# ARTICLES

# PESSIMISM AND UNCERTAINTY OF THE NEWS AND INVESTOR BEHAVIOR IN BRAZIL

*Pessimismo e incerteza das notícias e o comportamento dos investidores no Brasil*

*Pesimismo e incertidumbre de las noticias y comportamiento de los inversores en Brasil*

## ABSTRACT

How investors impound qualitative information released by the media into prices, especially in a less efficient market such as Brazil, helps understand the types of news most sensitive to investors. This study investigates the relationship between the content of the daily editions of specialized financial media in Brazil, captured by a metric of textual tone, and returns and volatility of market indexes. Our database contains 1,237 daily editions of the newspaper "Valor Econômico," between 01/02/2012 and 12/30/2016. The results indicate that the market put more weight on the words "uncertainty" and "negative" in the news. "Uncertainty" has negative relation to current market-returns and weak evidence that news with "negative" terms have positive associations with current market-volatility. The evidences obtained point to the existence of informative content in the news published by specialized media in Brazil, especially with the words "negative" and "uncertainty."

**KEYWORDS |** Sentiment analysis, textual analysis, financial media, Brazil, efficient markets.

## RESUMO

*Investidores formam suas expectativas sobre os fluxos de caixa futuros das empresas considerando as informações quantitativas e qualitativas a que têm acesso. O entendimento de como os preços de mercado incorporam as informações qualitativas divulgadas pela mídia, especialmente em um mercado com menor nível de eficiência como o Brasil, ajuda na compreensão de quais tipos de notícia mais sensibilizam os investidores. Nesse contexto, este trabalho estuda a relação entre o teor das edições diárias da mídia financeira especializada no Brasil, capturado por uma métrica de tom textual, e a rentabilidade e volatilidade dos índices de mercado. A base de dados estudada contém 1.237 edições diárias do jornal Valor Econômico, compreendendo o período entre 2/1/2012 e 30/12/2016. Os resultados indicam que o mercado avalia com maior peso palavras de incerteza e negativas divulgadas nas notícias. A aparição de termos do tipo "incerteza" tem relação negativa com a rentabilidade, e há indícios mais fracos de que termos relacionados a palavras "negativas" têm associação positiva com a sua volatilidade. Tomadas em conjunto, as evidências obtidas neste estudo apontam para a existência de conteúdo informativo nas notícias veiculadas pela mídia especializada no Brasil, especialmente notícias com palavras "negativas" e de "incerteza".*

***PALAVRAS-CHAVE |*** *Análise de sentimento, análise textual, mídia financeira, Brasil, mercado eficiente.*

**FERNANDO CAIO GALDI**
fernando.galdi@fucape.br
ORCID: 0000-0001-6231-0010

**ARTHUR MARTINS GONÇALVES**
arthgonc@gmail.com
ORCID: 0000-0003-2380-9809

*Fundação Instituto Capixaba de Pesquisa em Contabilidade, Economia e Finanças, Vitória, ES, Brazil*

## RESUMEN

*Los inversores forman sus expectativas sobre los flujos de caja futuros de las empresas, considerando la información cuantitativa y cualitativa a la que tienen acceso. La comprensión de cómo los precios de mercado incorporan las informaciones cualitativas divulgadas por los medios, especialmente en un mercado con menor nivel de eficiencia como Brasil, ayuda a la comprensión de qué tipos de noticias más sensibilizan a los inversores. En este contexto, este trabajo estudia la relación entre el tenor de las ediciones diarias de los medios de comunicación financieros especializados en Brasil, capturado por una métrica de tono textual, y la rentabilidad y volatilidad de los índices de mercado. La base de datos estudiada contiene 1.237 ediciones diarias del periódico "Valor Económico", que comprenden el período del 02/01/2012 al 30/12/2016. Los resultados indican que el mercado evalúa con mayor peso palabras de incertidumbre y negatividad divulgadas en las noticias. La aparición de términos como "incertidumbre" tiene una relación negativa con la rentabilidad, y hay indicios más débiles de que las palabras "negativas" tienen una asociación positiva con la volatilidad. Las evidencias obtenidas en este estudio muestran la existencia de contenido informativo en las noticias difundidas por los medios especializados en Brasil, especialmente noticias con palabras "negativas" y de "incertidumbre".*

***PALABRAS CLAVE |*** *Análisis de sentimiento, análisis textual, medios de comunicación financieros, Brasil, mercado eficiente.*

# INTRODUCTION

Specialized media is an important data source for companies, especially those involved in the capital market, where regulations require that information be disclosed to investors equitably. The association between published reports (journalistic texts, financial blog reviews, posts in social media, rumors, etc.) and market behavior became the subject of study for several researchers, including, Antweiler and Frank (2004), who related financial blog reviews to the return of certain stocks; Tetlock (2007), Tetlock, Saar-Tsechansky, and Mackskassy (2008), Fang and Peress (2009) and Chen et al. (2011), who studied the relationship between journalism and the profitability of certain companies; Porshnev, Redkin, and Shevchenko (2013) and Bogle and Potter (2015), who discussed the possibility of predicting the market based on the tone of Twitter posts; Rogers, Skinner, and Zechman (2015), who evaluated if the way in which news stories are spread by the media affects the response of asset prices; and Bushman, Williams, and Wittenberg-Moerman (2016), who investigated whether media coverage of a borrower influences syndicated loans.

In this context, this study investigates if it is possible to identify any relationship (positive or negative) between profitability and market volatility indices (São Paulo State Stock Exchange - Ibovespa - and Brazil Broad-Based Index - IBrA) and positive or negative content (referred to as "tone") extracted from reports published by primary news sources that specialize in economic issues in Brazil.

The database used to formulate the main variables of this research was the daily edition available on the *Valor Econômico* website, which is the largest specialized newspaper in Brazil. This was the only vehicle of communication used for this research because after *Brasil Econômico* stopped circulating their print versions in 2015, *Valor Econômico* remains the only player in the market and the only source that generates information on the economy, finance, and markets in Brazil on a daily basis. According to data from the Brazilian Association of Newspapers (ANJ, 2017), in 2015, the average daily circulation of *Valor Econômico* was 41,431 copies. A sample containing all the editorial sections was included, as well as a sub-sample, called a filtered sample, considering only the following sections: Brazil, Politics, International, and Finance. Both samples were extracted over a full five-year period, from 2012 to 2016. A computational method called sentiment analysis (or opinion mining), which extracts the opinion of a text (Liu & Zhang, 2012), was used to process the database.

The sentiment analysis was performed using an algorithm that, along with word dictionaries, processed the daily editions, transforming textual information into quantitative data. This made it possible to quantitatively evaluate the tone of news stories and perform statistical analysis. The algorithm and dictionaries were based on the equivalents developed by Pagliarussi, Aguiar, and Galdi (2016).

The results indicated a negative association between returns on the day that an edition of *Valor Econômico* is put into circulation (print or online) and a higher number of terms denoting uncertainty in the said edition. In addition, weaker evidence showed a positive relation between news with a negative tone and the increase of volatility of Ibovespa.

It is necessary to understand that this study contributed to the literature of the country in both Finance and Accounting, as this is a widely researched subject internationally and is gaining popularity in Brazil. Moreover, combining sentiment analysis with machine learning algorithms can help the investor and/or regulators predict market behavior (Cambria, 2016; Tripathy, Agrawal, & Rath, 2016).

# THEORETICAL FRAMEWORK

## News and their influence in the market

The theory says that the value of a company should be equal to the present value of their expected cash flow, considering the appropriate cost of capital (Cochrane & Culp, 2003). The projection of this cash flow is conditional on other sets of information, such as a qualitative description of the business environment of companies, their operations, and the prospects presented by the financial press (Tetlock et al., 2008). The literature shows ample evidence that specialized financial media discloses information relevant to capital and credit market participants. This is in addition to the information provided by market analysts and annual reports (Bushman et al., 2016; Tetlock et al., 2008).

Generally, new relevant information about a specific company, industry or economy, may change the market's view of implied risk and expected financial profitability. Consequently, the market may rebalance the value of these companies owing to their new expectation of financial returns (Tetlock et al., 2008).

As most market investors and participants have access to the media, they could predict changes in the projected cash flow based on the tone of collected information (for instance, positive or negative news about a company, industry, or economy, in particular). This would result in the valuation or devaluation of a

company (stock), and therefore, the exchange of equity holding of companies compromised by others with better performance (Mitra & Mitra, 2011). In this way, considering the market as an aggregate, it is expected that the tone of the news on a given day be associated with the performance and volatility of market indices.

## Sentiment analysis

Evaluating the influence of investor sentiment in making decisions goes back to early studies in the 1980s, when economists began using psychological tools to explain investor behavior (Boussaidi, 2013). Barberis, Shleifer, and Vishny (1998) were the first to model investor sentiment to show the formation of beliefs based on psychological evidence and extreme reactions (overreactions or underreactions). Such research was characterized within the field of Behavioral Finance.

As such, the sentiment analysis used in this study is different, as it can be defined as a computational study of opinions, evaluations, attitudes and emotions directed toward entities, individuals, editions, events, as well as topics and their attributes (Liu & Zhang, 2012).

For an average human being, it is a difficult task to follow and read the available news in the vehicles of information (i.e. specialized media, blogs, forums, social networks, etc.). In addition, given the amount of published information, news will not always be easily decoded (understood) by the reader, which further complicates the decision-making process (Liu & Zhang, 2012).

Furthermore, usually, people tend to pay more attention to information and opinions that go against their own preferences (Liu & Zhang, 2012).

Therefore, a big advantage of using computational methods is the ability to process large volumes of text very fast, yielding consistent results, and mitigating the effect of bias by individual opinions and predilections (Liu & Zhang, 2012).

## Previous studies

Antweiler and Frank (2004) affirmed that messages about the financial market in forums influence its behavior. While studying the effect of more than 1.5 million messages posted on Yahoo! Finance and the Raging Bull about 45 companies belonging to the Dow Jones Industrial Average and Dow Jones Internet Index, they verified that the tone of the comments helps predict market

volatility. Their results showed that the impact of the messages over the stock returns was statistically relevant, despite being economically small.

Tetlock (2007) evaluated the interaction between the media and the stock market by measuring the tone of The Wall Street Journal column "Abreast of the market" from 1984 to 1999. This column discusses the reasons for the market's behavior on the previous day, and includes predictions by analysts. The author found evidence indicating that high levels of pessimism predicted a negative effect on stock prices, while average, high, and low levels of pessimism predict a high volume of negotiations. He also suggested that low profitability in the financial market leads to increasing pessimistic news.

Tetlock et al. (2008) expanded the analysis by Tetlock (2007), measuring not only the tone of a column, but also the entirety of The Wall Street Journal and the Dow Jones News Service from 1980 to 2004. They followed stock returns and investigated whether a higher number of negative words could be used to increase the expectancy of future cash flows. The results found that a higher number of negative words in specific news relating to the firm predicted low financial results, especially if the news were related to the company's financial structure (Tetlock et al., 2008).

Fang and Peress (2009) begin with the hypothesis that media influences stock profitability, even when it conveys incoherent or exaggerated information. They measured the relationship between media coverage and stock returns, and found that stock in companies that are not frequently mentioned by vehicles of information tend to have higher returns than their counterparts.

Chen et al. (2011), in a similar study as Fang and Peress (2009), used the hypothesis that the media can bring new information to the market. They observed The Wall Street Journal's coverage of companies listed on the S&P 500 Index before the disclosure of financial reports, and the behavior of stocks in terms of their profitability. They argued that the greater the media coverage on a company, the lower the chance of their stocks obtaining abnormal gains, which leads to a smaller earning response coefficient (ERC).

Loughran and McDonald (2011) presented a new methodology to analyze text. They argued that the use of dictionaries produced in other fields, such as psychology, incorrectly classified the tone of financial texts. The authors developed a new word list (Fin-Neg), using texts from the field of Finance, and concluded that about three quarters of words classified as negative by the Harvard Psychosocial Dictionary did not receive the same classification in the new word list.

They also proposed using a mathematical equation that not only takes into account the frequency of words in texts, but also measures their weight (term weighting: $w_{i,j}$, in line with equation 1 (Loughran & McDonald, 2011; Pagliarussi et al., 2016).

$$w_{i,j} = \begin{cases} \frac{(1+\log(tf_{i,j}))}{(1+\log(a_j))} \log\frac{N}{df_i} \ if \ tf_{i,j} \geq 1 \\ \qquad 0 \qquad Otherwise \end{cases} \qquad (1)$$

Where,

$tf_{i,j}$ = Total occurrences of the word $i$ in a document $j$;

$a_j$ = Proportion of words counted in a document $j$;

$N$ = Total number of documents in the sample;

$df_i$ = Total number of documents with at least once occurrence of the word $i$.

The argument is that words that appear very frequently in the text do not necessarily provide more information than others that appear less frequently. One of the functions of the algorithm in Equation 1 is to decrease the importance of these terms (Loughran & McDonald, 2011).

Pagliarussi et al. (2016) used sentiment analysis to extract opinions from management reports of certain Brazilian companies from 1997 to 2009, and relate them to their abnormal returns, abnormal volume of business, and their stock price volatility. The authors did not find evidence indicating that the management reports influenced business in the stock market. The authors also developed an algorithm to analyze texts using the formula by Loughran and McDonald (2011) and created Portuguese word dictionaries that they think could be used for any text in the field of Finance.

The dictionaries built by Pagliarussi et al. (2016) are broad and consider some relevant points in adapting to the Portuguese language. Specifically, Pagliarussi et al. (2016) commented:

> With the final list containing 22,879 distinct words, we proceeded with their classification as positive, negative, contentious, uncertainty related and modal. Some words can be classified in two or more categories (Loughran and McDonald, 2011). So, the uncertainty-related words list might contain words also occurring in the list of negative words. Another point mentioned by the authors

is that when including a word in the list of negative words, for example, consideration should also be given to the inclusion of its variants. We considered these issues in examining the words contained in the dictionary before closing the lists. The list of negative words contained 1,080 words, such as "crise,", "endividar", "impacto", "risco", "limitado", "perder", "reduzir" and "prejuízo" (in English, "crisis", "debt", "impact", "risk", "limited", "lose", "reduce" and "loss"). In addition to the negative word list, we also classified words into four other categories: positive, litigious, uncertainty and modal. The list of positive words included 701 words. Positive words are usually expected to have little impact to evaluate a text's tone (Loughran and McDonald, 2011). Many of the apparently positive words have their classification jeopardized by ambiguity, since they frequently occur in a context of negation ("did not improve"), although it is more difficult to convey positive news using negation of negative words ("did not worsen"). The list of uncertainty-related words included 170 words, such as "assumir", "variações", "especulação", "eventualidade", "imaginava", "instabilidade" and "volatilidade" (in English, "to assume", "variations", "speculation", "eventuality", "imagined", "instability" and "volatility"). Words sought in this case are those usually employed in scenarios of uncertainty and risk. As in Loughran and McDonald's study (2011), some words from the uncertainty-related words list, such as "volatilidade", "instabilidade" and "risco" (in English, "volatility", "instability" and "risk"), are also present in the list of negative words. The litigious words list contained 492 words, such as "anulação", "contestação", "investigação", "legalidade", "legitimar", "processual", "recorrer" and "suborno" (in English, "annulment", "defense", "investigation", "legality", "to legitimize", "procedural", "appeal" and "bribery"). Finally, building of the modal word list took into consideration words that express degrees of certainty or obligation. Examples of modal words are "possível", "provável", "improvável", "necessário", "talvez", "deve", "claramente", and "compulsório", ("possible", "likely", "unlikely", "necessary", "maybe", "ought", "clearly" and

"compulsory"). The modal list contained 81 words. We prepared the lists out of a corpus that includes an excess of 8 million words occurring in texts directed primarily to the stakeholders of the Brazilian capital market. (p.57)

Even in Brazil, Nascimento, Osiek, and Xexéo (2015) used sentiment analysis to investigate the reaction of the population to news published by the media, and capture the reactions to comments posted on the social network Twitter.

## METHODOLOGY

### Data collection and treatment

Three computational programs were used to collect and treat the data. The first two, developed in Java, download the daily editions of the *Valor Econômico* newspaper (just the free part) from their website. The third was developed in Python by Pagliarussi et al. (2016), and was used to perform the sentiment analysis of the files produced by the first two.

The sample period extended from January 2, 2012 to December 30, 2016, due to the availability of the *Valor Econômico* newspaper on the website. Days on which the newspaper did not circulate or on which BM&FBovespa was not open were excluded from the database. Consequently, the sample consisted of 1,237 daily editions of the newspaper disseminated in both print and electronic media, from which the tones of the news stories were extracted.

The editorial sections of the newspaper, which were considered in the full sample of the news database, are Brazil, Politics, Finance, Companies, Agribusinesses, International, Opinion, Legislation, Careers, Culture, and Style. In addition, a sub-sample was considered, called a filtered edition, from which the sections that are not primarily related to the development of Brazilian capital markets were removed. The subsample with the filtered edition was composed considering the following sections: Brazil, Politics, International, and Finance.

Every edition is saved in a .txt format file named by the year, month, and day of release (for example, "20120307.txt" for complete editions and "20120307-BPIF.txt" for filtered editions.) This approach was chosen because it would optimize data organization.

With respect to the first two algorithms, Exhibits 1 and 2 provide an example of their functionality. The following news report was taken from the May 15, 2013 edition of the *Valor Econômico* newspaper.

### Exhibit 1. First news report from the 5/15/2013 edition (complete edition)

Auctions attract investments of RS 7 billion in petroleum
The planned investments of the companies that yesterday purchased exploratory areas for petroleum and gas are estimated at R$ 7 billion. The 11ANP bidding round, scheduled to end today, concluded in a single day and warranted strong government commemoration.

Source: *Valor Econômico (2013)*.
Note: Text originally written in Portuguese.

In the generated file ("20130515.txt"), presented in the following image, the text in Figure 1 is transformed into:

### Exhibit 2. Part of the text in the file "20130515.txt"

Auctions attract investments of R$ 7 billion in petroleum. The planned investments of the companies that yesterday purchased exploratory areas for petroleum and gas are estimated at R$ 7 billion. The 11 ANP bidding round, scheduled to end today, concluded in a single day and warranted strong government commemoration.

Source: *Valor Econômico (2013)*.
Note: Text originally written in Portuguese.

Comparing both images it can be observed in Exhibit 2 that all special characters or those with accents were removed or replaced with their non-accented counterparts. For example: "ã" was replaced with "a", "ê" with "e", and "ç" with "c". Meanwhile, "ª" was eliminated, and so on.

These replacements became necessary owing to compatibility issues with the compiler (Python) used. When these characters were present, the algorithm did not identify the words that they contained, indicating that they were missing from the sentiment analysis process. This would generate incorrect values for the primary variables. The same error was also reported by Pagliarussi et al. (2016).

The control variables used are the three factors provided by Fama and French (1993): Small Minus Big (SMB) and High Minus Low (HML), with Market Factor (market risk subtracted from a risk-free interest state) excluded, given that the dependent variable is directly related to market risk. Factors provided by Carhart (1997) and Amihud (2002) were also included: Winners Minus Losers (WML) and Illiquid Minus Liquid (IML), respectively.

The Bovespa and IBrA indices were chosen for the dependent variables. This choice is because the first is the most used in Brazil and that the second has the largest number of companies.

To illustrate, Table 1 presents a brief comparison of the two indexes.

Table 1. Comparison of Ibovespa and IBrA

|  | Ibovespa | | IBrA | |
|---|---|---|---|---|
| Number of companies | 59 | | 115 | |
| Total of top 10 companies | 55.14% | | 49.37% | |
| Total of top 20 companies | 75.04% | | 67.83% | |
| Greatest participation | ITUB4 | 11.29% | ITUB4 | 10.02% |
| Second greatest participation | BBDC4 | 7.99% | BBDC4 | 7.09% |
| Third greatest participation | ABEV3 | 7.31% | ABEV3 | 6.49% |

Source: Adapted from BM&FBovespa Ibovespa (2015a); BM&FBovespa IBrA (2015b, 2015c).

## Operationalization of the Loughran and McDonald (2011) equation

To simplify the construction of the tone variables, we present an operationalization of the calculations in this section.

In the following example, a dictionary of negative words was used that contains just the terms "mensalao" (a neologism roughly meaning "large monthly payments," which is linked to a vote-buying scandal in Brazil) and "loss" (as neither of these words have accents and when used, the Python version 2.7 does not identify them, excluding them from the analysis. This process was used in all analyses).

Exhibits 3, 4, and 5 show portions of the June 11, 12, and 13 editions, respectively, from 2012. Only a small portion of these newspapers was used, as the figures would otherwise be very long without affecting the mathematical analysis.

Exhibit 3. Part of file 20120611.txt

Brazil Congress Candidates use CPI against opponents in municipal elections. A coincidence in timing has encouraged politicians to use research to try to weaken opposing political groups. Revenue for a bad mayor. We can rule out bad mayoral candidates by seeing if they discuss topics fundamental to the city or if they avoid them. A 2014 dispute is motivating Alckmin and Kassab to search for allies for Serra. Current alliance is circumstantial and does not rule out confrontation between the two in the next elections. PT chooses name for vice of Lacerda. Tucano launches candidacy to the City of Rio. International Energy USA buys less petroleum abroad. Country reduces dependency on imported petroleum, thanks to an increase in domestic production, more efficient engines, and the use of ethanol. Shorts Data dispels pessimism over China. Numbers released over the weekend indicate a much stronger economy than what many in the market feared one week ago. Parallel dollar out of control in Venezuela. Bolivar suffers record losses in the parallel market in Venezuela while individuals race due to the shortage of American currency. March of anti-Chavez. Special Juncture Industrial production suffers but confidence improves in benefited sectors. Clothing and footwear manufacturing still skates, but entrepreneurs report improved confidence after payroll tax relief. Country needs to stimulate growth in the private sector, says Araujo valor.com.br. Arrogance reigns there as much as here. With a little tolerance, it is possible to look upon our southern neighbors without preconceived notions or conservative biases. Flu vaccination slows and specialists are already discussing the anticipation of campaigns. Flu vaccine delivery delayed and governments and experts anticipate upcoming vaccination campaigns. Infrastructure Public investment in transportation in 2012 is not jumping off the page.

Source: *Valor Econômico* (2012).
Note: Text originally written in Portuguese.

Exhibit 4. Part of file 20120612.txt

Some exchange any gift for a ride on wheels alongside the half-face. On Valentine's Day, we tell stories with a happy ending and hide those that are not publishable. Brazil Congress Ruralistas defend Mendes Ribeiro. Minister has been stripped of authorization by the president in negotiations relating to the Forest Code. Curtas Dirceu waiting for Armageddon 4. On the eve of the trial of the mensalao defendants, former minister calls on students for the "final battle." Federal pact commission maintains unanimity requirement in Confaz. Group of 14 experts called on by the president of the Senate, Jose Sarney, also favors the adoption of the IPCA for the indexation of state debts valor.com.br Commission will investigate Pimentel's use of plane. Committee decided to deepen investigations of consulting services provided by the ministry. Elections Freixo launches with appeal for a "Rio de Janeiro" spring. 1 PSOL candidate has the support of artists and composers, like Caetano Veloso and Chico Buarque. Truth Commission endorses calling agents of the dictatorship. Regional Truth Commissions will be allowed to call alleged torturers by the national body. At stake is information that can be used to sell advertising on mobile devices. Aviation TAM stock fell by 2.09% on the eve of the creation of Latam. Brazilian airline closes capital today and the Chilean Lan will negotiate BDRs on the BM&FBovespa. Europeans must have losses of 1.1 billion US dollars. Airlines offer iPads to passengers to reduce costs. Arezzo shoes starts expanding a more expensive brand. Schutz franchise project allows a looser gross margin to higher-end stores. Health Sabara must replicate AACD model. Eduardo Carneiro, from the Sabara hospital, calculates that in 2014, it will reach a balance. Vapza Foods reduces package and attracts women.

Source: *Valor Econômico* (2012).
Note: Text originally written in Portuguese.

## Exhibit 5. Part of file 20120613.txt

Sentencing for the <u>mensalao</u> is still delayed. None of the 38 accused in the <u>mensalao</u> will be arrested in the Federal Supreme Court decision, scheduled to begin on August 1, and unlikely to end before September. Brazil Government CGU prohibits Delta from closing new contracts. Company inequity, decided by CGU, extends to states and municipalities. The historical crossroads of the <u>mensalao</u> 6. The trial decided not only the future of the 38 accused, but also the route or path that Brazil will take. Perillo's testimony should lead CPI to back off from its breach of secrecy. PT and PSDB deny non-aggression agreement in relation to their governors. Cachoeira got a vote for habeas corpus. Chamber approves MP that simplifies PAC biddings. A new kind of European union. If it seems unlikely that current policies work, I suggest a combination of two ideas: insurance-based bonds and adjustment-based bonds. Judiciary Resources may take <u>mensalao</u> to 2013. In a process marked by uncertainty, the only certainty is that no one will go to jail due to the Supreme Court case. Ayres Britto denies pressure on STF. Companies Oil Comgas will be BG's client in the presal. Comgas will be able to buy the British part of the gas produced in the Lula field. Featured Bankruptcy motion Accounting Brazil sends more resources to the IFRS. Endowment increases 16%, but the contribution of the country to the foundation that governs international accounting is still small. Stock offer Facebook investor sues stock market.

Source: *Valor Econômico* (2012).
Note: Text originally written in Portuguese.

Using the formula proposed by Loughran and McDonald (2011) presented in Equation 1 to complete the sentiment analysis of the texts contained in the boxes above yields:

In Exhibit 1, neither of the two words was found; therefore, the result is obtained according to the equations below:

| | |
|---|---|
| $$w_{loss,\,20120611} = 0 \text{ (A)}$$ | (A) |

| | |
|---|---|
| $$w_{mensalao,\,20120611} = 0 \text{ (B)}$$ | (B) |

Meanwhile, in Figure 2, both words appear once in the text; solving the equation for this case yields:

For the calculation of the word "loss," the values are as follows:

$tf_{loss,\,20120612} = 1$ (There is just one occurrence in document 20120612);

$N = 3$ (Document total, three newspapers);

$df_{loss} = 1$ (Only occurs in one document).

Calculating $a_{20120612}$:

| | |
|---|---|
| $$a_{20120612} = \frac{nc}{nt}$$ | (C) |

Where:

$nc$ = Sum of occurrences of words "loss" and "mensalão"

$nt$ = Sum of a single presence of each dictionary word in the document analyzed.

| | |
|---|---|
| $$a_{20120612} = \frac{2}{2} = 1$$ | (D) |

Improving on the explanation of the calculation of $a_j$, in Exhibit 2, both the words "mensalão" and "loss" appear once, making equation D the result of this analysis. If, for example, "mensalão" were to have occurred twice, and "loss" just once, then the new value of $a_{20120612}$ is $a_{20120612} = 1.5$ because $nc = 3$ while $nt = 2$. For Exhibit 5, it is possible to find the word "mensalão" four times, while "loss" does not appear. In this way, the result of $a_{20120612}$ is

| | |
|---|---|
| $$a_{20120613} = \frac{4}{1} = 4$$ | (E) |

Replacing the values in equation (1)

| | |
|---|---|
| $$w_{loss,\,20120612} = \frac{(1+\ln(1))}{(1+\ln(1))}\, ln\, \frac{3}{1}$$ | (F) |

| | |
|---|---|
| $$w_{loss,\,20120612} = ln3$$ | (G) |

| | |
|---|---|
| $$w_{loss,\,20120612} = 1.098612289$$ | (H) |

For the calculation of the word "mensalão," the values are as follows:

$tf_{mensalao,\,20120612} = 1$ (There is just one occurrence in document 20120612);

$N = 3$ (Document total, three newspapers);

$df_{mensalao,\ 20120612} = 2$ (Occurs in two documents, 20120612 and 20120613).

$a_{mensalao,\ 20120612} = 1$ (Result of equation D)

Again, substituting the values in equation (1)

$$w_{mensalao,20120612} = \frac{(1+\ln(1))}{(1+\ln(1))} \ln \frac{3}{2} \qquad \text{(I)}$$

$$w_{mensalao,\ 20120612} = \ln 1.5 \qquad \text{(J)}$$

$$w_{mensalao,\ 20120612} = \ln 0.405465 \qquad \text{(K)}$$

Summing the two values yields

$$w_{loss,\ 20120612} + w_{mensalao,\ 20120612} = \ln 1.504077 \qquad \text{(L)}$$

The value found in equation K is the weight of the words (for the dictionary of negative words used in this example) for the day of June 12, 2012.

Finally, in Exhibit 3, the word "loss" does not occur, making the result $w_{loss,\ 20120612} = 0$. In this case, "mensalão" appears four times. The calculation in this case is as follows:

$tf_{mensalao,\ 20120612} = 4$, $a_{20120613} = 4$, $N=3$, $df_{mensalao} = 2$

Replacing these values in equation (1)

$$w_{mensalao,20120612} = \frac{(1+\ln(4))}{(1+\ln(4))} \ln \frac{3}{2} \qquad \text{(M)}$$

$$w_{mensalao,\ 20120612} = \ln 1.5 \qquad \text{(N)}$$

$$w_{mensalao,\ 20120612} = 0.405465 \qquad \text{(O)}$$

The value determined in equation M is the value of the weight of negative words on the day of June 13, 2013.

To calculate the weight of each word in the sample presented in Exhibits 1, 2, and 3, all of the $w_{i,j}$ must be summed for "loss" and "mensalão." The equations for P and Q simplify as

$$w_{loss,\ 20120611} + w_{loss,\ 20120612} + w_{loss,\ 20120613} = 1.504077 \qquad \text{(P)}$$

$$w_{mensalao,\ 20120611} + w_{mensalao,\ 20120612} + w_{mensalao,\ 20120613} = 1.81093 \qquad \text{(Q)}$$

## Econometric models

The dependent variables are ibov, ibra, ibov_vol, and ibra_vol. These represent the daily profitability (equations 4 and 12) and volatilities (equations 6 and 13) of Ibovespa and IBrA, respectively.

The independent variables are the weight of the negative, positive, litigious, uncertain, and modal (primary variables) words (term weighting) and SMB, HML, WML, IML, and riskfree (control variables). In addition, SMB_vol, HML_vol, WML_vol, IML_vol, and riskfree_vol, which are the respective volatilities of the control variables, were added when the dependent variables were the volatility of both the indices.

Equations 2 and 3 represent the econometric models in which the dependent variables are related to the Bovespa Index.

$$ibov = \beta_0 + \beta_1\ negatives + \beta_2\ positives + \beta_3\ litigious + \beta_4\ uncertainty + \beta_5\ modals + \beta_6\ SMB + \beta_7\ HML + \beta_8\ WML + \beta_9\ IML + \beta_{10}\ riskfree + u \qquad \text{(2)}$$

$$ibov\_vol = \beta_0 + \beta1\ negatives + \beta_2\ positives + \beta_3\ litigious + \beta_4\ uncertainty + \beta_5\ modals + \beta_6\ SMB + \beta_7\ HML + \beta_8\ WML + \beta_9\ IML + \beta_{10}\ riskfree + \beta_{11}\ SMB\_vol + \beta_{12}\ HML\_vol + \beta_{13}\ WML\_vol + \beta_{14}\ IML\_vol + \beta_{15}\ riskfree\_vol + u \qquad \text{(3)}$$

In which:

$$ibov_t = ln\frac{B_t}{B_{t-1}} \quad (4)$$

In which

$t$ = A date (business day), ranges from January 2, 2012 to December 30, 2014.

$B_t$ = Closing value of the Bovespa Index for a given day (ex.: $B_{01/08/13}$ = 49.140).

$B_{t-1}$ = Closing value of the Bovespa Index on the preceding day $t$ (ex.:$B_{31/07/13}$ = 48.234). With respect to the volatility ($ibov\_vol_t$), the formula is as follows:

$$ibov\_vol_t = \sqrt{\frac{1}{n-1}\sum_{i=t}^{n}(x_i - \overline{x})^2} \quad (5)$$

Similarly to Equations 2 and 3, Equations 6 and 7 now have a dependent variable related to the IBrA.

$$ibra = \beta_0 + \beta_1\,negatives + \beta_2\,positives + litigious + \beta_4\,uncertainty + \beta_5\,modals + \beta_6\,SMB + \beta_7\,HML + \beta_8\,WML + \beta_9\,IML + \beta_{10}\,riskfree + u \quad (6)$$

$$ibra\_vol = \beta_0 + \beta_1\,negatives + \beta_2\,positives + \beta_3\,litigious + \beta_4\,uncertainty + \beta_5\,modals + \beta_6\,SMB + \beta_7\,HML + \beta_8\,WML + \beta_9\,IML + \beta_{10}\,riskfree + \beta_{11}\,SMB\_vol + \beta_{12}\,HML\_vol + \beta_{13}\,WML\_vol + \beta_{14}\,IML\_vol + \beta_{15}\,riskfree\_vol + u \quad (7)$$

Where

$$ibra_t = ln\frac{A_t}{A_{t-1}} \quad (8)$$

In which:

$t$ = A date (business day), ranges from January 2, 2012 to December 30, 2014.

$A_t$ = Closing value of the IBrA for a given day $t$.

$A_{t-1}$ = Closing value of the IBrA before day $(t-1)$

With respect to the volatility $(ibra\_vol_t)$, the formula is as follows:

In which:

$n$ = Number of days (ex: $n$ = 60, value used in this work).

$i$ = A given day, begins on $t$ and goes until $t-59$ (ex: $t$ = August 1, 2013, $t-1$ = July 31, 2013, $t-2$ = July 30, 2013, …,, $t-59$ = May 8, 2013).

$x_i$ = Value of $ibov_t$ on a given date (ex:$ibov_{01/08/2013}$ = 1.86%).

$\overline{x}$ = Mean value of $ibov_t$ within a certain period $t$ until $t-59$ (60 business days).

The methodology used to calculate the variables SMB_vol, HML_vol, WML_vol, IML_vol, and riskfree_vol was the same as applied to obtain the result for ibov_vol, changing, evidently, the values for ibov in $x_i$ and $\overline{x}$ for the respective values of SMB, HML, WML, IML, and riskfree.

$$ibra\_vol_t = \sqrt{\frac{1}{n-1}\sum_{i=t}^{n}(x_i - \overline{x})^2} \quad (9)$$

In which:

$n$ = Number of days (ex: $n$ = 60, value used in this work).

$i$ = A given day, begins on $t$ and goes until $t-59$ (ex: $t$ = August 1, 2013, $t-1$ = July 31, 2013, $t-2$ = 7/30/2013, …, , $t-59$ = May 8, 2013).

$x_i$ = Value of $ibra_t$ on a given date (ex: $ibra_{01/08/2013}$ = 1,84%).

$\overline{x}$ = Mean value of $ibra_t$ within a certain period $t$ until $t-59$ (60 business days).

The daily Ibovespa closing values (used in $B_t$ and $B_{t-1}$) were extracted from the database of the Institute for Applied Economic Research (IpeaData, 2015), while those for IBrA (referred in $A_t$ and $A_{t-1}$) were retrieved from the BM&FBovespa database.

The variables SMB, HML, WML, IML, and riskfree were obtained from the website of the Center for Financial Economic Research (Nefin), connected to the Economics Department of the School of Economics, Business Administration, and Accounting of the University of São Paulo.

In all the econometric models, the applied method to estimate the parameters was that of ordinary least squares (OLS), using fixed effects for year and cluster-robust standard errors.

## Hypothesis and expected behavior

From the econometric equations presented in the previous section (profitability: Equations 2 and 6; and volatility: Equations 3 and 7), the tested variables are the weight of the negative; positive; litigious; uncertainty; and modal words (primary variables). Thus, using "negative" as an example, the hypothesis is that if $\beta_1$ is equal to zero, it means that the "negative" variable does not affect what returns, this possibly being, ibov, ibov_vol, ibra, and ibra_vol. Otherwise, it would not be possible to discard "negatives" from the previously mentioned equations. The same example works with the variables "positives", "uncertainty", "litigious", and "modals"; however, changing the value of Beta ($\beta$) for its respective value.

Staying with the primary variables, "negatives" have negative signs in Equations 2 and 6 (related to daily profitability of the indices), and positive signs in Equations 3 and 7 (related to volatility). The opposite effect is expected for the "positives" or rather, positive signs in Equations 2 and 6 and negatives in Equations 3 and 7 (Tetlock, 2007; Tetlock et al., 2008).

Like the "negative" variable, the "uncertain" and "litigious" ones yield a lower profitability (negative sign) in Equations 2 and 6, and positive in Equations 3 and 7. It is believed that a greater weight of these words in newspapers increases uncertainty about the direction the market will take.

Finally, there are no expectations on "modals" with respect to signs of profitability and volatility.

# RESULTS

## Word analysis (term weighting)

Table 2 shows the most heavily weighted words within the sample of collected newspapers. When comparing the results in the two tables, there is significant difference between the most heavily weighted words in both cases. For instance, the word "development," which belongs to the positive category, was the most weighted (significant) in Table 2 for both every individual year and all the years pooled together. Nonetheless, in Table 3, this word was not among the top five of the most significant words, in any case.

Another interesting example is the word "mensalão," which belongs to the negative category. Returning to the tables, "mensalão" was the most heavily weighted word in 2012, with a weight equal to 99.78, according to Table 2. However, in Table 3, for the same year, "mensalão" appeared in the second position, with a higher weight equal to 103.85. Respectively, "mensalão" appeared 313 times in the sample that contained all the editions and 255 times in the sample with filtered editions.

The difference between the results above can be explained by how Equation (1), used in the algorithm by Pagliarussi et al. (2016), treats the data.

## Descriptive statistics

Table 3 displays descriptive statistics of the main variables for the complete and filtered editions, respectively. For the complete editions, the negative words (in average) were weighted more, followed by the positive and litigious words. Words in the modal and uncertainty category exhibited the lowest values, respectively.

A similar situation occurs in the case of the filtered editions. However, these showed lower average, median, maximum, and minimum values. Besides, the standard deviation also presented lower values. The lower number of words within that sample explains this outcome.

Similarly to the study by Davolos, Rogers, Silva, and Oliveira (2013), to contribute to the understanding of the obtained results, Table 4 shows the main news published in *Valor Econômico* over 24 days with the highest returns and 24 days with the lowest returns in Ibovespa, in the analyzed period. In Panel A, on the days with the highest returns, reports with a positive tone, of a political nature, and relating to great impacts on the economy dominate the news. Similarly, Panel B is dominated by economic and political news with a negative tone.

## Regression analysis

Table 5 displays the results of the estimations from Equation 2 and 6. The dependent variables are: "ibov" and "ibra," which are

Fernando Caio Galdi | Arthur Martins Gonçalves

the daily profitability of the Bovespa Index and IBrA, and "ibov_vol" and "ibra_vol," which represent the volatility of both indices calculated over a period of 60 days.

To organize the table better, it was decided to differentiate between the dependent variables of the sample composed of all the editions and the sub-sample, in which only the following sections were analyzed, Brazil, Politics, International, and Finance.

The results show that the weight of words with "uncertainty" tone is statistically relevant in explaining Ibovespa returns when all editions are considered, as well as Ibovespa and IBrA returns when only the filtered editions are considered. Words with a "Negative," "Positive," and "Modal" tone were not significant in any of the equations.

Table 2. **Term weighting**

| | All editions | | Filtered editions | |
|---|---|---|---|---|
| | loss | 468.17 | crisis | 468.21 |
| | discount | 454.96 | inflation | 467.42 |
| Negatives | deficit | 444.88 | deficit | 460.46 |
| | low | 432.67 | risk | 447.71 |
| | cut | 431.63 | reduction | 445.41 |
| | development | 434.53 | growth | 484.31 |
| | grew | 422.64 | investments | 480.18 |
| Positives | invests | 421.81 | investment | 472.50 |
| | trust | 417.36 | grow | 454.24 |
| | gain | 417.03 | trust | 393.26 |
| | judicial | 434.78 | resources | 471.03 |
| | law | 431.58 | disputes | 446.57 |
| Litigious | contract | 429.32 | fiscal | 444.62 |
| | contracts | 424.17 | rules | 427.87 |
| | creditors | 420.01 | law | 406.82 |
| | risks | 447.18 | wait | 459.93 |
| | expectation | 438.80 | risk | 449.52 |
| Uncertainty | possible | 430.78 | expectation | 397.10 |
| | expectations | 423.16 | risks | 377.11 |
| | tendency | 417.41 | expectations | 350.87 |
| | high | 336.89 | less | 415.24 |
| | little | 328.24 | smaller | 413.83 |
| Modals | large | 322.16 | strong | 407.08 |
| | majority | 321.18 | large (pl.) | 370.45 |
| | greater | 314.71 | little | 354.36 |

## Table 3. Descriptive statistics

|  |  | Negatives | Positives | Uncertainty | Litigious | Modals |
|---|---|---|---|---|---|---|
| Complete editions | Mean | 59.381 | 33.807 | 14.576 | 24.317 | 6.853 |
|  | Median | 39.015 | 23.753 | 15.683 | 20.963 | 5.093 |
|  | Maximum | 107.302 | 69.974 | 34.512 | 58.325 | 22.062 |
|  | Minimum | 15.418 | 8.815 | 1.909 | 3.145 | 0.486 |
|  | Standard deviation | 13.988 | 10.603 | 5.711 | 9.001 | 3.157 |
| Filtered editions | Mean | 37.334 | 18.835 | 9.199 | 14.017 | 6.066 |
|  | Median | 24.620 | 15.317 | 5.048 | 11.821 | 3.282 |
|  | Maximum | 81.015 | 55.908 | 30.563 | 49.683 | 25.578 |
|  | Minimum | 10.229 | 0.000 | 0.000 | 0.000 | 0.000 |
|  | Standard deviation | 11.288 | 8.147 | 5.118 | 7.736 | 3.613 |

## Table 4. Main news stories published on the days that Ibovespa fluctuated the most between January 2012 and December 2016

| Panel A. Most positive returns over the period and the news of the day | | | |
|---|---|---|---|
| Date | Ibovespa | % day | News |
| Oct. 22, 2014 | 56,432.03 | +7.75% | Abstentions should reach record level in the 2nd round |
|  |  |  | "Kit elections" still have speculative gains |
| Mar. 17, 2016 | 50,886.40 | +6.66% | Clip of Lula and Dilma plunges country into political chaos |
|  |  |  | Federal Reserve cuts interest rate hike |
| Mar. 3, 2016 | 47,288.28 | +5.65% | Cunha must become defendant, but says that he will not resign |
|  |  |  | Distant from PT, Dilma seeks to save mandate |
| Nov. 21, 2014 | 56,055.06 | +4.96% | Weak economy and strong dollar decrease profits |
|  |  |  | Marina supports Aécio and cites "Letter to the Brazilians" |
| Jul. 27, 2012 | 56,515.35 | +4.70% | Government only agrees to linear increase to servant |
|  |  |  | Credit grows and default retreats discretely |
| Nov. 3, 2015 | 48,023.72 | +4.65% | Accounting of interest and dollar already impacts balance sheets |
|  |  |  | MP and PF open six fronts of investigation against Lula |
| Aug. 24, 2015 | 44,312.77 | +4.63% | Government plans to raise taxes next year |
|  |  |  | Midsize bank profit down 43% |

## Table 4. Main news stories published on the days that Ibovespa fluctuated the most between January 2012 and December 2016

| Panel A. Most positive returns over the period and the news of the day | | | |
|---|---|---|---|
| Date | Ibovespa | % day | News |
| Oct. 6, 2014 | 57,013.57 | +4.57% | Votes maintain PT-PSDB polarization |
| | | | Performance of Aécio Neves should boost the market |
| Jan. 29, 2016 | 40,263.28 | +4.53% | Oil field concessions will be extended |
| | | | Petrobras wants to pull out of various areas |
| Oct. 28, 2014 | 52,294.96 | +4.42% | Lula makes three nominations for Fazenda |
| | | | Companies request definitions from Dilma |
| Feb. 22, 2016 | 43,304.60 | +4.18% | Renegotiation of state debts raises deficit |
| | | | Commodity manufacturers drop GDP |
| Nov. 7, 2016 | 64,092.86 | +4.16% | Government wants to change law to intervene in Oi |
| | | | FBI files new case against Hillary |
| Oct. 31, 2014 | 54,666.09 | +4.14% | Interest should continue rising |
| | | | Vale loss of R$ 3 bn surprises |
| May 10, 2016 | 53,051.05 | +4.08% | Renan ignores Maranhão and continues with the impeachment |
| | | | BC president loses minister status |
| Mar. 4, 2016 | 49,168.99 | +3.98% | Recession spreads and threatens country |
| | | | Accusation of Delcídio encourages supporters of impeachment |
| Apr. 12, 2016 | 52,068.20 | +3.94% | Chamber Commission approves opening impeachment |
| | | | Market already indicates change of government |
| May 21, 2012 | 56,583.31 | +3.83% | CGU sees irregularities in FGTS applications |
| | | | Dividend already yields more than real interest |
| Jan. 3, 2012 | 59,224.75 | +3.82% | 2012 concessions will demand investments of R$ 90 bn |
| | | | Exports feel the weight of the crisis |
| Dec. 9, 2015 | 46,082.81 | +3.82% | Impeachment in the hands of the opposition |
| | | | Commodity decline deepens |
| Oct. 2, 2015 | 47,012.50 | +3.80% | Measure broad power of BC and CVM to investigate and punish |
| | | | Petrobras investment halves |

Table 4. **Main news stories published on the days that Ibovespa fluctuated the most between January 2012 and December 2016**

| | | | Panel A. Most positive returns over the period and the news of the day |
|---|---|---|---|
| **Date** | **Ibovespa** | **% day** | **News** |
| Dec. 17, 2014 | 48,722.59 | +3.67% | Russia crisis presses emerging markets |
| | | | For Levy, fiscal adjustment even more urgent |
| Aug. 27, 2015 | 47,689.07 | +3.60% | Government will propose the return of the CPMF to cover gap |
| | | | Entrepreneur listens to Dilma and presses Levy |
| Jan. 2, 2013 | 62,761.20 | +3.55% | Industry should have a "slack" in costs of 20% in 2013 |
| | | | Economists predict GDP above 3% |
| Sep. 13, 2012 | 61,979.58 | +3.45% | Deceleration is good for China and the world |
| | | | Companies default at record pace |
| | | | Panel B. Most negative returns over the period and the news of the day |
| **Date** | **Ibovespa** | **%** | **News** |
| Oct. 15, 2014 | 56,134.11 | -3.18% | Decay of petroleum puts an end to gasoline lag |
| | | | Dilma could veto new exonerations |
| May 17, 2012 | 54,022.06 | -3.28% | Crisis tends to slow recovery for 2013 |
| | | | Greek banks lose credit and deposits |
| Jul. 10, 2012 | 53,581.92 | -3.31% | Electricity companies push to define contracts |
| | | | Ibama speeds up licenses for petroleum and gas |
| Oct. 16, 2014 | 54,270.88 | -3.32% | Decay of petroleum threatens Petrobras investments |
| | | | Deceleration crashes markets around the world |
| Dec. 8, 2014 | 50,246.99 | -3.35% | Lack of water becomes a risk factor for credit |
| | | | Government predicts burden increase in LDO |
| Oct. 10, 2014 | 55,317.26 | -3.41% | Whistleblowers revealed tipping scheme at Petrobras |
| | | | Equating debt is the main objective of Oi |
| Jun. 10, 2016 | 49,337.70 | -3.42% | Fraga suggests a longer deadline to the goal of 4.5% |
| | | | Accusations to mark US election |
| Jun. 19, 2013 | 47,742.76 | -3.44% | Bad mood hits market and mega-deal is canceled |
| | | | Fed victims spend a day hoping |
| May 14, 2012 | 57,411.51 | -3.50% | Banks intensify the increase in tariffs |
| | | | Unpredictable performance of fixed income |

Table 4. Main news stories published on the days that Ibovespa fluctuated the most between January 2012 and December 2016

Conclusion

| Panel B. Most negative returns over the period and the news of the day | | | |
|---|---|---|---|
| Date | Ibovespa | % | News |
| Oct. 21, 2014 | 52,373.42 | -3.50% | Business profit expected to fall 10% in the 3rd quarter |
| | | | Fundos try to attract resources abroad |
| Dec. 12, 2014 | 48,012.83 | -3.55% | CEO of Petrobras says he warned Graça about abuse |
| | | | Electricity bill increases the January IPCA |
| Mar. 15, 2016 | 47,098.11 | -3.59% | Judge passes Moro's arrest warrant and Lula will be minister |
| | | | Congress discusses how to remove Dilma |
| Apr. 4, 2016 | 48,746.98 | -3.61% | Business debts grow 24% |
| | | | The results of the developers are falling |
| Oct. 27, 2014 | 50,083.17 | -3.63% | Dilma, re-elected, promises dialogue |
| | | | Tight win keeps Lula under the spotlight |
| Sep. 9, 2016 | 57,952.98 | -3.70% | Government employee earns up to 200% more than a private one |
| | | | Grows reaction to the increase in servers |
| Jul. 2, 2013 | 45,244.99 | -4.13% | By decree, BNDES helps in surplus |
| | | | Dilma denies ministerial reform |
| Oct. 13, 2015 | 47,346.20 | -4.19% | Agreements to reduce wages and work hours progress |
| | | | Cornered, Cunha dispatches impeachment today |
| Nov. 11, 2016 | 58,827.82 | -4.21% | "Trump Effect" increases the value of the dollar and generates turbulence |
| | | | China resumes offensive for TPP without United States |
| Dec. 1, 2016 | 59,466.33 | -4.23% | Economy melts and pressure grows due to interest cuts |
| | | | States may win 'recovery law' |
| Feb. 2, 2016 | 38,770.10 | -4.27% | Brazilian bonds pay high interest and see attraction |
| | | | Steel crisis |
| Dec. 1, 2014 | 52,218.40 | -4.43% | Levy wants new business financing model |
| | | | Remains to pay already exceed investment forecast |
| Sep. 29, 2014 | 54,599.67 | -4.57% | Government starts siege on employment fraud |
| | | | BC resists transaction between BTG and Nacional |
| Oct. 24, 2014 | 51,968.04 | -7.91% | Market reflects progress of Dilma |
| | | | Whistleblower also cites corruption at Eletrobras |
| Aug. 21, 2015 | 42,352.03 | -9.09% | Janot asks 184 years in prison for Cunha |
| | | | Fear will break out of government articulation |

Table 5. **Relationship between news tone and return**

|  | Complete editions | | Filtered editions | |
|---|---|---|---|---|
|  | Ibov | Ibra | Ibov | Ibra |
| Negative | 0.0000274 | 0.0000334 | 0.0000325 | 0.0000471 |
|  | (-0.99) | (-1.11) | (-1.29) | (-1.78) |
| Positive | 0.00000981 | 0.000011 | 0.0000337 | 0.0000391 |
|  | (-0.55) | (-0.62) | (-1.91) | (-1.75) |
| Uncertainty | -0.0000951** | -0.0000545 | -0.0000801* | -0.0000774* |
|  | (-3.51) | (-1.83) | (-2.52) | (-2.43) |
| Litigious | 0.0000435 | 0.0000482 | -0.0000142 | -0.00000187 |
|  | (-1.34) | (-1.34) | (-0.37) | (-0.04) |
| Modal | 0.000119 | 0.000132 | -0.0000605 | -0.0000659 |
|  | (-1.3) | (-1.76) | (-0.63) | (-0.73) |
| SMB | 0.324* | 0.195 | 0.321* | 0.192 |
|  | (-2.54) | (-1.23) | (-2.51) | (-1.21) |
| HML | 0.590** | 0.505** | 0.584** | 0.499** |
|  | (-3.90) | (-3.04) | (-3.85) | (-3.00) |
| WML | -0.357** | -0.258* | -0.357** | -0.256* |
|  | (-3.74) | (-2.43) | (-3.76) | (-2.43) |
| IML | -1.205*** | -0.963** | -1.203*** | -0.961** |
|  | (-6.48) | (-4.26) | (-6.35) | (-4.18) |
| Riskfree | 1.691 | -1.628 | 0.438 | -2.618 |
|  | (-0.21) | (-0.17) | (-0.05) | (-0.26) |
| Constant | -0.00313 | -0.00239 | -0.000468 | 0.000542 |
|  | (-0.69) | (-0.48) | (-0.09) | (-0.10) |
| Fixed year effect | yes | yes | yes | yes |
| N | 1237 | 1237 | 1237 | 1237 |
| $R^2$ | 0.496 | 0.436 | 0.494 | 0.434 |

Note: Coefficients in the first row and p-value statistics in parentheses in the second row of each variable ($\alpha < 10\%*$, $\alpha < 5\%**$, $\alpha < 1\%***$).

Generally, the control variables, SMB, HML, and WML and riskfree are statistically significant, demonstrating the estimation adequacy of the four-factor model by Fama and French (1993).

Table 6 shows the results for the estimates of Equations 3 and 7. These data display less robust evidence and demonstrate that only words with a negative tone in the sub-sample show a positive association with the volatility of Ibovespa. This result, not entirely as expected based on international literature (Tetlock, 2007; Tetlock et al., 2008), can be explained by the high volatility of the Brazilian market when compared to the volatility of the North American market.

Table 6. **Relationship between news tone and volatility**

| | Complete editions | | Filtered editions | |
|---|---|---|---|---|
| | Vol Ibov | Vol Ibra | Vol Ibov | Vol Ibra |
| Negative | 0.00000926 | 0.00000441 | 0.0000160** | 0.00000815 |
| | (-1.59) | (-0.92) | (-2.99) | (-1.10) |
| Positive | 0.0000016 | 0.00000124 | -0.00000547 | -0.00000639 |
| | (-0.26) | (-0.24) | (-0.72) | (-0.90) |
| Uncertainty | 0.00000504 | 0.00000892 | -0.0000104 | -0.00000177 |
| | (-0.99) | (-1.63) | (-1.82) | (-0.25) |
| Litigious | 0.00000551 | 0.00000737 | 0.00000116 | 0.00000374 |
| | (-0.62) | (-0.83) | (-0.09) | (-0.28) |
| Modal | -0.0000169 | -0.00000986 | 0.00000021 | -0.00000071 |
| | (-1.28) | (-1.38) | (-0.02) | (-0.08) |
| SMB_vol | -0.958 | -0.948 | -0.958 | -0.945 |
| | (-1.20) | (-1.30) | (-1.21) | (-1.31) |
| HML_vol | 0.418 | 0.766** | 0.422 | 0.774** |
| | (-1.26) | (-3.77) | (-1.28) | (-3.86) |
| WML_vol | 0.384 | 0.182 | 0.381 | 0.177 |
| | (-2.05) | (-1.34) | (-2.05) | (-1.31) |
| IML_vol | 1.783* | 1.416 | 1.783* | 1.414 |
| | (-2.14) | (-1.70) | (-2.16) | (-1.72) |
| Riskfree_vol | 284.0** | 164.5** | 284.2** | 163.2** |
| | (-3.18) | (-2.82) | (-3.18) | (-2.83) |
| Constant | 0.000519 | 0.000306 | 0.000824 | 0.000623 |
| | (-0.20) | (-0.13) | (-0.32) | (-0.27) |
| Fixed year effect | yes | yes | yes | yes |
| N | 1237 | 1237 | 1237 | 1237 |
| R² | 0.496 | 0.436 | 0.494 | 0.434 |

Note: Coefficients in the first row and p-value statistics in parentheses in the second row of each variable (α‹10%*, α‹5%**, α‹1%***).

# FINAL CONSIDERATIONS

The objective of this work was to apply the sentiment analysis technique to daily editions of the *Valor Econômico* newspaper, to investigate the existence of a relationship between profitability and the volatility of Ibovespa and IBrA and the tone of news published by daily, printed media, specializing in Economy and Finance in the Brazilian market.

There is some discrepancy in "negative" terms, between the values found on this study and those in the literature. It was expected that, in line with Tetlock (2007) and Tetlock et al. (2008), a higher number of negative terms would lead to unfavorable effects on profitability (decreasing) and volatility (increasing). This was the case in the Brazilian market, but only for volatility and with weaker evidence. On the other hand, as expected from the evidence in the North American market, the terms classified under "uncertainty" showed a negative association with daily profitability in both Ibovespa and IBrA

From the dictionaries used in this study, words classified under "negative" and "uncertainty" were relevant, in contrast with words with a "positive," "litigious," and "modal" tone. In this sense, a greater number of these terms in the daily edition of *Valor Econômico* could have consequences associated with market profitability and/or volatility. In other words, the evidence indicates that the market accords greater weight to words published by the specialized media that are negative or convey uncertainty.

This study aims to contribute to the research on the impacts of qualitative information from textual analysis within Brazil.

The evidence obtained from this study points to the relevance of the specialized media in Brazil, and to the existence of informative content in published news. The results can motivate capital market participants to use this method, along with others such as machine learning, to predict the behavior of variables in the market (Cambria, 2016; Tripathy et al., 2016). Additionally, investors can benefit from the results of this study, as a relationship is evident between the tones extracted from published news in the specialized media in Brazil and stock profitability and/or volatility during the day of the analysis.

This study has some limitations. Note that it investigates a relationship between weight/tone of words and the market as a whole (using the Bovespa Indices and Brazil Broad-Based Index). Therefore, a single negative news story, for instance, about a company with a great deal of weight on the aforementioned indices could elicit strong movements in profitability and/or volatility. However, if there were a potential market prediction based on these models, the effect would not be felt.

Future studies can attempt to directly establish the relationship between the tone of specific news stories about a company and stock returns and volatility. Moreover, they can adopt other vehicles of financial communication (for instance, Bloomberg, Google Finance, etc.), as well as information from social networks, such as Twitter or Facebook. They may also consider option developing word dictionaries to perform the textual analysis. Finally, the use of machine learning methods to create prediction techniques for market or company indicators is suggested. Therefore, a broad spectrum of research on the topic of this study is evident.

## ACKNOWLEDGEMENTS

## REFERENCES

Amihud, Y. (2002). Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets*, *5*(1), 31-56. doi:10.1016/S1386-4181(01)00024-6

Antweiler, W., & Frank, M. Z. (2004). Is all that talk just noise? The information content of internet stock message boards. *The Journal of Finance, 59*(3), 1259-1294.

Associação Nacional de Jornais. (2017, Junho 10). *Os maiores jornais do Brasil de circulação paga, por ano*. Retrieved from http://www.anj.org.br/maiores-jornais-do-brasil/

Barberis, N., Shleifer, A., & Vishny, R. (1998). *A model of investor sentiment. Journal of Financial Economics, 49*(3), 307-343. doi:10.1016/S0304-405X(98)00027-0

BM&FBovespa. (2015a). Índice Bovespa (Ibovespa). *Composição/Carteira do índice*. Retrieved from http://www.bmfbovespa.com.br/indices/ResumoIndice.aspx?Indice=Ibovespa&Idioma=pt-br

BM&FBovespa. (2015b). Índice Brasil Amplo (IBrA). *Composição/Carteira do índice*. Retrieved from http://www.bmfbovespa.com.br/indices/ResumoCarteiraTeorica.aspx?Indice=IBrA&idioma=pt-brr

BM&FBovespa. (2015c). Índice Brasil Amplo (IBrA). *Estatísticas históricas*. Retrieved from http://www.bmfbovespa.com.br/indices/ResumoEvolucaoDiaria.aspx?Indice=IBrA&idioma=pt-br

Bogle, S. A., & Potter, W. D. (2015). SentAMaL: *A sentiment analysis machine learning stock predictive model*. Proceedings on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing, Las Vegas, USA, WorldComp.

Boussaidi, R. (2013). Representativeness heuristic, investor sentiment and overreaction to accounting earnings: The case of the Tunisian stock market. *Procedia-Social and Behavioral Sciences*, *81*, 9-21. doi:10.1016/j.sbspro.2013.06.380

Bushman, R. M., Williams, C. D., & Wittenberg-Moerman, R. (2016). The informational role of the media in private lending. *Journal of Accounting Research, 55*(1), 115-152. doi:10.1111/1475-679X.12131

Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, *31*(2), 102-107. doi:10.1109/MIS.2016.31

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

Chen, K. T., Lu, H-M., Chen, T-J., Li, S-H., Lian, J-S., & Chen, H. (2011). Giving context to accounting numbers: The role of news coverage. *Decision Support Systems, 50*(4), 673-679. doi:10.1016/j.dss.2010.08.025

Cochrane, J. H., & Culp, C. L. (2003). Equilibrium asset pricing and discount factors: Overview and implications for derivatives valuation and risk management. In P. Field (Ed.), *The Growth of Risk Management: A History* (pp. 57-92). London, UK: Risk Books.

Davolos, L. C., Rogers, P., Silva, W. M., & Oliveira, M. A. (2013). O que determina o preço das ações? Exame empírico do mercado brasileiro pré-subprime (1994-2007). *REA-Revista Eletrônica de Administração, 12*(1), 48-67.

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

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The Journal of Finance*, *64*(5), 2023-2052. doi:10.1111/j.1540-6261.2009.01493.x

Instituto de Pesquisa Econômica Aplicada. (2015). *Índice de ações Ibovespa – Fechamento*. Retrieved from http://www.ipeadata.gov.br/

Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In C. Aggarwal & C. Zhai (Eds.), *Mining text data* (pp. 415-463). Boston, USA: Springer.

Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, *66*(1), 35-65. doi:10.1111/j.1540-6261.2010.01625.x

Mitra, G., & Mitra, L. (Eds.). (2011). *The handbook of news analytics in finance*. Hoboken, USA: John Wiley & Sons.

Nascimento, P., Osiek, BA, & Xexéo, G. (2015). Análise de sentimento de *Tweets* com foco em notícias. *Revista Eletrônica de Sistemas de Informação*, *14*(2), 1-14. doi:10.21529/RESI.2015.1402002

Pagliarussi, M. S., Aguiar, M. O., & Galdi, F. C. (2016). Sentiment analysis in annual reports from Brazilian companies listed at the BM&FBovespa. *BASE-Revista de Administração e Contabilidade da Unisinos, 13*(1), 53-64.

Porshnev, A., Redkin, I., & Shevchenko, A. (2013). *Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis*. 13th IEEE International Conference on Data Mining Workshops. Washington, USA: IEEE.

Rogers, J. L., Skinner, D. J., & Zechman, S. L. (2015). The role of the media in disseminating insider-trading activity (Working Paper, No. 13-34). University of Colorado, Boulder, USA.

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. doi:10.1111/j.1540-6261.2007.01232.x

Tetlock, P. C., Saar-Tsechansky, M., & Mackskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance, 63*(3), 1437-1467. doi:10.1111/j.1540-6261.2008.01362.x

Tripathy, A., Agrawal, A., & Rath, SK (2016). Classification of sentiment reviews using n-gram machine learning approach. *Expert Systems with Applications*, *57*, 117-126. doi:10.1016/j.eswa.2016.03.028

Valor Econômico. (2012, Maio 15). *Edição impressa*. Retrieved from http://www.valor.com.br/impresso/

Valor Econômico. (2013, Maio 15). *Edição impressa*. Retrieved from http://www.valor.com.br/impresso/