stock returns in the next quarter. The tone of analysts and company representatives can be further explored in the literature, investigating whether they are complementary or have similar explanatory power.

Information disclosed to a regulatory body, such as the SEC in the USA, before a company goes public makes textual analysis a potential tool to predict post-IPO returns. Loughran and McDonald (2013) showed that the tone of the first SEC registration in the IPO process could predict returns on the asset's first trading day. The findings showed that high levels of uncertain text have higher returns on the first trading day.

In the same sense, Ferris, Hao, and Liao (2013) revealed that the element of conservatism in the text of IPO prospectuses of non-technology companies is inversely related to the abnormal return on a company's shares after its IPO. These results show that both the tone of the S-1 report used by Loughran and McDonald (2013) and the prospective information have explanatory power in post-IPO stock returns.

Notably, research involving corporate disclosures, whether through 10-K or 8-K reports, earnings conference calls, and pre-IPO documents, does not have developed literature such as textual analysis in traditional and social media, so there are no varied methods, markets investigated, or the objectives proposed. Such a fact enables a range of research opportunities to investigate the capital markets of other countries with different time frames and objectives.

Studies specifically involving earnings conference calls show evidence that investors are aware of companies' voluntary disclosures, considering this corporate event influences the market. Because it represents an expanding field in the literature, there are research gaps in methodological aspects, mainly in measuring textual sentiment, which can be quantified differently from those already documented in studies investigating earnings conference calls.

4.2.3 Social Media

Social media is one way to capture investors' sentiments and opinions because it is where investors, analysts, and other stakeholders share their investment analyses and/or comment on other investors' opinions. Additionally, investors' sentiments broadcast on social media might refer to a specific asset or a broader opinion about the market's direction. Both company-specific and broader market opinions can predict price changes within a five-minute window, as found by Broadstock and Zhang (2019).

Chen et al. (2014) used the dictionary by Loughran and McDonald (2011) to understand whether the opinions shared by investors could predict variations in the capital market. The results showed that the fraction of negative words in papers and comments on a popular social network among investors in the United States could negatively predict stock returns over the next three months, even controlling for the effects of analyst recommendations and traditional news media.

The language adopted in social media sometimes differs from the usual formality adopted in companies' reports and the news published by traditional media. The reason is that publications and comments on social media are disseminated by the most diverse users using informal language. Meanwhile, the news in traditional media is published by journalists who communicate facts formally. Hence, Renault (2017) built a lexicon of words commonly used by investors in social media. This lexicon had a significantly higher predictive power than the dictionaries most frequently used in the literature. The author added more complex machine learning algorithms, and the results remained competitive.

Xu, Pang, and Han (2021) opted for a different alternative to the word list to measure textual sentiment. They used an index derived from Twitter posts through textual analysis as a proxy for online sentiment, the Daily Happiness Index (DHS). Their findings revealed a power-law cross-correlation between the financial market and online sentiment in some developed countries and all developing countries analyzed in the sample, represented by Brazil and India.

The social media in which the users' feelings were captured are shown in Table 4 below. Some studies selected only one media (Broadstock & Zhang, 2019; Chen et al., 2014), while others used more than one (Griffith Najand & Shen, 2020; Gan et al., 2020). Choosing media is explained by the fact that they are the most popular among investors in the USA.

Table 4
Social Media used

Social Media	
Seeking Alpha	Yahoo!Finance
StockTwits	Sina
Twitter	EastMoney

Source: study's data, 2022.

Studies analyzing the investors' sentiment expressed in social media are more recent than those that use traditional media as a source of information. Although research on social media represents approximately 21% of the papers analyzed in this review, we need to consider the growing expansion of this type of communication, which provides new research opportunities to analyze, for example, other media used by investors.

4.2.4 Other Documents

Among the studies analyzing other documents, the Central Bank's speeches stand out, which were the object of research in two papers by Anand, Basu, Pathak, and Thampy (2021) and Möller and Reichmann (2021). The language used by Central Bank officials in public press conferences and how this factor influences stock returns in the euro area was examined using the dictionary of Loughran and McDonald (2011). The results showed that feelings of uncertainty had a significant positive effect on the intraday returns of the index of the top 50 companies in the Eurozone before the global financial crisis; i.e., uncertainty is interpreted as a sign of future political accommodation, causing the market reacts positively (Möller & Reichmann, 2021).

The textual tone of qualitative information in analyst reports also demonstrated explanatory power over 12-month abnormal returns, as evidenced by Caylor, Cecchini, and Winchel (2017). Likewise, a study conducted in the Chinese market (Li, He, Chan, 2021) shows that the disclosure of qualitative risk information of companies can negatively affect the return on shares of their suppliers, corroborating the notion that textual sentiment in documents external to the company influences stock returns.

Credit rating reports are another document, the textual tone of which can influence variations in stock prices. Agarwal, Chen, and Zhang (2016) identified that net linguistic tone (negative tone minus positive tone) in reports is significantly and negatively related to abnormal stock returns, in addition to being able to predict future rating variations. On the other hand, the empirical evidence found by Kiesel (2021) among companies listed in the United States and Europe, does not provide clear conclusions about whether rating updates can affect stock returns.

Thus, regarding credit rating reports, future studies are suggested to use other alternatives to explore the measurement of textual sentiment, seeking to resolve this impasse on the potential influence of the tone of reports on stock prices. Additionally, all the studies investigating credit rating reports, analyst reports, and Central Bank speeches are concentrated on the American in European markets. For comparison purposes, future research may obtain empirical evidence on the relationship between textual tone and stock returns in the documents cited in emerging markets.

5. Final considerations

This systematic literature review provides an overview of how research addressing textual analysis as a tool for predicting stock returns in the capital market has evolved. A total of 78 studies addressing this topic were analyzed. They were all published in English in 50 different journals. Analysis was conducted using the VosViewer® software, and the papers were classified into categories representing the documents in which textual analysis was applied.

Four primary sources of information used in textual analysis emerged from this classification: financial news (31), corporate disclosures (29), social media (16), and a fourth source, which represents other documents (8). The main evidence found in each category was compiled, and potential directions are indicated for future research.

Analysis of the sample showed that most papers used a dictionary to classify words according to tone and capture the documents' textual sentiment. The dictionary by Loughran and McDonald (2011) was the most frequently adopted; 36 out of the 78 papers used it. Regarding the proxy applied to capture market reactions, the most frequently used was the gross return of stocks. Most studies included in this review were conducted in the American capital market, representing approximately 58% of the papers addressed here.

Studies classified under financial news and social media are the most representative in the sample (47 papers) and present a greater variety of methods, samples, and specificities than the other categories. Therefore, financial news and social media represent the primary sources of information for applying textual analysis in finance when the objective is to associate the textual tone with market reactions.

As for the other categories, there are documents involving the companies' disclosures, which may be voluntary (e.g., earnings conference calls) or mandatory (10-K, 8-K reports, and pre-IPO documents). In these cases, the studies have some differences, showing that more research is needed to explore the various gaps. Furthermore, investigating developing countries is an opportunity for future research, as well as studies comparing between developed and developing countries.

There are also studies classified under other documents (8), i.e., documents not disclosed by the companies and not referring to financial news or social media. The sentiment of the Central Bank's speeches stands out in this category, which proved to be related to the return of shares. However, only three studies were found, all in the European environment. Hence, there is the possibility of further studies addressing the relationship of such disclosure with stock prices.

In general, evidence found in the papers addressed here shows that textual sentiment can predict the return of shares, being captured from the most diverse sources of information. Therefore, the EMH cannot explain these results, considering that this theory is premised on the investors' rationality. However, the EMH can still be tested under conditions different from those addressed in the sample studies.

This study contributes to the dissemination of the systematic literature review technique in research conducted in the fields of finance and accounting. Although there is an increase in the number of publications, this technique is not widely disseminated in this field of knowledge. In the academic milieu, this study contributes to systematizing this topic's state of the art. It can be helpful to those intending to conduct similar studies in the future. As practical implications, investors can learn how textual sentiment affects their stock investments.

One of this study's limitations is that only papers written in English were included. Second, the Web of Science was the only database searched; hence, relevant studies not indexed by this database may have been omitted. Consequently, further research is suggested to investigate other databases and include studies written in other languages. Additionally, future studies are suggested to explore this topic by categorizing studies considering geographic regions, which could provide new insights and enrich the debate on the influence of textual sentiment on stock returns in different regional contexts.

References

Agarwal, S., Chen, V. Y. S., & Zhang, W. (2016). The information value of credit rating action reports: a textual analysis. *Management Science*, *62*(8), 2218-2240. Recuperado de https://doi.org/10.1287/mnsc.2015.2243

Ahmad, K., Han, J., Hutson, E., Kearney, C., & Liu, S. (2016). Media-expressed negative tone and firm-level stock returns. *Journal of Corporate Finance*, *37*, 152-172. Recuperado de https://doi.org/10.1016/j.jcorpfin.2015.12.014

Ahmad, S., & Omar, R. (2016). Basic corporate governance models: a systematic review. *International Journal of Law and Management*, *58*(1), 73-107. Recuperado de https://doi.org/10.1108/ijlma-10-2014-0057

Anand, A., Basu, S., Pathak, J., & Thampy, A. (2021). The impact of sentiment on emerging stock markets. *International Review of Economics & Finance,* 75(1), 161–177. Recuperado de https://doi.org/10.1016/j.iref.2021.04.005

Anand, A., Basu, S., Pathak, J., & Thampy, A. (2021). Whose speeches impact European markets: ECB's or the national central banks'? *European Financial Management*, 28(5), 1413–1476). Recuperado de https://doi.org/10.1111/eufm.12334

Ball, R. (1992). The earnings-price anomaly. *Journal of Accounting and Economics*, *15*(2-3), 319-345. Recuperado de https://doi.org/10.1016/0165-4101(92)90023-u

Bian, S., Jia, D., Li, R., Sun, W., Yan, Z., & Zheng, Y. (2021). Can management tone predict IPO performance? – Evidence from mandatory online roadshows in China. *Pacific-Basin Finance Journal,* 68(1), 101588). Recuperado de https://doi.org/10.1016/j.pacfin.2021.101588

Boudoukh, J., Feldman, R., Kogan, S., & Richardson, M. (2018). Information, trading, and volatility: evidence from firm-specific news. *The Review of Financial Studies*, *32*(3), 992-1033. Recuperado de https://doi.org/10.1093/rfs/hhy083

Box, T. (2018). Qualitative similarity and stock price comovement. *Journal of Banking & Finance*, *91*, 49-69. Recuperado de https://doi.org/10.1016/j.jbankfin.2018.04.010

Brau, J. C., Cicon, J., & McQueen, G. (2016). Soft strategic information and IPO underpricing. *Journal of Behavioral Finance*, *17*(1), 1-17. Recuperado de https://doi.org/10.1080/15427560.2016.1133619

Broadstock, D. C., & Zhang, D. (2019). Social-media and intraday stock returns: the pricing power of sentiment. *Finance Research Letters*, *30*, 116-123. Recuperado de https://doi.org/10.1016/j.frl.2019.03.030

Brochet, F., Kolev, K., & Lerman, A. (2018). Information transfer and conference calls. *Review of Accounting Studies*, *23*(3), 907–957. Recuperado de https://doi.org/10.1007/s11142-018-9444-4

Brock, W., Lakonishok, J., & Lebaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *The Journal of Finance*, *47*(5), 1731-1764. Recuperado de https://doi.org/10.1111/j.1540-6261.1992.tb04681.x

Brockman, P., Li, X., & Price, S. M. (2017). Conference Call tone and stock returns: evidence from the stock exchange of Hong Kong. *Asia-Pacific Journal of Financial Studies*, *46*(5), 667-685. Recuperado de https://doi.org/10.1111/ajfs.12186

Buehlmaier, M. M. M., & Whited, T. M. (2018). Are financial constraints priced? Evidence from textual analysis. *The Review of Financial Studies*, *31*(7), 2693-2728. Recuperado de https://doi.org/10.1093/rfs/hhy007

Campbell, J. L., Cecchini, M., Cianci, A. M., Ehinger, A. C., & Werner, E. M. (2018). Tax-related mandatory risk factor disclosures, future profitability, and stock returns. *Review of Accounting Studies*, *24*(1), 264-308. Recuperado de https://doi.org/10.1007/s11142-018-9474-y

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57-82. Recuperado de https://doi.org/10.1111/j.1540-6261.1997.tb03808.x

Carlini, F., Cucinelli, D., Previtali, D., & Soana, M. G. (2020). Don't talk too bad! stock market reactions to bank corporate governance news. *Journal of Banking & Finance* (Vol. 121, p. 105962). Recuperado de https://doi.org/10.1016/j.jbankfin.2020.105962

Caylor, M., Cecchini, M., & Winchel, J. (2017). Analysts' qualitative statements and the profitability of favorable investment recommendations. *Accounting, organizations and society*, *57*, 33-51. Recuperado de https://doi.org/10.1016/j.aos.2017.03.005

Chen, H., De, P., Hu, Y. (Jeffrey), & Hwang, B.-H. (2014). Wisdom of crowds: the value of stock opinions transmitted through social media. *Review of Financial Studies*, *27*(5), 1367-1403. Recuperado de https://doi.org/10.1093/rfs/hhu001

Chen, J., Demers, E., & Lev, B. (2018). Oh what a beautiful morning! Diurnal influences on executives and analysts: evidence from conference calls. *Management Science*, *64*(12), 5899-5924. Recuperado de https://doi.org/10.1287/mnsc.2017.2888

Chen, S., Liu, T., Peng, Q., & Zhao, Y. (2021). Manager Sentiment Bias and Stock Returns: Evidence from China. *Emerging Markets Finance and Trade*, 58(3), 823–836. Recuperado de https://doi.org/10.1080/1540496x.2021.1918543

Cohen, L., Malloy, C., & Nguyen, Q. (2020). Lazy Prices. *The Journal of Finance*, *75*(3), 1371-1415. Recuperado de https://doi.org/10.1111/jofi.12885

Debata, B., Ghate, K., & Renganathan, J. (2021). COVID-19 pandemic sentiment and stock market behavior: evidence from an emerging market. *Review of Behavioral Finance*. Recuperado de https://doi.org/10.1108/rbf-05-2021-008

Doukas, A. J., Guo, J. M., Lam, H. Y. T., & Xiao, S. H. (2014). Media endorsements of new product announcements: a new marketing strategy. *European Financial Management*, *22*(3), 394-426. Recuperado de https://doi.org/10.1111/eufm.12057

Eachempati, P., & Srivastava, P. R. (2021). Accounting for investor sentiment in news and disclosures. *Qualitative Research in Financial Markets*, *14*(1), 53-75. Recuperado de https://doi.org/10.1108/qrfm-01-2020-0006

Fama, E. F. (1970). Efficient capital markets: a review of theory and empirical work. *The Journal of Finance*, *25*(2), 383-417. Recuperado de https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56. Recuperado de https://doi.org/10.1016/0304-405x(93)90023-5

Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, *116*(1), 1-22. Recuperado de https://doi.org/10.1016/j.jfineco.2014.10.01

Feldman, R., Govindaraj, S., Livnat, J., & Segal, B. (2009). Management's tone change, post earnings announcement drift and accruals. *Review of Accounting Studies*, *15*(4), 915-953. Recuperado de https://doi.org/10.1007/s11142-009-9111-x

Ferris, S. P., Hao, (Grace) Qing, & Liao, (Stella) Min-Yu. (2012). The effect of issuer conservatism on IPO pricing and performance. *Review of Finance*, *17*(3), 993-1027. Recuperado de https://doi.org/10.1093/rof/rfs018

Feuerriegel, S., & Prendinger, H. (2016). News-based trading strategies. *Decision Support Systems*, 90, 65–74. Recuperado de https://doi.org/10.1016/j.dss.2016.06.020

Feuerriegel, S., & Pröllochs, N. (2018). Investor Reaction to Financial Disclosures across Topics: An Application of Latent Dirichlet Allocation. *Decision Sciences*, *52*(3), 608–628. Recuperado de https://doi.org/10.1111/deci.12346

Filip, A., Lobo, G. J., Paugam, L., & Stolowy, H. (2021). Disclosures About Key Value Drivers in M&A Announcement Press Releases: An Exploratory Study. *Abacus*, *58*(1), 62–104. Recuperado de https://doi.org/10.1111/abac.12234

Galdi, F. C., & Gonçalves, A. M. (2018). Pessimism and uncertainty of the news and investor's behavior in Brazil. *Revista de Administração de Empresas*, *58*(2), 130-148. Recuperado de https://doi.org/10.1590/s0034-759020180203

Gan, B., Alexeev, V., Bird, R., & Yeung, D. (2020). Sensitivity to sentiment: news vs social media. *International Review of Financial Analysis*, *67*, 101390. Recuperado de https://doi.org/10.1016/j.irfa.2019.101390

Garcia-Lopez, F. J., Batyrshin, I., & Gelbukh, A. (2018). Analysis of relationships between tweets and stock market trends [JB*]. Journal of Intelligent & Fuzzy Systems*, *34*(5), 3337-3347. Recuperado de https://doi.org/10.3233/JIFS-169515

Geva, T., & Zahavi, J. (2014). Empirical evaluation of an automated intraday stock recommendation system incorporating both market data and textual news. *Decision Support Systems*, *57*, 212–223. Recuperado de https://doi.org/10.1016/j.dss.2013.09.013

Glasserman, P., & Mamaysky, H. (2019). Does Unusual news forecast market stress? *Journal of Financial and Quantitative Analysis*, *54*(5), 1937-1974. Recuperado de https://doi.org/10.1017/s0022109019000127

Griffith, J., Najand, M., & Shen, J. (2019). Emotions in the stock market. *Journal of Behavioral Finance*, *21*(1), 42-56. Recuperado de https://doi.org/10.1080/15427560.2019.1588275

Gu, W., Zhang, L., Xi, H., & Zheng, S. (2021). Stock Prediction Based on News Text Analysis. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *25*(5), 581–591. Recuperado de https://doi.org/10.20965/jaciii.2021.p058

Hájek, P. (2017). Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. *Neural Computing and Applications*, *29*(7), 343-358. Recuperado de https://doi.org/10.1007/s00521-017-3194-2

Hassan, T. A., Hollander, S., van Lent, L., & Tahoun, A. (2019). Firm-level political risk: measurement and effects. *The Quarterly Journal of Economics*, *134*(4), 2135-2202. Recuperado de https://doi.org/10.1093/qje/qjz021

Hatefi Ghahfarrokhi, A., & Shamsfard, M. (2020). Tehran stock exchange prediction using sentiment analysis of online textual opinions. *Intelligent Systems in Accounting, Finance and Management*, *27*(1), 22-37. Recuperado de https://doi.org/10.1002/isaf.1465

He, Y., Qu, L., Wei, R., & Zhao, X. (2021). Media-based investor sentiment and stock returns: a textual analysis based on newspapers. *Applied Economics*, *54*(7), 774–792. Recuperado de https://doi.org/10.1080/00036846.2021.1966369

Hendricks, B. E., Lang, M., & Merkley, K. (2021). Through the eyes of the founder: CEO characteristics and firms' regulatory filings. *Journal of Business Finance &Amp; Accounting*, *49*(3–4), 383–422. Recuperado de https://doi.org/10.1111/jbfa.12572

Hillert, A., Jacobs, H., & Müller, S. (2018). Journalist disagreement. *Journal of Financial Markets*, *41*, 57-76. Recuperado de https://doi.org/10.1016/j.finmar.2018.09.002

Houlihan, P., & Creamer, G. G. (2017). Can sentiment analysis and options volume anticipate future returns? *Computational Economics*, *50*(4), 669-685. Recuperado de https://doi.org/10.1007/s10614-017-9694-4

Huang, A. G., Tan, H., & Wermers, R. (2020). Institutional Trading around Corporate News: Evidence from Textual Analysis. *The Review of Financial,* 33(10), 4627–4675). Recuperado de https://doi.org/10.1093/rfs/hhz136

Ibriyamova, F., Kogan, S., Salganik-Shoshan, G., & Stolin, D. (2016). Using semantic fingerprinting in finance. *Applied Economics*, *49*(28), 2719-2735. Recuperado de https://doi.org/10.1080/00036846.2016.1245844

Ibriyamova, F., Kogan, S., Salganik-Shoshan, G., & Stolin, D. (2018). Predicting stock return correlations with brief company descriptions. *Applied Economics*, *51*(1), 88-102. Recuperado de https://doi.org/10.1080/00036846.2018.1494377

Jiang, F., Lee, J., Martin, X., & Zhou, G. (2019). Manager sentiment and stock returns. *Journal of Financial Economics*, *132*(1), 126-149. Recuperado de https://doi.org/10.1016/j.jfineco.2018.10.001

Kahneman, D. (1994). New Challenges to the Rationality Assumption. *Journal of Institutional and Theoretical Economics (JITE)*, *150*(1), 18-36. Recuperado de http://www.jstor.org/stable/40753012

Kearney, C., & Liu, S. (2014). Textual sentiment in finance: a survey of methods and models. *International Review of Financial Analysis*, *33*, 171-185. Recuperado de https://doi.org/10.1016/j.irfa.2014.02.006

Kiesel, F. (2021). It's the tone, stupid! Soft information in credit rating reports and financial markets. *Journal of Financial Research*, *44*(3), 553–585. Recuperado de https://doi.org/10.1111/jfir.12250

Li, J., Li, J., & Zhu, X. (2020). Risk dependence between energy corporations: a text-based measurement approach. *International Review of Economics & Finance*, *68*, 33-46. Recuperado de 10.1016/j.iref.2020.02.009

Li, Q., Wang, J., Wang, F., Li, P., Liu, L., & Chen, Y. (2016). The role of social sentiment in stock markets: a view from joint effects of multiple information sources. *Multimedia Tools and Applications, 76*(10), 12315-12345. Recuperado de https://doi.org/10.1007/s11042-016-3643-4

Li, S., Ning, K., & Zhang, T. (2021). Sentiment-aware jump forecasting. *Knowledge-Based Systems*, *228*, 107292. Recuperado de https://doi.org/10.1016/j.knosys.2021.107292

Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: a case of Hong Kong. *Information Processing & Management*, *57*(5), 102212. Recuperado de https://doi.org/10.1016/j.ipm.2020.102212

Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. *Knowledge-Based Systems*, *69*, 14-23. Recuperado de https://doi.org/10.1016/j.knosys.2014.04.022

Li, Y., He, J., & Chan, K. C. (2021). Information transmission along supply chains: Stock price reaction of suppliers upon a customer's release of qualitative risk information. *International Journal of Production Economics*, *239*, 108189. Recuperado de https://doi.org/10.1016/j.ijpe.2021.108189

Liu, S., & Han, J. (2020). Media tone and expected stock returns. *International Review of Financial Analysis*, *70*, 101522. Recuperado de https://doi.org/10.1016/j.irfa.2020.101522

Lo, A. W. (2004). The adaptive markets hypothesis. *The Journal of Portfolio Management*, *30*(5), 15-29. Recuperado de https://doi.org/10.3905/jpm.2004.442611

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. Recuperado de https://doi.org/10.1111/j.1540-6261.2010.01625.x

Loughran, T., & McDonald, B. (2011b). Barron's Red Flags: Do They Actually Work? *Journal of Behavioral Finance*, *12*(2), 90-97. Recuperado de https://doi.org/10.1080/15427560.2011.575971

Loughran, T., & McDonald, B. (2013). IPO first-day returns, offer price revisions, volatility, and form S-1 language. *Journal of Financial Economics*, *109*(2), 307-326. Recuperado de https://doi.org/10.1016/j.jfineco.2013.02.017

Loughran, T., & McDonald, B. (2020). Textual Analysis in Finance. *Annual Review of Financial Economics*, *12*(1), 357-375. Recuperado de https://doi.org/10.1146/annurev-financial-012820-032249

Lutz, B., Pröllochs, N., & Neumann, D. (2020). Predicting sentence-level polarity labels of financial news using abnormal stock returns. *Expert Systems with Applications*, *148*, 113223. Recuperado de https://doi.org/10.1016/j.eswa.2020.113223

Lv, Y., Piao, J., Li, B., & Yang, M. (2021). Does online investor sentiment impact stock returns? Evidence from the Chinese stock market. *Applied Economics Letters*, *29*(15), 1434–1438. Recuperado de https://doi.org/10.1080/13504851.2021.1937490

Malaquias, R. F., & Júnior, D. M. B. (2021). Positive Tone in Management Reports and Volatility of Stock Returns. *Global Business Review*, 097215092110542. Recuperado de https://doi.org/10.1177/09721509211054279

Mangee, N. (2018). Stock Returns and the Tone of Marketplace Information: does Context matter? *Journal of Behavioral Finance*, *19*(4), 396–406. Recuperado de https://doi.org/10.1080/15427560.2018.1405268

Maragoudakis, M., & Serpanos, D. (2015). Exploiting financial news and social media opinions for stock market analysis using MCMC Bayesian Inference. *Computational Economics*, *47*(4), 589-622. Recuperado de https://doi.org/10.1007/s10614-015-9492-9

Marquez-Illescas, G., Zebedee, A. A., & Zhou, L. (2018). Hear Me Write: does CEO narcissism affect disclosure? *Journal of Business Ethics*, *159*(2), 401-417. Recuperado de https://doi.org/10.1007/s10551-018-3796-3

Milian, J. A., & Smith, A. L. (2017). An investigation of analysts' praise of management during earnings conference calls. *Journal of Behavioral Finance*, *18*(1), 65-77. Recuperado de https://doi.org/10.1080/15427560.2017.1276068

Möller, R., & Reichmann, D. (2021). ECB language and stock returns – A textual analysis of ECB press conferences. *The Quarterly Review of Economics and Finance*, *80*, 590–604. Recuperado de https://doi.org/10.1016/j.qref.2021.04.003

Myšková, R., & Hájek, P. The effect of managerial sentiment on market to-book ratio. *Transformations in Business & Economics*, *15*(2A), 498-513. Recuperado de https://dk.upce.cz/handle/10195/67653

Myšková, R., Hájek, P., & Olej, V. (2017) Predicting abnormal stock return volatility using textual analysis of news - A meta-learning approach. *Amfiteatru Economic*, *20*(47), 185-201. Recuperado de https://www.ceeol.com/search/article-detail?id=678584

Price, S. M., Doran, J. S., Peterson, D. R., & Bliss, B. A. (2012). Earnings conference calls and stock returns: the incremental informativeness of textual tone. *Journal of Banking & Finance*, *36*(4), 992–1011. Recuperado de https://doi.org/10.1016/j.jbankfin.2011.10.013

Renault, T. (2017). Intraday online investor sentiment and return patterns in the U.S. stock market. *Journal of Banking & Finance*, *84*, 25-40. Recuperado de https://doi.org/10.1016/j.jbankfin.2017.07.002

Rozeff, M. S., & Kinney, W. R., Jr. (1976). Capital market seasonality: the case of stock returns. *Journal of Financial Economics*, *3*(4), 379-402. Recuperado de https://doi.org/10.1016/0304-405x(76)90028-3

Sampaio, R.F., & Mancini, M.C. (2007). Estudos de revisão sistemática: um guia para síntese criteriosa da evidência científica. *Revista Brasileira de Fisioterapia, 11*(1),83-89. Recuperado de https://www.redalyc.org/articulo.oa?id=2350/235016477013

Sampath, V. S., O'Connor, A. J., & Legister, C. (2021). Moral leadership and investor attention: An empirical assessment of the potus's tweets on firms' market returns. *Review of Quantitative Finance and Accounting*, *58*(3), 881–910. Recuperado de https://doi.org/10.1007/s11156-021-01012-0

Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news. *ACM Transactions on Information Systems*, *27*(2), 1-19. Recuperado de https://doi.org/10.1145/1462198.1462204

Seng, J.-L., & Yang, H.-F. (2017). The association between stock price volatility and financial news – a sentiment analysis approach. *Kybernetes*, *46*(8), 1341-1365. Recuperado de https://doi.org/10.1108/k-11-2016-0307

Sun, L., Najand, M., & Shen, J. (2016). Stock return predictability and investor sentiment: a high-frequency perspective. *Journal of Banking & Finance*, *73*, 147–164. Recuperado de https://doi.org/10.1016/j.jbankfin.2016.09.010

Suslava, K. (2021). "Stiff Business Headwinds and Uncharted Economic Waters": The Use of Euphemisms in Earnings Conference Calls. *Management Science*, *67*(11), 7184–7213. Recuperado de https://doi.org/10.1287/mnsc.2020.3826

Takahashi, S., Takahashi, M., Takahashi, H., & Tsuda, K. (2006, october). Analysis of stock price return using textual data and numerical data through text mining. In *Lecture Notes in Computer Science*. Recuperado de https://doi.org/10.1007/11893004_40

Wang, Q., & Waltman, L. (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of science and scopus. *Journal of Informetrics*, *10*(2), 347-364. Recuperado de https://doi.org/10.1016/j.joi.2016.02.003

Wu, C.-H., & Lin, C.-J. (2017). The impact of media coverage on investor trading behavior and stock returns. *Pacific-Basin Finance Journal*, *43*, 151-172. Recuperado de https://doi.org/10.1016/j.pacfin.2017.04.001

Xu, K., Pang, Y., & Han, J. (2021). Dynamic Cross-Correlation between Online Sentiment and Stock Market Performance: A Global View. *Discrete Dynamics in Nature and Society*, *2021*, 1–11. Recuperado de https://doi.org/10.1155/2021/6674379

Yan, Y., Xiong, X., Meng, J. G., & Zou, G. (2019). Uncertainty and IPO initial returns: evidence from the tone analysis of China's IPO prospectuses. *Pacific-Basin Finance Journal*, *57*, 101075. Recuperado de https://doi.org/10.1016/j.pacfin.2018.10.004

Yang, A. S. (2020). Misinformation corrections of corporate news: corporate clarification announcements. *Pacific-Basin Finance Journal*, *61*, 101315. Recuperado de https://doi.org/10.1016/j.pacfin.2020.101315

Zhang, W., Wang, M., Zhu, Y., Wang, J., & Ghei, N. (2019). A hybrid neural network approach for fine-grained emotion classification and computing [JB]. *Journal of Intelligent & Fuzzy Systems*, *37*(3), 3081-3091. Recuperado de https://doi.org/10.3233/JIFS-179111

Zhao, W., Zhang, W., Xiong, X., & Zou, G. (2019). Share pledges, tone of earnings communication conferences, and market reaction: evidence from China. *Accounting & Finance*, *59*(5), 2817-2853. Recuperado de https://doi.org/10.1111/acfi.12585