

# ARTIGOS

Submetido 19.09.2016. Aprovado 14.08.2017

Avaliado pelo sistema *double blind review*. Editor Científico Convidado: Wesley Mendes-da-Silva

Versão original

DOI: <http://dx.doi.org/10.1590/S0034-759020180203>

## PESSIMISMO E INCERTEZA DAS NOTÍCIAS E O COMPORTAMENTO DOS INVESTIDORES NO BRASIL

*Pessimism and uncertainty of the news and investor's behavior in Brazil*

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

### 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.

### 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 indices. 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 the specialized media in Brazil, especially with the words “negative” and “uncertainty.”*

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

### 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.

**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, Brasil

## INTRODUÇÃO

A mídia especializada é uma importante fonte de informações sobre as empresas, especialmente para aquelas que utilizam o mercado de capitais, em que a regulação exige que as informações devem ser divulgadas aos investidores de maneira equitativa. A associação entre as notícias divulgadas (textos jornalísticos, opiniões de *blogs* financeiros, postagem em mídias sociais, rumores etc.) e o comportamento do mercado tornou-se objeto de estudo de vários pesquisadores, podendo-se citar: Antwiler e Frank (2004), que relacionam a opiniões de *blogs* financeiros ao retorno de determinadas ações; Tetlock (2007), Tetlock, Saar-Tsechansky e Mackskassy (2008), Fang e Peress (2009) e Chen et al. (2011), que trabalham com a mídia jornalística e a sua relação na rentabilidade de certas empresas; Porshnev, Redkin e Shevchenko (2013) e Bogle e Potter (2015), que discutem a possibilidade de se prever o mercado a partir do tom de posts do Twitter; Rogers, Skinner e Zechman (2015), que avaliam se a maneira como as notícias são veiculadas pela mídia afeta a resposta dos preços dos ativos; e Bushman, Williams e Wittenberg-Moerman (2016), que investigam se a cobertura da mídia sobre um tomador de recursos influencia empréstimos sindicalizados.

Nesse contexto, este trabalho investiga se é possível observar alguma associação (positiva ou negativa) entre a rentabilidade e volatilidade dos índices de mercado (Ibovespa e IBrA) com o teor positivo ou negativo (doravante denominado “tom”) extraído das notícias divulgadas pela principal mídia especializada em assuntos econômicos do Brasil.

A base de dados utilizada para a elaboração das variáveis primárias desta pesquisa consiste na edição diária disponível no site do *Valor Econômico*, maior jornal especializado no Brasil. Para esta pesquisa, considera-se a utilização apenas desse veículo de comunicação, pois desde 2015, com o fim da circulação da versão impressa do jornal *Brasil Econômico*, o jornal *Valor Econômico* é *player* único desse mercado e a principal fonte de geração diária de informações sobre economia, finanças e mercados no Brasil. De acordo com dados da Associação Nacional de Jornais (ANJ, 2017), em 2015, a tiragem média diária do *Valor Econômico* foi de 41.431 exemplares. Optou-se por uma amostra contendo todos os editoriais e uma subamostra, denominada amostra filtrada, considerando apenas os seguintes cadernos: Brasil, Política, Internacional e Finanças. Ambas foram extraídas dentro de um período de cinco anos completos (de 2012 a 2016). Para processar a base descrita, utilizou-se o método computacional denominado análise de sentimento (ou *opinion mining*), que tem por objetivo extrair a opinião de textos (Liu & Zhang, 2012).

A utilização da análise de sentimento foi feita por meio de um algoritmo que, junto a dicionários de palavras, processa as suas edições diárias, transformando essas informações textuais em dados quantitativos, possibilitando, assim, uma avaliação quantitativa do tom de uma notícia e a realização de análises estatísticas. Tanto o algoritmo quanto os dicionários foram baseados nos equivalentes desenvolvidos por Pagliarussi, Aguiar e Galdi (2016).

Os resultados encontrados apresentam indícios de que uma maior quantidade de termos com conotação de incerteza na edição diária do jornal *Valor Econômico* tem uma associação negativa com o retorno do dia da veiculação da edição (impressa e *on-line*). Adicionalmente, encontram-se evidências mais fracas sobre uma relação positiva entre notícias com tom negativo e aumento da volatilidade do Ibovespa.

Entende-se que este trabalho contribuiu para a literatura de Finanças e Contabilidade do País, pois trata de um assunto amplamente pesquisado em âmbito internacional e em ascensão no Brasil. Além disso, a utilização da análise de sentimento em conjunto com algoritmos de, por exemplo, aprendizado de máquina (*machine learning*) pode vir a auxiliar o investidor e/ou reguladores a prever algum comportamento do mercado (Cambria, 2016; Tripathy, Agrawal, & Rath, 2016).

## REFERENCIAL TEÓRICO

### Notícias e sua influência no mercado

A teoria diz que valor de uma empresa deverá ser igual ao valor presente de seus fluxos de caixa esperados considerando-se o adequado custo de capital (Cochrane & Culp, 2003). A projeção desses fluxos de caixa é condicional a outros conjuntos de informações, como: descrição qualitativa do ambiente de negócio das empresas, suas operações e as perspectivas apresentadas pela imprensa financeira (Tetlock et al., 2008). A literatura apresenta diversas evidências de que a mídia financeira especializada divulga informações relevantes para os participantes do mercado de capitais e do mercado de crédito, adicionais às informações providas pelos analistas de mercado e pelas demonstrações contábeis (Bushman et al., 2016; Tetlock et al., 2008).

Novas informações relevantes sobre uma determinada empresa, setor ou economia, em geral, podem alterar a ótica do mercado em relação ao risco implícito e à rentabilidade financeira prevista. Consequentemente, o valor dessas empresas poderá ser reequilibrado pelo mercado em função da nova expectativa de retorno financeiro delas (Tetlock et al., 2008).

Como a maior parte dos investidores e participantes do mercado tem acesso aos meios de comunicação, dependendo do tom (teor) das informações coletadas (por exemplo: notícias positivas ou negativas sobre determinada empresa, setor ou economia), eles poderão prever alterações no fluxo de caixa projetado. Isso teria como resultado a valorização ou desvalorização da empresa (ação) e, conseqüentemente, a troca de posições acionárias de companhias comprometidas por outras com desempenho mais robusto (Mitra & Mitra, 2011). Assim, considerando-se o mercado de maneira agregada, espera-se que o tom das notícias em um determinado dia tenha associação com o desempenho e a volatilidade dos índices de mercado.

## Análise de sentimento

A avaliação da influência do sentimento dos investidores em suas decisões remonta a estudos iniciais na década de 1980, quando economistas começaram a utilizar ferramentas da Psicologia para explicar o comportamento dos investidores (Boussaidi, 2013). Nessa linha, Barberis, Shleifer e Vishny (1998) foram os primeiros a modelar o sentimento dos investidores de maneira a mostrar como são formadas suas crenças baseados em evidências da Psicologia e de reações exageradas (*overreaction* ou *underreaction*). Esse tipo de pesquisa ficou caracterizada dentro da linha denominada Finanças Comportamentais.

Nessa linha, a análise de sentimento utilizada neste artigo se diferencia, pois pode ser definida como o estudo computacional de opiniões, avaliações, atitudes e emoções dirigidas a entidades, indivíduos, edições, eventos, temas e seus atributos (Liu & Zhang, 2012).

Para um ser humano comum, acompanhar e fazer a leitura das várias notícias disponíveis nos veículos de informação (ex.: mídia especializada, *blogs*, fóruns, redes sociais etc.), além de ser uma tarefa árdua, devido à quantidade de informações publicadas, nem sempre significa que essas notícias serão facilmente decodificadas (compreendidas) por parte do leitor, o que traz dificuldades para o seu processo de tomada de decisão (Liu & Zhang, 2012).

Além disso, sabe-se que uma pessoa normalmente tende a prestar mais atenção em informações e opiniões que vão ao encontro de suas próprias preferências (Liu & Zhang, 2012).

Assim, a grande vantagem de se utilizarem métodos computacionais é tratar grandes volumes de texto rapidamente, produzindo resultados consistentes e mitigando o efeito do viés por opiniões ou predileções individuais (Liu & Zhang, 2012).

## Estudos anteriores

Antwiler e Frank (2004) afirmam que mensagens em fóruns a respeito do mercado financeiro influenciam o seu comportamento. Estudando o efeito de mais de 1,5 milhão de mensagens postadas no *Yahoo! Finance* e no *Raging Bull* sobre 45 companhias do *Dow Jones Industrial Average* e do *Dow Jones Internet Index*, eles verificam que o tom delas ajuda a prever a volatilidade do mercado. Seus resultados mostram que o impacto das mensagens sobre os retornos da bolsa é estatisticamente relevante, apesar de ser economicamente pequeno.

Tetlock (2007) avaliou a interação entre a mídia e o mercado de ações medindo a entonação da coluna “Abreast of the market” do *Wall Street Journal* no período de 1984 a 1999. Essa coluna discute as razões do comportamento do mercado no dia anterior e também contém previsões de analistas sobre o futuro. O autor encontra evidências de que níveis altos de pessimismo preveem efeito negativo sobre os preços das ações e que níveis anormais, altos ou baixos, de pessimismo antecipam alto volume de negociações. Ele também sugere que baixas rentabilidades do mercado financeiro levam ao aumento de notícias pessimistas.

Tetlock et al. (2008) estendem a análise feita pelo autor acima, medindo agora não somente o tom de uma coluna, mas sim de todo o *Wall Street Journal* e do *Dow Jones News Service* no período de 1980 a 2004. Eles acompanham os retornos das ações e investigam se um maior número de palavras negativas pode ser usado para aumentar a expectativa de fluxos de caixa futuros.

Os resultados encontrados indicam que a maior quantidade de palavras negativas em notícias específicas sobre a firma prevê baixo resultado financeiro, principalmente se estas estiverem relacionadas com a estrutura financeira da empresa (Tetlock et al., 2008).

Fang e Peress (2009) partem da hipótese de que a mídia influencia a rentabilidade das ações, mesmo quando ela apresenta informações incoerentes ou exageradas. Eles medem a relação entre a cobertura da mídia e os retornos das ações, e encontram evidências de que ações de empresas menos citadas pelos veículos de comunicação tendem a ter retornos superiores que suas contrapartes.

Chen et al. (2011), de maneira semelhante à pesquisa de Fang e Peress (2009), utilizam a hipótese de que a mídia pode trazer novas informações ao mercado. Eles observam a cobertura do *Wall Street Journal* sobre as empresas listadas no *S&P 500* antes da divulgação dos relatórios financeiros, e como as ações se comportam em relação a sua rentabilidade. Em seus resultados, eles argumentam que a maior cobertura jornalística sobre as

empresas diminui a chance de as ações obterem ganhos anormais, o que leva a um menor coeficiente de resposta ao lucro (*Earnings Responce Coefficient* – ERC).

Loughan e McDonald (2011) apresentam uma nova metodologia para análise de textos. Eles argumentam que a utilização de dicionários produzidos por outras disciplinas, como o da Psicologia, classifica erroneamente o tom de textos financeiros. Os autores desenvolvem uma nova lista de palavras (Fin-Neg, em inglês), a partir de textos da área de Finanças, e concluem que aproximadamente três quartos das palavras categorizadas como negativas pela lista de Harvard (*Harvard Psychosocial Dictionary*) não receberam a mesma classificação pela nova lista de palavras.

Os pesquisadores acima também propõem a utilização de uma equação matemática que, em vez de apenas contar a frequência das palavras nos textos, mede o seu peso (*term weighting:  $w_{i,j}$* ) conforme Equação 1 (Loughran & McDonald, 2011; Pagliarussi et al., 2016).

$$w_{i,j} = \begin{cases} \frac{(1+\log(tf_{i,j}))}{(1+\log(a_j))} \log \frac{N}{df_i} \text{ se } tf_{i,j} \geq 1 \\ 0 \quad \text{Caso contrario} \end{cases} \quad (1)$$

Em que:

$tf_{i,j}$  Total de ocorrências de uma palavra  $i$  em um documento  $j$ ;

$a_j$  Proporção de palavras contadas em um documento  $j$ ;

$N$  Total de documentos na amostra;

$df_i$  Total de documentos com ao menos uma ocorrência da palavra  $i$ .

O argumento utilizado é que palavras muito frequentes no texto não são necessariamente mais informativas que outras com menos ocorrência. O logaritmo presente na Equação 1 tem justamente a função de reduzir a significância desses termos (Loughran & McDonald, 2011).

Pagliarussi et al. (2016) utilizam a análise de sentimento para extrair a opinião dos relatórios de Administração de determinadas empresas brasileiras no período de 1997 a 2009 e relacionar com o retorno anormal, o volume anormal de negócios e a volatilidade do preço de ações das firmas. Em seus resultados, os autores não foram capazes de encontrar evidências de que os relatórios da Administração tenham influência sobre os negócios realizados no mercado de ações.

Os autores também desenvolveram um algoritmo para análise dos textos empregando a fórmula de Loughan e McDonald (2011) e produziram dicionários de palavras em português que acreditam poderem ser utilizados para qualquer texto da área de Finanças.

Os dicionários construídos por Pagliarussi et al. (2016) são amplos e consideraram alguns pontos relevantes na adaptação para a língua portuguesa. Especificamente, Pagliarussi et al. (2016) comentam este ponto:

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 usual-

ly 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", "legítimar", "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)

Ainda no Brasil, Nascimento, Osiek e Xexéo (2015) utilizaram a análise de sentimento para investigar a reação da população em relação às notícias divulgadas na mídia de maneira a capturar as reações dos comentários realizados pelas pessoas na rede social Twitter.

## METODOLOGIA

### Coleta e tratamento dos dados

Para realização da coleta e do tratamento dos dados, três programas computacionais foram utilizados. Os dois primeiros, desenvolvidos em Java, fazem o *download* das edições diárias do jornal *Valor Econômico* (apenas a sua parte gratuita) a partir do seu *site*. O terceiro foi desenvolvido em Python por Pagliarussi et al. (2016) e tem por objetivo fazer a análise de sentimento dos arquivos que foram produzidos pelos dois primeiros.

O período de amostragem se estendeu entre 2/1/2012 e 30/12/2016, em razão da disponibilidade dos jornais pelo *site* do jornal *Valor Econômico*. Os dias em que não houve a circulação do jornal e em que a BM&FBovespa não funcionou

foram excluídos da base de dados. Consequentemente, a amostra conta com 1.237 edições diárias do jornal, divulgadas em mídia impressa e eletrônica, de onde foram extraídos os tons das notícias divulgadas.

As seções dos editoriais do jornal e que foram consideradas na amostra completa da base de notícias são: Brasil, Política, Finanças, Empresas, Agronegócios, Internacional, Opinião, Legislação, Carreira, Cultura e Estilo. Foi considerada, também, uma subamostra denominada de editorial filtrado, de onde foram retirados os cadernos cujas notícias não possuem relação de primeira ordem com o desempenho do mercado de capitais brasileiro de maneira geral. Assim, a subamostra com o editorial filtrado foi formada considerando os seguintes cadernos: Brasil, Política, Internacional e Finanças.

Cada edição é salva em um arquivo em formato ".txt" nomeado pelo ano, mês e dia de sua divulgação (exs.: "20120307.txt" para editoriais completos e "20120307-BPIF.txt" para editoriais filtrados). Optou-se por essa forma porque a organização dos dados ficaria otimizada.

Ainda sobre os dois primeiros algoritmos, os Quadros 1 e 2 mostram um exemplo de sua funcionalidade. A notícia a seguir foi extraída da edição do jornal *Valor Econômico* do dia 15 de maio de 2013.

#### Quadro 1. Primeira notícia da edição de 15/5/2013 (todos os editoriais)

##### Leilão atrai investimentos de R\$ 7 bilhões em petróleo

Os investimentos previstos empresas que arremataram ontem áreas para exploração de petróleo e gás são estimados em R\$ 7 bilhões. A 11ª. Rodada de licitações da ANP, prevista para terminar hoje, foi concluída em um único dia e mereceu forte comemoração do governo.

Fonte: Valor Econômico (2013).

No arquivo gerado ("20130515.txt"), apresentado pela próxima imagem, o texto do Quadro 1 se transforma em:

#### Quadro 2. Parte do texto no arquivo "20130515.txt"

##### Leilao atrai investimentos de R\$ 7 bilhoes em petroleo

Os investimentos previstos pelas empresas que arremataram ontem areas para exploracao de petroleo e gas sao estimados em R\$ 7 bilhoes. A 11 Rodada de licitacoes da ANP, prevista para terminar hoje, foi concluida em um unico dia e mereceu forte comemoracao do governo.

Fonte: Valor Econômico (2013).

Comparando ambas as imagens, percebe-se no Quadro 2 que todos os caracteres com acentos ou especiais foram substituídos pelas suas contrapartes sem acento ou removidos. Por exemplo: “ã” foi trocado por “a”, “ê” por “e”, ç por c e “ã” por “” (nenhum caractere) e assim por diante.

Essas substituições tornaram-se necessária por motivos de compatibilidade com o compilador (Python) utilizado. Observou-se que, quando esses caracteres estavam presentes, o algoritmo não identificava as palavras que os continham e estas ficavam ausentes do processo de análise de sentimento, gerando valores incorretos para as variáveis primárias. O mesmo erro também foi relatado por Pagliarussi et al. (2016).

As variáveis de controle utilizadas são os três fatores de Fama e French (1993): SMB (*Small Minus Big*) e HML (*High Minus Low*), excluindo-se o *Market Factor* ( $R_m - R_f$ ) risco de mercado subtraído de uma taxa de juros livre de riscos), em razão de a variável dependente ser diretamente relacionada ao risco de mercado. Foram incluídos também os fatores de Carhart (1997) e Amihud (2002): WML (*Winners Minus Losers*) e IML (*Illiquid Minus Liquid*) respectivamente.

Para as variáveis dependentes, escolheu-se fazer o uso dos índices Bovespa e do IBrA. Essa escolha deve-se ao fato de o primeiro ser o mais utilizado no Brasil e de o segundo ser o que contém maior número de papéis.

A fim de exemplificar, a Tabela 1 apresenta uma breve comparação entre esses dois índices:

**Tabela 1. Comparação entre Ibovespa e IBrA**

	Ibovespa		IBrA	
Número de papéis	59		115	
Soma dos 10 maiores papéis	55,14%		49,37%	
Soma dos 20 maiores papéis	75,04%		67,83%	
Maior participação	ITUB4	11,29%	ITUB4	10,02%
Segunda maior participação	BBDC4	7,99%	BBDC4	7,09%
Terceira maior participação	ABEV3	7,31%	ABEV3	6,49%

Fonte: Adaptado de BM&FBovespa Ibovespa (2015a); BM&FBovespa IBrA (2015b, 2015c).

## Operacionalização da equação de Loughan e McDonald (2011)

De maneira a se tornar mais simples a visualização da construção das variáveis de tom, apresentamos nesta seção uma operacionalização dos cálculos.

No exemplo abaixo, foi utilizado um dicionário de palavras negativas contendo apenas os termos “mensalao” e “prejuizo” (o motivo pelo qual ambas as palavras estão sem acentos é que, quando estes são utilizados, o Python na versão 2.7 não consegue identificá-los, e as palavras acabavam ficando fora da análise. Este procedimento foi utilizado em todas análises).

Os Quadros 3, 4 e 5 são respectivamente parte das edições de 11, 12 e 13 de junho de 2012. Foi utilizado apenas um pedaço desses jornais, pois, caso contrário, os quadros ficariam muito extensos sem afetar a análise matemática.

### Quadro 3. Parte do arquivo 20120611.txt

Brasil Congresso Candidatos usam CPI contra adversarios nas eleicoes municipais Coincidencia de calendarios estimulou politicos a usarem investigacao para tentar fragilizar grupos politicos opostos Receita para um mau prefeito Podemos descartar os maus candidatos a prefeito vendo se discutem os temas essenciais para a cidade ou se fogem deles Disputa de 2014 motiva Alckmin e Kassab a buscarem aliados para Serra Alianca atual e circunstancial e nao descarta confronto dos dois nas proximas eleicoes PT escolhe nome para vice de Lacerda Tucano lanca candidatura a Prefeitura do Rio Internacional Energia EUA compram menos petroleo no exterior Pais reduz dependencia de petroleo importado, gracias ao aumento da producao domestica, aos motores mais eficientes e ao uso de etanol Curtas Dados dissipam pessimismo sobre a China Numeros lancados no fim de semana indicam uma economia mais forte do que muitos no mercado temiam uma semana atras Dolar paralelo sai de controle na Venezuela Bolivar sofre quedas recordes no mercado paralelo na Venezuela, em meio a corrida de pessoas fisicas pela escassa moeda americana Marcha para o anti-Chavez Especial Conjuntura Producao da industria sofre, mas confianca melhora em setores beneficiados Producao de confeccao e calçados ainda patina, mas empresarios relatam melhora da confianca depois da desoneracao da folha de pagamentos Pais precisa estimular vigor do setor privado, diz Araujo valor.com.br A arrogancia campeia tanto la quanto ca Com um pouco de tolerancia, e possivel olhar para os vizinhos do sul sem visoes preconcebidas ou vieses conservadores Vacinacao antigripe atrasa e especialistas ja discutem a antecipacao de campanhas Entrega da vacina da gripe atrasa e governos e especialistas cobram antecipacao das proximas campanhas de vacinacao Infraestrutura Investimento publico em transportes em 2012 nao esta saindo do papel

Fonte: Valor Econômico (2012).

Nota: Extrato de texto igual ao analisado pelo software. Intencionalmente sem acentuação.

**Quadro 4. Parte do arquivo 20120612.txt**

Alguns trocam qualquer presente por um passeio sobre rodas ao lado da cara metade. No Dia dos Namorados, contamos historias com final feliz e escondemos as que nao sao publicaveis. Brasil Congresso Ruralistas defendem Mendes Ribeiro Ministro havia sido desautorizado pela presidente em negociacoes em torno do Codigo Florestal Curtas Dirceu a espera do armagedon 4 As vespuras do julgamento dos reus do mensalao, ex-ministro convoca estudantes para a "batalha final" Comissao do pacto federativo mantem exigencia de unanimidade no Confaz Grupo de 14 especialistas convocado pelo presidente do Senado, Jose Sarney, tambem e favoravel a adocao do IPCA para a indexacao das dividas estaduais valor.com.br Comissao investigara uso de aviao por Pimentel Colegiado decidiu aprofundar as investigacoes a respeito das consultorias prestadas pelo ministro Eleicoes Freixo lanca-se com apelo por 'primavera carioca' 1 Candidato do PSOL tem o apoio de artistas e compositores, como Caetano Veloso e Chico Buarque Comissao da Verdade avaliza convocacao de agentes da ditadura Comissoes da Verdade regionais terao aval do orgao nacional para convocar supostos torturadores Em jogo estao informacoes que podem ser usadas para a venda de publicidade em dispositivos moveis Aviacao Acao da TAM cai 2,09% na vespera da criacao da Latam Companhia aerea brasileira fecha capital hoje e a chilena Lan vai negociar BDRs na BM&FBovespa Europeias devem ter prejuizo de US\$ 1,1 bilhao Companhias aereas oferecem iPads aos passageiros para reduzir custos Calcados Arezzo comeca expandir marca mais cara Projeto de franquias da Schutz permite margem bruta mais folgada a lojas de maior faturamento Saude Sabara deve replicar modelo da AACD Eduardo Carneiro, do hospital Sabara, calcula que em 2014 podera atingir o equilibrio Alimentos Vapza reduz pacote e atrai mulheres

Fonte: Valor Econômico (2012).  
 Nota: Extrato de texto igual ao analisado pelo software. Intencionalmente sem acentuação.

**Quadro 5. Parte do arquivo 20120613.txt**

Punicoes no mensalao ainda devem demorar Nenhum dos 38 reus no processo do mensalao saira preso do julgamento no Supremo Tribunal Federal, marcado para comecar em 1 de agosto e que nao deve terminar antes de setembro. Brasil Governo CGU proibe Delta de fechar novos contratos Inidoneidade da empresa, decidida pela CGU, estende-se a Estados e municipios A encruzilhada historica do mensalao 6 Julgamento decidira nao apenas o futuro dos 38 acusados, mas tambem a rota ou caminho que o Brasil ira tomar Depoimento de Perillo deve levar CPI a recuar da quebra de seu sigilo PT e PSDB negam acordo de nao agressao em relacao aos seus governadores Cachoeira conseguiu um voto pelo habeas corpus Camara aprova MP que simplifica licitacoes do PAC Uma nova forma de uniao europeia Se parece improvavel que as politicas atuais funcionem, sugiro uma combinacao de duas ideias: uniao baseada em seguro e uniao baseada em ajustes. Judiciario Recursos podem levar mensalao para 2013 Num processo marcado por indefinicoes, a unica certeza e que ninguem saira preso do Supremo Ayres Britto nega pressao sobre STF Empresas Petroleo Comgas sera cliente da BG no pre-sal A Comgas podera comprar a parte da britanica do gas produzido no campo de Lula Destaque Movimento falimentar Contabilidade Brasil manda mais recursos para o IFRS Doacao aumenta 16%, mas contribuicao do pais para fundacao que rege contabilidade internacional ainda e pequena Oferta de acoes Investidor do Facebook processa bolsa

Fonte: Valor Econômico (2012).  
 Nota: Extrato de texto igual ao analisado pelo software. Intencionalmente sem acentuação.

Utilizando a fórmula proposta por Loughan e McDonald (2011) apresentada na Equação 1 para fazer a análise de sentimento dos textos contidos nos quadros acima, tem-se:

No Quadro 1, não foi encontrada nenhuma das duas palavras, logo o resultado encontrado é obtido conforme as equações abaixo:

$w_{prejuizo, 20120611} = 0$ (A)	(A)
$w_{mensalao, 20120611} = 0$ (B)	(B)

Já no Quadro 2, ambas as palavras aparecem uma única vez no texto; resolvendo a equação para esse caso, tem-se: Para o cálculo da palavra “prejuízo”, sendo os valores:

$tf_{prejuizo, 20120612} = 1$  (Só existe uma ocorrência no documento 20120612);

$N=3$  (Total de documentos, três jornais);

$df_{prejuizo} = 1$  (Só ocorre em um documento).

Calculando  $a_{20120612}$ :

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

Onde:

$nc$  = Soma de ocorrências da palavras “prejuízo” e “mensalão”;

$nt$  = Soma de apenas uma presença de cada palavra do dicionário dentro do documento analisado.

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

Melhorando a explicação do cálculo de  $a_j$ , no Quadro 2, ambas as palavras, “mensalão” e “prejuízo”, apareceram uma vez, sendo a equação D o resultado dessa análise. Se, por exemplo, “mensalão” tivesse ocorrido duas vezes, enquanto “prejuízo” apenas uma, então o novo valor de  $a_{20120612}$  é  $a_{20120612} = 1,5$  pois  $nc = 3$  enquanto  $nt = 2$ . Para o quadro C.3, é possível encontrar a palavra “mensalão” quatro vezes, enquanto “prejuízo”, nenhuma. Dessa forma, o resultado de  $a_{20120612}$  é:

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

Substituindo os valores na equação (1):

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

$w_{prejuizo, 20120612} = \ln 3$	(G)
----------------------------------	-----

$w_{prejuizo, 20120612} = 1,098612289$	(H)
--	-----

Para o cálculo da palavra “mensalão”, sendo os valores:

$tf_{mensalao, 20120612} = 1$  (Só existe uma ocorrência no documento 20120612);

$N = 3$  (Total de documentos, sendo três jornais);

$df_{mensalao, 20120612} = 2$  (Ocorre em dois documentos, 20120612 e 20120613).

$a_{mensalao, 20120612} = 1$  (Resultado da equação D)

Novamente substituindo os valores na equação (1):

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

$w_{mensalao, 20120612} = \ln 1,5$	(J)
------------------------------------	-----

$w_{mensalao, 20120612} = \ln 0,405465$	(K)
---	-----

Somando os dois valores, tem-se:

$w_{prejuizo, 20120612} + w_{mensalao, 20120612} = \ln 1,504077$	(L)
--	-----

O valor encontrado pela equação K é o peso das palavras (para o dicionário de palavras negativas utilizado neste exemplo) para o dia 12 de junho de 2012.

Por fim, no Quadro 3, a palavra “prejuízo” não ocorre, sendo o seu resultado  $w_{prejuizo, 20120612} = 0$ . Já “mensalão” aparece quatro vezes. O cálculo neste caso é:

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

Substituindo esses valores na equação (1):

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

$w_{mensalao, 20120612} = \ln 1,5$	(N)
------------------------------------	-----

$w_{mensalao, 20120612} = 0,405465$	(O)
-------------------------------------	-----

De modo que o valor encontrado na equação M é o valor do peso das palavras negativas para o dia 13 de junho de 2013.

Para calcular o peso de cada palavra para a amostra apresentada nos Quadros 1, 2 e 3, deve-se somar todos os  $w_{ij}$  para “prejuízo” e para “mensalão”. As equações a P e Q simplificam:

$w_{prejuizo, 20120611} + w_{prejuizo, 20120612} + w_{prejuizo, 20120613} = 1,504077$	(P)
$w_{mensalao, 20120611} + w_{mensalao, 20120612} + w_{mensalao, 20120613} = 1,81093$	(Q)

### Modelos econométricos

As variáveis dependentes são: ibov, ibra, ibov\_vol e ibra\_vol. Respectivamente, estas representam a rentabilidade diária (equações 4 e 12) do Ibovespa e do IBRA e as suas volatilidades (equações 6 e 13).

As variáveis independentes são: o peso das palavras (*term weighting*) negativas, positivas, litigiosas, incerteza, modais (variáveis primárias), SMB, HML, WML, IML e riskfree (variáveis de controle). Além destas, SMB\_vol, HML\_vol, WML\_vol, IML\_vol e riskfree\_vol, que são a volatilidade respectiva das variáveis de controle, foram adicionadas quando as variáveis dependentes foram a volatilidade de ambos os índices.

As Equações 2 e 3 representam os modelos econométricos em que as variáveis dependentes estão relacionadas ao Índice Bovespa.

$ibov = \beta_0 + \beta_1 \text{negativas} + \beta_2 \text{positivas} + \beta_3 \text{litigiosas} + \beta_4 \text{incerteza} + \beta_5 \text{modais} + \beta_6 \text{SMB} + \beta_7 \text{HML} + \beta_8 \text{WML} + \beta_9 \text{IML} + \beta_{10} \text{riskfree} + u$	(2)
$ibov\_vol = \beta_0 + \beta_1 \text{negativas} + \beta_2 \text{positivas} + \beta_3 \text{litigiosas} + \beta_4 \text{incerteza} + \beta_5 \text{modais} + \beta_6 \text{SMB} + \beta_7 \text{HML} + \beta_8 \text{WML} + \beta_9 \text{IML} + \beta_{10} \text{riskfree} + \beta_{11} \text{SMB\_vol} + \beta_{12} \text{HML\_vol} + \beta_{13} \text{WML\_vol} + \beta_{14} \text{IML\_vol} + \beta_{15} \text{riskfree\_vol} + u$	(3)

Em que:

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

Em que:

$t$  = Uma data (dia útil) que vai de 2/1/12 até 30/12/14.

$B_t$  = Pontuação de fechamento Índice Bovespa para um certo dia  $t$  (ex.:  $B_{01/08/13} = 49.140$ ).

$B_{t-1}$  = Pontuação de fechamento Índice Bovespa no dia  $t$  anterior (ex.:  $B_{31/07/13} = 48.234$ ).

Já para a volatilidade ( $ibov\_vol_t$ ), a formula é:

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

Em que:

$n$  = Números de dias (ex.:  $n = 60$ , valor utilizado neste trabalho).

$i$  = Certa data que começa em  $t$  e vai até  $t - 59$  (ex.:  $t = 1/8/13$ ,  $t - 1 = 31/7/13$ ,  $t - 2 = 30/07/13$ , ...,  $t - 59 = 8/5/13$ ).

$x_i$  = Valor de  $ibov_t$  em uma determinada data (ex.:  $ibov_{01/08/2013} = 1,86\%$ ).

$\bar{x}$  = Média dos valores de  $ibov_t$  dentro de um certo período  $t$  até  $t - 59$  (60 dias úteis).

A metodologia utilizada para calcular as variáveis SMB\_vol, HML\_vol, WML\_vol, IML\_vol e riskfree\_vol foi a mesma aplicada para se obter o resultado de  $ibov\_vol$ , alterando, evidentemente, os valores de  $ibov$  em  $x_i$  e  $\bar{x}$  pelos respectivos de SMB, HML, WML, IML e riskfree.

De maneira similar às Equações 2 e 3, as Equações 6 e 7 têm a variável dependente agora relacionada ao Índice Brasil Amplo (IBrA).

$$ibra = \beta_0 + \beta_1 \text{negativas} + \beta_2 \text{positivas} + \beta_3 \text{litigiosas} + \beta_4 \text{incerteza} + \beta_5 \text{modais} + \beta_6 \text{SMB} + \beta_7 \text{HML} + \beta_8 \text{WML} + \beta_9 \text{IML} + \beta_{10} \text{riskfree} + u \quad (6)$$

$$ibra\_vol = \beta_0 + \beta_1 \text{negativas} + \beta_2 \text{positivas} + \beta_3 \text{litigiosas} + \beta_4 \text{incerteza} + \beta_5 \text{modais} + \beta_6 \text{SMB} + \beta_7 \text{HML} + \beta_8 \text{WML} + \beta_9 \text{IML} + \beta_{10} \text{riskfree} + \beta_{11} \text{SMB\_vol} + \beta_{12} \text{HML\_vol} + \beta_{13} \text{WML\_vol} + \beta_{14} \text{IML\_vol} + \beta_{15} \text{riskfree\_vol} + u \quad (7)$$

Onde:

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

Em que:

$t$  = Uma data (dia útil) que vai de 2/1/12 até 30/12/14.

$A_t$  = Pontuação de fechamento Índice Brasil Amplo para um certo dia  $t$ .

$A_{t-1}$  = Pontuação de fechamento Índice Brasil Amplo no dia  $t$  anterior ( $t - 1$ )

Já para a volatilidade ( $ibra\_vol_t$ ), a fórmula é:

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

Em que:

$n$  = Números de dias (ex.:  $n = 60$ , valor utilizado neste trabalho).

$i$  = Certa data que começa em  $t$  e vai até  $t - 59$  (ex.:  $t = 1/8/13$ ,  $t - 1 = 31/7/13$ ,  $t - 2 = 30/7/13$ , ...,  $t - 59 = 8/5/13$ ).

$x_i$  = Valor de  $ibra_t$  em uma determinada data (ex.:  $ibra_{01/08/2013} = 1,84\%$ ).

$\bar{x}$  = Média dos valores de  $ibra_t$  dentro de um certo período  $t$  até  $t - 59$  (60 dias uteis).

Os valores de fechamento diários do Ibovespa (utilizados em  $B_t$  e  $B_{t-1}$ ) foram extraídos da base de dados do Instituto de Pesquisa Econômica Aplicada (IpeaData, 2015), enquanto os do IBrA (referentes a  $A_t$  e  $A_{t-1}$ ) foram retirados da base da BM&FBovespa.

As variáveis SMB, HML, WML, IML e riskfree foram obtidas no site do Núcleo de Pesquisa em Economia Financeira (Nefin), vinculado ao Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo.

Em todos os modelos econométricos, o método aplicado para estimar os parâmetros foi o dos mínimos quadrados ordinários (MQO) com utilização de efeitos fixos por ano e erros padrões robustos clusterizados.

## Hipóteses e comportamento esperado

Das equações econométricas apresentadas no tópico anterior (rentabilidade: Equações 2 e 6; e volatilidade: Equações 3 e 7), as variáveis testadas são o peso das palavras negativas; positivas; litigiosas; incerteza e modais (variáveis primárias). Dessa forma, utilizando “negativas” como exemplo, a hipótese é de que se  $\beta_1$  for igual a zero, significa que a variável “negativas” não tem efeito para o regressando, podendo este ser: ibov, ibov\_vol, ibra e ibra\_vol. Caso contrário, “negativas” não poderá ser descartado das equações anteriormente citadas. O mesmo exemplo serve para as variáveis: “positivas”, “incerteza”, “litigiosas” e “modais”, no entanto alterando o valor de Beta ( $\beta$ ) para o seu respectivo.

Ainda sobre as variáveis primárias, espera-se que “negativas” tenha sinal negativo nas Equações 2 e 6 (relacionadas com a rentabilidade diária dos índices estudados), e sinal positivo na 3 e na 7 (relacionadas à volatilidade). O efeito contrário é esperado para “positivas”, ou seja, sinal positivo em 2 e 6 e negativo em 3 e 7 (Tetlock, 2007; Tetlock et al., 2008).

Igualmente à variável “negativas”, de “incerteza” e “litigiosas” espera-se uma menor rentabilidade (sinal negativo) para as Equações 2 e 6, e positivo nas Equações 3 e 7. Acredita-se que um maior peso dessas palavras nos jornais aumente as incertezas no que diz respeito à direção que o mercado tomará.

Por fim, de “modais” nada é esperado quanto aos sinais no tocante à rentabilidade e à volatilidade.

## RESULTADOS

### Análise das palavras (*term weighting*)

A Tabela 2 apresenta as palavras que tiveram maior peso dentro da amostra de jornais coletados. Comparando-se os resultados evidenciados pelas duas tabelas, observa-se uma grande diferença entre as palavras que obtiveram maior peso nos dois casos. Por exemplo, a palavra da categoria positiva “desenvolvimento” obteve maior peso (significância) na Tabela 2 para todos os anos individualmente, como também para o *pooled*, enquanto na Tabela 3 esta não se apresentou entre as cinco posições de maior significância em nenhum caso.

Outro exemplo interessante é o da palavra da categoria negativa “mensalão”. Observando novamente as tabelas, constata-se que, em 2012, na Tabela 2, “mensalão” foi o termo mais significativo, com peso igual a 99,78. Já na Tabela 3, para o mesmo ano, “mensalão” aparece na segunda posição, com um peso maior igual a 103,85. Respectivamente, a palavra “mensalão” apareceu 313 vezes na amostra contendo todos os editoriais e 255 na com editoriais filtrados.

A diferença entre os resultados acima pode ser explicada pela forma como a equação (1), utilizada no algoritmo de Pagliarussi et al. (2016), trata os dados.

Tabela 2. *Term weighting*

	Todos os editoriais		Editoriais filtrados	
Negativas	prejuizo	468,17	crise	468,21
	desconto	454,96	inflacao	467,42
	deficit	444,88	deficit	460,46
	baixo	432,67	risco	447,71
	corte	431,63	reducao	445,41
Positivas	desenvolvimento	434,53	crescimento	484,31
	creceu	422,64	investimentos	480,18
	investe	421,81	investimento	472,50
	confianca	417,36	crece	454,24
	ganha	417,03	confianca	393,26
Litigiosas	judicial	434,78	recursos	471,03
	lei	431,58	disputa	446,57
	contrato	429,32	fiscal	444,62
	contratos	424,17	regras	427,87
	credores	420,01	lei	406,82

Continua

### Estatísticas descritivas

A Tabela 3 apresenta as estatísticas descritivas das variáveis primárias para os editoriais completos e filtrados, respectivamente. Para os editoriais completos, as palavras negativas tiveram, em média, maior peso, seguidas das positivas e das litigiosas. Modais e incerteza apresentaram os menores valores, respectivamente.

O mesmo aconteceu para o caso dos editoriais filtrados. No entanto, estes apresentaram valores de média, mediana, máximo, mínimo mais baixo. Além disso, o desvio padrão também apresentou valores inferiores. O menor número de palavras dentro dessa amostra explica esse efeito.

Similar ao apresentado em Davolos, Rogers, Silva e Oliveira (2013), de maneira a contribuir com o entendimento dos resultados obtidos, a Tabela 4 demonstra as principais notícias veiculadas no *Valor Econômico* nos 24 dias de maior retorno e nos 24 dias de menor retorno do Ibovespa ao longo do período analisado. No Painel A, percebe-se que, nos dias de maior retorno, as notícias de cunho político e referentes a grandes impactos na economia com tom positivo dominam o noticiário. De maneira similar, no Painel B, percebe-se uma dominância de notícias econômicas e políticas de tom negativo.

Tabela 2. *Term weighting*

Conclusão

	Todos os editoriais		Editoriais filtrados	
Incerteza	riscos	447,18	espera	459,93
	expectativa	438,80	risco	449,52
	possível	430,78	expectativa	397,10
	expectativas	423,16	riscos	377,11
	tendencia	417,41	expectativas	350,87
Modais	alto	336,89	menos	415,24
	pouco	328,24	menor	413,83
	grande	322,16	forte	407,08
	maioria	321,18	grandes	370,45
	maiores	314,71	pouco	354,36

Nota: As palavras aparecem sem acentos neste quadro porque foram assim avaliadas pelo software.

Tabela 3. Estatísticas descritivas

		Negativas	Positivas	Incerteza	Litigiosas	Modais
Editoriais completos	Média	59,381	33,807	14,576	24,317	6,853
	Mediana	39,015	23,753	15,683	20,963	5,093
	Máximo	107,302	69,974	34,512	58,325	22,062
	Mínimo	15,418	8,815	1,909	3,145	0,486
	Desv. padrão	13,988	10,603	5,711	9,001	3,157
Editoriais filtrados	Média	37,334	18,835	9,199	14,017	6,066
	Mediana	24,620	15,317	5,048	11,821	3,282
	Máximo	81,015	55,908	30,563	49,683	25,578
	Mínimo	10,229	0,000	0,000	0,000	0,000
	Desv. padrão	11,288	8,147	5,118	7,736	3,613

Tabela 4. Principais notícias veiculadas nos dias de maior oscilação do Ibovespa entre janeiro de 2012 e dezembro de 2016

Painel A. Retornos mais positivos no período e as notícias do dia			
Data	Ibovespa	% dia	Notícia
22-out-2014	56.432,03	+7,75%	Abstenções devem atingir nível recorde no 2º turno
			“Kit eleições” ainda tem ganhos da especulação
17-mar-2016	50.886,40	+6,66%	Grampo de Lula e Dilma faz país mergulhar em caos político
			Federal Reserve reduz ritmo de alta do juro

Continua

Tabela 4. Principais notícias veiculadas nos dias de maior oscilação do Ibovespa entre janeiro de 2012 e dezembro de 2016

Painel A. Retornos mais positivos no período e as notícias do dia			
Data	Ibovespa	% dia	Notícia
3-mar-2016	47.288,28	+5,65%	Cunha deve virar réu, mas diz que não deixará cargo
			Distante do PT, Dilma procura salvar mandato
21-nov-2014	56.055,06	+4,96%	Economia fraca e dólar forte derrubam lucros
			Marina apoia Aécio e cita 'Carta aos Brasileiros'
27-jul-2012	56.515,35	+4,70%	Governo só concorda com aumento linear a servidor
			Crédito cresce e calote tem discreto recuo
3-nov-2015	48.023,72	+4,65%	Conta de juros e do dólar já tem impacto nos balanços
			MP e PF abrem seis frentes de investigação contra Lula
24-ago-2015	44.312,77	+4,63%	Governo prevê elevar impostos no próximo ano
			Lucro de bancos de médio porte cai 43%
6-out-2014	57.013,57	+4,57%	Votos mantêm polarização PT-PSDB
			Desempenho de Aécio Neves deve animar mercado
29-jan-2016	40.263,28	+4,53%	Concessões de campos de petróleo serão prorrogadas
			Petrobras quer sair de várias áreas
28-out-2014	52.294,96	+4,42%	Lula faz três indicações de nomes para a Fazenda
			Empresas pedem definições a Dilma
22-fev-2016	43.304,60	+4,18%	Renegociação de dívidas dos Estados eleva déficit
			Fabricantes de commodities derrubam PIB
7-nov-2016	64.092,86	+4,16%	Governo quer mudar lei para intervir na Oi
			FBI arquiva de novo caso contra Hillary
31-out-2014	54.666,09	+4,14%	Movimento de elevação de juros deve continuar
			Prejuízo de R\$ 3 bi na Vale surpreende
10-mai-2016	53.051,05	+4,08%	Renan ignora Maranhão e segue com o impeachment
			Presidência do BC perde status de ministério
4-mar-2016	49.168,99	+3,98%	Recessão se estende e ameaça o país
			Delação de Delcídio anima os defensores do impeachment
12-abr-2016	52.068,20	+3,94%	Comissão da Câmara aprova a abertura do impeachment
			Mercado já conta com mudança de governo
21-mai-2012	56.583,31	+3,83%	CGU vê irregularidades em aplicações do FGTS
			Dividendo já rende mais que juro real
3-jan-2012	59.224,75	+3,82%	Concessões de 2012 exigirão investimentos de R\$ 90 bi
			Exportações sentem o peso da crise

Continua

Tabela 4. Principais notícias veiculadas nos dias de maior oscilação do Ibovespa entre janeiro de 2012 e dezembro de 2016

Painel A. Retornos mais positivos no período e as notícias do dia			
Data	Ibovespa	% dia	Notícia
9-dez-2015	46.082,81	+3,82%	Impeachment nas mãos da oposição
			Aprofunda-se a queda das commodities
2-out-2015	47.012,50	+3,80%	Medida amplia poder de BC e CVM para apurar e punir
			Investimento da Petrobras cai à metade
17-dez-2014	48.722,59	+3,67%	Crise russa pressiona emergentes
			Para Levy, ajuste fiscal é ainda mais urgente
27-ago-2015	47.689,07	+3,60%	Governo vai propor a volta da CPMF para cobrir rombo
			Empresariado ouve Dilma e pressiona Levy
2-jan-2013	62.761,20	+3,55%	Indústria deve ter 'folga' de custos de 20% em 2013
			Economistas preveem PIB acima de 3%
13-set-2012	61.979,58	+3,45%	Desaceleração é boa para China e para o mundo
			Calote das empresas é recorde
Painel B. Retornos mais negativos no período e as notícias do dia			
Data	Ibovespa	%	Notícia
15-out-2014	56.134,11	-3,18%	Queda do petróleo põe fim à defasagem da gasolina
			Dilma pode vetar novas desonerações
17-mai-2012	54.022,06	-3,28%	Crise tende a retardar recuperação para 2013
			Bancos gregos perdem crédito e depósitos
10-jul-2012	53.581,92	-3,31%	Elétricas pressionam por definição sobre concessões
			Ibama acelera licenças para petróleo e gás
16-out-2014	54.270,88	-3,32%	Petróleo em queda ameaça investimentos da Petrobras
			Desaceleração derruba bolsas pelo mundo
8-dez-2014	50.246,99	-3,35%	Escassez de água se torna fator de risco para crédito
			Governo prevê aumento da carga na LDO
10-out-2014	55.317,26	-3,41%	Delatores revelam esquema de propinas na Petrobras
			Equacionar dívida é maior objetivo da Oi
10-jun-2016	49.337,70	-3,42%	Fraga sugere mais prazo à meta de 4,5%
			Acusações vão marcar eleição nos EUA
19-jun-2013	47.742,76	-3,44%	Mau humor atinge mercado e megaoferta é cancelada
			Vítimas do Fed vivem dia de expectativa
14-mai-2012	57.411,51	-3,50%	Bancos intensificam o aumento de tarifas
			Rendimento imprevisível da renda fixa

Continua

Tabela 4. Principais notícias veiculadas nos dias de maior oscilação do Ibovespa entre janeiro de 2012 e dezembro de 2016

Conclusão

Painel B. Retornos mais negativos no período e as notícias do dia			
Data	Ibovespa	%	Notícia
21-out-2014	52.373,42	-3,50%	Lucro de empresas deve cair 10% no 3º trimestre
			Fundos tentam atrair recursos para o exterior
12-dez-2014	48.012,83	-3,55%	Gerente da Petrobras diz ter alertado Graça sobre abusos
			Conta de luz salga IPCA de janeiro
15-mar-2016	47.098,11	-3,59%	Juíza passa a Moro pedido de prisão e Lula será ministro
			Congresso discute como afastar Dilma
04-abr-2016	48.746,98	-3,61%	Dívidas de empresas crescem 24%
			Desabam os resultados das incorporadoras
27-out-2014	50.083,17	-3,63%	Dilma, reeleita, promete diálogo
			Vitória apertada mantém Lula sob os holofotes
9-set-2016	57.952,98	-3,70%	Funcionário público ganha até 200% mais que o privado
			Cresce reação ao aumento de servidores
2-jul-2013	45.244,99	-4,13%	Por decreto, BNDES ajuda no superávit
			Dilma nega reforma ministerial
13-out-2015	47.346,20	-4,19%	Acordos para redução de salário e jornada avançam
			Acuado, Cunha despacha hoje impeachment
11-nov-2016	58.827,82	-4,21%	“Efeito Trump” valoriza o dólar e gera turbulência
			China retoma ofensiva para uma 'TPP' sem Estados Unidos
1-dez-2016	59.466,33	-4,23%	Economia derrete e crescem pressões por corte de juros
			Estados podem ganhar 'lei de recuperação'
2-fev-2016	38.770,10	-4,27%	Bônus brasileiros pagam juro alto e viram atração
			Crise no aço
1-dez-2014	52.218,40	-4,43%	Levy quer novo modelo de financiamento a empresas
			Restos a pagar já superam previsão de investimento
29-set-2014	54.599,67	-4,57%	Governo inicia cerco à fraude no emprego
			BC resiste a transação entre BTG e Nacional
24-out-2014	51.968,04	-7,91%	Mercado reflete avanço de Dilma
			Delator cita corrupção também na Eletrobras
21-ago-2015	42.352,03	-9,09%	Janot pede 184 anos de prisão para Cunha
			Temer vai sair da articulação do governo

## Análise das regressões

A Tabela 5 apresenta os resultados encontrados para as estimações das Equações 2 e 6. As variáveis dependentes são: “ibov” e “ibra”, que são a rentabilidade diária dos índices Bovespa e do IBrA, e “ibov\_vol” e “ibra\_vol”, que são a volatilidade de ambos os índices calculada num período de 60 dias.

Tabela 5. Relação entre tom das notícias e retorno

	Editorial completo		Editorial filtrado	
	ibov	ibra	ibov	ibra
Negativas	0,0000274	0,0000334	0,0000325	0,0000471
	(-0,,99)	(-1,11)	(-1,29)	(-1,78)
Positivas	0,00000981	0,000011	0,0000337	0,0000391
	(-0,55)	(-0,62)	(-1,91)	(-1,75)
Incerteza	-0,0000951**	-0,0000545	-0,0000801*	-0,0000774*
	(-3,51)	(-1,83)	(-2,52)	(-2,43)
Litigiosas	0,0000435	0,0000482	-0,0000142	-0,0000187
	(-1,34)	(-1,34)	(-0,37)	(-0,04)
Modais	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)
Constante	-0,00313	-0,00239	-0,000468	0,000542
	(-0,69)	(-0,48)	(-0,09)	(-0,10)
Efeito fixo de ano	sim	sim	sim	sim
N	1237	1237	1237	1237
R <sup>2</sup>	0,496	0,436	0,494	0,434

Nota: Coeficientes na primeira linha e estatística p-valor em parênteses na segunda linha de cada variável ( $\alpha 10\%$ \*,  $\alpha 5\%$ \*\*\*,  $\alpha 1\%$ \*\*\*).

Para uma melhor organização da tabela, optou-se por diferenciar as variáveis dependentes da amostra em que estão todos os editoriais e outra subamostra na qual somente foram analisados os seguintes editoriais: Brasil, Política, Internacional e Finanças.

Os resultados evidenciam que o peso das palavras com tom de “incerteza” é estatisticamente relevante para explicar os retornos do Ibovespa quando se considera o editorial completo e para explicar os retornos do Ibovespa e do IBrA quando se consideram os editoriais filtrados. Palavras com tons “Negativos”, “Positivos” e “Modais” não apresentaram significância em nenhuma das equações.

De maneira geral, as variáveis de controle, SMB, HML e WML e riskfree são estatisticamente significantes, demonstrando a adequação da estimação do modelo de quatro fatores de Fama e French (1993).

A Tabela 6 apresenta os resultados para as estimações das Equações 3 e 7. Esses resultados apresentam evidências menos robustas e demonstram que somente palavras com tom negativo na subamostra de editoriais filtrados apresentam associação positiva com a volatilidade do Ibovespa. Esse resultado, não totalmente alinhado como o esperado com base na literatura internacional (Tetlock, 2007; Tetlock et al., 2008), pode ser explicado pela alta volatilidade do mercado brasileiro quando comparada com a volatilidade do mercado norte-americano.

**Tabela 6. Relação entre tom das notícias e a volatilidade**

	Editorial completo		Editorial filtrado	
	Vol Ibov	Vol Ibra	Vol Ibov	Vol Ibra
Negativas	0,00000926	0,00000441	0,0000160**	0,00000815
	(-1,59)	(-0,92)	(-2,99)	(-1,10)
Positivas	0,0000016	0,00000124	-0,00000547	-0,00000639
	(-0,26)	(-0,24)	(-0,72)	(-0,90)
Incerteza	0,00000504	0,00000892	-0,0000104	-0,00000177
	(-0,99)	(-1,63)	(-1,82)	(-0,25)
Litigiosas	0,00000551	0,00000737	0,00000116	0,00000374
	(-0,62)	(-0,83)	(-0,09)	(-0,28)
Modais	-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)
Constante	0,000519	0,000306	0,000824	0,000623
	(-0,20)	(-0,13)	(-0,32)	(-0,27)
Efeito fixo de ano	sim	sim	sim	sim
N	1237	1237	1237	1237
R <sup>2</sup>	0,496	0,436	0,494	0,434

Nota: Coeficientes na primeira linha e estatística p-valor em parênteses na segunda linha de cada variável ( $\alpha$ 10%\*,  $\alpha$ 5%\*\*\*,  $\alpha$ 1%\*\*\*).

## CONSIDERAÇÕES FINAIS

Este trabalho teve por objetivo aplicar a técnica de análise de sentimento para as edições diárias do jornal *Valor Econômico* de maneira a investigar a existência de associação entre a rentabilidade e a volatilidade do Ibovespa e do IBrA e o tom das notícias divulgadas pela mídia diária impressa especializada em Economia e Finanças no mercado brasileiro.

Dos resultados encontrados, observou-se certa discrepância entre os valores encontrados e a literatura no caso dos termos “negativos”. Esperava-se, conforme Tetlock (2007) e Tetlock et al. (2008), que uma maior quantidade de termos negativos trouxesse efeitos desfavoráveis tanto para a rentabilidade (diminuindo) quanto para a volatilidade (aumentando), o que só ocorreu no mercado brasileiro para o caso da volatilidade, e ainda assim com evidências mais fracas. Por sua vez, de acordo com o esperado pelas evidências no mercado norte-americano, os termos do tipo “incerteza” apresentaram relação negativa com a rentabilidade diária do Ibovespa e do IBrA.

Dos dicionários aplicados neste estudo, entende-se que aqueles contendo palavras do tipo “negativas” e de “incerteza” apresentaram relevância, em contrapartida às palavras com tons “positivos”, “litigiosos” e “modais”. Dessa forma, observa-se que uma maior quantidade desses termos dentro da edição diária do *Valor Econômico* poderia trazer alguma consequência associada à rentabilidade e/ou volatilidade do mercado. Em outras palavras, as evidências apontam que o mercado avalia com maior peso palavras negativas e de incerteza divulgadas na mídia especializada.

Este trabalho busca contribuir para a pesquisa dos impactos de informações qualitativas provenientes de análise textual dentro do Brasil.

Tomadas em conjunto, as evidências obtidas neste estudo apontam para a relevância da mídia especializada no Brasil, bem como para a existência de conteúdo informativo nas notícias veiculadas. Os resultados podem incentivar os participantes do mercado de capitais a empregar esse método junto a outros de aprendizado de máquina para realização de previsões do comportamento de variáveis no mercado (Cambria, 2016; Tripathy et al., 2016). Adicionalmente, os investidores podem se beneficiar dos resultados apresentados nesta pesquisa, pois se evidencia que a extração do tom de notícias veiculadas pela mídia especializada no Brasil guarda relação com a rentabilidade dos papéis e/ou com a sua volatilidade no dia corrente da análise.

Sobre as limitações deste trabalho, deve-se lembrar que a relação investigada foi entre o peso/tons das palavras e o mercado como um todo (neste caso, sendo os índices

Bovespa e Brasil Amplo). Nesse contexto, apenas uma notícia, por exemplo, negativa de uma empresa com peso muito grande dentro dos índices citados poderia trazer movimentos fortes para a rentabilidade e/ou volatilidade, enquanto numa possível previsão do mercado, a partir desses modelos, esse efeito não seria sentido.

Dessa forma, sugere-se, para outros trabalhos, realizar a relação direta dos tons das notícias específicas por empresa e o retorno e a volatilidade dos papéis. Também se sugere a adoção de outros veículos de comunicação financeira existentes (por exemplo: Bloomberg, Google Finance etc.), além de conteúdos de informações nas redes sociais como Twitter ou Facebook. Pode-se também evoluir no desenvolvimento do dicionário de palavras para a realização da análise textual. Por fim, sugere-se a utilização de métodos de aprendizado de máquina (*machine learning*) para criar métodos de previsão de algum indicador do mercado ou empresa. Vê-se, portanto, amplo espectro de pesquisa sobre temas correlatos com este artigo.

### NOTA DE AGRADECIMENTO

Fernando Caio Galdi agradece à Fundação de Amparo à Pesquisa e Inovação do Espírito Santo (FAPES) pelo apoio financeiro para a execução de pesquisas.

## REFERÊNCIAS

- 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*. Recuperado de <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*. Recuperado de <http://www.bmfbovespa.com.br/indices/ResumoIndice.aspx?Indice=Ibovespa&Idioma=pt-br>
- BM&FBovespa. (2015b). *Índice Brasil Amplo (IBrA). Composição/ Carteira do índice*. Recuperado de <http://www.bmfbovespa.com.br/indices/ResumoCarteiraTeorica.aspx?Indice=IBrA&Idioma=pt-br>

- BM&FBovespa. (2015c). *Índice Brasil Amplo (IBrA). Estatísticas históricas*. Recuperado de <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. Da, & 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*. Recuperado de <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, B. A., & 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., & Macksassy, 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, S. K. (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*. Recuperado de <http://www.valor.com.br/impresso/>
- Valor Econômico. (2013, Maio 15). *Edição impressa*. Recuperado de <http://www.valor.com.br/impresso/>