What drives the release of material facts for Brazilian stocks?

Marcos Henrique Reichert[†]
Marcelo Scherer Perlin[‡]

Abstract In this study we look at the determinants of material facts (*Fatos Relevantes*) in the Brazilian Market. Following local legislation, material facts should be released to the public right after its occurrence, and preferably, after trading time. We investigate the randomness of the release of material facts—and release strategies by executives—and test whether there is a particular time period where more or fewer material facts are published. We also investigate whether the content of the material fact—positive or negative sentiment—explains different strategies regarding the release of news to the market. Lastly, using Vector Auto Regressions, we test for a feedback effect between material facts and the financial data, that is, if material facts publishing drives the returns, volume and volatility and vice versa. Finally, our results show that volume, volatility and returns (to a lesser extent) are determinants for material facts publishing and material facts.

Keywords: Material facts; Investor sentiment; Determinants.
JEL Code: G14, G41.

1. Introduction

According to Brazilian Corporate law—*Leis das SA* of 1976, article 157, paragraph 3—companies traded at B3 must immediately report to the stock exchange and disclose to the press any resolution passed by its general meeting, or by its board of directors, or any material fact occurred or related to its business that may have a significant influence in the market investors' decision to sell or buy any equity issued by the firm. Some examples of material facts include company divisions, mergers, regulatory interventions, debt renegotiations, the discovery of new resources or technologies, forecast revisions and approvals or cancellations of investment projects. The Security and Exchange Commission (CVM, which is the regulatory authority of security markets in Brazil) requires that publicly traded companies fully disclose material facts. Publicly traded firms must send to CVM structured files describing the material facts, electronically and before disclosures in other communication channels. After that, the information is immediately displayed on the websites of B3 and CVM, and made available to market participants.

Submitted on January 30, 2022. Accepted on February 23, 2022. Published online in June 2022. Editor in charge: Bruno Giovannetti.
[†]Escola de Administração, UFRGS, Brazil: marcoshenriquereichert@gmail.com
[‡]Escola de Administração, UFRGS, Brazil: marcelo.perlin@ufrgs.br

The mandatory disclosure of information reduces insider trading, which arises from informational asymmetries in stock markets (de Carvalho et al., 2016). Price volatility results the interaction between the vast number of participants in Financial Markets. Upon receiving new information, it's expected that an investor will act accordingly to it. However, in reality, it's impossible to know how each individual will act after receiving this new information. This casts doubt upon how long it takes the investor to receive and react to this new information. If part of the investors gets this information earlier than other parties, or even before it is publicly released, there is a source of informational asymmetry or insider trading.

As such, companies' disclosures are recommended to be sent before or after the trading hours to avoid excessive price volatility. Although, this is not always the case as a large number of material facts are published during trading hours. It casts doubts on whether material facts can be published as a means to raise returns or volatility, or even as a response to their variation. In the literature, we find that de Carvalho et al. (2016) results suggest there is some anticipation before releasing new material facts, with a surge of shares traded and price reactions up to 4 minutes before their disclosure. This hints that material facts releases aren't published immediately or randomly.

Previous studies displayed the effects of material facts on price both in Brazil (de Carvalho et al., 2016; Marques et al., 2011; Silva and da Silva Felipe, 2010; Damascena et al., 2017) and internationally (Patell and Wolfson, 1984; Barclay and Litzenberger, 1988). There are even studies on the material facts' readability (Silva and Fernandes, 2009), but the authors of this article found no previous research on material facts' determinants. Thus, this article uses high frequency trading data, daily data, and material facts set by publicly traded companies to test for material facts' determinants. We have also checked for the market's sentiment towards material facts.

The motivation for the present paper is based on the well documented effect of material facts on stock prices and volume. It's interesting to note that managers can publish positive material facts when the stock prices are low or wait to publish negative material facts when the stock is high. Managers can also wait to publish positive material facts at the end of fiscal quarters, before investor meetings or wait even longer to not publish a material fact during these months. Managers could also schedule material facts publishing to Monday or Friday if they want the market to absorb (or not) the new information. Besides, we'll also check if there is a feedback effect between the material facts and stock returns, volume and volatility.

This article aims to identify the determinants for the publishing of material facts. For example, we want to test whether there are any preferences for month, weekday or hour for a new material fact publishing. We also want to check for feedback effect between stock's volume, returns and volatility and material facts' publishing and sentiment. This research is the first to be focused on these subjects.

Unlike the previous literature on the impact of new information using high frequency trading data, this research uses text mining for material facts' sentiment classification, instead of a subjective one (de Carvalho et al., 2016). Along with de Carvalho et al. (2016), this article is one of the few that uses high frequency trading data in the Brazilian Market, and it is the first one to use this high amount of data.

Our paper is organized: Section two discusses the literature about the market efficiency hypothesis and investor perception, highlighting studies of intraday data; Section three describes the database; Section four focuses on the methodology and section five presents the results. Finally, section six holds our final considerations.

2. Review

One landmark in the financial literature is the Efficient Market Hypothesis (EMH) (Malkiel and Fama, 1970), which defines that investors are assumed to be rational in valuing financial securities by incorporating all the available information. It also defines that if irrational investors are present, they should trade randomly; therefore, their trades cancel each other out without affecting the prices. The effect of irrational investors on prices is also eliminated by the trading activities of arbitrageurs. Also, according to the EMH, stock market prices are mostly driven by new information, rather than present and past prices. Since there is no way to predict the news, stock market prices would follow a random walk pattern and cannot be wholly predicted (Fama, 1965; Fama et al., 1969).

The work of de Carvalho et al. (2016) has a lot of similarities to our study. They have made an event study to analyze a small sample of material facts, searching for abnormal returns and testing EMH's semi-strong market efficiency. What they have discovered was that material facts reveal new information to investors. The results showed that the investor could take up to 50 minutes to react to information in the Material Fact and that some investors used this time frame to profit. Also, they had found an increase in negotiations before new material facts were released to the market.

Marques et al. (2011) test if material facts impact Bovespa Novo Mercado's stocks.[1] They analyzed 78 material facts and discovered that only

[1] B3 was previously called Bovespa.

15 of them had any effect on the stock prices. This goes against the semi-strong EMH. Silva and da Silva Felipe (2010) analyze how the wording of the material facts affects the stock prices. That is, they categorized material facts as optimistic and pessimistic and tested the period before and after their publication. The results show that the Brazilian Stock Market didn't react to optimistic material facts, while the stock prices decreased their abnormal returns after pessimistic material facts. But Damascena et al. (2017) support EMH's semi-strong form. The new information in Material Facts had an initial impact on the studied stock, but the stock price went back to normal as the market absorbed this new information.

While previous studies analyzed daily data, using High Frequency Data is justified by verifying the different intervals of the day and events' immediate effects. The following studies use High Frequency Data to test the effect of new information in the stock market. Patell and Wolfson (1984) analyzed the effects of announcements of revenues and dividends in the New York Stock Exchange (NYSE) using event studies. But, Barclay and Litzenberger (1988) analyzed announcements of debt and equity issues, while Busse and Green (2002) analyzed the stocks of companies 15 seconds after a daily announcement about their situation on American television shows. Drienko and Sault (2013) inspected companies' abnormal returns after Australian Stock Exchange announcements. In Brazil, the article by de Carvalho et al. (2016) is one of the few cases of studies that use High Frequency Trading Data to assess the impact of new information on investors.

However, how does the investor perceive this new information? To answer this, we can use textual analysis. The first instances of textual analysis use trace back to the 1300s, but the first cases of using textual analysis in finance are more recent. For example, one use of textual analysis is the readability of companies' reports and material facts. Li (2008), finds that companies with lower reported earnings publish annual reports with lower readability, i.e., are harder to read, which could be explained by the need for longer sentences to explain the situation to investors. Lehavy et al. (2011) find that companies with better readability have more analysts following them and a better forecast of the annual earnings. Using a different index, Silva and Fernandes (2009) measures the readability of Brazilian Material Facts from 2002 to 2006 and finds that most of the material facts are hard to read. From the 4533 material facts studied, only ten percent were easy to read.

More important to the article than readability are the textual analysis methods that attempt to extract the meaning – or in our case – the sentiment from the message. One method is a word dictionary. Loughran and McDonald (2016) list some advantages of using word dictionaries to measure sentiment. First, by using a dictionary, researcher subjectivity is avoided. Second, it is easier to scale the method to larger samples. Third, due to the public nature of word dictionaries, it is easier to replicate other research. Tetlock (2007), using a Harvard dictionary, shows that the pessimism of a Wall Street Journal daily column is linked to lower returns and higher volatility in the following days, and this downward pressure is not caused by new information on the company valuations. Garcia (2013) measures the sentiment of two New York Times financial columns from 1905 to 2005, and finds that, controlling for well-known time series patterns, the news sentiment predicts daily returns, especially during recessions. Other articles measuring sentiment to predict the financial market include Bollen et al. (2011), who test if it's possible to forecast the stock market by using messages on Twitter - running a text mining tool to distinguish tweets between different emotions, and Gilbert and Karahalios (2010), which extract the sentiment of *LiveJournal's* microblogs to forecast the stock market.

Likewise, some articles use other data to test the investors' perception of new information, like Bordino et al. (2012), which take hold of data related to ticker queries on Yahoo! and correlate it to the volume of transactions. This work is very similar to Da et al. (2011).

But, for a material fact to have a determinant related to the stock performance, we need the management to purposely withhold information intending to publish it during the optimal interval. This idea of managers waiting to disclose information at the best timing is nothing new in the finance literature. The Strategic Timing of Disclosure theory says that the optimal disclosure strategy is determined by the costs and benefits of the disclosure. Based on their assessment, management should decide the nature, content, timing, way of publication, etc. This theory suggests that managers should decide the disclosure date of earnings report based on their evaluation of potential benefits, or costs, of early(late) publication (Sengupta, 2004). According to the literature, one of the principal determinants to the publication is a simple market intuition, managers publish good news earlier and bad news later (Givoly and Palmon, 1982; Patell and Wolfson, 1982; Chambers and Penman, 1984; Bagnoli et al., 2002). An important thing to notice is that the Strategic Timing of Disclosure shows that the timing of a publication usually is a management decision (Bagnoli et al., 2002). A survey by Graham et al. (2005) with 400 managers (composed by CEOs, CFOs and other management roles) shows they time voluntary disclosures to improve stock performance and reduce asymmetric information.

To verify the feedback effect, we propose using a Vector Auto Regression. Some papers that use this approach for financial data are available in the literature. The feedback effect comes from the feedback traders, investors whose demand is based on the history of past returns rather than the expectation of future fundamentals (Cutler et al., 1990). Grossmann et al. (2014) investigate the link between exchange rate volatility and macroeconomic and financial variables using Panel VAR models. They find that the volatility of developing countries has a bigger feedback effect than the volatility in developed countries. Perlin et al. (2017) investigates if queries of finance-related words impact the returns, volatility and volume of the stock market of 4 anglophone countries. Bozza et al. (2019) tests for a feedback effect between the return and volume of different cryptocurrencies using high frequency trading data. They find that is a negative feedback effect, and they suggest this effect results from the use of bots.

Table 1 shows a summary of part of the body of literature. As we can observe, the vast majority of the empirical results show that new information given to the investor impacts the stock prices. However, there are results against the effects of the new information in material facts relating to stock prices. The table also shows how different emotions and new information the investor perceives get different results in different studies.

Compared to the body of literature present in table 1, this article is the only one to use such diverse material facts, or even more generally, events, to analyze the effect of new information in stock prices, especially in Brazil. There is only one article using High Frequency Trading Data in the Brazilian Market. Not a single article focuses on the determinants of material facts. Next, we'll discuss the Data used in the present study.

3. Data

Our data comprises two databases: The first is a collection of material facts (*Fatos Relevantes*) extracted directly from CVM's site, and the second is a High Frequency Trading data originally from *B3's* FTP site (Perlin and Ramos, 2016).[2] We also use daily financial data obtained through the Economatica software, a reliable source of information, constantly used for research in the local literature.

CVM provides a web interface to any material fact of any company on B3 that has been published since 2010. All material facts are extracted using this research's proprietary web scrapping process. The authors developed a custom robot to sequentially gather all available information from CVM's website. It works as follows: the automated browser session searches for ev-

[2]Be aware that, as of 2020-06-29, the ftp site has been shut down by B3 during the unification of their BM&FBovespa and Cetip websites as explained in this note

Table 1
Synopsis of the literature

Authors	Study Object	Main results
Patell and Wolfson (1984)	Earnings and dividends releases. Dow Jones returns.	Holding period return in the first thirty minutes from releases was higher than any other thirty-minute returns verified in the same announcement day. The highest numbers of extreme price changes were identified in the first four minutes following earnings announcements.
Barclay and Litzenberger (1988)	Announcements of new equity or debt issues. Return by firms listed on the New York or American Stock Exchange.	Finds atypically large number of transactions recorded during the first fifteen minutes following the announcements of new equity issues. Stock prices fell on average 1.34% during this time interval. They calculated an average cumulative return of -2.44% between one hour before the disclosure of new equity issues and two hours after.
Busse and Green (2002)	US Markets. Morning Call or Midday Call on CNBC (TV shows).	Finds that cumulative average returns started to grow two minutes before the disclosure of positive news by the Midday Call but stabilized at 0.5% three minutes after the first mention of the stock in this television report. Negative on-air stock reports provoked a longer reaction in the cumulative average return: statistically significant effects lasted thirteen minutes from the moment bad news was given.
Silva and da Silva Felipe (2010)	Brazilian stocks. Pessimistic and optimistic Material Facts.	Finds that the Brazilian stock market is conservative. Investors don't react to positive material facts,but react to negative material facts.
Marques et al. (2011)	Bovespa's Novo Mercado Stocks Material Facts	Only 15 of 78 material facts had a significant effect on the stock's prices. Results goes against the semi-strong EMH.
Bollen et al. (2011)	Dow Jones Industrial Average. Twitter Messages.	Tests if it's possible to predict the stock-market with Twitter messages. They were read by a mood-tracking tool distinguishing messages according to different moods. This data was used to measure the relationship of the different moods of the general public with the closing values of the Dow Jones Industrial index. It was found that some of the moods intensities can help explain variations in the market index.
Drienko and Sault (2013)	Firms' announcements in response to requirements from the Australia Securities Exchange. Australian stocks.	Used intraday data to examine abnormal returns' behavior and found evidence that the market takes up to sixty minutes to reflect the information released.
de Carvalho et al. (2016)	Material facts. HFT returns from B3.	Findings indicate that material facts reported by firms reveal unexpected information to investors. They also suggest that stock prices take up to fifty minutes to incorporate the new information.
Gilbert and Karahalios (2010)	S&P 500 and Monte Carlo Simulations. Weblogs.	Shows that by estimating emotions from weblogs, specifically, the website LiveJournal, it is possible to predict future stock market prices. The data showed that increases in expressions of anxiety predicted downward pressure on the SP& 500 index. These results were confirmed via Monte Carlo Simulation.
Bordino et al. (2012)	Trading volume in US Stocks. Trading Queries.	Found that an increase in the search frequency of stocks tickers can explain changes in the traded volume of the same stock.
Perlin et al. (2017)	American, British, Canadian and Australian stocks. Google search queries of finance-related words.	Finds that increases in search queries including the word stock predict increased volatility and decreased index returns over the next week. Created a market timing based strategy based on the search frequency with positive results which are particularly stronger during the global crisis of 2009
Damascena et al. (2017)	Two years of a single stock prices.	Material facts initially affected stock prices, but soon after they returned to their initial levels. These results corroborate to the semi-strong EMH.

ery material fact of each company. Then, from the most recent to the oldest, the program will scrape all available information, including the 6-digit protocol number that index the document. Based on this number, it is possible to reconstruct the internet link to the original document. Therefore, instead of downloading every material fact, we download only the ones in the categories we are interested in.

The high frequency financial data contains price and volume information and is compounded in 5-minute intervals. The data is from January 1st, 2010 to May 28th, 2018, with short gaps in the financial data between 2010 and 2014. Companies with an average number of daily trades lower than 5000 are removed from the sample. When both the common and preferred stocks of a company have been filtered, we selected the stocks with the highest mean traded volume over the sample. This is necessary because a low volume of data may bias the results and compromise the study due to lack of price volatility. The data consists of information about each trade of the day, with these items:

- Session date: The day, month and year of the data;
- Instrument identifier: The local ticker symbol representing the negotiated asset;
- Trade price: The price of the trade;
- Traded quantity: The number of traded shares;
- Trade time: The hours in which the contracts were negotiated with the format HH:MM:SS.NNNNNN;
- Trade indicator: A dummy in which 1 represents a trade and 2 stands for a cancelled trade;
- Aggressor Buy Order Indicator: A dummy in which 0 represents that an order was not executed, 1 when it was an aggressive order and 2 when it was passive order;

After filtering for the stocks, we have used our custom web scrapping algorithm to obtain all the required information about the material facts of the 33 chosen companies. This information includes:

- Category: The category of news, which can be any one of 67 categories. For the paper, we only keep the ones classified as “material fact” and “announcements to the market”;
- Reference date: The date that the document is referring to, such as a year or the schedule of a meeting;
- Delivery date: The date on which the material fact was published, with the format dd/mm/yyyy HH:MM:SS;

- Version;
- Protocol: The protocol is a sequence of numbers that can be used to access the documents' links;
- Company ID: Which is the company ID on CVM used to filter the material facts listed on the website;
- Company Name: The firm's official and complete name;
- Ticker: The company's ticker is the unique stock symbol assigned for the firm's security traded at B3;

Even though normative rules state that material facts should not be sent during trade time, there are no legal impediments or fines for disclosures amid trading time, so companies still publish material facts during it. This way, the resulting information is then filtered so we only keep material facts published during the working hours of B3 (Delivery date later than 10:00 AM and sooner than 18:00 PM) and that are categorized as *Fato Relevante* or *Comunicado ao Mercado* Using the protocols from the filtered database, we access the page containing the PDF file and download it, which results in a database of 3514 material facts for 33 companies in 2007 unique dates.[3].

Next, this database goes through a process of text mining using the R packages tm (Feinerer et al., 2008), tidytext (Silge and Robinson, 2016) and Oplexicon (Souza and Vieira, 2012) a Portuguese sentiment lexicon special for Brazil. This process creates a sentiment analysis for a large part of the material facts data. [4]

Sentiment analysis can be described as a use of text mining and natural language processing (NLP) to identify and extract the subjective content by analyzing user's opinion, evaluation, sentiments, attitudes and emotions (Bhardwaj et al., 2015; Feldman, 2013; Medhat et al., 2014). Our process of sentiment analysis is the traditional bag of words approach. First, we extract the raw text from each material fact. This text is then prepared for sentiment analysis, formatting the entire text to lower case, removing punctuation and numbers and trimming white spaces. Next, we remove the stop words, also known as the language's most common words. Finally, the text is prepared to be put through the process of sentiment analysis. The Oplexicon has a rating of -1, 0 or 1 for almost every word in the Portuguese language, This rating is given to every word of every material fact. After the evaluation of each word is done, we calculate the sentiment of a material fact by dividing the sum of its rating by the total number of words. The resulting database has 2281 material facts with sentiment analysis.

[3] Material fact and Announcement to the market in Portuguese.

[4] Oplexicon was created by PUCRS' Natural language Processing Team and we thank the authors for providing the code, which greatly helped our research.

Figure 1
Different summaries of data

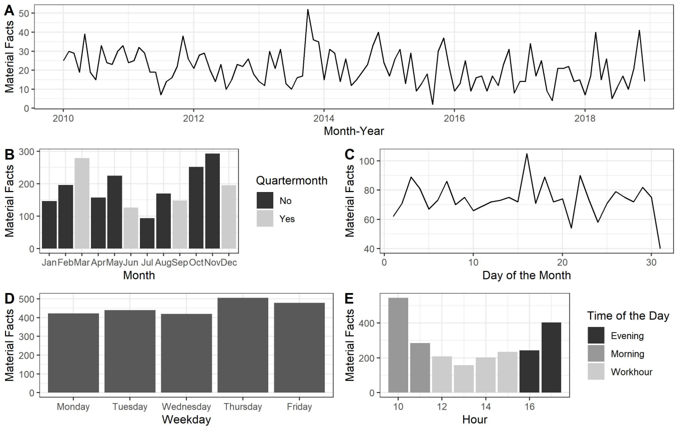

Figure 1 shows that the publishing of Material Facts, at least by the 33 companies studied in this article, does not appear to have a seasonal effect. But, as we can see in Table 2 and graph B from figure 1, the companies publish fewer material facts in the middle of the year during working hours. We should also highlight the higher number of material facts during November. The database doesn't seem to have a downwards or upwards yearly or even a quarterly movement, but the last three years of the Database have a lower number of Material Facts released, as shown in Table 2. Graph B also shows that it doesn't appear to have any significant spike or fall in material facts publications during the common fiscal quarters. Graph C shows that most material quarters are published during the middle of the month, and fewer Material facts are published during their end. Graph D shows that most material facts were published during Thursdays and Wednesdays had fewer material facts published. The most interesting fact in figure 1 is present in graph E, where we can see that the distribution of the hours in which a material fact is published is concave, being heavy tailed, with most of the Material Facts published during the working hours being published during the beginning of the shifts.

In Table 3, the standard deviation in the sentiment is lower for negative material facts when compared to positive material facts. It's also important to notice there are more negative material facts than positive material facts, although most of the material facts are not really negative, since the material facts with a sentiment rating of less than one are in the first quartile of the data.

Table 2
Material facts per month

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2010	25	30	29	19	39	19	15	33	24	23	30	33
2011	24	25	32	29	19	19	7	14	16	22	38	26
2012	21	28	29	20	14	23	10	15	23	22	26	18
2013	14	12	30	21	31	13	10	16	17	54	36	35
2014	15	31	29	14	26	12	15	19	23	33	40	24
2015	17	26	31	13	29	9	13	18	2	30	37	22
2016	9	13	25	9	16	17	9	17	12	25	31	8
2017	14	14	34	17	25	9	4	21	21	22	14	15
2018	7	17	40	15	26	5	11	17	10	21	41	14

Table 3
Descriptive stats of material facts' sentiment

Statistic	Total	Positive	Negative
N	2,281	1,026	1,255
Mean	0.153	0.294	0.037
St. Dev.	0.173	0.136	0.097
Min	−0.562	0.153	−0.562
Pctl(25)	0.053	0.196	0.000
Pctl(75)	0.234	0.350	0.111
Max	1.000	1.000	0.153

Table 4
Descriptive statistics of high frequency trading data

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Number of Trades	3,524,132	152.100	167.022	3	52	193	10,762
Period Return	3,524,132	−0.00001	0.003	−0.072	−0.001	0.001	0.069
Period Volatility	3,524,132	0.0005	0.0005	0.000	0.0002	0.001	0.045
Traded Quantity	3,524,132	85,511.490	237,703.100	300	12,900	80,400	57,596,600
Traded Volume	3,524,132	1,579,135.000	4,489,040.000	294	244,158	1,448,803.0	1,321,683,289
HF Ibovespa	3,524,079	−0.00001	0.001	−0.019	−0.001	0.001	0.021

The effect on returns of publishing a new material fact is measured using the nominal returns of the B3's assets resulting from the former liquidity test. We consider the trading volume for the same time frame used in the return analysis for liquidity. The volatility is also calculated from the HFT data. Table 4 shows a summary of the HFT financial data.

Table 4 shows that the High Frequency Trading data is abundant, with almost 3.5 million observations. As we can note, the Number of Trades has a high number which guarantees a certain degree of liquidity but, as the standard deviation and max amount indicate, the distribution is heavily tailed. In

Traded Quantity, we have the total number of shares negotiated during the 5-minute period with an average of 1 million shares traded during this interval. The Traded Volume is the Traded Quantity multiplied by the share price. Each one of these variables refers to each individual firm. The next variable HFT Ibovespa, on the other hand, is the Ibovespa Portfolio constructed in High Frequency Trading. This variable is used in the regressions in the next section.

4. Method

In this paper, we test two hypotheses. First, we'll discuss the main hypothesis of the article: the determinants of material facts publishing and sentiment. Next, we'll show the method used to test for a feedback system between the returns, volume and volatility and the publishing and sentiment of Material Facts. Before debating the hypotheses, the financial data is all normalized, so for each financial variable, we subtract the average of the sample and then divide by the standard deviation. This process is done to facilitate the computation of the model.

Hypothesis 1. Material facts have determinants: There is a preferred month, weekday and hour to publish Material Facts. More specific, Managers prefer to publish positive Material Facts near the end of fiscal quarters and avoid publishing negative Material Facts. Companies publish new positive material facts after negative material facts and negative material facts after positive material facts. The stock returns and volume also affect material facts publishing.

Since there is no previous literature about Material Facts' determinants, the null hypothesis is that material facts do not have determinants and the alternative hypothesis is that firms prefer to publish material facts in specific months, weekdays or hours. Companies publish new positive material facts after negative material facts and negative material facts after positive material facts, and the stock returns, volume and volatility also affect material facts publishing. As discussed before, it is expected by law that companies immediately report any material fact that occurred or related to its business that may significantly influence in the market investors' decision to sell or buy any equity issued by the firm. By having determinants, firms may be waiting for the best moment to report bad news, or even good news, causing asymmetrical information between their managers and stakeholders. One example is waiting for the end of the day to publish negative material facts or until Friday to publish a negative Material Fact or for Monday to publish a positive one. This is especially important since it would cause overnight volatility, which is notably bad for the foreign investor. This would make it harder to raise foreign capital. On the statistical side, we use a probit model to test hypothesis 1 as follows:

$$\Pr(M_{i,t} = 1|X) = \Phi(X^T \beta), \tag{1}$$

where $M_{i,t}$ has a value of 1 if there is a material fact published by firm i in interval t, otherwise 0. Parameter X^T is a vector of regressors, containing the return, volatility and volume of firm i in interval t, the Ibovespa returns in the same interval, dummies for the different periods of the day – more specifically the Morning and Evening (see graph E in figure 1), a dummy for Fridays and for Mondays, and a dummy for the months in which the fiscal quarters end which is called Quartermonth (see Graph B in Figure 1). The regressors vector also contains the cumulative returns, volatility volume and Ibovespa returns in the last 30 minutes for the High Frequency Trading data and the last 5 days for the daily data.

Alternatively, we also test for determinants of Material Facts sentiment, where $M_{i,t}$ has a value of 1 if there is a positive material fact published by firm i in interval t and 0 if there is a negative material fact. It's important to notice that we're only testing during intervals in which B3 is open, since companies are expected to publish material facts before and after the trading time. In the second hypothesis, we test for a feedback effect between material facts publishing and sentiment and stocks returns, volatility and volume.

Hypothesis 2. There is a feedback system between the stocks and the material facts. Stocks returns, volume and volatility affect material facts publishing and sentiment and vice versa.

The null hypothesis is that the material facts publishing does not affect the firm's stocks price, and the stock does not affect the material facts publishing. The alternative hypothesis is that material facts publishing and sentiment affect the firm's stock returns, volume and volatility and the firm's stock returns, volume and volatility affect the material facts publishing and sentiment. To test this, we use a structural vector auto regression (VAR) similar to Härdle et al. (1998) and Perlin et al. (2017), which provides insights regarding the endogenous relationship between the material facts publishing and sentiment and the dependent variables. This model tests the effect of a publishing material fact and the inverse — that is, the effect that the financial markets can have in material facts publishing. For the model we'll use three dependent variables - volatility, return and traded volume:

$$y_{i,t} = \alpha_1 + \sum_{p=1}^{BICLag} \beta_p y_{i,t-p} + \sum_{p=1}^{BICLag} \lambda_p M_{i,t-p} + \varepsilon_{1t}, \tag{2}$$

$$M_{i,t} = \alpha_2 + \sum_{p=1}^{BICLag} \gamma_p y_{i,t-p} + \sum_{p=1}^{BICLag} \phi_p M_{i,t-p} + \varepsilon_{2t}, \tag{3}$$

In the system of equations (2) and (3), variable $y_{i,t}$ is a placeholder for $\Delta Volat_{i,t}$, $Ret_{i,t}$ and $\Delta Vol_{i,t}$. To determine the lag of the system we use the Bayesian information Criterion (BIC). We selected the BIC method for optimal lag selection to guarantee that the model selection is optimal maintaining a low number of variables, since the BIC method places a heavier penalty on models with many variables, and usually selects smaller models over alternative methods, as, for example, the AIC.

5. Results

Table 5 shows the results for the regression using a publishing dummy with daily data, equation (1) and hypothesis 1. As we can see from Table 5, the returns seem to affect publishing material facts, at least on a lower level, as its coefficient is weakly statistically significant, but not especially for positive or negative material facts. This could mean that when stock performance is good, the management does not find it necessary to publish new material facts to correct informational asymmetries. On the other hand, the volatility and the cumulative (weekly) volatility affect the material facts' publishing, especially positive material facts, which could mean that managers publish good news to turn their stocks less risky. It is also interesting to note that the daily volume positive relates to material facts publishing. The Quartermonth determinant is not statistical significant, invalidating the theory that companies publish material facts closer to the end of the fiscal quarters. This is specifically noted by the regression for publishing material facts with positive sentiment as if companies published more material facts close to the end of the fiscal quarters, they would rather publish mainly positive material facts, which does not happen. There is also not a correlation between the days before and after the weekend and the Material Facts publishing. Overall, the results in Table 5 show that the volatility, volume and returns, on a lower level, are determinants for material publishing using daily data.

Next, we'll test if this scenario repeats itself when using the High Frequency Trading data.

Table 6 shows the results for the probit regression using High Frequency Trading data. Contrary to the daily data, the returns do not have weak statistical significance for publishing material facts. Similar to the daily data, the accumulated volatility does affect material facts publishing for the complete and positive cases. Still, this time, the effect is the contrary, being negative, which

Table 5
Daily results for Material Facts publishing

	Dependent variable:		
	Material Fact publishing		
	Complete	Positive	Negative
	(1)	(2)	(3)
Returns	-0.023^{*}	-0.019	-0.025
	(0.012)	(0.013)	(0.023)
C. Returns	0.009	0.010	0.004
	(0.013)	(0.013)	(0.024)
Volume	0.129^{***}	0.118^{***}	0.118^{***}
	(0.015)	(0.016)	(0.026)
C. Volume	-0.145^{***}	-0.133^{***}	-0.133^{***}
	(0.017)	(0.018)	(0.032)
Volatility	0.022^{**}	0.021^{**}	0.012
	(0.009)	(0.010)	(0.015)
C. Volatility	0.030^{***}	0.026^{**}	0.034^{*}
	(0.011)	(0.012)	(0.020)
Quartermonth	0.018	0.016	0.022
	(0.021)	(0.022)	(0.041)
Dummy Friday	0.036	0.031	0.045
	(0.024)	(0.025)	(0.045)
Dummy Monday	0.033	0.025	0.053
	(0.024)	(0.025)	(0.045)
Ibovespa	0.602	0.524	0.582
	(0.796)	(0.847)	(1.492)
C. Ibovespa	-0.088	-0.047	-0.238
	(0.381)	(0.405)	(0.731)
Constant	-1.817^{***}	-1.895^{***}	-2.540^{***}
	(0.013)	(0.014)	(0.026)

This table shows the results from equation (1) using daily financial data. Column 1, "Complete", shows the results in which the $M_{i,t}$ dummy has a value of 1 for every publishing of a material fact. Column Positive shows the results in the $M_{i,t}$ dummy has a value of 1 for every publishing of a Material Fact with positive sentiment. Column Negative shows the results in the $M_{i,t}$ dummy has a value of 1 for every publishing of a Material Fact with negative sentiment. Every variable refers to the interval t, which consists of a day, except for the variables C. Returns, C. Volume, C. Volatility, C. Ibovespa which refer to the cumulative variable from the last $t-5$ days. *Note:* $^{*}: p<0.1$; $^{**}: p<0.05$; $^{***}: p<0.01$

Table 6
HFT results for Material Facts publishing

	Dependent variable:		
	Material Fact publishing		
	Complete	Postive	Negative
	(1)	(2)	(3)
Returns	−0.028	0.001	−0.033
	(0.021)	(0.027)	(0.022)
C. Returns	0.019	−0.012	0.028
	(0.020)	(0.027)	(0.022)
Volume	0.001	0.004	−0.045
	(0.015)	(0.008)	(0.031)
C. Volume	0.002	0.007	0.001
	(0.016)	(0.016)	(0.019)
Volatility	0.015	−0.011	0.021
	(0.015)	(0.020)	(0.015)
C. Volatility	-0.036^{**}	-0.046^{*}	−0.010
	(0.018)	(0.024)	(0.019)
Quartermonth	0.005	−0.003	0.005
	(0.035)	(0.046)	(0.041)
Dummy Friday	0.034	0.053	0.008
	(0.038)	(0.051)	(0.045)
Dummy Monday	−0.010	0.050	−0.058
	(0.040)	(0.051)	(0.048)
Dummy Morning	9.322	3.656^{***}	4.115^{***}
	(82.805)	(0.067)	(0.066)
Dummy Evening	9.284	3.937^{***}	3.781^{***}
	(94.730)	(0.074)	(0.073)
Ibovespa	18.593	13.667	14.221
	(17.153)	(22.946)	(19.540)
C. Ibovespa	−3.283	−10.948	4.494
	(6.987)	(9.146)	(7.795)
Constant	-3.696^{***}	-3.932^{***}	-3.819^{***}
	(0.022)	(0.031)	(0.026)

This table shows the results from equation (1) using High Frequency Trading Data (5-minutes interval). Column 1, "Complete", shows the results in which the $M_{i,t}$ dummy has a value of 1 for every publishing of a material fact. Column Positive shows the results in the $M_{i,t}$ dummy has a value of 1 for every publishing of a Material Fact with positive sentiment. Column Negative shows the results in the $M_{i,t}$ dummy has a value of 1 for every publishing of a Material Fact with negative sentiment. Every variable refers to the interval t, which consists of a 5-minute interval, except for the variables C. Returns, C. Volume, C. Volatility, C. Ibovespa which refer to the cumulative variable from the last $t-6$ 5-minute intervals, which consists of the last 30 minutes. *Note:* $^{*}: p<0.1$; $^{**}: p<0.05$; $^{***}: p<0.01$

could imply that on the High Frequency level the managers are not preoccupied with the stock risk, as they would not need to withhold information and publish positive news in the interval in which the volatility is lower. Meanwhile, volume does not affect material facts publishing, also going against the results using daily data. Again, the dummy Quartermonth does not have statistical relevance, which, is the opposite of the idea that companies release more material facts during the end of the fiscal quarters. Contrary to the idea that good news is published earlier and bad news is published later, in the high frequency scenario, both dummies Friday and Monday are not statistically significant, which means that firm management don't deliberately wait to publish positive material facts on Mondays and bad ones on Fridays.

The dummies evening and Morning are statistically significant in positive and negative cases. Upon expecting graph B in figure 1, this was expected. But it seems that Positive material facts have a higher probability of being published during the evening compared to the morning, while the contrary happens to negative material facts. This, again, goes against the idea that good news is disclosed early and bad news is disclosed later. In the following table, we'll show the regression results for material facts sentiment in daily data, in which we use a dummy that is zero if there is a publishing of a negative material fact and one when it is positive.

If the previous results indicate determinants for Material Facts publishing in the Brazilian Stock Market, Table 7 indicates there aren't determinants for material facts sentiment. This could be expected for the returns and volume variables, as Table 5 doesn't have special results for publishing of positive and negative material facts. In contrast with the volatility results, which could've been expected to be positive and statistically significant, but are mixed and not significant instead. Also, there is no significance for the weekday or material fact month variables.

Now, we'll see the results for the High Frequency Trading Data.

Table 8 shows there isn't a clear financial determinant to material facts sentiment using high frequency trading data. Returns, volatility and volume are not a determinant of the sentiment of the published Material Facts. The dummy Quartermonth once again does not have statistical significance. Again, Mondays and Friday are also statistically insignificant.

The most interesting result from Table 8 is that both the dummies evening and morning are not significant. This means that managers do not wait to publish positive material facts during the end of the day so the investor could absorb overnight or during the morning, to affect the entirety of trading time. These results follow the results from Table 6, where there isn't a special sentiment where the dummies are significant. We calculated the results for the

Table 7
Daily results for Material Facts sentiment

	Dependent variable:
	Material Fact sentiment
Returns	0.004
	(0.038)
C. Returns	0.013
	(0.041)
Volume	−0.020
	(0.047)
C. Volume	0.024
	(0.057)
Volatility	0.006
	(0.027)
C. Volatility	−0.029
	(0.039)
Quartermonth	−0.032
	(0.071)
Dummy Friday	−0.016
	(0.080)
Dummy Monday	−0.065
	(0.080)
Ibovespa	0.121
	(2.806)
C. Ibovespa	0.109
	(1.342)
Constant	0.996^{***}
	(0.045)

This table shows the results from equation (1) using High Frequency Trading Data in which the $M_{i,t}$ dummy has a value of 1 for every publishing of a material fact with positive sentiment and 0 for publishing of negative material facts. Every variable refers to the interval t, which consists of a 5-minute interval, except for the variables C. Returns, C. Volume, C. Volatility, C. Ibovespa which refer to the cumulative variable from the last $t-6$ 5-minute intervals, which consists of the last 30 minutes. *Note:* $^{*}: p<0.1$; $^{**}: p<0.05$; $^{***}: p<0.01$

Table 8
HFT results for Material Facts sentiment

	Dependent variable:
	Material Fact sentiment
Returns	0.013
	(0.049)
C. Returns	−0.060
	(0.049)
Volume	0.051
	(0.047)
C. Volume	0.032
	(0.035)
Volatility	−0.020
	(0.031)
C. Volatility	−0.037
	(0.043)
Quartermonth	0.004
	(0.094)
Dummy Friday	0.069
	(0.104)
Dummy Monday	0.162
	(0.108)
Dummy Morning	−0.045
	(0.103)
Dummy Evening	0.157
	(0.111)
Ibovespa	−6.034
	(46.905)
C.Ibovespa	−21.967
	(18.239)
Sigma	0.377***
	(0.086)
Constant	−0.222*
	(0.119)

This table shows the results from equation (1) using High Frequency Trading Data in which the $M_{i,t}$ dummy has a value of 1 for every publishing of a material fact with positive sentiment and 0 for publishing of negative material facts. Every variable refers to the interval t, which consists of a day, except for the variables C. Returns, C. Volume, C. Volatility, C. Ibovespa which refer to the cumulative variable from the last $t-5$ days. *Note:* $^{*}: p<0.1$; $^{**}: p<0.05$; $^{***}: p<0.01$

models before using stacked data for a simple probit model. The results are somewhat similar.

For hypothesis 2, we will use the following Vector Auto Regression test.[5] First, we compute the individual VAR to each individual stock in the sample, using the variables return, volatility and volume, which are normalized, as our "X". We use the dummies for the day or interval of a publication and for the polarity of the material fact's sentiment as our "Y", the same dummies used in probit regressions. The models are the same for each asset, except for the number of lags chosen by the BIC. Next, we do a Granger-causality test for each one of them. The null hypothesis of our Granger test is that a variable does not grange-cause the other and vice-versa. Next, we'll show a summary of the results which showing the sample percentage where the null hypothesis is not rejected, where the sum of the coefficients is positive, and the maximum max lag.

Table 9
VAR summary

Panel A: High Frequency Trading Results						
$X \longrightarrow Y$	P-value > .1	Sum > 0	$Y \longrightarrow X$	P-value > .1	Sum > 0	Max max Lag
Returns	78.78%	0%	Publishing	87.87%	54.54%	7
Returns	84.84%	0%	Sentiment	72.72%	51.51%	7
Volume	66.66%	100%	Publishing	84.84%	100%	20
Volume	75.75%	100%	Sentiment	75.75%	100%	20
Volatility	63.63%	100%	Publishing	57.57%	100%	20
Volatility	66.66%	100%	Sentiment	63.63%	100%	20
Panel B: Daily Results						
$X \longrightarrow Y$	P-value > .1	Sum > 0	$Y \longrightarrow X$	P-value > .1	Sum > 0	Max max Lag
Returns	72.72%	30.30%	Publishing	72.72%	30.30%	14
Returns	69.69%	36.36%	Sentiment	87.87%	48.48%	16
Volume	81.81%	100%	Publishing	75.75%	100%	20
Volume	90.90%	100%	Sentiment	78.78%	100%	20
Volatility	84.84%	100%	Publishing	81.81%	100%	20
Volatility	84.84%	100%	Sentiment	87.87%	100%	19

We compute the individual VAR to each individual stock in the sample, using the variables return, volatility and volume, which are normalized, as our "X" and dummies for the day or interval of a publication and for the polarity of the material fact's sentiment as our "Y", the same dependent variables used in Probit regressions. The models are the same for each asset, except for the number of lags chosen by the BIC. We do a Granger-causality test for each one of them. The null hypothesis of our Granger test is that a variable does not grange-cause the other and vice-versa. Columns 2 and 5 show the percentage of stocks where the P-value of the Granger test is bigger than 0.1. Columns 3 and 6 show the percentage of stocks where the sum of the coefficients is bigger than 0. Last column shows the largest max lag chosen by the BIC.

Panel A of Table 9 shows that most of the stocks do not have a Granger-causality between both variables. Also, most of the coefficients don't have their sum higher than 0, if you exclude the volume and volatility regressions.

[5] Originally, the intention of the authors was to use a Panel Vector Auto Regression, but the computer was unable to complete the model due to problems related to the package.

The X variable with the higher number of statistically significant Granger-causality is volatility. Meanwhile, for the Y variables, publishing has a higher number. Next, we have the results for the Vector Auto Regressions using daily data.

Panel B of Table 9 shows that even fewer stocks have a Granger-causality between both variables using daily data. Again, excluding the volume and volatility regressions, most stocks do not have the sum of their coefficients higher than 0. For daily data, the x variable with the higher number of statistically significant Granger-causality is returns, as opposed to the High Frequency and for the Y variable is Publishing, again. Considering this, it seems troublesome to establish the existence of a feedback system between stocks and material facts.

Contrary to part of the literature, which finds compelling evidence that material facts released during trading hours generate abnormal returns, our results indicate the contrary, as they show that most stocks studied do not show a Granger-causality. This can be explained first by the difference between the utilized methodology, as Granger-causalities have the intention of surpassing simple correlation between variables (Granger, 1969). Another possible reason is the bulk of material facts used in both studies, as we have over two thousand Material Facts, a huge difference compared to other articles in the literature. Another reason for this difference could be related to our bulk of articles, as there is a trade-off between having a lot of Material Facts in the sample, which could avoid biases and outliers, and having just a few but meaningful material facts, as they would have a real impact on stock performance. Also, there is the question if a material fact with a sentiment rating of -0.01 would really be that negative to have a meaningful impact on stocks returns, in which we could argue that the possible researcher bias is a necessary evil to filter which material facts are in fact positive and in fact negative.

6. Conclusion

This article analyzed the effects of material facts publishing and sentiment in stock returns, volume and volatility and vice versa. Following part of the literature showing that material facts incorporate new information to the stock market, which usually takes from up to one hour until the market completely reacts, this article finds weak evidence of returns affecting material facts publishing. However, we discovered that volume and especially are stronger determinants for material fact publishing.

In the study based on daily data, we find that returns (on a lower level), volume and volatility are determinants for material facts publishing. The results show that material facts sentiment does not have a clear determinant regarding material fact sentiment. For High Frequency Trading Data, the only financial determinant for material facts publishing is volatility, while the sentiment is the same as the daily data, without significant determinants. In relation to the hours of the day, as we find there is not a specific time of day for material facts publishing, unlike expectations. This could mean that companies are not withholding or timing information to the stock market. Although there is a higher probability of positive material facts being published during the evening than in the morning, the opposite is true for negative material facts. This article is the only known work in Brazilian literature to test a large bulk of material facts and check for a feedback effect between material facts and the stock market statistics.

This article serves as an extension of the study of the effects of material facts in the Brazilian Stock Market using High Frequency Trading Data. This article may be expanded by having a different approach to the effects of material facts in the stock returns and by how long the new information given by the material facts takes to be incorporated into the stock price, as seen in the work of de Carvalho et al. (2016).

Surely, this study may be improved, since it could have simply used a Panel Vector Auto Regression for the feedback system tests as the database was too big, which prevented the use of the desired methods due to the lack of computational capacity. Last, another improvement for this investigation would be a bigger focus on the volatility, especially considering that volatility had an effect on material facts publishing in daily and high frequency trading data.

Acknowledgments

We are grateful to the Coordenação de Aperfeiçoamento de Pessoal de Nível Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for financing this work.

References

Bagnoli, M., Kross, W. and Watts, S. G. (2002). The information in management's expected earnings report date: A day late, a penny short, *Journal of Accounting Research* **40**(5): 1275–1296.

Barclay, M. J. and Litzenberger, R. H. (1988). Announcement effects of new equity issues and the use of intraday price data, *Journal of Financial Economics* **21**(1): 71–99.

Bhardwaj, A., Narayan, Y., Dutta, M. et al. (2015). Sentiment analysis for indian stock market prediction using sensex and nifty, *Procedia Computer Science* **70**: 85–91.

Bollen, J., Mao, H. and Zeng, X. (2011). Twitter mood predicts the stock market, *Journal of computational science* **2**(1): 1–8.

Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A. and Weber, I. (2012). Web search queries can predict stock market volumes, *PloS one* **7**(7): e40014.

Bozza, C. C., Klotzle, M. C., Pinto, A. C. F. and da Gama Silva, P. V. J. (2019). Efeito feedback trading em criptomoedas com dados de alta frequência, *Revista De Gestao, Financas E Contabilidade* **9**(1): 80–98.

Busse, J. A. and Green, T. C. (2002). Market efficiency in real time, *Journal of Financial Economics* **65**(3): 415–437.

Chambers, A. E. and Penman, S. H. (1984). Timeliness of reporting and the stock price reaction to earnings announcements, *Journal of accounting research* pp. 21–47.

Cutler, D. M., Poterba, J. M. and Summers, L. H. (1990). Speculative dynamics and the role of feedback traders.

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

Damascena, D. S., da Silva, S. E. P., dos Santos, J. O. and de Godoi, A. F. (2017). Análise do impacto de fatos relevantes no retorno das ações de empresa de educação superior, *Revista Brasileira de Administração Científica* **8**(3): 1–12.

de Carvalho, A., Suen, A. S., Gallo, F. et al. (2016). Market efficiency in brazil: Some evidence from high-frequency data, *Bcb working papers 431*, Banco Central do Brasil.

Drienko, J. and Sault, S. J. (2013). The intraday impact of company responses to exchange queries, *Journal of Banking & Finance* **37**(12): 4810–4819.

Fama, E. F. (1965). The behavior of stock-market prices, *Journal of Business* **38**(1): 34–105.

Fama, E. F., Fisher, L., Jensen, M. C. and Roll, R. (1969). The adjustment of stock prices to new information, *International Economic Review* **10**(1): 1–21.

Feinerer, I., Hornik, K. and Meyer, D. (2008). Text mining infrastructure in r, *Journal of Statistical Software* **25**(5): 1–54.
URL: https://www.jstatsoft.org/v25/i05/

Feldman, R. (2013). Techniques and applications for sentiment analysis, *Communications of the ACM* **56**(4): 82–89.

Garcia, D. (2013). Sentiment during recessions, *The Journal of Finance* **68**(3): 1267–1300.

Gilbert, E. and Karahalios, K. (2010). Widespread worry and the stock market, *Fourth International AAAI Conference on Weblogs and Social Media*.

Givoly, D. and Palmon, D. (1982). Timeliness of annual earnings announcements: Some empirical evidence, *Accounting review* pp. 486–508.

Graham, J. R., Harvey, C. R. and Rajgopal, S. (2005). The economic implications of corporate financial reporting, *Journal of accounting and economics* **40**(1-3): 3–73.

Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods, *Econometrica: journal of the Econometric Society* pp. 424–438.

Grossmann, A., Love, I. and Orlov, A. G. (2014). The dynamics of exchange rate volatility: A panel var approach, *Journal of International Financial Markets, Institutions and Money* **33**: 1–27.

Härdle, W., Tsybakov, A. and Yang, L. (1998). Nonparametric vector autoregression, *Journal of Statistical Planning and Inference* **68**(2): 221–245.

Lehavy, R., Li, F. and Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts, *The Accounting Review* **86**(3): 1087–1115.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence, *Journal of Accounting and economics* **45**(2-3): 221–247.

Loughran, T. and McDonald, B. (2016). Textual analysis in accounting and finance: A survey, *Journal of Accounting Research* **54**(4): 1187–1230.

Malkiel, B. G. and Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work, *The journal of Finance* **25**(2): 383–417.

Marques, V. A., Lara, C. O., Ayala, L. C. and Lamounier, W. M. (2011). O impacto dos fatos relevantes no mercado de capitais: um estudo dos fatos publicados pelas empresas participantes do novo mercado do segmento bovespa, *Sociedade, Contabilidade e Gestão* **5**(3).

Medhat, W., Hassan, A. and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey, *Ain Shams engineering journal* **5**(4): 1093–1113.

Patell, J. M. and Wolfson, M. A. (1982). Good news, bad news, and the intraday timing of corporate disclosures, *Accounting review* pp. 509–527.

Patell, J. M. and Wolfson, M. A. (1984). The intraday speed of adjustment of stock prices to earnings and dividend announcements, *Journal of financial economics* **13**(2): 223–252.

Perlin, M. and Ramos, H. (2016). Gethfdata: Ar package for downloading and aggregating high frequency trading data from bovespa, *Available at SSRN 2824058* .

Perlin, M. S., Caldeira, J. F., Santos, A. A. and Pontuschka, M. (2017). Can we predict the financial markets based on google's search queries?, *Journal of Forecasting* **36**(4): 454–467.

Sengupta, P. (2004). Disclosure timing: Determinants of quarterly earnings release dates, *Journal of Accounting and Public Policy* **23**(6): 457–482.

Silge, J. and Robinson, D. (2016). tidytext: Text mining and analysis using tidy data principles in r, *JOSS* **1**(3).
URL: http://dx.doi.org/10.21105/joss.00037

Silva, C. A. T. and da Silva Felipe, E. (2010). Avaliação da influência de textos narrativos de fatos relevantes no preço das ações de empresas brasileiras, *Revista Contabilidade e Controladoria* **2**(2).

Silva, C. A. T. and Fernandes, J. L. T. (2009). Legibilidade dos fatos relevantes no brasil, *Revista de Administração Contemporânea-RAC Electronica* **3**(1): 142–159.

Souza, M. and Vieira, R. (2012). Sentiment analysis on twitter data for portuguese language, *Proceedings of the 10th International Conference*

on Computational Processing of the Portuguese Language, PROPOR'12, Springer-Verlag, Berlin, Heidelberg, pp. 241–247.
URL: http://dx.doi.org/10.1007/978-3-642-28885-2_28

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.