XXIV SemeAd
Seminários em Administração

novembro de 2021
ISSN 2177-3866

What Drives the Release of Material Facts for Brazilian Stocks?

MARCOS HENRIQUE REICHERT
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

MARCELO SCHERER PERLIN
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (UFRGS)

Agradecimento à orgão de fomento:
This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001

What Drives the Release of Material Facts for Brazilian Stocks?

1 INTRODUCTION

According to Brazilian 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's investors decision to sell or buy any equity issued by the firm. Some examples of material facts include company's divisions, mergers, regulatory interventions, debt re-negotiations, discovery of new resources or technologies, forecast revisions and approvals or cancellations of investment projects. CVM (Comissão de Valores Mobiliários), the regulatory authority of security markets in Brazil, requires that publicly traded companies fully disclose material facts to reduce insider trading and informational asymmetries. In practice, publicly traded firms must, electronically and prior to disclosures in other communication channels, send to CVM structured files describing the material facts. After that, the information is immediately displayed on the websites of B3 and CVM, and available to market participants. The mandatory disclosure of information reduces insider trading arising from informational asymmetries in stock markets (CARVALHO et al., 2016)

Price volatility is the result of 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 but, 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 for the investor to receive and react to this new information. If a part of the investors receive this information sooner than another or even before it is publicly released, there is a source of informational asymmetry or insider trading.

As such, company's disclosures are recommended to be sent after the trading hours or before it to avoid excessive price volatility. But this is not always the case, as a large number of material facts are published during the trading hours. This raises questions whether material facts can be published as a mean to raise returns or the volatility or even as a response to their variation. In the literature, we find that Carvalho et al. (2016) results suggest that there is indeed some kind of anticipation before the release of 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 (CARVALHO et al., 2016; MARQUES et al., 2011; SILVA; FELIPE, 2010; DAMASCENA et al., 2017) and internationally (PATELL; WOLFSON, 1984; BARCLAY; LITZENBERGER, 1988). There is even studies on the material facts' readability (SILVA; FERNANDES, 2009), but the authors of this article didn't find any previous study on material facts' determinants. Thus, this article uses high frequency trading data, daily data and a set of material facts by publicly traded companies to test for material facts' determinants. We also test specifically for material facts with positive sentiment and negative sentiment.

The motivation for the paper is based on the well documented effect of material facts on stock prices and volume. It's interesting to note that managers can choose to 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 during the end of fiscal quarters, before investors meeting or wait even longer to not publish a material fact during these months. Managers could also time material facts publishing to Monday or Friday if they want the market to absorb or not the new information given by the material fact. This way, we'll also check if there is a feedback effect between the material facts, and stock returns, volume and volatility.

This article has the following objective: Identify the determinants for material facts' publishing. As an example, we want to test if there is any preference for month, weekday or hour to publish a new material fact. We also want to check for a feedback effect between stocks volume, returns and volatility and material facts' publishing and sentiment. This research is the first one to have a special focus on Material Facts determinants and to check for a feedback between material facts publishing and stock returns, volume and volatility. Contrary to the previous literature, this research uses text mining for material facts' sentiment classification, instead of a subjective one (CARVALHO et al., 2016). Last, despite the literature about the impact of new information using high frequency trading data , along with Carvalho et al. (2016), this article is one of the few to use high frequency trading data in the Brazilian Market, and the first one to use this high amount of data. As the mentioned literature demonstrates, the use of high-frequency data allows to examine with accuracy the effects of the studied events and reduce sample noise.

The remainder of this paper is organized as follows. Section two discusses the literature about market efficiency hypothesis and investor perception, highlighting studies with intraday data. Section three describes the database and section four the methodology. Section five presents the results. And section six will conclude the paper.

2 REVIEW

One of the landmarks in the financial literature is the Efficient Market Hypothesis (EMH) (MALKIEL; FAMA, 1970), which defines that investors are assumed to be rational in valuing financial securities by incorporating all the available information. It defines that irrational investors, if present, trade randomly and therefore their trades cancel each other out without affecting the prices and 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 fully predicted (FAMA, 1965; FAMA et al., 1969).

A very similar work to this article is the one by Carvalho et al. (2016). They make an event study to analyze a small sample of material facts, searching for abnormal returns and testing EMH's semi-strong market efficiency. They find that material facts reveal new information to investors. The results show that the investor can take up to 50 minutes to react to information in the Material Fact and that some investors use this time frame to profit. Also, they find a rise in negotiations before new material facts are released to the market.

Marques et al. (2011) test if material facts have an impact in Bovespa[1] Novo Mercado's stocks. They analyzed 78 material facts and found that only 15 of them had any effect in the stock prices. This goes against the semi strong EMH. Silva and Felipe (2010) analyze how the wording of the material facts affect the stock prices. That is, they categorized material facts as optimistic and pessimistic and test 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 had a decrease in their abnormal returns. Damascena et al. (2017) contrary to the previous study, has results that corroborate to EMH's semistrong form. The new information in Material Facts have an initial impact in the studied stock, but the stock price goes back to normal as this new information is absorbed by the market.

While previous studies analyzed daily data, the use of High Frequency Data is justified by the ability to verify 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. On the other hand, 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.

However, how does the investor perceives this new information? To answer this, we can use textual analysis. The first instances of use of the textual analysis 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 to extract the meaning - or in our case - the sentiment from the message. One of this methods is a word dictionary. Loughran and McDonald (2016) lists some of the advantages of using word dictionaries to measure sentiment. First, 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 researches. 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 company valuations. Garcia (2013) measures the sentiment of two financial columns of the New York Times from 1905 to 2005 and finds that controlling for well-known time series patterns the news sentiment predicts daily returns, especially during recessions.

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

3 DATA

The data used in this study is composed of 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[2] (PERLIN; RAMOS, 2016). We also use daily financial data obtained through the Economatica software, a reliable source of information, constantly used in the local literature.

CVM provides a web interface to any material fact of any company listed on B3 that has been published since 2010. For this research, all material facts are extracted using a proprietary web scrapping process. A custom robot was developed in order to sequentially gather all available information from CVM's website. It works as follows: the automated browser session searches for every 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. So, instead of downloading every material fact, we download only the ones in the categories we are interested in.

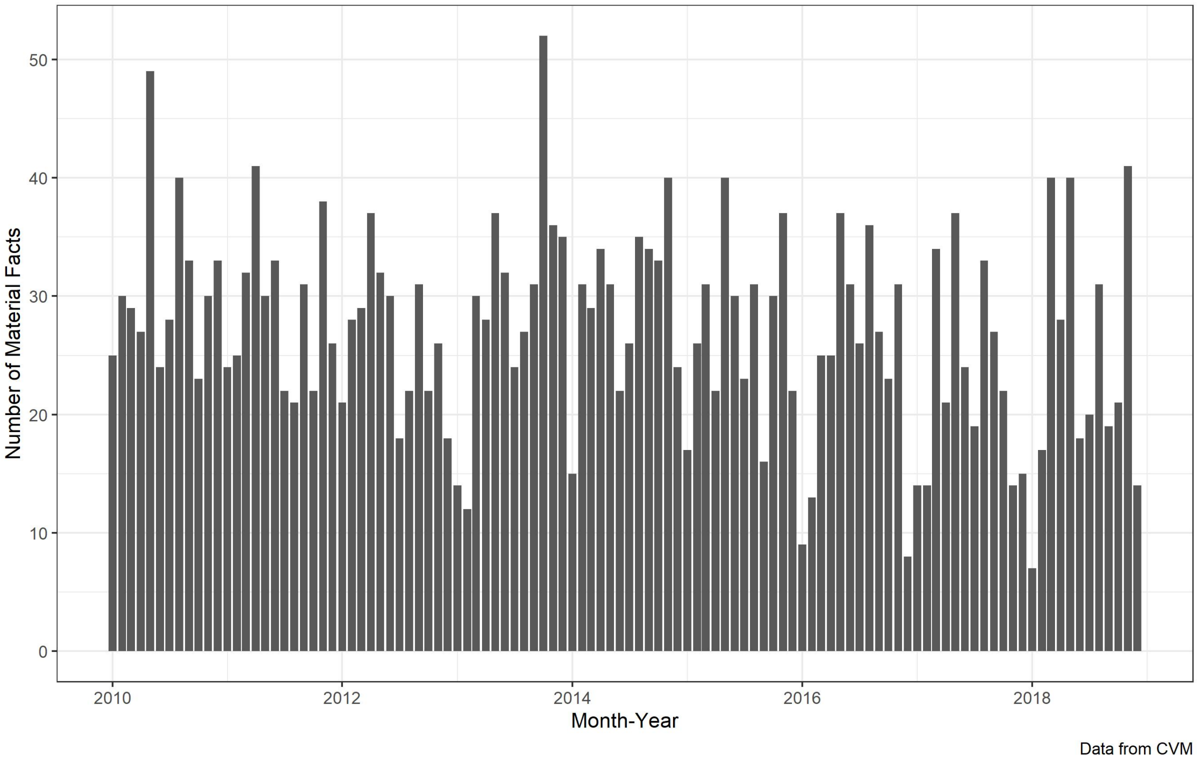
The high frequency financial data contains price and volume information and is compounded in 5 minute intervals. The data is from January 1st, 2010 to May 28th, 2018, with small gaps in the financial data between 2010 and 2014. Companies with an average number of daily trades lower than 5000 are removed from the sample. When both the common and preferred stocks of a company passed the filter, 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 can compromise the study due to extreme or lack of price volatility. The data consists of information about each trade of the day, including the following: Session date, Ticker, Trade Price and Traded Quantity. After filtering for the stocks, we use the custom web scrapping algorithm to obtain all the necessary information about the material facts of the 33 chosen companies. This information includes: Category, Delivery Date, Version, Protocol, Company ID, Company Name and Ticker.

Even though normative rules state that material facts should not be sent during trade time, to avoid excessive price volatility, there are no legal impediments or fines for disclosures during trading time, so companies still publish material facts during it. This way, the resulting information is then filtered so that we only keep material facts published during the working hours of B3 (Delivery date later than 10:00 AM and sooner than 17:00 PM) and that are categorized as *Fato Relevante* or *Comunicado ao Mercado*[3]. Using the protocols from the filtered database we access the page containing the PDF file and download it. This results in a database of 3514 material facts for 33 companies in 2007 unique dates.

Next, this database goes through a process of text mining using the R packages tm (FEINERER; HORNIK; MEYER, 2008), tidytext (SILGE; ROBINSON, 2016) and Oplexicon (SOUZA; VIEIRA, 2012)[4], a Portuguese sentiment lexicon special for Brazil. This process creates a sentiment analysis for a large part of the material facts data.

A sentiment analysis can be described as a use of of text mining and natural language processing (NLP) in order to identify and extract the subjective content by analyzing user's opinion, evaluation, sentiments, attitudes and emotions (BHARDWAJ et al., 2015; FELDMAN, 2013; MEDHAT; HASSAN; KORASHY, 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 to the process of sentiment analysis, formatting the entire text to lower case, removing punctuation and numbers and trimming white spaces. Next, we remove the stopwords, also known as the most common words of the language. 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 2928 material facts with sentiment analysis.

Figure 1 – Material facts per Month-Years



Source – Elaborated by the author

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. In table 1 we find that the standard deviation in the sentiment is lower for negative material facts compared to the positive material facts. It's also important to notice that we have 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.

In table 1 we find that the standard deviation in the sentiment is lower for negative material facts compared to the positive material facts. It's also important to notice that we have 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.

The effect on returns of the publishing of a new material fact is measured using the nominal returns of the B3's assets resulting from the former liquidity test. For liquidity we consider the trading volume for the same time frame used in the return analysis. The

Table 1 – Descriptive stats of Material Facts' sentiment

Statistic	Total	Positive	Negative
N	2,932	1,318	1,614
Mean	0.158	0.302	0.040
St. Dev.	0.174	0.133	0.098
Min	−0.562	0.158	−0.562
Pctl(25)	0.055	0.205	0.000
Pctl(75)	0.250	0.357	0.115
Max	1.000	1.000	0.158

Source – Elaborated by the author

Table 2 – 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

Source – Elaborated by the author

Volatility is also calculated from the HFT data. Table 2 shows a brief summary of the HFT financial data.

Table 2 shows that the High Frequency Trading data is plentiful, with almost 3.5 million observations. As we can see, 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 price of the share. It should be noted that every one of these variables refers to the individual firms. The next variable HF 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 different hypothesis. First we'll talk about 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 discussing the hypotheses, it is important to note that the financial data is all **normalized**. Which means that, 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 model's computation.

Hypothesis 1 (H1) *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 isn't any 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 month, weekday or hour to publish Material Facts. Companies publish new positive material facts after negative material facts and negative material facts after positive material facts and the stock returns and volume also affect material facts publishing. As discussed before, it is expected by law that companies *immediately* report any material fact occurred or related to its business that may have a significant influence in the market's 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 a bad news, or even good news, causing asymmetrical information between their managers and stakeholders. One example is to wait for the end of the day to publish negative material facts or to wait for 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 especially bad for the foreign investor. In turn, this would make it harder to raise foreign capital. In 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 intervals t and $t-1$, the Ibovespa returns in the same intervals, dummies for the different hours of the day, a dummy for Fridays and for Mondays, and a dummy for the months in which the fiscal quarters end which is called quartermonth.

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 it is expected for companies 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 (H2) *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 do not affect the firm's stocks price and the stock do 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, Tsybakov and Yang (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 not only the effect of a publishing material but also for the inverse — that is, the effect that the financial markets can have in material facts publishing. For the model we'll use three different dependent variables: volatility, return and traded volume:

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

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

In the system of equations 2 and 3 the 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 than alternative methods, as, for example, the AIC.

5 RESULTS

Table 3 shows the results for the regression using a publishing dummy with daily data, equation 1 and Hypothesis 1. As we can see from Table 3, the returns do not seem to affect the publishing of a material fact, as their coefficients are not statistically significant. On the other hand, the volatility and the lagged volatility do affect the material facts' publishing. It is also interesting to note that the daily volume has a positive relation with 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 especially noted by the regression for the publishing of material facts with positive sentiment. However, if companies publish more material facts close to the end of the fiscal quarters, they would like to publish mainly positive material facts, which doesn't 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 3 show that the volatility, even the lagged volatility, and the volume are determinants for material publishing using daily data.

Table 3 – Daily results for Material Facts publishing

	Dependent variable:		
	Material Fact publishing		
	Complete	Positive	Negative
	(1)	(2)	(3)
Returns	−0.016	−0.013	−0.019
	(0.011)	(0.011)	(0.020)
Lagged Returns	−0.002	−0.006	0.016
	(0.012)	(0.012)	(0.022)

Volume	0.065***	0.057***	0.071***
	(0.012)	(0.012)	(0.020)
Lagged Volume	0.009	0.005	0.018
	(0.009)	(0.010)	(0.015)
Volatility	0.034***	0.033***	0.021
	(0.008)	(0.008)	(0.013)
Lagged Volatility	−0.066***	−0.056***	−0.081***
	(0.013)	(0.014)	(0.026)
Quartermonth	0.015	0.012	0.021
	(0.021)	(0.022)	(0.040)
Dummy Friday	0.036	0.030	0.048
	(0.024)	(0.025)	(0.045)
Dummy Monday	0.025	0.017	0.050
	(0.024)	(0.025)	(0.045)
Ibovespa	0.383	0.417	0.023
	(0.720)	(0.766)	(1.351)
Lagged Ibovespa	−0.257	0.079	−1.511
	(0.725)	(0.771)	(1.393)
Constant	−1.810***	−1.887***	−2.537***
	(0.013)	(0.014)	(0.026)
Note:			*p<0.1; **p<0.05; ***p<0.01

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

Table 4 – HFT results for Material Facts publishing

	Dependent variable:		
	Material Fact publishing		
	Complete	Positive	Negative
	(1)	(2)	(3)
Returns	0.015	0.007	0.020*
	(0.009)	(0.014)	(0.012)
Lagged Returns	−0.006	−0.004	−0.007
	(0.009)	(0.014)	(0.012)
Volume	0.004	0.007	−0.010
	(0.005)	(0.004)	(0.015)

Lagged Volume	0.004	0.010	0.001
	(0.009)	(0.012)	(0.011)
Volatility	0.010	−0.021	0.025***
	(0.008)	(0.014)	(0.009)
Lagged Volatility	0.007**	0.008**	0.005
	(0.003)	(0.003)	(0.005)
Quartermonth	0.023	0.027	0.017
	(0.020)	(0.028)	(0.025)
Dummy Friday	0.039*	0.057*	0.022
	(0.022)	(0.031)	(0.028)
Dummy Monday	0.003	0.051*	−0.039
	(0.022)	(0.031)	(0.030)
Dummy Morning	−0.008	−0.013	−0.003
	(0.047)	(0.068)	(0.060)
Dummy Evening	−0.142***	−0.161***	−0.117***
	(0.019)	(0.028)	(0.024)
Ibovespa	−0.716	−6.387	3.303
	(8.369)	(12.102)	(10.646)
Lagged Ibovespa	11.283	7.781	13.040
	(8.371)	(12.116)	(10.640)
Constant	−3.311***	−3.545***	−3.463***
	(0.014)	(0.020)	(0.018)

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

Table 4 shows the results for the Probit regression using High Frequency Trading data. Contrary to the daily data, this time, the returns were weakly statistic significant for material facts' publishing, at least for the negative case. It is interesting to see that returns have a higher effect on negative material facts publishing, which could mean that managers would wait for intervals when the price is high to publish new material facts, but since the lagged returns aren't significant, the manager would wait for the exact five minute interval. This makes hard to support this theory.

Contrary to the daily data, Volatility does not affect material facts publishing for the complete and positive cases, but it is statistical significant for the negative case. It's curious that the lagged volume has the opposite effects, as they're statistically significant for the complete and positive cases. Since we have the previous five minute interval in the lagged cases, this gives a bigger support for the volume being a determinant for the complete and negative cases, even if the coefficient for the volume in interval t is stronger and more significant. Meanwhile, volume does not affects material facts publishing, also going against the results using daily data. The dummy Quartermonth, again, does not have statistical relevance, which, again, is contrary to the idea that companies release more material facts during the end of the fiscal quarters. It is noted that both the Morning and Evening dummies are statistical significant and negative. This also goes against the notion that companies would intentionally publish good material facts during the market hours.

With that in mind, the dummies Friday and Monday have a weakly significant positive coefficient for positive material facts, which could mean that managers indeed wait for Monday to publish a positive material fact, or maybe they want investor to absorb the information and wait to publish them on Friday. However, there is also a significant coefficient for the dummy Fridays in the complete case. The most interesting result in table 4 is the coefficient for the dummy evening. It was expected for it to be significant, but, contrary to the expectations, it is also negative. This means that is less probable for a material fact to be published from 16 to 18 p.m. The dummy for publishing during the morning is not significant. In the following table we'll show the results for the regression for material facts sentiment in High Frequency Trading 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. There isn't results for daily data for computational reasons.

Table 5 – HFT results for Material Facts sentiment

	Dependent variable:
	Material Fact sentiment
Returns	-0.058
	(0.044)
Lagged Returns	0.006
	(0.044)
Volume	0.037
	(0.042)
Lagged Volume	0.031
	(0.036)
Volatility	-0.048
	(0.033)
Lagged Volatility	0.026
	(0.032)
Quartermonth	-0.027
	(0.093)
Friday	0.133
	(0.103)

Monday	0.246^{**}
	(0.108)
Morning	-0.183
	(0.222)
Evening	-0.074
	(0.092)
Ibovespa	-31.711
	(43.761)
Lagged Ibovespa	-22.754
	(44.348)
Last Negative	0.970^{**}
	(0.487)
Last Positive	1.071^{**}
	(0.486)
Sigma	0.579^{***}
	(0.121)
Constant	-1.188^{**}
	(0.494)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 5 shows that there isn't a clear financial determinant to material facts sentiment using high frequency trading data. Returns, Volatility and volume are not a determinant to the sentiment of the published Material Facts. The dummy Quartermonth does not have statistical significance again, opposite to the dummy Monday, which is positive. This results corroborates with the results from last table, in which Positive Material Facts where published more frequently during Mondays and Fridays.

The most interesting result from Table 5 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 that the investor could absorb overnight or during the morning, to affect the entirety of trading time. This results follow the results from 4, where there isn't a special sentiment where the dummies are significant. We calculated the results for the models before, including the sentiment results for daily data, 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" 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 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 shows the percentage of the sample where the null hypothesis is not rejected, where the sum of the coefficients is positive, and the maximum max lag.

Table 6 – VAR summary

$X \longrightarrow Y$	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

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 were the P-value of the Granger test is bigger than 0.1. Columns 3 and 6 shows the percentage of stocks were the sum of the coefficients is bigger than 0. Last column shows the largest max lag chosen by the BIC.

Table 6 shows that most of the stocks do not have a granger-causality between both variables. Also, most of the coefficients doesn't have their sum higher than 0, if you exclude the volume and volatility regressions. The only exception is for the VAR for Sentiment by the returns, which is strange, considering the results from the Probit regression. The X variable with the higher number of statistical 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.

Table 7 – VAR summary Daily

$X \longrightarrow Y$	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

This table is identical to table 6, except it is showing the results for daily data.

Table 7 shows that using daily data, even fewer stocks have a granger-causality between both variables. Again, excluding the volume and volatility regressions, most of the stocks do not have the sum of their coefficients higher than 0. For daily data, the x variable with the higher number of statistical significant granger-causality is returns, contrary to the High Frequency and for the Y variable is Publishing, again. With this in mind, is hard to establish the existence of a feedback system between stocks and material facts.

6 CONCLUSION

This article analyzed the effects of material facts publishing and sentiment in stock returns, volume and volatility and the inverse. Even if part of the literature shows that material facts incorporate new information to the stock market and usually this takes from up to one hour until the market fully reacts, this article didn't find strong evidence of material facts affecting the returns of the stock market. But we found that volatility and volume are determinants for material fact publishing.

In the study with high frequency trading data, we find that returns are a determinant for negative material facts publishing, and that volatility is also a determinant. Another important point to notice is that, regarding material facts sentiment, we find that positive material facts have a higher chance of being published during Mondays and Fridays and Material Facts are more optimistic on Mondays. In relation to the the hour of the day, we find that companies are less prone to publish material facts during the last hours of the evening, contrary to the expected. This could mean that companies are not withholding or timing information to the stock market. 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 a continuation for the study of material facts effects in the Brazilian Stock Market using High Frequency Trading Data. This article can be expanded by having a different approach to the effects of material facts in the stock returns and how long the new information given by the material facts takes to be incorporated into the stock price, more similar to the work from Carvalho et al. (2016).

There is room for improvement in this article, since it could have simply used a Panel Vector Auto Regression for the feedback system tests but the dataset was too big, which prevented the methods use due to the lack of computational capacity. Lastly, other improvement for this dissertation would be a bigger focus in the volume, especially considering that volume had an effect in material facts publishing and sentiment.

NOTES

[1]B3 was previously called Bovespa.

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

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

[4]Oplexicon was created by PUCRS' Grupo de processamento de Linguagem natural and we thank the authors for providing the code, which greatly helped our research.

[5]Originally, the intent 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.

BIBLIOGRAPHY

BARCLAY, M. J.; LITZENBERGER, R. H. Announcement effects of new equity issues and the use of intraday price data. *Journal of Financial Economics*, Elsevier, v. 21, n. 1, p. 71–99, 1988.

BHARDWAJ, A. et al. Sentiment analysis for indian stock market prediction using sensex and nifty. *Procedia Computer Science*, Elsevier, v. 70, p. 85–91, 2015.

BORDINO, I. et al. Web search queries can predict stock market volumes. *PloS one*, Public Library of Science, v. 7, n. 7, p. e40014, 2012.

BUSSE, J. A.; GREEN, T. C. Market efficiency in real time. *Journal of Financial Economics*, Elsevier, v. 65, n. 3, p. 415–437, 2002.

CARVALHO, A. de et al. *Market efficiency in Brazil: Some evidence from high-frequency data*. [S.l.], 2016.

DA, Z.; ENGELBERG, J.; GAO, P. In search of attention. *The Journal of Finance*, Wiley Online Library, v. 66, n. 5, p. 1461–1499, 2011.

DAMASCENA, D. S. et al. Análise do impacto de fatos relevantes no retorno das ações de empresa de educação superior. *Revista Brasileira de Administração Científica*, v. 8, n. 3, p. 1–12, 2017.

FAMA, E. F. The behavior of stock-market prices. *The journal of Business*, JSTOR, v. 38, n. 1, p. 34–105, 1965.

FAMA, E. F. et al. The adjustment of stock prices to new information. *International economic review*, JSTOR, v. 10, n. 1, p. 1–21, 1969.

FEINERER, I.; HORNIK, K.; MEYER, D. Text mining infrastructure in r. *Journal of Statistical Software*, v. 25, n. 5, p. 1–54, March 2008. Disponível em: <https://www.jstatsoft.org/v25/i05/>.

FELDMAN, R. Techniques and applications for sentiment analysis. *Communications of the ACM*, ACM New York, NY, USA, v. 56, n. 4, p. 82–89, 2013.

GARCIA, D. Sentiment during recessions. *The Journal of Finance*, Wiley Online Library, v. 68, n. 3, p. 1267–1300, 2013.

HÄRDLE, W.; TSYBAKOV, A.; YANG, L. Nonparametric vector autoregression. *Journal of Statistical Planning and Inference*, Elsevier, v. 68, n. 2, p. 221–245, 1998.

LOUGHRAN, T.; MCDONALD, B. Textual analysis in accounting and finance: A survey. *Journal of Accounting Research*, Wiley Online Library, v. 54, n. 4, p. 1187–1230, 2016.

MALKIEL, B. G.; FAMA, E. F. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, Wiley Online Library, v. 25, n. 2, p. 383–417, 1970.

MARQUES, V. A. et al. 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*, v. 5, n. 3, 2011.

MEDHAT, W.; HASSAN, A.; KORASHY, H. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, Elsevier, v. 5, n. 4, p. 1093–1113, 2014.

PATELL, J. M.; WOLFSON, M. A. The intraday speed of adjustment of stock prices to earnings and dividend announcements. *Journal of financial economics*, Elsevier, v. 13, n. 2, p. 223–252, 1984.

PERLIN, M.; RAMOS, H. Gethfdata: Ar package for downloading and aggregating high frequency trading data from bovespa. *Available at SSRN 2824058*, 2016.

PERLIN, M. S. et al. Can we predict the financial markets based on google's search queries? *Journal of Forecasting*, Wiley Online Library, v. 36, n. 4, p. 454–467, 2017.

SILGE, J.; ROBINSON, D. tidytext: Text mining and analysis using tidy data principles in r. *JOSS*, The Open Journal, v. 1, n. 3, 2016. Disponível em: <http://dx.doi.org/10.21105/joss.00037>.

SILVA, C. A. T.; FELIPE, E. da S. Avaliação da influência de textos narrativos de fatos relevantes no preço das ações de empresas brasileiras. *Revista Contabilidade e Controladoria*, v. 2, n. 2, 2010.

SILVA, C. A. T.; FERNANDES, J. L. T. Legibilidade dos fatos relevantes no brasil. *Revista de Administração Contemporânea-RAC Electronica*, v. 3, n. 1, p. 142–159, 2009.

SOUZA, M.; VIEIRA, R. Sentiment analysis on twitter data for portuguese language. In: *Proceedings of the 10th International Conference on Computational Processing of the Portuguese Language*. Berlin, Heidelberg: Springer-Verlag, 2012. (PROPOR'12), p. 241–247. ISBN 978-3-642-28884-5. Disponível em: <http://dx.doi.org/10.1007/978-3-642-28885-2_28>.

TETLOCK, P. C. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, Wiley Online Library, v. 62, n. 3, p. 1139–1168, 2007.