

**UNIVERSIDADE DE BRASÍLIA
PROGRAMA MULTIINSTITUCIONAL E INTER-REGIONAL DE PÓS-
GRADUAÇÃO EM CIÊNCIAS CONTÁBEIS UnB/UFPB/UFRN
DOUTORADO EM CIÊNCIAS CONTÁBEIS**

MARIA DANIELLA DE OLIVEIRA PEREIRA DA SILVA

**O EFEITO DO SENTIMENTO DAS NOTÍCIAS SOBRE O COMPORTAMENTO
DOS PREÇOS NO MERCADO ACIONÁRIO BRASILEIRO**

Orientador: Prof. Dr. Márcio André Veras Machado

João Pessoa/PB
2017

MARIA DANIELLA DE OLIVEIRA PEREIRA DA SILVA

**O EFEITO DO SENTIMENTO DAS NOTÍCIAS SOBRE O COMPORTAMENTO
DOS PREÇOS NO MERCADO ACIONÁRIO BRASILEIRO**

Tese apresentada ao Programa Multi-institucional e Inter-regional de Pós-graduação em Ciências Contábeis da Universidade de Brasília, Universidade Federal da Paraíba e Universidade Federal do Rio Grande do Norte, como requisito parcial para a obtenção do título de Doutor em Ciências Contábeis.

Orientador: Prof. Dr. Márcio André Veras Machado.

Área de concentração: Mensuração Contábil.

Linha de Pesquisa: Contabilidade e Mercado Financeiro.

S586e Silva, Maria Daniella de Oliveira Pereira da.

O efeito do sentimento das notícias sobre o comportamento dos preços no mercado acionário brasileiro / Maria Daniella de Oliveira Pereira da Silva. - João Pessoa, 2018.

211 f. : il.

Orientação: Márcio André Veras Machado.
Tese (Doutorado) - UNB-UFPB/CCSA.

1. Notícias Financeiras. 2. Investidor - Sentimento Textual. 3. Mercado Acionário. 4. Previsão de risco - Volatilidade. I. Machado, Márcio André Veras. II. Título.

UFPB/BC

UNIVERSIDADE DE BRASÍLIA – UnB

Reitor:

Prof. Dra. Márcia Abrahão Moura

Vice-Reitor:

Prof. Dr. Enrique Huelva

Decano de Pós-Graduação:

Prof. Dr. Cláudio Henrique Soares Del Menezzi

**Diretor da Faculdade de Administração, Contabilidade, Economia e Gestão de Políticas
Públicas - FACE:**

Prof. Dr. Roberto de Goes Ellery Junior

Chefe do Departamento de Ciências Contábeis e Atuariais:

Prof. Dr. José Antônio de França

**Coordenador Geral do Programa Multi-Institucional e Inter-Regional de Pós-
Graduação em Ciências Contábeis da UnB, UFPB e UFRN:**

Prof. Dr. Jorge Katsumi Niyama

MARIA DANIELLA DE OLIVEIRA PEREIRA DA SILVA

**O EFEITO DO SENTIMENTO DAS NOTÍCIAS SOBRE O COMPORTAMENTO
DOS PREÇOS NO MERCADO ACIONÁRIO BRASILEIRO**

Tese apresentada ao Curso de Doutorado do Programa Multi-institucional e Inter-regional de Pós-graduação em Ciências Contábeis da Universidade de Brasília, Universidade Federal da Paraíba e Universidade Federal do Rio Grande do Norte, como requisito parcial à obtenção do grau de Doutor em Ciências Contábeis.

Prof. Dr. Márcio André Veras Machado

Universidade Federal da Paraíba
(Orientador)

Prof. Dr. Paulo Roberto Nóbrega Cavalcante

Universidade Federal da Paraíba
(Avaliador Interno)

Prof. Dr. Luiz Renato Regis de Oliveira Lima

Universidade Federal da Paraíba / *University of Tennessee*
(Avaliador Interno)

Prof. Dr. Rafael Lopes Rogo

University of British Columbia
(Avaliador Externo)

À minha tão querida avó, Maria,
aos meus pais, Carlos e Ivanilda,
à minha irmã, Deborah e
ao meu amor, Ricardo.

AGRADECIMENTOS

Uma tese é fruto de uma longa caminhada, marcada por muitas horas de trabalho e por muitas reflexões. É um árduo trabalho construído mediante o apoio e a colaboração de muitas pessoas, pois sozinho não se faz uma tese. Desse modo, agradeço:

Primeiramente a Deus por iluminar os meus caminhos e por me dar forças para concluir esta tese.

Ao meu orientador, Márcio Machado, pelo incentivo à pesquisa na área de *text regress*, pela confiança depositada, por sua infinita disponibilidade, comprometimento e dedicação ao longo de toda a elaboração do trabalho.

Aos professores Jorge Katsumi Niyama (UnB), Luiz Renato Lima (UFPB-UTK), Paulo Roberto Cavalcante (UFPB) e Rafael Rogo (UBC), membros avaliadores, por aceitarem prontamente o convite para avaliar o trabalho, por incentivar a realização deste estudo, e pelas valiosas críticas e sugestões apresentadas para o refinamento desta tese.

Aos professores Wesley Mendes-da-Silva (FGV) e Roy Martelanc (USP), membros avaliadores do Consórcio Doutoral Anpad – 2016, pelas contribuições e por todos os comentários sobre a proposta da tese, que contribuíram com a evolução da pesquisa.

Aos programadores Aldrey Galindo e Jorge Dias (UFPB) por contribuírem na construção das programações, utilizadas para a obtenção dos dados e para a leitura automatizada dos textos.

Aos professores Aléssio Almeida (UFPB), César Augusto Tibúrcio (UnB), Edilson Paulo (UFPB), Fátima Freire (UnB), José Dionísio Gomes da Silva (UFRN), José Matias-Pereira (UnB), Jorge Katsumi Niyama (UnB), Luiz Renato Lima (UFPB-UTK), Márcio Machado (UFPB), Otávio Ribeiro de Medeiros (UnB), Paulo Amilton Maia (UFPB), Paulo Roberto Barbosa Lustosa (UnB), Paulo Roberto Cavalcante (UFPB), Rodrigo Gonçalves (UnB) e Wenner Lucena (UFPB) pelo aprendizado em suas disciplinas.

Ao Prof. André Callado (UFRPE), meu orientador do mestrado, pelo incentivo à realização do doutorado no Programa Multi.

Aos professores Aldo Leonardo Callado (UFPB) e Márcia Reis Machado (UFPB) pelo apoio e pelo incentivo dado ao longo do doutorado.

Ao Prof. César Augusto Tibúrcio (UnB) por partilhar a sua experiência em trabalhar com narrativas e pelas instigantes discussões sobre economia da informação.

Ao Prof. Luiz Renato Lima (UFPB-UTK) por todos os seus ensinamentos sobre modelagens econométricas e pelos seus valiosos comentários sobre *text regress*, que contribuíram com a evolução da tese.

Ao Prof. Rodrigo Gonçalves (UnB) por seus profundos questionamentos sobre a sustentação teórica da tese.

Ao Prof. Wenner Lucena (UFPB), que foi meu orientador na graduação e com quem tive aula na graduação e na pós-graduação, meu mentor na área acadêmica, agradeço profundamente por ter me apresentado à pesquisa contábil.

À Coordenação do Curso e às secretarias, especialmente à Inez e a Wilma.

Ao Departamento de Gestão Pública (CCSA-UFPB), por conceder o afastamento de todas as minhas atribuições de docente ao longo do doutorado.

Aos meus amigos de turma do multi: Aline, Anna Paola, Antônio Firmino, Atelmo, Charline, Christianne, Daniel, Ednilto, José Emerson, Ivone, Joana, Luzivalda, Maurício, Renato e Robério, pelo companheirismo e pela paciência que tiveram para ouvir as minhas extensas reflexões. Em especial a: Aline, Anna Paola, Christianne, Joana, Luzivalda e Robério, por ouvirem as minhas inquietações de pesquisa, pelas prazerosas discussões sobre a tese e por todas as palavras de incentivo e conforto.

Aos meus amigos da UFPB: Adriana Vasconcelos, Dimas Queiroz, Douglas Franco, Karla Katiúscia, Kleber Formiga, Lauana Cabral, Lucas Godeiro, Maria José Paiva e Vanessa Meneses, por partilharem suas experiências e aprendizado.

À minha amiga Jéssica Silva pelo apoio e por todas as horas dedicadas à leitura da tese.

À minha família por todo o apoio concedido e por compreenderem as minhas constantes ausências. Em especial, a minha Vó Maria, aos meus pais Carlos e Ivanilda, e a minha irmã, Deborah, por todo o carinho e incentivo.

Ao meu namorado Ricardo, pela paciência, pelo companheirismo, apoio, incentivo, pelas motivações e por ser meu grande esteio ao longo de todo o doutorado.

RESUMO

O objetivo deste estudo foi investigar o efeito do sentimento textual das notícias financeiras sobre o comportamento dos preços no mercado acionário brasileiro, buscando respaldo teórico no Teorema da Concordância, no viés da negatividade e na Hipótese do Mercado Eficiente. Para analisar o efeito do tom das notícias sobre o comportamento de oscilação dos preços no mercado brasileiro, foi verificada a influência que o sentimento textual das notícias exercia sobre o retorno do índice IBOVESPA e sobre o risco (volatilidade do retorno) e, analisada a influência do sentimento textual na previsão do risco do mercado. Para alcançar o objetivo do trabalho, foram utilizados os valores diários dos índices: IBOVESPA, DJIA, VIX e S&P500, e analisados os textos das notícias financeiras do Jornal Valor Econômico (45.304 matérias), no período de 25 de julho de 2011 a 30 de junho 2017, correspondendo a 1.470 observações diárias. A escolha das variáveis preditoras da volatilidade foi realizada pelo método de regressão quantílica ℓ_1 -penalizada - LASSO, e o impacto do sentimento sobre o mercado foi avaliado por regressão quantílica autoregressiva. Os resultados levantados mostram que o sentimento de pessimismo da mídia prever reduções do retorno do IBOVESPA em -28.72 pontos-base e o sentimento de otimismo prever aumento de 22 pontos-base sobre o retorno do mercado, quando o mercado apresenta baixo desempenhos (baixo retorno). Em relação à volatilidade, foi observado que o pessimismo da mídia corrobora com a redução da volatilidade. Nos períodos de alta incerteza econômica foi verificado que o sentimento de pessimismo induz a reduções nos retornos, nas circunstâncias de melhor ou pior desempenho do mercado, chegando a prever quedas de -76.21 pontos-base sobre os menores retornos do mercado (q.05), e reduções de -81.94 pontos-base (q.95) sobre os maiores retornos do mercado (melhor desempenho). Os resultados sobre a previsão da volatilidade, mostram que o sentimento textual é um preditor da volatilidade, que contribui com a redução dos erros de previsão da volatilidade. Portanto, conclui-se que, as informações provenientes dos jornais, no Brasil, afetam a percepção dos investidores nos momentos que existe uma maior incerteza no mercado e na economia. O trabalho contribuiu com as discussões sobre o papel da mídia no mercado acionário de países emergentes, levantando evidências de que o sentimento textual das notícias pode fornecer informações importantes para o gerenciamento de risco.

Palavras-Chave: Notícias Financeiras; Sentimento Textual; Mercado Acionário; Previsão do risco.

ABSTRACT

This paper aimed to investigate the effect of the textual sentiment of the financial news on asset prices in the Brazilian Stock Exchange, seeking theoretical support in the agreement theorem, in the negativity bias and in the efficient market hypothesis. In order to analyze the effect of the tone of the news on asset prices. I verified the influence the textual sentiment of the news exerted on the IBOVESPA index return and the risk (volatility of the return); also I analyzed the influence of textual sentiment on the forecast market risk. This study used to daily data of the following index: IBOVESPA, DJIA, VIX and S&P500. As for the texts of financial columns from the newspaper Valor Econômico (45.304 articles). I analyzed the period of July 25, 2011 to June 30, 2017, corresponding to 1.470 daily observations. The choice of predictors of volatility was performed by ℓ_1 -penalized (LASSO) quantile regression method and the impact of the sentiment on the market was evaluated by quantile autoregression. As for the main results, this study shows that the sentiment of media pessimism predict reductions in the return of IBOVESPA of -28.72 basis points and optimism sentiment predict increase of 22 basis points on the market return when low performances market (low return). As for the volatility, I observed that media pessimism corroborates with the reduction of volatility. In the periods of high uncertainty, the sentiment of pessimism led to reductions in returns, in the circumstances of high or low market performance, predicting -76.21 basis points on the low performance ($q.05$), and reductions of -81.94 points ($q.95$) on the high performance. Regarding the forecast of volatility, the results that the textual sentiment is a predictor of volatility, which contributes to minimizes the forecast errors of volatility. Therefore, we can conclude that the information coming from newspapers in Brazil affect the perception of investors in moments of greater uncertainty in the market and in the economy. The work has contributed to the discussions about the role of the media in the stock market of emerging countries, raising evidence that textual sentiment of news can provide information for risk management.

Keywords: Financial News; Textual Sentiment; Stock Market; Forecast risk.

LISTA DE QUADROS

Quadro 1 - Síntese dos principais estudos empíricos da reação defasada do mercado no ajuste dos preços	41
Quadro 2 - Síntese dos principais estudos empíricos sobre notícias da mídia	63
Quadro 3 - Síntese dos principais estudos empíricos sobre a incerteza econômica	70
Quadro 4 - Síntese dos sinais esperados para o Sentimento Textual	90
Quadro 5 - Síntese das Hipóteses Analisadas Sobre a Influência do Sentimento Textual Sobre o Mercado Acionário Brasileiro	156

LISTA DE ILUSTRAÇÃO

Figura 1 - Desenho da Pesquisa.....	25
Figura 2 - Distribuição de Zipf das palavras da obra literária Ulisses	43
Figura 3 - Síntese das etapas de coleta, armazenamento e extração de informações	83
Figura 4 - Notícia do caderno de Finanças do Jornal Valor Econômico – 11/08/2017.....	84
Figura 5 - Distribuição das séries temporais analisadas	116
Figura 6 - Erros de Previsões da Volatilidade em Modelos Univariados.....	159
Figura 7 - Erros de Previsões da Volatilidade em Modelos de Regressão Quantílica Combinada (RQC), Regressão Quantílica fixa (RQF) e OLS.	161
Figura 8 - Estimativas dos coeficientes do sentimento textual ao longo da distribuição condicional do retorno do IBOVESPA	206
Figura 9 - Estimativas dos coeficientes do sentimento textual ao longo da distribuição condicional da volatilidade do IBOVESPA	207
Figura 10 - Estimativas dos coeficientes do sentimento de pessimismo e otimismo ao longo da distribuição condicional do Retorno do IBOVESPA	208
Figura 11 - Estimativas dos coeficientes do sentimento de pessimismo e otimismo ao longo da distribuição condicional da Volatilidade do IBOVESPA	209
Figura 12 - Estimativas dos coeficientes do sentimento de pessimismo em dias incerteza ao longo da distribuição condicional da retorno do IBOVESPA.....	210
Figura 13 - Estimativas dos coeficientes do sentimento de pessimismo em dias incerteza ao longo da distribuição condicional da volatilidade do IBOVESPA	211

LISTA DE TABELAS

Tabela 1 – Volume de matérias financeiras analisadas	77
Tabela 2 - Síntese das matérias usadas na estruturação do dicionário	102
Tabela 3 - Total de Palavras das Matérias	103
Tabela 4 - Palavras mais e menos citadas nas matérias.....	106
Tabela 5 - Estatística Descritivas das Variáveis – 2011 a 2017	111
Tabela 6 - Testes De Especificação Das Variáveis	113
Tabela 7 - Estimativas do Modelo GARCH(1,1)	114
Tabela 8 - Descrição das palavras presentes nas notícias financeiras	118
Tabela 9 - Impacto do sentimento textual sobre o retorno do IBOVESPA.....	124
Tabela 10 - Impacto do sentimento textual sobre a volatilidade realizada do IBOVESPA ...	128
Tabela 11 - Impacto do sentimento textual sobre o retorno em dias de mídia pessimista e otimista	135
Tabela 12 - Impacto do sentimento sobre a volatilidade em dias de mídia pessimista e otimista	140
Tabela 13 - Impacto do sentimento de pessimismo sobre o retorno do IBOVESPA em dias de alta e baixa incerteza.....	147
Tabela 14 - Impacto do sentimento de pessimismo sobre a volatilidade do IBOVESPA em dias de alta e baixa incerteza.....	152
Tabela 15 - Preditores da Volatilidade escolhidos por Lasso.....	158
Tabela 16 - Modelos condicionais de previsão da volatilidade.....	163

LISTA DE ABREVIATURAS E SIGLAS

AEX- *Amsterdam Exchange index*

AMEX – *American Stock Exchange*

BMF&BOVESPA – *Bolsa de Mercadorias e Futuros e Bolsa de Valores de São Paulo*

BM - *book-to-market*

CEO - *chief executive officer*

CDS - *credit default swap*

CVM - *Comissão de Valores Mobiliários*

DJIA - *Dow Jones Industrial Average*

EM - *Emotion Lexicon*

GI - *General Inquirer*

HE - *Henry's Finance*

HME - *Hipótese do Mercado Eficiente*

IIE-Br – *índice de incerteza*

IBRE – *Instituto Brasileiro de Economia*

IBOVESPA - *Índice IBOVESPA*

IPCA - *pelo índice de preço ao consumidor ampliado*

IPEA - *Instituto de Pesquisa Econômica Aplicada*

IVC - *Instituto Verificador de Comunicação*

Nasdaq – *National Association of Securities Dealers Automated Quotations*

NYSE - *New York Stock Exchange*

LIWC- *Linguistic Inquiry and Word Count*

LPN - *Linguagem de Processamento Natural*

L&M – *Dictionary Loughran and McDonald*

MD&A- *Management Discussion and Analysis*

PIB - *produto interno bruto*

LISTA DE ABREVIATURAS E SIGLAS

Russell 3000 – Russell 3000 index

SIRCA - Securities Industry Research Centre of the Asia Pacific

S&P 500– Standard & Poor’s 500 index

TRNA- Thomson Reuters News Analytics

VIX -Chicago Board Options Exchange Volatility Index

Wilshire 5000 - Wilshire 5000 Total Market Index

WSJ - Wall Street Journal

SUMÁRIO

1 INTRODUÇÃO	17
1.1 PROBLEMATIZAÇÃO	17
1.2 OBJETIVOS	23
1.2.1 Objetivo Geral	23
1.2.2 Objetivos Específicos	23
1.3 A TESE	24
1.4 ORIGINALIDADE E JUSTIFICATIVA.....	26
1.5 LIMITAÇÕES DA PESQUISA	28
1.6 ESTRUTURA DO TRABALHO	29
2 REVISÃO DA LITERATURA	30
2.1 TEOREMA DA CONCORDÂNCIA.....	30
2.2 HIPÓTESE DO MERCADO EFICIENTE.....	33
2.3 LEI DE ZIPF	42
2.4 SENTIMENTO TEXTUAL.....	43
2.4.1 Evidências Empíricas do Sentimento Textual das Notícias Financeiras	48
2.5 INCERTEZA ECONÔMICA	67
2.6 SÍNTESE DO CAPÍTULO E HIPÓTESES DE PESQUISA	71
3 METODOLOGIA	76
3.1 DADOS	76
3.2 DEFINIÇÃO DE VARIÁVEIS.....	78
3.2.1 Sentimento Textual	78
3.2.2 Retorno, Volatilidade e Incerteza Econômica	80
3.3 PROCEDIMENTOS DE ESTIMAÇÃO TEXTUAL.....	82
3.4 ESPECIFICAÇÕES DOS MODELOS ECONÔMÉTRICOS	84
3.4.1 Impacto	84
3.4.2 Previsão da Volatilidade	91
3.4.2.1 Modelos de Previsão da Volatilidade	95
3.4.2.2 Análise da Previsão	98
3.5 LISTA DE TERMOS FINANCEIROS	100
3.5.1 Construção do banco de dados para estruturação do dicionário de termos financeiros	100
3.5.2 Estruturação do Dicionário	102
3.5.3 Procedimentos de validação da lista de termos proposta	107

4 ANÁLISE DOS RESULTADOS	110
4.1 ESTATÍSTICA DESCRITIVA	110
4.2 ESTIMAÇÃO DA VOLATILIDADE – GARCH (1,1)	113
4.3 ANÁLISE GRÁFICA DAS VARIÁVEIS	115
4.4 DESCRIÇÃO DAS PALAVRAS	117
4.5 IMPACTO DO SENTIMENTO TEXTUAL.....	119
4.5.1 Influência do sentimento textual sobre o retorno e sobre a volatilidade	119
4.5.2 Influência do Sentimento Pessimista Sobre o Retorno e Sobre a Volatilidade do Mercado.....	130
4.5.3 Influência do Sentimento de Pessimismo Sobre o Retorno e Sobre a Volatilidade do Mercado em Períodos de Incerteza Econômica	143
4.5.4 Síntese dos Resultados da Influência do Sentimento Textual Sobre o Risco e Retorno do Ibovespa	155
4.6 PREVISÃO DA VOLATILIDADE.....	157
4.6.1 Análises das variáveis preditoras da volatilidade.....	157
4.6.2 Análises dos Modelos de Estimação da Previsão da Volatilidade.....	159
4.6.3 Síntese dos Resultados da Previsão de Volatilidade.....	166
5 CONSIDERAÇÕES FINAIS.....	170
REFERÊNCIAS	175
APÊNDICE A – ALGORITMO PARA CONTAGEM GERAL DAS PALAVRAS EM UM REPOSITÓRIO – LINGUAGEM R.....	185
APÊNDICE B - ALGORITMO PARA CONTABILIZAÇÃO DE NOVAS PALAVRAS ENCONTRADAS POR TEXTO	186
APÊNDICE C – ALGORITMO PARA ESTIMAÇÃO DO SENTIMENTO TEXTUAL DE UMA NOTÍCIA	187
APÊNDICE D – LISTAS DAS PALAVRAS NEGATIVAS.....	191
APÊNDICE E – LISTAS DAS PALAVRAS POSITIVAS	198
ANEXO A – EXEMPLO DA ESTRATIFICAÇÃO DO BANCO DE DADOS USADOS NA PROPOSTA DO DICIONÁRIO	202
ANEXO B – CORRELOGRAMA DA VARIÂNCIA CONDICIONAL DO IBOVESPA205	
ANEXO C – ESTIMATIVAS SOBRE O SENTIMENTO TEXTUAL.....	206

1 INTRODUÇÃO

O presente capítulo trata da delimitação e da definição dos objetivos de estudo, bem como apresenta os argumentos e as limitações da pesquisa realizada.

1.1 PROBLEMATIZAÇÃO

Desde meados da década de 1930, procura-se compreender a dinâmica do mercado acionário. Um dos precursores nas explicações sobre as oscilações dos preços das ações foi Keynes (1936), que defendia a ideia de que as movimentações selvagens dos preços não poderiam ser explicadas por fundamentos econômicos, pois o mercado era lastreado por incertezas, decorrentes do processo decisório não ser estruturado essencialmente por expectativas matemáticas, mas por sofrer a influência das emoções, fazendo alusão a tal comportamento como o “espírito animal” dos investidores. No entanto, a visão de Keynes (1936) sobre o comportamento dos preços não perdurou, diante do fortalecimento do conhecimento racional e da mensuração do risco.

Ao final da década de 1980, surgiram especulações de que as notícias divulgadas na mídia apresentariam conexões com os movimentos dos preços das ações. O estudo de Cutler, Poterba e Summers (1989) foi um dos pioneiros na investigação sobre o poder informativo das notícias macroeconômicas para o mercado acionário. Nessa obra, foi investigado se notícias de importantes acontecimentos mundiais estariam associadas às grandes oscilações de preços no mercado. Apesar da literatura da época atribuir que as variações dos preços das ações estavam atreladas à chegada de novas informações, os resultados alcançados no estudo de Cutler, Poterba e Summers (1989) mostraram que as notícias apresentavam um poder explicativo relativamente pequeno, para justificar os grandes movimentos do mercado. Desde então, a literatura tem investigado se os movimentos dos preços poderiam ser explicados pelo conteúdo informacional das notícias.

As pesquisas realizadas até o final da década de 1990 tratavam as notícias geralmente como uma variável latente, visto que as métricas para a mensuração das informações qualitativas apresentavam critérios subjetivos. Desse modo, as pesquisas empíricas sobre o comportamento do mercado acionário, tradicionalmente, eram realizadas com bases de dados essencialmente quantitativas. Em virtude disso, foi gerada uma volumosa literatura voltada para a análise das oscilações dos preços das ações, baseada em informações meramente numéricas

(TETLOCK; SAAR-TSECHANSKY; MACSKASSY, 2008; HUANG; TEOH; ZHANG, 2014).

Com o aperfeiçoamento da Linguagem de Processamento Natural (LPN), surgiram diversas técnicas voltadas para a análise de informações não numéricas, possibilitando, assim, a otimização das investigações empíricas que utilizam variáveis qualitativas. A LPN tem favorecido a realização de estudos que fazem uso de medidas qualitativas, como as pesquisas que tratam sobre narrativas (BEATTIE, 2014; FISHER; GARNSEY; HUGHES, 2016).

Pode-se dizer que as informações de origem qualitativas são tão importantes quanto às de origem quantitativa, pois, dentre as suas contribuições, ressalta-se o conteúdo informacional incremental sobre o desempenho das organizações (ENGELBERG, 2008; KERNEY; LIU, 2014; HUANG; TEOH; ZHANG, 2014).

No âmbito financeiro, parte das informações qualitativas pode ser descrita como informações linguísticas, expressas mediante textos ou por meio do discurso. No que diz respeito às informações textuais, as principais características analisadas nos estudos, geralmente, são: (a) legibilidade, definida como a complexidade sintática da comunicação, ou seja, nível de compreensão do texto (LEHAVY; LI; MERKLEY, 2011; BEATTIE, 2014); e (b) tom ou sentimento textual¹, que faz referência à semântica do processo de comunicação, quanto à interpretação da informação, podendo esta apresentar um sentido otimista ou pessimista, conforme as palavras empregadas no texto (BEATTIE, 2014; HUANG; TEOH; ZHANG, 2014).

Nesse sentido, a análise do sentimento textual envolve o contexto de suposições relacionadas ao comportamento dos investidores, ao desempenho das organizações ou à dinâmica do mercado acionário. Em geral, as pesquisas procuram avaliar o poder informacional que o tom da notícia pode apresentar. Desse modo, muitas informações textuais têm sido o foco das investigações empíricas, tais como: jornais (TETLOCK, 2007; GRIFFIN; HIRSCHEY; KELLY, 2011; DOUGAL et al., 2012; GARCÍA, 2013; LIU, 2014; HENDERSHOTT; LIVDAN; SCHÜRHOFF, 2015; STRAUß; VLIEGENTHART; VERHOEVEN, 2016); relatórios da administração ou relatórios específicos (LI, 2010b; LOUGHRAN; McDONALD, 2011; HUANG; TEOH; ZHANG, 2014); mensagens de *chat*, *Twitter*, *Facebook* e mensagens de *spam* relacionadas aos ativos (ANTWEILER; FRANK, 2004; BOLLEN; MAO; ZENG,

¹ Sentimento textual e tom são considerados, neste estudo, como sinônimos.

2011; HU; MCINISH; ZENG, 2010; KARABULUT, 2013); e previsão dos analistas (TWEDT; REES, 2012).

As principais fontes informacionais do mercado acionário são constituídas por: (a) informações contábeis; (b) previsão dos analistas; e (c) mídia financeira (TETLOCK; SAAR-TSECHANSKY; MACSKASSY, 2008). Ao analisar a contribuição dessas informações para o investidor, Tetlock, Saar-Tsechansky e Macskassy (2008) explicam que os dados provenientes da contabilidade e da previsão dos analistas são informações que podem ser incompletas na apresentação do ambiente financeiro, além de serem tendenciosas, e, com isso, ressaltam a importância da utilização de fontes informacionais qualitativas, baseadas em textos, como as informações provenientes da mídia especializada em finanças.

Pode-se dizer que a mídia financeira é uma das grandes responsáveis pela disseminação de informação em massa. Consequentemente, para o mercado acionário, ela desempenha um papel fundamental, atuando, principalmente, na redução da assimetria informacional e na minimização dos problemas de agência, naturalmente existentes entre investidores e gestores (KOTHARI; LI; SHORT, 2009; TETLOCK, 2010, 2014; ENGELBERG; REED; RINGGENBERG, 2012; LIU; McCONNELL, 2013; AHERN; SOSYURA, 2015). Acredita-se que o conteúdo da mídia financeira tem a capacidade de fornecer informações aos investidores, possibilitando a atualização de suas diferentes crenças e influenciando os movimentos dos preços (DOUGAL et al., 2012; ENGELBERG, REED; RINGGENBERG, 2012; TETLOCK, 2014; STRAUß; VLIEGENTHART; VERHOEVEN, 2016; MANELA; MOREIRA, 2017). Além disso, supõe-se que a mídia tem um menor incentivo para a realização de gerenciamentos de impressão a favor das organizações, pois não está diretamente atrelada a elas.

Dentre as informações veiculadas na mídia financeira, as informações financeiras configuram-se como uma importante fonte informativa, visto que fornece uma perspectiva sobre o estado da economia, possibilitando a fundamentação de decisões de investimento, produção e consumo. Em decorrência dessa importância, surgiram numerosos estudos voltados para a análise do impacto que as notícias financeiras desencadeiam sobre o mercado financeiro (TETLOCK, 2007; TETLOCK; SAAR-TSECHANSKY; MACSKASSY, 2008; GARCÍA, 2013).

A literatura que trata sobre o conteúdo informacional das notícias tem investigado se o mercado acionário reage ao volume e ao sentimento das notícias. Para a realização dessas investigações, são utilizadas várias hipóteses, que fazem alusão à suposição de que os investidores seriam sensíveis ao conteúdo informacional das notícias (ENGELBERG;

PARSONS, 2011; GARCÍA, 2013; LIU; McCONNELL, 2013). Evidências Empíricas mostram indícios da sensibilidade dos investidores às notícias, ao fazerem inferências relacionadas à dinâmica do mercado, constatando que o sentimento das notícias da mídia tem valor preditivo sobre os retornos das ações, na volatilidade e sobre os volumes de negociações (TETLOCK, 2007; TETLOCK; SAAR-TSECHANSKY; MACSKASSY, 2008; GARCÍA, 2013; LIU, 2014; FERGUSON et al., 2015).

O respaldo teórico que explica a reação do mercado ao sentimento textual das notícias está atrelado ao teorema da concordância (AUMANN, 1976), à Hipótese do Mercado Eficiente - HME (FAMA, 1970) e ao viés da negatividade (PEETERS; CZAPINSKI, 1990; ROZIN, ROYMAN, 2001). Desse modo, compreende-se que as crenças dos investidores racionais são atualizadas mediante o acesso à informação, seguindo o *teorema de bayes*, e as expectativas geradas sobre o risco e sobre o retorno dos ativos são refletidas nos preços, conforme prega a HME.

No entanto, os investidores apresentam falhas no processo de ajustes das crenças, associadas a uma limitação da sua racionalidade, comprometendo, assim, o ajuste dos preços das ações diante das informações disponíveis (KAHNEMAN; TVERSKI, 1979). Essa explicação vai de encontro à Hipótese do Mercado Eficiente, proposta por Fama (1970), ao considerar que existe uma defasagem no tempo para incorporação total da informação nos preços, provocada pela presença de investidores desinformados e por restrições cognitivas, como o viés de negatividade.

Nesse sentido, a compreensão do processo de absorção das informações no mercado configura-se como uma importante discussão teórica e empírica, e está fundamentalmente atrelada à eficiência dos mercados e ao viés de negatividade (PEETERS; CZAPINSKI, 1990; ROZIN, ROYMAN, 2001; BAUMEISTER, BRATSLAVSKY, FINKENAUER, VOHS, 2001; ALVES, KOCH, UNKELBACH, 2017).

Os estudos que se propuseram a investigar esse processo de incorporação das informações nos preços apontam para possíveis ineficiências temporais do mercado, tais como o estudo de Jegadeesh e Titman (1993), em que foram analisadas as estratégias de aquisição de ações ganhadoras e perdedoras, encontrando evidências de reações retardadas no processo de ajuste dos preços, sustentado a suposição de uma possível defasagem temporal no processo de transmissão das informações. No tocante ao Brasil, Barbosa e Medeiros (2007) investigaram os choques de notícias favoráveis e desfavoráveis, verificando reações exageradas no mercado

para ambos os choques, levantando indícios de ineficiências temporais no mercado acionário brasileiro.

Dentre os estudos voltados para a análise do conteúdo informacional da mídia financeira, que adotaram como *proxy* para a notícia o sentimento textual, Tetlock (2007) foi um dos precursores. Em sua pesquisa, Tetlock (2007) procurou analisar o efeito do pessimismo de uma coluna financeira presente no *Wall Street Journal* sobre as atividades do mercado acionário, encontrando evidências de que níveis elevados de pessimismo previam quedas nos preços, mas que eram revertidos em curtos intervalos temporais.

Outro importante estudo que procurou analisar os efeitos do conteúdo da mídia sobre o mercado acionário foi realizado por García (2013). Seguindo as evidências levantadas por Tetlock (2007) e embasado pela literatura de economia e da psicologia, García (2013) levantou indícios de que a maior reação do mercado ao conteúdo da mídia era concentrada nos períodos de recessão, visto que, nos momentos de instabilidade econômica, os investidores estariam mais vulneráveis a serem movidos pelo sentimento das notícias da mídia. Assim, constatou que oscilações no sentimento das notícias previam mudanças no retorno do *Dow Jones Industrial Average* (DJIA), sendo tais mudanças mais acentuadas nos períodos de recessão.

Já o estudo de Liu (2014), concentrou-se na análise do conteúdo informacional de notícias específicas sobre finanças públicas que saíram ao longo da crise Europeia (2009-2012), para investigar o poder da mídia sobre o mercado de títulos públicos da Europa, constatando que o pessimismo da mídia e o número de notícias ajudavam na previsão dos *spreads* dos títulos avaliados.

Na pesquisa de Bianconi, Hua e Tan (2015), foi avaliado o efeito das notícias financeiras sobre o risco sistêmico de instituições financeiras, porém, os resultados levantados sugerem que o sentimento das notícias não é uma medida capaz de predizer o risco extremo das instituições. Corroborando com as investigações dentro do âmbito de empresas, Ahmad et al. (2016) exploraram notícias divulgadas pela mídia sobre as maiores empresas americanas, verificando indícios de que o tom das notícias por longos períodos não provocava efeitos sobre o retorno dos ativos, e, ocasionalmente, quando ocorriam impactos significativos, estes eram revertidos rapidamente.

Nesse sentido, esta pesquisa está relacionada à literatura que busca compreender o papel da mídia no mercado financeiro, ao procurar avaliar o efeito do sentimento das notícias sobre o comportamento dos preços no mercado acionário brasileiro com o intuito de verificar o impacto

da mídia sobre o mercado de um país emergente, visto que são levantados indícios de que a mídia em países emergentes não seria sofisticada ao ponto de provocar reações no mercado (GRIFFIN, HIRSCHEY; KELLY, 2011; FRAIBERGER, 2016).

O presente estudo apresenta algumas características que o distinguem dos estudos já realizados neste âmbito. Em primeiro lugar, foram utilizadas as notícias financeiras divulgadas em jornal especializado em discussões financeiras, do qual foram extraídas as matérias divulgadas na seção de finanças, sem fazer distinções quanto à coluna de publicação, ao contrário de Tetlock (2007) e García (2013), que analisaram apenas as notícias de colunas específicas de jornais. Pretendeu-se, com isso, avaliar um maior número de informações, que tendem a provocar impactos sobre as atividades do mercado.

Em segundo lugar, foi avaliado o mercado acionário de um país emergente, respeitando o idioma oficial do país, visto que existem poucos estudos explorando o papel da mídia financeira em mercados emergentes e, com isso, pouco se sabe sobre o valor informacional das notícias para esses mercados, uma vez que a maioria dos estudos concentra-se em investigar mercados desenvolvidos, como os Estados Unidos (TETLOCK; SAAR-TSECHANSKY; MACSKASSY, 2008; DOUGAL et al., 2012; ENGELBERG, REED; RINGGENBERG, 2012; LIU; MCCONNELL, 2013; CARRETTA et al., 2013; AHERN; SOSYURA, 2015; AHMAD et al., 2016).

Em terceiro lugar, foi explorado *big data* de notícias financeiras divulgadas em jornais na língua portuguesa, a fim de identificar padrões de comportamento atrelados aos movimentos de oscilação do mercado, quanto ao risco (volatilidade) e ao retorno, que possam contribuir com o gerenciamento de risco e a otimização da tomada de decisão no mercado acionário do Brasil.

Em quarto lugar, foi verificado o efeito do sentimento das notícias financeiras ao longo de toda a distribuição do retorno e da volatilidade do mercado, por meio do método de regressão quantílica, com o intuito de verificar a existência de comportamentos distintos da informação, conforme a euforia do mercado. A necessidade dessa investigação decorre dos questionamentos que recentemente tem surgido na literatura quanto ao valor informacional que, de fato, as notícias teriam para explicar os movimentos de oscilação do mercado. Desde a obra de Keynes (1936), são levantadas especulações sobre o que moveriam os preços. Parte das explicações atribui à mídia um enorme impacto sobre os movimentos dos preços.

No entanto, Tetlock (2014), ao analisar o estado da arte da transmissão das informações no mercado financeiro, mostra que, apesar da existência de impactos da notícia sobre os movimentos do mercado, eles são baixos em termos de magnitude. García (2013) corrobora empiricamente com esse pensamento, ao levantar que, apesar da existência do impacto, o efeito não é persistente, e ele tende a desaparecer nos dias seguintes. Os achados empíricos de Bianconi, Hua e Tan (2015) e Ahmad et al. (2016) também lançam dúvidas quanto à capacidade informacional do sentimento textual em prever o risco e o retorno das ações.

Portanto, diante do contexto apresentado, despertou-se o anseio de estudar o reflexo do sentimento textual no mercado acionário brasileiro. Desse modo, a pesquisa visou responder a seguinte indagação: *Qual o efeito do sentimento textual das notícias financeiras sobre o comportamento dos preços no mercado acionário brasileiro?*

1.2 OBJETIVOS

1.2.1 Objetivo Geral

Investigar o efeito do sentimento textual das notícias financeiras sobre o comportamento dos preços no mercado acionário brasileiro.

1.2.2 Objetivos Específicos

Para alcançar o objetivo geral, foram estabelecidos os seguintes objetivos específicos:

- Propor uma lista de termos financeiros para a categorização das notícias da mídia financeira, visando à construção de um instrumento de mensuração do sentimento das notícias;
- Estimar um índice de sentimento baseado em notícias financeiras;
- Verificar a influência que o sentimento textual das notícias exerce sobre o comportamento dos preços no mercado acionário brasileiro;
- Analisar o poder de previsão do risco (volatilidade) do mercado, com a incorporação da informação relacionada ao sentimento das notícias.

1.3 A TESE

As notícias, de modo geral, contribuem com a atualização das crenças dos investidores sobre os valores fundamentais das organizações, possibilitando um alinhamento das expectativas futuras sobre os ativos, e esse comportamento é refletido no mercado, no processo de ajuste dos preços às informações. Admitindo que esse alinhamento das expectativas tem um sentido *bayesiano* e pode ser explicado segundo o Teorema da Concordância de Aumann (1976).

Ao tratar os investidores como agentes racionais, espera-se que o acesso às informações possibilite uma atualização convergente das crenças, em função de apresentarem expectativas homogêneas. Com isso, os preços no mercado revelariam as crenças dos investidores e, portanto, os preços seriam o melhor sistema informacional para aqueles que operam no mercado.

No entanto, os agentes estão vulneráveis a desvios cognitivos, os quais limitam a sua racionalidade, levando o ajuste das crenças a não seguirem a prescrição do modelo racional, conforme pode ser observado na Figura 1. Assim, os agentes ficam sujeitos a apresentarem divergências no processo de atualização das suas crenças, em decorrência da racionalidade limitada, atrelada à influência de fatores psicológicos, como o viés da negatividade.

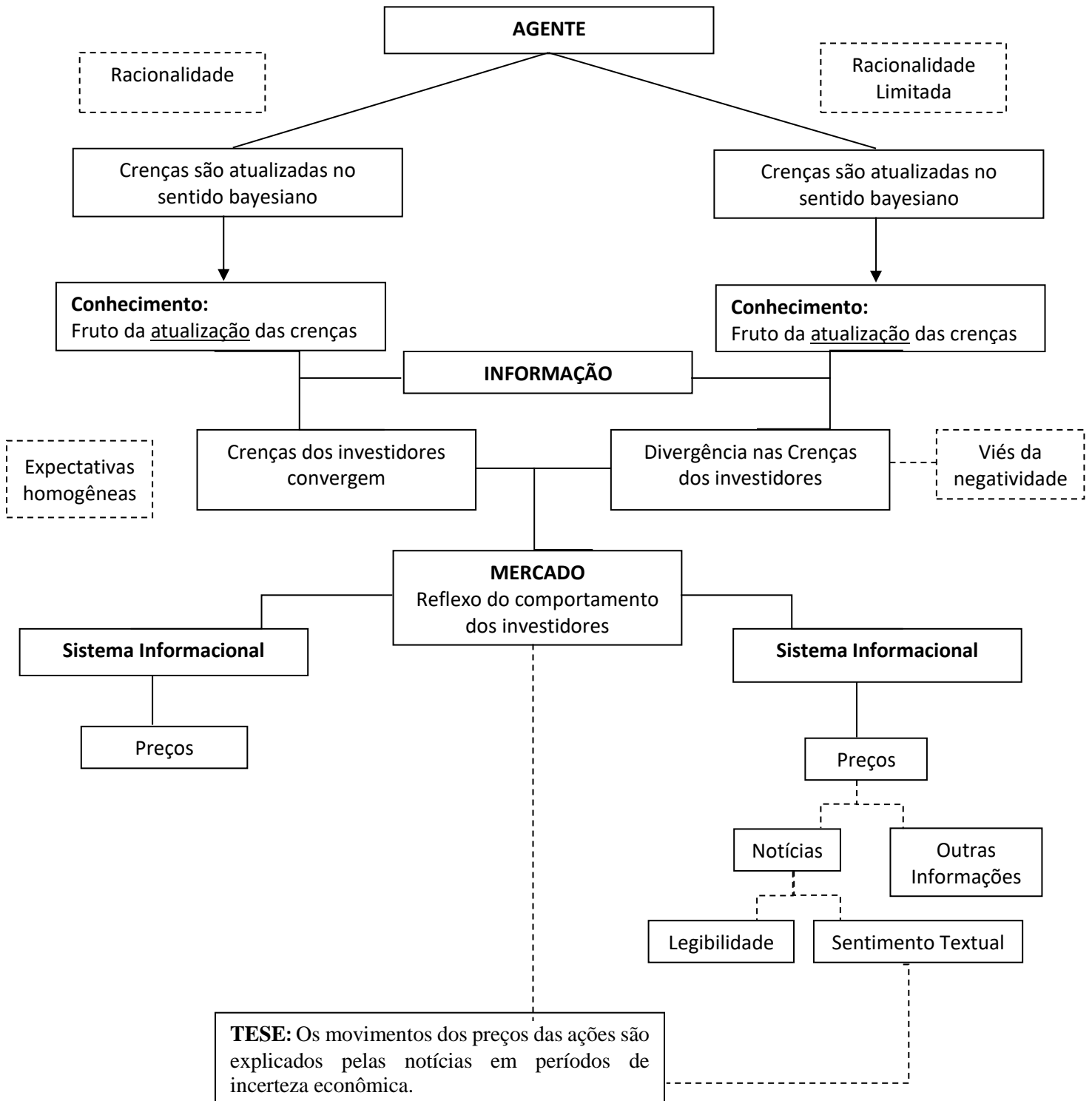
Desse modo, a falta de alinhamento das crenças provoca ruídos no processo de ajuste dos preços às informações, comprometendo a atualização dos preços e provocando ineficiências temporais no mercado, com reações *underreaction* e *overreaction*. Em função disso, os investidores buscam por informações que apresentem conexões com preços e que possam explicar os movimentos de oscilação do mercado, visto que os preços tendem a não incorporar as informações disponíveis. Dentre o conjunto informacional que respalda o processo de tomada de decisão dos agentes encontram-se as notícias da mídia financeira, que correspondem a uma importante fonte de informações para o mercado, por atuar na disseminação de informações e por apresentar capacidade de influenciar os agentes, principalmente em períodos que remetem a maior uma incerteza econômica.

As notícias podem ser analisadas sob diferentes perspectivas. No âmbito do sentimento textual, o conteúdo informacional analisado é o posicionamento de otimismo ou pessimismo adotado na construção da notícia. Portanto, considerando que as notícias contribuem com o processo de atualização das crenças dos indivíduos e que o mercado apresenta ineficiências no

processo de ajuste dos preços as informações, a tese proposta é a de que: Os movimentos dos preços das ações são explicados pelas notícias em períodos de maior incerteza econômica.

A Figura 1 apresenta uma síntese dos principais aspectos que levam à proposta de tese.

Figura 1 - Desenho da Pesquisa



Fonte: Elaborado pelo autor.

1.4 ORIGINALIDADE E JUSTIFICATIVA

Os avanços tecnológicos têm proporcionado a difusão de uma variedade de informações, cada vez mais rápida, permitindo o seu armazenamento e gerando um imenso volume de dados, que são chamados de *Big Data*. Atualmente, o grande desafio reside no processamento desses dados, para a extração de informações que favoreçam à tomada de decisão. A linguagem de processamento natural tem apresentado soluções, levando em consideração os aspectos cognitivos da leitura humana na construção de algoritmos, possibilitando a leitura mecânica dos textos e favorecendo a extração de informações, como o sentimento textual.

Diante da possibilidade de análises de volumosos bancos de dados textuais, a literatura tem buscado investigar o papel da mídia nos mercados financeiros. No âmbito de estudos relacionados ao sentimento textual, as evidências empíricas ressaltam que o tom dos textos da mídia apresenta relação com as oscilações do mercado, em termos do risco e do retorno dos ativos (TETLOCK, 2007; TETLOCK, SAAR-TSECHANSKY; MACSKASSY, 2008; ENGELBERG; PARSONS, 2011; LOUGHRAN; McDONALD, 2011; ENGELBERG; REED; RINGGENBERG, 2012; DOUGAL et al., 2012; CARRETTA et al., 2013; GARCÍA, 2013; LIU; McCONNELL, 2013; LIU, 2014; HUANG; TEOH; ZHANG, 2014; FERGUSON et al., 2015; AHERN; SOSYURA, 2015; STRAUß; VLIEGENTHART; VERHOEVEN, 2016; FRAIBERGER, 2016).

Apesar das evidências mostrarem que existe relação entre o sentimento textual e a movimentação do mercado, sintetizada geralmente na análise do risco e do retorno, a magnitude dos coeficientes das regressões é pequena, indicando um baixo impacto, levando Tetlock (2014) a questionar o impacto da mídia sobre a movimentação dos preços, sugerindo a realização de estudos que possam aprofundar melhor essa discussão.

Além disso, as investigações, em sua maioria, concentram-se no mercado acionário dos Estados Unidos, que é reconhecido por apresentar uma mídia com ampla cobertura de notícias voltadas para o mercado financeiro (GRIFFIN; HIRSCHHEY; KELLY, 2011). Porém, acredita-se que o sentimento das notícias pode desempenhar um papel importante na compreensão da movimentação dos mercados em outros países desenvolvidos, assim como em países emergentes (KEARNEY; LIU, 2014; LOUGHRAN; MCDONALD, 2016; KHUU, DURAND; SMALES, 2016),

Desse modo, a realização da tese é motivada por: (a) necessidade de explorar *big data* de notícias financeiras, a fim de identificar padrões de comportamento que possam contribuir com o gerenciamento de risco e a otimização da tomada de decisão no mercado acionário brasileiro; (b) necessidade de aprofundar as explicações sobre o impacto que o conteúdo informacional das notícias provoca sobre as atividades do mercado acionário; e (c) necessidade de compreender o papel da mídia de um país emergente para o mercado acionário.

Assim, a presente pesquisa contribui com a literatura, ao: (a) propor uma lista de termos financeiros em português, possibilitando a avaliação do sentimento textual, respeitando as características linguísticas do Brasil, visto que a lista de termos mais utilizada nos estudos é a lista proposta por Loughran e McDonald (2011), porém, encontra-se em inglês. Desse modo, visando evitar uma perda informacional na tradução dos termos, foi proposta uma lista de termos financeiros em português; (b) investigar a relação do sentimento textual nos movimentos do mercado, analisando toda a distribuição do retorno e da volatilidade do mercado, por meio da regressão quantílica, método pouco utilizado em estudos dessa natureza, com a finalidade de detalhar o comportamento e a magnitude dos impactos que as notícias provocam no mercado; e, (c) investigar o comportamento do mercado acionário de um país emergente, como o Brasil, que, apesar de apresentar um mercado pequeno, em termos de volume de negociações e de número de investidores, comparado aos maiores mercados, é um mercado que atrai o ingresso de capitais externos (GAY, 2016).

Em relação ao aspecto de originalidade da pesquisa, o estudo torna-se inédito, na medida em que busca explorar a capacidade do sentimento textual de predizer os movimentos do mercado, buscando explicações sobre o impacto que o tom das notícias exerce, principalmente sobre os momentos de euforia do mercado. Vale salientar que é uma abordagem empírica pouco explorada na literatura internacional e que existe uma escassez de estudos abordando o tema no mercado acionário brasileiro.

Dessa forma, a pesquisa apresenta importância, por buscar um meio de processamento de *big data* de textos em língua portuguesa, visando a identificação de padrões de comportamento atrelados ao sentimento textual, que possam favorecer a compreensão sobre as oscilações do mercado acionário brasileiro.

O estudo é viável, na medida em que existe a disponibilidade dos dados, sejam: (a) textuais, referentes às notícias financeiras divulgadas por jornais, visto que existem acervos digitais para consulta das matérias, como o Jornal Valor Econômico; e (b) dados do mercado acionário, que são de domínio público, pois a Comissão de Valores Mobiliários (CVM) obriga

a divulgação das informações financeiras. Além disso, a LNP possibilita trabalhar com um amplo volume de textos, os quais foram examinados com análises de conteúdo computadorizadas, mediante a utilização de algoritmos que realizam a leitura mecânica dos textos, conforme *scripts* de programação divulgados no apêndice.

Desse modo, o trabalho contribui teoricamente com a discussão sobre o papel da mídia no mercado financeiro de países emergentes, pois as evidências sobre o impacto das notícias da mídia, nesse tipo de mercado, ainda são inconclusivas. Griffin, Hirschey e Kelly (2011), ao analisar as notícias da mídia de 56 países, levantaram indícios dos países apresentarem comportamentos divergentes, assim como Fraiberger (2016), que também constatou comportamentos distintos, ao analisar a influência do tom das notícias na previsibilidade do produto interno bruto (PIB) de alguns países. Nesse sentido, Tetlock (2014), Kearney e Liu (2014) e Loughran e McDonald (2016) demonstram, em seus estudos, que existe um grande interesse na literatura por investigações que explorem o papel da mídia financeira no mercado financeiro, levando em consideração a língua oficial dos países.

No que diz respeito às contribuições práticas, acredita-se que os resultados beneficiarão os agentes do mercado, no âmbito referente às explicações sobre a conexão do sentimento textual das notícias com a dinâmica de oscilação dos preços dos ativos, de modo que os investidores poderão analisar a viabilidade de fazer uso do sentimento textual das notícias, dentro das estratégias de investimento, visando minimizar a sua exposição ao risco.

1.5 LIMITAÇÕES DA PESQUISA

A elaboração do trabalho encontra-se limitada aos seguintes aspectos:

- Algoritmo de leitura mecânica – a extração do sentimento das notícias foi realizada com base em programação, levando em consideração a cognição humana na segregação de palavras que remetem a aspectos positivos e negativos. Porém, os algoritmos utilizados não tem o objetivo de replicar o procedimento de leitura humana, visando a sua substituição. Os algoritmos configuram-se como uma ferramenta para extração de informação de *Big Data* relacionados a dados textuais.
- Índice de sentimento – a mensuração do índice de sentimento foi realizada com base no dicionário proposto nesta tese, composto por uma lista de termos positivos e negativos, os quais poderão sofrer alterações ao longo do tempo, visto que o processo de comunicação muda, com

a incorporação e exclusão de palavras. Logo, os resultados do índice de sentimento deste estudo são restritos às listas de termos proposta.

- Amostra – o estudo investigou as matérias do Jornal Valor Econômico, por levar em consideração a sua relevância para o mercado financeiro na disseminação de informações, conforme dados do Instituto Verificador de Comunicação – IVC (2017). Desse modo, as conclusões do estudo ficam restritas às notícias divulgadas nas matérias do Jornal Valor Econômico.
- Período avaliado – para a elaboração desta tese, foi considerado um período de seis anos, restrito aos anos de 2011 a 2017, em função de ser o período disponibilizado para a consulta do acesso digital do Jornal Valor Econômico.

1.6 ESTRUTURA DO TRABALHO

O trabalho encontra-se dividido em cinco seções, sumarizadas da seguinte forma: i) a primeira trata da apresentação da problematização, tese, objetivos, justificativa e limitações; ii) a segunda trata da fundamentação teórica e das evidências empíricas sobre sentimento textual; iii) a seção três trata dos aspectos metodológicos, com a apresentação dos modelos e a descrição dos dados utilizados na execução do estudo; iv) a seção quatro apresenta a análise dos resultados; e, por fim, v) a seção cinco apresenta a conclusão do trabalho. Além das cinco seções, o trabalho apresenta as referências, o apêndice, com a apresentação dos algoritmos utilizados no estudo, bem com o dicionário proposto, e anexos.

2 REVISÃO DA LITERATURA

O propósito do capítulo é apresentar os fundamentos teóricos que respaldam a discussão sobre sentimento textual, e as principais evidências empíricas que sustentam as hipóteses da pesquisa realizada.

2.1 TEOREMA DA CONCORDÂNCIA

O teorema da concordância, apresentado no estudo seminal de Aumann (1976), explica o processo da interação racional dos indivíduos, seguindo o teorema de *bayes* e em meio ao conhecimento comum, que leva ao ajuste das crenças. Com isso, é apresentada uma modelagem de como o conhecimento dos indivíduos é estruturado.

Assim, dado um espaço de probabilidade (Ω, B, p) , em que Ω é o espaço amostral ou espaço de estados, B representa os eventos e P corresponde à probabilidade dos eventos, também chamado de partições dos indivíduos P_i , que representa o conhecimento que um dado indivíduo i apresenta sobre a ocorrência de um estado ω . Então, para cada estado $\omega \in \Omega$, existirá uma única partição representada pelo elemento $P_i(\omega)$ do indivíduo P_i , logo, ao ocorrer o estado ω , o indivíduo terá conhecimento apenas sobre $P_i(\omega)$ da partição. Desse modo, um estado ω do espaço Ω terá um evento E com conhecimento comum, quando as partições dos indivíduos P_1 e P_2 , que estão contidas em ω , tornarem-se públicas.

Considerando que a atualização do conhecimento dos indivíduos é dada no sentido *bayesiano*, e antes que um evento ocorra os indivíduos terão as suas respectivas crenças sobre o estado de um evento, que é a probabilidade *a priori*, e que, diante da ocorrência de um evento essa probabilidade inicial é ajustada, em função da partição do indivíduo $P_i(\omega)$, sendo chamada de probabilidade *a posteriori*. Assim, dada a ocorrência de um evento A , a probabilidade *a posteriori* $P(A|P_i(\omega))$ do evento será dada no sentido *bayesiano*, conforme Equação 1:

$$P(A|P_i(\omega)) = \frac{P(A \cap P_i(\omega))}{P(P_i(\omega))} \quad (1)$$

A Equação 1 evidencia a lógica de como os indivíduos atualizariam o seu conhecimento em termos de probabilidade, ou seja, a crença inicial sobre o evento A expresso pela probabilidade *a priori* $P(A)$ está condicionado a informação obtida com a realização do evento $P_i(\omega)$. Desse modo, a probabilidade *a posteriori* representa a atualização do processo de conhecimento sobre um evento. Assim, considerando que q é a probabilidade *a*

posteriori $P(A|P_i(\omega))$ de um dado evento, e que existe um conhecimento comum perfeito entre os indivíduos, ou seja, cada indivíduo conhece todas as partições individuais de um estado ω , a proposição de Aumann (1976) é que:

Proposição: Seja $\omega \in \Omega$, e dado q_1 e q_2 como as probabilidades numéricas, se ω é de conhecimento comum perfeito, então, $q_1 = q_2$.

Em outras palavras, a proposição do teorema de concordância mostra que, se duas pessoas, 1 e 2, conhecem um dado evento, e se ambas têm conhecimento sobre essa situação, de que a pessoa 1 sabe o que a pessoa 2 sabe, e a pessoa 2 sabe o que o indivíduo 1 sabe, e a pessoa 1 sabe o que o indivíduo 2 conhece sobre o que a pessoa 1 sabe, e assim sucessivamente. Desse modo, se os indivíduos apresentarem os mesmos antecedentes em relação às suas crenças, e se o seu conhecimento sobre um dado evento tornar-se conhecimento comum, então, os indivíduos não poderão ser levados a concordar em discordar um do outro.

No contexto do mercado financeiro, o teorema da concordância, apresentado por Aumann (1976), pode ser utilizado na compreensão de como os agentes atualizam as suas crenças em meio às informações disponíveis no mercado. O estudo de Milgrom e Stokey (1982) procura fazer essa relação entre a lógica do teorema da concordância no mercado financeiro, mostrando que, na presença de expectativas racionais e diante de um ambiente no qual as informações sobre as negociações são de conhecimento comum, os agentes não teriam incentivos para realizar especulações, mesmo em posse de informações privadas, pois os preços no mercado revelariam qualquer alteração informacional. Nesse sentido, os preços configuram-se como um sistema informativo, possibilitando a atualização das expectativas dos agentes do mercado.

A perspectiva dos preços como um sistema informacional perfeito, conforme é apresentado por Milgrom e Stokey (1982), seria possível se o mercado detivesse apenas investidores informados. Porém, o mercado apresenta investidores desinformados, que provocam ruídos no sistema de preços, por não compreenderem ou por não atualizarem o seu conhecimento, diante das oscilações provocadas nos preços decorrentes das negociações dos investidores informados (GROSSMAN; STIGLITZ, 1980). Em função disso, os preços podem não revelar completamente as crenças dos agentes, em função de falhas no processo de incorporação das informações junto aos preços, levando os investidores a procurarem por informações que expliquem a movimentação do mercado (TETLOCK, 2014).

Além da existência de investidores desinformados no mercado, um outro aspecto que provoca ruídos na atualização dos preços está relacionado ao modo como os indivíduos processam as informações. Hirshleifer e Teoh (2003) apontam que, em geral, os investidores apresentam um processamento imperfeito das informações publicamente disponíveis. Nesse sentido, acredita-se que o processamento e interpretação das informações sejam influenciados por restrições cognitivas, as quais levam ao comprometimento na revisão das crenças e provocam falhas no processo de tomada de decisão (CAMPBELL; LOUMIOTI; MOERMAN, 2017).

Estudos na área de psicologia mostram que o sistema perceptivo dos indivíduos não apresenta um processamento informacional simétrico, quando as pessoas avaliam se uma informação é positiva ou negativa, levando-as a apresentarem reações distintas. Em geral, os indivíduos tendem a ponderar mais os eventos que remetem a emoções negativas, devido à percepção de que os maus eventos apresentam uma pressão psicológica maior que os bons eventos (PEETERS; CZAPINSKI, 1990; ROZIN, ROYMAN, 2001; BAUMEISTER et al., 2001; ALVES, KOCH, UNKELBACH, 2017).

Isso não quer dizer que o mau sobrepeça-se ao bom, quer dizer que eventos ruins apresentam um impacto psicológico maior que o impacto provocado pelos eventos bons. Para Alves et al. (2017), essa ênfase atribuída aos eventos negativos ocorre em função da ameaça que eles podem provocar sobre o bem-estar, assim, os indivíduos naturalmente atribuem uma maior importância aos eventos ruins, como um comportamento cognitivo de defesa, em decorrência da possível perda potencial, explicado por Rozin e Royman (2001), como um viés de negatividade.

As evidências sobre a influência que a negatividade exerce sobre os indivíduos podem ser observadas no processo de linguagem humana, em que existem indícios de um vocabulário muito mais rico em termos utilizados para descrever fenômenos negativos, comparado ao vocabulário de termos positivo (PEETERS, 1971; ROZIN, ROYMAN, 2001; BAUMEISTER et al., 2001), sugerindo, com isso, que a cognição humana para os aspectos negativos possa ser mais complexa e elaborada (PEETERS; CZAPINSKI, 1990; ROZIN, ROYMAN, 2001; ALVES et al., 2017).

Esse viés de negatividade também é encontrado no processo de decisão dos indivíduos, no qual as pessoas costumam atribuir uma maior importância para as perdas que para os ganhos, devido a aversão à perda que elas apresentam, que é a discussão central da Teoria do Prospecto proposta por Kahneman e Tversky (1979).

No mercado acionário, os agentes tendem a atribuir uma maior atenção para os eventos negativos, por exemplo, Allen, McAleer e Singh (2017), ao investigarem o comportamento do índice americano *Dow Jones Industrial Average* (DJIA) em relação ao índice de sentimento da *Thomson Reuters News Analytics* (TRNA), verificaram que o mercado atribui uma notoriedade às más notícias. Liu, McConnell e Xu (2017) levantaram que o tom negativo das notícias da mídia financeira reduz a probabilidade de um diretor executivo, chamado de *Chief Executive Officer* – CEO, assumir assento no conselho de outras diretorias após o encerramento do seu mandato. Cox, Dayanandan e Nofsinger (2017), ao investigar os comunicados de lucros, constataram que as notícias negativas levavam o mercado a reações negativas, porém, quando as notícias negativas ocorriam em períodos de expansão econômica, os preços apresentavam uma reação significativamente maior, em comparação aos períodos de recessão econômica.

A divulgação de informações, segundo Tetlock (2014), deveria levar a uma convergência das crenças dos investidores, porém, são observadas divergências no alinhamento das crenças, em decorrências de aspectos cognitivos, que são refletidas em desacordos na compreensão do processo de precificação dos ativos, lançando dúvidas quanto à eficiência do mercado no ajuste das informações.

2.2 HIPÓTESE DO MERCADO EFICIENTE

A necessidade de explicar o que move os preços no mercado acionário e os anseios por identificar padrões de comportamento atrelados aos preços foram fatores motivadores para a realização de investigações empíricas sobre o comportamento dos preços no mercado, de modo que o acúmulo dessas evidências empíricas culminou no surgimento de explicações teóricas sobre mercado eficiente.

Uma das primeiras contribuições para a compreensão da eficiência do mercado foi apresentada por Bachelier (1900), que procurou analisar o comportamento dos preços das opções no mercado francês, buscando respaldo teórico na descrição de um processo de especulação e na teoria da probabilidade para explicar a movimentação dos preços dos ativos.

Segundo Bachelier (1900), no mercado acionário, o investidor não apresenta posição de vantagem ou desvantagem diante do conhecimento das informações passadas, pois, o movimento dos preços seria um reflexo da especulação do investidor. Assim, a expectativa do retorno de um especulador seria nula, remetendo a um *fair game*.

Ao relacionar o movimento dos preços dos ativos a um processo de especulação dos investidores, Bachelier (1900) formulou uma das primeiras ideias sobre o comportamento aleatório dos preços ou *random walk* (FAMA, 1970). Nesse sentido, considerar que os preços são aleatórios significa dizer que não é possível prever o comportamento futuro dos preços utilizando informações históricas, visto que os preços já incorporaram essa informação.

Outro estudo que contribuiu para a fundamentação de uma teoria sobre os mercados eficientes foi o estudo de Osborne (1959), no qual foram levantados indícios de que os preços dos ativos da *New York Stock Exchange* (NYSE) apresentavam uma independência nos seus valores, em função da existência de variáveis aleatórias, as quais seriam responsáveis por provocar alterações nos preços. Com isso, no processo de tomada de decisão, os investidores fariam suas escolhas, levando em consideração a expectativa do valor esperado para aquela transação, existindo uma independência nas transações futuras. Em função disso, os preços apresentariam um comportamento aleatório.

Até o final da década de 1950, os estudos que tratavam sobre análises do comportamento dos preços e a sua capacidade preditiva, em geral, procuravam respaldo na modelagem *fair game*. No entanto, tais estudos não exprimiam rigor nas explicações teóricas e apresentavam pressupostos fortes quanto à independência na tomada de decisão, sem que existissem explicações para admitir que os indivíduos tomavam decisões sem sofrer influências de outros investidores (FAMA, 1965, 1970).

Nesse contexto de críticas, Samuelson (1965) procurou estudar o comportamento dos preços, a partir de observações relacionadas a *commodities* do trigo. Ao longo do estudo, Samuelson (1965) levanta uma série de indagações sobre possíveis interferências dos aspectos sazonais, intrínsecos ao processo produtivo do trigo, nos preços das suas *commodities* e, com isso, questiona em que medida as séries de preços apresentariam aleatoriedade. Diferentemente das modelagens adotadas na literatura anterior, que procuravam explicar o comportamento dos preços, por meio do *fair game*, considerando que a expectativa do retorno seria nula, Samuelson (1965) explica que o comportamento dos preços passados não prediz os preços futuros, por meio da modelagem *martingale*, na qual a expectativa do retorno de um investimento é que o preço seja igual ou maior ao preço atual.

Fama (1965), ao estudar o comportamento dos preços de 30 ações integrantes da *Dow Jones Industrial Average* (DJIA), no período de 1957 a 1962, encontrou, nos retornos diários dos ativos analisados, uma autocorrelação positiva de primeira ordem, em 23 ativos dos 30 analisados, mas, em virtude das variações serem pequenas, do ponto de vista econômico,

considerou que o comportamento dos preços remetia a um *random walk*. Assim, é ressaltado na obra que essa não seria uma dependência importante para as avaliações de investimento, visto que o comportamento dos preços passados das ações não poderia ser utilizado como estratégia para aumentar os lucros dos investidores. No entanto, a dependência encontrada poderia ser uma informação importante para a compreensão do processo de determinação dos preços no mercado de capitais.

Diante das evidências empíricas sobre a aleatoriedade dos preços dos ativos, Fama (1970) realizou uma revisão da literatura teórica e empírica sobre a eficiência do mercado, expondo as principais limitações dos modelos utilizadas até aquele período. Assim, partindo das ideias difundidas nas modelagens *fair game* e *martingale*, Fama (1970) propõe uma teoria sobre os mercados eficientes, chamada de Hipótese do Mercado Eficiente, defendendo que os preços refletem a qualquer momento todas as informações disponíveis.

Nesse sentido, o ajuste dos preços, segundo Fama (1970), ocorre mediante três diferentes níveis ou subconjuntos de informações, que são categorizados como segue: (a) forma fraca – conjunto de informações passadas dos preços; (b) forma semiforte - informações de domínio público; e (c) forma forte – informações de domínio público e informações privilegiadas.

A análise da eficiência do mercado, segundo Fama (1970), pressupõe as condições de que: (a) não há custos de transação nas negociações dos títulos; (b) todas as informações estão disponíveis para todos os participantes do mercado; e que (c) todos os investidores apresentam expectativas homogêneas quanto aos efeitos da informação relacionada aos preços.

Em sua essência, a HME admite que os indivíduos apresentam racionalidade nas suas decisões econômicas, visando uma maximização da utilidade. Desse modo, Fama (1970), ao pressupor que os investidores são racionais, admite que as novas informações sejam sinais para ajustar as crenças dos investidores sobre a expectativa do fluxo de caixa futuro do ativo, seguindo o teorema de *bayes*. Com isso, quando a expectativa é frustrada, os investidores reagem no mercado, comprando ou vendendo os ativos, ajustando os preços às novas informações. Assim, as informações estariam refletidas dentro de um curto espaço de tempo nos preços e o mercado seria eficiente, por não ser possível utilizar a informação para obter ganhos anormais.

Com isso, a eficiência do mercado está diretamente relacionada à ideia de um mercado ideal, no qual os preços fornecem sinais precisos sobre a escassez dos recursos, possibilitando

a tomada de decisão sobre os investimentos, visto que os preços refletem as propriedades reais da empresa (FAMA, 1970).

No entanto, a definição da eficiência do mercado é generalizada e remete a uma hipótese sobre o possível comportamento dos preços, de modo que a eficiência do mercado não é em si empiricamente testável. Assim, as investigações sobre os preços em equilíbrio são realizadas em termos de análises sobre o retorno esperado, mediante a utilização de modelos de precificação de ativos (FAMA, 1970, 1991).

Diante da formulação de uma teoria sobre mercados eficientes e mediante as suas limitações conceituais, surgem questionamentos quanto à suposição da racionalidade, culminando em uma ruptura das convicções de que os investidores apresentariam racionalidade ilimitada, que perdurou até a década de 1970.

Com isso, surgem estudos comportamentais defendendo a racionalidade limitada dos investidores, como o estudo de Kahneman e Tversky (1979), que discorre sobre a influência dos aspectos psicológicos (desvios cognitivos) na tomada de decisão, levando a limitações no processo de escolha racional, visto que as pessoas seriam avessas não só ao risco, mas também à perda.

No contexto do mercado acionário, os estudos comportamentais, apesar de criticarem teoricamente e empiricamente a HME, não explicam o que seria o mercado em meio a uma racionalidade limitada. Com isso, os estudos, em geral, procuram invalidar ou fazer adaptações à HME (GROSSMAN; STIGLITZ; 1980; BARBOSA; MEDEIROS, 2007).

Grossman e Stiglitz (1980) fazem uma reflexão sobre a ineficiência do mercado, apontando problemas teóricos na HME. Segundo os autores, as informações não estariam refletidas nos preços a todo o momento, em decorrência dos custos de arbitragem e da presença de informações privadas no mercado. Assim, apenas uma parte das informações estaria refletida nos preços, as quais seriam oriundas dos investidores conhecedores dos dados, aqueles que, para obter retornos anormais, arcariam com os custos para obtenção de informações que não são de domínio público. Logo, os investidores, informados ou não, em média, deixariam de apresentar expectativas homogêneas, levando a um desequilíbrio no mercado e, consequentemente, à sua ineficiência.

Fama (1991), ao analisar críticas teóricas à HME, assevera que os custos de arbitragem e os custos para a obtenção de informações não estariam entre os problemas mais graves. De acordo com o autor, um dos maiores obstáculos para a realização das inferências sobre a

eficiência do mercado seria o problema de hipótese conjunta, ligada ao fato da eficiência do mercado não ser empiricamente testável, sendo necessária a utilização de modelos de precificação de ativos. Com isso, a eficiência do mercado e os modelos de precificação são testados conjuntamente. Em função disso, as conclusões empíricas são ambíguas, de forma que, quando são encontrados retornos anormais, não é possível afirmar se o resultado é decorrente da ineficiência do mercado ou se é em função de um mau modelo de precificação.

Na literatura, são encontrados estudos empíricos sobre o comportamento dos preços, mostrando indícios de que não se ajustariam imediatamente às novas informações (JEGADEESH; TITMAN, 1993; MOSKOWITZ; GRINBLATT, 1999; SCHNUSENBERG; MADURA, 2001; HONG; TOROUS; VALKANOV, 2007; BARBOSA; MEDEIROS, 2007; TETLOCK, 2011; FENG; HU, 2014; SINHA, 2016).

Jegadeesh e Titman (1993) analisaram o movimento dos preços ligados às estratégias de compra de ações que obtiveram bons resultados (ações vendedoras) e venda de ações que apresentaram maus resultados (ações perdedoras). Desse modo, eles investigaram um comportamento anômalo para a HME, denominado na literatura de efeito momento. Assim, utilizando dados da NYSE e da *American Stock Exchange* (AMEX), entre os anos 1965 a 1989, analisaram 16 estratégias de negociação em períodos de 12, 24 e 36 meses. Os resultados empíricos mostraram uma correlação positiva na série de retornos históricos com os retornos presentes, indicando que a estratégia de comprar ações vencedoras e vender ações perdedoras gerava retornos anormais positivos, de modo que as carteiras formadas após o acompanhamento do retorno de seis meses geravam um retorno acumulado médio de 9,5% ao longo dos doze meses seguintes ao investimento. No entanto, o retorno caía para menos da metade nos 24 meses seguintes ao investimento, chegando a 4% no último mês. Portanto, chegaram a constatação de que a existência de retornos anormais, por meio dessas estratégias de investimento, seria proveniente de uma reação retardada na incorporação de informações específicas sobre o anúncio dos lucros da organização nos preços dos seus ativos financeiros.

Moskowitz e Grinblatt (1999) estudaram o efeito momento na indústria, analisando estratégias de investimento com ações vencedoras e perdedoras de 20 carteiras de investimento com ativos provenientes da indústria, com negociação na NYSE, AMEX e Nasdaq, no período de 1963 a 1995. Assim, foi analisado o retorno obtido entre o período de seis meses a um ano, após a formulação da estratégia, encontrando evidências de que existia uma persistência nos retornos de investimentos. As carteiras analisadas apresentavam correlação serial positiva, principalmente no curto prazo, o que foi associado a uma reação retardada dos preços na

incorporação de novas informações, provocada, sobretudo por empresas pequenas, nas quais as notícias levariam certo tempo para serem difundidas entre outras empresas, entre os analistas e investidores, provocando, conseqüentemente, um efeito *lead-lag* entre os participantes do mercado, repercutido nos preços dos ativos. Além disso, os autores levantam que esse comportamento estaria relacionado à confiança que os investidores poderiam apresentar por certos tipos de indústrias, fazendo com que o ajuste às novas informações fosse lento.

Schnusenberg e Madura (2001) procuraram analisar as reações excessivas e fracas que ocorrem no mercado com a chegada de novas informações, conhecidas na literatura como *overreaction* (excessiva) e *underreaction* (fraca). Os autores analisaram o comportamento do mercado com base nos seguintes índices: DJIA, S&P 500, *Nasdaq*, *New York Stock Exchange* (NYSE), *Russell 3000* e *Wilshire 5000*, buscando compreender a reação exagerada às boas (más) notícias relacionadas ao retorno dos índices. Assim, utilizaram o modelo autoregressivo integrado de médias móveis – ARIMA e o modelo de retorno ajustado à média, para estimação dos retornos dos índices.

No estudo, foram encontradas evidências da existência de uma reação fraca no mercado, no dia subsequente ao resultado de um retorno médio positivo (vencedores), o mesmo comportamento foi observado para o retorno médio negativo (perdedores). Assim, foi constatado um comportamento *underreaction* em todos os índices no curto prazo, que, segundo os autores, estaria atrelado a um desvio cognitivo no comportamento dos investidores que interpretariam positivamente as notícias muito pessimistas, e negativamente as notícias muito otimistas.

No entanto, ao analisar o prazo de 60 dias, o comportamento para os perdedores modificou, ou seja, a média acumulada do retorno anormal tende a tornar-se positiva no longo prazo para a maioria dos índices na condição inicial de perdedores, com exceção do índice *Russell 3000* e do *Wilshire 5000*. Desse modo, no médio prazo, são encontradas evidências de uma reversão, analisada pelos autores como uma reação exagerada para os índices perdedores. Já para os vencedores, foi observado o comportamento *underreaction*.

Zhang (2006) procurou investigar se a incerteza informacional levaria os preços a não se ajustarem rapidamente às novas informações, adotando como *proxies* para a incerteza informacional: (a) tamanho da empresa; (b) idade da empresa; (c) cobertura dos analistas; (d) variação da previsão dos analistas; (e) volatilidade do retorno; e (e) volatilidade do fluxo de caixa. A amostra do estudo foi formada por empresas com ações negociadas na NYSE, AMEX e Nasdaq, no período de 1983 a 2001. O autor analisou o comportamento da revisão dos preços

mediante a previsão dos analistas e a variação do retorno no curto prazo, considerando como boa (má) notícia o resultado do anúncio dos ganhos, e o retorno médio mensal acumulado no intervalo temporal de 11 meses.

Desse modo, o estudo investigou se a resposta do preço às novas informações era mais lenta quando existiam ambiguidades (incertezas) sobre os fundamentos da empresa, justificando que isso ocorreria em função dos investidores apresentarem vieses cognitivos, os quais seriam mais evidentes diante da incerteza. As evidências empíricas mostraram que o mercado reagia mais rapidamente às informações, quando existia baixa incerteza informacional, no entanto, a previsibilidade do retorno era pequena, visto que os preços refletiam as informações disponíveis. Já na presença de maior incerteza informacional, o mercado reagia mais lentamente, e os retornos eram mais elevados que o retorno esperado, quando existiam divulgações de boas notícias, e quando eram divulgadas más notícias, os retornos eram relativamente menores. Portanto, o autor chega à constatação de que o mercado apresentava um efeito *underreaction* em períodos de incerteza informacional.

Hong, Torous e Valkanov (2007) verificaram em que medida o comportamento de ativos financeiros provenientes da indústria poderiam prever o movimento do mercado acionário. Para isso, analisaram o retorno de carteira de ativos da indústria, negociados em bolsas dos Estados Unidos, no período de 1946 a 2002, e contrastaram os resultados com a análise do comportamento dos oito maiores mercados (Japão, Canadá, Austrália, Reino Unido, Holanda, Suíça, França e Alemanha), excluído os Estados Unidos, no período de 1973 a 2002. No que diz respeito ao mercado acionário dos Estados Unidos, dentre os 34 ativos avaliados, 14 apresentaram capacidade de prever o comportamento do mercado por até um mês. Para os demais mercados, foi observado um comportamento semelhante, com exceção do Japão. Assim, é levantada a conclusão de que o mercado reage com atraso às informações sobre os fundamentos da indústria, sustentando a hipótese de que as informações se difundem gradualmente no mercado. Uma das explicações para esse comportamento dos investidores seria o fato de não serem especialistas no mercado geral, mas em determinados mercados. Com isso, informações valiosas acabariam levando certo tempo até serem processadas por todo o mercado.

No Brasil, Barbosa e Medeiros (2007) estudaram o comportamento do mercado acionário, visando verificar a existência da eficiência do mercado após choques de notícias favoráveis e desfavoráveis, no período de 2001 a 2005, utilizando como *proxy* para notícia (favoráveis/desfavoráveis) a classificação do valor do desvio padrão dos resíduos da regressão

do índice do retorno do índice Bovespa (IBOVESPA) regredido contra o retorno do índice Dow Jones (*proxy* para o índice do mercado mundial).

Assim, foi realizado um estudo de eventos para a eficiência do mercado, considerando que os preços dos ativos financeiros deveriam reagir rapidamente à divulgação de qualquer notícia que afetasse os seus fundamentos, de forma que não existiria tendência ao exagero, nem reversão dos preços após a divulgação. Além da HME, os autores procuraram analisar o comportamento *underreaction* e *overreaction* no mercado brasileiro. As evidências empíricas mostraram que tanto os choques de notícias positivas, quanto os choques de notícias negativos, não levavam a uma eficiência do mercado, mas a uma reação exagerada aos dois tipos de choques de notícias. Portanto, foram encontrados indícios de *overreaction* no mercado acionário brasileiro.

Feng e Hu (2014) analisam a reação dos investidores no mercado acionário chinês, nas datas de divulgação dos anúncios de lucros, utilizando respaldo teórico na teoria da atenção limitada. A amostra do estudo foi formada por empresas listadas no mercado acionário da China, no período de 2005 a 2010. Assim, realizaram um estudo de eventos para investigar como os investidores reagiam à chegada das informações relacionadas aos lucros, em termos de volume de negociação e em relação ao preço dos ativos financeiros. Os autores encontraram evidências de que o mercado chinês reagia lentamente aos anúncios dos lucros no curto prazo, ou seja, o impacto imediato da informação sobre os preços dos ativos e sobre o volume de negociações no mercado era fraco. Os autores explicam que esse comportamento está associado a uma tendência comportamental do investidor, que, ao processar os vários sinais da informação, teria a sua atenção limitada por fatores cognitivos. Com isso, as informações sobre a divulgação dos resultados não seriam processadas em tempo hábil, levando a uma má precificação temporária dos ativos. Portanto, a atenção limitada dos investidores provocaria impactos sobre a eficiência do mercado.

O Quadro 1 apresenta uma síntese dos principais estudos empíricos que levantaram evidências de defasagens temporais na incorporação das informações junto aos preços.

Diante das evidências empíricas expostas sobre a eficiência do mercado, verifica-se que, em geral, o mercado apresenta ineficiências, provocadas por uma defasagem temporal no processo de incorporação das informações juntos aos preços, levando a reações excessivas (*overreaction*) ou retardadas (*underreaction*), que tendem a serem revertidas com o passar do tempo.

Quadro 1 - Síntese dos principais estudos empíricos da reação defasada do mercado no ajuste dos preços

Autor	Período	País	Contribuição
Jegadeesh e Titman (1993)	1965 a 1989	Estados Unidos	Analisa estratégias de investimento baseadas em retornos históricos, associando o comportamento do mercado a uma reação retardada na incorporação das informações junto aos preços.
Moskowitz e Grinblatt (1999)	1963 a 1995	Estados Unidos	Analisa estratégias de investimento baseadas no retorno histórico da indústria, encontrando uma reação retardada no ajuste dos preços, provocadas por empresas pequenas.
Schnusenberg e Madura (2001)	1928 a 1997 – <i>DJIA e S&P 500</i> ; 1966 a 1997- <i>NYSE</i> ; 1971 a 1997 – <i>Nasdaq</i> ; 1979 a 1997 - <i>Russell 3000 e Wilshire 5000</i> .	Estados Unidos	Analisa o comportamento do mercado diante da chegada de novas informações por meio de índices setoriais ou do mercado, encontrando reações <i>underreaction</i> no curto prazo para todos os índices.
Zhang (2006)	1983 a 2001	Estados Unidos	Analisa o processo de ajuste dos preços em condições de incerteza, levantando o efeito <i>underreaction</i> em períodos de incerteza informacional.
Hong, Torous e Valkanov (2007)	1946 a 2002	Estados Unidos, Japão, Canadá, Austrália, Reino Unido, Holanda, Suíça, França e Alemanha.	Investigam a capacidade do retorno dos ativos da indústria em prever o movimento de mercados globais, evidenciando que as informações se difundem gradualmente no mercado.
Barbosa e Medeiros (2007)	2001 a 2005	Brasil	Investigam o comportamento do mercado acionário após choques de notícias, encontrando indícios de <i>overreaction</i> tanto para os choques de notícias positivas, quanto para os choques de negativas.
Feng e Hu (2014)	2005 a 2010	China	Analisa a eficiência do mercado sob a ótica da atenção limitada do investidor, evidenciando que o mercado reagia lentamente ao anúncio dos lucros e que a atenção limitada contribuiria para a ineficiência do mercado.

Fonte: Elaborado pelo autor com base nos estudos analisados (2017).

As explicações para a ineficiência do mercado estão atreladas à ideia de que o investidor não apresenta a racionalidade pressuposta pela HME, mas uma racionalidade limitada, atrelada à influência de fatores psicológicos, também chamados de vieses cognitivos, como o viés da negatividade. Desse modo, seguindo o teorema de *bayes*, o ajuste das crenças sobre as expectativas futuras do fluxo de caixa dos ativos sofreria influência de uma valorização excessiva ou retardada das novas informações, em virtude dos vieses cognitivos, levando os preços a uma má precificação temporária (SCHNUSENBERG; MADURA, 2001).

Na presente tese, investigou-se o comportamento dos preços no mercado acionário brasileiro diante do sentimento textual das notícias, utilizando o viés cognitivo de negatividade

e a hipótese do mercado eficiente, como respaldo teórico, considerando que o sentimento textual é uma informação importante e que provoca reações no mercado de país emergente. Para isso, buscou-se analisar o comportamento das palavras presentes nas matérias financeiras de jornais.

2.3 LEI DE ZIPF

A lei de Zipf trata de uma lei de potência, que apresenta explicações sobre o comportamento de distribuição das palavras usadas nos processos de comunicação. Zipf (1949), ao analisar a obra literária *Ulisses*, observou a existência de uma correlação entre a frequência e o ranking das palavras encontradas no texto. Além disso, verificou que o produto da frequência das palavras pelo seu ranking resultava em um valor aproximadamente constante. Diante dessas evidências, Zipf (1949) chegou a constatação de que o produto da ordem de classificação de uma palavra pelo número de sua ocorrência era aproximadamente um valor constante, formulando, assim, a primeira Lei de Zipf, conforme Equação 2:

$$fr = c \quad (2)$$

Em que:

f é o número de ocorrência de uma palavra ou *ranking*; r é a ordem de classificação da palavra; e c é uma constante relacionada ao evento analisado.

A primeira Lei de Zipf é válida para explicações sobre o comportamento das palavras com alta frequência em um texto. Para as palavras menos citadas, a explicação da distribuição é dada pela segunda lei de Zipf, que foi apresentada por Booth (1967), e considera que as palavras menos citadas em um texto tendem a apresentar uma mesma frequência, conforme Equação 3.

$$I_n = \frac{I_1 k}{n(n+1)} \quad (3)$$

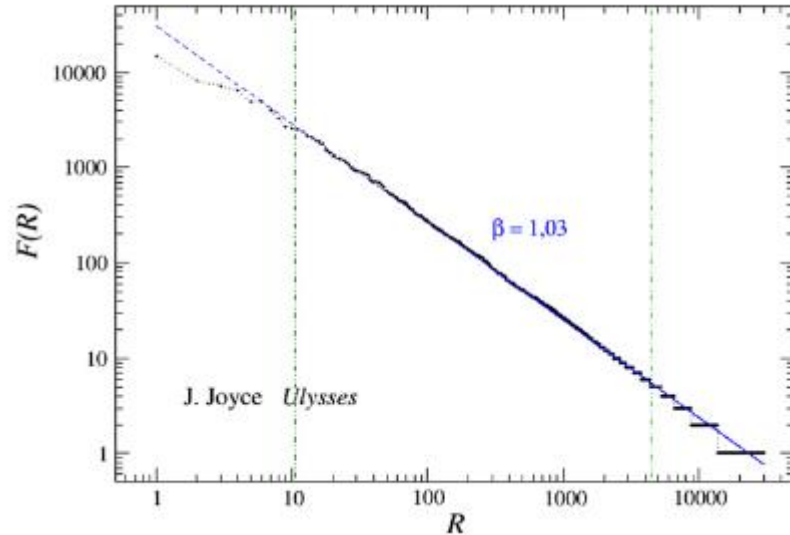
Em que: I_n número de palavras que tem frequência n ;

I_1 número de palavras com frequência 1;

n palavras que apresentam a frequência n ;

k é uma constante da língua a ser tratada; Booth (1976) sugere que, para a língua, inglesa seja considerada uma constante no valor 2.

Figura 2 - Distribuição de Zipf das palavras da obra literária *Ulysses*



Fonte: Kwapień e Drożdż (2012, p.186).

Nota: A figura trata da distribuição da frequência das palavras encontradas na obra *Ulysses*, escrita por Joyce em inglês. A linha contínua trata da função da lei de Zipf e a linha tracejada corresponde aos resultados empíricos da pesquisa desenvolvida por Kwapień e Drożdż (2012).

A lei Zipf mostra o comportamento da curva de distribuição das palavras presentes em textos longos e pode ser utilizada como um instrumento auxiliar para indexação de termos e estruturação de listas de palavras-chave com maior relevância em análises textuais.

No processo de determinação dos termos mais representativos em um texto, Pao (1978) sugere que devem ser desconsideradas as palavras localizadas nas extremidades da curva de distribuição, pois as palavras com maior frequência costumam apresentar baixo valor semântico, pois são palavras classificadas como artigos, preposições e conjunção, e as palavras pouco utilizadas, geralmente, não apresentam conteúdo significativo para o texto. Portanto, as palavras com maior conteúdo semântico, que devem ser utilizadas para a estruturação de listas de termos representativos, geralmente, não estão localizadas nas extremidades das distribuições de frequência, e são termos relacionados a substantivos, adjetivos e verbos.

2.4 SENTIMENTO TEXTUAL

No processo de comunicação escrita ou verbal, o sentimento ou o tom das mensagens pode ser compreendido como um dos atributos do discurso (KEARNEY; LIU, 2014). Nesse

contexto, o sentimento está atrelado à opinião subjetiva presente nos textos ou no discurso verbal, ou seja, seria o posicionamento de otimismo/positivo ou de pessimismo/negativo apresentado no processo de comunicação (BEATTIE, 2014; KEARNEY; LIU, 2014).

Nesse sentido, o sentimento ou tom textual pode ser entendido como uma análise linguística das palavras empregadas nos textos, visando capturar o posicionamento de otimismo ou pessimismo. Desse modo, a avaliação do sentimento textual, em geral, está atrelada às características linguísticas das palavras, ou dos aspectos léxicos, não sendo levada em consideração a estrutura das sentenças (sintaxe) ou o seu significado, conforme o contexto no qual a palavra está inserida (semântica) (BEATTIE, 2014).

Assim, para a mensuração do sentimento textual, é preciso que os textos sejam submetidos a análises de conteúdo, sejam elas manuais ou computadorizadas, para ser extraído o tom do texto (LI, 2010a). As análises manuais são realizadas, basicamente, por meio de técnicas que exigem a participação direta das pessoas no processo de classificação dos textos, enquanto na análise de conteúdo computadorizada o processo ocorre de modo automatizado, com a utilização de algoritmos que fazem as análises, levando em consideração características previamente definidas.

O estudo realizado por Niederhoffer (1971) está dentre as pesquisas que procuraram investigar o impacto das notícias na movimentação do mercado acionário. Nesse estudo, foi analisada a relação entre eventos mundiais e o comportamento dos preços das ações, por meio de uma análise de conteúdo manual das principais matérias do jornal *New York Times*, entre os anos de 1950 a 1966. Assim, foram analisados 432 eventos, com o auxílio de leitores, que classificaram os títulos das matérias, como notícias boas/más, por meio de uma escala. Os resultados indicaram sinais de um comportamento exagerado no mercado diante das más notícias. No entanto, a pesquisa de Niederhoffer (1971) trabalhou com uma amostra pequena, restringindo, com isso, a realização de inferências generalizadas sobre o comportamento do mercado. Em geral, a limitação do tamanho da amostra é uma característica peculiar aos estudos que trabalham com análises de conteúdo manuais, devido à impossibilidade operacional de usar grandes volumes de textos. Outro problema que pode existir é o subjetivismo na condução das análises, que impossibilita a realização de estudos em outros contextos.

Na década de 1950, a maior acessibilidade aos computadores possibilitou o aperfeiçoamento da Linguagem de Processamento Natural, originando técnicas sofisticadas para a realização de análises de conteúdo, as quais fortaleceram os estudos sobre o sentimento textual (LOUGHRAN; McDONALD, 2011).

As análises de conteúdo automatizadas que realizam a mensuração do sentimento textual, em geral, ocorrem por meio de: (a) aprendizagem baseada em máquina; e (b) aprendizagem baseada em dicionário. Em relação à aprendizagem baseada em máquina, pode-se dizer que é uma abordagem embasada em técnicas estatísticas, que possibilitam a realização de análises sobre o conteúdo dos textos, assim como permite a classificação dos documentos, por meio de inferências estatísticas (LI, 2010b). Um dos algoritmos utilizados nesse tipo de abordagem é o *Naïve Bayesian*, que estima a relação entre o conteúdo do documento e a frequência de cada palavra, capturando a correlação existente entre eles, com base em uma distribuição de probabilidade condicional. No entanto, para a utilização desse tipo de algoritmo, é imprescindível uma prévia categorização das características analisadas, como, por exemplo, a classificação manual das palavras positivas e negativas de uma amostra da análise, possibilitando, assim, o treinamento do algoritmo, para posteriores inferências sobre o sentimento textual dos documentos (KEARNEY; LIU, 2014; BUEHLMAIER, 2013).

Na literatura, são encontrados estudos empíricos que analisaram o sentimento textual por meio da aprendizagem em máquina, como Li (2010b), que utilizou o *Naïve Bayesian* para examinar o tom das declarações de projeções das organizações, presente na seção *Management Discussion and Analysis* (MD&A) dos relatórios 10-k e 10-Q. Outro estudo que adotou esse algoritmo para o levantamento do sentimento textual foi Huang, Zang e Zheng (2014), ao analisar o tom das previsões dos analistas. Já Khuu, Durand e Smales (2016) analisaram o sentimento textual das notícias apresentado no banco de dados da *Thomson Reuters News Analytics* (TRNA), fazendo uma associação entre a aprendizagem em máquina com rede neural.

A aprendizagem baseada em dicionário analisa as informações textuais por meio de algoritmos, que fazem a leitura e a classificação automatizada das palavras ou frases, utilizando dicionários ou de listas de palavras previamente definidas (LI, 2010b). Segundo Kearney e Liu (2014), para a estimação do sentimento textual baseado nesse tipo de abordagem, é preciso levar em consideração: (a) a estruturação dos dicionários ou das listas de palavras; e (b) a ponderação das palavras.

Uma das primeiras listas de palavras utilizadas nas análises de informações financeiras foi a *Harvard General Inquirer* (GI/Harvard), proposta na área de psicologia, com o objetivo de avaliar o comportamento social, mas que teve sua aplicabilidade estendida para a área de finanças, principalmente no que se refere à utilização das listas de palavras positivas e negativas da *Harvard-IV* (LOUGHRAN; McDONALD, 2011; KEARNEY; LIU, 2014).

No entanto, a utilização das listas de palavras oriundas da psicologia passou a ser criticada, em função dos erros de classificação e das omissões que ocorriam ao serem analisadas informações financeiras. Nesse sentido, Loughran e McDonald (2011), ao observarem as limitações do dicionário GI/Havard, desenvolveram listas de palavras com ênfase em finanças, por meio de uma expansão das ideias apresentadas na GI/Havard, ficando conhecidas na literatura como listas *L&M*. Desse modo, a proposta da lista de palavras para o contexto financeiro estava voltada para uma adequada classificação das palavras, visando capturar melhor o tom das informações financeiras.

Diante das evidências empíricas de Loughran e McDonald (2011), que indicavam erros na classificação das palavras negativas, ao utilizar o GI/Havard – IV para a análise do relatório 10-K, e mediante o surgimento das listas *L&M* como uma alternativa para as análises de conteúdo no contexto financeiro, as listas *L&M* passaram a predominar entre os estudos empíricos sobre o sentimento textual, podendo ser citada a sua utilização nas obras de Carretta et al. (2013), García (2013), Liu (2014), Ferguson et al. (2015), Ahern e Sosyura (2015) e Ahmad et al. (2016).

No contexto da aprendizagem baseada em dicionário, para ser levantado o sentimento textual, é preciso classificar as palavras seguindo dicionários e definir critérios de ponderação para elas. Em geral, os estudos costumam analisar as palavras com base em sua frequência, sem estabelecer padrões de ponderações (TETLOCK, 2007; KOTHARI; LI; SHORT, 2009). Porém, ao analisar as palavras apenas com base em sua frequência, podem existir perdas informacionais na estimação do tom textual, em função da valoração excessiva das palavras que apresentam alta frequência comparada às que apresentam baixa frequência (LOUGHRAN; McDONALD, 2011).

Diante disso, Loughran e McDonald (2011) sugerem que as palavras sejam analisadas com base em duas ponderações, as quais devem considerar a importância da palavra para o texto, no qual está inserida (seria o peso da palavra para o documento), e a importância da palavra dentro do conjunto de documentos analisados (peso da palavra dentro da amostra), sendo normalizados por algum fator relacionado ao volume dos textos. Com esse tipo de ponderação, é possível adequar o impacto da frequência das palavras ao tamanho dos documentos, capturando melhor o impacto de uma palavra para o contexto no qual está sendo avaliada.

No entanto, a mensuração do sentimento textual dentro da aprendizagem baseada em dicionário, geralmente, não utiliza ponderações e é obtida por meio da diferença entre o

peso/frequência das palavras positivas e negativas (CARRETTA et al., 2013; DAVIS; PIGER; SEDOR, 2012), ou ainda pela a diferença entre peso/frequência das palavras positivas e negativas dividido pelo total de palavras (ENGELBERG; REED; RINGGENBERG, 2012; GARCÍA, 2013; LIU; McCONNELL, 2013; BIANCONI; HUA; TAN, 2015).

No que tange às informações passíveis de extração do sentimento textual, são constatados na literatura que os estudos empíricos procuram analisar informações textuais relacionadas à: (a) *disclosures* corporativos (LOUGHRAN; McDONALD, 2011; HUANG; TEOH; ZHANG, 2014; JEGADEESH; WU, 2013); (b) previsão dos analistas (TWEDT; REES, 2012; HUANG; ZANG; ZHENG, 2014); (c) Artigos de Jornais (GARCÍA, 2013; LIU, 2014; FERGUSON et al., 2015; AHERN; SOSYURA, 2015; FRAIBERGER, 2016; STRAUß, VLIEGENTHART; VERHOEVEN, 2016); e (d) *post* – mensagens oriundas de *blogs*, *Facebook*, *Twitter*, dentre outras (ANTWEILER; FRANK, 2004; HU; MCINISH; ZENG, 2010; BOLLEN; MAO, ZENG, 2011).

Pode-se dizer que o sentimento textual tem sido uma área de estudo emergente entre as pesquisas empíricas realizadas tanto em finanças, quanto em contabilidade. As possíveis explicações para os avanços dessas investigações estão atreladas ao reconhecimento do sentimento textual como uma fonte informacional tão importante quanto às tradicionais informações quantitativas (TETLOCK; SAAR-TSECHANSKY; MACSKASSY, 2008; LIU, 2014).

Dentre as informações textuais, observa-se que as notícias provenientes da mídia financeira desempenham um papel importante para o mercado financeiro, pois atuam na distribuição de informações, atenuando os problemas de agência e de assimetria informacional (TETLOCK, 2010; GRIFFIN; HIRSCHHEY; KELLY, 2011; LIU; McCONNELL, 2013; FERGUSON et al., 2015; AHERN; SOSYURA, 2015). Além disso, por não estarem diretamente vinculadas às organizações, passam certa credibilidade para os investidores (KOTHARI; LI; SHORT, 2009). Com isso, acredita-se que as notícias transmitem informações que contribuem no processo de tomada de decisão dos investidores.

Segundo Kothari, Li e Short (2009), a demanda e a divulgação de informações no mercado têm o sentido de minimizar a assimetria informacional e os problemas de agência, existentes entre gestores, investidores e intermediários da informação. Tetlock (2007), ao analisar as notícias provenientes do *Wall Street Journal*, chama a atenção para o fascínio que as notícias provocavam entre os investidores, ao serem relatadas explicações sobre a movimentação do mercado, levantando, com isso, a importância que as informações

apresentadas nos jornais teriam para o mercado. Dentro dessa mesma linha de investigação, García (2013) ressalta a relevância das notícias financeiras para o mercado acionário, principalmente nos momentos de recessão econômica. Por fim, Liu (2014), ao analisar o mercado europeu em um período de crise financeira, chama a atenção para o papel da mídia na disseminação de informações importantes sobre os riscos de crédito dos países.

No âmbito do sentimento textual, as pesquisas têm levantado evidência de que o tom das notícias provoca reações no mercado, em função de transmitirem informações importantes para os investidores (TETLOCK, 2014). Além disso, acredita-se que as informações provenientes das notícias são utilizadas na estimação dos valores fundamentais dos ativos (GRIFFIN; HIRSCHEY; KELLY, 2011; TETLOCK, 2014). Em outras palavras, os investidores utilizariam as notícias como fonte de informação para alinhar as suas crenças sobre as organizações, logo, a chegada de notícias otimistas ou pessimistas provocaria reações no mercado.

2.4.1 Evidências Empíricas do Sentimento Textual das Notícias Financeiras

Uma das primeiras pesquisas a discorrer sobre a interação entre o sentimento textual das notícias no mercado acionário, utilizando análise de conteúdo automatizada, foi o estudo de Tetlock (2007), no qual foi analisada a coluna financeira *Abreast of the Market* presente no *Wall Street Journal*, no período de 1984 a 1999. Segundo o autor, a coluna financeira estaria dentre as mais populares dos Estados Unidos, visto que nela seriam encontradas informações sobre as movimentações do mercado, assim como as previsões dos analistas e o posicionamento de corretoras, explicando o que ocorreu no mercado e projetando comportamentos futuros. No estudo, investigou-se a correlação entre o conteúdo da coluna diária do WSJ e as atividades do mercado de ações, com o intuito de verificar se o pessimismo da mídia poderia prever os movimentos no mercado, em termos de volume das negociações e de retornos anormais.

Para a mensuração do fator de pessimismo, foi utilizada uma análise de componentes principais, baseada nas palavras negativas que foram encontradas ao longo dos textos, e, para a classificação e mensuração da frequência de palavras, foi utilizado o *General Inquirer's Harvard IV-4 psychosocial*. Como *proxy* para o comportamento do mercado, foi utilizada a rentabilidade diária do índice DJIA. Assim, a estimação da relação entre o fator de pessimismo

da mídia e os movimentos do mercado foi obtida mediante o modelo Vetor Autorregressivo (VAR). Os principais resultados mostraram que níveis elevados de pessimismo na mídia preveem uma pressão sobre os preços de mercado, os quais reagem com quedas, mas que são revertidas rapidamente. Além disso, foram encontrados indícios de que o baixo pessimismo na mídia prediz um elevado volume de negociação no mercado, e que baixos retornos no mercado tendem a elevar o pessimismo na mídia. Portanto, o autor sugere que o indicador de pessimismo, baseado no sentimento textual das notícias, seria uma *proxy* para o sentimento dos investidores ou para a aversão ao risco.

Tetlock, Saar-Tsechansky e Macskassy (2008) analisaram em que medida a linguagem negativa explicava lucros contábeis e o retorno das ações. Assim, analisaram os impactos das informações apresentadas no *WSJ* e no *Dow Jones News Service* (DJNS), quanto às notícias sobre as empresas integrantes do índice S&P500, no período de 1980 a 2004. Os autores partiram da suposição de que palavras negativas apresentavam uma relação forte com o retorno dos ativos financeiros, pois elas provocariam um maior impacto nos investidores. Desse modo, o estudo restringiu-se à análise do impacto das palavras negativas utilizadas nos textos das notícias que faziam referências às empresas integrantes da amostra, fazendo uso do *General Inquirer's Harvard IV-4* para a classificação e mensuração da frequência das palavras. Assim, foi investigado se as análises linguísticas das palavras negativas poderiam melhorar a compreensão sobre o fluxo de caixa futuro da empresa e se os preços incorporavam eficientemente essa informação, fazendo o uso de regressão linear, com estimação dos parâmetros por Mínimos Quadrados Ordinários (MQO).

Os resultados evidenciaram que as palavras negativas utilizadas nas notícias analisadas transmitem informações negativas sobre os ganhos e que poderiam predizer ganhos além do que seria possível captar com a previsão dos analistas e com as medidas tradicionais de avaliação dos preços. Ao analisarem especificamente o processo de incorporação das informações textuais junto aos preços, observaram que o mercado reage à informação com um atraso de um dia. Com isso, verificou-se que as informações textuais negativas apresentam potencial para a formulação de estratégias de investimento. No entanto, os custos de transação poderiam eliminar a rentabilidade da estratégia.

Kothari, Li e Short (2009) estudaram o efeito sistemático do risco com a disseminação de informações financeiras, no contexto do mercado acionário dos Estados Unidos. Para isso, foram adotadas como *proxies* de risco: o custo do capital, a variação da previsão dos analistas e a volatilidade do retorno. Além disso, realizaram uma análise de conteúdo para classificação

das informações favoráveis (otimistas) e desfavoráveis (pessimistas), por meio do *General Inquirer's Harvard*, em 326.357 documentos textuais, relacionados à: (a) *disclosures* das empresas (MD&A); (b) textos com previsão dos analistas; e (c) notícias financeiras, partindo da suposição de que as informações textuais podem afetar o mercado, e modo que, as informações desfavoráveis (favoráveis) tendem a aumentar (reduzir) o risco do fluxo de caixa.

Ao analisar especificamente as informações relacionadas aos *disclosures*, foi constatado que tons de otimismo não afetaram significativamente o custo de capital da empresa, sugerindo que divulgações otimistas da gestão não gerariam credibilidade. Já os tons de pessimismo estavam relacionados ao aumento da volatilidade do retorno e da variação da previsão dos analistas. Ao analisarem as informações textuais relacionadas às previsões dos analistas, observaram que o mercado ponderava fortemente essas informações e que o posicionamento de otimismo ou pessimismo seria insignificante para o custo de capital. Já na análise quanto aos impactos das notícias da mídia, foram encontradas evidências de que essas informações impactavam o custo de capital, a volatilidade e a variação da previsão dos analistas, de modo que notícias otimistas diminuía significativamente as *proxies* de risco, enquanto as notícias desfavoráveis eram acompanhadas por um aumento do risco. Com isso, os autores concluem que as notícias da imprensa teriam mais credibilidade, visto que seriam menos afetadas pelos incentivos de agência, comparado aos gestores (*disclosures*) e aos analistas (previsão dos analistas).

Dougal et al. (2012) investigaram a relação existente entre as informações apresentadas na coluna *Abreast of the Market* do *WSJ* e o desempenho do mercado acionário dos Estados Unidos, no período de 1970 a 2007, com a premissa de que a escrita das notícias tem um efeito causal sobre o comportamento dos investidores e, assim, foi analisada a dinâmica do retorno do DJIA em relação aos jornalistas que escrevem as notícias analisadas, levando em consideração características peculiares da escrita dos textos relacionadas: (a) ao posicionamento de otimismo ou pessimismo – avaliando mediante as palavras positivas e negativas utilizadas no texto, classificadas de acordo com o dicionário L&M; e (b) à complexidade na construção das frases. A relação entre as variáveis foi realizada mediante regressão linear, com estimação dos parâmetros por MQO. Ao analisar as notícias levando em consideração o jornalista que a escreveu, os resultados sugerem que as matérias dos jornalistas fixos conseguem prever o retorno, e, quando o efeito do jornalista fixo é analisado com os retornos mais recentes a previsibilidade ainda é maior. Constatou-se, também, que o tom pessimista influenciava temporariamente o preço das ações, no entanto, o estudo não se concentrou em investigar a

influência das ferramentas retóricas utilizadas pelos jornalistas. Portanto, constatou-se que os jornalistas apresentam potencial para influenciar o comportamento dos investidores no curto prazo.

Engelberg, Reed e Ringgenberg (2012) analisaram o processo de negociação de derivativos, investigando as vantagens que as informações de domínio público poderiam possibilitar nesse mercado. Os autores partem da suposição de que a vantagem das negociações em derivativos encontrava-se na capacidade de analisar informações publicamente disponíveis. Com isso, é realizado um cruzamento entre dados relacionados às negociações diárias de derivativos na NYSE com notícias específicas sobre as empresas, chamadas de comunicados de imprensa, pertencentes à *Dow Jones News Service* e ao *WSJ*, no período de 2005 a 2007, considerando para o tratamento dos textos os dicionários Harvard-IV e L&M.

Os autores procuraram analisar a relação entre o volume de negociação, retorno e o sentimento dos textos mediante regressão linear. As principais evidências mostraram que existia um aumento nas negociações após a divulgação de notícias, sugerindo que estratégias baseadas em notícias de domínio público poderiam gerar vantagens competitivas. Assim, ao ser analisada a interação entre o volume de negociações e o sentimento das notícias, observou-se que as notícias negativas (coeficiente -1.805 , *t-statistic* -4.26) comparadas às notícias positivas (coeficiente -1.113 , *t-statistic* -2.40), apresentavam efeitos mais fortes sobre os retornos, e a magnitude dos coeficientes mostraram o tamanho da previsibilidade dos retornos futuros, indicando que, em dias de notícias ruins, quase quadruplica a previsibilidade, enquanto nos dias com notícias boas chegava a quase a triplicar a previsibilidade. Desse modo, uma estratégia de negociação baseada em notícias negativas teria um ganho de 180% ao longo do período analisado no estudo (2,5 anos). Assim, os autores chegaram à constatação de que esse resultado seria um reflexo dos investidores qualificados, os quais teriam uma capacidade de processamento informacional diante das informações divulgadas, gerando vantagens competitivas nesse mercado. Desse modo, levantaram que derivativos embasados por informações não criavam mercado, mas geravam clientes. Portanto, as evidências sugerem que as notícias públicas geram oportunidades valiosas na venda de derivativos para aqueles que são qualificados no processamento das informações.

García (2013) verificou o efeito do sentimento textual das notícias sobre o preço dos ativos, no período de 1905 a 2005, no mercado acionário dos Estados Unidos. O estudo faz inferências ao comportamento dos investidores, com embasamento na literatura de psicologia e economia, sugerindo que os investidores estão mais sensíveis às notícias em períodos de

recessão. Assim, investigou o poder informacional do sentimento das notícias no mercado acionário, ao longo dos ciclos econômicos. O autor utilizou como *proxy* para notícia o sentimento textual de duas colunas financeiras do *New York Times*, adotando o dicionário de *L&M* para a categorização das palavras (positivas/negativas). Como *proxy* para o comportamento do mercado foi utilizado o retorno do índice DJIA.

Desse modo, a relação entre o comportamento do mercado e o sentimento textual foi obtida mediante o modelo de heterocedasticidade condicional autoregressivo generalizado - GARCH, que mostrou evidências de que, nos períodos de recessão, as notícias teriam um impacto maior sobre os retornos do DJIA, pois mudanças no desvio padrão do índice de pessimismo das notícias provocavam mudanças de 12 pontos base, enquanto nos períodos de expansão as alterações no desvio padrão do índice de pessimismo provocavam um efeito de apenas 3,5 pontos base. Com isso, foram encontrados padrões de previsibilidade no retorno dos ativos financeiros atrelados ao conteúdo das notícias, principalmente nos períodos de recessão.

Liu e McConnell (2013) investigaram a sensibilidade dos gestores das empresas de capital aberto norte-americanas à mídia pública, nos momentos de decisão de alocação de capital, que foram acompanhados por uma reação negativa do mercado. Assim, foram analisados 636 anúncios referentes às fusões e às propostas de aquisições, com transações no valor mínimo de US\$100 milhões, entre 01 de janeiro de 1990 a 31 de dezembro de 2010. As informações sobre os anúncios foram obtidas no banco de dados de fusões e aquisições da *Thomson Financial Securities Data Company's*. A relação entre as variáveis foi estimada por regressão probit, tendo como variável dependente a análise da proposta de aquisição, verificando se ela foi abandonada ou não. As principais variáveis independentes foram: (a) preço das ações da empresa adquirente; (b) nível de atenção da mídia ao anúncio de negociação; (c) tom de cobertura da mídia; e (d) valor da participação acionária do CEO.

Para a estimação do nível de atenção, foi considerado o número de notícias específicas sobre as empresas integrantes da amostra, ao longo dos dez dias seguintes a divulgação da proposta de negociação, que saíram no *WSJ*, *New York Times* e no DJNS. Já o tom dos anúncios foi obtido com base na relação entre a contagem de palavras negativas pelo total de palavras contadas nos textos, classificadas segundo o dicionário de dicionário L&M.

Os resultados mostraram que as propostas de aquisição estão mais propensas de serem abandonadas quando o mercado reage com preços negativos ao anúncio das negociações. Além disso, observa-se que os gestores são sensíveis aos efeitos da decisão de abandono sobre a sua riqueza, pois, ao serem incluídas as mudanças no capital tangível (ações) do CEO, observa-se

que tanto a mudança dos preços, quanto a mudança do capital do CEO, são negativamente correlacionadas com o abandono das negociações.

Ao serem incluídas no modelo as variáveis nível de atenção da mídia e o tom dos anúncios, como *proxies* para o efeito da negociação sobre a reputação do CEO, constatou-se que sozinhas não eram significantes, porém, quando as variáveis estavam juntas no modelo, apresentam significância estatística e sinais negativos, o que significa dizer que existe uma propensão ao abandono quando predomina o tom negativo dos anúncios com uma ampla disseminação na mídia. Já a reação dos preços perde a significância com a inclusão das *proxies* sobre reputação, o que é consistente com a ideia de que a preocupação do CEO com o efeito da sua decisão sobre o seu capital tangível e sobre a sua reputação o induz a ouvir o mercado. Sendo assim, nos processos de negociações, não está envolvido apenas o capital da empresa, mas fica em evidência o capital tangível do CEO e a sua reputação.

Desse modo, propostas que levam o mercado a reagir negativamente em termos de preço e do tom, e que apresentam uma ampla atenção na mídia, levam os gestores a ficarem mais propensos a abandonarem as propostas de negociação, pois a reversão da decisão reverte às perdas de capital associadas ao anúncio. Portanto, os resultados sugerem que o nível de atenção da mídia e o tom dos anúncios desempenham um papel importante na decisão dos gestores, contribuindo positivamente com a governança corporativa.

Carretta *et al.* (2013) investigaram a relação existente entre as notícias do *Wall Street Journal* junto ao mercado acionário, analisando especificamente as ofertas de *spin off*, também conhecida como oferta de ações provenientes de uma cisão empresarial. Nesse contexto, os autores explicam que o comportamento do mercado com a chegada de novas informações pode apresentar reações *underreaction* ou *overreaction*, as quais seriam explicadas pelos vieses cognitivos do comportamento humano, que interferem no nível de atenção dos investidores. Assim, analisaram como o sentimento das notícias e o nível de atenção dos investidores influenciavam o desempenho das ofertas de *spin off* no mercado, utilizando um modelo VAR estrutural.

Como *proxy* para o comportamento do mercado, utilizaram o índice DJIA, já para o nível de atenção dos investidores foi considerado o percentual da frequência de procura no *google insights* para os casos de *spin off* analisados, e para o sentimento das notícias foi utilizado a *proxy* sentimento textual, obtido mediante uma relação entre palavras positivas e negativas encontradas nos textos e classificadas conforme as listas *L&M*. Assim, a amostra do estudo foi formada por 16 ofertas de *spin off* negociadas na NYSE e publicadas no *Wall Street*

Journal entre os anos de 2004 a 2010. Os resultados mostraram evidências de que as notícias sobre *spin off* que apresentavam um sentimento positivo provocavam, em média, um impacto de 0,24% na variação do retorno, enquanto notícias com um sentimento negativo levavam a uma variação que chegou a 1,09%, constatando, assim, que o sentimento das notícias interferia no comportamento do investidor, influenciando suas preferências. Além disso, foi observado que um aumento no nível de atenção provocado pela oferta de *spin off* aumentava o volume de negociação no mercado.

Liu (2014) investigou a relação existente entre o sentimento de pessimismo das notícias, a concentração/volume de notícias e os *spreads* dos rendimentos dos *sovereign bond* (títulos do governo), ao longo dos anos de 2009 a 2012, período marcado por uma crise na Europa e por uma ampla divulgação de notícias, fazendo referência a situação de créditos dos países. Nesse contexto, a análise do estudo foi restrita aos seguintes países: (a) Grécia; (b) Irlanda; (c) Itália; (d) Portugal; e (e) Espanha. O autor procurou analisar em que medida o sentimento textual se aplicava aos mercados de dívidas, que apresentavam os títulos de longo prazo, *sovereign bond*.

A amostra do estudo foi formada por 18.610 notícias, dentre as quais foram extraídos o sentimento textual, por meio de uma análise de conteúdo baseada nas palavras negativas, classificadas segundo o dicionário de Loughran e McDonald (2011). Além do pessimismo das notícias, concentração/volume e *spreads* dos rendimentos dos títulos avaliados, o estudo utilizou variáveis relacionadas ao crédito, tais como: (a) risco do crédito – utilizando como *proxy* o prêmio *credit default swap* (CDS), que consiste em um prêmio que o investidor recebe em casos de problemas com o título; (b) liquidez – mensurada pela *proxy* valor de mercado do *sovereign bond*; e (c) aversão ao risco – avaliado mediante a diferença entre rendimento do índice de performance dos títulos das organizações com vencimento entre 1 a 10 anos (BOFA ML US *High-Yield* 1–10Y) e o rendimento do índice de performance dos títulos do tesouro, com vencimento entre 1 a 10 anos (BOFA ML US 1–10Y). As relações entre as variáveis foram analisadas mediante a utilização de regressões lineares, estimadas com o método dos momentos generalizados e regressões logísticas.

Os resultados das regressões lineares indicaram que, à medida que o pessimismo e o número de notícias aumentavam, os preços tendiam a cair, enquanto, os resultados das regressões logísticas mostraram que a probabilidade de expansão *spreads* dos rendimentos dos títulos aumentavam em 21%, quando o pessimismo da mídia se encontrava acima da média. Ademais, constataram que o aumento simultâneo do pessimismo e do volume de notícias também contribuiu com rendimentos maiores. Além disso, foi observado que a intensidade do

pessimismo na mídia estava associada ao volume de notícias, sugerindo que, um baixo volume de notícia naturalmente seria uma boa notícia sobre o mercado. Com isso, constatou-se que o pessimismo da mídia e o número de notícias ajudavam na previsão dos *spreads* dos *sovereign bond*.

Ferguson et al. (2015) examinam o poder informacional da combinação do tom (positivo/negativo) e do volume de notícias específicas sobre as empresas, analisando a capacidade de previsão dos retornos futuros das ações. O estudo restringiu-se à análise das notícias financeiras do Reino Unido, publicadas nos jornais *The Financial Times*, *The Times*, *The Guardian* e *Mirror*, referentes às empresas integrantes do índice FTSE 100, no período de 1981 a 2010, totalizando uma amostra de 264.647 notícias. Assim, para a análise do tom das notícias, foi utilizada uma medida quantitativa de análise de conteúdo, baseada na frequência das palavras positivas e negativas, classificadas conforme o dicionário de *L&M*, enquanto, o retorno das ações foi obtido mediante o modelo três fatores de Fama e French (1993). Os autores segregaram as empresas em dois grupos, conforme o tamanho e valor do *book-to-market* (BM), visando a identificação das empresas com maior/menor visibilidade.

Os resultados mostraram indícios das notícias positivas predizerem retornos futuros maiores, enquanto notícias negativas indicavam retornos menores. Além disso, ressaltou-se que as notícias que faziam referência aos fundamentos da empresa conseguiam gerar retornos anormais no dia de sua publicação, e, o impacto do tom das notícias foi mais significativo entre as empresas com menor visibilidade (menores empresas do índice FTSE 100 e alto *BM*). Portanto, o tom e o volume das notícias específicas sobre as empresas apresentaram informações valiosas que contribuiriam para a predição dos retornos futuros.

Ahern e Sosyura (2015) analisaram como as notícias sensacionalistas afetavam o processo de fusão das organizações. Segundo os autores, a mídia, visando chamar atenção dos leitores, publica notícias sensacionalistas sobre fusões, que podem contribuir, distorcendo a crença dos investidores e impactando o preço dos ativos financeiros. Assim, foram investigadas as características das notícias que trataram sobre rumores de fusões, que se tornaram realidade, e analisaram o impacto da precisão dessas notícias nos preços das ações.

Para isso, os autores procuraram notícias com a especulação de fusões, nos períodos de 2000 a 2011, levantando 501 rumores de fusões. Para estimar as características que direcionariam rumores a se tornarem realidade, foi utilizada uma regressão *logit* associada ao comportamento de quatro conjuntos de fatores, quais sejam: (a) empresas alvo (tamanho da empresa, reconhecimento da marca e gastos com publicidade); (b) características do jornalista

(idade, experiência e formação); (c) características do texto (ambiguidade do texto e detalhes específicos); e (d) atributos dos jornais (ano de fundação e estatísticas de circulação diária).

Em relação ao primeiro conjunto de fatores, foi constatado que notícias sobre as empresas alvo dos jornalistas (maiores e com grande reconhecimento) nem sempre estão propensas a se tornarem realidade. Já em relação às características dos jornalistas, observou-se que elas podem prever, significativamente, a precisão dos rumores analisados, ou seja, jornalistas mais velhos, com formação em jornalismo e especializados no setor alvo, tendem a apresentar uma maior precisão nas especulações.

Em relação às características do texto, avaliadas mediante o dicionário de palavras proposto por Loughran e McDonald (2011), verificou-se que textos com a utilização de poucas palavras modais, como as palavras: (a) concebível e (b) talvez, tendem a não fornecer detalhes e são mais subjetivas, indicando que a especulação era menos provável de se tornar realidade. Quanto às características dos jornais, os autores levantaram que não seriam atributos importantes para explicar a precisão dos rumores. Ao analisar o impacto das notícias sobre os preços, observaram que, nas datas de divulgação de notícias sobre possíveis fusões, os retornos anormais chegaram a 6,9% para rumores precisos e 3% para rumores imprecisos. Ao analisar os 20 dias seguintes à publicação, o retorno dos ativos atrelado aos rumores precisos chegou a 492 pontos base, maior que o retorno de rumores imprecisos. Desse modo, os autores avaliaram que o mercado exagera aos rumores de uma fusão, e que os investidores não conseguiriam distinguir perfeitamente os rumores, por uma limitação do seu nível de atenção.

Bianconi, Hua e Tan (2015) mensuraram o efeito potencial das notícias no risco sistêmico das instituições financeiras, partindo da conjectura de que o sentimento do mercado é impulsionado pelo conteúdo da mídia e diante da possibilidade de mensuração do risco pelo *Value at risk* (VaR). Desse modo, o estudo analisou o processo de divulgação e disseminação da informação aos investidores como um elemento preditivo do risco sistêmico em instituições financeiras. Assim, foi investigado se o pessimismo previa o risco, não em termos do retorno médio condicionado ao pessimismo do investidor, mas em relação ao efeito do pessimismo sobre as caudas de distribuição do risco, visto que nelas é possível mensurar essencialmente o risco, e as formas de contágios às outras instituições financeiras.

No estudo, foram utilizados dois tipos de informações, sendo elas: (a) as informações de ampla circulação, propagadas pelos jornais, que seriam uma *proxy* para o sentimento de pessimismo dos investidores; e (b) informações sobre as expectativas futuras da volatilidade do mercado, que seria uma *proxy* para o índice de medo do investidor. Desse modo, foi analisado

o mercado americano no período de 1992 a 2006, considerando os dados do índice de pessimismo do estudo de García (2013) como uma medida para o sentimento do mercado, e para o índice de medo foram utilizados dados do *Chicago Board Options Exchange Volatility Index*, chamado de VIX. Assim, o risco foi mensurado com base no *Conditional value at risk* (CoVaR) e com base no VaR, mediante um modelo de regressão linear quantílica, no qual foi estimado o quantil 5%. Para a análise da previsibilidade, foram utilizadas as modelagens ARCH(1)-GARCH(1).

As evidências empíricas sugerem que as duas formas de informações apresentam efeitos distintos sobre o risco, visto que, ao fazer a inclusão do pessimismo do investidor na estimação por VaR, os resultados são significativos para metade dos casos e no *CoVaR* não existe significância. Já a inclusão do índice de medo tanto no *VaR*, quanto no *CoVaR*, apresenta significância para a maioria dos casos. Desse modo, é constatado que o efeito do índice de medo tem um alto impacto sobre o risco sistêmico das instituições financeiras.

Ao investigar a bolha de *internet* na década de 1990, é levantado que o pessimismo prevê um risco sistêmico maior, ou seja, um maior esforço financeiro ao longo de todo o período, enquanto o índice de medo prevê um risco sistêmico mais acentuado no período prestes ao estouro da bolha. Portanto, os autores concluem que o pessimismo dos investidores seria uma boa medida na previsão dos retornos médios, entretanto, não é uma medida que prediz o risco extremo (risco nas caudas).

Strauß, Vliegthart e Verhoeven (2016) investigaram a emoção apresentada nas matérias de jornais holandeses com grande circulação, analisando os seus efeitos no preço de algumas ações listadas no índice *Amsterdam Exchange* (AEX), entre os anos de 2002 a 2013, com o objetivo de avaliar o papel dos meios de comunicação no processo de transmissão de emoções ao mercado, por meio de uma análise dos efeitos recíprocos existentes entre o tom dos artigos e o preço de abertura de ações do índice AEX.

O estudo foi fundamentado na visão de que o mercado não seria eficiente, conforme prega a HME, em função dos investidores serem influenciados por aspectos emocionais. Em decorrência disso, os investidores estariam sujeitos ao efeito manada, e a tomada de decisão seria realizada não com base no valor real da ação, mas sobre a conjectura que é formada no mercado sobre aquele ativo. Nesse contexto, é atribuído à mídia o papel de fonte informacional que exprime o consenso das discussões do mercado, com poder informacional capaz de influenciar os investidores e, conseqüentemente, explicar os movimentos dos preços. Assim,

no estudo, foram analisadas as notícias que faziam referência a 21 empresas com ações integrantes do índice AEX.

As notícias foram coletadas na base de dados *LexisNexis*, totalizando 128.507 matérias que faziam referência às empresas da amostra, as quais foram avaliadas mediante um índice de emoção, estimado com base na frequência de palavras positivas e negativas, que foram classificadas seguindo o dicionário do programa *Linguistic Inquiry and Word Count (LIWC)*, em uma versão adaptada para a Holanda. O método de análise das relações entre as variáveis foi dado por regressão linear com a estimação por MQO, bem como por VAR.

As evidências sugerem que a emoção positiva das notícias apresenta efeitos no preço de abertura das ações em 12 empresas da amostra, dentre as quais prevaleceu um efeito negativo sobre os preços dos ativos. Ao tratar das emoções negativas, observa-se que apenas nove empresas apresentam resultados significativos, dentre os quais provocaram reações negativas nos preços no dia seguinte. Diante da baixa significância encontrada nas evidências, os autores concluíram que não existem indícios das notícias provocarem efeitos consistentes sobre os preços de abertura das ações. Portanto, não existem evidências que levem a concluir que a mídia apresente um efeito forte sobre o movimento de oscilação no mercado.

Jandl (2016) explorou o mercado de securitizações imobiliárias, analisando o papel do sentimento dos jornais do Reino Unido e o nível de atenção dos investidores, como fontes de informações capazes de explicar os movimentos do mercado. Teoricamente, o preço das ações deveria ser igual ao valor do patrimônio líquido (PL) por ação. Entretanto, isso não é observado, levando a busca por explicações que justifiquem tal comportamento. Os autores apontam respostas respaldadas por duas visões, sendo a primeira delas atrelada a um comportamento racional, no qual o tamanho da empresa, o setor, a qualidade da gestão e a transparência seriam aspectos que justificariam a distinção entre os valores, porém, os resultados empíricos elucidados na obra são heterogêneos e não são consistentes, indicando fragilidades.

Já a segunda visão, diz respeito ao comportamento irracional, postulada na ideia de que existem ruídos no processo de negociação dos investidores, os quais justificariam a diferença entre o valor de mercado e o valor contábil, chamado de *spreads* do PL por ação. Os autores se respaldam nessa segunda visão e apontam uma lacuna na literatura sobre a compreensão de sentimento dos textos e do nível de atenção do investidor para o mercado imobiliário securitizado.

Assim, é proposta uma investigação para averiguar a interferência do pessimismo dos jornais e do nível de atenção dos investidores nos *spreads* do patrimônio líquido por ação e no retorno dos fundos. A análise textual é realizada com base nos jornais do Reino Unido, que tratavam sobre o mercado imobiliário, entre os anos de 2001 a 2012, coletadas na base de dados *LexisNexis*.

A estimação do sentimento foi dada com base na frequência das palavras, seguindo a mesma visão de García (2013), com a utilização de dicionários de psicologia, que foram: (a) *Emotion Lexicon* – EM; (b) *General Inquirer* – GI; e (c) *Subjectivity* – MPQA. E também dicionários de finanças, que foram: (a) *Henry's Finance* – HE; (b) L&M; e (c) *L&M máster dictionary*, uma versão atualizada do dicionário L&M. Quanto ao nível de atenção dos investidores, os autores adotaram como *proxy* o índice do volume de pesquisa *on-line* oriunda do *Google*. A relação entre as variáveis foi dada mediante regressão linear.

As evidências empíricas sugerem que o pessimismo da mídia reduz os retornos dos fundos de investimento imobiliário e aumenta os *spreads* do PL por ação. Já o nível de atenção dos investidores apresenta efeitos diversificados para o retorno dos fundos analisados, no entanto, para os *spreads* do PL por ação, observa-se que o pessimismo apresenta interferências no curto prazo. Desse modo, os autores concluem que o sentimento textual e a atenção dos investidores podem servir como *proxies* confiáveis para explicar os movimentos de curto prazo do mercado de securitizações imobiliárias. Portanto, apontam que o sentimento extraído da mídia é um instrumento potencialmente poderoso para explicar fenômenos econômicos.

Fraiberger (2016) analisou a flutuação da economia, relacionando o indicador de crescimento dos países, o produto interno bruto (PIB), ao sentimento textual extraído das notícias produzidas pela *Reuters News* entre os anos de 1987-2013, dos seguintes países: Argentina, Austrália, Brasil, Índia, Indonésia, Japão, Nova Zelândia, Suécia, Suíça, África do Sul, Coreia do Sul, e Reino Unido.

O estudo procura mostrar como a opinião dos jornais está relacionada com as flutuações econômicas. Na mensuração do sentimento textual das notícias, a abordagem adotada foi a *bag-of-words*, utilizando para a classificação das palavras uma combinação de dicionários, de modo que os textos classificados como financeiros foram analisados com base na lista de palavras de *L&M*, enquanto para os textos políticos e econômicos foi realizada com base na lista de Young e Soroka (2012).

O índice de sentimento foi calculado, levando em consideração a frequência das palavras ponderadas pelo termo inverso da frequência, que é um tipo de ponderação muito utilizada nesse tipo de análise. A estimação da relação entre o sentimento textual e o indicador de crescimento PIB foi dada mediante o modelo de defasagem autoregressiva – $AR_{(p)}$ e o modelo *autoregressive distributed lag* - $ADL_{(p,q)}$.

As evidências empíricas sugerem que o sentimento negativo das notícias é um indicador de crescimento para 9 dos 12 países analisados. Já o sentimento positivo foi um indicativo de crescimento apenas para 6 países da amostra. Em relação aos erros de previsão, foi constatado que, ao incluir a medida de sentimento das notícias no modelo de previsibilidade $ADL_{(p,q)}$ e comparando ao processo $AR_{(p)}$, é observado que os países em média diminuem os seus erros de previsão em 9,1%. Ao analisar os cenários econômicos, segregando em bons e ruins, observa-se que os erros de previsão tendem a ser maiores em períodos econômicos ruins, período no qual os analistas demorariam mais para incorporar as informações nas suas previsões, sugerindo um comportamento *underreaction*.

Ao fazer a inclusão do sentimento textual, os erros de previsão apresentam uma redução de 21% nos períodos ruins, já em períodos bons a redução desses erros foi de 13%. Portanto, o índice de sentimento textual consegue rastrear com sucesso as flutuações da economia, sendo considerado como um importante indicador do crescimento futuro do PIB, por conter informações que nem sempre são capturadas pelos analistas ao prepararem as previsões sobre o estado futuro da economia.

Liu, McConnell e Xu (2017) analisam a influência da mídia sobre as oportunidades de trabalho de ex-CEOs, verificando se o tom e a cobertura da mídia dada as empresas de ex-CEOs afeta a probabilidade de que o mesmo seja nomeado para a diretoria de outras empresas após o seu mandato. Para isso, foram investigados os CEOs das empresas integrantes do S&P500 que saíram do mandato de gestão entre 1996 a 2009.

As notícias sobre as organizações foram extraídas do *WSJ*, *New York Times*, e *Dow Jones News Service*, coletadas no banco de dados *Factiva*. Para a estimação do tom das notícias, foi utilizada a lista de palavras negativas de Loughran e McDonald (2011) e realizada uma contagem das palavras negativas presentes nos textos avaliados. Desse modo, a medida de estimação do tom foi a fração de palavras negativas em relação ao total de palavras do texto, e a medida de estimação da cobertura da mídia foi o número de notícias divulgadas sobre as

empresas. As relações entre as variáveis foi capturada por regressão linear com a estimação dos parâmetros por MQO.

As evidências empíricas sugerem que o nível de cobertura da mídia dado a empresa, enquanto o CEO está no controle, está positivamente relacionada ao número de assentos no conselho de outras organizações, após o encerramento do seu mandato. Economicamente, o aumento da atenção da mídia está associado a um aumento na probabilidade em 35% do CEO assumir novas diretorias. Ao verificar o tom das notícias, é constatado que notícias negativas leva a uma redução em 28%, na probabilidade do CEO assumir assento no conselho de outras diretorias após seu mandato, e, quando existe um aumento da atenção da mídia, com notícias positivas para as organizações, isso leva, a um aumento na possibilidade novos assentos em conselhos para os CEOs. Portanto, foi verificado que existe um vínculo econômico entre a cobertura da mídia e o desempenho dos CEOs atual em relação às oportunidades futuras para os CEOs.

Wei et al. (2017) construíram um índice de sentimento, baseado em notícias financeiras das empresas listadas na bolsa de valores de Taiwan, e analisaram o comportamento desse índice em relação as movimentações do mercado, visando levantar se as notícias chinesas forneciam sinais precisos para os investidores. Os autores delimitaram-se a investigar as notícias que foram publicadas no *China Times* e no *Commercial Times*, entre os anos de 2003 a 2012, totalizando uma amostra de 1.155.757 textos.

A proposta de estimação do índice de sentimento foi obtida mediante a contagem dos termos de otimismo e pessimismo, oriundos do dicionário de Loughran e McDonald (2011), adaptado para a língua chinesa. A modelagem utilizada para verificar o reflexo do índice de sentimento no mercado foi o VAR.

Os resultados alcançados sugerem que o sentimento das notícias tem impactos, significativos, no retorno das ações, no índice de volatilidade e no volume de negociações do mercado, pois, foi verificado que, quanto maior o nível do sentimento das notícias, menor é o medo dos investidores e, conseqüentemente, existem efeitos diretos nas atividades de negociação do mercado acionário.

Os autores procuraram investigar o efeito do sentimento das notícias, estruturando um índice com frequência diária, semanal e mensal, e constataram que o mercado sofre a influência das notícias no curto prazo, quando são analisados cenários de períodos de tempo superiores a uma semana os efeitos são reduzidos. Além de analisar a relação do índice nas

movimentações do mercado, os autores estruturaram carteiras de investimento, adotando o índice como uma estratégia de negociação, constatando que o gerenciamento de carteiras baseadas em altos (baixos) índices de sentimento poderiam levar a maiores (menores) retornos para os investidores, no entanto não foi levado em consideração os custos de transação. Diante das evidências levantadas, os autores concluem que as notícias de domínio público contêm informações relevantes sobre as organizações, e, se forem devidamente filtradas, podem sinalizar informações importantes para os investidores.

O Quadro 2 apresenta uma síntese dos principais estudos empíricos sobre o sentimento textual de notícias da mídia.

Quadro 2 - Síntese dos principais estudos empíricos sobre notícias da mídia

(Continua)

Autor	Período	País	Informação	Dicionário	Método	Principais resultados	Contribuição
Tetlock (2007)	1984 a 1999	Estados Unidos	Notícias	GI/Havard	VAR	Os níveis elevados de pessimismo na mídia preveem o declínio dos preços no mercado.	Analisa o reflexo do pessimismo da mídia no mercado acionário, relacionando o sentimento da mídia como uma <i>proxy</i> para a aversão ao risco.
Tetlock, Saar-Tsechansky e Macskassy (2008)	1980 a 2004	Estados Unidos	Notícias	GI/Havard	Regressão Linear	As palavras negativas utilizadas nas notícias sobre as empresas transmitem informações negativas sobre os ganhos.	Analisa em que medida o sentimento textual negativo consegue explicar ganhos nas organizações.
Kothari, Li; Short (2009)	1996 a 2001	Estados Unidos	<i>Disclosures</i> ; Previsão dos analistas; e Notícias.	GI/Havard	Regressão Linear	Divulgações desfavoráveis levam a um aumento das medidas de risco.	Estudam o efeito que o tom de otimismo/pessimismo das informações textuais provoca no risco sistemático das empresas.
Dougal et al. (2012)	1970 a 2007	Estados Unidos	Notícias	<i>L&M</i>	Regressão Linear	O retorno do DJIA pode ser previsto no curto prazo utilizando apenas informações de uma coluna do WSJ.	Analisa os efeitos causais da mídia sobre o movimento do mercado.
Engelberg et al. (2012)	2005 a 2007	Estados Unidos	Notícias	GI/Havard <i>L&M</i>	Regressão Linear	As notícias negativas comparadas às positivas apresentam efeitos mais fortes sobre os retornos, sugerindo indícios de maior previsibilidade.	As notícias geram oportunidades de negócios para os investidores que apresentam capacidade de processamento das informações de domínio público.
García (2013)	1905 a 2005	Estados Unidos	Notícias	<i>L&M</i>	GARCH	As notícias apresentam um impacto maior sobre os retornos do DJIA, nos períodos de recessão econômica.	Verificou o efeito do sentimento das notícias no mercado de acionário ao longo dos ciclos econômicos, encontrados padrões de previsibilidade nos retornos, principalmente em períodos de recessão.

Quadro 2 – Síntese dos principais estudos empíricos sobre notícias da mídia

(Continuação)

Autor	Período	País	Informação	Dicionário	Método	Principais resultados	Contribuição
Liu e McConnell (2013)	1990 a 2010	Estados Unidos	Notícias	<i>L&M</i>	Probit	O nível de atenção da mídia e o tom dos anúncios desempenham um papel importante na decisão dos gestores nos processos de fusão.	Verificam que a mídia pode desempenhar um papel positivo na governança corporativa
Carretta <i>et al.</i> (2013)	2004 a 2010	Estados Unidos	Notícias	<i>L&M</i>	VAR Estrutural	As notícias sobre <i>spin off</i> que apresentam um sentimento negativo provocam impactos maiores no retorno, que as notícias com sentimento positivo.	Observam o reflexo do sentimento das notícias e do nível de atenção dos investidores no comportamento do mercado acionário, analisando especificamente casos de <i>spin off</i> .
Liu (2014)	2009 a 2012	Grécia; Irlanda; Itália; Portugal; Espanha.	Notícias	<i>L&M</i>	Regressões Lineares	Verifica que o aumento do pessimismo e do número de notícias provoca reduções nos preços das ações.	Analisam o sentimento textual em mercados de dívidas (créditos).
Ferguson <i>et al.</i> (2015)	1981 a 2010	Reino Unido	Notícias	<i>L&M</i>	Regressão Linear	As notícias positivas predizerem retornos futuros maiores, enquanto notícias negativas indicam retornos menores.	Estudam o impacto do tom e do volume das notícias na previsibilidade do retorno das ações de empresa com maior/menor visibilidade no mercado.
Ahern e Sosyura (2015)	2000 a 2011	Estados Unidos	Notícias	<i>L&M</i>	Regressão Logit	A precisão das notícias apresenta um impacto significativo sobre os preços, no entanto, o Mercado exagera aos rumores de uma fusão.	Investigam as características das notícias que remetem a rumores precisos/imprecisos sobre fusões empresariais e o impacto dessas notícias sobre o preço dos ativos financeiros.

Quadro 2 – Síntese dos principais estudos empíricos sobre notícias da mídia

(Continuação)

Autor	Período	País	Informação	Dicionário	Método	Principais resultados	Contribuição
Bianconi et al. (2015)	1992 a 2006	Estados Unidos	Notícias	<i>L&M</i>	CoVaR VaR ARCH- GARCH	O pessimismo dos jornais não prediz o risco extremo.	Utilização de variáveis que remetem a forma e a disseminação de informação para prever o risco sistêmico de instituições financeiras.
Strauß, Vliegenthart e Verhoeven (2016)	2002 a 2003	Holanda	Notícias	LIWC	VAR	As emoções positivas e as emoções negativas dos jornais não apresentam efeitos consistentes nos preços de abertura das ações.	Análise do impacto da emoção das notícias sobre o preço de abertura das ações.
Jandl (2016)	2001 a 2012	Reino Unido	Notícias	<i>EM;</i> <i>GI;</i> <i>MPQA;</i> <i>HE;</i> <i>L&M;</i> <i>L&M master.</i>	Regressão Linear	O sentimento textual e a atenção dos investidores podem servir como <i>proxies</i> confiáveis para explicar os movimentos de curto prazo do mercado de securitizações imobiliárias.	Análise do reflexo do sentimento textual no mercado imobiliário securitizado.
Fraiberger (2016)	1987 a 2013	Argentina; Austrália; Brasil; Índia; Indonésia; Japão; Nova Zelândia; Suécia; Suíça; África do Sul; Coréia do Sul; e Reino Unido.	Notícias	<i>L&M e Young e Soroka (2012)</i>	AR e ADL	A inclusão do sentimento textual reduz os erros de previsão.	Apresentam o sentimento textual das notícias como um importante indicador do crescimento econômico dos países.

Quadro 2 – Síntese dos principais estudos empíricos sobre notícias da mídia

(Conclusão)

Autor	Período	País	Informação	Dicionário	Método	Principais resultados	Contribuição
Liu, McConnell e Xu (2017)	1996 a 2009.	Estados Unidos	Notícias	<i>L&M</i>	Regressão Linear	Existe um vínculo econômico entre a cobertura da mídia e o desempenho dos CEOs atual em relação às oportunidades futuras para os CEOs.	Relatam uma ligação direta entre a cobertura da mídia e o capital humano dos CEOs.
Wei et al. (2017)	2003-2012	Taiwan	Notícias	<i>Tradução do L&M para a língua chinesa.</i>	VAR	O sentimento das notícias tem impactos significados sobre o comportamento do mercado e em estratégias de investimento.	O sentimento das notícias pode fornecer um valioso apoio na gestão de carteiras de investimento.

Fonte: Elaborado pelo autor com base nos estudos analisados (2017).

Os estudos empíricos apresentados mostram evidências de que o sentimento das notícias, divulgadas em jornais, afetam os movimentos do mercado acionário. Dentre os estudos que procuraram analisar o impacto do sentimento das notícias sobre os preços, foi possível constatar que as notícias têm um efeito mais intenso sobre o mercado em maus momentos econômicos, como períodos de crises e de recessões, apresentando efeitos com magnitudes significativamente expressivas, em comparação aos resultados encontrados para bons momentos econômicos.

Desse modo, os resultados levantados sobre o impacto do sentimento textual em momentos de crises e de recessão/ascensão econômica lançam dúvidas sobre a importância das notícias em períodos de maior incerteza econômica. Nesse sentido, o presente estudo buscou investigar os impactos das notícias sobre o comportamento dos preços no mercado acionário brasileiro, levando em consideração a incerteza econômica do país.

2.5 INCERTEZA ECONÔMICA

A incerteza está relacionada ao processo de tomada de decisão dos indivíduos. Uma das suas primeiras definições foi apresentada por Knight (1921), mediante uma dissociação entre risco e incerteza. Segundo Knight (1921), a incerteza pode ser descrita como a incapacidade de prever a ocorrência de eventos, em função da probabilidade de sua distribuição ser desconhecida, enquanto o risco corresponde a distribuição da probabilidade sobre o comportamento dos eventos que é conhecida.

Para Bloom (2014), a incerteza apresenta um conceito amorfo, porém, sabe-se que a incerteza está relacionada às dúvidas dos investidores, gestores e políticos sobre o futuro, e que apresenta uma forte relação com fenômenos macroeconômicos (crescimento do PIB), fenômenos microeconômicos (taxas de crescimento das empresas), bem como com eventos não econômicos (mudanças climáticas). Já Jurado, Ludvigson e Serena (2015) caracterizam a incerteza como uma volatilidade condicional de uma perturbação imprevisível que ocorre sobre as expectativas dos agentes econômico.

Conforme observado, a incerteza recebe uma definição ampla e, por isso, não existe uma medida objetiva para mensurá-la. Diante disso, os estudos empíricos costumam utilizar distintas *proxies* para captar o seu comportamento, baseadas, geralmente, em medidas de

volatilidade (BLOOM, 2009; CHULIÁ; GUILLÉN; URIBE, 2017) e em medidas de dispersão (SCOTTI, 2016).

A incerteza está relacionada às flutuações da economia e do mercado acionário. O estudo seminal de Bloom (2009) evidenciou essa relação, ao analisar a incerteza após grandes choques econômicos (como o ataque terrorista de 11 de setembro), verificando que elevados níveis de incerteza, avaliada perante a volatilidade do mercado, estavam associados a quedas no emprego e na produção, em função das empresas reduzirem investimentos e contratação de mão de obra.

Chuliá et al. (2017) também procuraram analisar a incerteza com base na volatilidade do mercado, com a proposta de uma *proxy* para a incerteza financeira dos mercados, utilizando o retorno diário de ativos, negociados na NYSE, AMEX e NASDAQ, identificando que as oscilações do índice proposto permitiram identificar vários períodos de incerteza, os quais coincidiam com eventos de alto impacto econômico, como as grandes recessões, guerras e turbulências políticas.

A estimação da incerteza com base na volatilidade do mercado pode apresentar algumas implicações, pois a volatilidade pode mudar ao longo do tempo, sem necessariamente existir uma mudança na incerteza sobre os fundamentos econômicos; a alteração na flutuação dos ativos pode ser direcionada por mudanças na alavancagem ou por alterações no sentimento sobre o mercado (aversão ao risco) (JURADO et al., 2015).

Em decorrências das limitações que as *proxies* baseadas na volatilidade do mercado podem apresentar, Jurado et al. (2015) propõe uma medida de incerteza variável no tempo, com possibilidade de monitoramento das flutuações das atividades e do mercado, utilizando séries de dados macroeconômicos, financeiros e séries de dados sobre empresa, para a construção de uma *proxy* sobre incerteza. Ao analisar as oscilações dos ciclos econômicos do Estados Unidos, Jurado et al. (2015) verificaram que a *proxy* apresentava uma estimativa mais persistente da incerteza dos períodos de recessão, em comparação aos resultados da incerteza baseada em volatilidade do mercado.

Scotti (2016) buscou investigar a incerteza existente sobre o estado real da economia, por meio da expectativa dos analistas, na tentativa de apresentar uma *proxy* com informações mais específicas sobre a economia. Assim, Scotti (2016) analisou a incerteza existente sobre os Estados Unidos, Zona do Euro, Reino Unido, Canadá e Japão, adotando o erro de previsão

diária dos analistas sobre aspectos macroeconômicos, como *proxy* para a incerteza, levantando evidências gerais de que a incerteza tende a ser elevada nos períodos de recessão.

Outra *proxy* utilizada para a estimação da incerteza é a frequência de palavras que remetem a ideia de incerteza. Alexopoulos e Cohen (2015) utilizaram o volume de notícias que apresentavam palavras-chave relacionadas à economia e à incerteza para investigar as flutuações de incerteza nos Estados Unidos, verificando que o choque de incerteza diminuía o nível da atividade econômica, aumentava a volatilidade do mercado e reduzia os retornos do mercado. Uma abordagem semelhante foi utilizada por Baker, Bloom e Davis (2016) para analisar a incerteza política.

Baker et al. (2016) estudaram a incerteza política dos países integrantes do G-10, entre 1900 a 2010, por meio de uma *proxy* baseada em de artigos de jornais norte-americanos. Os resultados alcançados mostraram que os choque de incerteza tendem a provocar efeitos econômicos negativos, levantando indícios de que a incerteza política pode ter prejudicado o desempenho macroeconômico dos Estados Unidos e da Europa, além de ter provocado efeitos negativos sobre os preços das ações, volatilidade do mercado, taxas de emprego e de investimento.

O impacto da incerteza sobre a atividade econômica no Brasil foi estudado por Costa (2014), mediante utilização de três *proxies* para incerteza, quais sejam: frequência da palavra “incerteza” nas seções econômicas de jornais; desvio padrão das expectativas do crescimento econômico; e volatilidade do mercado acionário. As evidências mostraram que a incerteza, em geral, é anticíclica, de modo que a maior incerteza ocorre em períodos de atividade econômica deprimida. Além disso, constataram que os choques de incerteza são relevantes para explicar a dinâmica da atividade econômica no Brasil.

Godoi e Lima (2017) também procuraram investigar a incerteza da economia brasileira, propondo um indicador de incerteza, baseado em uma série de dados macroeconômicos, seguindo a metodologia do estudo de Jurado et al. (2015). Godoi e Lima (2017) constataram que o índice proposto apresentava elevações nos períodos precedentes às recessões de 2009 e de 2014. Além disso, o índice mostrou-se elevado nas mudanças da política macroeconômica que ocorreu entre 2010-2011, levantando, com isso, fortes indícios de que o índice seria uma boa *proxy* para previsões de recessões e para a análise do impacto das políticas macroeconômicas no Brasil.

O Quadro 3 apresenta uma síntese dos principais estudos empíricos que levantaram evidências sobre a incerteza econômica.

Quadro 3 - Síntese dos principais estudos empíricos sobre a incerteza econômica

Autor	Período	País	Proxy	Contribuição
Bloom (2009)	1962 a 2008	Estados Unidos	Volatilidade do mercado acionário	Analisa o impacto da incerteza em grandes choques econômicos.
Alexopoulos e Cohen (2015)	1986 a 2007	Estados Unidos	Frequência de palavras de jornais	Estimam a incerteza econômica e a incerteza política mediante análise textual.
Jurado et al. (2015)	1960 a 2011	Estados Unidos	Série de dados financeiros, macroeconômicos e empresariais	Propõem uma proxy para incerteza mais persistente na identificação dos movimentos de incerteza econômica.
Scotti (2016)	2003 a 2016	Estados Unidos; Zona do Euro; Reino Unido; Canadá; Japão.	Previsão dos Analistas	Estima a incerteza econômica existente sobre o estado real da economia.
Costa (2014)	2003 a 2013	Brasil	Frequência da palavra incerteza; Variabilidade das expectativas do crescimento econômico; Volatilidade do mercado acionário.	Mostra evidências de que choques de incerteza são relevantes para explicar a dinâmica da atividade econômica no Brasil.
Baker et al. (2016)	1900 a 2010	Estados Unidos e demais países integrantes do G-10.	Frequência de palavras presente em matérias de jornais	Propõem um indicador de incerteza de políticas econômicas e levantaram evidências de que elevada incerteza política estava relacionada ao desempenho macroeconômico dos países analisados.
Godeiro e Lima (2017)	2002 a 2015	Brasil	Série de dados macroeconômicos	Propõem uma <i>proxy</i> para incerteza com potencial para prever recessões e analisar políticas macroeconômicas.
Chuliá et al. (2017)	1927 a 2014	Estados Unidos	Retornos dos Ativos	Propõem uma <i>proxy</i> para a incerteza financeira, que possibilita avaliações sobre a dinâmica da economia.

Fonte: Elaborado pelo autor com base nos estudos analisados (2017).

As evidências empíricas apresentadas mostram que a incerteza é estimada por várias *proxies*, no entanto, os resultados empíricos parecem ser convergentes quanto ao aspecto de que a incerteza é anticíclica, e tende a aumentar em períodos de recessão (BLOOM, 2009; COSTA, 2014; JURADO et al., 2015; SCOTTI, 2016). Além disso, é verificado que a incerteza

aumenta na presença de choques de eventos, como ataques terroristas, mudanças políticas e crises de crédito (CHULIÁ et al., 2017; GODOI; LIMA, 2017).

Desse modo, é possível constatar que a incerteza apresenta relação com a dinâmica da economia, porém, as economias de países emergente e de países desenvolvidos costumam apresentar um comportamento distinto em relação à incerteza. Segundo Bloom (2014), os países emergentes chegam a ter um terço a mais de incerteza macroeconômica que os países desenvolvidos.

A existência de uma maior incerteza nos países emergentes está relacionada à: (a) presença de uma economia menos diversificada, levando a uma maior vulnerabilidade atrelada às flutuações da produção e do preço dos seus produtos; (b) maior dependência de commodities, que influenciam o preço de produtos, como açúcar e petróleo; (c) maior vulnerabilidade a choque políticos, como golpes e revoluções; (d) suscetibilidade a problemas naturais, como epidemias e inundações; (e) presença de políticas de estabilização fiscal e monetária menos eficazes (KOREN; TENREYRO, 2007; BLOOM, 2014).

Diante do contexto apresentado, os países emergentes estão mais susceptíveis a incertezas econômicas, em decorrência de uma maior vulnerabilidade à existência de eventos que possam desencadear incertezas. Sendo assim, o presente estudo se propôs a investigar o impacto que o sentimento das notícias apresenta sobre o comportamento dos preços em períodos marcados por uma maior incerteza econômica.

2.6 SÍNTESE DO CAPÍTULO E HIPÓTESES DE PESQUISA

O mercado acionário apresenta indício de falhas no processo de incorporação das informações junto aos preços, levando-o a uma ineficiência temporária, que tende a desaparecer com o passar do tempo, conforme é apontado por Fama (1998). Essa não é uma peculiaridade que está restrita aos mercados acionários dos países emergentes ou dos países desenvolvidos, as evidências empíricas mostram indícios de ineficiências em ambos (JEGADEESH; TITMAN, 1993; MOSKOWITZ; GRINBLATT, 1999; SCHNUSENBERG; MADURA, 2001; ZHANG, 2006; HONG; TOROUS; VALKANOV, 2007; BARBOSA; MEDEIROS, 2007; FENG; HU, 2014).

Considerando que, temporariamente, os preços tendem a apresentar uma má precificação, é possível que os preços sejam influenciados por notícias e pelas emoções, como

é especulado, desde a década de 1930, por Keynes (1936) e evidenciado nos estudos empíricos de Tetlock (2007), Tetlock; Saar-Tsechansky; Macskassy (2008); Dougal et al. (2012); Engelberg et al. (2012); García (2013); Carretta et al. (2013); Liu (2014); Ferguson et al. (2015); Ahern; Sosyura (2015); Fraiberger (2016); Strauß, Vliegthart e Verhoeven (2016); Wei et al. (2017); Allen et al. (2017) e Shapiro et al. (2017).

Nas explicações sobre o que movem os preços no mercado, a discussão sobre o sentimento textual das notícias tem sido enaltecida, em decorrência de ser considerada uma importante medida informacional, capaz de capturar informações mais refinadas sobre questões financeiras e sobre o estado da economia, que nem sempre ficam evidentes nos dados quantitativos ou levam um certo tempo até serem incorporadas nos dados quantitativos (TETLOCK et al., 2008; WEI et al., 2017; SHAPIRO et al., 2017).

Nesse sentido, os preços podem ser influenciados pelo sentimento das notícias, em função do impacto que elas podem provocar na expectativa dos investidores sobre os ativos, fazendo com que os investidores atualizem, de algum modo, as suas crenças sobre os fundamentos das empresas. Segundo Tetlock (2014), as notícias divulgadas em mídias, como jornais, contribuem com esse processo de atualização das crenças dos investidores, e os preços no mercado tendem a expressar o comportamento médio dos investidores quanto às suas expectativas. Diante do exposto e considerando que a movimentação do mercado pode ser avaliada em termos de retorno e risco (volatilidade), tem-se a primeira e segunda hipóteses do estudo:

H₁: O sentimento das notícias exerce influência sobre o retorno das ações do mercado acionário brasileiro.

H₂: O sentimento das notícias exerce influência sobre a volatilidade do retorno das ações do mercado acionário brasileiro.

Acredita-se que os indivíduos atualizam as suas crenças com base nas informações disponíveis no mercado, seguindo a lógica do teorema da concordância de Aumann (1976). Assim, se os participantes do mercado apresentassem expectativas homogêneas, os preços seriam um sistema informacional perfeito (MILGROM; STOKEY, 1982). No entanto, os indivíduos apresentam divergências no alinhamento das suas expectativas, provocando falhas no processo de precificação dos ativos, conforme é apontado por Grossman e Stiglitz (1980).

Tetlock (2014) explica que a chegada de informações, em geral, provoca a convergência das crenças de alguns investidores, porém, outros investidores tendem a apresentar

divergências, provocadas por motivos racionais ou não. Uma das explicações atribuídas a existência dessas divergências são as restrições provocadas pelos vieses cognitivos dos indivíduos (KAHNEMAN; TVERSKY, 1979; CAMPBELL et al. 2017).

No processamento informacional, é comum as pessoas atribuírem uma maior importância às notícias negativas (PEETERS; CZAPINSKI, 1990; ANDERSEN; BOLLERSLEV; DIEBOLD; VEGA, 2003; ROZIN, ROYMAN, 2001; BAUMEISTER et al., 2001; FERGUSON et al., 2012; ALVES et al., 2017; LIU et al., 2017; COX et al., 2017). Isso ocorre devido a um viés cognitivo do processamento informacional, que leva os indivíduos a processarem as informações positivas e negativas de forma assimétrica.

No contexto financeiro, as informações negativas despertam uma maior atenção, em função da sua ameaça potencial à sobrevivência econômica (SVENSSON et al., 2017). Os estudos que procuraram investigar o valor informacional do sentimento das notícias apontam indícios de que as notícias negativas apresentam um maior valor informacional para o mercado (TETLOCK, 2007; TETLOCK et al., 2008; KOTHARI et al., 2009; DOUGAL et al., 2012; ENGELBERG et al., 2012; GARCÍA, 2013; CARRETTA et al., 2013; LIU, 2014; FERGUSON et al., 2015; WEI et al., 2017). A maior valoração atribuída às informações negativas se deve ao fato de que bons eventos tendem a dissipar muito mais rapidamente do que os eventos ruins, conforme é explicada por Baumeister et al. (2001).

Assim, a presença de uma maior concentração de notícias pessimistas na mídia pode transmitir informações importantes sobre os preços dos ativos. Nesse sentido, Tetlock (2007), Kothari et al., (2009), García (2013), Ferguson *et al.* (2015), mostram evidências de que o pessimismo da mídia tende a provocar quedas nos retornos das ações e a elevar a volatilidade.

Tendo em vista que o mercado é mais sensível às informações pessimistas, a terceira e quarta hipóteses de pesquisa são:

H₃: O sentimento pessimista das notícias reduz o retorno no mercado acionário brasileiro.

H₄: O sentimento pessimista das notícias eleva a volatilidade do retorno no mercado acionário brasileiro.

No mercado acionário, as decisões dos investidores refletem a previsão realizada sobre o comportamento futuro dos ativos, e são reveladas no sistema de preços do mercado. Assim, os preços evidenciam as expectativas médias dos agentes, e na medida que ocorrem eventos inesperados, que frustram as expectativas dos agentes, os preços sofrem alterações.

Nesse sentido, a incerteza contribui com as flutuações do mercado acionário, em função de provocar perturbações sobre as expectativas dos agentes (JURADO et al., 2015). O impacto provocado pela incerteza costuma diferir entre os países, as evidências empíricas mostram que países emergentes apresentam mais incertezas em sua economia, comparado às economias de países desenvolvidos (KOREN; TENREYRO, 2007; BLOOM, 2014). Isso significa que países emergentes apresentam uma maior vulnerabilidade à ocorrência de eventos que possam desencadear incertezas econômicas.

Ao investigar os períodos de recessão/ascensão econômica dos Estados Unidos, García (2013) levanta indícios de que as notícias da mídia contribuía com explicações sobre as movimentações do mercado, especialmente nos períodos de recessão econômica. Resultados semelhantes foram encontrados por Liu (2014) ao analisar a crise Europeia, entre 2009-2012, verificando que o pessimismo da mídia contribuía com a previsão dos *spreads* dos títulos avaliados.

Os indivíduos tendem a apresentar uma maior reação às notícias negativas, porém, as notícias negativas da mídia parecem apresentar uma maior importância nos períodos de incerteza econômica.

Considerando que o sentimento das notícias apresenta um maior valor informacional em períodos econômicos ruins, como crises (LIU, 2014) e de recessões (GARCÍA, 2013); tendo em vista que os períodos de maior incerteza estão correlacionados a eventos ruins (BLOOM, 2009; COSTA, 2014; JURADO et al., 2015; SCOTTI, 2016; BAKER et al., 2016; CHULIÁ et al., 2017); e levando em consideração que países emergentes apresentam uma maior incerteza econômica (KOREN; TENREYRO, 2007; BLOOM, 2014), a quinta e a sexta, hipóteses de pesquisa são:

H₅: O impacto do sentimento de pessimismo das notícias sobre o retorno no mercado acionário brasileiro é maior em períodos de incerteza econômica.

H₆: O impacto do sentimento de pessimismo das notícias sobre a volatilidade no mercado acionário brasileiro é maior em períodos de incerteza econômica.

Teoricamente, os preços dos ativos financeiros refletem as informações disponíveis (FAMA, 1970). No entanto, as evidências empíricas mostram que o mercado apresenta falhas temporárias no ajuste dos preços às informações (JEGADEESH; TITMAN, 1993; MOSKOWITZ; GRINBLATT, 1999; SCHNUSENBERG; MADURA, 2001; ZHANG, 2006; HONG; TOROUS; VALKANOV, 2007; BARBOSA; MEDEIROS, 2007; FENG; HU, 2014).

Em função de os preços não incorporarem todas as informações disponíveis, portanto, é possível encontrar padrões de previsibilidade nas séries de preços.

Os estudos que analisaram a conexão das notícias com a movimentação dos preços no mercado mostram evidências de que o tom das notícias apresenta valor informacional capaz de prever retornos no mercado (TETLOCK, 2007; TETLOCK; SAAR-TSECHANSKY; MACSKASSY, 2008; GARCÍA, 2013; LIU, 2014; FERGUSON et al., 2015; BIANCONI et al., 2015; SHAPIRO et al., 2017).

Ao ser investigado o comportamento do sentimento das notícias em modelos de previsão, Bianconi et al. (2015) levantaram indícios de que o sentimento era uma boa medida para a previsão de retornos médios. Já Shapiro et al. (2017) encontraram evidências de que o sentimento das notícias prever o comportamento de índices do mercado financeiro.

Assim, ao considerar que o mercado acionário brasileiro apresenta indícios de não reagir eficientemente à chegada de notícias otimistas/pessimistas (BARBOSA; MEDEIROS, 2007; MÜLLER; RIGHI; CERETTA, 2015), que o sentimento das notícias apresenta valor informacional capaz de prever os movimentos dos ativos (BIANCONI et al. 2015; SHAPIRO et al., 2017), e que o risco do mercado (volatilidade) está relacionado ao sentimento (BROWN, 1999; CAMPOS; REYES, 2017), acredita-se que o mercado brasileiro pode apresentar um padrão de previsibilidade na volatilidade do retorno associado ao tom das notícias. Isso leva a hipótese de que:

H7: O sentimento textual das notícias financeiras contribui com a previsão da volatilidade do retorno no mercado acionário brasileiro.

3 METODOLOGIA

O presente capítulo trata dos instrumentos metodológicos adotados para a realização do estudo quanto à apresentação dos dados, métrica de mensuração do sentimento textual e modelagens econométricas.

3.1 DADOS

O estudo visou investigar a relação entre o sentimento textual das notícias e as oscilações provocadas nos preços das ações do mercado acionário brasileiro. Desse modo, foram utilizados dados sobre o mercado acionário brasileiro, dados sobre o mercado acionário dos Estados Unidos e notícias financeiras textuais.

Para analisar o comportamento do mercado acionário brasileiro, foi adotado como *proxy* o IBOVESPA, por ser um índice que melhor representa as atividades do mercado brasileiro, uma vez que ele sintetiza o desempenho das maiores ações, em termos de negociabilidade e representatividade. Além disso, é um índice extremamente sensível às notícias e às especulações, que, apesar de não estarem diretamente relacionadas à sua estrutura, influenciam a oscilação do seu retorno. Como *proxies* para o comportamento do mercado acionário americano foram utilizados os índices: DJIA, índice de volatilidade das ações do S&P500, chamado de *CBOE Volatility Index* – VIX, e o índice S&P500.

Os dados sobre a série histórica dos índices: IBOVESPA, DJIA, VIX e S&P500 foram coletados no banco de dados da *Thomson Reuters*, no qual, foi extraída a série diária relacionada ao valor do fechamento dos índices em pontos, entre o período de 25 de julho de 2011 a 30 de junho 2017.

As informações referentes às notícias financeiras textuais foram obtidas mediante uma consulta às matérias publicadas no Jornal Valor Econômico e disponibilizadas para leitura na área de assinante do referido jornal. A seleção das notícias que estruturaram a amostra foi realizada com base no filtro da seção de publicação das notícias, como foi proposto por Tetlock (2007) e García (2013). Assim, foram analisadas as matérias publicadas no caderno de finanças, que se encontravam disponíveis para consulta entre os meses de julho e agosto de 2017, referentes ao período de 25 de julho de 2011 a 30 de junho de 2017, totalizando uma amostra de 45.304 matérias, conforme disposto na Tabela 1.

Tabela 1 – Volume de matérias financeiras analisadas

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total
2011							12	33	17	1034	1042	926	3.064
2012	1004	965	1085	728	980	788	808	780	626	756	650	593	9.763
2013	719	628	684	599	662	563	682	635	568	688	577	501	7.506
2014	604	620	609	563	614	559	729	685	684	703	640	629	7.639
2015	652	575	655	595	613	562	629	669	635	650	565	593	7.393
2016	563	563	560	511	539	601	587	564	537	545	522	568	6.660
2017	565	508	597	475	581	553							3.279
Total													45.304

Fonte: Notícias Financeiras do Jornal Valor Econômico

*Matérias disponibilizada no campo de assinante do Valor Econômico em jul-ago 2017.

** Amostra relacionada ao período de 25/07/2011 a 30/06/2017.

A escolha das matérias financeiras do Jornal Valor Econômico como *proxy* para as notícias do mercado foi motivada em função do mesmo ser o principal jornal brasileiro voltado para discussões financeiras e econômicas, e, por ter ênfase na apresentação das tendências e das principais movimentações do mercado financeiro brasileiro e internacional (VALOR, 2017 a,b). Desse modo, acredita-se que as matérias financeiras do jornal Valor Econômico sintetizam os acontecimentos relevantes para o mercado brasileiro.

Para fins de análise das notícias, foi levado em consideração as notícias divulgadas nos dias em que ocorrerem movimentações na bolsa de valores do Brasil - B3, bem como foram analisadas as notícias divulgadas em feriados e em finais de semana, por acreditar que os investidores tendem a reagir aos informativos no primeiro dia útil de movimentação no mercado acionário.

Desse modo, as notícias divulgadas em feriados e finais de semana foram acumuladas e os seus impactos foram analisados no primeiro dia útil de negociação da B3, seguinte à publicação. Esse mesmo tratamento foi realizado no estudo de García (2013), ao analisar o impacto das notícias do *New York Times* no mercado acionário dos Estados Unidos.

Portanto, a série temporal analisada no estudo foi de 25 de julho de 2011 a 30 de junho 2017, totalizando 1.470 observações diárias.

3.2 DEFINIÇÃO DE VARIÁVEIS

3.2.1 Sentimento Textual

As notícias foram analisadas com base nas palavras presentes nos textos, as quais foram submetidas a uma análise de conteúdo computadorizada, por meio da Linguagem de Processamento Natural. Desse modo, para a análise do sentimento textual, foi utilizada a técnica do *vector space model*, que faz o reconhecimento das palavras como vetores de identificação, para estimação do peso de uma palavra em um texto, utilizando a frequência das palavras/vetores como base para estimação do peso.

Na literatura, são encontradas algumas métricas para a estimação do peso dos vetores. Na presente tese, foi adotada a métrica apresentada por Chisholm e Kolda (1999), que faz uso de ponderações na frequência das palavras, visando melhorar o desempenho da técnica *vector space model*. Assim, a estimação do peso das palavras foi realizada por meio de uma função logarítmica, conforme Equação 4:

$$P_{i,j} \begin{cases} \frac{(1+\log(Tf_{i,j}))}{(1+\log(a_j))} \times \log \frac{N}{df_i}, & \text{se } Tf_{i,j} \geq 1, \\ 0, & \text{se } Tf_{i,j} = 0 \end{cases} \quad (4)$$

Em que:

$P_{i,j}$ - peso da palavra i no documento j ;

$Tf_{i,j}$ - total de ocorrências de uma palavra i em um documento j ;

a_j - média² 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 peso das palavras foi obtido por meio de um sistema de ponderação da frequência das palavras encontradas no conjunto de textos analisados, de modo que o termo da Equação 4,

² A média das palavras será obtida pela seguinte Equação: $a_j = \frac{1}{n} \sum_{i=1}^n FP_j$, em que n corresponde ao número de palavras do dicionário, proposto no estudo, que estão presentes no texto; e FP é a frequência das palavras que existem no texto j .

expresso por $\frac{(1+\log(Tf_{i,j}))}{(1+\log(a_j))}$, representa o peso que a palavra i apresenta no documento j , enquanto o termo $\log \frac{N}{df_i}$ faz referência ao peso global que a palavra representa dentre o conjunto de textos da amostra.

Ao analisar a frequência das palavras atribuindo uma ponderação, é possível minimizar a discrepância entre as palavras que apresentam uma alta frequência das palavras que apresentam baixa frequência, contribuindo com a minimização de *outlier*. Loughran e McDonald (2011), ao analisarem os textos dos relatórios 10-Ks, constataram empiricamente que esse tipo de métrica atenua o impacto das palavras com alta frequência.

O sentimento textual foi estimado com base no peso das palavras positivas/negativas presentes nas notícias. O procedimento de análise da categorização das palavras como positivas ou negativas foi dado mediante a utilização do dicionário de termos financeiros positivos/negativos propostos no presente estudo. Assim, para a mensuração do sentimento textual, foram consideradas apenas as palavras presentes do dicionário, sendo as demais palavras desprezadas.

Desse modo, após o cálculo do peso das palavras positivas e negativas encontradas nas notícias, foi estimado o sentimento textual das notícias mediante o índice do sentimento textual da notícia, conforme Equação 5:

$$ST_j = \frac{\sum \text{Peso das Palavras Positivas} - \sum \text{Peso das Palavras Negativas}}{\sum \text{Peso Palavras Positivas} + \sum \text{Peso Palavras Negativas}} \quad (5)$$

Assim, sentimento textual ST da notícia j foi considerado otimista, quando o índice se encontrava de próximo de 1 (um), foi neutro, quando o resultado do índice era 0 (zero), e pessimista, quando o índice apresentou um resultado próximo de -1 (menos um).

O sentimento textual da notícia foi acumulado por dia de publicação, para possibilitar a construção do índice diário do sentimento das notícias, visto que a seção de notícias analisada na tese, em média, divulgou mais de uma notícia financeira por dia. Assim, o índice do sentimento diário foi estimado com base em uma média do sentimento das notícias que saíram no dia t , conforme Equação 6:

$$ST_t = \frac{1}{n} \sum_{i=1}^n ST_{j,t} \quad (6)$$

Em que:

ST_t - é o sentimento textual médio das notícias no dia t ;

n - número de notícias; e

$ST_{j,t}$ - é o sentimento textual por notícia no dia t .

Desse modo, o índice do sentimento diário foi obtido mediante a média aritmética do sentimento textual das notícias financeiras do caderno de finanças do Jornal Valor Econômico.

3.2.2 Retorno, Volatilidade e Incerteza Econômica

Para o cálculo do retorno dos índices, foi utilizado o logaritmo natural do valor de fechamento diário do índice, mediante a subtração do valor do índice na data t pelo valor encontrado na data $t - 1$. Em relação ao IBOVESPA, o retorno foi dado pelo log-retorno do IBOVESPA, e deflacionado pelo índice de preço ao consumidor ampliado-IPCA³.

A volatilidade do retorno do índice IBOVESPA foi estimada com base no modelo de heterocedasticidade condicional autoregressivo generalizado – GARCH, proposto por Bollerslev (1986), em função de ser um modelo eficiente na estimação da volatilidade de ativos financeiros. O modelo GARCH foi definido conforme a Equação 7:

$$r_t = \mu_{t|t-1} + \varepsilon_t = \mu_{t|t-1} + \sigma_{t|t-1}z_t, \quad z_t \sim i.i.d(0,1);$$

$$\sigma_{t|t-1}^2 = \omega_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-1}^2 + \sum_{j=1}^p \beta_j \sigma_{t-1|t-j}^2 \quad (7)$$

Em que:

r_t – são os retornos de uma série financeira;

$\mu_{t|t-1}$ – média condicional do retorno em t dado um conjunto de informações passadas;

ε_t - termos de erro heterocedástico;

$\sigma_{t|t-1}$ – desvio padrão condicional da série;

z_t - corresponde ao erro, que é independente e identicamente distribuído, com distribuição de probabilidade apresentando média zero e variância um;

³ O IPCA foi coletado que foi coletado, na frequência mensal, no Instituto de Pesquisa Econômica Aplicada – IPEADATA (2017).

$\sigma_{t|t-1}^2$ - corresponde a variância condicional em t ;

ε_{t-1}^2 - corresponde ao quadrado dos erros passados ou inovações; $\sigma_{t-1|t-j}^2$ - volatilidade passada; ω_0 ; α_i e β_j - são os parâmetros.

As restrições impostas aos coeficientes do modelo GARCH são: (a) os parâmetros não podem ser negativos, portanto, espera-se que: $\omega_0 > 0$; $\alpha_p > 0$, $\alpha_i \geq 0$, $i = 1, \dots, p - 1$; $\beta_q > 0$, $\beta_j \geq 0$, $j = 1, \dots, q - 1$, pois não existe dispersão negativa; e (b) deve ser admitido que a soma dos parâmetros α_i e β_j seja inferior a 1, $(\sum_{i=1}^q \alpha_i + \sum_{j=1}^p \beta_j) < 1$, para satisfazer a condição de estacionariedade.

Desse modo, o $GARCH_{(p,q)}$ admite que a variância oscile no tempo e a sua modelagem é realizada por meio de um modelo autoregressivo, no qual a variância depende do quadrado dos erros do passado. No presente estudo, foi utilizado o GARCH (1,1).

A incerteza econômica foi avaliada mediante a utilização de duas *proxies*, quais sejam: a volatilidade do mercado e o índice de incerteza IIE-Br da Fundação Getúlio Vargas - FGV. A volatilidade do mercado corresponde a volatilidade do retorno do índice IBOVESPA, e pode ser considerada como uma medida de incerteza, em função de ser sensível aos eventos que desencadeiam expressivas flutuações sobre a economia e sobre o mercado acionário, como turbulências políticas e recessões, conforme foi identificado nos estudos realizados por Bloom (2009) e Chuliá et al. (2016).

Já o índice de incerteza IIE-Br da FGV é um indicador da incerteza da economia brasileira, proposto pelo Instituto Brasileiro de Economia - IBRE da FGV, estimado pela composição de três medidas: (a) índice de mídia – IIE-BR- Mídia, que é calculado com base na frequência de notícias da mídia que remetem à incerteza; (b) índice de expectativa – IIE-Br- Expectativa, que é baseado na dispersão das previsões sobre a taxa de câmbio e IPCA; e (c) índice de mercado - IIE-BR-Mercado, construído com base na volatilidade do mercado financeiro. Segundo o IBRE (2016), o IIE-Br é um índice estimado pela ponderação dos indicadores de mídia, de expectativa e mercado, conforme Equação 8:

$$IIE - Br = 0.7 \times IIBr_{Mídia} + 0.2 \times IIBr_{Expectativa} + 0.1 \times IIBr_{Mercado} \quad (8)$$

O índice de incerteza IIE-Br apresenta frequência mensal e é divulgado gratuitamente pelo IBRE. Para a identificação dos períodos de maior incerteza no Brasil, foi analisado o

comportamento da distribuição do índice, IIE-Br, por quartis. Admitindo que os períodos de alta incerteza foram os meses em que o índice apresentou um valor superior ao valor encontrado no terceiro quartil (75° percentil) da amostra analisada, correspondendo aos 25% dos valores mais altos da série do índice. Assim, os períodos de maior(alta)/menor(baixa) incerteza foram analisados mediante uma *dummy*, que recebeu o valor “1” nos dias pertencentes aos meses, em que, o IIE-Br foi igual ou superior ao valor do 75° percentil, e, recebeu “0” para os períodos em que o IIE-Br foi inferior ao valor do 75° percentil da amostra. A série temporal do índice IIE-Br foi de junho de 2011 a junho de 2017.

3.3 PROCEDIMENTOS DE ESTIMAÇÃO TEXTUAL

Para a mensuração do sentimento textual das notícias financeiras do valor econômico, inicialmente, foi realizada uma leitura automatizada dos textos, para extração e armazenamento de informações a serem utilizadas para o cálculo do tom das notícias.

Assim, foi realizada uma consulta aos textos financeiros, disponível no formato HTML, no ambiente de assinante, eliminando acentuações (como: ! ? ^ ~), caracteres especiais (& *), caracteres matemáticos e lógicos (como: = + - /) < >), com o intuito de evitar conflitos com os a linguagem de programação dos algoritmos utilizados para coleta dos dados e estimação do sentimento textual.

Após a limpeza das notícias, com a extração de acentuações e caracteres, as palavras presentes em cada notícia foram armazenadas no formato documento de texto (.txt), compondo um banco de dados textuais, estruturado com informações sobre a data da notícia, autor da matéria, caderno de publicação, palavras encontradas na matéria e número geral de palavras por notícia.

O banco de dados passou por um procedimento de mineração de dados textuais, mediante a utilização de algoritmos escritos em linguagem R⁴, com o auxílio de uma biblioteca de termos financeiros, buscando extrair o sentimento das notícias, visando identificar possíveis padrões de comportamento atrelados às informações textuais analisadas.

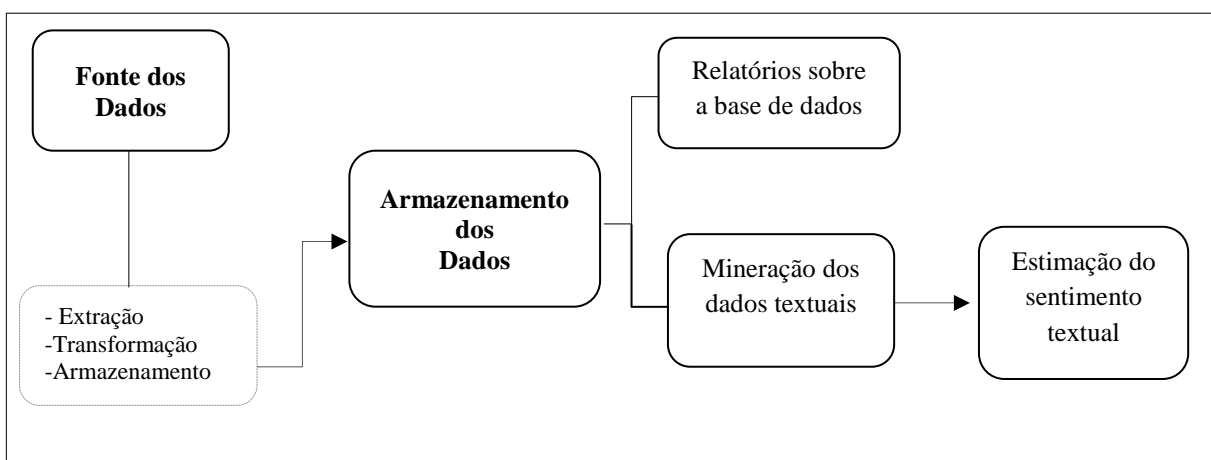
⁴ Os algoritmos propostos no estudo foram validados, como base, em resultados de estimações manuais realizadas pelo pesquisador.

Desse modo, as informações textuais do banco de dados foram submetidas a técnica do *vector space model*, para o cálculo dos pesos das palavras, conforme algoritmo escrito em linguagem R, disponível no apêndice C. Inicialmente, a programação fez uma comparação entre as palavras encontradas no banco de dados junto a uma biblioteca de termos chave em finanças, que foi estruturada para o R, correspondendo a lista de termos financeiros positivos e negativos proposta neste estudo. Em seguida, o algoritmo verificou a recorrência das vezes em que os termos surgiram ao longo das notícias analisadas e, com base nessa informação, foi estimado o peso das palavras. Por fim, foi calculado o sentimento textual das notícias, com base no peso das palavras positivas e negativas.

A programação usada na estimação do tom das notícias levou em consideração que as palavras positivas poderiam ser antecedidas por advérbios de negatividade, levando um termo, essencialmente, positivo a apresentar um sentido negativo no contexto do texto. Assim, para evitar classificações erradas, a programação⁵ localizou os dois termos que antecediam as palavras positivas e, quando os termos antecedentes faziam referência aos advérbios de negatividade: “jamais”; “não”; “nada”, “ninguém”, “nunca”, “nem”, “nenhum” e “tampouco”, as palavras positivas eram classificadas, no texto, como termos negativos. O intuito de investigar os termos precedentes das palavras positivas foi evitar a presença de um falso positivo na estimação do peso das palavras.

A Figura 3 apresenta uma síntese das etapas de coleta e extração das informações textuais.

Figura 3 - Síntese das etapas de coleta, armazenamento e extração de informações



⁵ Na leitura automatizada, escrita em linguagem R, as palavras são *strings*, que são contabilizados respeitando o modo como a palavra encontra-se escrita, segregando letras maiúsculas e minúsculas. A intenção de admitir essa regra visou evitar que termos, como, por exemplo, “Temer” e “temer” fossem contabilizados como a mesma palavra, visto que a diferença encontra-se apenas na pronúncia da palavra.





Fonte: Elaborado pelo autor.

O procedimento de análise textual adotado pode ser exemplificado na Figura 4, que apresenta um recorte de uma notícia financeira coletada no caderno de finanças do Jornal Valor Econômico, que foi publicada no dia 11 de agosto de 2017. Os termos grifados na notícia seriam palavras utilizadas no procedimento de estimação do tom da notícia.

Figura 4 - Notícia do caderno de Finanças do Jornal Valor Econômico – 11/08/2017*.

11/08/2017 às 17h44 1

Ibovespa acompanha bolsas americanas e fecha semana em alta

Por Chrystiane Silva | Valor





SÃO PAULO - Em uma semana marcada pelo **risco** de um **conflito** entre os Estados Unidos e a Coreia do Norte, o Ibovespa oscilou, mas encerrou a semana com **alta** acumulada de 0,69%. Hoje, o índice **ganhou** 0,55% aos 67.359 pontos, acompanhando o **desempenho positivo** das bolsas americanas. O giro financeiro, no entanto, foi considerado **fraco**, de R\$ 5,7 bilhões, abaixo da média diária do mês, que é de R\$ 5,8 bilhões.

Fonte: SILVA, (2017).

*Adaptação da matéria divulgada no site Valor Econômico, o texto na íntegra encontra-se em: <http://www.valor.com.br/financas/5077320/IBOVESPA-acompanha-bolsas-americanas-e-fecha-semana-em-alta>

3.4 ESPECIFICAÇÕES DOS MODELOS ECONOMETRÍCOS

3.4.1 Impacto

Para analisar o efeito que o sentimento textual das notícias financeiras provoca sobre o comportamento dos preços no mercado acionário brasileiro, foi verificado o valor informacional que o sentimento apresenta para explicar as oscilações do mercado, em termos de retorno e volatilidade, bem como foi investigada a sua capacidade de prever os movimentos dos preços no mercado.

O presente estudo concentra-se na análise da eficiência do mercado em sua forma fraca, partindo da hipótese de Fama (1970; 1991), de que os retornos passados não preveem o retorno presente, podendo ser expresso por um modelo autoregressivo, conforme Equação 9:

$$Ret_t = \beta_0 + \beta_1 \mathcal{L}_s(Ret_t) + v_t \quad (9)$$

Em que:

Ret_t - corresponde ao log-retorno do IBOVESPA no tempo t ;

β_0 - constante do modelo;

β_1 - coeficiente de correlação do retorno passado;

\mathcal{L}_s = Operador de *lag*, em que s representa o número de defasagens ou *lags*;

v_t - termo de erro ou resíduos, que espera que seja um ruído branco,

$v_t \sim N(0, \sigma^2)$, com distribuição normal, média zero e variância constante.

Desse modo, a Equação 9 descreve uma das possíveis formas de testar a eficiência em sua forma fraca, considerando que retorno presente é uma função do retorno passado adicionado de um termo de erro, que segue um ruído branco. Porém, os dados financeiros têm particularidades em suas séries, como, por exemplo, a presença de *outliers*. Essas particularidades, geralmente, levam os resíduos a não apresentarem distribuições normais, além de ausência de homocedasticidade, decorrente de oscilações na sua variância. Em função disso, os modelos passaram a incorporar as peculiaridades dos dados financeiros, dando origem aos modelos de heterocedasticidade condicional, como o modelo de heterocedasticidade condicional autoregressivo generalizado – GARCH, proposto por Bollerslev (1986), definido conforme a Equação 7.

Para testar a primeira hipótese de pesquisa, foi utilizado um modelo de regressão linear para séries temporais, com a estimação dos parâmetros por regressão quantílica e por mínimos quadrados ordinários. Assim, o modelo geral, utilizado para testar a influência do sentimento textual das notícias sobre o retorno do mercado admitiu que o retorno presente é uma função autoregressiva, que depende da volatilidade do retorno passado e do sentimento das notícias do passado, definido conforme Equação 10:

$$Ret_t = \beta_1 \mathcal{L}_s Sent_t + \gamma_1 \mathcal{L}_s (Ret_t) + \psi_1 \mathcal{L}_s (\sigma_t^2) + \eta X_t + v_t \quad (10)$$

Em que:

Ret_t - corresponde ao log-retorno do IBOVESPA;

\mathcal{L}_s - Operador de *lag*, em que s representa o número defasagens ou *lag*;

$Sent_t$ - Sentimento textual;

σ_t^2 - variância condicional do retorno do IBOVESPA;

X_t - conjunto de variáveis, que são: a constante do modelo, *dummy* para os dias da semana e *dummy* para o mês de janeiro;

v_t - termo de erro.

Na Equação 10, o coeficiente β_1 capta o efeito do sentimento textual das notícias sobre o retorno do índice que representa o mercado acionário brasileiro; Já γ_1 mostra o efeito do retorno passado sobre o retorno presente, enquanto ψ_1 estima os efeitos da volatilidade, obtida mediante um processo GARCH_(1,1), conforme Equação 7. Admitiu-se um número de 5 lags de defasagem ($s=5$), visando analisar se informações com um atraso de cinco dias provocariam efeitos sobre os preços.

Para verificar a influência que o sentimento textual das notícias financeiras exerce sobre a volatilidade do retorno do mercado, considerou-se que a volatilidade do retorno é uma função autoregressiva, da volatilidade do retorno passado e do sentimento das notícias passadas, definido conforme Equação 11:

$$\sigma_t^2 = \beta_1 \mathcal{L}_s Sent_t + \psi_1 \mathcal{L}_s (\sigma_t^2) + \eta X_t + v_t \quad (11)$$

Em que:

σ_t^2 = corresponde a volatilidade do retorno do IBOVESPA;

\mathcal{L}_s = Operador de *lag*, em que s representa o número de *lag*;

$Sent_t$ = Sentimento textual;

X_t = conjunto de variáveis, que são: a constante do modelo, *dummy* para os dias da semana e *dummy* para o mês de janeiro;

v_t = termo de erro.

Na Equação 11, o coeficiente β_1 capta o efeito do sentimento textual das notícias sobre a volatilidade do índice que representa o mercado acionário brasileiro; Já ψ_1 capta o efeito da volatilidade passada sobre a volatilidade presente.

No sentido de investigar se as notícias pessimistas e otimistas impactam o retorno e a volatilidade do retorno do IBOVESPA, visando responder as hipóteses de pesquisa 3 e 4, os modelos 10 e 11 sofreram alterações com a inclusão de uma na variável *dummy* D_{pes} , que recebeu o valor “1” para os dias, em que, a mídia foi pessimista, ou seja, dias em que o a variável sentimento textual, em média, apresentou um valor inferior a zero ($Sent < 0$), e recebeu o valor “0” quando o sentimento textual foi igual ou superior a zero ($Sent \geq 0$). Assim, os modelos passaram a ser definidos conforme a Equação 12 e Equação 13:

$$Ret_t = \beta_0 + (D_{pes})(\beta_1 \mathcal{L}_s Sent_t + \gamma_1 \mathcal{L}_s (Ret_t) + \psi_1 \mathcal{L}_s (\sigma_t^2) + \eta X_t) + (1 - D_{pes})(\beta_2 \mathcal{L}_s Sent_t + \gamma_2 \mathcal{L}_s (Ret_t) + \psi_2 \mathcal{L}_s (\sigma_t^2) + \eta X_t) + v_t \quad (12)$$

$$\sigma_t^2 = \beta_0 + (D_{pes})(\beta_1 \mathcal{L}_s Sent_t + \psi_1 \mathcal{L}_s (\sigma_t^2) + \eta X_t) + (1 - D_{pes})(\beta_2 \mathcal{L}_s Sent_t + \psi_2 \mathcal{L}_s (\sigma_t^2) + \eta X_t) + v_t \quad (13)$$

Em que:

$Sent_t$ - corresponde ao sentimento textual das notícias financeiras do dia t ;

D_{pes} - *dummy* para classificação dos dias de mídia pessimista.

A *dummy* sobre dias de mídia pessimista, D_{pes} , entrou na equação de regressão interagindo com todas as variáveis, com a finalidade de possibilitar que o modelo extraísse o efeito dos dias pessimistas. Além disso, foi realizada a inclusão da expressão “1- D_{pes} ” para extrair o efeito dos dias de mídia otimista. Esse procedimento de inclusão das *dummies* “ D_{pes} ”

e “1- D_{pes} ”, interagindo com todas as variáveis, foi realizado, com o intuito de evitar que as estimações dos coeficientes, relacionados aos períodos de média pessimista e otimista, fossem realizadas com uma subdivisão da amostra.

Portanto, no modelo 12, β_1 capta o efeito do sentimento de pessimismo das notícias financeiras sobre o retorno do IBOVESPA, e β_2 capta o efeito do otimismo das notícias sobre o retorno. Enquanto, no modelo 13, β_1 detecta o efeito do tom de péssimo das notícias sobre a volatilidade do mercado, e β_2 detecta o efeito do otimismo sobre a volatilidade.

Para analisar a influência do sentimento de pessimismo sobre o mercado nos períodos de maior incerteza econômica, foi extraída da amostra as observações de dias pessimistas, e foram realizadas adequações às Equações 12 e 13, seguindo a proposta de modelagem de García (2013), mediante análise da incerteza econômica, baseada no índice IIE-Br da FGV. Os períodos de incerteza econômicas foram inclusos no modelo mediante a utilização de uma *dummy*, D_I , que recebeu o valor “1” para os dias de maior incerteza, e “0” para os dias de menor incerteza, conforme o comportamento do índice IIE-Br⁶. Assim, o impacto do sentimento de pessimismo sobre o retorno do IBOVESPA foi avaliado conforme Equação 14:

$$Ret_t = \beta_0 + (D_I)(\beta_1 \mathcal{L}_s Sent_{P_t} + \gamma_1 \mathcal{L}_s (Ret_t) + \psi_1 \mathcal{L}_s (\sigma_t^2)) + (1 - D_I)(\beta_2 \mathcal{L}_s Sent_{P_t} + \gamma_2 \mathcal{L}_s (Ret_t) + \psi_2 \mathcal{L}_s (\sigma_t^2)) + v_t \quad (14)$$

Em que:

Ret_t - corresponde ao log-retorno do IBOVESPA;

D_I - variável *dummy* para períodos de maior incerteza econômica, classificada segundo índice IIE-Br, que assume o valor 1 em períodos de maior incerteza, e 0 para períodos com menor incerteza;

$Sent_{P_t}$ - Sentimento de pessimismo das notícias financeiras.

\mathcal{L}_s - Operador de *lag*, em que s representa o número de defasagens;

⁶ Na amostra do IIE-Br o valor encontrado para o terceiro quartil (75° percentil) foi 122.7. Portanto, nos meses em que o IIE-Br apresentou um valor superior a 122.7 os dias foram classificados como alta (maior) incerteza, recebendo uma *dummy* “1”. Nos meses em que o índice foi igual ou inferior a 122.7, foi inclusa a *dummy* “0” para classificar os dias pertencentes aos meses de baixa (menor) incerteza.

σ_t^2 - variância condicional do retorno;

u_t - termo de erro.

Na Equação 14, o coeficiente β_1 capta o efeito do sentimento de pessimismo das notícias sobre o retorno do IBOVESPA em períodos de maior incerteza e β_2 capta o efeito do sentimento de pessimismo das notícias em períodos que existe uma menor incerteza. Já γ_1 e γ_2 mostra o efeito do retorno passado sobre o retorno presente, nos períodos com maior incerteza e com menor incerteza, respectivamente, enquanto, ψ_1 e ψ_2 estima os efeitos da volatilidade sobre o retorno em períodos de maior e menor incerteza.

Os coeficientes que são alvo das inferências do modelo correspondem a β_1 e β_2 , visto que, por meio deles, é possível investigar o valor sentimento de pessimismo sobre o retorno em períodos de maior e menor incerteza econômica.

A análise da influência do sentimento de pessimismo sobre a volatilidade do mercado nos períodos de maior incerteza econômica foi realizada conforme Equação 15:

$$\sigma_t^2 = \alpha_0 + (D_t)(\beta_1 \mathcal{L}_s \text{Sent}_t + \psi_1 \mathcal{L}_s(\sigma_t^2)) + (1 - D_t)(\beta_2 \mathcal{L}_s \text{Sent}_t + \psi_2 \mathcal{L}_s(\sigma_t^2)) + u_t \quad (15)$$

Em que:

σ_t^2 - corresponde a volatilidade do retorno do IBOVESPA;

D_t - Variável *dummy* para períodos de maior incerteza econômica classificando segundo índice IIE-Br, que assume o valor 1 em períodos de maior incerteza, e 0 para períodos com menor incerteza;

\mathcal{L}_s - Operador de *lag*, em que s representa o número de defasagens;

Sent_t - medida referente aos índices do sentimento de pessimismo;

u_t - termo de erro.

Na Equação 15, o coeficiente β_1 capta o efeito do sentimento de pessimismo das notícias sobre a volatilidade em períodos de maior incerteza e β_2 capta o efeito do sentimento de pessimismo das notícias em períodos que existe uma menor incerteza. Já ψ_1 e ψ_2 estima os

efeitos da volatilidade passada sobre a volatilidade contemporânea em períodos de maior e menor incerteza.

Os principais coeficientes para as inferências do modelo correspondem a β_1 e β_2 , visto que eles mostram o impacto que o sentimento de pessimismo exerce sobre a volatilidade do mercado em períodos de maior e menor incerteza econômica, segundo o índice IIE-Br.

Em relação ao impacto do sentimento das notícias sobre o retorno e sobre a volatilidade do retorno do IBOVESPA, apresenta-se uma síntese dos sinais esperados para a variável explicativa sentimento textual no Quadro 4.

Quadro 4 - Síntese dos sinais esperados para o Sentimento Textual

Hipótese	Modelo	Sinais Esperados		Explicação
		Quantil 0.05	Quantil 0.95	
H1	Equação 10	—	—	Espera-se que o Sentimento textual provoque impacto positivo ou negativo sobre o retorno.
H2	Equação 11	—	—	Espera-se que o sentimento textual provoque impacto positivo ou negativo sobre a volatilidade, principalmente em períodos de maior oscilação (quantil 0.95).
H3	Equação 12	-	-	Espera-se que o pessimismo da mídia reduza o retorno do mercado.
H4	Equação 13	+	+	Espera-se que o pessimismo na mídia esteja associado a um aumento da volatilidade, principalmente nos momentos de maior oscilação do mercado (quantil 0.95).
H5	Equação 14	-	-	Espera-se que o pessimismo da mídia reduza o retorno do mercado nos períodos de maior incerteza econômica.
H6	Equação 15	+	+	Espera-se que o pessimismo da mídia contribua com o aumento da volatilidade, principalmente em períodos de maior incerteza econômica.

Fonte: Elaborado pelo autor.

Portanto, as Equações 10, 11, 12, 13, 14 e 15 foram utilizadas para testar a influência do sentimento textual ao longo dos menores e maiores *quantis* de distribuição do retorno e da volatilidade do retorno, com o intuito de verificar se existem comportamentos distintos da

informação, conforme a oscilação do retorno e da volatilidade, e levando em consideração períodos marcados por uma maior incerteza econômica.

O comportamento das variáveis foi analisado quando a estacionariedade das séries, pelos testes de *Dickey-Fuller Aumentado* e *Phillips-Perron*, análise da distribuição das séries, pelo teste de normalidade de *Jarque-Bera* e a autocorrelação pelo teste *Ljung-Box Q-statistic*. Na estimação dos coeficientes por regressão quantílica⁷ o comportamento das variáveis no modelo foi analisado mediante a significância do teste *t* e para analisar a diferença entre o comportamento das variáveis interquantil foi realizado o teste *wald*.

3.4.2 Previsão da Volatilidade

A análise da contribuição do sentimento textual das notícias na previsão da volatilidade foi realizada com base em um modelo autoregressivo quantílico. Assim, para determinação do modelo de previsão, admitiu-se que um modelo autoregressivo geral de ordem p pode ser expresso conforme Equação 16:

$$y_t = \theta_0 + \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + v_t, \quad t = 1, \dots, n \quad (16)$$

Em que:

y_t - volatilidade no tempo t ;

y_{t-p} - variável com p defasagens;

v_t - é um termo de erro.

Modelos autoregressivos, como o apresentado na Equação 16, segundo Koenker e Xiao (2006), podem expressos por quantis condicionais. Logo, o modelo autoregressivo quantílico pode ser descrito, conforme Equação 17:

$$Q_{y_t}(\tau | y_{t-1}, \dots, y_{t-p}) = X_t' \alpha + X_t' \gamma F_v^{-1}(\tau) \quad (17)$$

Em que:

⁷ Na regressão quantílica, representada pelo modelo geral $Q_y(\tau | X) = \beta_0 + X\beta_1 + F_v^{-1}(\tau)$, não são feitas suposições à distribuição dos erros, $(F_v^{-1}(\tau))$ (KOENKER, 2005). Desse modo, em regressão quantílica existe o relaxamento de pressupostos, como, os pressupostos de homocedasticidade e normalidade.

$Q_{y_t}(\tau|y_{t-1}, \dots, y_{t-p})$ - quantil τ da variável y_t condicionado a um conjunto de informações passadas $\{y_{t-1}, \dots, y_{t-p}\}$;

X_t' - vetor de variáveis defasadas $\{y_{t-1}, \dots, y_{t-p}\}$; $F_v^{-1}(\tau)$ - função quantílica dos erros, supondo que os erros v que sejam independentes e identicamente distribuídos i.i.d., com uma distribuição F_v ;

α e γ são parâmetros, considerando que $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{k-1})'$ e $\gamma = (\gamma_0, \gamma_1, \dots, \gamma_{k-1})'$ são vetores com $k \times 1$ parâmetros, em que α_0 e γ_0 são interceptos.

Admitindo que $Q_{y_t}(\tau|y_{t-1}, \dots, y_{t-p}) = Q_{y_t}(\tau|I_{t-1})$; $\theta(\tau) = \alpha + \gamma F_v^{-1}$, e impondo, conforme Koenker e Xiao (2006), que as variáveis são definidas pelo vetor $X_t' = \{1, y_{t-1}, \dots, y_{t-p}\}$, a equação quantílica autoregressiva condicional pode ser simplificada, conforme Equação 18:

$$Q_{y_t}(\tau|I_{t-1}) = X_t' \theta(\tau) \quad (18)$$

No presente estudo, o intuito foi prever a volatilidade do retorno das ações do IBOVESPA. Assim, foi considerado que y_{t+1} correspondeu a variável a ser prevista, conforme um conjunto de informações observadas no tempo t , determinadas por um vetor de variáveis predictoras X_t' , admitindo que $X_{t+1,t}' = X_t'$. O modelo preditivo estimou a média condicional de y_{t+1} por quantis condicionais de y_{t+1} , admitindo que $\tau \in (0,1)$ e que $E(y_{t+1}|X_t) = X_t' \alpha$, conforme Equação 19:

$$Q_\tau(y_{t+1}|X_t) = X_t' \alpha + X_t' F_\eta^{-1}(\tau) = X_t' \beta(\tau) \quad (19)$$

Em que:

$Q_\tau(y_{t+1}|X_t)$ - quantil condicional da variável volatilidade do retorno do IBOVESPA, dado um conjunto de informações X_t ;

X_t' - vetor de variáveis predictoras composto composto pela volatilidade, pelo sentimento textual, pelo índice de medo do mercado americano VIX; *pelo DJIA e S&P500*.

$F_\eta^{-1}(\tau)$ - função quantílica incondicional dos erros no tempo $t+1$, com distribuição $F_\eta(0,1)$, admitindo que tem média zero, variância 1, e que a distribuição que não depende do conjunto de informações em t .

Na previsão de uma variável, costumam ser empregada uma série de modelos de estimação, visando possibilitar a escolha de um modelo que forneça uma previsão mais precisa de uma variável. Desse modo, para realizar a escolha de um melhor modelo de previsão deve

ser analisada a função perda L , que corresponde a função perda do erro de previsão $L(e_{t,h})$, definida por Gonzalez-Rivera (2016) como uma medida de avaliação dos custos associados ao erro de previsão. De um modo geral, os erros de previsão são avaliados conforme a Equação 20:

$$e_{t,h} = Y_{t+h} - f_{t,h} \quad L = L(e_{t,h}) \quad (20)$$

Em que:

$e_{t,1}$ – erro de previsão, h passos à frente;

Y_{t+h} – variável observada no tempo $t + h$;

$f_{t,h}$ – valor da previsão da variável, obtido por meio de modelos de estimação h passos à frente; L – função perda;

$L(e_{t,h})$ - função perda do erro de previsão.

Os modelos utilizados na previsão de uma variável costumam admitir combinação de variáveis, defasagens temporais, e ainda admitem, que os coeficientes possam ser estimados, por exemplo, com OLS – *Ordinary Least Squares* e Regressão Quantílica. O melhor modelo de previsão é aquele que minimiza o valor esperado da função perda, ou seja, é o modelo que apresenta o menor erro de previsão.

Neste estudo, a estimação do modelo de previsão para a média condicional seguiu a visão de Lima e Meng (2017), admitindo que $X'_{t+1,t} = X'_t$ e considerando que a previsão $f_{t,h} = \hat{y}_{t+1}$ vem de uma estimação por regressão quantílica. Assim, o modelo de previsão para a média condicional foi determinado conforme a Equação 21:

$$\begin{aligned} \hat{y}_{t+1} &= Q_{\tau}(y_{t+1}|X_t) \\ \hat{y}_{t+1} &= E(y_{t+1}|X_t) + k_{\tau} \end{aligned} \quad (21)$$

Em que:

\hat{y}_{t+1} - previsão da volatilidade representada pela variável y_{t+1} condicionada a um conjunto de informações X_t ;

$k_{\tau} = X'_t F_{\eta}^{-1}(\tau)$ – é a uma medida de viés em relação a previsão baseada na média condicional por quantil τ . Considerando que existe uma dependência de X_t , de uma distribuição F_{η} e de uma função perda L .

A média condicional costuma ser contaminada por valores extremos, e em dados financeiros é comum que as séries apresentem uma alta concentração de *outliers*, por esses

motivos, Meligkotsidou et al. (2014) e Lima e Meng (2017) enfatizam que as estimações das previsões devem ser realizadas com instrumentos robustos a valores extremos, como a Regressão Quantílica, para que sejam minimizados os erros de previsão. Nesse sentido, Lima e Meng (2017) sugerem que previsão de variáveis financeiras seja realizada por meio de regressão quantílica combinada, conforme Equação 22:

$$\sum_{\tau=\tau_{min.}}^{\tau_{max.}} w_{\tau} Q_{\tau}(y_{t+1}|X_t) = E(y_{t+1}|X_t) + \sum_{\tau=\tau_{min.}}^{\tau_{max.}} w_{\tau} k_{\tau}$$

$$\sum_{\tau=\tau_{min.}}^{\tau_{max.}} w_{\tau} Q_{\tau}(y_{t+1}|X_t) = E(y_{t+1}|X_t) + X_t' \gamma \sum_{\tau=\tau_{min.}}^{\tau_{max.}} w_{\tau} F_n^{-1}(\tau) \quad (22)$$

Em que:

w_{τ} – peso atribuído ao quantil condicional $Q_{\tau}(y_{t+1}|X_t)$ de y_{t+1} , dados um conjunto de informações X_t .

A previsão quantílica combinada é dada pelo somatório das previsões estimadas por quantis, admitindo que a previsão tende a oscilar entre os quantis, e que, na estruturação do modelo de previsão quantílica o maior peso deverá ser atribuído ao quantil que apresentar a menor variação na estimação do erro de previsão de y .

Neste estudo, o conjunto de variáveis preditoras X_t , formado por: (i) volatilidade; (ii) sentimento textual; (iii) índice VIX; (iv) DJIA, e (v) VIX, foram submetidas a uma investigação quanto ao seu poder preditivo sobre a volatilidade, mediante a estimação de regressão quantílica ℓ_1 –penalizada (LASSO), seguindo as propostas de Lima e Meng (2017), Caporin e Poli (2017) e Belloni e Chernozhukov (2011), e utilizando uma adaptação da rotina desenvolvida por Lima e Meng (2017), para a para estimação da regressão quantílica ℓ_1 –penalizada, escrita em linguagem R.

A regressão quantílica ℓ_1 –penalizada (LASSO) possibilita uma seleção automatizada dos preditores mais fortes/fracos por quantis, ao longo do tempo. Isso permite que os modelos de previsão incorporem preditores com alto poder explicativo, removendo, conseqüentemente, preditores que são totalmente fracos, que são variáveis que não preevem nenhum quantil. Os preditores parcialmente fracos, que correspondem as variáveis preditoras da volatilidade em alguns quantis, são incorporadas na estimação da previsão, porém é levado em consideração a sua fraqueza ao ser incorporada no modelo de previsão.

Assim, admitindo que um modelo de previsão é definido conforme Equação 19, $Q_\tau(y_{t+1}|X_t) = X_t' \alpha + X_t' F_\eta^{-1}(\tau) = X_t' \beta(\tau)$, o estimador $\beta(\tau)$ da regressão quantílica ℓ_1 –penalizada (LASSO) pode ser definido conforme Equação 23:

$$\beta(\tau) = \min_{\beta_0, \beta_1} \sum_t \rho_\tau(r_{t+1} - \beta_0(\tau) - X_t' \beta_1(\tau)) + \frac{\lambda \sqrt{\tau(1-\tau)}}{m} \|\beta_1(\tau)\|_{\ell_1} \quad (23)$$

Em que:

$\sum_t \rho_\tau(r_{t+1} - \beta_0(\tau) - X_t' \beta_1(\tau))$ – corresponde a estimação do coeficiente β_τ , dentro da visão clássica de estimação por Regressão quantílica;

ρ_τ corresponde a função perda associada ao quantil τ , que equivale a $\rho_\tau(\varepsilon) \equiv \varepsilon[\tau - I_{-\infty,0}(\varepsilon)]$, em que $I_A(\cdot)$ corresponde a uma função indicadora;

$\frac{\lambda \sqrt{\tau(1-\tau)}}{m} \|\beta_1(\tau)\|_{\ell_1}$ - representa a função de penalização;

m – corresponde ao tamanho da amostra;

$\lambda \sqrt{\tau(1-\tau)}$ é a penalidade que depende do quantil escolhido, considerando que λ será o nível de penalidade admitido;

$\|\cdot\|_{\ell_1}$ função de penalidade para o vetor de preditores X_t' , admitindo que $\|\beta_1\|_{\ell_1} = \sum_{i=1}^{k-1} |\beta_{1i}|$; $X_t = (x_{1t}, \dots, x_{(k-1)t})'$.

Desse modo, o estimador de regressão quantílica penalizada corresponde a uma solução de otimização para a estimação do coeficiente $\beta(\tau)$, que possibilita a escolha dos melhores preditores por quantis em um conjunto de variáveis.

3.4.2.1 Modelos de Previsão da Volatilidade

Após a seleção dos melhores preditores da volatilidade, por quantis, ao longo do tempo por LASSO, a etapa pós-LASSO consistiu na estimação da previsão por quantil, com base no vetor de variáveis selecionadas por LASSO, admitindo que a previsão foi obtida conforme a Equação 24:

$$f_{t+1,t}^\tau = \beta_0(\tau) + X_t'^* \beta(\tau) \quad (24)$$

Em que:

$f_{t+1,t}^\tau$ - corresponde a previsão da volatilidade um passo à frente para o quantil τ ;

$X_t^{*'} -$ vetor de variáveis preditoras selecionadas por LASSO para o quantil τ , considerando um nível de significância de 5%.

Uma vez estimada a previsão de todos os quantis, definidos no intervalo $\tau \in (0,1)$, a previsão da volatilidade foi estruturada por uma combinação das previsões realizadas para os quantis, admitindo pesos fixos e variáveis para os quantis de previsão, na estimação da previsão da volatilidade por regressão quantílica combinada – RQC.

Para a análise das primeiras combinações da previsão quantílica, considerou-se pesos fixos para os quantis, assumindo dois intervalos para análise das previsões, que foram: (i) três quantis - $\tau \in (0.3; 0.5; 0.7)$; e (ii) cinco quantis- $\tau \in (0.3; 0.4; 0.5; 0.6; 0.7)$.

Desse modo, a previsão da volatilidade com pesos fixos para três quantis, foi realizada pelo modelo RQC_PF_3q, apresentado conforme Equação 25:

$$y_{t+1} = \frac{1}{3}f_{t+1,t}^{0.3} + \frac{1}{3}f_{t+1,t}^{0.5} + \frac{1}{3}f_{t+1,t}^{0.7} \quad (25)$$

A previsão da volatilidade com pesos fixos para cinco quantis, foi realizada pelo modelo RQC_PF_5q, apresentado conforme Equação 26:

$$y_{t+1} = \frac{1}{5}f_{t+1,t}^{0.3} + \frac{1}{5}f_{t+1,t}^{0.4} + \frac{1}{5}f_{t+1,t}^{0.5} + \frac{1}{5}f_{t+1,t}^{0.6} + \frac{1}{5}f_{t+1,t}^{0.7} \quad (26)$$

A segunda etapa de análise das combinações da previsão quantílica considerou pesos variáveis para os quantis, admitindo que os maiores pesos foram atribuídos aos quantis que apresentam o menor erro de estimação e assumindo dois intervalos de quantis, que foram: (i) três quantis - $\tau \in (0.3; 0.5; 0.7)$; e (ii) cinco quantis- $\tau \in (0.3; 0.4; 0.5; 0.6; 0.7)$.

Assim, a previsão da volatilidade com pesos variáveis para o intervalo de três quantis, foi realizada pelo modelo RQC_PV_3q, conforme Equação 27:

$$y_{t+1} = \sum_{\tau=\tau_1}^{\tau_3} \omega_\tau f_{t+1,t}^\tau + \varepsilon_{t+1} \quad \tau \in (0.3; 0.5; 0.7) \quad \omega_{\tau_1} + \omega_{\tau_2} + \omega_{\tau_3} = 1 \quad (27)$$

A previsão da volatilidade com pesos variáveis para o intervalo de cinco quantis, foi realizada pelo modelo RQC_PV_5q, conforme Equação 28:

$$y_{t+1} = \sum_{\tau=\tau_1}^{\tau_5} \omega_\tau f_{t+1,t}^\tau + \varepsilon_{t+1} \quad \tau \in (0.3; 0.4; 0.5; 0.6; 0.7) \quad (28)$$

$$\omega_{\tau_1} + \omega_{\tau_2} + \omega_{\tau_3} + \omega_{\tau_4} + \omega_{\tau_5} = 1$$

A seleção dos pesos nos modelos foi realizada com base em uma estimação por OLS, por meio da regressão: $y_{t+1} = \beta_0 + \beta_1 f_{t+1}^{0.3} + \dots + \beta_n f_{t+1}^\tau + \varepsilon_{t+1}$ $\tau \in (0.3, \dots, 0.7)$, de modo que, o valor dos coeficientes β_n de cada previsão por quantil f_{t+1}^τ corresponderam ao peso utilizado da combinação das previsões.

Para fins de comparação, além da estimação da regressão quantílica com a utilização de preditores selecionados por LASSO, foi realizada a estimação com base em preditores fixos para todos os quantis, admitindo os mesmos intervalos de quantis adotados para os modelos RQC. Assim, a previsão da regressão quantílica com preditores fixos no intervalo de três quantis, com pesos fixos de $\omega_\tau = \frac{1}{3}$ e com $x_{n,t}$ variáveis, foi realizada pelo modelo RQF_PF_3q, conforme Equação 29:

$$f_{t+1} = \sum_{j=1}^3 \omega_{\tau_j} \beta_0(\tau_j) + \sum_{j=1}^3 \omega_{\tau_j} \beta_1(\tau_j) x_{1,t} + \sum_{j=1}^3 \omega_{\tau_j} \beta_n(\tau_j) x_{n,t} \quad (29)$$

A estimação quantílica com preditores fixos no intervalo de cinco quantis, considerando peso fixo de $\omega_\tau = \frac{1}{5}$, foi realizado pelo modelo RQF_PF_5q, conforme Equação 30:

$$f_{t+1} = \sum_{j=1}^5 \omega_{\tau_j} \beta_0(\tau_j) + \sum_{j=1}^5 \omega_{\tau_j} \beta_1(\tau_j) x_{1,t} + \sum_{j=1}^5 \omega_{\tau_j} \beta_n(\tau_j) x_{n,t} \quad (30)$$

Além da estimação com pesos fixos, foi analisado o comportamento dos preditores fixos, admitindo que os pesos oscilam no tempo. Assim, a estimação do intervalo de três quantis foi realizada pelo modelo RQF_PV_3q e para cinco quantis utilizou-se o modelo RQF_PV_5q, conforme Equações 31 e 32:

$$f_{t+1} = \sum_{j=1}^3 \omega_{\tau_j} \beta_0(\tau_j) + \sum_{j=1}^3 \omega_{\tau_j} \beta_1(\tau_j) x_{1,t} + \sum_{j=1}^3 \omega_{\tau_j} \beta_n(\tau_j) x_{n,t} \quad (31)$$

$$f_{t+1} = \sum_{j=1}^5 \omega_{\tau_j} \beta_0(\tau_j) + \sum_{j=1}^5 \omega_{\tau_j} \beta_1(\tau_j) x_{1,t} + \sum_{j=1}^5 \omega_{\tau_j} \beta_n(\tau_j) x_{n,t} \quad (32)$$

Neste estudo, também foi estimada a previsão da volatilidade por OLS, considerando o conjunto de preditores selecionados por LASSO. Assim, foram estimados os modelos OLS_3 e OLS_5, os quais diferem-se, apenas, quanto ao conjunto das variáveis predictoras $X_{n,t}$ selecionadas no intervalo de três (0.3; 0.5; 0.7) e cinco (0.3; 0.4; 0.5; 0.6; 0.7) quantis condicionais. Desse modo, os modelos OLS_3 e OLS_5 foram estimados, conforme Equação 33:

$$y_{t+1} = \alpha_i + \beta_1 X_{1,t} + \dots + \beta_n X_{n,t} + \varepsilon_{t+1} \quad (33)$$

O poder preditivo de cada uma das variáveis preditoras da volatilidade, também, foi analisado por uma estimação de previsão, baseada em OLS univariado, seguindo a Equação 34:

$$y_{t+1} = \alpha_i + \beta_i X_{i,t} + \varepsilon_{i,t+1} \quad (34)$$

Lima e Meng (2017) destacam que dentre os modelos de previsão, RQC, RQF e OLS, a previsão realizada por regressão quantílica combinada tende a apresentar melhores estimações de previsão, pois os modelos de regressão quantílica baseada em preditores fixos e as previsões por OLS não excluem e não ponderam a incorporação de preditores fracos nos modelos de previsão, e o OLS ainda apresenta o agravante de não ser uma estimação robusta a valores extremos.

Portanto, com a finalidade de testar a hipótese de pesquisa 7, foi analisado a capacidade preditiva das variáveis preditoras da volatilidade por LASSO, e foi analisado a função perda das estimações da previsão da volatilidade, realizadas com base nos modelos: Regressão Quantílica Combinada (Eq.25; Eq. 26; Eq. 27; e Eq. 28), Regressão Quantílica com preditores fixos (Eq. 29; Eq. 30; Eq. 31; e Eq.32); OLS multivariado (Eq.33); e OLS univariado (Eq.34).

3.4.2.2 Análise da Previsão

Para realizar a previsão da volatilidade, a série temporal utilizada, neste estudo, foi dividida em duas partes, que foram: (i) amostra de estimação da previsão; e (ii) amostra de análise da previsão, chamada de previsão fora da amostra (*out-of-sample*).

A estimação dos modelos de previsão da volatilidade foi realizada com base nos dados da amostra de estimação, composta por 507 observações, com dados iniciais de 25/07/2011 a 9/08/2013. O objetivo do estudo foi realizar a previsão da volatilidade em $t+1$, ou seja, um passo à frente. Para isso, utilizou-se uma janela de estimação recursiva, que faz a previsão adicionando uma nova observação e desprezando a observação mais antiga a cada estimação realizada. Segundo Gonzalez-Rivera (2016), a vantagem desse esquema de previsão é que ele incorpora muitas informações na estimativa do modelo.

Neste estudo, a escolha pela estimação da previsão por janela recursiva foi realizada em função de ser uma estimação que agrega informações nas estimativas de previsão, e também por ser, uma estimação que costuma ser adotada em estudos sobre previsões de variáveis relacionadas ao mercado financeiro. Na pesquisa desenvolvida por Carretta et al. (2013) a

janela recursiva foi utilizada nas estimações das previsões sobre o comportamento de *spin-off* negociados na NYSE, admitindo a inclusão do sentimento textual nos modelos de previsão. A janela recursiva também foi adotada por Meligkotsidou et al. (2014) e Lima e Meng (2017) em previsões sobre o prêmio pelo risco do mercado.

A previsão fora da amostra, corresponde a janela utilizada para analisar da capacidade preditiva dos modelos de previsão da volatilidade. Neste estudo, ela foi composta por 963 observações, no período de 12/08/2013 a 30/06/2017. Porém, foi admitido, em alguns modelos, que os pesos das previsões oscilassem no tempo (modelos: *RQC_PV_3q* - Eq. 27; *RQC_PV_5q* - Eq.28; *RQF_PV_3q*- Eq.31; *RQF_PV_5q* – Eq. 32). Assim, foram desprezados 160 dias após a janela de estimação, para realizar a estimação dos pesos dos modelos, conforme foi realizado por Lima e Meng (2017). Desse modo, a previsão fora da amostra foi composta por 803 observações.

Para fins de análise, a amostra de previsão foi analisada levando em consideração três períodos, que foram: (i) Amostra Completa – período abril de 2014 a junho de 2017; (ii) Amostra Parcial I – período abril de 2014 a dezembro de 2015; e (iii) Amostra Parcial II – período janeiro de 2016 a junho de 2017.

A segregação da análise da previsão visou averiguar a robustez das previsões e o seu respectivo comportamento em períodos marcados por uma maior incerteza econômica, como o período da Amostra Parcial II, que foi contemplado por muitos escândalos políticos, tais como as denúncias de esquemas de corrupção.

Para analisar a acurácia das previsões da volatilidade, fora da amostra de previsão, foram realizadas comparações dos resultados dos modelos de previsão com um modelo padrão, baseado na média histórica das *proxies* da volatilidade, conforme Equação 35:

$$\bar{y}_{t+1} = \frac{1}{t} \sum_{m=1}^t y_m \quad (35)$$

A adoção de um modelo padrão baseado na média histórica da variável prevista como um modelo padrão de comparação segue a visão de análise de Lima e Meng (2017). Desse modo, considerando que as variáveis preditoras são informações que contribuem com a previsão da volatilidade, espera-se que os modelos estimados apresentem resultados superiores ao modelo padrão.

A análise das previsões da volatilidade foi realizada pelo teste⁸ R^2 , que teve o objetivo de investigar o poder explicativo do modelo em comparação ao modelo padrão. Para complementar a análise, foi realizada uma comparação do desempenho dos modelos previsões condicionais de \hat{y}_{t+1} , em relação ao modelo padrão \bar{y}_{t+1} . Para isso, foi realizado o teste proposto por Clark e West (2007), que é MSPE *ajust.*, um teste sobre o erro quadrático médio da previsão, que tem a hipótese nula de que os erros de previsões entre dois modelos são iguais.

Além do teste MSPE *ajust.*, a acurácia das previsões da volatilidade foi avaliada mediante o teste Diebold e Mariano (1995), que faz uma comparação do custo do erro de previsão entre modelos concorrentes, com base na função perda do erro de cada modelo, sob hipótese nula de que não existe diferença na precisão da previsão de dois modelos concorrentes.

Portanto, os testes MSPE *ajust.*, proposto por Clark e West (2007) e o teste de Diebold e Mariano (1995) foram realizados com a finalidade de verificar se os modelos de previsões da volatilidade condicionados a um conjunto de informações, como o sentimento textual, superaram a média histórica da volatilidade.

3.5 LISTA DE TERMOS FINANCEIROS

3.5.1 Construção do banco de dados para estruturação do dicionário de termos financeiros

A métrica escolhida para analisar o conteúdo dos textos foi a aprendizagem baseada em dicionário, que consiste em uma técnica computacional de leitura de informações, com base em uma lista de palavras-chave previamente definidas. Desse modo, foi estruturado um dicionário composto por duas listas de termos, formadas por palavras positivas e negativas, extraídas de jornais, uma vez que é considerado como uma fonte importante de disseminação de informações para o mercado (TETLOCK; SAAR-TSECHANSKY; MACSKASSY, 2008).

Um outro aspecto que motivou a utilização das matérias jornalísticas como base para extração de palavras-chaves sobre o contexto financeiro se deve ao fato de que as matérias costumam descrever eventos financeiros e econômicos que impactam o mercado acionário. Nesse sentido, as palavras utilizadas pela imprensa, com ênfase em finanças e economia, podem refletir uma preocupação média dos investidores (MANELA; MOREIRA, 2017).

⁸ O cálculo do teste R^2 fora da amostra foi dado por: $R^2 = \frac{1 - \text{Var}(y_t - \hat{y}_t)}{\text{Var } y_t}$, é uma medida de comparação do modelo estimado com a média histórica.

Para a elaboração da lista de termos financeiros, foi realizado um levantamento dos principais jornais do Brasil, utilizando como critério de seleção a existência do acervo digital das matérias na Biblioteca Digital do Senado Federal do Brasil e a importância informacional do Jornal para o Brasil, seguindo o Instituto Verificador de Comunicação – IVC. Desse modo, os jornais selecionados para compor a amostra para estruturação do dicionário foram: Estadão, O Globo e Valor Econômico.

O Estadão, também chamado de O Estado de São Paulo, é um diário de notícias que está em circulação desde 1875, correspondendo ao jornal mais antigo da cidade de São Paulo que ainda se encontra em circulação (ESTADÃO, 2017). Segundo dados do IVC, ele é o quarto jornal com maior circulação impressa e digital no Brasil, apresentando, em fevereiro de 2017, uma tiragem média de 211.322 por dia (IVC, 2017).

O Jornal O Globo foi fundado em 1925, no Rio de Janeiro. Inicialmente, era um noticiário com duas edições diárias, porém, na década de 1970, o jornal foi reformulado, tornando-se um noticiário matutino, com uma única edição por dia (OGLOBO, 2017). O jornal ganhou uma grande importância no país, sendo considerado o segundo maior jornal em circulação impressa e digital do Brasil (IVC, 2017) e, segundo os dados do relatório do IVC, em fevereiro de 2017, a média de exemplares do jornal em circulação impressa e digital era de 240.962 exemplares por dia (IVC, 2017).

O Valor Econômico é um jornal existente desde 2000, especializado em finanças, economia e negócios, sendo considerado como um importante canal de informações para desenvolvimento do mercado de capitais, visto que se propõe a fornecer informações relacionada aos acontecimentos sobre a esfera produtiva, visando auxiliar a tomada de decisão dos agentes econômicos e do público em geral (VALOR, 2017 a, b). O Valor Econômico corresponde ao oitavo maior jornal em circulação do Brasil, com uma média de 59.156 tiragens diárias, de exemplares impressos e digitais, segundo dados de fevereiro de 2017 presentes no relatório do IVC (IVC, 2017).

Diante da seleção dos jornais brasileiros, foram coletadas as matérias divulgadas na Biblioteca Digital do Senado Federal do Brasil, que faziam menção às discussões financeiras e econômicas. Assim, foram coletados 44.942 textos do Jornal Estadão, 36.368 textos do Jornal O Globo e 40.125 textos do Jornal Valor Econômico, totalizando uma amostra de 121.435 textos jornalísticos para a estruturação do dicionário, com matérias publicadas entre anos de 1975 a 2017.

Tabela 2 - Síntese das matérias usadas na estruturação do dicionário

Jornal	Período	Coluna	Acervo	Acervo coletado*
Estadão	1975 a 2017	Economia; Economia & negócios	68.664	44.942
O globo	1974 a 2017	Economia; O País	56.085	36.368
Valor Econômico	2000 a 2017	Brasil; Empresas; Nacional; Empresas & Tecnologia; Finanças	46.063	40.125

Fonte: Biblioteca Digital do Senado Federal do Brasil (2017)

*Volume de matérias coletadas

**Coleta de dados realizada entre 01.12.2016 a 19.05.2017

As matérias dos jornais selecionados foram coletadas diretamente no sítio da Biblioteca Digital do Senado Federal do Brasil, mediante a utilização de uma programação em linguagem *Hypertext Preprocessor* - PHP, que realizou a captura das matérias que se encontravam no formato HTML (html).

3.5.2 Estruturação do Dicionário

As matérias do Estadão, Valor Econômico e O globo, coletadas no sítio da Biblioteca Digital do Senado Federal do Brasil, foram armazenadas em um repositório de dados. Os textos originais das matérias, que estavam no formato HTML (.html), foram convertidos para o formato documento de texto (.txt). Em seguida, foram extraídos dos arquivos todas as informações numéricas, as acentuações das palavras, os sinais de pontuação do texto, tais como: (“”, ; : ? ! ‘’-), bem como foram eliminados todos os símbolos matemáticos (+, -, /, %, \$, = < > # *) e demais caracteres relacionados à linguagem de programação (\ & @ { [_).

A eliminação das acentuações, pontuações e caracteres foram necessárias para possibilitar a leitura automatizada das matérias. Assim, os textos isentos de caracteres e valores numérico passaram por uma leitura automatizada, para a identificação e contagem da frequência das palavras, mediante o algoritmo A, escrito em linguagem R, conforme apêndice A, que reconhece as palavras como *strings*, armazenando as palavras encontradas e o total de vezes que a palavra surgiu ao longo de toda a amostra de dados. Os resultados do total de palavras encontradas nas matérias estão sintetizados na Tabela 3.

Tabela 3 - Total de Palavras das Matérias

Total de palavras encontradas	5.946.365.180.478
Palavras distintas	249.181

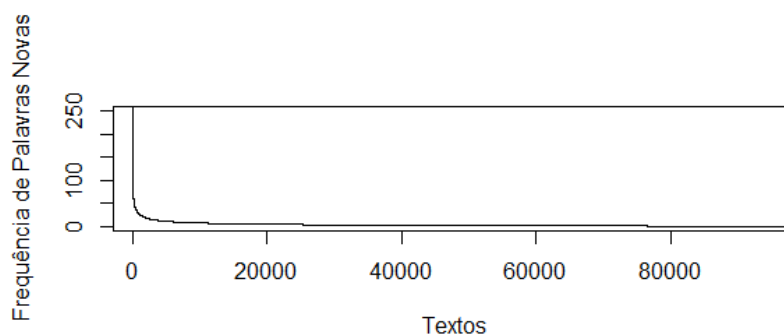
Fonte: Elaborado com base nas notícias da Biblioteca Digital do Senado Federal do Brasil

*Contagem com base no algoritmo A

**Desempenho do algoritmo 4h15min

A leitura automatizada dos 121.435 textos jornalísticos identificou um total de 5 trilhões de palavras, dentre as quais existiam 249.181 palavras distintas. Com o objetivo de identificar se o montante de palavras distintas encontradas seriam palavras suficientes para estruturar uma lista de palavras-chave sobre finanças, procurou-se analisar a diferença de palavras novas que surgiam ao longo da amostra de textos jornalísticos. Desse modo, foi construído um algoritmo para a contagem da frequência de palavras novas que surgiram ao longo da amostra de textos analisados, seguindo o *script* da programação apresentada no apêndice B. Os resultados da frequência acumulada de palavras novas que surgiram ao longo da amostra podem ser visualizados no Gráfico 1.

Gráfico 1 - Surgimento de palavras novas nas matérias jornalísticas*



Fonte: Elaborado com base nas notícias da Biblioteca Digital do Senado Federal do Brasil

Contagem de palavras realizada com o algoritmo B;

**Desempenho do algoritmo aproximadamente 120h.

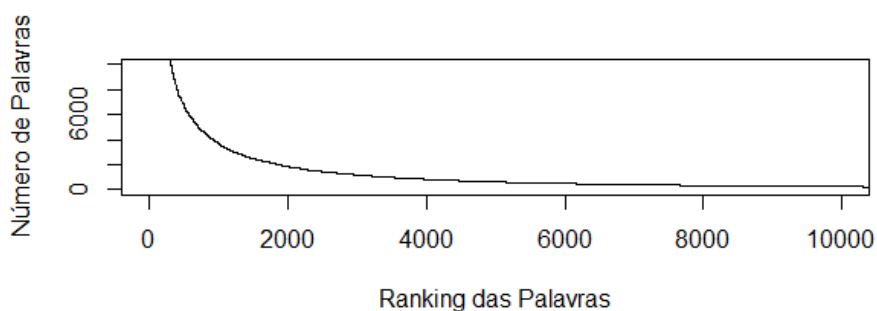
Conforme o resultado apresentado no gráfico 1, observa-se que a concentração de palavras novas está presente nos primeiros 20.000 textos analisados. Na medida em que são incorporados mais textos para análise, observa-se um decaimento no surgimento de novas palavras, chegando ao ponto de não ser encontrada nenhuma palavra nova, quando a amostra chega próximo a leitura de 80.000 textos.

Diante desses resultados, observa-se que, na amostra de 121.435 textos, o conjunto de palavras utilizadas no processo de comunicação de informações financeiras foi definida nos primeiros 80.000 textos analisados. Isso significa que, no processo de comunicação, existe um conjunto limitado de palavras que determinam o vocabulário utilizado e, portanto, a incorporação de um volume maior de textos dessa natureza, tende a não aumentar

significativamente o vocabulário. Dessa forma, acredita-se que a amostra utilizada neste estudo apresentou um valor ideal de textos para estruturação de uma lista de termos financeiros.

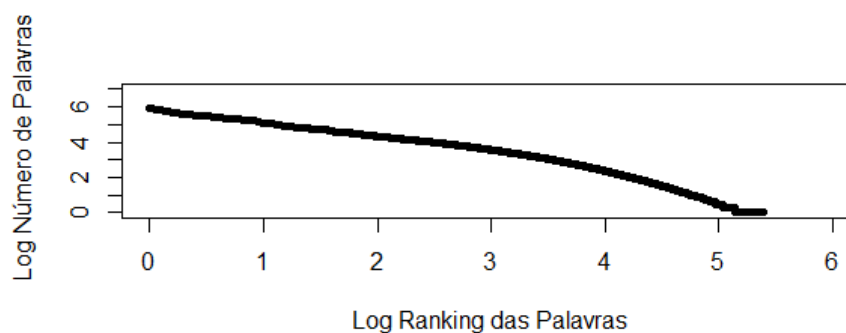
Vale salientar que o vocabulário tende a mudar ao longo dos anos, pois algumas palavras caem em desuso e novas palavras são incorporadas. Com isso, é natural que as listas de palavras-chave necessitem de refinamentos e atualizações com o passar do tempo. Por exemplo, a lista de termos financeiros mais utilizada na língua inglesa, chamada de L&M, propostas por Loughran e McDonald (2011), desde o seu surgimento da sua primeira versão, em 2009, já passou por três revisões, nos anos de 2011, 2012 e 2014. Em relação ao comportamento da distribuição de frequência de palavras novas na amostra, é possível observar, no Gráfico 1, indícios da presença de um padrão de comportamento atrelado às explicações oriunda da lei de Zipf. Para melhor investigar a distribuição das palavras, foi realizada uma análise da relação entre a frequência das palavras e o seu *ranking* dentro do conjunto total de palavras encontradas. Os resultados são apresentados no Gráfico 2 e no Gráfico 3.

Gráfico 2 - Curva da distribuição da frequência de palavras*



Fonte: Elaborado com base nas notícias da Biblioteca Digital do Senado Federal do Brasil

Gráfico 3 - Curva de Zipf para a amostra de matérias jornalísticas*



Fonte: Elaborado com base nas notícias da Biblioteca Digital do Senado Federal do Brasil

Ao observar o comportamento da distribuição das palavras na amostra de matérias dos jornais analisados, conforme Gráficos 2 e 3, verifica-se que a escolha das palavras no processo

de comunicação não é aleatória, existindo um padrão de comportamento, que pode ser observado ao ordenar as palavras, levando em consideração a sua frequência e o *ranking* dessa frequência.

Conforme pode ser observado nos Gráficos 2 e 3, as palavras com alta concentração são aquelas com os menores *rankings*, as quais tendem a não apresentar frequências iguais e estruturam o ápice da distribuição, enquanto as palavras com baixa concentração são aquelas que apresentam os maiores *rankings*, tendem a apresentar uma mesma frequência e estruturam a extremidade inferior da curva de distribuição.

Desse modo, seguindo a avaliação gráfica da curva de distribuição, tanto com os valores reais (gráfico 2), quanto os valores em logaritmo base 10 (Gráfico 3) constata-se que a distribuição das palavras da amostra tende a seguir o comportamento observado na lei de Zipf, em que existem poucas palavras que são muito citadas nos textos e muitas palavras que são pouco utilizadas. Assim, a frequência de palavras da amostra de textos analisados apresenta uma distribuição que pode ser aproximada pelas duas leis de Zipf.

Sabendo que a distribuição das palavras tende a seguir a Lei de Zipf, no procedimento de estruturação das listas de palavras-chave sobre termos financeiros, foram desprezadas as palavras situadas nas extremidades da distribuição, pois observou-se que as palavras localizadas nos pontos extremos não apresentavam significância semântica para o contexto avaliado. Constatou-se que as palavras com maiores frequências eram formadas por artigos, conjunções e preposições, enquanto as palavras com menores frequências não apresentavam conteúdo semântico, pois, eram constituídas por muitos substantivos próprios, conforme pode ser observado na Tabela 4. Esse comportamento, também, foi observado no estudo de Pao (1978), levando-o a sugerir que as extremidades da distribuição das palavras deveriam ser abandonadas nas análises de termos para determinação de palavras-chave.

Desse modo, com a finalidade de obter uma lista de palavras-chave com relevância semântica para a avaliação de informações financeiras, foram desprezadas as 100 palavras mais citadas, bem como foram eliminadas da amostra as palavras que surgiram menos de cinco vezes no conjunto de textos, totalizando 173.624 palavras. Assim, após a eliminação das extremidades, visando a eliminação de artigos, conjunções, pronomes, proposições, preposições e substantivos próprios, restaram 75.557 palavras, dentre as quais foram analisados o significado etimológico e semântico das 30.000 palavras mais citadas.

Tabela 4 - Palavras mais e menos citadas nas Matérias

Palavras	Frequência	Ranking
que	893499	1
para	407567	2
com	294548	3
nao	285288	4
uma	223609	5
Dos	202044	6
Por	176457	7
Mais	161671	8
Como	132044	9
Das	119598	10
Zypries	1	249180
Zyuganov	1	249181

Fonte: Elaborado com base nas notícias da Biblioteca Digital do Senado Federal do Brasil

A análise do significado e da origem das palavras visou a identificação de termos relacionados aos aspectos positivos e negativos. Para isso, o significado das palavras na língua portuguesa foi avaliado pelo Priberam dicionário (PRIBERAM, 2017), dicionário Michaelis (MICHAELIS, 2017) e pelo dicionário etimológico da língua portuguesa (CUNHA, 2010).

A lista de termos positivos foi estruturada por palavras que remetiam a um contexto de positividade. Para isso, analisou-se o significado e o sentido das palavras, fazendo exclusão de termos positivos que apresentavam um duplo sentido, como por exemplo: “Agregado”, “Proteger”, “Favorecer” e “Facilitar”. Assim, foi analisado o sentido da raiz das palavras nas 30.000 palavras, verificando a presença de 953 termos positivos, tais como: “Acalmar”, “Acelerar”. Em seguida, foi realizada uma varredura ao longo de toda a distribuição de palavras, com a finalidade de verificar os desmembramentos da palavra original, ou seja, para verificar a existência da palavra no plural, adjetivos ou tempos verbais distintos, sem considerar a frequência apresentada. Por exemplo, a palavra “Acelerar” encontrava-se no intervalo analisado, porém, foi realizado uma varredura para localizar as derivações das palavras que remetiam ao mesmo sentido, localizando na amostra os termos “Acelera”, “Aceleração”, “Acelerações”, “Acelerada”, “Aceleradamente”. Desse modo, foram localizadas 500 derivações das palavras fora do intervalo analisado. Portanto, a lista de termos positivos foi composta por 1.453 palavras.

A estruturação da lista de termos negativos seguiu os mesmos parâmetros utilizados na análise das palavras positivas, e foi composta por palavras que remetiam a negatividade, tais como: “Abandonar”, “Crises”, “Massacrar”. Além das palavras que exprimem negatividade, foram incorporados termos que remetiam a incerteza, como: “Medo”, “Suspeitar”, “Incerteza”,

“Volatilidade”. As palavras consideradas negativas, que poderiam remeter a um aspecto positivo, dependendo do contexto em que fossem apresentadas, como “Prender” e “Processar”, não integraram a lista de termos negativos. Assim, no intervalo analisado, foram encontradas 2.289 palavras essencialmente negativas, tais como: “Colapso”, “Corrupção”, “Déficit”, “Fraude”. Ao analisar os desmembramentos das palavras ao longo de toda a distribuição, foram localizadas mais 863 palavras, como os seguintes termos: “Corruptos”, “Fraudulento”.

Os advérbios de negação, como “Não”, “Jamais”, “Nunca”, “Ninguém” e “Nenhum”, não integraram a lista de termos negativos, pois são termos que modificam o sentido das palavras positivas. Desse modo, os termos não integraram a lista de palavras negativas. Assim, a lista de termos negativos foi composta por um total de 3.152 palavras.

O dicionário proposto, nesta tese, apresentou um total de 4.605 palavras, dentre as quais a maior parcela de termos é formada por palavras de origem negativa (3.152 palavras). Esse comportamento de maior concentração de termos negativos também foi observado na principal lista de termos financeiros da língua inglesa, proposta por Loughran e McDonald (2011), e na lista de palavras proposta por Aguiar (2012) para a língua portuguesa. Geralmente, existe um prevailecimento de termos negativos, na estruturação de listas com palavras-chave, no contexto financeiro, em função do valor informacional que as palavras negativas apresentam para os investidores.

O baixo valor incremental que as palavras positivas apresentam na descrição do contexto financeiro, segundo Liu, McConnell e Xu (2017), faz com que as palavras negativas sejam valorizadas. Além disso, as palavras negativas parecem incorporar mais elementos aos preços no mercado, conforme é apresentado por Tetlock et al. (2008), e são mais poderosas nas explicações sobre o comportamento do mercado, segundo Kearney e Liu (2014). Desse modo, pode-se dizer que a maior concentração de termos negativos ocorre em função da ameaça que eventos negativos podem provocar sobre os investidores. Com isso, a própria organização do vocabulário está, naturalmente, sujeita a um viés de negatividade, decorrente da necessidade de redução de incertezas.

3.5.3 Procedimentos de validação da lista de termos proposta

O processo de validação das listas de palavras positivas e negativas foi realizado em três etapas. A primeira etapa consistiu em uma análise do significado e da origem das palavras, por meio de uma consulta manual a dicionários da língua portuguesa. Para a análise do significado das palavras, foram utilizados Priberam dicionário (PRIBERAM, 2017) e dicionário Michaelis

(MICHAELIS, 2017). Já para a análise da origem da palavra, foi realizada uma consulta ao dicionário etimológico da língua portuguesa, de autoria de Cunha (2010). A escolha dos termos mediante consulta à dicionário configurou-se como um procedimento de validação e foi realizado pelo pesquisador.

Diante da verificação da origem e do significado das palavras presentes na lista, foi realizado um segundo procedimento de validação, que consistiu em uma análise comparativa entre o dicionário proposto e a principal lista de termos financeiros para a língua inglesa, proposta por Loughran e McDonald (2011). A lista proposta, também, foi comparada com a lista de termos positivos e negativos para a língua portuguesa de Aguiar (2012). O segundo procedimento de validação visou a verificação de possíveis erros no processo de classificação dos termos como positivos ou negativos.

As listas de termos de Loughran e McDonald (2011) surgiram após a realização de uma análise do *Harvard Dictionary*, que foi um dicionário proposto na área psicologia para a análise contextual de informações, mas não estava adequado para o contexto financeiro. Assim, Loughran e McDonald (2011) estruturam uma lista de termos, que consideraram mais apropriadas para investigações financeiras, nas seguintes categorias: (a) positivos - 354 palavras; (b) negativos - 2.354 palavras; (c) incerteza - 297 palavras; (d) litigiosos - 903 palavras; (e) restrição - 184 palavras; (f) supérfluos - 56 palavras; (g) interesse - 68 palavras; e, (h) modal - 60 palavras.

Em virtude das listas de palavras apresentadas por L&M (2011) encontrar-se na língua inglesa, foi necessário realizar a tradução dos termos e expressões positivas e negativas, com auxílio do dicionário *linguee* (2017) e do dicionário *WordNet* (2017), levando em consideração o contexto financeiro.

Em relação à lista de palavras positivas de L&M (2011), entre as 354 palavras e expressões positivas, 274 termos foram traduzidos e classificados como positivo na língua portuguesa. Já da lista de palavras negativas, comporta por 2.354 palavras e expressões, foram traduzidos 1.999 termos. A perda de termos no processo de tradução se deve ao fato de que algumas palavras perdem o sentido de positividade e negatividade, ao serem traduzidas para a língua portuguesa, como por exemplo, *Advantaged*, que é uma palavra classificada como positiva na língua inglesa, mas que perde o sentido ao ser traduzida para o português, por estar relacionada a termos, tais como: “Privilegiada”, “Favorecida”, “Vantagem”. Além disso, termos que não apresentam uma tradução com significado para o contexto financeiro foram

desconsiderados, como: *Spammers* e *Pleasure*. Desconsiderou-se, também, palavras que, ao serem traduzidas, remetiam a expressões, como: *Cybercrime* (crime cibernético).

Ao realizar uma comparação entre as listas de palavras de L&M (2011) e a lista de palavras propostas nesta tese, constatou-se que 307 palavras negativas encontradas na lista L&M estavam presentes na lista de palavras negativas propostas, e existiam 48 palavras positivas comuns entre as listas.

A lista de Aguiar (2012) corresponde a primeira lista de termos com ênfase financeira, proposta para a língua portuguesa, construída com palavras oriundas de relatórios anuais, relacionadas às seguintes categorias: (a) positivas - 701 palavras; (b) negativas – 1.080 palavras; (c) litigiosas – 492 palavras; (d) incerteza – 170 palavras; e (e) verbos modais – 81 palavras. Ao realizar uma comparação entre a lista proposta nesta tese e as listas de palavras positivas e negativas propostas por Aguiar (2012), constatou-se a presença de 556 termos negativos iguais, e 563 termos comuns entre a lista de termos proposta e a lista de termos positivos de Aguiar (2012).

Após a finalização da comparação de classificação entre as listas de termos financeiros existentes em relação às listas propostas, foi realizada a terceira etapa de validação do dicionário. Desse modo, as listas foram submetidas para um procedimento de averiguação de classificação por um lexicografista, que consiste em um profissional especialista na estruturação de dicionários e vocabulários. Assim, foi realizada uma avaliação geral das classificações de positividade e negatividade atribuídas aos termos, visando corrigir possíveis erros de classificação.

Portanto, o processo de validação das listas de palavras positivas e negativas, nesta tese, foi realizado de modo manual, por meio de três instrumentos de validação, quais sejam: (a) uso de dicionários na língua portuguesa; (b) comparação com listas de termos financeiros já existentes; e (c) validação por especialista.

4 ANÁLISE DOS RESULTADOS

Com o objetivo de responder ao problema de pesquisa proposto, mediante a não rejeição ou refutação das hipóteses levantadas, esta seção apresenta a estatística descritiva das variáveis, descreve o processo de modelamento da volatilidade, e os resultados sobre a influência do sentimento textual das notícias financeiras sobre o comportamento dos preços no mercado acionário brasileiro, assim como, apresenta os resultados obtidos com a investigação sobre o valor informacional do sentimento das notícias, sobre a previsão da volatilidade do mercado.

4.1 ESTATÍSTICA DESCRITIVA

A Tabela 5 apresenta o resultado da estatística descritiva da amostra das variáveis em estudo. Os dados foram segregados conforme o comportamento do sentimento das notícias da mídia, além de terem sido realizadas fragmentações relacionadas aos períodos de incerteza.

Ao analisar o sentimento textual, observa-se que a mídia, em geral, apresenta um comportamento de pessimismo em suas matérias, uma vez que -0.057723 foi o sentimento médio e a sua mediana foi -0.05804 , conforme os dados do painel A da Tabela 5.

Quando a amostra foi segregada conforme a incerteza do período, encontrou-se que, até nos períodos de menor incerteza no Brasil, o sentimento das notícias, em média (-0.11350), é pessimista. Diante desse contexto, percebe-se que, em média, existe uma maior utilização de palavras negativas na descrição das notícias financeiras, o que pode ser explicado pelo viés de negatividade, naturalmente existente na linguagem humana (PEETERS, 1971; ROZIN, ROYMAN, 2001; BAUMEISTER et al., 2001).

Essa predominância de um comportamento de pessimismo na mídia, observada tanto na amostra geral (Painel A), quanto nas segregações amostrais realizadas (Paineis D e E, da Tabela 5), está de acordo com os estudos anteriores dessa natureza, tais como os estudos de Tetlock (2007) e García (2013), que identificaram a existência de uma forte tendência ao pessimismo nas matérias financeiras do *WSJ* e *New York Times*, respectivamente.

Tabela 5 - Estatística Descritivas das Variáveis – 2011 a 2017

Variável	Mínimo	Média	Mediana	Máximo	Desvio Padrão	Assimetria	Curtose	Observações
Painel A: Todos os Dados								
<i>RET_IBOV*</i>	-9.2106	0.002914	-0.01431	6.38866	1.523994	-0.15245	4.95850	1470
<i>SENT</i>	-1	-0.05772	-0.05804	1	0.133039	0.387055	11.2705	1470
<i>RET_DJIA*</i>	-5.7061	0.035436	0.03443	4.15333	0.88027778	-0.4127	7.50593	1470
<i>RETS&P500*</i>	-6.8958	0.037930	0.03585	4.63174	0.9273088	-0.53524	8.64304	1470
<i>VIX*</i>	9.75	16.59893	14.78	48	5.79201585	2.139078	8.16595	1470
Painel B: Dias de Média Pessimista								
<i>RET_IBOV*</i>	-9.2106	-0.14050	-0.15837	6.38866	1.532489	-0.04924	4.84860	1041
<i>SENT_PES</i>	-1	-0.11500	-0.09704	-1.4E-05	0.094825	-2.39469	14.6249	1041
Painel C: Dias de Média Otimista								
<i>RET_IBOV*</i>	-8.4307	0.350934	0.33746	4.64731	4.647311	-0.39249	5.94632	429
<i>SENT_OTIM</i>	0	0.081286	0.05372	1	0.107464	3.864039	24.5757	429
Painel D: Períodos de Alta Incerteza** - Dados Gerais								
<i>RET_IBOV*</i>	-9.2106	0.005062	0.05419	6.38866	1.7681982	-0.43221	6.19506	400
<i>SENT</i>	-0.5750	-0.06107	-0.06053	1	0.135251	1.519708	14.1395	400
Painel E: Períodos de Baixa Incerteza** - Dados Gerais								
<i>RET_IBOV*</i>	-4.9469	0.002111	-0.03997	4.900237	1.422902	0.050993	3.46552	1070
<i>SENT</i>	-1	-0.05646	-0.0567	0.820315	0.132245	-0.06345	10.16427	1070
Painel F: Períodos de Alta Incerteza** - Média Pessimista								
<i>RET_IBOV*</i>	-9.2106	-0.13185	-0.15903	6.388665	1.773769	-0.17002	5.928464	284
<i>SENT_P</i>	-0.5750	-0.11900	-0.10641	-0.00021	0.087259	-1.20819	5.804733	284
Painel G: Períodos de Baixa Incerteza** - Média Pessimista								
<i>RET_IBOV*</i>	-4.9469	-0.14375	-0.15739	4.900237	1.432818	0.034129	3.534166	757
<i>SENT_P</i>	-1	-0.11350	-0.09462	-1.4E-05	0.097525	-2.71439	16.70832	757

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

*Valores dos índices encontra-se em pontos percentuais.

**Períodos de incerteza foram determinados conforme o índice de incerteza do Brasil - IIE_BR.

Nota: *RET_IBOV* – retorno do índice IBOVESPA, estimado pelo log-retorno e deflacionados pelo IPCA. Vale salientar que nos modelos ecométricos a variável *RET_IBOV* equivale a *Ret*; *SENT* – sentimento geral da mídia

especializada em finanças; *SENT_P*- sentimento de pessimismo da mídia, classificado de acordo com o sinal apresentado no indicador. Assim, foram considerados como *SENT_P* todos os valores do índice em que $SENT < 0$; *SENT_OTIM* - sentimento de otimismo da mídia, classificando nos casos em que $SENT \geq 0$; *RET_DJIA* – retorno do índice *Dow Jones Industrial Average*; *RETS&P500* – retorno do índice *Standard & Poor's 500*; *VIX* – índice de volatilidade baseado no preço das opções do S&P500.

Outro ponto importante a ser destacado na Tabela 5, é a estatística descritiva do Retorno do IBOVESPA (*RET_IBOV*) que, em média, foi positivo em 0.002914 pontos percentuais, porém, ao ser considerado o resultado da mediana da variável, que é uma medida robusta a valores extremos, o resultado do *RET_IBOV* foi negativo, em -0.01431 pontos percentuais. Isto sugere a presença de valores extremos, contaminando a média do retorno, que é confirmada ao ser analisada a distribuição da variável, com base nos resultados das medidas de assimetria (-0.15245) e de curtose (4.958508), de acordo com os dados do Painel A da Tabela 5, e diante da ausência de normalidade na distribuição da variável, conforme resultado do teste *Jarque-Bera*, que consta na Tabela 6.

Desse modo, os resultados encontrados na estatística descritiva do *RET_IBOV* mostram a presença de valores extremos, em todas as fragmentações amostrais, sugerindo a utilização de instrumentos de análise que sejam robustos a *outliers*. Nas séries financeiras do Brasil, é comum uma alta concentração de valores extremos nos retornos e na volatilidade, segundo Balcilar et al. (2016), isso faz com que estimações médias para o Brasil não sejam robustas e confiáveis para explicar o comportamento de variáveis financeiras. Diante disso, neste estudo, as estimações foram realizadas por meio de regressão quantílica, tendo em vista tratar-se de um instrumento de análise mais eficiente e mais robusto, que estimações médias por meio do OLS, quando existe ausência de normalidade e presença de *outliers* (KOENKER, 2005; MELIGKOTSIDOU et al., 2014).

A Tabela 6 apresenta a estatística dos testes realizados sobre o comportamento da distribuição das variáveis, como a estacionariedade das séries e dependência temporal.

Conforme os resultados apresentados na Tabela 6, em nenhuma das séries foram encontradas evidências de distribuição normal, uma vez que foi rejeitada a hipótese nula do teste *Jarque-Bera*. Os testes *Dickey-Fuller Aumentado* e *Phillips-Perron* foram realizados, a fim de verificar a estacionariedade das séries, e os resultados levam à rejeição da hipótese nula de existência de raiz unitária, concluindo que todas as séries são estacionárias.

Tabela 6 - Testes De Especificação Das Variáveis

	RET_IBOV	SENT	RET_DJIA	RET_S&P500	VIX
<i>Jarque-Bera</i>	240.634 (0.000)	4226.334 (0.000)	1285.313 (0.000)	2020.632 (0.000)	2755.622 (0.000)
<i>Dickey-Fuller Aumentado</i>	-38.498 (0.000)	-37.330 (0.000)	-40.529 (0.000)	-40.218 (0.000)	-5.241 (0.000)
<i>Phillips-Perron</i>	-38.550 (0.000)	-38.073 (0.000)	-40.899 (0.000)	-41.273 (0.000)	-4.669 (0.000)
<i>Q1* (1)</i>	0.041 (0.840)	2.946 (0.086)	4.662 (0.031)	3.488 (0.062)	1364.500 (0.000)
<i>Q1* (2)</i>	0.161 (0.923)	5.686 (0.058)	7.423 (0.024)	5.878 (0.053)	2648.900 (0.000)

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

*Q1 (k) refere-se a estatística do teste *Ljung-Box Q-statistic* para as séries temporais com *k* defasagens

Nota: Os valores entre parênteses correspondem ao *p-valor* da estatística dos testes

Ao ser realizado o teste *Ljung-Box* para verificar a existência de dependência temporal nas variáveis. Conforme a estatística *Q* do *Ljung-Box*, observam-se evidências de dependência temporal nas séries do *RET_DJIA* e do *VIX*, a uma significância de 5%. Não foram encontradas evidências de dependência temporal na amostra analisada do *RET_IBOV*, sugerindo que não é possível realizar a previsão do retorno do IBOVESPA na frequência diária.

4.2 ESTIMAÇÃO DA VOLATILIDADE – GARCH (1,1)

A estimação da volatilidade do retorno do IBOVESPA foi realizada com base no modelo de heterocedasticidade condicional GARCH (1,1), em função de ser uma medida eficiente para estimações da volatilidade de ativos financeiros e por ser muito utilizada para modelar a volatilidade do mercado acionário brasileiro (GODEIRO; LIMA, 2017).

A Tabela 7 apresenta os resultados das estimativas e dos testes do GARCH (1,1), assumindo que o retorno tem média constante e a volatilidade oscila no tempo, conforme a Equação 7:

$$\sigma_{t|t-1}^2 = \omega_0 + \sum_{i=1}^q \alpha_i \varepsilon_{t-1}^2 + \sum_{j=1}^p \beta_j \sigma_{t-1|t-j}^2 \quad (7)$$

Os resultados para o GARCH (1,1), admitindo distribuição normal, evidenciam significância em todos os coeficientes e satisfazem as restrições ausência de dispersões

negativas, pois não foram encontrados parâmetros negativos, uma vez que a constante (ω_0) foi superior a zero, assim como, o termo de inovação (α_1) e o termo autoregressivo (β_1); os resultados do modelo também satisfazem a condição de estacionariedade, pois o somatório do termo de inovação com o termo autoregressivo foi inferior a 1 ($\alpha_1 + \beta_1 = 0.9679$), atendendo à condição de estacionariedade. Os testes dos resíduos, quanto à normalidade (*Jarque-Bera*), heterocedasticidade (*ARCH-LM(1)*) e autocorrelação (*Ljung-Box Q-statistic*), apontam problemas apenas para a ausência de distribuição normal, originada pela presença de *outliers* na série do retorno.

Tabela 7 - Estimativas do Modelo GARCH(1,1)

	GARCH(1,1)		GARCH(1,1)	
	Distribuição Normal		Distribuição <i>t-Student</i>	
	<i>coef.</i>	<i>t-stat</i>	<i>coef.</i>	<i>t-stat</i>
Constante, ω_0	0.072548	3.197486**	0.054872	2.107722**
Termo de Inovação, α_1	0.047684	4.781767***	0.043226	3.472527***
Termo Autoregressivo, β_1	0.920106	51.33966***	0.93223	45.39043***
Critério Akaike	3.629784		3.603274	
Critério Schwarz	3.644187		3.617677	
Critério Hannan-Quinn	3.635155		3.608645	
	<i>Estatística do teste</i>	<i>p-valor</i>	<i>Estatística do teste</i>	<i>p-valor</i>
<i>Jarque-Bera</i>	258.978	0.0000	277.9532	0.0000
<i>ARCH(1)</i>	0.391599	0.5316	0.235968	0.6272
<i>QI(2)</i>	0.0531	0.974	0.041	0.98

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro, disponível na Thomson Reuters.

Nota: Estimação para a equação da variância; *Jarque-Bera*: teste de normalidade dos resíduos; *ARCH-LM(1)*: teste de heterocedasticidade; *QI(k)* refere-se à estatística do teste *Ljung-Box (Q-statistic)* para testar a autocorrelação dos resíduos com *k* defasagens

* Significante ao nível de 10%.; ** significativo ao nível de 5%; ***significante ao nível de 1%.

O GARCH (1,1) com distribuição *t-Student* também apresenta significância em todos os parâmetros e a modelagem satisfaz às condições de ausência de dispersões negativas, pois, os coeficientes apresentaram significância estatística e foram superiores a zero. Além disso, foi satisfeita a condição de estacionariedade ($\alpha_1 + \beta_1 = 0.9754$). Em relação aos testes dos resíduos, observou-se ausência de heterocedasticidade e de autocorrelação, sendo o único problema levantado a ausência distribuição normal, certamente provocada pela presença de valores extremos na série temporal do retorno do IBOVESPA.

Diante dos resultados, percebe-se que ambos os modelos captam a volatilidade. No entanto, analisando os critérios informacionais, observa-se que o modelo com distribuição *t-*

Student modela melhor a volatilidade, pois apresenta os menores critérios de *Akaike*, *Schwarz* e *Hannan-Quinn*.

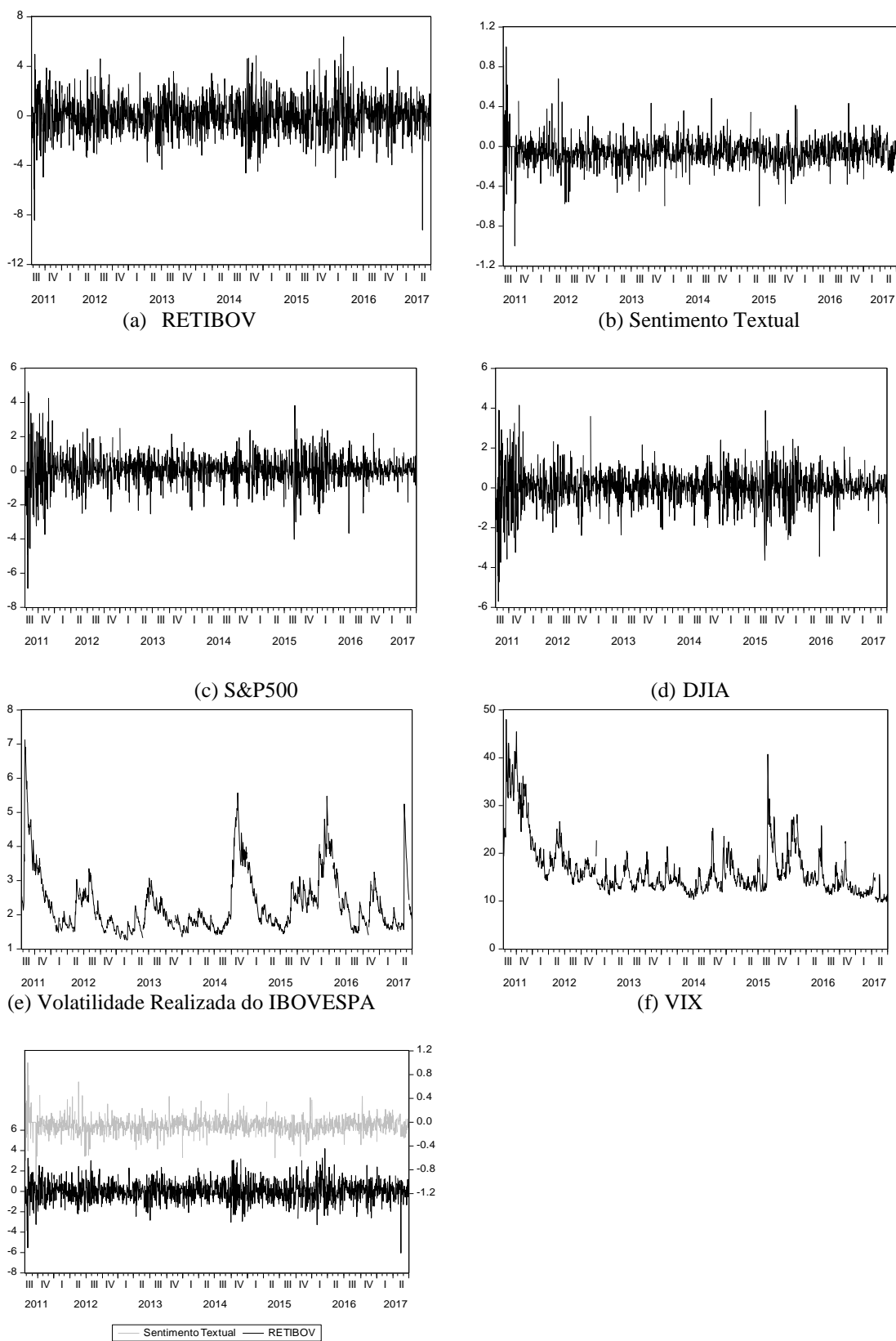
Considerando que a volatilidade do retorno do IBOVESPA pode ser modelada por um GARCH (1,1), admitiu-se a série da variância condicional do retorno do IBOVESPA, oriunda do modelo GARCH (1,1) com distribuição *t-Student*, como uma *proxy* para a volatilidade realizada do IBOVESPA. Na literatura que trata sobre sentimento textual, o GARCH (1,1) costuma ser empregado para modelar a volatilidade realizada do retorno (GARCÍA, 2013; BIANCONE et al., 2015).

A série temporal da volatilidade realizada do IBOVESPA passou por uma análise, no intuito de verificar a existência de dependência temporal. Para isso, foi realizada uma análise gráfica da função de autocorrelação (AC) e da função de correlação parcial (PAC), além de ser realizado o teste de autocorrelação de *Ljung-Box Q-statistic*, conforme o correlograma que consta no Anexo B. Os resultados encontrados mostram evidências de dependência temporal, portanto, é possível realizar a previsão da volatilidade diária do IBOVESPA.

4.3 ANÁLISE GRÁFICA DAS VARIÁVEIS

As séries apresentadas na Figura 5 mostram, claramente, um comportamento estacionário com a presença de alguns picos de oscilação. Realizando uma comparação entre o Retorno do IBOVESPA, (a) *RET_IBOV*, com o retorno dos principais índices do mercado americano, (c) *S&P500* e (d) *DJIA*, percebe-se que o índice do mercado brasileiro apresenta mais oscilações, esse resultado, também, poder ser constatado ao comparar o resultado do desvio padrão dos índices, conforme Painel A da Tabela 5. Consequentemente, ao serem comparadas, as volatilidades do retorno do IBOVESPA, gráfico (e), e do mercado dos Estados Unidos, com o índice (f) *VIX*, constata-se a presença de maiores picos de oscilação na volatilidade do mercado brasileiro, que é um comportamento característico dos mercados de países emergentes.

Figura 5 - Distribuição das séries temporais analisadas



Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

As fortes oscilações da volatilidade estão concentradas nos períodos de: agosto de 2011 (rebaixamento da nota de crédito dos Estados Unidos e reflexos da crise internacional); novembro de 2014 (especulações sobre a estagnação da indústria chinesa); março de 2016 (denúncia de corrupção da operação “Lava Jato” envolvendo a Petrobrás); e maio de 2017 (divulgação de envolvimento do Presidente do Brasil, em exercício, Michael Temer, em esquema de corrupção). Diante dessa relação, entre a volatilidade e os principais fatos do período, é possível perceber o quanto o mercado acionário brasileiro é sensível às notícias de ordem econômica e política.

Ao analisar especificamente o comportamento do índice do sentimento das notícias, observa-se que as maiores oscilações do pessimismo da mídia ocorreram em: agosto e setembro de 2011, diante de matérias sobre a crise financeira internacional; em meados de 2012, com a divulgação de notícias financeiras, trazendo discussões sobre a incerteza política no Brasil; junho e outubro de 2015, com matérias contendo discussões sobre uma possível crise financeira no Brasil, e especulações da saída de Dilma Rousseff da Presidência do Brasil; Março de 2016, com notícias sobre as denúncias de corrupção da operação “Lava jato”; e Maio de 2017, com matérias sobre o envolvimento do Presidente, Michael Temer, em um escândalo de corrupção.

Diante do cenário apresentado, percebe-se que o período temporal analisado neste estudo, foi marcado por importantes acontecimentos econômicos e políticos, ocorridos no âmbito nacional e internacional.

4.4 DESCRIÇÃO DAS PALAVRAS

A Tabela 8 apresenta uma descrição da frequência das palavras mais citadas nas matérias financeiras. O Painel A faz uma descrição das palavras gerais, sem fazer qualquer relação ao dicionário proposto neste estudo, enquanto o Painel B apresenta apenas as palavras presentes no dicionário proposto e que foram as mais citadas nas matérias avaliadas.

Assim, na Tabela 8, o Painel A, evidencia-se que as palavras mais citadas nas notícias financeiras analisadas, foram: “que” e “para”, as quais correspondem a termos que podem ser classificados como conjunções e preposições. Já as palavras menos citadas estão associadas a substantivos próprios, como: “Zurich” e “Zutphen”. Assim, pode-se dizer que as palavras mais (menos) citadas são termos que não apresentam conteúdo semântico para extração do sentimento textual. Esse é um comportamento que também foi observado por Pao (1978), ao

verificar a existência de um baixo valor informacional nas palavras com maior (menor) frequência em textos.

Tabela 8 - Descrição das palavras presentes nas notícias financeiras

Painel A: Palavras Gerais					
<i>Mais citadas</i>			<i>Menos citadas</i>		
<i>Palavras</i>	<i>Frequência</i>	<i>Ranking</i>	<i>Palavras</i>	<i>Frequência</i>	<i>Ranking</i>
que	576.509	1	Zurich	1	139735
para	381.523	2	Zuricher	1	139736
com	253.578	3	Zurrer	1	139737
uma	169.075	4	Zutphen	1	139738
dos	167.241	5	Zuzzi	1	139739
mais	149.498	6	Zwolle	1	139740
não	130.518	7	ZWZZT	1	139741
por	115.279	8	Zynga	1	139742
mercado	102.396	9	Zynler	1	139743

Painel B: Palavras do Dicionário Proposto					
<i>Palavras negativas mais citadas</i>			<i>Palavras positivas mais citadas</i>		
<i>Palavras</i>	<i>Frequência</i>		<i>Palavras</i>	<i>Frequência</i>	
Juros	54.583	1	Crescimento	31.422	1
Inflação	33.869	2	Investimentos	20.676	2
Risco	30.036	3	Expectativa	16.884	3
Dívida	20.666	4	Lucro	13.297	4
Caiu	16.445	5	Desempenho	12.994	5
Crise	16.059	6	Recuperação	12.495	6
Redução	14.564	7	Ganhos	9.857	7
Contra	10.470	8	Expectativas	9.587	8
Perdas	9.530	9	Positivo	9.471	9
Reduzir	9.368	10	Valorização	9.178	10
Riscos	9.022	11	Expansão	7.761	11

Fonte: Baseado nas notícias de finanças do Jornal Valor Econômico e nos dicionários de termos positivos e negativos, que constam nos apêndices D e E.

*Contagem realizada com o algoritmo disponível no apêndice C.

**Desempenho, médio, do algoritmo 4h.

Ao analisar as palavras que realmente apresentam valor informacional para a estruturação do sentimento textual, conforme dados apresentados no Painel B da Tabela 8, observa-se que existe uma maior utilização de palavras negativas, uma vez que os termos mais citados foram: “Juros” (54.583) e “Inflação” (33.869). Fazendo uma comparação entre a frequência das palavras negativas e positivas, verifica-se que as palavras negativas apresentaram frequências superiores às palavras positivas mais citadas, tais como: “Crescimento (31.422) e “Investimentos” (20.676). Isso significa que, na redação das notícias

financeiras, existe uma valorização de termos negativos, explicada pelo viés de negatividade, que se faz presente no vocabulário dos indivíduos, e que pode ser percebido, ao verificar que, em geral, existe uma maior riqueza de termos no vocabulário para descrever aspectos negativos (PEETERS, 1971; ROZIN, ROYMAN, 2001; BAUMEISTER et al., 2001).

Vale salientar que a valorização dos termos negativos na linguagem humana não está atrelada necessariamente ao pessimismo dos indivíduos, mas sim a um comportamento de defesa às possíveis ameaças contra o seu patrimônio. Logo, os indivíduos tendem a utilizar mais termos negativos e atribuir uma maior importância às informações negativas, com o intuito de minimizar perdas.

4.5 IMPACTO DO SENTIMENTO TEXTUAL

4.5.1 Influência do sentimento textual sobre o retorno e sobre a volatilidade

Na literatura sobre finanças, são encontradas evidências de falhas no processo de incorporação das informações juntos aos preços, desencadeando uma ineficiência temporária no mercado, que tende a ser revertida com o passar o tempo (JEGADEESH; TITMAN, 1993; FAMA, 1998; BARBOSA; MEDEIROS, 2007).

Ao ser admitida a existência temporária de ativos mal precificados no mercado, é provável que os que os preços dos ativos financeiros, no Brasil, possam ser influenciados por notícias, conforme especulações realizadas desde Keynes (1936). Um dos primeiros a investigar o poder informacional da mídia sobre o mercado acionário foi Tetlock (2007), evidenciando que era possível analisar os movimentos do mercado, com base em variáveis de sentimento, estruturadas pela contagem das palavras utilizadas nas matérias dos jornais. Nesse contexto, procurou-se analisar o poder informacional da mídia sobre os movimentos do risco (volatilidade) e do retorno do mercado brasileiro, adotando uma estratégia de estimação para o sentimento textual das notícias, que respeita as peculiaridades da língua portuguesa. Assim, as hipóteses de pesquisa avaliadas, nesta seção, foram:

H₁: O sentimento das notícias exerce influência sobre o retorno das ações do mercado acionário brasileiro.

H₂: O sentimento das notícias exerce influência sobre a volatilidade do retorno das ações do mercado acionário brasileiro.

A Tabela 9, Painei A, apresenta os resultados da regressão do retorno do IBOVESPA com o sentimento textual, em um modelo respaldado na ideia da eficiência do mercado, na forma fraca, conforme Fama (1970; 1991). Nesse sentido, foi investigado se o desempenho do retorno poderia ser explicado pelo sentimento textual⁹ das notícias financeiras, admitindo uma defasagem de 5 *lags*, com o intuito de analisar se as informações da mídia provocavam algum efeito sobre os preços, e se o efeito era persistente ao longo do tempo, ou, se o comportamento, apresentado, era temporário (apenas em t ou $t-1$). A escolha por realizar uma defasagem de 5 períodos (5 *lags*), foi feita em função das evidências empíricas (TETLOCK, 2007; GARCÍA, 2013, FERGUSON et al., 2015) apontarem que, em média, a informação divulgada na mídia apresenta efeitos sobre o mercado em até quatro dias após a divulgação das notícias, depois disso, existe uma tendência dos preços incorporarem as informações e, conseqüentemente, deixam de serem encontrados efeitos significativos do tom das notícias sobre o mercado.

Além do sentimento textual, foi considerado que as diferenças do desempenho do retorno do mercado poderiam ser explicadas pelo retorno passado, Ret_{t-n} , e pela volatilidade do mercado, $Volat_{t-n}$, admitindo, também, uma defasagem de 5 períodos. Adicionalmente, foram inclusas *dummies* temporais, relacionadas aos dias do mês de janeiro e aos dias da semana, para controlar os possíveis efeitos de oscilações atreladas à anomalia de calendário.

A análise das estimações dos coeficientes realizadas por Regressão Quantílica ocorreu sobre os quantis que representam os menores desempenhos do mercado, que são os retornos mais baixos (*quantil 0.05* e *quantil 0.25*), e os melhores desempenhos do mercado, retornos mais altos (*quantil 0.75* e *quantil 0.95*). Além do comportamento dos extremos da distribuição do desempenho do mercado acionário brasileiro, observou-se, também, o comportamento da estimação da mediana (*quantil 0.5*), visando verificar a consistência do comportamento das variáveis explicativas ao longo da distribuição condicional do retorno do IBOVESPA.

Os resultados levantados e sintetizados na Tabela 9 apontam que o sentimento textual das notícias financeiras provoca efeitos sobre o retorno do mercado. Analisando especificamente o efeito do sentimento contemporâneo, $Sent_t$, constata-se que existe uma

⁹ O sentimento textual foi normalizado, com média zero e variância 1, para avaliar os coeficientes de regressão em termos de alterações no desvio padrão do sentimento, possibilitando análises da magnitude econômica dos resultados, conforme sugerido na literatura (GARCÍA, 2013; FERGUSON et al., 2015).

reação positiva e persistente do mercado, em relação às notícias divulgadas na mídia, pois os coeficientes apresentaram sinais positivos e foram estatisticamente significantes em todos os quantis de distribuição condicional do retorno estimados.

Analisando a magnitude econômica do impacto do sentimento, $Sent_t$, observa-se uma oscilação do sentimento textual, em termos de seu desvio-padrão, provocando um aumento de 32.14 pontos-base¹⁰ ($q.05$) nos menores retornos do IBOVESPA, contra, 15 pontos-base de aumento sobre os maiores retornos ($q.95$). Realizando um teste de igualdade do efeito do sentimento sobre o desempenho do mercado, por meio dos coeficientes estimados nos quantis $q.05$ e $q.95$, verifica-se que, de fato, os efeitos do sentimento sobre os melhores/piiores desempenho não são iguais, conforme o resultado do teste *wald* (4.0811**), ao nível de significância de 5%, conforme Painel B da Tabela 9. Assim, a análise inicial aponta que as notícias contemporâneas apresentam uma maior importância quando o mercado apresenta um baixo desempenho.

Ao serem consideradas defasagens temporais, verifica-se que existe uma inversão do efeito do sentimento sobre o mercado, pois o impacto passa a ser negativo, e só é encontrada significância estatística do sentimento, com uma ($t-1$), e, com quatro ($t-4$) defasagens temporais, em alguns quantis. Ao observar o sentimento textual defasado em um período, $Sent_{t-1}$, verifica-se que o impacto ocorre somente sobre os quantis que representam os menores retornos, $q.05$ (-15.72 pontos-base) e $q.25$ (-10.66 pontos-base). Analisando uma defasagem de quatro períodos, $Sent_{t-4}$, observa-se que o impacto é estatisticamente significativo, apenas nos quantis com melhores desempenhos, $q.75$ (-9.77 pontos-base) e $p.0.95$ (-23.56 pontos base).

Desse modo, é possível afirmar que o impacto da informação proveniente da mídia tende a persistir sobre as extremidades da distribuição do retorno, ou seja, sobre os piores/melhores desempenhos do mercado. Na estimação dos coeficientes por OLS, observa-se que, em média, o sentimento textual contemporâneo e o sentimento defasado em quatro períodos provocam impactos sobre o retorno.

Diante das evidências estatisticamente significantes sobre o sentimento textual, é possível analisar que o tom das notícias financeiras provoca efeitos distintos sobre o mercado brasileiro. Existem indícios dos investidores reagirem exageradamente às informações disponíveis, mas, rapidamente, parece existir um realinhamento das crenças, que é refletido nos altos/baixos retornos. Nas circunstâncias em que não existe um bom desempenho no mercado

¹⁰ Na escala de pontos-base um ponto base corresponde a 0.01 pontos percentuais.

(baixo retorno), o ajuste dos preços é rápido, pois o efeito do sentimento textual só é estatisticamente significativo até $t-1$, com uma redução de -15.72 pontos-base. Já nos momentos em que o mercado apresenta um bom desempenho (alto retorno), as evidências estatísticas apontam que a reação de ajustes dos preços em relação ao sentimento textual demora um pouco mais para ocorrer, visto que só é encontrada significância estatística em $t-4$, porém, a redução é de -23,56 pontos-base.

As evidências encontradas sugerem que o sentimento das notícias provoca impactos não só quando o mercado apresenta baixos desempenhos, como também, nas situações em que o mercado apresenta bons desempenhos. Esses resultados são consistentes com as evidências levantadas por García (2013), para o mercado norte americano, e, por Ferguson et al. (2015), para o mercado acionário do Reino Unido, porém, no mercado brasileiro, a magnitude econômica do impacto do sentimento das notícias sobre o mercado é bem superior aos impactos levantados por García (2013) e por Ferguson et al. (2015). Isso, provavelmente, ocorre em função de mercados acionários de países emergentes serem mais vulneráveis ao choque de notícias, assim, é de se esperar que, em mercados emergentes a magnitude do impacto das notícias sobre o mercado seja alta.

Analisando as demais variáveis, observa-se que, em média, não existem efeitos dos retornos passados sobre o desempenho do mercado. No entanto, nas estimações realizadas por regressão quantílica, são encontradas evidências de que os retornos passados provocam impactos sobre os retornos contemporâneos, e os valores mais expressivos são observados nos quantis $q.05 Ret_{t-2}$ (11.7 pontos-base) e $q.95 Ret_{t-3}$ (-14.4 pontos-base), ou seja, essas evidências sugerem que as informações passadas não estão totalmente incorporadas nos preços, e isso é refletido, principalmente, nos quantis que expressam os piores e melhores desempenhos do mercado. No entanto, existe uma tendência de informações passadas serem rapidamente incorporadas nos preços, assim como ocorre com as informações relacionadas ao sentimento textual, corroborando com as evidências levantadas por Ferguson et al. (2015). Além disso, foi verificado que nos quantis inferiores, as relações dos retornos passados tendem a serem positivas, enquanto nos quantis superiores apresentam um impacto negativo, esse padrão de comportamento, também foi observado por Baur e Dimpfl (2012) ao analisarem os Estados Unidos.

Em relação à volatilidade, constata-se que o maior impacto da informação sobre os retornos do mercado ocorre quando este apresenta um menor desempenho ($q.05$), de modo que, oscilações na volatilidade podem levar a um aumento de 95 pontos-base ($Volat_{t-4}$) e reduções

de -85.9 pontos-base ($Volat_{t-5}$) do retorno do índice IBOVESPA, sugerindo que a volatilidade passada interfere no comportamento do IBOVESPA.

Já em relação às *dummies* inclusas no modelo para capturar os efeitos dos dias da semana, os resultados evidenciam que as anomalias financeiras de calendário são observadas apenas quando o mercado apresenta um desempenho ruim. No período analisado, os dias de segunda-feira apresentaram quedas de -26.8 pontos-base, nos dias de terça-feira as quedas persistem, chegando a uma redução de -30.47 pontos-base no retorno do IBOVESPA, e nos dias de quarta-feira foram encontrados aumentos positivos significantes do IBOVESPA de 47.9 pontos-base. Esses resultados corroboram as evidências levantadas por French (1980) para o mercado norte-americano, quanto à tendência de retornos mais baixos em dias de segunda-feira e retornos mais altos a partir dos dias de quarta-feira. Os resultados também foram consistentes com as evidências levantadas para o mercado brasileiro, por Bone e Ribeiro (2002), quanto à alta oscilação do retorno encontrada nos dias de terça-feira, chegando a serem superiores às oscilações dos dias de segunda-feira, comportamento explicado como “efeito Brasília”¹¹.

¹¹ Nos dias de terça-feira, iniciam as atividades parlamentares no Congresso Nacional do Brasil, assim, as oscilações relacionadas aos dias de terça-feira estariam associadas a interpretações de questões políticas, para maiores detalhes consultar Bone e Ribeiro (2002).

Tabela 9 - Impacto do sentimento textual sobre o retorno do IBOVESPA

Painel A: Estimação das Regressões												
Variáveis	Quantis											
	$\tau - 0.05$		$\tau - 0.25$		$\tau - 0.50$		$\tau - 0.75$		$\tau - 0.95$		OLS	
	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão
<i>Ret</i> _{t-1}	0.07835***	0.02868	-0.0394	0.0308	-0.0615**	0.02566	-0.0009	0.0335	0.0088	0.0370	-0.01901	0.026483
<i>Ret</i> _{t-2}	0.11792***	0.0601	-0.0176	0.0407	0.0136	0.0302	-0.0698*	0.04056	-0.0598	0.0546	-0.00411	0.02658
<i>Ret</i> _{t-3}	0.0097	0.0333	0.0198	0.0334	-0.0376	0.0261	-0.1023***	0.03476	-0.1448***	0.04081	-0.04267	0.026545
<i>Ret</i> _{t-4}	-0.0112	0.0454	-0.0216	0.0387	-0.0115	0.0270	0.0033	0.0362	0.0393	0.0564	-0.0096	0.026659
<i>Ret</i> _{t-5}	0.0131	0.0250	0.0839***	0.0246	0.0517*	0.02781	-0.0013	0.0338	-0.0152	0.0723	0.030092	0.026642
<i>Sent</i>	0.32141***	0.0641	0.39375***	0.0509	0.3735***	0.04633	0.3363***	0.05592	0.1500**	0.05855	0.298478***	0.041143
<i>Sent</i> _{t-1}	-0.1572***	0.04487	-0.1066**	0.04599	-0.0094	0.0376	-0.0473	0.0495	0.0060	0.0679	-0.04057	0.041815
<i>Sent</i> _{t-2}	0.0605	0.0611	0.0557	0.0397	-0.0033	0.0348	-0.0071	0.0536	-0.0740	0.0516	-0.01551	0.04161
<i>Sent</i> _{t-3}	0.0519	0.0533	-0.0200	0.0508	-0.0301	0.0441	-0.0330	0.0493	0.0302	0.0628	-0.00106	0.041332
<i>Sent</i> _{t-4}	-0.0667	0.0544	-0.0189	0.0479	-0.0340	0.0417	-0.0977*	0.05115	-0.2356***	0.07125	-0.07798*	0.04117
<i>Sent</i> _{t-5}	0.0613	0.0445	0.0190	0.0508	0.0198	0.0408	-0.0014	0.0522	-0.0542	0.0533	0.006988	0.040176
<i>Volat</i> _{t-1}	-1.0287	0.9194	-0.2107	0.4605	-0.0595	0.2346	0.1393	0.4233	0.5164	0.4919	-0.25463	0.176648
<i>Volat</i> _{t-2}	0.6693	0.9403	0.1733	0.5391	0.3418	0.3011	0.3950	0.5059	0.5171	0.6053	0.627513**	0.245257
<i>Volat</i> _{t-3}	-0.1175	0.5295	-0.0778	0.5183	-0.1229	0.2127	-0.4832	0.4559	-0.6258	0.9221	-0.29403	0.246076
<i>Volat</i> _{t-4}	0.95041*	0.5165	0.3107	0.4449	0.0369	0.2305	0.2566	0.4763	0.3884	0.9878	0.170057	0.245531
<i>Volat</i> _{t-5}	-0.8596***	0.2447	-0.3113	0.2387	-0.1815	0.2209	-0.0922	0.3503	-0.4083	0.7604	-0.20868	0.175924
<i>d_jan</i>	0.0447	0.2033	0.0001	0.2083	0.0407	0.1442	-0.1251	0.1094	-0.1975	0.3168	-0.0268	0.140924
<i>d_seg</i>	-0.1620	0.1672	-0.2682*	0.15846	-0.0323	0.1410	0.3064**	0.15369	-0.0377	0.1569	-0.04291	0.124573
<i>d_ter</i>	0.0885	0.1654	-0.3047*	0.17507	-0.0428	0.1197	0.0059	0.1321	-0.2595	0.1743	-0.06948	0.124529
<i>d_quar</i>	0.47948***	0.1499	-0.1717	0.1108	-0.0827	0.1254	0.0061	0.1487	-0.1402	0.1113	0.005493	0.124866
<i>d_sex</i>	-0.1347	0.1400	-0.2983	0.13342	-0.1771	0.1193	-0.0170	0.1273	-0.0346	0.3019	-0.11836	0.124997
<i>Intercepto</i>	-2.3637***	0.14154	-0.7207*	0.10307	0.0216	0.0899	0.8856***	0.10036	2.5356***	0.14891	0.051575	0.088958
<i>Observações</i>	1470		1470		1470		1470		1470		1470	
<i>R²</i>	0.10026		0.04479		0.03223		0.03929		0.08279		0.03219	
<i>Testes de Significância Conjunta</i>												
<i>Teste F</i>	-		-		-		-		-		3.319	
<i>Teste Wald</i>	3.0754***											-

Painel B: Teste de igualdade dos coeficientes

Teste de igualdade		Estatística do teste <i>wald</i>	p-valor	Teste de igualdade		Estatística do teste <i>wald</i>	p-valor
<i>t</i>	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	4.0811	0.0434571**	<i>t-3</i>	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	0.0722	0.7881751
<i>t-1</i>	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	4.1793	0.0410094**	<i>t-4</i>	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	3.7198	0.0538672*
<i>t-2</i>	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	2.9685	0.0850070*	<i>t-5</i>	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	2.8963	0.0888906*

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

Modelo Geral – Eq. 10: $Ret_t = \beta_1 \mathcal{L}_s Sent_t + \gamma_1 \mathcal{L}_s (Ret_t) + \psi_1 \mathcal{L}_s (\sigma_t^2) + \eta X_t + v_t$.

. Ret_{t-n} – retorno do IBOVESPA com n defasagens;

. $Volat_{t-n}$ - volatilidade realizada com n defasagens, descrita no modelos com o símbolo σ_t^2 ;

. $Sent_{t-n}$. Sentimento textual com n defasagens;

. d_{jan} - *Dummy* para meses de janeiro;

. $d_{seg}, d_{ter}, d_{quar}, d_{sex}$ – *dummies* para dias da semana.

* Significante ao nível de 10%.; ** significativo ao nível de 5%; ***significante ao nível de 1%.

Nota: regressão nos quantis (τ): 0,05; 0,25; 0,5; 0,75; 0,95, admitindo que os resíduos não eram identicamente distribuídos, com base na matriz de *Huber Sandwich Standard*. A estatística R^2 para a estimação OLS foi o R^2 ajustado, enquanto, para as estimações em regressão quantílica foi o pseudo R^2 , conforme proposta de Koenker e Machado (1999). Variável sentimento e volatilidade foram normalizadas, com variância 1, para ser possível analisar a magnitude do impacto econômico sobre o retorno. A análise da igualdade dos coeficientes foi obtida por meio do teste *wald*.

A Tabela 10 apresenta os resultados do impacto do sentimento das notícias financeiras, sobre a volatilidade, mediante a utilização de um modelo autoregressivo da volatilidade, no qual foram inclusas uma variável sobre o sentimento textual e *dummies* temporais relacionadas aos dias da semana e aos meses de janeiro. A estimação dos coeficientes foi realizada por OLS e por regressão quantílica autoregressiva, admitindo que os menores quantis (*quantil 0.05* e *quantil 0.25*) representavam as menores volatilidades e os maiores quantis (*quantil 0.75* e *quantil 0.95*) estavam associados às maiores volatilidades, que seriam períodos de turbulência no mercado acionário.

Os resultados, apresentados no Painel A da Tabela 10, mostram que existem impactos dos sentimentos das notícias financeiras sobre o comportamento da volatilidade do retorno do IBOVESPA. Foram encontradas evidências de reações, estatisticamente significativas, em todas as defasagens temporais da variável e em vários quantis de distribuição condicional da volatilidade. No entanto, os impactos mais expressivos são observados quando o sentimento textual está defasado em um período, $Sent_{t-1}$, exercendo um impacto negativo e persistente sobre os períodos de maior turbulência (maiores volatilidades).

Outro aspecto que chama atenção é a magnitude do impacto provocado pelo sentimento, $Sent_{t-1}$, pois, na mediana, a cada oscilação no desvio-padrão do sentimento das notícias existe uma redução de -0,45 pontos-base (q.50), chegando a -1.17 pontos-base no quantil 0.75, e, entre os dias com maior volatilidade, o sentimento das notícias provoca uma redução de -3.95 pontos-base (q.95) na volatilidade do IBOVESPA. O teste de igualdade dos coeficientes, realizado para o q.05 em relação a q.95, teste de *wald* (4.0049**), confirma, ao nível de significância de 5%, que existem diferenças no comportamento do sentimento sobre os momentos de baixa e alta volatilidade do mercado. Isso significa que o tom das notícias financeiras que saíram um dia antes da movimentação analisada pode levar à redução da volatilidade, à medida que o mercado apresenta uma maior turbulência.

Diante dos resultados sobre o sentimento textual, é possível inferir que o tom das informações financeiras divulgadas em jornal, tendem a influenciar os investidores nos períodos de maior volatilidade do mercado. Assim, pode-se dizer que o sentimento das notícias contribui com a atualização das crenças dos investidores em períodos que remetem a uma maior incerteza no mercado, visto que a maior concentração de impactos, estatisticamente significativa, foi observada nos maiores quantis de distribuição da volatilidade.

Na estimação do impacto da volatilidade passada sobre a volatilidade contemporânea, são encontradas evidências de impactos significativos até a terceira *lag*, porém os maiores impactos foram observados na primeira *lag*, com persistência ao longo de toda a distribuição condicional da volatilidade, e com impactos que aumentam ao longo dos quantis estimados. No quantil relacionado à distribuição das menores volatilidades do IBOVESPA (q.05), a volatilidade passada, $Volat_{t-1}$, exerce um impacto de 93.27 pontos-base, enquanto no quantil relacionado às maiores volatilidades, o impacto chega a 103.35 pontos-base. Ao analisar as demais defasagens temporais, estatisticamente significativas, observa-se que existe uma perda considerável do efeito da informação passada sobre a informação contemporânea da volatilidade, a partir da segunda defasagem. Os resultados encontrados demonstram que o comportamento de altas/baixas volatilidades tendem a serem seguidos por altas/baixas volatilidades, sendo consistentes com as explicações de Fama (1965) sobre a existência de agrupamentos de volatilidade nas análises realizadas sobre distribuições de retorno.

Em relação às *dummies* temporais, observa-se que as maiores oscilações da volatilidade (q.95), estatisticamente significantes, ocorrem nos dias de: segunda-feira (9,48 pontos-base); terça-feira (13,80 pontos-base); e sexta-feira (11,36 pontos-base). Vale destacar que a maior oscilação da volatilidade foi encontrada nos dias de terça-feira, o que pode estar associado à sensibilidade do mercado às especulações de notícias políticas, relacionadas ao “efeito Brasília”. Esses resultados corroboram com as evidências levantadas por Ceretta e Vieira (2010), para o mercado brasileiro, de que, em média, os dias de segunda e terça tendem a apresentar volatilidades superiores à volatilidade encontrada nos dias de sexta-feira.

Tabela 10 - Impacto do sentimento textual sobre a volatilidade realizada do IBOVESPA

Painel A: Estimação das Regressões												
Variáveis	$\tau - 0.05$		$\tau - 0.25$		$\tau - 0.50$		$\tau - 0.75$		$\tau - 0.95$		OLS	
	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão
<i>Volat</i> _{t-1}	0.93272***	0.0006	0.92792***	0.0037	0.9340***	0.0142	0.93584***	0.0207	1.10357***	0.1459	0.96846***	0.0263
<i>Volat</i> _{t-2}	0.0015	0.0013	0.022**	0.0093	0.0420	0.0322	0.1378***	0.0364	0.5401*	0.2812	0.10030***	0.0366
<i>Volat</i> _{t-3}	-0.0004	0.0015	-0.0080	0.0108	-0.0194	0.0349	-0.0360	0.0665	-0.536*	0.28628	-0.069*	0.03661
<i>Volat</i> _{t-4}	-0.0008	0.0011	-0.0070	0.0086	-0.0195	0.0226	-0.0623	0.0746	0.0113	0.2196	-0.0601	0.0366
<i>Volat</i> _{t-5}	-0.0005	0.0007	0.0013	0.0059	0.0107	0.0099	0.0101	0.0502	-0.0309	0.1200	0.0321	0.0263
<i>Sent</i>	0.0000	0.0001	-0.0001	0.0009	0.0006	0.0030	-0.0087	0.0058	0.0054	0.0152	-0.0079	0.0062
<i>Sent</i> _{t-1}	0.0000	0.0001	-0.0004	0.0008	-0.0047*	0.00286	-0.0117**	0.00474	-0.0395**	0.01974	0.0004	0.0062
<i>Sent</i> _{t-2}	0.00014*	0.0001	0.0008	0.0008	0.0030	0.0032	0.0070	0.0067	-0.0045	0.0155	0.0059	0.0061
<i>Sent</i> _{t-3}	-0.0001	0.0001	-0.0013	0.0008	-0.0045*	0.00264	-0.0084	0.0059	-0.0268	0.0175	-0.0091	0.0061
<i>Sent</i> _{t-4}	0.0000	0.0001	0.0004	0.0008	0.0000	0.0028	-0.0088	0.0068	-0.028*	0.01631	0.0014	0.0061
<i>Sent</i> _{t-5}	-0.0002**	0.00009	-0.0018**	0.00074	-0.0045*	0.00263	-0.0023	0.0061	-0.0137	0.0126	0.0014	0.0059
<i>d_jan</i>	0.0000	0.0003	-0.0014	0.0030	-0.0062	0.0065	-0.0368**	0.01695	-0.0271	0.0388	-0.0217	0.0212
<i>d_seg</i>	0.0000	0.0002	-0.0013	0.0027	0.0038	0.0079	0.0169	0.0179	0.0948**	0.0386	0.0309*	0.0187
<i>d_ter</i>	-0.0004**	0.00018	0.0016	0.0026	0.0149*	0.0084	0.0210	0.0161	0.1380*	0.0761	0.0409**	0.0141
<i>d_quar</i>	-0.0001	0.0003	-0.0008	0.0025	0.0014	0.0079	0.0052	0.0185	0.0334	0.0386	0.0192	0.0188
<i>d_sex</i>	-0.0003	0.0002	-0.0021	0.0021	-0.0026	0.0068	0.0107	0.0230	0.1136**	0.0516	0.0474**	0.0188
<i>Intercepto</i>	-0.1155***	0.00021	-0.10516***	0.00197	-0.0747***	0.00514	0.02394*	0.01344	0.26527***	0.04172	-0.0261**	0.0133
<i>Observações</i>	1470		1470		1470		1470		1470		1470	
<i>R²</i>	0.88806		0.86943		0.85435		0.83514		0.79347		0.94910	
Testes de Significância Conjunta												
<i>Teste F</i>	-		-		-		-		-		1708	
<i>Teste Wald</i>	2.9113***											-

Painel B: Teste da igualdade dos coeficientes

Testes de Igualdade	Estatística do Teste Wald	p-valor	Teste de Igualdade	Estatística do Teste Wald	P-valor
t	$\beta_1(\tau_{0.5}) = \beta_1(\tau_{9.5})$	0.1286	t-3	$\beta_1(\tau_{0.5}) = \beta_1(\tau_{9.5})$	2.3458
t-1	$\beta_1(\tau_{0.5}) = \beta_1(\tau_{9.5})$	4.0049	t-4	$\beta_1(\tau_{0.5}) = \beta_1(\tau_{9.5})$	2.9475
t-2	$\beta_1(\tau_{0.5}) = \beta_1(\tau_{9.5})$	0.0909	t-5	$\beta_1(\tau_{0.5}) = \beta_1(\tau_{9.5})$	1.1475

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

Modelo – Eq. 11: $\sigma_t^2 = \beta_1 \mathcal{L}_S \text{Sent}_t + \psi_1 \mathcal{L}_S (\sigma_t^2) + \eta X_t + v_t$.

- . $Volat_{t-5}$ – volatilidade realizada com n defasagens, descrita no modelo como σ_t^2 ;
- . $Sent_{t-n}$ sentimento textual com n defasagens;
- . d_{jan} – *dummy* para meses de janeiro;
- . d_{seg} – *dummy* para dias de segunda-feira;
- . d_{ter} – *dummy* para dias de terça-feira;
- . d_{quar} – *dummy* para dias de quarta-feira;
- . d_{sex} – *dummy* para dias de sexta-feira.

* Significante ao nível de 10%.; ** significativo ao nível de 5%; ***significante ao nível de 1%.

Nota: regressão nos quantis (τ): 0,05; 0,25; 0,5; 0,75; 0,95, admitindo que os resíduos não eram identicamente distribuídos, com base na matriz de *Huber Sandwich Standard*. A estatística R^2 para a estimação OLS foi o R^2 ajustado, enquanto, para as estimações em regressão quantílica foi o pseudo R^2 , conforme proposta de Koenker e Machado (1999). Variável sentimento textual foi normalizadas, com variância 1, para ser possível analisar a magnitude do impacto econômico sobre a volatilidade. A análise da igualdade dos coeficientes foi obtida por meio do teste Wald.

As evidências empíricas levantadas sobre o efeito do sentimento das notícias financeiras sobre o retorno e sobre o risco (volatilidade), no período de julho de 2011 a junho de 2017, levam à conclusão de que os movimentos acionários do mercado brasileiro são influenciados pelas notícias financeiras, principalmente nos períodos marcados por um baixo desempenho do mercado (baixos retornos) e por altas turbulências (alta volatilidade). Desse modo, as evidências encontradas não permitem a rejeição das hipóteses de que o sentimento das notícias exerce influência sobre o retorno das ações do mercado acionário brasileiro (H1); e de que o sentimento das notícias exerce influência sobre a volatilidade do retorno das ações do mercado acionário brasileiro (H2).

4.5.2 Influência do Sentimento Pessimista Sobre o Retorno e Sobre a Volatilidade do Mercado

No mercado acionário, os preços seriam um sistema informacional perfeito, se existissem expectativas homogêneas (MILGROM; STOKEY, 1982). No entanto, existem divergências no processo de alinhamento das expectativas dos investidores, provocando falhas na precificação dos ativos (GROSSMAN, STIGLITZ, 1980). Essas divergências são explicadas na literatura como uma restrição provocada por vieses cognitivos existentes no processamento informacional dos indivíduos (KAHNEMAN; TVERSKY, 1979; CAMPBELL et al. 2017), como o viés da negatividade, que está relacionado ao comportamento dos indivíduos em atribuir uma maior atenção e importância a eventos negativos (PEETERS; CZAPINSKI, 1990; ROZIN, ROYMAN, 2001).

Assim, admitindo que os investidores apresentem comportamentos distintos diante de notícias otimistas e pessimistas (ANDERSEN et al. 2003), e que, em geral, costuma ser atribuída uma maior importância às notícias pessimistas, procurou-se investigar o impacto do sentimento de pessimismo das notícias financeiras sobre os movimentos do mercado acionário brasileiro. Diante do contexto apresentado, as hipóteses de pesquisa avaliadas, nesta seção, foram:

H₃: O sentimento pessimista das notícias reduz o retorno no mercado acionário brasileiro.

H4: O sentimento pessimista das notícias tende a elevar a volatilidade do retorno no mercado acionário brasileiro.

Na Tabela 11, são apresentados os resultados quanto ao impacto do sentimento de otimismo e pessimismo das notícias financeiras sobre o retorno do IBOVESPA. Para isso, foi estimada uma regressão autoregressiva do retorno do IBOVESPA, com a inclusão das variáveis sentimento textual e volatilidade, considerando 5 *lags* de defasagens em todas as variáveis. Além disso, o modelo foi estruturado por *dummies*, estas, relacionadas aos dias da semana e aos meses de janeiro. Para capturar o efeito do sentimento de pessimismo, foi incluída no modelo uma *dummy*, D_{pes} , que recebeu o valor “1” para os dias em que a mídia foi pessimista, ou seja, dias em que a variável sentimento textual, em média, apresentou um valor inferior à zero ($Sent < 0$), e recebeu o valor “0” nos casos contrários. Assim, a *dummy* sobre pessimismo D_{pes} , entrou na equação de regressão interagindo com todas as variáveis, com a finalidade de possibilitar que o modelo extraísse o efeito dos dias pessimistas. Além disso, foi realizada a inclusão da expressão “ $1-D_{pes}$ ” para extrair o efeito dos dias de mídia otimista. Esse procedimento de inclusão das *dummies*, “ D_{pes} ” e “ $1-D_{pes}$ ”, interagindo com todas as variáveis, foi realizado com o intuito de evitar que as estimações dos coeficientes relacionados aos períodos de mídia pessimista e otimista fossem realizadas com uma subdivisão da amostra analisada neste estudo.

Desse modo, as estimações dos coeficientes dos modelos foram realizadas por OLS e por regressão quantílica, admitindo que os menores quantis condicionais estimados, (*quantil 0.05* e *quantil 0.25*), correspondiam ao comportamento dos piores desempenhos do mercado (menores retornos), enquanto os maiores quantis correspondiam aos melhores desempenhos do mercado (altos retornos). Assim, a distribuição condicional do retorno do IBOVESPA foi avaliada quanto ao impacto do pessimismo/otimismo sobre os piores/melhores desempenhos do mercado.

O painel A da Tabela 11 mostra que os sentimentos de pessimismo e de otimismo da mídia apresentam impactos distintos sobre o mercado. Analisando especificamente os dias de mídia pessimista, observa-se que alterações no sentimento de pessimismo, $D_{pes} * Sent_t$, provocam um aumento sobre o retorno do IBOVESPA, chegando a 44,73 pontos-base (q.05), quando o mercado apresenta um desempenho ruim (retornos baixos), e um aumento de 42,46 pontos-base (q.95) sobre o retorno do IBOVESPA, nos momentos em que o mercado apresenta

um bom desempenho. No entanto, ao ser testada a igualdade do efeito do sentimento de pessimismo contemporâneo sobre os quantis extremos (q.05) e (q.95), a estatística do teste de *wald* (0.903666) não foi significativa, revelando que o efeito do sentimento de pessimismo das notícias contemporâneas sobre os piores/melhores desempenhos do mercado são iguais, ou seja, estatisticamente não existem diferenças entre no impacto provocado pelas notícias pessimistas contemporâneas sobre as distribuições extremas do retorno.

O sentimento de pessimismo apresenta importância sobre o mercado quando foram admitidas defasagens temporais, de modo que alterações no sentimento de pessimismo da mídia, $D_{pes} * Sent_{t-1}$, levam à reduções do retorno do IBOVESPA em -28.72 pontos-base (q.05), enquanto no quantil (q.25) a redução é -19.38 pontos-base. Ao realizar o teste de igualdade de *wald* (30.2218***) confirma-se que o sentimento de pessimismo exerce um impacto diferente entre os menores e maiores retornos. Observam-se, também, impactos negativos significativos na quarta *lag*, após isso, não foram encontradas evidências significativas.

É importante observar que a maior magnitude econômica do impacto do sentimento de pessimismo sobre retorno do IBOVESPA ocorre sobre o menor quantil de distribuição condicional do retorno. Isso significa que a importância do sentimento de pessimismo da mídia sobre o mercado brasileiro é dada nos períodos de desempenho ruim (baixos retornos), quando o sentimento das notícias do dia anterior pode predizer uma redução -28.72 pontos-base no retorno do IBOVESPA, no primeiro dia útil de movimentação no mercado, seguinte à divulgação das notícias.

Ao ser analisado o sentimento de otimismo contemporâneo e defasado, em 5 *lags*, $1 - D_{pes} * Sent_{t-n}$, percebe-se que o otimismo da mídia provoca aumentos sobre o retorno do IBOVESPA. O sentimento de otimismo contemporâneo impacta o retorno do IBOVESPA em 23.12 pontos-base (q.75). Ao serem admitidas defasagens temporais, o efeito do otimismo persiste até uma *lag*, levando a um aumento de 22 pontos-base (q.05) e 15.74 pontos-base (q.25) sobre o retorno.

Desse modo, pode-se dizer que o sentimento de pessimismo e o sentimento de otimismo das notícias financeiras provocam impactos sobre os movimentos de retorno do mercado.

Os resultados apresentados, na Tabela 11, sobre o sentimento de otimismo e pessimismo da mídia financeira do Brasil, analisados neste estudo, demonstram que os sentimentos das notícias apresentam valor informacional capaz de predizer os movimentos de retorno do

mercado. Esses resultados são contrários aos indícios de que a mídia de países emergentes não seria sofisticada ao ponto de provocar reações significativas no mercado (GRIFFIN, HIRSCHEY; KELLY, 2011; FRAIBERGER, 2016).

As evidências sobre o sentimento das notícias mostram que o tom das matérias financeiras é uma informação que impacta o retorno do mercado brasileiro e, rapidamente, são incorporadas aos preços, visto que, são observados efeitos do pessimismo até a quarta *lag*, enquanto, os efeitos do otimismo são encontrados apenas até a primeira *lag*. Pode-se dizer, que as informações otimistas tendem a se dissipar mais rapidamente no mercado, em comparação com as informações pessimistas, corroborando com a visão Baumeister et al. (2001).

Em relação ao efeito que o sentimento provoca sobre o mercado, os resultados mostram que o sentimento de pessimismo reduz o retorno, corroborando com as evidências encontradas por: Tetlock (2007), Dougal et al. (2012) e García (2013), para o mercado norte americano, Jandl (2016) para o Reino Unido e, Wei, et al. (2017), para o mercado chinês. Já as evidências sobre o sentimento de otimismo, apontam que um aumento do tom de otimismo na mídia eleva o retorno do IBOVESPA, contribuindo com as evidências levantadas por Tetlock (2007) e García (2013) sobre a importância dos tons de otimismo na previsão dos movimentos dos preços no mercado norte-americano.

Com relação à magnitude econômica do impacto do sentimento das notícias no pior momento do mercado (q.05), observa-se que a magnitude do pessimismo, em comparação ao otimismo, é maior em 6.72 pontos-base ($|-28.72| - (22.00) = 6.72$). Isso significa que, no mercado brasileiro, é atribuída uma maior importância aos eventos negativos. Essa evidência está respaldada na literatura sobre o viés de negatividade (PEETERS; CZAPINSKI, 1990; ROZIN, ROYMAN, 2001; BAUMEISTER et al., 2001) e corrobora com as evidências levantadas por Engelberg et al. (2012). Outro aspecto que chama atenção é o valor da magnitude do impacto do sentimento das notícias encontrado no Brasil, em comparação aos impactos levantados para os Estados Unidos e Reino Unido. García (2013) identificou que o pessimismo da mídia levava a uma redução média, de -5.5 pontos-base no retorno do índice DJIA, enquanto, Ferguson et al. (2015) verificaram, no mercado acionário do Reino Unido, que, em média, existiam impactos de 19 pontos-base sobre o retorno, ao serem utilizadas estratégias de investimento baseadas em palavras. A alta magnitude encontrada neste estudo poder estar relacionada ao fato do mercado brasileiro ser mais sensível às notícias, por ser um país

emergente, assim como, pode ser uma peculiaridade da amostra analisada, visto que o período de análise deste estudo contemplou importantes eventos que desencadearam incertezas sobre o mercado acionário.

Em relação às variáveis de controle do modelo, o retorno mostrou significâncias estatísticas até a quinta *lag*. Os impactos mais expressivos são observados nos dias de mídia pessimista. Vale destacar que, na segunda *lag* no quantil q.05, existe um aumento de 11.74 pontos-base sobre o retorno contemporâneo, e, na terceira *lag* do quantil q.95, foi encontrada uma redução de -28.10 pontos-bases. Os resultados evidenciam indícios de que o mercado acionário brasileiro apresenta ineficiências nos preços, visto que informações passadas provocam impactos sobre o mercado.

Já a volatilidade só apresenta significância nos dias de mídia pessimista, sendo o maior impacto identificado sobre o período de melhor desempenho do mercado (q.95), o qual, provocou uma oscilação de 151.74 pontos-base sobre o retorno do mercado na primeira *lag*. Em relação às *dummies* temporais, os maiores impactos significativos foram encontrados nos dias de quarta-feira, tanto para dias de mídia pessimista, quanto otimista. Vale destacar, que nos dias em que a mídia foi otimista, a oscilação do retorno na quarta-feira (93.35 pontos-bases - q.05), foi bem superior, às oscilações provocadas nos períodos de mídia pessimista (43.68 pontos-base - q.05). Essas altas oscilações do retorno, em dias de quarta, corroboram com as evidências sobre as anomalias de calendário, como os resultados levantados por French (1980), para o mercado norte-americano.

Em relação às estimações dos coeficientes por OLS, observa-se que é um modelo sem significância estatística (Teste F 2.617), e as estimações dos coeficientes médios foram realizadas para fins de comparação com coeficientes estimados para os quantis condicionais do retorno. A comparação entre os parâmetros do sentimento textual, pessimista e otimista, obtidos por regressão quantílica e por regressão linear, pode ser observada na Figura 10 do Anexo C.

Tabela 11 - Impacto do sentimento textual sobre o retorno em dias de mídia pessimista e otimista

(Continua)

Painel A: Estimação das Regressões												
Variáveis	Quantis											
	$\tau - 0.05$		$\tau - 0.25$		$\tau - 0.50$		$\tau - 0.75$		$\tau - 0.95$		OLS	
	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão
D _{pes} *Ret _{t-1}	0.07101**	0.0353	-0.0464	0.0349	-0.0592**	0.03238	-0.0026	0.0364	-0.0501	0.0405	-0.03203	0.03124
D _{pes} *Ret _{t-2}	0.1174*	0.0606	0.0295	0.0396	0.0025	0.0377	-0.1086**	0.04634	-0.0486	0.0549	-0.01647	0.032221
D _{pes} *Ret _{t-3}	0.0211	0.0465	0.0137	0.0277	-0.0366	0.0309	-0.1065**	0.04266	-0.2810***	0.05421	-0.0689**	0.032079
D _{pes} *Ret _{t-4}	-0.0564	0.0455	-0.0342	0.0397	-0.0258	0.0339	0.0156	0.0428	-0.0426	0.1036	-0.02681	0.032214
D _{pes} *Ret _{t-5}	0.0191	0.0546	0.06267*	0.0358	0.08254**	0.0369	0.0105	0.0307	-0.0199	0.0618	0.035064	0.0318
D _{pes} *Sent _t	0.44731***	0.1116	0.66773***	0.0972	0.50442***	0.0921	0.37741***	0.1097	0.42463***	0.1556	0.53418***	0.0704
D _{pes} *Sent _{t-1}	-0.2872***	0.04669	-0.1938***	0.03957	-0.0714	0.0474	-0.0528	0.0546	0.0448	0.0397	-0.06316	0.0492
D _{pes} *Sent _{t-2}	0.0423	0.0739	-0.0151	0.0472	-0.0693	0.0603	-0.0345	0.0631	0.0068	0.0873	-0.0609	0.0532
D _{pes} *Sent _{t-3}	0.0766	0.0794	-0.0512	0.0462	-0.0666	0.0508	-0.0243	0.0624	0.0532	0.0817	-0.00973	0.0499
D _{pes} *Sent _{t-4}	-0.0743	0.0559	-0.0013	0.0530	-0.0285	0.0429	-0.1090*	0.03484	-0.2246***	0.08422	-0.07322	0.050
D _{pes} *Sent _{t-5}	0.0468	0.0778	-0.0597	0.0580	-0.0326	0.0578	0.0054	0.0587	-0.0632	0.0765	-0.02771	0.0509
D _{pes} *Volat _{t-1}	-0.9497	0.7449	-0.5297	0.3643	-0.2744	0.3638	-0.0141	0.5995	1.51742*	0.8024	-0.19827	0.2253
D _{pes} *Volat _{t-2}	0.8816	0.8689	0.95504***	0.3487	0.5183	0.3519	0.5116	0.6934	-0.0632	0.8858	0.663323**	0.299
D _{pes} *Volat _{t-3}	-0.3018	0.6764	-0.3393	0.4031	-0.0851	0.3163	-0.4661	0.5197	-1.6905	1.2356	-0.43842	0.2932
D _{pes} *Volat _{t-4}	0.9896	0.7966	0.0612	0.4093	-0.1903	0.3897	-0.0979	0.4297	1.0963	1.2079	0.141573	0.2914
D _{pes} *Volat _{t-5}	-0.9798*	0.56421	-0.2807	0.1871	-0.0022	0.3095	0.1698	0.3477	-0.3825	0.5675	-0.16752	0.2131
D _{pes} *D_JAN	0.1768	0.1921	0.0522	0.1756	-0.0269	0.1612	-0.2632***	0.10105	-0.6766***	0.18464	-0.14315	0.1605
D _{pes} *D_SEG	0.1481	0.3410	0.0722	0.1783	0.0443	0.1695	0.3783**	0.18796	0.60903***	0.2304	0.161846	0.1467
D _{pes} *D_TER	0.2121	0.2184	-0.2264	0.1798	-0.1456	0.1315	-0.0328	0.1434	-0.2722	0.2236	-0.15201	0.1374
D _{pes} *D_QUAR	0.43686*	0.2273	-0.2695**	0.11728	-0.0947	0.1634	-0.0082	0.2058	0.1467	0.2339	-0.10483	0.1398
D _{pes} *D_SEX	-0.2041	0.2233	-0.2903**	0.12727	-0.2225*	0.13144	-0.1409	0.1338	0.1145	0.3915	-0.2012	0.1381
1-D _{pes} *Ret _{t-1}	-0.1097	0.1055	-0.0183	0.0585	-0.0814	0.0525	-0.0505	0.0668	0.0679	0.1015	-0.03128	0.0515
1-D _{pes} *Ret _{t-2}	0.1042	0.1308	-0.0196	0.0751	0.0233	0.0556	-0.0747	0.0781	-0.1217	0.1600	0.007412	0.0473
1-D _{pes} *Ret _{t-3}	-0.0020	0.1197	0.0600	0.0532	0.0064	0.0814	-0.0411	0.0561	0.1087	0.1197	0.011387	0.0481
1-D _{pes} *Ret _{t-4}	-0.0934	0.0939	-0.0723*	0.04304	-0.0248	0.0581	0.0651	0.0740	0.1379	0.1165	0.017939	0.0500
1-D _{pes} *Ret _{t-5}	0.0631	0.0958	0.0624	0.0584	0.0235	0.0468	-0.0450	0.0585	-0.0051	0.1437	0.018416	0.0509
1-D _{pes} *Sent _t	-0.1063	0.3277	0.0225	0.1157	0.1220	0.1540	0.2312**	0.1082	0.1251	0.3486	-0.03714	0.0996
1-D _{pes} *Sent _{t-1}	0.22006*	0.1267	0.15744*	0.0924	0.0902	0.0812	0.0492	0.1030	-0.1766	0.2931	0.017008	0.0804
1-D _{pes} *Sent _{t-2}	0.1206	0.1385	0.1477	0.1169	0.0420	0.0441	-0.0664	0.0788	-0.0162	0.1208	0.042844	0.0702
1-D _{pes} *Sent _{t-3}	0.1080	0.0934	-0.1140	0.1005	-0.0501	0.0724	-0.1394	0.1157	0.0385	0.2389	0.001681	0.0793
1-D _{pes} *Sent _{t-4}	-0.1806	0.1754	-0.1227	0.0876	-0.0245	0.0680	-0.0903	0.1003	-0.2169	0.2087	-0.10986	0.0759
1-D _{pes} *Sent _{t-5}	0.2020	0.1373	0.0146	0.0880	0.0483	0.0864	0.0108	0.1046	-0.1270	0.2230	-0.00165	0.0677
1-D _{pes} *Volat _{t-1}	-2.7836	2.4089	-0.2719	1.0032	0.1933	0.4508	0.1527	0.7857	0.8023	1.3781	-0.17802	0.3026
1-D _{pes} *Volat _{t-2}	1.6167	2.5064	-0.3542	1.1193	0.0582	0.9603	0.6028	1.0448	-0.6659	1.5084	0.405279	0.4477
1-D _{pes} *Volat _{t-3}	2.3117	1.8326	0.8373	0.8349	-0.4386	0.9793	-1.2547	1.0870	0.6255	2.4029	0.165776	0.4640

Tabela 11 - Impacto do sentimento textual sobre o retorno em dias de mídia pessimista e otimista

(Conclusão)

Painel A: Estimação das Regressões												
Variáveis	$\tau - 0.05$		$\tau - 0.25$		$\tau - 0.50$		$\tau - 0.75$		$\tau - 0.95$		OLS	
	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão
1-D _{pes} *Volat _{t-4}	-2.3017	2.1287	0.0271	0.9380	0.6828	0.7559	0.7543	1.2473	0.3010	3.0596	-0.03152	0.4843
1-D _{pes} *Volat _{t-5}	0.7645	1.3413	-0.3722	0.8186	-0.4435	0.6022	0.0473	0.9230	-0.5900	2.8507	-0.27001	0.3332
1-D _{pes} *D_JAN	0.5543	1.2408	-0.2647	0.3762	0.0256	0.3768	0.3211	0.6599	0.8900	2.0515	0.186698	0.2900
1-D _{pes} *D_SEG	0.4382	0.4310	-0.0814	0.1975	-0.0713	0.2354	0.0027	0.2195	-0.1208	0.6256	0.074227	0.2252
1-D _{pes} *D_TER	-0.0243	0.4647	-0.1185	0.2045	0.1935	0.2018	0.2695	0.2502	0.0754	0.4727	0.181635	0.2047
1-D _{pes} *D_QUAR	0.93354***	0.3594	0.32243*	0.1822	0.1384	0.1383	-0.0011	0.1646	-0.2188	0.4826	0.266214	0.1924
1-D _{pes} *D_SEX	0.0514	0.3231	-0.0309	0.2329	-0.0814	0.1828	0.3590	0.2518	-0.2343	0.5096	0.141296	0.2031
Intercepto	-2.3493***	0.21951	-0.6598***	0.09645	0.1068	0.0967	0.90895***	0.1153	2.41187***	0.2391	0.171915*	0.09290
Observações	1470		1470		1470		1470		1470		1470	
R ²	0.14345		0.07102		0.04435		0.04896		0.11780		0.04434	
Testes de significância conjunta												
Teste F	-		-		-		-		-		-	2.617
Teste Wald	3.8906***											-

Painel B: Teste da igualdade dos coeficientes

Teste – Pessimismo			Teste – Pessimismo		
	Estatística do Teste Wald	p-valor		Estatística do Teste Wald	p-valor
t	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	0.0147	t-3	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	0.833601
t-1	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	30.2218	t-4	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	0.129195
t-2	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	0.1013	t-5	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	0.301324
Teste – Otimismo			Teste – Otimismo		
	Estatística do Teste Wald	p-valor		Estatística do Teste Wald	p-valor
t	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.2439	t-3	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.783732
t-1	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	1.5729	t-4	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.891466
t-2	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.5756	t-5	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.199989

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

Modelo geral – Eq. 12: $Ret_t = \beta_0 + (D_{pes})(\beta_1 L_s Sent_t + \gamma_1 L_s (Ret_t) + \psi_1 L_s (\sigma_t^2) + \eta X_t) + (1 - D_{pes})(\beta_2 L_s Sent_t + \gamma_2 L_s (Ret_t) + \psi_2 L_s (\sigma_t^2) + \eta X_t) + v_t$

. D_{pes} – dummy relacionada aos dias de mídia pessimista, que recebe o valor “1” quando o *Sent* < 0, e “0” quando *Sent* ≥ 0.

. D_{pes}*Ret_{t-n} – Retorno em dias de mídia pessimista no n lags;

. D_{pes}*Sent_{t-1} – sentimento da mídia em dias pessimistas com n lags;

. D_{pes}*Volat_{t-1} – volatilidade do mercado com n lags, em dias de mídia pessimista, expressa no modelo como σ_t^2 ;

. 1-D_{pes}*Ret_{t-n} – retorno em dias de mídia otimista;

. 1-D_{pes}*Sent_{t-n} – sentimento das notícias quando a mídia está otimista;

- . $1-D_{pes} * Volat_{t-n}$ – volatilidade em dias de média otimista, expressa no modelo como σ_t^2 ;
 - . $D_{pes} * d_{jan}$ e $1-D_{pes} * d_{jan}$ - *dummy* para meses de janeiro, com média pessimista e otimista, respectivamente;
 - . $D_{pes} * d_{seg}$ e $1-D_{pes} * d_{seg}$ - *dummy* para dias de segunda-feira, com média pessimista e otimista, respectivamente;
 - . $D_{pes} * d_{ter}$ e $1-D_{pes} * d_{ter}$ - *dummy* para dias de terça-feira, com média pessimista e otimista, respectivamente;
 - . $D_{pes} * d_{quar}$ e $1-D_{pes} * d_{quar}$ - *dummy* para dias de quarta-feira, com média pessimista e otimista, respectivamente;
 - . $D_{pes} * d_{sex}$ e $1-D_{pes} * d_{sex}$ - *dummy* para dias de sexta-feira, com média pessimista e otimista, respectivamente;
- * Significante ao nível de 10%.; ** significativo ao nível de 5%; ***significativo ao nível de 1%.

Nota: regressão nos quantis (τ): 0,05; 0,25; 0,5; 0,75; 0,95, admitindo que os resíduos não eram identicamente distribuídos, com base na matriz de *Huber Sandwich Standard*. A estatística R^2 para a estimação OLS foi o R^2 ajustado, enquanto, para as estimações em regressão quantílica foi o pseudo R^2 , conforme proposta de Koenker e Machado (1999). Variável sentimento e volatilidade foram normalizadas, com média zero e variância 1, para ser possível analisar a magnitude do impacto econômico sobre o retorno. A análise da igualdade dos coeficientes foi obtida por meio do teste wald.

Na Tabela 12 são apresentados os resultados de um modelo autoregressivo da volatilidade com a inclusão do sentimento textual e de *dummies* temporais, que diferem do modelo anteriormente apresentado, na discussão sobre volatilidade (equação 11), apenas quanto ao aprofundamento da discussão do sentimento textual sob a ótica da mídia em dias pessimistas e otimistas. Desse modo, para ser possível analisar o impacto do sentimento de otimismo e pessimismo, foi incorporado ao modelo, uma *dummy*, D_{pes} , em que, foi atribuído “1” para os dias em que a mídia foi pessimista ($Sent < 0$), e “0” para nos casos contrários, procedimento semelhante ao realizado na avaliação do retorno (Equação 12). Assim, foram investigados os impactos dos dias otimistas/pessimistas sobre os períodos de baixas e altas volatilidades (períodos de turbulência) do mercado acionário brasileiro.

Os resultados do reflexo do sentimento das notícias sobre a volatilidade do retorno do IBOVESPA mostraram que o tom pessimista das notícias contemporâneas aumenta a volatilidade em 0.014 pontos-base, quando o mercado está calmo (pequena volatilidade) (q.05). No entanto, quando são admitidas defasagens temporais, o sentimento de pessimismo tende a reduzir a volatilidade, pois foram encontrados impactos negativos e estatisticamente significantes de -0.94 pontos-base no quantil q.5, -1.42 pontos-base no quantil q.75 e -4.86 pontos-base no quantil q.95, sugerindo que o sentimento de pessimismo reduz a volatilidade, à medida que aumenta a turbulência do mercado. Por outro lado, o sentimento de otimismo, só provoca efeitos sobre a volatilidade a partir da segunda *lag*, com aumento de 0.22 pontos-base no quantil q.025, e um aumento de 0.81 pontos-base, no quantil q.50. Ao serem analisados outros níveis de defasagem, observa-se que existe uma reversão do impacto, como na terceira *lag*, quando o sentimento de otimismo passa a reduzir a volatilidade em -1.29 (q.50).

Diante dos resultados apresentados sobre o impacto do tom de pessimismo e otimismo sobre a volatilidade, é possível dizer que as informações negativas são incorporadas rapidamente à volatilidade, em comparação às informações otimistas. Ademais, o sentimento de pessimismo da mídia tende a reduzir a volatilidade do retorno do IBOVESPA, contrariando as evidências empíricas de Kothari et al. (2009) de que os tons de pessimismo da mídia eram acompanhados por aumentos da volatilidade do retorno no mercado norte americano. Esse comportamento do pessimismo da mídia estar associado a uma redução da volatilidade sugere indícios de que o tom das notícias contribui com a redução das expectativas dos investidores sobre o risco do mercado acionário, levando à redução da volatilidade (observada pelo sinal negativo), tal comportamento pode ser uma característica do mercado brasileiro, visto que, uma relação semelhante foi encontrada por Marins e Vicente (2017) ao analisar a volatilidade das

taxas de juros diante das notícias oriundas do Comitê de Políticas Monetárias do Brasil (Copom).

Em relação à volatilidade com defasagens temporais, observa-se um comportamento semelhante ao anteriormente avaliado, de modo que, a volatilidade passada exerce influência sobre a volatilidade contemporânea até a quarta *lag*, porém os maiores impactos são observados na primeira *lag*, apresentando uma persistência ao longo de toda a distribuição dos quantis condicionais da volatilidade, com os impactos apresentando tendência crescente ao longo dos quantis. Esse comportamento é observado tanto nos dias de mídia pessimista, quanto nos dias de mídia otimista. Desse modo, mais uma vez, foram observadas evidências, neste estudo, que são consistentes com as explicações de Fama (1965), sobre a existência de agrupamentos de volatilidade, visto que, altas/baixas volatilidades tendem a serem seguidas por altas/baixas volatilidades.

As evidências estatisticamente significantes sobre as *dummies* temporais mostram que os impactos mais expressivos foram encontrados nos dias de mídia pessimista e nos períodos em que o mercado apresentava uma alta turbulência, levando à uma oscilação da volatilidade de 14.04% nos dias de terça-feira, e nos dias de sexta-feira a oscilação chegou a 11.84%. Uma possível explicação para a maior oscilação encontrada nos dias de terça-feira seria a influência do “efeito Brasília” sobre o mercado. Além disso, esses resultados estão de acordo com as evidências levantadas por Ceretta e Vieira (2010), quanto a constatação de altas oscilações nos dias de terça-feira.

Tabela 12 - Impacto do sentimento sobre a volatilidade em dias de média pessimista e otimista

(Continua)

Painel A: Estimação das Regressões												
Variáveis	Quantis											
	$\tau - 0.05$		$\tau - 0.25$		$\tau - 0.50$		$\tau - 0.75$		$\tau - 0.95$		OLS	
	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão
D _{pes} *Volat _{t-1}	0.93184***	0.0005	0.92823***	0.0049	0.94795***	0.0246	0.93373***	0.0538	0.97396***	0.1692	0.932743***	0.0339
D _{pes} *Volat _{t-2}	0.00269**	0.0012	0.02494**	0.0100	-0.0003	0.0392	0.13275**	0.0597	0.82334*	0.4576	0.162563***	0.0447
D _{pes} *Volat _{t-3}	-0.0008	0.0014	-0.0206**	0.00992	-0.0007	0.0387	-0.1092	0.0907	-0.7357*	0.43129	-0.1184***	0.0439
D _{pes} *Volat _{t-4}	-0.0007	0.0011	0.0046	0.0086	-0.0104	0.0242	-0.0013	0.0980	0.1103	0.1995	-0.0181	0.0440
D _{pes} *Volat _{t-5}	-0.0006	0.0006	-0.0012	0.0075	0.0109	0.0111	0.0361	0.0638	-0.0691	0.0913	0.0169	0.0322
D _{pes} *Sent _t	0.00014*	0.0001	0.0010	0.0016	0.0040	0.0051	-0.0028	0.0106	0.0019	0.0310	-0.0158	0.0106
D _{pes} *Sent _{t-1}	-0.0001	0.0001	-0.0009	0.0009	-0.0094***	0.00357	-0.0142**	0.00597	-0.0486**	0.02106	0.0073	0.0073
D _{pes} *Sent _{t-2}	-0.0002*	0.00008	-0.0005	0.0008	0.0025	0.0040	0.0072	0.0076	-0.0110	0.0177	-0.0019	0.0078
D _{pes} *Sent _{t-3}	-0.0001	0.0001	-0.0007	0.0008	-0.0027	0.0027	-0.0070	0.0074	-0.0274	0.0217	-0.0074	0.0073
D _{pes} *Sent _{t-4}	0.0000	0.0000	0.0001	0.0009	-0.0012	0.0029	-0.0127	0.0082	-0.0280	0.0195	0.0053	0.0074
D _{pes} *Sent _{t-5}	-0.0002	0.0001	-0.0018***	0.00058	-0.0045	0.0037	-0.0018	0.0072	-0.0020	0.0188	0.0071	0.0075
D _{pes} *D_JAN	0.0003	0.0003	0.0018	0.0036	-0.0042	0.0062	-0.0426*	0.0248	-0.0319	0.0434	-0.0179	0.0243
D _{pes} *D_SEG	0.0000	0.0003	-0.0006	0.0028	0.0022	0.0099	0.0273	0.0229	0.0874	0.0599	0.0218	0.0222
D _{pes} *D_TER	-0.0004	0.0002	0.0011	0.0023	0.0132	0.0092	0.0167	0.0183	0.14047**	0.0676	0.039631*	0.0208
D _{pes} *D_QUAR	-0.0001	0.0003	-0.0017	0.0024	-0.0034	0.0105	-0.0058	0.0227	0.0415	0.0534	0.0197	0.0211
D _{pes} *D_SEX	-0.0003	0.0002	-0.0011	0.0023	-0.0033	0.0069	0.0062	0.0272	0.11848*	0.0678	0.049297**	0.0209
1-D _{pes} *Volat _{t-1}	0.93268***	0.0027	0.92585***	0.0069	0.91557***	0.0208	0.98744***	0.1144	1.14155**	0.4462	1.006453***	0.0445
1-D _{pes} *Volat _{t-2}	0.0015	0.0025	0.0153	0.0104	0.08637**	0.0424	0.0931	0.1761	-0.0674	0.6302	0.0030	0.0670
1-D _{pes} *Volat _{t-3}	-0.0011	0.0028	0.0159	0.0118	0.0249	0.0496	0.0033	0.1690	-0.0174	0.8753	0.0453	0.0691
1-D _{pes} *Volat _{t-4}	0.0004	0.0038	-0.0280***	0.01017	-0.1051**	0.0508	-0.1898	0.1650	-0.0499	1.0463	-0.1520**	0.069697
1-D _{pes} *Volat _{t-5}	-0.0009	0.0021	0.0057	0.0104	0.0255	0.0246	0.0806	0.0957	0.0737	0.7129	0.0624	0.0483
1-D _{pes} *Sent _t	-0.0002	0.0004	-0.0006	0.0017	0.0026	0.0041	-0.0124	0.0189	-0.0287	0.0870	-0.0040	0.0150
1-D _{pes} *Sent _{t-1}	0.0000	0.0004	0.0015	0.0014	-0.0012	0.0038	0.0032	0.0138	-0.0496	0.0727	-0.0097	0.0118
1-D _{pes} *Sent _{t-2}	0.0003	0.0002	0.00227**	0.0011	0.00812***	0.0037	0.0132	0.0105	-0.0156	0.0334	0.0141	0.0104
1-D _{pes} *Sent _{t-3}	0.0000	0.0003	-0.0016	0.0015	-0.0129**	0.00506	-0.0025	0.0134	-0.0311	0.0629	-0.0039	0.0118
1-D _{pes} *Sent _{t-4}	0.0000	0.0002	0.0002	0.0009	0.0034	0.0038	-0.0127	0.0137	-0.0437	0.0626	-0.0080	0.0114
1-D _{pes} *Sent _{t-5}	-0.0003	0.0004	-0.0026***	0.00083	-0.0030	0.0041	-0.0012	0.0107	-0.0373	0.0896	-0.0088	0.0102
1-D _{pes} *D_JAN	0.0002	0.0011	-0.0041*	0.00237	-0.0222*	0.01209	-0.0428	0.0406	-0.0290	0.6800	-0.0294	0.0438
1-D _{pes} *D_SEG	0.0005	0.0006	0.0007	0.0056	-0.0005	0.0127	0.0114	0.0513	0.1113	0.2181	0.0340	0.0340
1-D _{pes} *D_TER	-0.0004	0.0007	0.0009	0.0038	0.0098	0.0135	0.0146	0.0439	0.0226	0.1563	0.0286	0.0308

Tabela 12 - Impacto do sentimento sobre a volatilidade em dias de média pessimista e otimista

(Conclusão)

Painel A: Estimação das Regressões												
Variáveis	Quantis										OLS	
	$\tau - 0.05$		$\tau - 0.25$		$\tau - 0.50$		$\tau - 0.75$		$\tau - 0.95$		coef.	Erro padrão
1-D _{pes} *D_QUAR	0.0006	0.0006	-0.0011	0.0029	0.0054	0.0107	-0.0004	0.0257	0.0052	0.1222	0.0181	0.0291
1-D _{pes} *D_SEX	0.0003	0.0007	-0.0044	0.0027	-0.0047	0.0086	-0.0083	0.0293	0.1423	0.1181	0.0448	0.0305
Intercepto	-0.1155***	0.00024	-0.1049***	0.00174	-0.0718***	0.00508	0.02785*	0.0152	0.2745***	0.0595	-0.0279***	0.014025
Observações	1470		1470		1470		1470		1470		1470	
R ²	0.88813		0.86983		0.85536		0.83581		0.79864		0.94910	
Testes de significância conjunta												
Teste F	-		-		-		-		-		854.7	
Teste Wald	7.9081***										-	
Painel B: Teste da igualdade dos coeficientes												
Teste - Pessimismo		Estatística do Teste Wald	p-valor	Teste - Pessimismo		Estatística do Teste Wald	p-valor					
t	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	0.0034	0.95338	t-3	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	1.5803	0.20881					
t-1	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	5.3201	0.02115**	t-4	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	2.0488	0.15243					
t-2	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	0.3801	0.53758	t-5	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	0.0097	0.92169					
Teste - Otimismo		Estatística do Teste Wald	p-valor	Teste - Otimismo		Estatística do Teste Wald	p-valor					
T	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.1074	0.74319	t-3	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.2449	0.62070					
t-1	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.4645	0.49557	t-4	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.4881	0.48484					
t-2	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.2264	0.63423	t-5	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.1701	0.68001					

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

Modelo Geral – Eq. 13: $\sigma_t^2 = \beta_0 + (D_{pes})(\beta_1 \mathcal{L}_s \text{Sent}_t + \psi_1 \mathcal{L}_s(\sigma_t^2) + \eta X_t) + (1 - D_{pes})(\beta_2 \mathcal{L}_s \text{Sent}_t + \psi_2 \mathcal{L}_s(\sigma_t^2) + \eta X_t) + v_t$

. D_{pes} – dummy relacionada aos dias de média pessimista, que recebe o valor “1” quando o *Sent* < 0, e “0” quando *Sent* ≥ 0.

. D_{pes}*Sent_{t-1} – sentimento da média em dias pessimistas com n lags;

. D_{pes}*Volat_{t-1} – volatilidade do mercado com n lags, em dias de média pessimista, expressa no modelo como σ_t^2 ;

. 1-D_{pes}*Sent_{t-n} – sentimento das notícias quando a média está otimista;

. 1-D_{pes}*Volat_{t-n} – volatilidade em dias de média otimista, expressa no modelo como σ_t^2 ;

. D_{pes}*d_{jan} e 1-D_{pes}*d_{jan} – dummy para meses de janeiro, com média pessimista e otimista, respectivamente;

. D_{pes}*d_{seg} e 1-D_{pes}*d_{seg} – dummy para dias de segunda-feira, com média pessimista e otimista, respectivamente;

. D_{pes}*d_{ter} e 1-D_{pes}*d_{ter} – dummy para dias de terça-feira, com média pessimista e otimista, respectivamente;

. D_{pes}*d_{quar} e 1-D_{pes}*d_{quar} – dummy para dias de quarta-feira, com média pessimista e otimista, respectivamente;

. D_{pes}*d_{sex} e 1-D_{pes}*d_{sex} – dummy para dias de sexta-feira, com média pessimista e otimista, respectivamente;

* Significante ao nível de 10%.; ** significante ao nível de 5%; ***significante ao nível de 1%.

Nota: regressão nos quantis (τ): 0,05; 0,25; 0,5; 0,75; 0,95, admitindo que os resíduos não eram identicamente distribuídos, com base na matriz de *Huber Sandwich Standard*. A estatística R^2 para a estimação OLS foi o R^2 ajustado, enquanto para as estimações em regressão quantílica foi o pseudo R^2 , conforme proposta de Koenker e Machado (1999). Variável sentimento foi normalizada, com média zero e variância 1, para ser possível analisar a magnitude do impacto econômico sobre o retorno. A análise da igualdade dos coeficientes foi obtida por meio do teste *wald*.

Neste tópico, procura-se ampliar a discussão do sentimento textual, segregando o impacto dos tons de pessimismo e otimismo das notícias financeiras. As evidências empíricas encontradas quanto ao reflexo do sentimento de pessimismo sobre os movimentos de risco (volatilidade) e retorno do mercado não permitem a rejeição da hipótese de que o sentimento pessimista das notícias reduz o retorno no mercado acionário brasileiro (H3), porém, os resultados levantados sobre a volatilidade, levam à rejeição, da hipótese de que o sentimento pessimista das notícias tende a elevar a volatilidade do retorno no mercado acionário brasileiro (H4), pois as evidências apontaram que o sentimento de pessimismo das notícias financeiras, no Brasil, reduz a volatilidade do mercado, principalmente nos períodos de maior turbulência do mercado.

4.5.3 Influência do Sentimento de Pessimismo Sobre o Retorno e Sobre a Volatilidade do Mercado em Períodos de Incerteza Econômica

A dinâmica de qualquer mercado está suscetível às incertezas econômicas, porém, os países emergentes são mais vulneráveis aos eventos que gerem incertezas (KOREN; TENREYRO, 2007; BLOOM, 2014), conseqüentemente, nos mercados acionários dos países emergentes costumam existir mais flutuações (KOREN, TENREYRO, 2007, BLOOM, 2014, JURADO et al., 2015).

A mídia desempenha um importante papel para os mercados acionários, principalmente em períodos marcados por recessões econômicas (GARCÍA, 2013), visto que, atua na disseminação de informações em massa, conforme é ressaltado por Griffin et al. (2011). Assim, admitindo que as notícias apresentam valor informacional para o mercado acionário, principalmente em períodos de recessões e crises (GARCÍA, 2013 e LIU 2014), considerando que as notícias contribuem com a atualização das expectativas dos agentes econômicos sobre os preços dos ativos e que os indivíduos costumam atribuir uma maior importância aos eventos negativos (ROZIN; ROYMAN, 2001), nesta seção, as hipóteses de pesquisa avaliadas foram:

H₅: O impacto do sentimento de pessimismo das notícias sobre o retorno no mercado acionário brasileiro é maior em períodos de incerteza econômica.

H₆: O impacto do sentimento de pessimismo das notícias sobre a volatilidade no mercado acionário brasileiro é maior em períodos de incerteza econômica.

A Tabela 13 sintetiza os resultados das estimações sobre o impacto do sentimento de pessimismo na mídia, sobre o retorno do IBOVESPA, em períodos marcados por uma maior/menor incerteza econômica. A classificação da incerteza foi realizada com base no índice IIE-Br da FGV, conforme o comportamento da sua distribuição por quartis. Assim, foram considerados como períodos de alta incerteza, as datas em que o índice IIE-Br apresentou um valor superior aos valores encontrado no terceiro quartil (75° percentil) da amostra, que corresponde aos 25% dos valores mais altos de um conjunto de dados. Na amostra analisada foram considerados como momentos de alta incerteza, os períodos em que o índice IIE-Br foi igual ou superior a 122.7 (valor do 75° percentil da amostra analisada). Para analisar os períodos de alta/baixa incerteza, foi incorporada ao modelo, uma *dummy*, D_i , que recebeu o valor “1” para os períodos, em que, o IIE-Br foi igual ou superior a 122.7 (valor do 75° percentil), e, recebeu “0” para os períodos em que o IIE-Br foi inferior ao valor do 75° percentil da amostra (122.7).

Assim, a regressão de estimação do retorno do IBOVESPA, apresenta como principal variável explicativa o sentimento de pessimismo da mídia e como variáveis de controle, o retorno e a volatilidade do mercado, admitindo 5 *lags* de defasagens em todas as variáveis. Para averiguar o efeito dos períodos de maior incerteza sobre os movimentos do mercado, a *dummy*, D_i , foi incorporada ao modelo, interagindo com todas as variáveis e, para verificar os efeitos dos dias de menor incerteza, foi incluída a expressão “ $1-D_i$ ”, também, interagindo com todas as variáveis do modelo. Esse procedimento é semelhante ao utilizado anteriormente, na segregação de dias de mídia otimista e pessimista, com a finalidade de evitar a separação da amostra para realizar a estimação das regressões dos períodos de alta e baixa incerteza. Assim, foi analisando o impacto do sentimento de pessimismo da mídia em períodos de alta e baixa incerteza sobre os piores e melhores desempenhos do mercado acionário brasileiro.

As evidências sobre o pessimismo das notícias financeiras, apresentadas no Painel A da Tabela 13 mostram que existem efeitos significativos do sentimento da mídia em períodos de maior e menor incerteza. Dentre as evidências encontradas nos períodos de alta incerteza, observa-se que o sentimento das notícias apresenta impactos persistentes nos momentos em que o mercado apresenta um baixo desempenho (baixos retornos). No entanto, o efeito do

sentimento sobre o mercado oscila, pois os sinais dos coeficientes apresentam mudanças entre as defasagens temporais. Com uma *lag*, foi observado que mudanças no desvio-padrão do sentimento de pessimismo, $Di*Sent_{t-1}$, provocaram um aumento de 75.47 pontos-base (q.05) no retorno, já na segunda *lag* o pessimismo da mídia provocou uma redução do retorno em -50.66 pontos-base (q.05), na terceira *lag* a redução passou a ser de -76.21 pontos-base (q.05), sendo encontrados efeitos até a quinta *lag*. Essas evidências mostram indícios da presença de ruídos no processo de ajuste dos preços às informações disponíveis.

Ao analisar os períodos de melhor desempenho do mercado (altos retornos), verificou-se, que alterações no pessimismo da mídia levaram a redução de -56.19 (q.95) pontos-base com uma *lag* de defasagem, e o ápice do impacto, ocorre na quarta *lag*, com uma redução de -81.94 pontos-base (q.95). Assim, pode-se dizer que, em períodos de alta incerteza, o sentimento de pessimismo da mídia atua no mercado, reduzindo os retornos nas circunstâncias de melhor ou pior desempenho do mercado. Os testes de igualdade dos coeficientes para os períodos de maior incerteza, que são apresentados no painel B da Tabela 13, confirmam que existem diferenças entre os impactos que o sentimento de pessimismo provoca sobre os melhores e piores desempenhos do mercado, com única exceção ocorrendo na segunda *lag*, em que não foi observada significância estatística, sugerindo que o pessimismo não apresentou impactos distintos entre as extremidades do desempenho do mercado.

Nos períodos de menor incerteza, os impactos mais significativos são encontrados na primeira *lag*, com uma redução de -34.44 pontos-base (q.05), e, na terceira *lag*, com um aumento de 23.67 pontos-base (q.05). Isso significa que na presença de uma baixa incerteza, o impacto do sentimento de pessimismo é mais evidente sobre os períodos de baixo desempenho do mercado.

Ao comparar as evidências levantadas sobre os períodos de incerteza, observa-se que o pessimismo da mídia, seja alta ou baixa incerteza, sempre provoca impactos sobre os piores desempenhos do mercado (retornos mais baixos). Cabe destacar que os impactos sobre o retorno, em termos de magnitude, são maiores na alta incerteza. Isso significa que a mídia, em termos informacionais, apresenta uma maior importância para o mercado nos períodos de alta incerteza, sugerindo que o mercado acionário, quando a economia está sob pressão, é mais sensível às notícias negativas (pessimismo da mídia), corroborando com as evidências de Caporale, Spagnolo e Spagnolo (2016), ao analisar o comportamento do mercado acionário de

alguns países europeus, antes e após a crise de 2008. Essas evidências são consistentes, também, com as análises de García (2013) sobre a influência do pessimismo da mídia nos ciclos econômicos nos Estados Unidos, bem como os resultados do estudo de Liu (2014) sobre a contribuição do pessimismo das notícias no período da crise Européia (2009-2012).

Além disso, observa-se que, na presença de alta incerteza, os efeitos do pessimismo da mídia demoraram a se dissipar, pois, até a quinta *lag*, foram encontradas evidências estatisticamente significativas de o pessimismo impactar o retorno do mercado. Já nos períodos de baixa incerteza, existe uma incorporação mais rápida da informação nos preços, fazendo com que, os efeitos desapareçam rapidamente. Essa evidência encontra respaldo nas explicações sobre o viés de negatividade, corroborando com a visão de que os investidores tendem a atribuir uma maior importância e atenção aos eventos negativos, com isso, informações negativas tendem a não se dissipar rapidamente no mercado (BAUMEISTER et al., 2001).

Em relação ao comportamento das variáveis de controle, constata-se que o retorno defasado apresentou um comportamento semelhante entre os períodos de alta e baixa incerteza, de modo que, foram encontradas significâncias estatísticas até a quinta *lag*. Vale destacar que os maiores impactos foram encontrados quando o mercado apresentava um baixo desempenho, chegando a provocar uma redução de -20,36 pontos bases (q.05) em períodos de alta incerteza. Assim, pode-se dizer, que em períodos de baixo desempenho são encontrados mais indícios de ineficiências no mercado acionário brasileiro. Já a volatilidade apresenta efeitos distintos entre os períodos de incerteza avaliados. Nos momentos de alta incerteza, os impactos mais expressivos da volatilidade sobre o retorno ocorrem sobre os quantis condicionais relacionados a um baixo desempenho do mercado, chegando a provocar, na segunda *lag*, um impacto de 85.85 pontos-base (q.25), enquanto nos períodos de baixa incerteza, o impacto ocorre sobre os maiores quantis de distribuição do retorno, provocando um impacto de 148 pontos-base (q.95), com duas defasagens temporais.

Tabela 13 - Impacto do sentimento de pessimismo sobre o retorno do IBOVESPA em dias de alta e baixa incerteza

(Continua)

Painel A: Estimação das Regressões												
Variáveis	Quantis											
	$\tau-0.05$		$\tau-0.25$		$\tau-0.50$		$\tau-0.75$		$\tau-0.95$		OLS	
	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão
$Di*Ret_{t-1}$	0.0832	0.0610	-0.0126	0.0506	-0.1081**	0.04655	-0.1411*	0.08475	-0.0699	0.1400	-0.0554	0.0547
$Di*Ret_{t-2}$	-0.2036*	0.05583	-0.0945*	0.05462	-0.0084	0.0460	-0.0412	0.0966	0.2539	0.1856	-0.0376	0.0566
$Di*Ret_{t-3}$	-0.0114	0.0805	-0.0837	0.0532	-0.0508	0.0561	0.0249	0.0793	-0.2076	0.1427	-0.0420	0.0569
$Di*Ret_{t-4}$	-0.1357***	0.04741	-0.1024*	0.06063	0.0315	0.0441	-0.0223	0.0918	-0.0375	0.1756	-0.0509	0.0569
$Di*Ret_{t-5}$	0.0718	0.0847	0.08133**	0.0406	0.0413	0.0366	0.0207	0.0687	-0.0585	0.1277	0.0164	0.0569
$Di*Sent_P_{t-1}$	0.75474***	0.1848	0.2019	0.1857	0.0448	0.1293	0.0430	0.2969	-0.5619**	0.26658	0.0873	0.1396
$Di*Sent_P_{t-2}$	-0.5066***	0.08573	-0.1459	0.1443	-0.0076	0.1234	-0.0203	0.2386	-0.1298	0.3449	-0.2155	0.1411
$Di*Sent_P_{t-3}$	-0.7621***	0.14383	-0.4091	0.13046	-0.206*	0.11507	-0.1815	0.1228	0.3503	0.2990	-0.2379	0.1478
$Di*Sent_P_{t-4}$	0.69649***	0.1507	0.36881**	0.1446	0.1030	0.1679	-0.3405*	0.20212	-0.8194***	0.25114	0.0717	0.1498
$Di*Sent_P_{t-5}$	-0.2016*	0.11878	0.0098	0.1509	-0.0058	0.0894	0.0875	0.1481	0.5235	0.3672	0.0542	0.1442
$Di*Volat_{t-1}$	-0.5634	0.7469	-0.4695**	0.22849	-0.3828	0.2944	-0.0908	0.7183	2.6840	2.7791	-0.3920*	0.2195
$Di*Volat_{t-2}$	1.0307	0.8404	0.85859**	0.4355	0.6336*	0.3244	0.3245	0.8088	-2.1209	3.2062	0.6315**	0.3015
$Di*Volat_{t-3}$	-0.0341	0.5396	-0.0162	0.4525	0.1918	0.2149	0.0832	0.8634	0.9629	1.6042	-0.0003	0.3438
$Di*Volat_{t-4}$	0.1038	0.6615	-0.0284	0.3087	-0.3445*	0.20814	0.1336	0.6824	-1.2255	2.4321	0.0205	0.3768
$Di*Volat_{t-5}$	-0.6485	0.5645	-0.3322*	0.17072	-0.0187	0.1592	-0.3252	0.2371	0.5662	1.7870	-0.1220	0.2708
$1-Di*Ret_{t-1}$	0.14465***	0.0466	0.0047	0.0394	-0.0248	0.0364	-0.0081	0.0417	0.0322	0.0499	-0.0076	0.0397
$1-Di*Ret_{t-2}$	0.09532**	0.0445	0.0777*	0.04378	-0.0157	0.0387	-0.0100	0.0502	-0.0989*	0.05093	-0.0064	0.0398
$1-Di*Ret_{t-3}$	-0.0499	0.0663	0.0223	0.0503	0.0257	0.0331	-0.0062	0.0362	0.0352	0.0676	-0.0007	0.0398
$1-Di*Ret_{t-4}$	-0.1058*	0.05611	0.0613	0.0382	-0.0081	0.0425	0.0389	0.0522	-0.1084	0.0687	0.0079	0.0396
$1-Di*Ret_{t-5}$	0.09056*	0.0496	0.0335	0.0412	0.0063	0.0386	-0.0119	0.0457	-0.1310**	0.05549	0.0154	0.0391
$1-Di*Sent_P_{t-1}$	-0.3444**	0.14599	-0.1363*	0.07196	-0.0694	0.0712	-0.1571**	0.07806	0.0337	0.1219	-0.0993	0.0803
$1-Di*Sent_P_{t-2}$	-0.0288	0.0767	-0.2064***	0.06989	-0.0470	0.0794	-0.1246	0.0905	-0.2349***	0.11602	-0.1583**	0.0791
$1-Di*Sent_P_{t-3}$	0.23671*	0.1298	-0.0370	0.0933	-0.0217	0.0711	0.0547	0.0713	0.0510	0.1425	0.0178	0.0784
$1-Di*Sent_P_{t-4}$	0.0657	0.0795	0.16027**	0.0648	0.0479	0.1014	-0.0974	0.1074	-0.1875	0.1310	-0.0134	0.0766
$1-Di*Sent_P_{t-5}$	-0.0551	0.0822	-0.0170	0.1033	0.0666	0.0654	-0.0621	0.0866	0.1320	0.1562	0.0151	0.0778
$1-Di*Volat_{t-1}$	-0.6719***	0.35306	-0.3440	0.4139	0.0294	0.3026	0.2308	0.3889	-0.4339	0.5869	-0.2476	0.3057
$1-Di*Volat_{t-2}$	-0.2321	0.5933	0.3587	0.5142	-0.2054	0.3942	0.1760	0.4553	1.48051**	0.7092	0.1905	0.4130
$1-Di*Volat_{t-3}$	0.4930	0.6670	-0.2028	0.4649	0.0391	0.5057	0.1625	0.3504	-0.8456	0.9211	-0.0238	0.3493

Tabela 13 - Impacto do sentimento de pessimismo sobre o retorno do IBOVESPA em dias de alta e baixa incerteza

(Conclusão)

Painel A: Estimação das Regressões												
Variáveis	Quantis										OLS	
	$\tau - 0.05$		$\tau - 0.25$		$\tau - 0.50$		$\tau - 0.75$		$\tau - 0.95$		coef.	Erro padrão
	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão
$I-Di * Volat_{t-4}$	-0.4337	0.5760	-0.7664	0.5919	-0.8705	0.5454	-1.1927***	0.40943	1.3051	1.1896	-0.786**	0.30687
$I-Di * Volat_{t-5}$	0.3594	0.3715	0.6177	0.4727	0.8928***	0.3330	0.63352**	0.3082	-1.2265	0.8390	0.7512***	0.2262
Intercepto	-2.5073***	0.11171	-1.2155***	0.0906	-0.212***	0.07209	0.5818***	0.0938	2.1739***	0.1511	-0.2692***	0.0769
Observações	1041		1041		1041		1041		1041		1041	
R ²	0.15658		0.06037		0.02972		0.02674		0.12123		0.01192	
Testes de significância conjunta												
Teste F	-	-	-	-	-	-	-	-	-	-	-	1.416
Teste Wald	4.0753*											-

Painel B: Teste da igualdade dos coeficientes											
Teste		Estatística do Teste Wald		p-valor		Estatística do Teste Wald		p-valor			
t-1	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	17.2274		3.450e-05 ***		t-4	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	27.5523		1.685e-07***	
t-2	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	1.1440		0.284941		t-5	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	3.6154		0.057385*	
t-3	$\beta_1(\tau_{0.05}) = \beta_1(\tau_{0.95})$	11.6143		0.000667***							
Teste		Estatística do Teste Wald		p-valor		Estatística do Teste Wald		p-valor			
t-1	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	4.1215		0.042469**		t-4	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	2.8537		0.091312*	
t-2	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	2.2990		0.129610		t-5	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	1.1693		0.279662	
t-3	$\beta_2(\tau_{0.05}) = \beta_2(\tau_{0.95})$	0.9785		0.322686							

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

Modelo Geral – Eq. 14: $Ret_t = \beta_0 + (D_t)(\beta_1 \mathcal{L}_s Sent_P_t + \gamma_1 \mathcal{L}_s (Ret_t) + \psi_1 \mathcal{L}_s (\sigma_t^2)) + (1 - D_t)(\beta_2 \mathcal{L}_s Sent_P_t + \gamma_2 \mathcal{L}_s (Ret_t) + \psi_2 \mathcal{L}_s (\sigma_t^2)) + v_t$

. D_t – dummy relacionada aos dias de alta incerteza econômica no Brasil, com base no índice IIE-Br, recebendo o valor “1” quando o IIE-Br foi igual ou superior a 122.7 (valor do 75º percentil na amostra analisada), e, recebeu “0” para os períodos em que o IIE-Br foi inferior ao valor do 75º percentil da amostra.

. $Di * Ret_{t-n}$ – retorno do mercado em dias de alta incerteza, com n lags;

. $Di * Sent_P_{t-n}$ – sentimento de pessimismo média em dias de alta incerteza, com n lags, representado no modelo como SP;

. $Di * Volat_{t-n}$ – volatilidade do mercado de alta incerteza, com n lags, representado no modelo por σ_t^2 ;

. $I-Di * Ret_{t-1}$ – retorno do mercado em dias de baixa incerteza, com n lags;

. $I-Di * Sent_P_{t-n}$ – sentimento de pessimismo das notícias quando existe baixa incerteza, com n lags, representado no modelo como SP;

. $I-Di * Volat_{t-n}$ – volatilidade em dias de baixa incerteza, com n lags, representado no modelo por σ_t^2 .

*Significante ao nível de 10%.; ** significativo ao nível de 5%; ***significante ao nível de 1%.

Nota: regressão nos quantis (τ): 0,05; 0,25; 0,5; 0,75; 0,95, admitindo que os resíduos não eram identicamente distribuídos, com base na matriz de *Huber Sandwich Standard*. A estatística R^2 para a estimação OLS foi o R^2 ajustado, enquanto, para as estimações em regressão quantílica foi o pseudo R^2 , conforme proposta de Koenker e Machado (1999). Variável sentimento foi normalizada, com média zero e variância 1, para ser possível analisar a magnitude do impacto econômico sobre o retorno. A análise da igualdade dos coeficientes foi obtida por meio do teste wald.

A Tabela 14 apresenta os resultados da estimação de uma regressão autoregressiva da volatilidade, admitindo a inclusão de uma variável sobre o tom de pessimismo das notícias financeiras, chamada de sentimento de pessimismo da mídia, com a finalidade de avaliar o impacto do tom das notícias sobre a volatilidade na presença de incertezas econômicas no Brasil. Assim, para a determinação dos períodos de incerteza econômica foi incluída uma *dummy*, D_i , classificada segundo o comportamento do índice IIE-Br da FGV, em que recebeu o valor “1” para os períodos de alta incerteza, e “0”, para períodos de baixa incerteza.

Adotou-se um procedimento semelhante ao utilizado na estimação do retorno em períodos de incerteza, a fim de evitar uma subdivisão na amostra analisada. Com isso, a *dummy*, D_i , foi incorporada ao modelo, interagindo com todas as variáveis, para capturar o efeito das variáveis em períodos de alta incerteza e, para analisar o impacto dos períodos com menor incerteza, foi incluída a expressão “ $1-D_i$ ”, também, interagindo com todas as variáveis do modelo.

As evidências sobre o sentimento de pessimismo mostram que os impactos mais significativos sobre a volatilidade do retorno ocorreram nos períodos de turbulência no mercado, comportamento que foi observado, tanto nos momentos de alta incerteza, quanto, nos períodos de baixa incerteza. Nos períodos de alta, incerteza o impacto mais significativo foi obtido na quinta *lag*, $D_i * Sent_P_{t-5}$, provocando uma redução de -7.71 pontos-base sobre a volatilidade do retorno no quantil q.75. Já nos períodos de baixa incerteza, o impacto mais expressivo foi observado na primeira *lag*, $1-D_i * Sent_P_{t-1}$, com uma redução, no quantil q.95, de -7.36 pontos-base sobre a volatilidade do retorno. Esses resultados levam à constatação de que, na alta incerteza existe um maior impacto do pessimismo da mídia sobre a volatilidade. No entanto, a persistência do impacto do sentimento sobre a volatilidade ocorre sobre os períodos de baixa incerteza. Ademais contrário às evidências empíricas de Kothari et al. (2009) para o mercado norte americano, os resultados encontrados neste estudo, apontam que, no mercado acionário brasileiro, um aumento no pessimismo da mídia provoca a redução da volatilidade do retorno do IBOVESPA, reforçando a constatação anteriormente levantada.

Em relação à volatilidade defasada, verifica-se que nos períodos de alta incerteza econômica, o impacto significativo da volatilidade é encontrado com uma *lag* de defasagem. A partir da segunda defasagem, não são encontrados efeitos significativos. Já nos períodos de baixa incerteza, foram encontradas significâncias estatísticas, não só na primeira defasagem,

mas também na quarta e quinta defasagem temporal, porém, a magnitude do impacto é sempre maior na primeira defasagem, $I-Di*Volat_{t-1}$, chegando a apresentar um impacto de 99.48 pontos-base no quantil q.75. Esses resultados mostram indícios de agrupamentos de volatilidade, como foi levantado em todas as estimações realizadas neste estudo. Essas evidências seguem as explicações de Fama (1965) e Brooks (2014), de que a volatilidade contemporânea apresenta uma correlação positiva com a volatilidade de períodos anteriores. Com isso, altas/baixas volatilidades tendem a serem seguidas por altas/baixas volatilidades.

Tabela 14 - Impacto do sentimento de pessimismo sobre a volatilidade do IBOVESPA em dias de alta e baixa incerteza

(Continua)

Painel A: Estimação das Regressões												
Variáveis	Quantis											
	$\tau - 0.05$		$\tau - 0.25$		$\tau - 0.50$		$\tau - 0.75$		$\tau - 0.95$		OLS	
	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão
$Di*Volat_{t-1}$	0.82739***	0.0798	0.88905* **	0.0342	0.94845***	0.0356	0.92114** *	0.0676	1.1544	0.7557	0.8649**	0.0376
$Di*Volat_{t-2}$	0.0878	0.0850	0.0286	0.0293	-0.0364	0.0573	0.0774	0.1580	0.2468	0.9703	0.0516	0.0499
$Di*Volat_{t-3}$	0.0467	0.0908	0.0231	0.0230	0.0366	0.0692	-0.0819	0.1495	-0.5138	1.1162	-0.0684	0.0595
$Di*Volat_{t-4}$	-0.0222	0.1098	-0.0009	0.0187	-0.0188	0.0667	-0.0218	0.0861	0.1078	0.6726	0.0849	0.0665
$Di*Volat_{t-5}$	-0.0630	0.0712	-0.0164	0.0100	-0.0054	0.0409	0.0562	0.0493	0.0371	0.1990	-0.0251	0.0481
$Di*Sent_P_{t-1}$	-0.0233	0.0197	-									
			0.0125** *	0.00207	-0.0234*	0.01418	-0.0078	0.0137	-0.0094	0.1485	-0.0257	0.0242
$Di*Sent_P_{t-2}$	0.0020	0.0192	0.0006	0.0019	-0.0034	0.0151	-0.0390	0.0246	-0.1955	0.1711	-0.045*	0.0243
$Di*Sent_P_{t-3}$	-0.0291	0.0281	0.0022**	0.0009	-0.0012	0.0120	0.0232***	0.0069	0.0233	0.2154	0.0361	0.0254
$Di*Sent_P_{t-4}$	0.0012	0.0190	-0.0018	0.0030	0.0166	0.0119	0.0305	0.0301	-0.0090	0.3979	-0.0644**	0.0256
$Di*Sent_P_{t-5}$	-0.0173	0.0175	0.0010	0.0053	-0.0246	0.0168	-0.0771**	0.03335	-0.1893	0.1983	-	
											0.1055***	0.02458
$I-Di*Volat_{t-1}$	0.84344***	0.0468	0.9218** *	0.0091	0.91263***	0.0251	0.99487** *	0.0772	0.9801** *	0.0521	0.9033***	0.0547
$I-Di*Volat_{t-2}$	0.0600	0.0602	0.0197	0.0134	0.0449	0.0339	0.0278	0.1090	-0.1433	0.1077	0.0385	0.0739
$I-Di*Volat_{t-3}$	-0.0708	0.0644	-0.0186	0.0144	-0.0178	0.0439	-0.0225	0.1414	0.3183	0.2154	-0.0216	0.0616
$I-Di*Volat_{t-4}$	-0.1538**	0.07542	0.0051	0.0166	-0.0338	0.0331	-0.0813	0.1493	0.2017	0.2144	-0.1055**	0.0530
$I-Di*Volat_{t-5}$	0.1878***	0.0534	0.0048	0.0107	0.03961**	0.0177	0.0858	0.0999	-0.2626	0.1753	0.1484***	0.0398
$I-Di*Sent_P_{t-1}$	0.0086	0.0127	-0.0005	0.0022	-0.0148**	0.00714	-0.0406**	0.01888	-			
									0.0736** *	0.02381	-0.027*	0.01410
$I-Di*Sent_P_{t-2}$	-0.0153*	0.00831	-0.0021	0.0017	0.0020	0.0058	0.0041	0.0131	0.0499** *	0.0174	0.0016	0.0140
$I-Di*Sent_P_{t-3}$	-0.0028	0.0067	0.0008	0.0014	0.00714*	0.0038	0.0053	0.0174	-0.0048	0.0243	0.0056	0.0138

Tabela 14 - Impacto do sentimento de pessimismo sobre a volatilidade do IBOVESPA em dias de alta e baixa incerteza

(Conclusão)

Painel A: Estimação das Regressões													
Variáveis	$\tau - 0.05$		$\tau - 0.25$		$\tau - 0.50$		$\tau - 0.75$		$\tau - 0.95$		OLS		
	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	coef.	Erro padrão	
$I-Di*Sent_P_{t-4}$	0.0114*	0.0069	0.0012	0.0019	0.0049	0.0043	0.0080	0.0177	-0.0275	0.0273	0.0052	0.0135	
$I-Di*Sent_P_{t-5}$	-0.0176***	0.00494	0.0049***	0.0021	-0.0205***	0.00638	-0.0197	0.0140	-0.0154	0.0277	-0.024*	0.01378	
Intercepto	-0.2212***	0.01208	0.1142***	0.00195	-0.0804***	0.00667	0.0305*	0.0184	0.3124**	*	0.0422	-0.0335**	0.013864
Observações	1041		1041		1041		1041		1041		1041		
R ²	0.77125		0.81119		0.81255		0.79215		0.75150		0.9113		
Testes de significância conjunta													
Teste F												532.5	
Teste Wald	6.9736***												

Painel B: Teste da igualdade dos coeficientes

Teste		Estatística do Teste Wald	p-valor	Teste		Estatística do Teste Wald	p-valor
t-1	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	0.0087	0.925823	t-4	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	0.0007	0.979486
t-2	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	1.3306	0.248834	t-5	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	0.7504	0.386438
t-3	$\beta_1(\tau_{.05}) = \beta_1(\tau_{.95})$	0.0588	0.808390				
Teste		Estatística do Teste Wald	p-valor	Teste		Estatística do Teste Wald	p-valor
t-1	$\beta_2(\tau_{.05}) = \beta_2(\tau_{.95})$	9.6504	0.001919 ***	t-4	$\beta_2(\tau_{.05}) = \beta_2(\tau_{.95})$	1.9657	0.161054
t-2	$\beta_2(\tau_{.05}) = \beta_2(\tau_{.95})$	11.8408	0.000591 ***	t-5	$\beta_2(\tau_{.05}) = \beta_2(\tau_{.95})$	0.0061	0.937835
t-3	$\beta_2(\tau_{.05}) = \beta_2(\tau_{.95})$	0.0065	0.935771				

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

Modelo Geral – Eq. 15: $\sigma_t^2 = \alpha_0 + (D_I)(\beta_1 \mathcal{L}_s Sent_P_t + \psi_1 \mathcal{L}_s(\sigma_t^2)) + (1 - D_I)(\beta_2 \mathcal{L}_s SentP_t + \psi_2 \mathcal{L}_s(\sigma_t^2)) + v_t$

. Di – *dummy* relacionada aos dias de alta incerteza econômica no Brasil, com base no índice IIE-Br, recebendo o valor “1” quando o IIE-Br foi igual ou superior a 122.7 (valor do 75º percentil na amostra analisada), e, recebeu “0” para os períodos em que o IIE-Br foi inferior ao valor do 75º percentil da amostra.

. $Di * Sent_P_{t-n}$ – sentimento da mídia em dias de alta incerteza, com n lags, representado por SP na equação;

. $Di * Volat_{t-n}$ – volatilidade do mercado de alta incerteza, com n lags, representado no modelo por σ_t^2 ;

. $1-Di * Sent_P_{t-n}$ – sentimento das notícias quando existe baixa incerteza, com n lags, representado por SP na equação

. $1-Di * Volat_{t-n}$ – volatilidade em dias de baixa incerteza, com n lags, representado no modelo por σ_t^2 .

* Significante ao nível de 10%.; ** significante ao nível de 5%; ***significante ao nível de 1%.

Nota: regressão nos quantis (τ): 0,05; 0,25; 0,5; 0,75; 0,95, admitindo que os resíduos não eram identicamente distribuídos, com base na matriz de *Huber Sandwich Standard*. A estatística R^2 para a estimação OLS foi o R^2 ajustado, enquanto para as estimações em regressão quantílica foi o pseudo R^2 , conforme proposta de Koenker e Machado (1999). Variável sentimento foi normalizada, com média zero e variância 1, para ser possível analisar a magnitude do impacto econômico sobre a volatilidade. A análise da igualdade dos coeficientes foi obtida por meio do teste *wald*.

Diante das evidências encontradas sobre o efeito do sentimento de pessimismo da mídia sobre os movimentos do mercado quanto ao retorno do IBOVESPA em períodos marcados por uma maior incerteza econômica, não pôde ser rejeitada a hipótese de que o impacto do sentimento de pessimismo das notícias sobre o retorno no mercado acionário brasileiro é maior em períodos de incerteza econômica (H5). No entanto, as evidências levantadas sobre o impacto do tom das notícias financeiras sobre o risco (volatilidade) levam à rejeição da hipótese de que o impacto do sentimento de pessimismo das notícias sobre a volatilidade no mercado acionário brasileiro é maior em períodos de incerteza econômica (H6), visto que, no mercado acionário brasileiro, são encontrados indícios do sentimento de pessimismo das notícias ser mais persistentes sobre períodos de baixa incerteza econômica.

4.5.4 Síntese dos Resultados da Influência do Sentimento Textual Sobre o Risco e Retorno do Ibovespa

Na análise da influência do sentimento textual sobre o comportamento dos preços no mercado acionário, busca-se verificar se as principais palavras utilizadas na descrição das notícias financeiras poderiam fornecer informações capazes de explicar os movimentos do risco (volatilidade) e do retorno do mercado acionário brasileiro.

Na literatura são apontados indícios de que a mídia de países emergentes não seria especializada o suficiente ao ponto de provocar reações no mercado (GRIFFIN, HIRSCHEY; KELLY, 2011; FRAIBERGER, 2016). Porém, não foram encontradas evidências conclusivas sobre tal comportamento. Pelo contrário, foram encontrados incentivos na literatura (KERNEY; LIU, 2014) para a execução de análises textuais, respeitando as peculiaridades linguísticas da língua oficial de um país.

Diante disso, procurou-se analisar o conteúdo informacional das notícias financeiras escritas em português, com técnicas semelhantes às utilizadas para os mercados desenvolvidos, porém com as devidas adaptações para língua portuguesa, em termos de código de programação. Assim, foram estruturadas listas de termos financeiros e foi realizada a construção de algoritmos para leitura de textos em português, com a finalidade de capturar informações relacionadas ao sentimento das notícias publicadas na língua portuguesa.

Mediante a estruturação de uma variável sobre o sentimento das notícias, buscou-se investigar o reflexo do sentimento dos textos sobre os movimentos do mercado acionário do Brasil.

No Quadro 5, é apresentada uma síntese das hipóteses e dos resultados levantados quanto à influência do sentimento textual das notícias financeiras sobre o mercado acionário brasileiro.

Quadro 5 - Síntese das Hipóteses Analisadas Sobre a Influência do Sentimento Textual Sobre o Mercado Acionário Brasileiro

Hipóteses	Resultado	Decisão
H₁ : O sentimento das notícias exerce influência sobre o retorno das ações do mercado acionário brasileiro.	Observou-se que o sentimento das notícias provoca impactos sobre o retorno, quando o mercado apresenta os piores (baixos retornos) e melhores (altos retornos) desempenhos.	Não rejeita H ₁
H₂ : O sentimento das notícias exerce influência sobre a volatilidade do retorno das ações do mercado acionário brasileiro.	Verificaram-se impactos do sentimento textual nos períodos de maior turbulência do mercado (maiores volatilidades).	Não rejeita H ₂
H₃ : O sentimento pessimista das notícias reduz o retorno no mercado acionário brasileiro.	Observou-se que o sentimento de pessimismo e de otimismo das notícias financeiras provocam impactos sobre os movimentos do retorno do mercado. O sentimento de pessimismo reduz o retorno e o sentimento de otimismo aumenta o retorno.	Não rejeita H ₃
H₄ : O sentimento pessimista das notícias eleva a volatilidade do retorno no mercado acionário brasileiro.	Observou-se que o sentimento de pessimismo da mídia tende a reduzir a volatilidade do retorno do IBOVESPA.	H₄ Rejeita
H₅ : O impacto do sentimento de pessimismo das notícias sobre o retorno no mercado acionário brasileiro é maior em períodos de incerteza econômica.	Verificou-se que os impactos sobre o retorno, em termos de magnitude, são maiores na alta incerteza.	H ₅ Não rejeita
H₆ : O impacto do sentimento de pessimismo das notícias sobre a volatilidade no mercado acionário brasileiro é maior em períodos de incerteza econômica.	Constatou-se que na alta incerteza existe um maior impacto do pessimismo da mídia sobre a volatilidade, no entanto, a persistência do impacto do sentimento sobre a volatilidade ocorre sobre períodos de baixa incerteza. Além disso, verificou-se que um aumento no pessimismo da mídia provoca reduções na volatilidade do retorno.	H₆ Rejeita

Fonte: Resultados da Pesquisa, elaborado com base nos dados sobre os índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

4.6 PREVISÃO DA VOLATILIDADE

4.6.1 Análises das variáveis preditoras da volatilidade

A volatilidade não é passível de uma observação direta, com isso, faz-se necessária a utilização de medidas para estimá-la. Neste estudo, a volatilidade do mercado foi modelada por um GARCH (1,1) com distribuição *t-Student*, e a série temporal da variância condicional dos retornos do IBOVESPA, obtidas na estimação do GARCH (1,1), foi utilizada como uma *proxy* para a volatilidade realizada do IBOVESPA, que representou o comportamento da volatilidade do mercado acionário brasileiro. A escolha por essa modelagem ocorreu em função de ser uma boa medida de estimação da volatilidade de ativos financeiros, conforme é apontado por Bollerslev (1986), bem como por ser uma medida muito utilizada para modelar a volatilidade do mercado brasileiro, devido à sua eficiência (GODEIRO; LIMA, 2017).

A Tabela 15 sintetiza os resultados da escolha das variáveis preditoras da volatilidade mediante o método de regressão quantílica ℓ_1 -penalizada - LASSO, admitindo que o poder de predição das variáveis pudesse oscilar ao longo da distribuição da volatilidade e no decorrer do tempo. Assim, os resultados apresentados na Tabela 15, mostram a frequência de escolha de cada variável na predição da volatilidade, nos quantis pertencentes ao intervalo τ [0,1].

Os resultados dispostos na Tabela 15 apresentam a frequência percentual da escolha das variáveis preditoras, dada pelo método de penalização por LASSO, considerando o período fora da amostra. Desse modo, uma variável foi considerada como um forte preditor da volatilidade, quando realizava previsões em todos os quantis. Quando realizavam previsões somente em alguns quantis, as variáveis foram classificadas como preditores parcialmente fracos. Por fim, foram consideradas como preditores fracos quando não conseguiam fazer previsões em nenhum quantil.

A capacidade preditiva do sentimento textual das notícias financeiras, *SENT*, sobre a volatilidade se encontra após o 5º decil, evidenciando que o poder informativo do sentimento das notícias sobre a volatilidade tende a ocorrer nas extremidades da distribuição da volatilidade. Desse modo, pode-se dizer que as informações provenientes dos jornais contribuem com a previsão da volatilidade nos momentos em que existe uma maior oscilação da volatilidade, o que seria um reflexo da importância que as informações provenientes dos jornais apresentam para os investidores nos momentos de maior incerteza.

Tabela 15 - Preditores da Volatilidade escolhidos por Lasso

T	$SENT$	VIX	$S\&P500$	$DJIA$	$VOLAT_{t-1}$
$\tau = 0.1$	-	-	-	-	100%
$\tau = 0.2$	-	-	-	-	100%
$\tau = 0.3$	-	-	-	-	100%
$\tau = 0.4$	-	-	-	-	100%
$\tau = 0.5$	-	96.01%	-	-	100%
$\tau = 0.6$	22.78%	98.50%	-	-	100%
$\tau = 0.7$	17.06%	99.37%	0.74%	0.99%	100%
$\tau = 0.8$	7.22%	74.09%	-	-	100%
$\tau = 0.9$	9.71%	70.98%	-	-	100%

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

Nota: *Sent* – índice de sentimento textual diário; *VIX* – índice de volatilidade americano da S&P500; *S&P500* retorno do S&P500; *DJIA* – índice da Dow Jones; *Volat_{t-1}* - volatilidade do IBOVESPA defasada.

Resultados mostram a escolha dos preditores fora da amostra, janela temporal 04/2014 a 06/2017, com estimação da previsão realizada por regressão quantílica combinada (RQC).

*nível de significância adotado na estimação foi de 5%.

Apesar dos resultados indicarem que o sentimento textual é um preditor da volatilidade, ele se configura como um preditor parcialmente fraco, visto que, a frequência percentual de seleção da variável, *SENT*, está concentrada somente em alguns decis. Dentre os decis estimados, a maior importância do sentimento textual se encontra no 6° decil, apresentando uma frequência de escolha de 22.78% na previsão da volatilidade do referido decil. Assim, pode-se dizer que o percentual de escolha está relacionado ao nível de importância atribuído à variável para realizar a previsão da volatilidade.

O índice de medo do mercado americano, *VIX*, também configurou-se como um preditor parcialmente fraco da volatilidade do IBOVESPA, assim como os indicadores *S&P500* e *DJIA*. Porém, realizando uma comparação entre as *proxies* do comportamento do mercado americano, em termos de importância como preditores da volatilidade do mercado brasileiro, constata-se que os indicadores *S&P500* e *DJIA* têm uma pequena importância, pois a frequência de seleção foi inferior a 1% no 7° decil, ao contrário do índice *VIX*, que apresenta um alto percentual de escolha na previsão, entre o 5° e 9° decil, chegando a apresentar uma frequência de escolha de 99,37% no 7° decil. Desse modo, dentre as *proxies* do mercado americanos, o *VIX* se configura como um importante preditor da volatilidade do IBOVESPA.

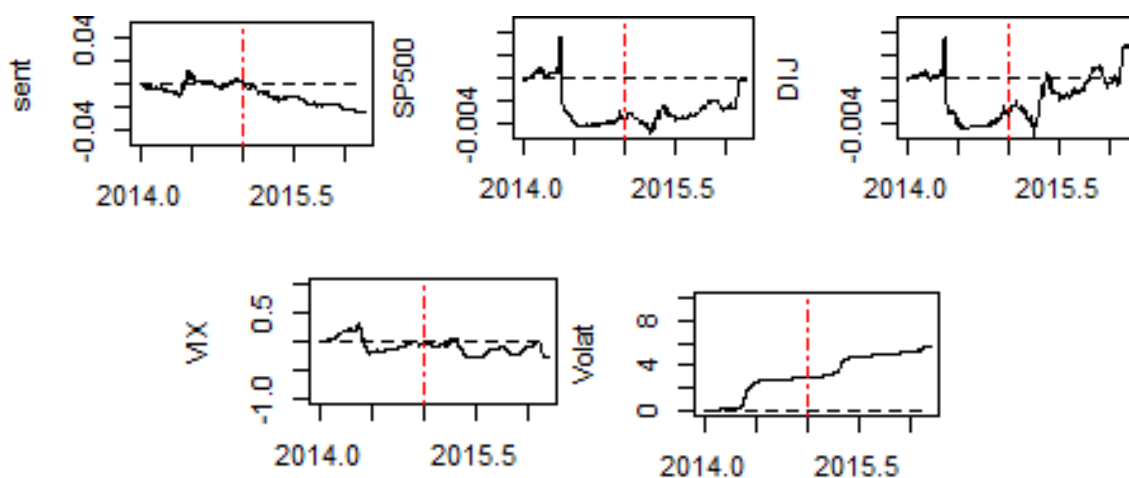
Já a volatilidade passada, *VOLAT_{t-1}*, não só é um forte preditor da volatilidade, como também apresenta uma alta importância em todos os decis analisados. Esse resultado reflete a

eficiência da modelagem de estimação da volatilidade, devido ao agrupamento da volatilidade (BROOKS, 2014).

4.6.2 Análises dos Modelos de Estimação da Previsão da Volatilidade

A Figura 6 trata sobre desempenho acumulado dos modelos de previsão univariados em relação ao desempenho do modelo geral, que foi estimado com base na média histórica da volatilidade. A estimação do desempenho, expressa na figura 6, é dada pela diferença entre o erro quadrático da previsão do modelo geral e o erro quadrático dos modelos univariados. Assim, quando o erro quadrático da previsão do modelo univariado é inferior ao erro quadrático da previsão do modelo geral, a curva tende a ser ascendente, já quando os erros do modelo univariado são superiores aos do modelo geral, a curva tem um comportamento descendente.

Figura 6 - Erros de Previsões da Volatilidade em Modelos Univariados



Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

Nota: A figura mostra o erro quadrático da previsão do modelo geral menos o erro quadrático da previsão dos modelos condicionais univariados acumulado ao longo do período temporal, relacionado à janela de fora da amostra, dado entre os dias 03/04/2014 a 30/06/2017.

Desse modo, analisando o comportamento dos modelos univariados, sintetizados na Figura 6, observa-se que, em geral, as variáveis sozinhas não conseguiram prever a volatilidade, visto que, os erros de previsão do modelo univariado superaram os erros do modelo

geral, levando as curvas a apresentarem um comportamento descendente ao longo do tempo. As exceções para esse comportamento, são observadas nas variáveis $VOLAT_{t-1}$ e o $DJIA$.

Em relação ao modelo univariado da volatilidade defasada, $VOLAT_{t-1}$, constata-se que os erros de previsão são inferiores aos erros de previsão do modelo geral, ao longo do período analisado, e tal comportamento mostra indícios de que a volatilidade defasada se configura como um importante preditor da volatilidade contemporânea, pois o desempenho do modelo superou a média histórica ao longo de todo o período de análise

A variável relacionada ao índice *Dow Jones*, $DJIA$, inicialmente mostra que os erros de previsão são inferiores ao modelo geral, porém, existe um enfraquecimento acentuado do modelo em meados de 2014, e este, volta a apresentar erros inferiores aos erros do modelo de previsão apenas ao final do período analisado.

Observando o desempenho acumulado do sentimento textual, $SENT$, nota-se que na primeira parte da amostra, período entre 04/2014 a 12/2015, os erros quadráticos são inferiores aos erros dos modelos gerais, no entanto, após 2015, existe um enfraquecimento do modelo, observado pelo comportamento descendente que persistiu de 2015 até o final do período analisado, mostrando, com isso, que a previsibilidade do sentimento textual perde a capacidade preditiva ao longo do tempo. Esse resultado corrobora com a visão de Carporale et al. (2016) de que o impacto das notícias sobre o mercado muda ao longo do tempo.

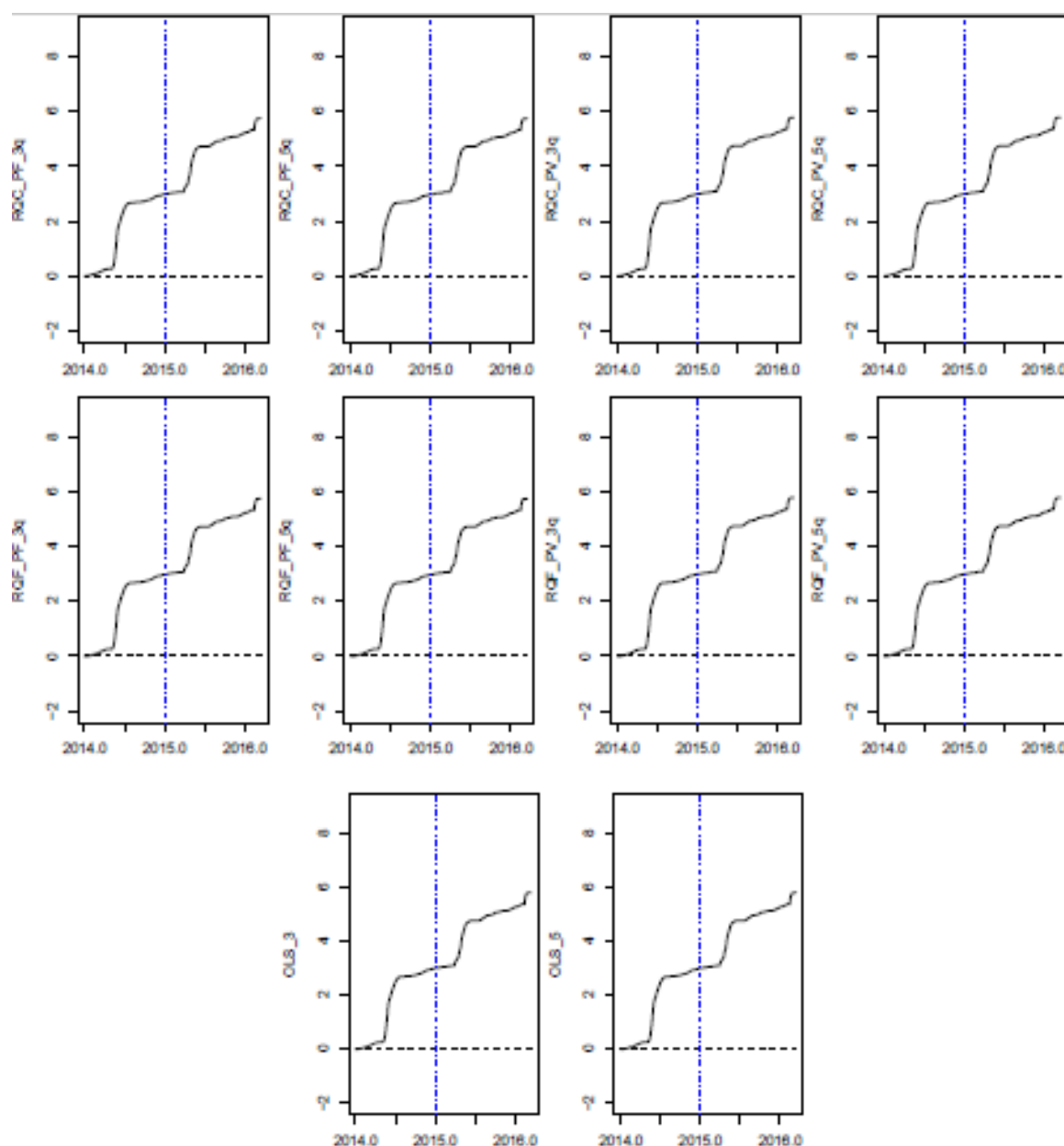
Nos modelos univariados das variáveis VIX e $S\&P500$, é observada uma baixa capacidade preditiva. O VIX apresenta um enfraquecimento na previsão em meados de 2014, exibindo uma leve recuperação em 2015, porém o modelo não demonstra uma minimização dos erros que persista ao longo do tempo. Já o modelo do $S\&P500$, apresenta um acentuado enfraquecimento da previsão em 2014 que persiste até meados de 2015, quando o modelo volta a retomar a sua capacidade de predição.

Diante do desempenho dos modelos univariados apresentados neste estudo, observa-se que os modelos preditivos formados por uma única variável costumam não realizar previsões da volatilidade que superem a média histórica.

A Figura 7 sintetiza os resultados do desempenho dos modelos de previsão da volatilidade com a junção das variáveis, fazendo uma análise semelhante à da Figura 6, de modo que, os resultados expressam o erro quadrático de previsão do modelo de geral menos o erro quadrático de previsão dos modelos utilizados para prever a volatilidade, que foram:

RQC_PF_3q; RQC_PF_5q; RQC_PV_3q; RQC_PV_5q; RQF_PF_3q; RQF_PF_5q; RQF_PV_3q; RQF_PV_5q; OLS_3; e OLS_5q. Da mesma forma, curvas ascendentes nas estimativas dos modelos condicionais evidenciam que os erros de previsão acumulados ao longo do tempo são inferiores aos erros de previsão do modelo baseado na média histórica da volatilidade, chamado neste estudo de modelo geral.

Figura 7 - Erros de Previsões da Volatilidade em Modelos de Regressão Quantílica Combinada (RQC), Regressão Quantílica fixa (RQF) e OLS.



Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

Nota: A figura mostra o erro de previsão quadrático da previsão do modelo geral menos o erro quadrático de previsão dos modelos de previsão RQC, RQF e OLS acumulado ao longo do período temporal relacionado a janela de fora da amostra, dado entre os dias 03/04/2014 a 30/06/2017.

Modelos:

$$\begin{aligned}
 & \cdot \text{RQC_PF_3q Eq.25: } y_{t+1} = \frac{1}{3} f_{t+1,t}^{0.3} + \frac{1}{3} f_{t+1,t}^{0.5} + \frac{1}{3} f_{t+1,t}^{0.7} \\
 & \cdot \text{RQC_PF_5q - Eq. 26: } y_{t+1} = \frac{1}{5} f_{t+1,t}^{0.3} + \frac{1}{5} f_{t+1,t}^{0.4} + \frac{1}{5} f_{t+1,t}^{0.5} + \frac{1}{5} f_{t+1,t}^{0.6} + \frac{1}{5} f_{t+1,t}^{0.7} \\
 & \cdot \text{RQC_PV_3q - Eq.27: } y_{t+1} = \sum_{\tau=\tau_1}^{\tau_3} \omega_{\tau} f_{t+1,t}^{\tau} + \varepsilon_{t+1} \\
 & \cdot \text{RQC_PV_5q - Eq. 28: } y_{t+1} = \sum_{\tau=\tau_1}^{\tau_5} \omega_{\tau} f_{t+1,t}^{\tau} + \varepsilon_{t+1} \\
 & \cdot \text{RQF_PF_3q - Eq. 29: } f_{t+1} = \sum_{j=1}^3 \omega_{\tau_j} \beta_0(\tau_j) + \sum_{j=1}^3 \omega_{\tau_j} \beta_1(\tau_j) x_{1,t} + \sum_{j=1}^3 \omega_{\tau_j} \beta_n(\tau_j) x_{n,t} \\
 & \cdot \text{RQF_PF_5q - Eq. 30: } f_{t+1} = \sum_{j=1}^5 \omega_{\tau_j} \beta_0(\tau_j) + \sum_{j=1}^5 \omega_{\tau_j} \beta_1(\tau_j) x_{1,t} + \sum_{j=1}^5 \omega_{\tau_j} \beta_n(\tau_j) x_{n,t} \\
 & \cdot \text{RQF_PV_3q - Eq. 31: } f_{t+1} = \sum_{j=1}^3 \omega_{\tau_j} \beta_0(\tau_j) + \sum_{j=1}^3 \omega_{\tau_j} \beta_1(\tau_j) x_{1,t} + \sum_{j=1}^3 \omega_{\tau_j} \beta_n(\tau_j) x_{n,t} \\
 & \cdot \text{RQF_PV_5q - Eq.32: } f_{t+1} = \sum_{j=1}^5 \omega_{\tau_j} \beta_0(\tau_j) + \sum_{j=1}^5 \omega_{\tau_j} \beta_1(\tau_j) x_{1,t} + \sum_{j=1}^5 \omega_{\tau_j} \beta_n(\tau_j) x_{n,t}
 \end{aligned}$$

Analisado o comportamento das curvas de desempenho dos modelos de previsão, apresentados na Figura 7, constata-se que todos os modelos condicionais superaram o modelo geral, ou seja, o erro da previsão dos modelos condicionais foi inferior ao erro de previsão do modelo geral. Além disso, a curva do desempenho acumulado de cada um dos modelos, mostra-se sempre ascendente, indicando que os modelos condicionais não só superaram o modelo geral, como também ganharam força em prever a volatilidade ao longo do tempo, ao contrário, da maioria dos modelos univariados. Porém, a análise gráfica não possibilita uma observação detalhada sobre quais modelos apresentaram os melhores desempenhos, visto que, existem muitas semelhanças entre as curvas de distribuição do desempenho dos modelos estimados por OLS e por regressão quantílica.

Assim, para uma melhor investigação sobre o desempenho dos modelos de previsão da volatilidade, foi analisado o resultado do poder explicativo do modelo, com base no teste R^2 , e mediante os testes MSPE *ajust* e o teste de Diebold e Mariano (1995), conforme resultados apresentados na Tabela 16.

Tabela 16 - Modelos condicionais de previsão da volatilidade

	Amostra Geral: Abril de 2014 a Junho 2017			Amostra I: Abril de 2014 – Dezembro de 2015			Amostra II: Janeiro de 2016 a Junho de 2017		
	R ²	MSPE <i>ajust</i>	Diebold e Mariano	R ²	MSPE <i>ajust</i>	Diebold e Mariano	R ²	MSPE <i>ajust</i>	Diebold e Mariano
Previsão Univariada									
<i>Sent</i>	-0.004250	0.999561	0.8822004	-0.00218129	0.9858754	0.498036545	-0.006481416	0.999033349	0.997217879
<i>VIX</i>	-0.044293	1.56E-13	1.02E-08	-0.02164014	9.33E-13	5.55E-08	-0.068711740	0.000959013	0.006548155
<i>S&P500</i>	-0.000012	0.31738	0.4541784	-0.00122247	0.3073098	0.820023794	0.0012922806	0.43107799	0.057952195
<i>DJIA</i>	0.0004751	0.243385	0.240238	-0.00095958	0.2212121	0.726267057	0.0020215743	0.396248215	0.042377905
<i>Volat_{t-1}</i>	0.9428361	2.79E-121	1.08E-41	0.966709119	1.67E-61	3.29E-20	0.9171040187	4.28E-62	2.55E-24
Previsão Multivariada									
<i>OLS_3q</i>	0.9426211	2.46E-123	1.41E-42	0.965330184	5.15E-63	1.25E-20	0.9181434237	1.54E-62	9.37E-25
<i>OLS_5q</i>	0.9426261	2.45E-123	1.41E-42	0.965339822	5.13E-63	1.26E-20	0.9181434237	1.54E-62	9.37E-25
<i>RQF_PF_3q</i>	0.9390976	7.18E-125	5.32E-43	0.962397342	1.28E-63	9.41E-21	0.9139832804	1.73E-63	3.49E-25
<i>RQF_PF_5q</i>	0.9384073	5.49E-125	4.68E-43	0.961628959	9.83E-64	8.92E-21	0.9133772611	1.76E-63	3.28E-25
<i>RQF_PV_3q</i>	0.943165	2.48E-122	4.26E-42	0.966428536	5.79E-62	2.34E-20	0.9180899045	1.13E-62	1.04E-24
<i>RQF_PV_5q</i>	0.9431109	2.63E-122	3.59E-42	0.966319070	5.86E-62	2.08E-20	0.9180952984	1.19E-62	1.06E-24
<i>RQC_PF_3q</i>	0.9391634	8.01E-125	6.30E-43	0.962502042	1.37E-63	1.05E-20	0.9140071993	1.80E-63	3.58E-25
<i>RQC_PF_5q</i>	0.9385665	7.05E-125	6.17E-43	0.961908846	1.16E-63	1.05E-20	0.9134063714	1.91E-63	3.43E-25
<i>RQC_PV_3q</i>	0.9431303	1.69E-122	4.61E-42	0.966370459	4.32E-62	2.46E-20	0.9180803011	1.05E-62	1.03E-24
<i>RQC_PV_5q</i>	0.9431285	1.66E-122	4.62E-42	0.966364866	4.18E-62	2.44E-20	0.9180826734	1.07E-62	1.04E-24

Fonte: Elaborado com base nos dados sobre índices do mercado brasileiro e do mercado americano, disponíveis na Thomson Reuters, e com base nas notícias de finanças do Jornal Valor Econômico.

Nota: A tabela apresenta o *p*-valor do teste MSPE *ajust* e do teste Diebold e Mariano, os quais foram utilizados para investigação a significância estatística dos modelos, admitindo um nível de significância de 0.05 nas análises.

Na Tabela 16, os resultados sobre a amostra geral evidenciam que, dentre os modelos univariados, o que consegue prever a volatilidade é o modelo estruturado pela volatilidade defasada, $Volat_{t-1}$, que apresenta um poder explicativo de 94,28%, com erros de previsão que diferem dos erros do modelo geral, conforme o resultado dos testes MSPE *ajust* e teste de Diebold e Mariano. Nos demais modelos univariados, as variáveis sozinhas não conseguiram fazer previsões estatisticamente significativas que superassem a média histórica da volatilidade.

Ao analisar os modelos compostos pelo conjunto de variáveis predictoras da volatilidade, observou-se que, dentre os modelos estruturados no intervalo de três quantis, todos os modelos condicionais de previsão superaram a média histórica da volatilidade. Os modelos estimados por regressão quantílica com peso variável, RQF_PV_3q (94,316%) e RQC_PV_3q (94,313%), apresentaram um desempenho superior ao desempenho da previsão realizada por OLS_3q (94,26%). Possivelmente, isso ocorreu, em função dos modelos com pesos variáveis admitirem uma maior importância (maior peso) aos quantis que apresentaram os menores erros de previsão, levando os modelos, RQF_PV_3q e RQC_PV_3q , a apresentarem um maior refinamento na previsão da volatilidade.

Analisando especificamente os modelos de regressão quantílica, dentre os quais, a principal diferença encontra-se na composição dos coeficientes, que podem ser fixos (RQF) ou podem apresentar coeficientes variáveis (RQC), constatou-se que os modelos com pesos variáveis RQF_PV_3q (94,316%) e RQC_PV_3q (94,313%) apresentaram melhores desempenhos, enquanto os modelos com coeficientes fixos foram ligeiramente superiores aos modelos com coeficientes variáveis, visto que, na modelagem com coeficientes fixos, os modelos são estatisticamente significativos, conforme testes MSPE *ajust* e Diebold e Mariano (1995), e o R^2 foi superior em 0.003%.

Analisando os modelos estruturados para o intervalo de cinco quantis, observa-se que o desempenho dos modelos foi superior à média histórica da volatilidade, e os melhores modelos de previsão da volatilidade foram as estimações realizadas por regressão quantílica com pesos variáveis, RQF_PV_5q (94,311%) e RQC_PV_5q (94,312%), uma vez que, eles superaram o desempenho dos modelos de regressão quantílica com pesos fixos, RQF_PF_5q (93,84%) e RQC_PF_5q (93,85%), e o desempenho do modelo de previsão estimado por OLS_5 (94,26%).

Desse modo, ao fazer uma comparação entre as estimações das previsões realizadas no intervalo temporal da amostra geral, entre três e cinco quantis, constatou-se que não ocorreram melhorias na previsão com cinco quantis, o que se deve ao fato de o aumento da quantidade de quantis nas estimações não terem agregado informações capazes de reduzir os erros de previsão.

Além disso, constatou-se que os modelos que admitiram pesos variáveis apresentaram os menores erros de previsão, alcançando, conseqüentemente, um melhor desempenho na previsão da volatilidade. Com isso, pode-se dizer que a melhor estimação da previsão da volatilidade foi obtida ao atribuir pesos variáveis aos quantis, atribuindo uma maior importância aos quantis que apresentarem os menores erros de previsão.

Analisando especificamente o poder do sentimento textual para a previsão da volatilidade, na amostra geral, pode-se dizer que o sentimento textual sozinho perde poder preditivo ao longo do tempo, de modo que o sentimento textual, como única variável, não consegue prever a volatilidade. Isso é evidenciado na Tabela 18, nos resultados da amostra geral para o modelo univariado *SENT*, que apresenta um R^2 negativo (-0.425), e não apresenta significância estatística nos testes realizados, evidenciando, com isso, que a média histórica da volatilidade consegue realizar uma melhor previsão que o modelo univariado *SENT*. Porém, quando é levado em consideração o poder preditivo do sentimento por quantis, e quando são atribuídos pesos variáveis a estes, constata-se que existem melhorias na previsão da volatilidade, pois os modelos *RQC_PV_3q* (94,3135%) e *RQC_PV_5q* (94,3128%) apresentam um desempenho superior à estimação univariada do modelo $VOLAT_{t-1}$, estimações por *OLS* e as estimações que consideram pesos fixos. Portanto, é possível dizer que na amostra geral, o sentimento textual contribuiu parcialmente com a redução do erro de previsão da volatilidade.

Além da análise geral, a amostra foi segregada em dois grupos, denominados de Amostra I (período 04/2014 a 12/2015) e Amostra II (01/2016 a 06/2017). Os resultados da Amostra I apresentaram uma melhoria, geral, nos modelos de predição da volatilidade. Dentre os modelos univariados, a variável $VOLAT_{t-1}$ permanece como um forte preditor da volatilidade. Dentre os modelos, que utilizam o conjunto de variáveis predictoras, observa-se que os melhores modelos de previsão foram aqueles estimados por regressão quantílica com pesos variáveis, *RQF_PV_3q* (96,642%); *RQF_PV_5q* (96,637%); *RQC_PV_3q* (96,639%); *RQC_PV_5q* (96,636%), os quais superaram as previsões por *OLS* e as previsões realizadas por regressão quantílica com pesos fixos. Constatou-se, também, que existiram melhorias na previsão ao serem incluídos mais quantis, ou seja, as previsões realizadas dentro do intervalo de cinco quantis foram superiores às previsões realizadas para três quantis.

Ao analisar a Amostra II, observou-se uma queda no poder de previsão dos modelos, pois, o R^2 apresentou valores percentuais menores em todos os modelos, em comparação aos

resultados levados na Amostra I e na Amostra Geral. A amostra II contemplou a janela temporal, entre Janeiro de 2016 a junho de 2017, período que foi marcado por uma grande incerteza econômica no país, desencadeada por muitas denúncias de corrupção no Brasil, envolvendo empresas e políticos. Desse modo, os resultados da queda na previsão da volatilidade na amostra II, estão de acordo com a literatura, pois segundo Bloom (2014), em períodos de maior incerteza, os dados financeiros tornam-se mais voláteis e, com isso, existe uma maior dificuldade em realizar previsões.

Analisando especificamente os modelos na amostra II, observa-se que modelos de regressão quantílica com pesos variáveis apresentam os melhores desempenhos, *RQF_PV_3q* (91,808%); *RQF_PV_5q* (91,809%); *RQC_PV_3q* (91,808%); e *RQC_PV_5q* (91,8082%), a inclusão de mais quantis nas estimações melhorou as previsões, logo, as previsões no intervalo de cinco quantis foram superiores às previsões realizadas no intervalo de três quantis. No entanto, as estimações por OLS, mostraram-se ligeiramente superiores às estimações realizadas por regressão quantílica, especificamente neste caso, a utilização de modelos robustos a valores extremos não apresentou vantagens.

Portanto, pode-se dizer que, de modo geral, os modelos de regressão quantílica com pesos variáveis mostraram resultados superiores na estimação da previsão da volatilidade. Em relação à segregação das estimações por regressão quantílica, foi admitido que pudessem existir coeficientes fixos (RQF), que estruturaram os modelos robustos a preditores fracos, e os modelos com coeficientes variáveis (RQC). Os resultados encontrados, de modo geral, não mostraram que os modelos robustos a preditores fracos apresentavam as melhores previsões. Isso pode ocorrer em função da existência de preditores fortes ao ponto de anular o efeito da retirada dos preditores fracos, como foi observado no estudo de Lima e Meng (2017), ao estudar o prêmio pelo risco. Desse modo, a incorporação de preditores fortes e de preditores parcialmente fracos, como o sentimento textual, ampliou a capacidade de previsão da volatilidade.

4.6.3 Síntese dos Resultados da Previsão de Volatilidade

A economia de países emergentes, como o Brasil, está vulnerável ao comportamento do mercado financeiro de países desenvolvidos, assim como se mostra sensível a choques de notícias e às instabilidades políticas. Nesse sentido, Koren e Tenreyro (2007) refletem que a

volatilidade está interligada ao desenvolvimento de um país, de modo, que países emergentes tendem a apresentar maiores volatilidades no crescimento e no mercado acionário, em comparação aos países desenvolvidos.

Em um estudo sobre a volatilidade desenvolvido por Bloom (2014), são levantadas evidências de que os países da América do Sul e da África apresentam uma incidência de alta volatilidade no mercado acionário, decorrente desses países estarem vulneráveis a uma maior incerteza provocada por choques de más notícias, do âmbito local e internacional.

Considerando que os mercados acionários tendem a apresentarem uma maior volatilidade diante do choque de notícias (BLOOM, 2009; CHULIÁ et al., 2017), neste estudo, admitiu-se que a volatilidade do mercado brasileiro sofre a influência dos choques de notícias locais, como notícias sobre mudanças de políticas econômicas, e é sensível, ao comportamento do mercado acionário dos Estados Unidos, conforme foi ressaltado no estudo desenvolvido por Costa (2014).

Diante disso, foram utilizadas *proxies* para captar o efeito das notícias sobre o mercado brasileiro. Assim, considerou-se como *proxies* de notícias locais o índice de sentimento textual, baseado em notícias financeiras brasileiras, e como *proxies* do comportamento do mercado americanos, foram utilizados o índice de volatilidade *VIX*, o índice *S&P500* e o índice *DJIA*.

A volatilidade não é uma variável diretamente observável, em função disso, foi utilizada a variância condicional dos retornos do IBOVESPA, obtida mediante a modelagem GARCH (1,1), como uma *proxy* para a volatilidade realizada do mercado acionário brasileiro. Neste estudo, não existiu a intenção de discutir melhores instrumentos para estimar e prever a volatilidade, mas sim investigar o valor informacional do sentimento textual sobre a previsão da volatilidade.

Assim, mediante a utilização do método de regressão quantílica ℓ_1 -penalizada foi investigado o poder preditivo de variáveis que são citadas na literatura como possíveis preditores da volatilidade, com o objetivo de incorporar aos modelos previsão da volatilidade, variáveis que, de fato, contribuíssem com a sua previsão, visto que previsões multivariadas com previsores fracos levam a previsões imprecisas da volatilidade.

Os resultados levantados mostraram que o sentimento textual, de fato, é um preditor da volatilidade, visto que, as notícias dos jornais estão interligadas aos movimentos dos preços no mercado, conforme as evidências empíricas encontradas na literatura (TETLOCK, 2007;

TETLOCK et al., 2008; GARCÍA, 2013; LIU, 2014; FERGUSON et al., 2015; BIANCONI et al., 2015; SHAPIRO et al., 2017). Além disso, os resultados corroboram com a visão de Biancone et al. (2015) e Manela et al. (2017), de que a construção de variáveis baseadas em palavras positivas e negativas oriundas de jornais, podem se configurar como um importante preditor da volatilidade do mercado.

A importância do sentimento textual na previsão da volatilidade se concentrou na extremidade direita da distribuição da volatilidade do IBOVESPA, levantando indícios de que a informação dos jornais é importante para os investidores quando o mercado está mais volátil. Esse resultado é consistente com as discussões empíricas que apontam que o risco do mercado apresenta uma relação com o sentimento (BROWN, 1999; CAMPOS; CORTAZAR; REYES, 2017; CAPORIN; POLI, 2017) e que o sentimento das notícias é importante em períodos de maior incerteza para os investidores, como crises (LIU, 2014) e recessões econômicas (GARCÍA, 2013).

Analisando a contribuição do sentimento textual, em termos de erros de previsão fora da amostra, constatou-se que a variável relacionada ao sentimento, sozinha, não conseguiu realizar previsões ao longo do tempo que superassem a média histórica da volatilidade. Desse modo, o sentimento, como única variável de predição, não foi consistente na previsão da volatilidade, indo de encontro às evidências do estudo realizado por Manela et al. (2017), no qual foi verificado que o comportamento da volatilidade realizada do mercado acionário dos Estados Unidos poderia ser previsto por uma variável baseada em palavras de jornais.

No entanto, ao integrar o sentimento textual à outras variáveis preditoras da volatilidade, e ao considerar o poder preditivo do sentimento por quantis, determinado por LASSO, observam-se melhorias no desempenho dos modelos de previsão da volatilidade, em comparação aos erros de previsão de um modelo baseado na média histórica da volatilidade, tanto nas estimações realizadas com instrumentos robustos a valores extremos, que foram estimações feitas por regressão quantílica, quanto nas inferências médias, realizadas por OLS. Assim, os resultados dos modelos levaram à constatação de que o sentimento textual das notícias financeiras, atrelado aos índices financeiros internacionais, minimizam os erros de previsão da volatilidade, comportamento que foi observando tanto na amostra geral, quanto na segregação da amostra em dois subperíodos. Esses resultados corroboram com os achados de Caporin e Poli (2017), que, ao analisarem a previsão da volatilidade de empresas, com a

utilização da metodologia de LASSO para escolha dos preditores, verificaram que os modelos de previsão de volatilidade que incorporaram *proxies* baseadas em notícias, e conseguiram melhorar substancialmente a previsão da volatilidade.

Portanto, diante das evidências encontradas na seleção das variáveis preditoras da volatilidade por regressão quantílica ℓ_1 -penalizada – LASSO, e mediante os resultados levantados para os modelos de previsibilidade da volatilidade, a hipótese de pesquisa de que o sentimento textual das notícias financeiras contribui com a previsão da volatilidade do retorno no mercado acionário brasileiro (H7), não pode ser rejeitada, uma vez que os modelos condicionais de previsão apresentam evidências de que o sentimento textual contribui com melhorias na previsão da volatilidade analisadas neste estudo.

5 CONSIDERAÇÕES FINAIS

Neste trabalho foi investigado o efeito do sentimento textual das notícias financeiras sobre o comportamento dos preços no mercado acionário brasileiro, buscando respaldo teórico no Teorema de Concordância, no Viés da Negatividade e na Hipótese do Mercado Eficiente.

A pesquisa partiu da premissa de que as notícias, de um modo geral, contribuem com a atualização das crenças dos investidores, possibilitando um alinhamento das expectativas sobre os ativos, que, em média, é um comportamento refletido no mercado, no processo de ajuste dos preços às informações. No entanto, os investidores estão vulneráveis aos desvios cognitivos, que limitam a sua racionalidade, levando o ajuste das crenças a não seguirem a prescrição do modelo racional. Com isso, a atualização dos preços, geralmente, apresenta ruídos, que levam às ineficiências temporais no mercado. Em função disso, os preços tendem a não incorporarem as informações disponíveis, sendo possível que a compreensão dos investidores a respeito do sentimento textual das notícias possa provocar reflexos nos movimentos do mercado.

Assim, admitindo que as crenças dos investidores sejam atualizadas mediante o acesso à informação, conforme o teorema da concordância (AUMANN, 1976), partindo das evidências empíricas sobre a importância do sentimento textual para os investidores em períodos de recessões econômicas (GARCÍA, 2013), em momentos de crises financeiras (LIU, 2014) e sobre o risco extremo (BIANCONI et al., 2015), e levando em consideração que países emergentes, naturalmente, apresentam uma maior incerteza econômica (KOREN; TENREYRO, 2007; BLOOM, 2014), a hipótese central deste trabalho é que os movimentos dos preços das ações são explicados pelas notícias em períodos de maior incerteza econômica.

Para analisar o reflexo do tom das notícias sobre o comportamento de oscilação dos preços no mercado brasileiro, foi verificada a influência que o sentimento textual das notícias exerce sobre o retorno do índice IBOVESPA e sobre o risco (volatilidade do retorno), bem como analisado o poder de previsão do risco do mercado, com a incorporação do sentimento das notícias.

A primeira etapa da análise consistiu na estruturação de uma lista de termos financeiros, para possibilitar a estimação do sentimento textual das notícias financeiras escritas em língua portuguesa. A lista de palavras foi estruturada com embasamento em uma amostra de 121.435 textos jornalísticos escritos entre os anos de 1975 a 2017, da qual foram extraídos os principais termos que apresentavam conteúdo semântico para análise de discussões financeiras, com base

na Lei de Zipf. A lista proposta neste estudo apresentou um total de 4.605 palavras, dentre as quais, a maior parcela de termos foi formada por palavras de origem negativa (3.152 palavras), visto que, foi verificado que a organização do vocabulário está, naturalmente, sujeita a um viés de negatividade decorrente da necessidade de redução de incertezas. Comportamento que também foi observado na estruturação de outras listas de termos financeiros, como as de Loughran e McDonald (2011), e Aguiar (2012).

A segunda etapa do estudo consistiu na investigação sobre a influência do sentimento textual sobre o risco e o retorno do mercado. As principais evidências empíricas sobre essa discussão mostraram que o tom das notícias financeiras provoca, sempre, algum tipo de reação no mercado. Nas investigações sobre o poder do sentimento de otimismo e pessimismo, foi verificado que uma alteração no sentimento de pessimismo da mídia, nos períodos em que o mercado apresentava um baixo desempenho (q.05), chegava a levar o retorno do IBOVESPA à uma queda de -28.72 pontos-base, enquanto, nos períodos de mídia otimista, o sentimento de otimismo levava a aumentos de 22 pontos-base sobre o retorno do mercado (q.05). Além disso, verificou-se que a importância do sentimento de pessimismo da mídia exercido sobre o mercado brasileiro era mais evidente sobre os períodos de desempenho ruim (baixos retornos), no entanto, notou-se que os sentimentos avaliados, sentimentos de pessimismo e otimismo, apresentaram valor informacional capaz de prever os movimentos de retorno do mercado, contrariando os indícios de que a mídia de países emergentes não seria sofisticada ao ponto de provocar reações significativas no mercado (GRIFFIN, HIRSCHHEY; KELLY, 2011; FRAIBERGER, 2016).

Em relação à volatilidade do retorno do mercado, verificou-se que a influência do pessimismo sobre a volatilidade era maior, em termos de magnitude, sobre os períodos de turbulência do mercado (q.95), no entanto, o pessimismo da mídia corrobora com a redução da volatilidade, e não com o seu aumento, conforme era esperado. Nas evidências sobre o sentimento de otimismo é que foram observados indícios de que o sentimento está relacionado ao aumento da volatilidade em períodos de calma no mercado (baixa volatilidade – q.05), chegando a provocar impactos de 0.22 pontos-base. Diante dos resultados levantados, concluiu-se que o sentimento de pessimismo reduz a volatilidade à medida que aumenta a turbulência do mercado.

As evidências levantadas sobre o reflexo do sentimento de pessimismo das notícias sobre os períodos de incerteza (classificadas de acordo com o índice IIE-Br) apresentaram indícios de que em períodos de alta incerteza, o sentimento de pessimismo da mídia induz a reduções nos retornos, nas circunstâncias de melhor ou pior desempenho do mercado, visto que foram observadas reduções de -76.21 pontos-base sobre os menores retornos do mercado (q.05), e reduções de -81.94 pontos-base (q.95) sobre os maiores retornos do mercado (melhor desempenho). Na existência de uma baixa incerteza, a evidência mais expressiva é observada sobre os baixos retornos do mercado, em que o sentimento de pessimismo leva a uma redução de -34.44 pontos-base (q.05). Com isso, conclui-se que o impacto do sentimento de pessimismo é mais evidente sobre os períodos de baixo desempenho do mercado, e com isso, é possível que a mídia, em termos informacionais, apresente uma maior importância para o mercado nos períodos de alta incerteza.

Em relação à volatilidade, as evidências apontam que o pessimismo da mídia apresenta impactos mais significativos quando o mercado apresenta certa turbulência (q.75). No entanto, os impactos provocados levam a uma redução da volatilidade em -7.71 pontos-base, contrariando as evidências de que aumentos no pessimismo da mídia estariam correlacionados positivamente a aumentos na volatilidade do mercado (KOTHARI et al. 2009). Além disso, foi verificado que o impacto do pessimismo sobre a volatilidade é mais persistente nos períodos de turbulência do mercado, quando existem evidências de uma baixa incerteza econômica. Assim, conclui-se que no mercado brasileiro existem indícios do sentimento do pessimismo contribuir com a redução da volatilidade do mercado.

Por fim, a terceira parte da análise consistiu em uma investigação sobre o poder de previsão do risco do mercado, com a incorporação da informação relacionada ao sentimento das notícias. A escolha das variáveis preditoras da volatilidade foi realizada pelo método de regressão quantílica ℓ_1 -penalizada - LASSO, o qual identificou que o sentimento textual das notícias é um preditor da volatilidade, porém configura-se como parcialmente fraco, já que a importância do sentimento textual na previsão, concentrou-se na extremidade direita da distribuição da volatilidade do IBOVSPA, apresentando uma frequência de escolha de 22.78% no no 6º decil (q.60). Esses resultados sugerem indícios de que a informação dos jornais é importante para os investidores quando o mercado está mais volátil.

Ao ser analisada a contribuição do sentimento textual, em termos de erros de previsão fora da amostra, constatou-se que a variável do sentimento textual, sozinha, não conseguiu realizar previsões ao longo do tempo que superassem a média histórica da volatilidade, porém, ao integrar o sentimento textual à outras variáveis preditoras da volatilidade, respeitando o poder de predição das variáveis, determinado por LASSO, foram observadas melhorias no desempenho dos modelos de previsão da volatilidade, em comparação aos erros de previsão de um modelo baseado na média histórica da volatilidade. Diante dos resultados, chegou-se a constatação de que o sentimento textual contribui com melhorias na previsão da volatilidade analisadas neste estudo.

Com base nos resultados levantados sobre o impacto do sentimento textual das notícias financeiras sobre os movimentos dos preços no mercado, e, diante das evidências sobre a previsão do risco do mercado, com a incorporação do sentimento das notícias, conclui-se que, as informações provenientes dos jornais afetam a percepção dos investidores nos momentos que existe uma maior incerteza no mercado e na economia. Portanto, não se rejeita a hipótese, central do estudo, de que, os movimentos dos preços das ações são explicados pelas notícias em períodos de maior incerteza econômica.

Diante das constatações apresentadas, observa-se que existe uma maximização da utilidade dos indivíduos no processo de comunicação, visto que as palavras tendem a ser repetidas quando são importantes, ou seja, o procedimento de escolha das palavras por um indivíduo não é aleatório, e tende a mudar ao longo do tempo. Assim, ao ser identificado o conjunto de palavras que apresentam conteúdo semântico em uma determinada área de discussão, é possível extrair informações provenientes dos textos. Esse comportamento foi observado nos textos jornalísticos da mídia brasileira, permitindo levantar evidências de que a mídia de um país emergente apresenta uma importância na discussão sobre o que movem os preços no mercado acionário.

Portanto, acredita-se que o trabalho contribuiu com a discussão do papel da mídia em países emergentes, levantando evidências de que o sentimento das notícias apresenta conteúdo informacional capaz de explicar uma parcela das oscilações dos preços no mercado. Desse modo, o sentimento textual pode contribuir em estratégias de investimentos, atuando no gerenciamento da exposição ao risco.

Cabe destacar que os resultados obtidos, neste estudo, restringiram-se ao comportamento do índice IBOVESPA, no período de Julho de 2011 a Junho de 2017. Desse modo, as conclusões ficam restritas à amostra utilizada para o Brasil. Sugere-se que trabalhos futuros investiguem o poder informacional das notícias sobre os movimentos dos preços dos ativos financeiros das empresas, com a utilização do sentimento textual e com outras *proxies* baseadas em informações textuais.

REFERÊNCIAS

- AGUIAR, Marcelo Otone. **Sentiment analysis em relatórios da administração divulgados por firmas brasileiras.** 2012. 92 f. Dissertação (Mestrado em Administração de Empresas)-Fundação Instituto Capixaba de Pesquisas em Contabilidade, Economia e Finanças, Vitória, 2012.
- AHERN, Kenneth R.; SOSYURA, Denis. **Rumor Has It: Sensationalism in Financial Media.** *Review of Financial Studies*, v. 28, n 7, p. 2050-2093, jul. 2015. DOI:10.1093/rfs/hhv006.
- AHMAD, K. et al.. **Media-expressed negative tone and firm-level stock returns.** *Journal of Corporate Finance*, v. 37, p. 152-172, 2016.
- ALLEN, David E.; MCALEER, Michael; SINGH, Abhay K. **An entropy-based analysis of the relationship between the DOW JONES Index and the TRNA Sentiment series.** *Applied Economics*, v. 49, n. 7, p. 677-692, 2017.
- ALEXOPOULOS, Michelle; COHEN, Jon. **The power of print: Uncertainty shocks, markets, and the economy.** *International Review of Economics & Finance*, v. 40, p. 8-28, 2015.
- ALVES, Hans; KOCH, Alex; UNKELBACH, Christian. **Why Good Is More Alike Than Bad: Processing Implications.** *Trends in Cognitive Sciences*, 2017.
- ANDERSEN, T. G; BOLLERSLEV, T; DIEBOLD, F. X; VEGA, C. **Micro effects of macro announcements: Real-time price discovery in foreign exchange.** *The American economic review*, v. 93, n. 1, p. 38-62, 2003.
- ANTWEILER, Werner; FRANK, Murray Z. Is all that talk just noise? **The information content of internet stock message boards.** *The Journal of Finance*, v. 59, n. 3, p. 1259-1294, 2004.
- AUMANN, Robert J. **Agreeing to disagree.** *The annals of statistics*, p. 1236-1239, 1976.
- BACHELIER, Louis. **Théorie de la spéculation.** [S.l.]: Gauthier-Villars, 1900. p. 21-86. Traduzido em inglês por D. May *Annales scientifiques de l'École Normale Supérieure*.
- BALCILAR, M., Gupta, R., Kyei, C., & Wohar, M. E. **Does economic policy uncertainty predict exchange rate returns and volatility? Evidence from a nonparametric causality-in-quantiles test.** *Open Economies Review*, v. 27, n. 2, p. 229-250, 2016.
- BARBOSA, Glauber de Castro; MEDEIROS, Otávio Ribeiro de. **Teste empírico da eficiência do mercado brasileiro na ocorrência de eventos favoráveis e desfavoráveis.** *Revista de Negócios*, Blumenau, v. 12, n. 4, p. 44-54, out/dez, 2007.
- BAUR, Dirk G.; DIMPFL, Thomas; JUNG, Robert C. **Stock return autocorrelations revisited: A quantile regression approach.** *Journal of Empirical Finance*, v. 19, n. 2, p. 254-265, 2012.

BAUMEISTER, R. F., BRATSLAVSKY, E., FINKENAUER, C., & Vohs, K. D. (2001). **Bad is stronger than good**. *Review of general psychology*, 5(4), 323.

BAKER, Scott R.; BLOOM, Nicholas; DAVIS, Steven J. **Measuring economic policy uncertainty**. *The Quarterly Journal of Economics*, v. 131, n. 4, p. 1593-1636, 2016.

BEATTIE, V. **Accounting Narratives and the Narrative Turn in Accounting Research: Issues, Theory, Methodology, Methods and a Research Framework**. *The British Accounting Review*, 2014.

BELLONI, Alexandre; CHERNOZHUKOV, Victor. **ℓ_1 -penalized quantile regression in high-dimensional sparse models**. *The Annals of Statistics*, v. 39, n. 1, p. 82-130, 2011.

BIANCONI, Marcelo; HUA, Xiabin; TAN, Chih Ming. **Determinants of systemic risk and information dissemination**. *International Review of Economics & Finance*, v. 38, p. 352-368, 2015.

BIBLIOTECA DIGITAL DO SENADO FEDERAL DO BRASIL. **Acervo de jornais**. Portal O Senado. Recuperado em 17 de Maio, 2017, de <http://www2.senado.leg.br/bdsf/handle/id/5>

BLOOM, Nicholas. **The impact of uncertainty shocks**. *Econometrica*, v. 77, n. 3, p. 623-685, 2009.

BLOOM, N. (2014). **Fluctuations in uncertainty**. *The Journal of Economic Perspectives*, 28(2), 153-175.

BOLLEN, Johan; MAO, Huina; ZENG, Xiaojun. **Twitter mood predicts the stock market**. *Journal of Computational Science*, v. 2, n. 1, p. 1-8, 2011.

BONE, R. B.; RIBEIRO, E. P. **Eficiência fraca, Efeito dia-da-semana e efeito feriado no mercado acionário brasileiro: Uma Análise Empírica Sistemática e Robusta**, *Revista de Administração Contemporânea*. Rio de Janeiro: ANPAD, v.6, jan/abr, 2002.

BOOTH, Andrew D. A **“Law” of occurrences for words of low frequency**. *Information and control*, v. 10, n. 4, p. 386-393, 1967.

BOLLERSLEV, T. **Generalized autoregressive conditional heteroscedasticity**. *Journal of Econometrics*, 31:307-27, 1986.

BROWN, Gregory W. **Volatility, sentiment, and noise traders**. *Financial Analysts Journal*, v. 55, n. 2, p. 82-90, 1999.

BROOKS, Chris. **Introductory econometrics for finance**. Cambridge university press, 2014.

BUEHLMAYER, Matthias M. M. **The role of the media in takeovers: Theory and evidence**. 2013. DOI: <http://dx.doi.org/10.2139/ssrn.1673164>

CAMPBELL, Dennis; LOUMIOTI, Maria; MOERMAN W., Regina. **Making Sense of Soft Information: Interpretation Bias and Ex-Post Lending Outcomes (March 2017)**.

Available at SSRN: <https://ssrn.com/abstract=2939881>

CAMPOS, I.; CORTAZAR, G.; REYES, T. **Modeling and predicting oil VIX: Internet search volume versus traditional variables**. *Energy Economics*, v. 66, p. 194-204, 2017.

CAPORALE, Guglielmo Maria; SPAGNOLO, Fabio; SPAGNOLO, Nicola. **Macro news and stock returns in the Euro area: A VAR-GARCH-in-mean analysis**. *International Review of Financial Analysis*, v. 45, p. 180-188, 2016.

CAPORIN, Massimiliano; POLI, Francesco. **Building News Measures from Textual Data and an Application to Volatility Forecasting**. *Econometrics*, v. 5, n. 3, p. 35, 2017.

CARRETTA, A. et al., **Does investor attention influence stock market activity? The case of spin-off deals**. In: *Asset Pricing, Real Estate and Public Finance over the Crisis*. Palgrave Macmillan UK, 2013. p. 7-24.

CERETTA, Paulo Sergio; VIEIRA, Kelmara Mendes. **Efeito dia-da-semana no mercado brasileiro: uma análise sob a ótica da liquidez, do retorno e da volatilidade**. *REAd-Revista Eletrônica de Administração*, v. 16, n. 3, p. 544-563, 2010.

CHISHOLM, Erica; KOLDA, Tamara G. **New term weighting formulas for the vector space method in information retrieval**. Computer Science and Mathematics Division, Oak Ridge National Laboratory, 1999.

CLARK, Todd E.; WEST, Kenneth D. **Approximately normal tests for equal predictive accuracy in nested models**. *Journal of econometrics*, v. 138, n. 1, p. 291-311, 2007.

COSTA FILHO, A. E. D. (2014). **Incerteza e atividade econômica no Brasil**. *Economia Aplicada*, 18(3), 421-453.

CHULIÁ, Helena; GUILLÉN, Montserrat; URIBE, Jorge M. **Measuring uncertainty in the stock market**. *International Review of Economics & Finance*, v. 48, p. 18-33, 2017.

COX, R. A., DAYANANDAN, A., DONKER, H., & Nofsinger, J. (2017). **The Bad, the boom and the bust: Profit warnings over the business cycle**. *Journal of Economics and Business*, 89, 13-19.

CUTLER, David M.; POTERBA, James M.; SUMMERS, Lawrence H. **What moves stock prices?**. *The Journal of Portfolio Management*, v. 15, n. 3, p. 4-12, 1989.

CUNHA, Antônio Geraldo da. **Dicionário etimológico da língua portuguesa**. 4 ed. Rio de Janeiro: Lexikon, 2010.

DAVIS, Angela K.; PIGER, Jeremy M.; SEDOR, Lisa M. **Beyond the Numbers: Measuring the Information Content of Earnings Press Release Language***. *Contemporary Accounting Research*, v. 29, n. 3, p. 845-868, 2012.

DIEBOLD, Francis X.; MARIANO, Roberto S. **Comparing Predictive Accuracy**. Journal of Business & Economic Statistics, v. 13, n. 3, p. 253-263, 1995.

DOUGAL, C. et al.. **Journalists and the stock market**. Review of Financial Studies, p. hhr133, 2012.

ESTADÃO. **O estado de São Paulo: história**. Disponível em: <http://www.estadao.com.br/historico/print/resumo.htm> Acesso em: Set 2017.

ENGELBERG, Joseph. **Costly information processing: Evidence from earnings announcements**. In: AFA 2009 San Francisco Meetings Paper. 2008. DOI: <http://dx.doi.org/10.2139/ssrn.1107998>.

ENGELBERG, Joseph E.; PARSONS, Christopher A. **The causal impact of media in financial markets**. The Journal of Finance, v. 66, n. 1, p. 67-97, 2011.

ENGELBERG, Joseph E.; REED, Adam V.; RINGGENBERG, Matthew C. **How are shorts informed? Short sellers, news, and information processing**. Journal of Financial Economics, v. 105, n. 2, p. 260-278, 2012.

FAMA, E. F. **Efficient markets: a review of theory and empirical work**. The Journal of Finance, v. 25, n. 2, p. 383-417, 1970.

FAMA, Eugene F. **Efficient capital markets: II**. The journal of Finance, v. 46, n. 5, p. 1575-1617, 1991.

FAMA, Eugene F. **The behavior of stock-market prices**. Journal of Business, p. 34-105, 1965.

FAMA, E. F.; FRENCH, K. R. **Common risk factors in the returns on stocks and bonds**. Journal of Financial Economics, v.33, n.1, p. 3-56, 1993.

FAMA, Eugene F. **Market efficiency, long-term returns, and behavioral finance**. Journal of financial economics, v. 49, n. 3, p. 283-306, 1998.

FENG, Xunan; HU, Na. **Are individual investors affected by attention?: Evidence from the earning announcement effect in China**. China Finance Review International, v. 4, n. 3, p. 289-304, 2014. DOI: <http://dx.doi.org/10.1108/CFRI-09-2013-0114>.

FERGUSON, N. J., PHILIP, D., LAM, H. Y., & GUO, J. M. **Media content and stock returns: The predictive power of press**. Multinational Finance Journal, 19(1), 1-31, 2015.

FISHER, Ingrid E.; GARNSEY, Margaret R.; HUGHES, Mark E. **Natural Language Processing in Accounting, Auditing and Finance: A Synthesis of the Literature with a Roadmap for Future Research**. Intelligent Systems in Accounting, Finance and Management, 2016.

FRAIBERGER, Samuel P. **News Sentiment and Cross-Country Fluctuations**. Available at SSRN, 2016.

FRENCH, K R. **The Stock returns and the weekend effect**. Journal Finance of Economics, Amsterdam, v.8, n.1 p. 55-69, Mar. 1980.

GAY, Robert D. **Effect Of Macroeconomic Variables On Stock Market Returns For Four Emerging Economies: Brazil, Russia, India, And China**. The International Business & Economics Research Journal (Online), v. 15, n. 3, p. 119, 2016.

GARCÍA, Diego. **Sentiment during recessions**. The Journal of Finance, v. 68, n. 3, p. 1267-1300, 2013.

GRIFFIN, John M.; HIRSCHEY, Nicholas H.; KELLY, Patrick J. **How important is the financial media in global markets?**. Review of Financial Studies, p. hhr099, 2011.

GROSSMAN, S.; STIGLITZ, J. E. **On the impossibility of informationally efficient markets**. American Economic Review, v.70, n. 3, p. 393-408, jun. 1980.

GODEIRO, Lucas Lúcio; LIMA, Luiz Renato Régis de Oliveira. **Medindo incerteza macroeconômica para o Brasil**. Economia Aplicada, v. 21, n. 2, p. 311, 2017.

GONZALEZ-RIVERA, Gloria. **Forecasting for economics and business**. Routledge, 2016.

HENDERSHOTT, Terrence; LIVDAN, Dmitry; SCHÜRHOFF, Norman. **Are institutions informed about news?**. Journal of Financial Economics, v. 117, n 2, 2015.

HIRSHLEIFER, David; TEOH, Siew Hong. **Limited attention, information disclosure, and financial reporting**. Journal of accounting and economics, v. 36, n. 1, p. 337-386, 2003.

HONG, Harrison; TOROUS, Walter; VALKANOV, Rossen. **Do industries lead stock markets?**. Journal of Financial Economics, v. 83, n. 2, p. 367-396, 2007.

HU, Bill; MCINISH, Thomas; ZENG, Li. **Gambling in penny stocks: the case of stock spam e-mails**. International Journal of Cyber Criminology, v. 4, n. 1/2, p. 610, 2010.

HUANG, Allen H.; ZANG, Amy Y.; ZHENG, Rong. **Evidence on the information content of text in analyst reports**. The Accounting Review, v. 89, n. 6, p. 2151-2180, 2014.

HUANG, Xuan; TEOH, Siew Hong; ZHANG, Yinglei. **Tone Management**. The Accounting Review, v.89, nº3. p.1083-1113, 2014.

INSTITUTO BRASILEIRO DE ECONOMIA – IBRE. **Indicador de Incerteza da Economia Brasil**. 2016. Disponível em: <http://portal.fgv.br/noticias/novo-indicador-mensura-incerteza-economia-brasileira> Acesso em 06 de outubro de 2017.

INSTITUTO VERIFICADOR DE CIRCULAÇÃO - IVC. **Dados oficiais do IVC relacionados a circulação de jornais impressos e digitais de Fevereiro 2017**. Disponível

em: <https://www.poder360.com.br/midia/jornais-e-revistas-continuam-avancando-em-2017-com-suas-edicoes-digitais/> Acesso: Set 2017.

INSTITUTO DE PESQUISA ECONÔMICA APLICADA - IPEADATA. **Série temporal do Índice de Preço ao Consumidor Ampliado**. 2017. Disponível em: <http://www.ipeadata.gov.br/Default.aspx> Acesso: Jul 2017.

JURADO, Kyle; LUDVIGSON, Sydney C.; NG, Serena. **Measuring uncertainty**. *The American Economic Review*, v. 105, n. 3, p. 1177-1216, 2015.

JANDL, Jan-Otto. **Information Processing in Securitized Real Estate Markets--How Newspaper Content and Online Search Behavior Help Explaining Market Movements**. In: 2016 49th Hawaii International Conference on System Sciences (HICSS). IEEE, 2016. p. 1092-1101.

JEGADEESH, Narasimhan; TITMAN, Sheridan. **Returns to buying winners and selling losers: Implications for stock market efficiency**. *The Journal of finance*, v. 48, n. 1, p. 65-91, 1993.

JEGADEESH, Narasimhan; WU, Di. **Word power: A new approach for content analysis**. *Journal of Financial Economics*, v. 110, n. 3, p. 712-729, 2013.

KAHNEMAN, Daniel; TVERSKY, Amos. **Prospect theory: An analysis of decision under risk**. *Econometrica*, v. 47, n. 2, p. 263-291, 1979. DOI: 10.2307/1914185

KARABULUT, Yigitcan. **Can Facebook predict stock market activity?**. In: AFA 2013 San Diego Meetings Paper. 2013.

KEARNEY, Colm; LIU, Sha. **Textual sentiment in finance: A survey of methods and models**. *International Review of Financial Analysis*, v. 33, p. 171-185, 2014.

KEYNES, John M. **The General Theory of Employment, Interest, and Money**. Londres: MacMillan, 1936.

KHUU, Joyce; DURAND, Robert B.; SMALES, Lee A. **Melancholia and Japanese stock returns, 2003 to 2012**. *Pacific-Basin Finance Journal*, 2016.

KOTHARI, S. P.; LI, Xu; SHORT, James E. **The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis**. *The Accounting Review*, v. 84, n. 5, p. 1639-1670, 2009.

KNIGHT, Frank H. **Risk, Uncertainty and profit**. 1921. Courier Corporation, 2012.

KOREN, Miklós; TENREYRO, Silvana. **Volatility and development**. *The Quarterly Journal of Economics*, v. 122, n. 1, p. 243-287, 2007.

KOENKER, Roger; XIAO, Zhijie. **Quantile autoregression**. *Journal of the American Statistical Association*, v. 101, n. 475, p. 980-990, 2006.

KOENKER, R., and A. Machado. 1999. **Goodness of fit and related processes for quantile regression**. *Journal of American Statistical Association* 94: 1296–1310.

KOENKER, R. **Quantile Regression**. Cambridge: Cambridge University Press, 2005.

KWAPIEŃ, Jarosław; DROŹDŹ, Stanisław. **Physical approach to complex systems**. *Physics Reports*, v. 515, n. 3, p. 115-226, 2012.

LINGUEE. **Dicionário Inglês-Português**. Disponível em: <http://www.linguee.com/english-portuguese> Acesso em: Set 2017.

LEHAVY, Reuven; LI, Feng; MERKLEY, Kenneth. **The effect of annual report readability on analyst following and the properties of their earnings forecasts**. *The Accounting Review*, v. 86, n. 3, p. 1087-1115, 2011.

LI, Feng. **Survey of the Literature**. *Journal of accounting literature*, v. 29, p. 143-165, 2010a.

LI, Feng. **The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach**. *Journal of Accounting Research*, v. 48, n. 5, p. 1049-1102, 2010b. DOI:10.1111/j.1475-679X.2010.00382.x

LIMA, Luiz Renato; MENG, Fanning. **Out - of - Sample Return Predictability: A Quantile Combination Approach**. *Journal of Applied Econometrics*, v. 32, n. 4, p. 877-895, 2017.

LIU, Baixiao; MCCONNELL, John J. **The role of the media in corporate governance: Do the media influence managers' capital allocation decisions?**. *Journal of Financial Economics*, v. 110, n. 1, p. 1-17, 2013.

LIU, Baixiao; MCCONNELL, John J.; XU, Wei. **The power of the pen reconsidered: The media, CEO human capital, and corporate governance**. *Journal of Banking & Finance*, v. 76, p. 175-188, 2017.

LIU, Sha. **The impact of textual sentiment on sovereign bond yield spreads: Evidence from the Eurozone crisis**. *Multinational Finance Journal*, Forthcoming, v. 18, n. 3/4, p. 215–248, 2014.

LOUGHRAN, T; McDONALD, B. When Is a Liability Not a Liability? **Textual Analysis, Dictionaries**, and 10-Ks. *The Journal of Finance*, v. 66, n. 1, p. 35-65, 2011. DOI:10.1111/j.1540-6261.2010.01625.x

_____; _____. **Textual analysis in accounting and finance: A survey**. *Journal of Accounting Research*, 2016.

MANELA, Asaf; MOREIRA, Alan. **News implied volatility and disaster concerns**. *Journal of Financial Economics*, v. 123, n. 1, p. 137-162, 2017.

MARINS, Jaqueline Terra Moura; VICENTE, José Valentim Machado. **Do the central bank actions reduce interest rate volatility?**. *Economic Modelling*, 2017.

MELIGKOTSIDOU, Loukia et al.. **A quantile regression approach to equity premium prediction**. *Journal of Forecasting*, v. 33, n. 7, p. 558-576, 2014.

MILGROM, Paul; STOKEY, Nancy. **Information, trade and common knowledge**. *Journal of Economic Theory*, v. 26, n. 1, p. 17-27, 1982.

MOSKOWITZ, Tobias J.; GRINBLATT, Mark. **Do industries explain momentum?**. *The Journal of Finance*, v. 54, n. 4, p. 1249-1290, 1999.

MÜLLER, Fernanda Maria; RIGHI, Marcelo Brutti; CERETTA, Paulo Sergio. **IBOVESPA Market efficiency analysis: A quantile autoregressive model approach**. *BASE-Revista de Administração e Contabilidade da Unisinos*, v. 12, n. 2, p. 122-134, 2015.

MICHAELIS. **Dicionário Brasileiro da Língua Portuguesa on-line**. Disponível em: <http://michaelis.uol.com.br> Acesso em: Set 2017.

OGLOBO. **História Grupo Globo**. Disponível em: <http://historiagrupoglobo.globo.com/hgg/index.htm> Acesso em Set 2017.

NIEDERHOFFER, Victor. **The analysis of world events and stock prices**. *Journal of Business*, p. 193-219, 1971.

OSBORNE, M. F. Maury. **Brownian motion in the stock market**. *Operations research*, v. 7, n. 2, p. 145-173, 1959.

PAO, Miranda Lee. **Automatic text analysis based on transition phenomena of word occurrences**. *Journal of the Association for Information Science and Technology*, v. 29, n. 3, p. 121-124, 1978.

PRIBERAM. **Dicionário da Língua Portuguesa**. Disponível em: <https://www.priberam.pt> Acesso: Set 2017.

PEETERS, Guido. **The positive - negative asymmetry: On cognitive consistency and positivity bias**. *European Journal of Social Psychology*, v. 1, n. 4, p. 455-474, 1971.

PEETERS, Guido; CZAPINSKI, Janusz. **Positive-negative asymmetry in evaluations: The distinction between affective and informational negativity effects**. *European review of social psychology*, v. 1, n. 1, p. 33-60, 1990.

ROZIN, Paul; ROYZMAN, Edward B. **Negativity bias, negativity dominance, and contagion**. *Personality and social psychology review*, v. 5, n. 4, p. 296-320, 2001.

SAMUELSON, Paul A. **Proof that properly anticipated prices fluctuate randomly**. *Industrial management review*, v. 6, n. 2, p. 41-49, 1965.

SCHNUSENBERG, Oliver; MADURA, Jeff. Do U.S. **stock market indexes over - or underreact?**. *Journal of Financial Research*, v. 24, n. 2, p. 179-204, 2001. DOI: 10.1111/j.1475-6803.2001.tb00764.x

SILVA, Chrystiane. **IBOVESPA acompanha bolsas americanas e fecha semana em alta**. Disponível em: <http://www.valor.com.br/financas/5077320/IBOVESPA-acompanha-bolsas-americanas-e-fecha-semana-em-alta> Acesso em 6 de outubro de 2017.

SINHA, Nitish Ranjan. **Underreaction to news in the US stock market**. *Quarterly Journal of Finance*, v. 6, n. 02, p. 1650005, 2016.

STRAUß, Nadine; VLIEGENTHART, Rens; VERHOEVEN, Piet. **Lagging behind? Emotions in newspaper articles and stock market prices in the Netherlands**. *Public Relations Review*, 2016.

SVENSSON, H. M., ALBÆK, E., VAN DALEN, A; DE VREESE, C. **Good News in Bad News: How Negativity Enhances Economic Efficacy**. *International Journal of Communication*, 11, 17, 2017.

SHAPIRO, Adam Hale, Moritz Sudhof, and Daniel Wilson. 2017. **Measuring News Sentiment**. FRB San Francisco Working Paper 2017-01. <http://www.frbsf.org/economic-research/publications/working-papers/2017/wp2017-01.pdf>

SCOTTI, Chiara. **Surprise and uncertainty indexes: Real-time aggregation of real-activity macro-surprises**. *Journal of Monetary Economics*, v. 82, p. 1-19, 2016.

TETLOCK, Paul C. **Does public financial news resolve asymmetric information?**. *Review of Financial Studies*, v. 23, n. 9, p. 3520-3557, 2010.

_____. **All the news that's fit to reprint: Do investors react to stale information?**. *Review of Financial Studies*, v. 24, n. 5, p. 1481-1512, 2011.

_____. **Information transmission in finance**. *Annu. Rev. Financ. Econ.*, v. 6, n. 1, p. 365-384, 2014.

_____. **Giving content to investor sentiment: The role of media in the stock market**. *Journal of Finance*, v. 62, n. 3, p. 1139-1168, 2007. DOI:10.1111/j.1540-6261.2007.01232.x

TETLOCK, P. C; SAAR-TSECHANSKY, M.; MACSKASSY, S. **More than words: Quantifying language to measure firms' fundamentals**. *Journal of Finance*, v. 63, n. 3, p. 1437-1467, 2008.

TWEDT, Brady; REES, Lynn. **Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports**. *Journal of Accounting and Public Policy*, v. 31, n. 1, p. 1-21, 2012.

VALOR ECONÔMICO. **Valor, 15 anos de bom jornalismo e reinvenções**. Disponível em: <http://www.valor.com.br/valor15anos> Acesso: Set 2017.

VALOR ECONÔMICO. **Missão Editorial**. Disponível em:
<http://midiakit.valor.com.br/index.php?p=2> Acesso em: Set 2017.b

YOUNG, Lori; SOROKA, Stuart. **Affective news: The automated coding of sentiment in political texts**. Political Communication, v. 29, n. 2, p. 205-231, 2012.

WEI, Y. C.; LU, Y. C.; CHEN, J. N.; HSU, Y. J. **Informativeness of the market news sentiment in the Taiwan stock market**. The North American Journal of Economics and Finance, 39, 158-181, 2017.

WORDNET. **WordNet a lexical database of English**. Disponível em:
<http://wordnetweb.princeton.edu/perl/webwn> Acesso em: Set 2017.

ZHANG, X. **Information uncertainty and stock returns**. The Journal of Finance, v. 61, n. 1, p. 105-137, 2006.

ZIPF, G.K. **Human Behaviour and the Principle of Least Effort**. Boston, Addison-Wesley; 1949.

APÊNDICE A – Algoritmo para contagem geral das Palavras em um repositório – Linguagem R

```
#####
##### Algoritmo para contagem geral de palavras em um texto #####
#####

## Bibliotecas Necessárias
library(tcltk)

## Diretório dos Dados
diretorio <- tclvalue(tkchooseDirectory(title="Abrir Banco de Dados"))

## Armazenamento das palavras em um texto
arquivos = dir(diretorio)
quantidade_arquivos = length(arquivos)
texto = c()

## Armazenamento das palavras de todos os textos
todos = c()

### Função de Armazenamento
for (i in 1:quantidade_arquivos) {
  caminho = paste0(diretorio,"/", arquivos[i])
  texto <- read.delim(file = caminho, header = F, sep = " ", quote = "\"", dec = ".", fill =
TRUE, comment.char = "")
  text_aux = c()
  for(j in 1:length(texto)){
    text_aux = c(text_aux,as.vector(texto[1,j],mode="any"))
  }
  texto = text_aux
  texto=ifelse(is.na(texto),"NA",texto)
  todos = c(todos, texto)
}

### Arquivamento da contagem das palavras
tabela = data.frame(table(todos))
write.csv(tabela, file = "tabela.csv")
```

APÊNDICE B - Algoritmo para contabilização de novas palavras encontradas por texto

```
#####
##### Algoritmo para contagem de palavras novas em um texto #####
#####

### Bibliotecas
library(tcltk)

### Diretório
diretorio <- tclvalue(tkchooseDirectory(title="Abrir Banco de Dados"))
arquivos = dir(diretorio)
quantidade_arquivos = length(arquivos)

### Armazenamento das palavras em textos escolhidos aleatoriamente
texto = c() #armazena as palavras de um texto
todos = c() #armazena todas as palavras de todos os textos
diferenca = c()
anterior = 0
depois = 0
ptm = proc.time()
for (i in 1:quantidade_arquivos) {
  caminho = paste0(diretorio,"/", arquivos[i])
  texto <- read.delim(file = caminho, header = F, sep = " ", quote = "\"", dec = ".", fill =
TRUE, comment.char = "")
  text_aux = c()
  for(j in 1:length(texto)){
    text_aux = c(text_aux,as.vector(texto[1,j],mode="any"))
  }
  texto = text_aux
  texto=ifelse(is.na(texto),"NA",texto)
  todos = c(todos, texto)
  antes = depois
tabela = data.frame(table(todos))
  depois = nrow(tabela)
  diferenca[i] = depois - antes
}

### Tempo de processamento
proc.time() – ptm

### Gráfico da Distribuição da Frequência das Palavras – Lei de Zipft
plot(diferenca)

### Armazenamento dos resultados
write.csv(diferenca,"palavras_novas_por_textos.csv")
```

APÊNDICE C – Algoritmo para estimação do Sentimento Textual de uma notícia

```
#####
##### Algoritmo para estimar o Sentimento Textual #####
#####

### Bibliotecas
library(tcltk)

#####
##### Diretórios #####

## Diretório – Palavras Positivas
local <- tclvalue(tkgetOpenFile(title="Abrir Banco de Dados"))
lista_positivas <- read.table(file = local, header=FALSE, dec=",")
lista_positivas = as.vector(lista_positivas$V1,mode = "any")

## Diretório – Palavras Negativas
local <- tclvalue(tkgetOpenFile(title="Abrir Banco de Dados"))
lista_negativas <- read.table(file = local, header=FALSE, dec=",")
lista_negativas = as.vector(lista_negativas$V1,mode = "any")

### Diretório de Adverbios de negação
local <- tclvalue(tkgetOpenFile(title="Abrir Banco de Dados"))
lista_adverbios <- read.table(file = local, header=FALSE, dec=",")
lista_adverbios = as.vector(lista_adverbios$V1,mode = "any")

#####
##### Leitura dos Textos #####

### Diretório
diretorio <- tclvalue(tkchooseDirectory(title="Abrir Banco de Dados"))

### Matriz com o armazenamento da frequência das palavras
arquivos = dir(diretorio)
quantidade_arquivos = length(arquivos)
texto = c()
texto_positivas = c()
texto_negativas = c()
tipo = "P"
positivas = data.frame(palavra = lista_positivas, tipo)
tipo = "N"
negativas = data.frame(palavra = lista_negativas, tipo)
positivas_negativadas = data.frame(palavra = lista_positivas, tipo)
tabelao = rbind(positivas, negativas, positivas_negativadas)
for(j in 1:quantidade_arquivos) {
  nomeArquivo = arquivos[j]
```

```

    tabelao[,nomeArquivo] = 0
}

```

```

for (i in 1:quantidade_arquivos)
{
    caminho = paste0(diretorio,"/", arquivos[i])
    texto <- read.delim(file = caminho, header = F, sep = " ", quote = "\"", dec = ".", fill =
TRUE, comment.char = "")
    text_aux = c()
    for(j in 1:length(texto)){
        text_aux = c(text_aux,as.vector(texto[1,j],mode="any"))
    }
    texto = text_aux
    texto=ifelse(is.na(texto),"NA",texto)

    lista_completa = c(lista_positivas, lista_negativas)

    for(j in 1:length(texto)) {
        if (length(subset(lista_positivas, lista_positivas == texto[j])) > 0) {
            if (j >= 2) {
                if (length(subset(lista_adverbios, lista_adverbios == texto[j-1]))
> 0 | length(subset(lista_adverbios, lista_adverbios == texto[j-2])) > 0) {
                    anterior = tabelao[tabelao$palavra == texto[j] &
tabelao$tipo == "N", arquivos[i]]
                    tabelao[tabelao$palavra == texto[j] & tabelao$tipo ==
"N", arquivos[i]] = anterior + 1
                }
            } else {
                anterior = tabelao[tabelao$palavra == texto[j] &
tabelao$tipo == "P", arquivos[i]]
                tabelao[tabelao$palavra == texto[j] & tabelao$tipo ==
"P", arquivos[i]] = anterior + 1
            }
        } else {
            anterior = tabelao[tabelao$palavra == texto[j] & tabelao$tipo ==
"P", arquivos[i]]
            tabelao[tabelao$palavra == texto[j] & tabelao$tipo == "P",
arquivos[i]] = anterior + 1
        } else {
            if (length(subset(lista_negativas, lista_negativas == texto[j])) > 0) {
                anterior = tabelao[tabelao$palavra == texto[j] &
tabelao$tipo == "N", arquivos[i]]
                tabelao[tabelao$palavra == texto[j] & tabelao$tipo ==
"N", arquivos[i]] = anterior + 1
            }
        }
    }
}

```

```

    }
  }
}

### Gerando matriz com a quantidade de textos em que uma palavra apareceu
soma = c()
t = quantidade_arquivos+2
for (i in 1:nrow(tabelao)) {
  soma[i] = 0
  for (j in 3:t) {
    if (tabelao[i,j]>0) {
      soma[i] = soma[i] + 1
    }
  }
}

### Armazenamento geral das frequências das palavras
tabelao = cbind(tabelao, soma)

#####
##### Estimação dos Pesos das Palavras #####
#Cálculo dos pesos
peso_positivo_texto = c()
peso_negativo_texto = c()

t = quantidade_arquivos+2
for (i in 1:t-2) {
  peso_positivo_texto[i] = 0
  peso_negativo_texto[i] = 0
}
peso_positivo_total = 0
pos = subset(tabelao, tipo=="P")
for(i in 1:nrow(pos)) {
  for(j in 3:t) {

    if (pos[i,j] != 0) {
      x = subset(tabelao, tabelao[, j]>0 & tipo == "P")

      num = 1 + log(pos[i,j])

      m = sum(pos[, j])/nrow(subset(pos, pos[, j] > 0 & tipo == "P"))
      den = 1 + log(m)
      W = (num) / (den) * (log(quantidade_arquivos / pos$soma[i]))

      peso_positivo_total = peso_positivo_total + W
      peso_positivo_texto[j-2] = peso_positivo_texto[j-2] + W
    }
  }
}

```

```

    }
  }
}

peso_negativo_total = 0
pos = subset(tabelao, tipo=="N")
for(i in 1:nrow(pos)) {
  for(j in 3:t) {
    if (pos[i,j] != 0) {
      num = 1 + log(pos[i,j])
      m = sum(pos[,j])/nrow(subset(pos, pos[,j] > 0))
      den = 1 + log(m)
      W = (num) / (den) * (log(quantidade_arquivos / pos$soma[i]))
      peso_negativo_total = peso_negativo_total + W
      peso_negativo_texto[j-2] = peso_negativo_texto[j-2] + W
    }
  }
}

#####
##### Estimação do Sentimento #####

#TOM = (peso_positivo_total - peso_negativo_total) / (peso_positivo_total +
peso_negativo_total)

#TOM por texto
TOM_texto = (peso_positivo_texto - peso_negativo_texto) / (peso_positivo_texto +
peso_negativo_texto)
print ("TOM_texto: ");
print(TOM_texto);
peso_negativo_texto
peso_positivo_texto

### Armazenamento dos Resultados
write.csv(TOM_texto, file = "tabela.csv")

```

APÊNDICE D – Listas das Palavras Negativas

abaixo; abaixou; abala; abalada; abaladas; abalado; abalados; abalam; abalar; abalaram; abalo; abalos; abalou; abandona; abandonada; abandonadas; abandonado; abandonados; abandonam; abandonamos; abandonando; abandonar; abandonaram; abandonarem; abandonaria; abandonasse; abandone; abandonem; abandono; abater; abateu; abatido; abdicacao; abdicado; abdicam; abdicando; abdicar; abdicaram; abdicarem; abdicou; abismo; abismos; aborrece; aborrecer; aborreceu; aborrecido; aborrecimento; aborrecimentos; abrupta; abruptamente; abruptas; abrupto; abruptos; abusa; abusada; abusadas; abusado; abusam; abusando; abusar; abusaram; abusiva; abusivamente; abusivas; abusivo; abusivos; abuso; abusos; abusou; acaba; acabada; acabado; acabados; acabam; acabando; acabar; acabaram; acabaria; acabasse; acabassem; acabava; acabavam; acabou; acarreta; acarretar; acidental; acidente; acidentes; acusa; acusacao; acusacoes; acusada; acusadas; acusado; acusador; acusadores; acusados; acusam; acusando; acusar; acusaram; acusatoria; acusatorio; acusava; acusavam; acusou; adia; adiada; adiadas; adiado; adiados; adiam; adiamento; adiamentos; adiando; adiar; adiararam; adiou; adulteracao; adulterados; adulteras; adversa; adversario; adversarios; adversas; adversidade; adversidades; adverso; adversos; adverte; advertencia; advertencias; advertia; advertida; advertidas; advertido; advertidos; advertindo; advertir; advertira; advertiram; advertiu; afasta; afastados; afastamento; afastamos; afastando; afastar; afastara; afastaram; afastarem; afastaria; afastasse; afaste; afastou; afeta; afetada; afetadas; afetado; afetados; afetam; afetando; afetar; afetara; afetaram; afetarao; afetaria; afete; afetem; afetou; aflicao; aflicoes; aflige; afligem; aflita; aflito; aflitos; afunda; afundado; afundam; fundamento; afundando; afundar; afundou; agitacao; agrava; agravada; agravadas; agravado; agravados; agravam; agravamento; agravando; agravante; agravantes; agravar; agravara; agravaram; agravaria; agrave; agravo; agravos; agravou; agredida; agredidas; agredido; agredidos; agredindo; agredir; agrediram; agrediu; agressao; agressiva; agressivamente; agressivas; agressividade; agressivo; agressivos; agressoes; agressor; agressores; agride; agridem; alarmado; alarmados; alarmante; alarmantes; alerta; alertado; alertar; alertaram; alertava; alertavam; alertou; aliciamento; ameaca; ameacada; ameacadas; ameacado; ameacador; ameacadora; ameacadoras; ameacadores; ameacados; ameacam; ameacando; ameacar; ameacaram; ameacarem; ameacaria; ameacas; ameacava; ameacavam; ameacou; angustia; angustiante; angustiantes; angustias; anomalia; anomalias; anormais; anormal; anormalidade; anormalidades; anormalmente; ansiedade; ansiedades; ansiosamente; ansiosos; anula; anulacao; anulada; anuladas; anulado; anulados; anulam; anulando; anular; anulara; anularam; apelacao; apelacoes; apelar; apelaram; apelava; apelo; apelos; apelou; apertado; apertados; apertando; apertar; apertaram; aperto; apertos; apertou; apreendem; apreendemos; apreender; apreenderam; apreendeu; apreendida; apreendidas; apreendido; apreendidos; apreensao; apreensiva; apreensivo; apreensivos; apreensoes; quem; arrisca; arriscada; arriscadas; arriscado; arriscados; arriscam; arriscando; arriscar; arriscaram; arriscaria; arrisco; arriscou; assalta; assaltada; assaltado; assaltados; assaltante; assaltantes; assaltar; assalto; assaltos; assaltou; assassinados; assassinar; assassinaram; assassinato; assassinatos; assediada; assediado; assediados; assedio; assusta; assustada; assustadas; assustado; assustador; assustadora; assustadoramente; assustadoras; assustadores; assustados; assustam; assustando; assustar; assustaram; assustava; ataca; atacada; atacadas; atacou; ataque; ataques; atentados; atingir; atingiram; atingiu; atrapalhar; atrasada; atrasadas; atrasado; atrasados; atrasam; atrasando; atrasar; atrasara; atrasaram; atraso; atrasos; atrasou; atrito; atritos; atrocidade; atrocidades; ausencia; ausencias; ausente; ausentes; ausentou; autuacao; autuacoes; autuada; autuadas; autuado; autuados; autuar; aversao; avessa; avessas; avesso; avessos; baixa; baixada; baixadas; baixado; baixando; baixaram; baixas; baixissima; baixissimo; baixo; baixos; banal; banalidade; banalizacao; banalizar; barreira; barreiras; bloqueada; bloqueadas; bloqueado; bloqueados; bloqueando; bloquear; bloqueio; bloqueios; boicotado; boicotando; boicotar; boicotaram; boicote; briga; brigam; brigar; brigaram; brigas; brutal; brutalmente; cai; caiam; caido; caindo; cair; cairam; cairao; cairem; cairia; cairiam; caisse; caiu; calamidade; calamidades; calotes; calunias; cancelada; canceladas; cancelado; cancelados; cancelamento; cancelamentos; cancelar; cancelou; caotica; caotico; carencia; carencias; cassados; cassar; cassaram; catastrophe; catastrofes; catastrophica; catastrophicas; catastrophico; catastrophicos; censura; censurado; censurar; censuraram; censuras; censurou; chantagea; chantageado; chantagear; chantagem;

chantagens; clandestino; clandestinos; coacao; coacoes; coagindo; cobram; cobranca; cobrancas; cobrar; cobraram; coercacao; coercitiva; coercitivas; coercitivo; coercitivos; colapso; colapsos; colisao; coliso; complexa; complexas; complexidade; complexo; complexos; complica; complicacao; complicacoes; complicada; complicadas; complicadissima; complicado; complicador; complicadores; complicados; complicam; complicando; complicar; complicara; complicaria; complicou; compromete; comprometem; comprometer; compulsiva; condena; condenacao; condenacoes; condenada; condenadas; condenado; condenados; condenam; condenamos; condenando; condenar; condenara; condenaram; condenasse; condenatoria; condenatorias; condenatorio; condenava; condenavam; condenaveis; condenavel; condenou; confiscadas; confiscado; confiscados; confiscar; confiscaram; confisco; conflitantes; conflito; conflitos; confrontacao; confrontada; confrontado; confrontados; confrontam; confrontando; confrontantes; confrontar; confronto; confrontos; confrontou; confunde; confundem; confundia; confundiam; confundido; confundindo; confundir; confundiram; confundiu; confusa; confusao; confusas; confuso; confusoes; coniventes; conluio; conluios; conspira; conspiracao; conspiracoes; conspiram; conspirando; conspirar; conspiratorio; constrange; constringedor; constringedora; constringedoras; constringedores; constrangem; constringendo; constranger; constringeu; constringida; constringido; constringidos; constringimento; constringimentos; contamina; contaminacao; contaminada; contaminadas; contaminado; contaminados; contaminar; contaminou; contencao; contencioso; contenciosos; contesta; contestacao; contestacoes; contestada; contestadas; contestado; contingencia; contingenciados; contingenciar; contra; contrabandeados; contrabandear; contrabando; contracao; contradicao; contradicoes; contraditoria; contraditoriamente; contraditorias; contraditorio; contraditorios; contradiz; contrapoe; contrapor; contraposicao; contraria; contrariam; contrariamente; contrariando; contrariar; contrariaram; contrariaria; contrarias; contrariava; contrariedade; contrario; contrarios; contrariou; contraventor; controversa; controversia; controversias; controverso; contundente; contundentes; conturbado; corroendo; corroer; corroi; corrompe; corrompem; corromper; corrompeu; corrompida; corrompido; corrompidos; corrupcao; corrupta; corrupto; corruptor; corruptores; corruptos; corta; cortada; cortadas; cortado; cortados; cortam; cortamos; cortando; cortar; cortaram; corte; cortes; credor; credoras; credores; crime; crimes; criminais; criminal; criminalidade; criminalizacao; criminalmente; criminosa; criminosas; criminoso; criminosos; crise; crises; critica; criticada; criticadas; criticado; criticados; criticam; criticando; criticar; critica; criticaram; criticarem; criticas; criticava; criticavam; critico; criticos; criticou; cuidado; culpa; culpabilidade; culpada; culpado; culpados; culpam; culpando; culpar; culparam; culpas; culposa; culposo; culpou; custado; custam; custando; custar; custara; custaram; custarao; custaria; custas; custava; custavam; custo; custos; danificam; dano; danos; danosa; danoso; decadencia; decepcao; decepcao; decepcao; decepcionada; decepionado; decepcionados; decepcionante; decepcionantes; decepcionar; decepcionaram; decepcionou; decepcoes; declinante; declinio; declinou; decrescente; decrescimo; defeito; defeitos; defeituosas; defeituoso; defeituosos; deficiencia; deficiencias; deficiente; deficientes; deficit; deficitaria; deficitarias; deficitario; deficitarios; deficits; deformacoes; deformadores; degenera; degeneracao; degenerada; degenerativa; degenerativas; degradacao; degradada; degradadas; degradado; degradados; degradam; degradando; degradante; degradantes; degradar; degradou; delacao; delacoes; delinquencia; delinquente; delinquentes; delito; delitos; delituosa; delituosas; delituoso; delituosos; demagogico; demissao; demissoes; demitem; demitido; demitidos; demitindo; demitir; demitiram; demitiu; demolicao; demolidas; demoliram; demora; demorada; demorado; demoram; demorando; demorar; demorou; denegrir; denegriu; depreciacao; depreciativo; depreciativos; depressao; deprimem; deprimente; deprimentes; deprimidos; derrota; derrotada; derrotadas; derrotado; derrotados; derrotando; derrotar; derrotaram; derrotas; derrotou; derruba; derrubada; derrubadas; derrubado; derrubados; derrubam; derrubando; derrubar; derrubara; derrubaram; derrubaria; derrubou; desabastecimento; desabou; desabrigadas; desabrigados; desaceleracao; desaceleradas; desacelerado; desacelerando; desacelerar; desaceleraram; desacelerou; desacerto; desacertos; desacordo; desacordos; desacreditado; desacreditados; desacreditar; desagrada; desagradado; desagradam; desagradando; desagradar; desagradaram; desagradaveis; desagradavel; desagrado; desagradou; desagravo; desajuste; desajustes; desaparecer; desaparecimento; desaprova; desaprovação; desaprova; desaquecimento; desastre; desastres; desastrosa; desastrosas; desastroso; desastrosos; desatencao; desativacao; desativada; desativadas; desativado; desativados; desatualizados; desavenca; desavenças; descartada; descartadas;

descartado; descartados; descartam; descartamos; descartando; descartar; descartaram; descartava;
 descaso; desce; descem; descer; desceram; desceu; descompasso; desconfiou; desconfortaveis;
 desconfortavel; desconforto; desconfortos; desconhece; desconhecem; desconhecemos; desconhecendo;
 desconhecer; desconhecia; desconheciam; desconhecimento; desconsidera; desconsideracao;
 desconsiderando; descontada; descontadas; descontado; descontados; descontando; descontar;
 descontentamento; descontinuidade; desconto; descontos; descontrolada; descontrolado; descontrole;
 descredito; descrenca; descrente; descrentes; descuidado; descuidar; descuido; descumpre;
 descumprem; descumprida; descumpridas; descumprido; descumprimento; descumprimentos;
 descumprir; descumpriram; descumprirem; descumpriu; desembolsado; desembolsados; desembolsar;
 desembolso; desembolsos; desembolsou; desempregado; desempregados; desemprego; desentendeu;
 desentendimento; desentendimentos; disequilibra; disequilibrada; disequilibradas; disequilibrado;
 disequilibrando; disequilibrar; disequilibrio; disequilibrios; disequilibrou; desesperada;
 desesperadamente; desesperadas; desesperado; desesperadora; desesperados; desesperanca; desespero;
 desestabiliza; desestabilizacao; desestabilizador; desestabilizadora; desestabilizadoras; desestabilizar;
 desestatizacao; desestimula; desestimulado; desestimulados; desestimulam; desestimulando;
 desestimular; desestimularia; desestimulo; desestimulou; desestruturacao; desestruturadas; defalques;
 defavoraveis; defavoravel; defavorecida; defavorecidos; defaz; defazem; defazendo; defazer;
 desgasta; desgastada; desgastadas; desgastado; desgastam; desgastando; desgastante; desgastantes;
 desgastar; desgaste; desgastes; desgastou; desgraca; desgracas; desiguais; desigual; desigualdade;
 desigualdades; desigualmente; desinformados; desinteressado; desinteressados; desinteressante;
 desinteresse; desiste; desistem; desistencia; desistencias; desistido; desistindo; desistir; desistiram;
 desistira; desistirem; desistiu; desleais; desleal; deslealdade; desligado; desligados; desligamento;
 desligar; deslizamento; deslizamentos; deslizar; deslize; deslizes; desmoraliza; desmoralizacao;
 desmoralizada; desmoralizadas; desmoralizado; desmoralizados; desmoralizam; desmoralizando;
 desmoralizante; desmoralizar; desnecessariamente; desnecessario; desonestidade; desonesto;
 desonestos; desonra; desonrada; desordem; desordenada; desordenado; desordens; despeito; despejo;
 despenca; despenca; despencando; despencar; despencaram; despencarem; despencou; desperdicado;
 desperdicadores; desperdicam; desperdicando; desperdicar; desperdicio; desperdicios; desperdicou;
 despesa; despesas; desprezar; desprezaram; desprezaveis; desprezavel; desprezo; desprovido;
 desqualifica; desqualificacao; desqualificadas; desqualificam; desqualificar; desrespeito; desrespeitos;
 destituicao; destituir; destroem; destroi; destrua; destrui; destituicao; destruida; destruidas; destruido;
 destruidor; destruidores; destruidos; destruindo; destruir; destruiram; destruiria; destruiu; destrutiva;
 destrutivo; desvaloriza; desvalorizacao; desvalorizacoes; desvalorizada; desvalorizadas; desvalorizado;
 desvalorizados; desvalorizando; desvalorizar; desvalorizaram; desvalorize; desvalorizou; desvantagem;
 desvantagens; desvantajosa; desvia; desviada; desviadas; desviado; desviados; desviam; desviando;
 desviar; desviaram; desviava; desvio; desvios; desviou; detencao; detencoes; deter; detera; deterioracao;
 deteriorando; deteriorar; deteriorou; detida; detidas; detido; detidos; detinha; detinham; detonou;
 devastacao; devastada; devastadas; devastado; devastador; devastadora; devastadoras; devastadores;
 devastados; devastando; devastar; devastaram; devastou; devedor; devedora; devedores; devendo;
 devolucao; devolver; devolveu; devolvida; devolvido; devolvidos; dificeis; dificil; dificilimo;
 dificilmente; dificuldade; dificuldades; dificulta; dificultada; dificultado; dificultam; dificultando;
 dificultar; dificultaram; dificultaria; dificultava; dificultou; diminua; diminua; diminuem; diminua;
 diminuicao; diminuida; diminuido; diminuindo; diminuir; diminuiria; diminuiria; discorda; discordam;
 discordancia; discordancias; discordando; discordante; discordantes; discordar; discordaram;
 discordava; discordavam; discordia; discordias; discussao; discussoes; discutem; discuti; discuti;
 discuti; discuti; dispendio; dispendios; dispendiosa; dispendioso; dispendiosos; disputa; disputada; disputado;
 dissidencia; dissidencias; dissidente; dissidentes; dissociar; dissolucao; dissolver; distorciao; distorce;
 distorcem; distorcer; distorcida; distorcoes; disturbios; divergem; divergencia; divergencias;
 divergentes; divergir; divida; dividas; doenca; doencas; doente; doloso; drastica; drasticamente;
 drásticas; drastico; drásticos; duramente; duvida; duvidam; duvidar; duvidas; duvidosa; duvidosas;
 duvidoso; duvidosos; elimina; eliminacao; eliminada; eliminadas; eliminado; eliminados; eliminam;
 eliminamos; eliminando; eliminar; eliminaram; embargo; embargos; emergencia; emergenciais;
 emergencial; emergenciais; empecilho; empecilhos; encarece; encarecem; encarecendo; encarecer;

encareceram; encareceria; encareceu; encarecimento; encargo; encargos; encolheu; encolhido; encolhimento; endivida; endividada; endividadas; endividado; endividados; endividam; endividamento; endividando; endividar; endividaram; endividou; endurecer; endureceu; endurecimento; enfraqueca; enfraquece; enfraquecem; enfraquecendo; enfraquecer; enfraquecera; enfraqueceram; enfraqueceria; enfraqueceu; enfraquecida; enfraquecidas; enfraquecido; enfraquecidos; enfraquecimento; enfureceu; enfurecida; enfurecido; engana; enganacao; enganada; enganado; enganadora; enganados; enganam; enganando; enganar; engano; enganos; enganosa; enganosas; enganoso; enganosos; enganou; epidemia; epidemias; equivocada; equivocadamente; equivocadas; equivocado; equivocados; equivoco; equivocos; erosao; erra; errada; erradamente; erradas; errado; errados; errando; errar; erraram; erro; erronea; erroneamente; erroneas; erros; errou; escandalo; escandalos; escandalosa; escandalosamente; escandalosas; escandaloso; escandalosos; escapar; escaparam; escapou; escassez; escombros; esconder; escondeu; escondia; escondiam; escondida; escondidas; escondido; escondidos; esgota; esgotada; esgotadas; esgotado; esgotados; esgotam; esgotamento; esgotando; esgotar; esgotaram; esgotarem; espera; esquecem; esquecendo; esquecer; esqueceram; esqueceu; esquecida; esquecidas; esquecido; esquecidos; esquecimento; estagnacao; estagnada; estagnadas; estagnado; estagnados; estagnar; estagnou; estraga; estragada; estragar; estrago; estragos; estragou; estressado; estressados; estressante; estressar; estresse; esvazia; esvaziada; esvaziadas; esvaziado; esvaziados; esvaziamento; esvaziando; esvaziar; esvaziou; evasao; evasivo; evita; evitada; evitadas; evitado; evitados; evitam; evitando; evitar; evitara; evitaram; evite; evitem; evitou; excluem; exclui; excluida; excluidas; excluido; excluidos; excluindo; excluir; excluiu; exclusao; exoneracao; exoneracoes; exonerada; exonerado; exonerar; exorbitantes; exorbitar; explode; explodem; explodido; explodindo; explodir; explodiram; explodiu; exploracao; explorada; exploradas; explorado; explorador; expropria; expropriacao; expropriacoes; expropriada; expropriar; expropriou; expulsa; expulsado; expulsando; expulsao; expulsar; expulsaram; expulso; expulsos; expulsou; extincao; extorsao; extravio; falcatrua; falcatruas; falencia; falencias; falha; falhado; falham; falhar; falharam; falhas; falho; falhou; falida; falidas; falido; falidos; falir; faliram; faliu; falsa; falsamente; falsas; falsidade; falsidades; falsificacao; falsificacoes; falsificadores; falsificados; falsificar; falsos; falta; faltando; faltar; faltaram; faltas; faltava; faltou; fardo; farsa; farsante; fatais; fatal; fatalidade; fatalmente; fechada; fere; ferem; ferido; feridos; ferir; feriram; feriu; fianca; fiancas; ficticio; ficticios; finge; fingir; flagrado; flagrados; flagrante; forcada; forçado; forçados; forçando; forçar; forçaram; forçou; fracassa; fracassada; fracassadas; fracassado; fracassados; fracassam; fracassando; fracassar; fracassaram; fracassarem; fracasso; fracassos; fracassou; fraco; fracos; fragilidade; fragilidades; fragilizacao; fragilizada; fragilizadas; fragilizado; fragilizados; fragilizar; fraqueza; fraquezas; fraudada; fraudadas; fraudado; fraudador; fraudadores; fraudados; fraudando; fraudar; fraudaram; fraudava; fraude; fraudes; fraudou; fraudulenta; fraudulentamente; fraudulentas; fraudulento; fraudulentos; frustra; frustracao; frustracoes; frustrada; frustradas; frustrado; frustrados; frustrante; frustrantes; frustrar; frustraram; frustrou; fugido; fugindo; fugir; fugiram; fugitivo; fugitivos; fugiu; furia; furiosa; furiosamente; furioso; furiosos; gargalo; gargalos; gasta; gastado; gastador; gastadora; gastadores; gastam; gastamos; gastanca; gastando; gastar; gastara; gastaram; gastaria; gastava; gastavam; gasto; gastos; golpe; golpes; golpismo; golpista; golpistas; gravidade; gravissima; gravissimas; gravissimo; gravissimos; greve; greves; hiperinflacao; horrenda; horrendo; horriveis; horrivel; horror; horrores; horrorizado; horrorosa; horroroso; hostile; hostilidade; hostilidades; hostilizada; hostilizado; hostilizar; hostis; humilhacao; humilhacoes; humilhada; humilhado; ignora; ignorada; ignoradas; ignorado; ignorados; ignoram; ignorancia; ignorando; Ignorando; ignorante; ignorantes; ignorar; ignoraram; ilegais; ilegal; ilegalidade; ilegalidades; ilegalmente; ilicita; ilicitamente; ilicitas; ilicito; ilicitos; ilicitude; ilicitudes; imorais; imoral; imoralidade; impede; impedem; impedia; impediam; impedido; impedidos; impedimento; impedimentos; impedindo; impedir; impedira; impediram; impedirem; impediria; impediriam; impedisse; impedissem; impediu; imperfeicao; imperfeicoes; impoe; impoem; impondo; impor; impos; imposicao; imposicoes; impositiva; impositivo; impossibilidade; impossibilita; impossibilitada; impossibilitado; impossibilitados; impossibilitando; impossibilitar; impossibilitou; impossiveis; impossivel; imposta; impostas; imposto; impostos; impotencia; impotente; improbidade; improprio; improprios; imprudencia; imprudente; impunidade; inaceitavel; inadequacao; inadequada; inadequadamente; inadequadas; inadequado; inadequados; inadimplencia; inadimplente; inadimplentes;

inadmissível; inaptos; incapacidade; incapaz; incapazes; incerta; incerteza; incertezas; incerto; incidente; incidentes; incomoda; incomodado; incomodar; incomodava; incomodavam; incomodo; incomodos; incompatibilidade; incompatíveis; incompatível; incompetencia; incompetente; incompetentes; inconcebível; inconsistência; inconsistências; inconsistente; inconsistentes; inconveniente; inconvenientes; incorreto; incorretos; indefinição; indefinições; indefinida; indefinidamente; indefinidas; indefinido; indesejada; indesejadas; indesejado; indesejados; indesejáveis; indesejável; indevida; indevidamente; indevidas; indevido; indevidos; indiciado; indiciados; indiciamento; indignação; indignada; indignadas; indignado; Indignado; indignados; indignar; indignidade; indignou; ineficácia; ineficaz; ineficazes; ineficiência; ineficiências; ineficiente; ineficientes; inércia; infeliz; infelizes; infelizmente; inferior; inferiores; infidelidade; infieis; infiel; inflação; inflacionar; inflacionária; inflacionárias; inflacionário; inflacionários; inflações; infraco; infracoés; infrator; infratoras; infratores; infringindo; inibe; inibem; inibicao; inibido; inibidor; inibidores; inibindo; inibir; inibiu; injúria; injúrias; injusta; injustamente; injustas; injusta; injusticada; injusticado; injusticados; injusticas; injustificável; injusto; injustos; inquerito; inqueritos; insatisfação; insatisfações; insatisfatória; insatisfatórias; insatisfatório; insatisfatórios; insatisfeita; insatisfeitas; insatisfeito; insatisfeitos; insegura; insegurança; inseguranças; inseguras; inseguro; inseguros; insolvência; instabilidade; instáveis; instável; insucesso; insucessos; insuficiência; insuficiências; insuficiente; insuficientemente; insuficientes; insuperáveis; insuperável; insuportáveis; insuportável; insustentabilidade; insustentáveis; insustentável; interdição; interfere; interferem; interferência; interferências; interferindo; interferir; interferiu; interrompe; interrompendo; interromper; interromperam; interrompeu; interrompida; interrompidas; interrompido; interrompidos; interrupção; interrupções; intimação; intimado; intimida; intimidação; intimidações; intimidada; intimidado; intimidados; intimidam; intimidando; intimidar; intimidou; inúteis; inútil; inutilidade; inutilmente; invadida; invadidas; invadido; invadidos; invadir; invadiram; invadirem; invadiu; inválida; inválida; invalidar; inválidas; invalidez; inválido; inválidos; invasão; invasiva; invasivas; invasivo; invasivos; vasos; investiga; investigação; Investigação; investigações; Investigações; investigada; investigadas; investigado; investigados; investigam; investigando; investigar; investigaram; investigativa; investigativas; investigativo; investigava; investigavam; investigou; investigue; investiguem; inviabiliza; inviabilização; inviabilizada; inviabilizado; inviabilizados; inviabilizam; inviabilizando; inviabilizar; inviabilizaram; inviabilizaria; inviabilizem; inviabilizou; inviáveis; inviável; irregular; irregulares; irregularidade; irregularidades; irregularmente; irreparáveis; irreparável; irresponsabilidade; irresponsabilidades; irresponsáveis; irresponsável; irresponsavelmente; irreversíveis; irreversível; irreversivelmente; irritação; irritada; irritadas; irritado; irritados; irritam; irritando; irritante; irritar; irritaram; irritava; irritou; juro; lamenta; lamentações; lamentado; lamentam; lamentamos; lamentando; lamentar; lamentaram; lamentava; lamentáveis; lamentável; lamentavelmente; lamento; lamentou; lentidão; lento; lentos; lesa; lesada; lesado; lesados; lesao; lesar; lesões; limita; limitação; limitações; limitada; limitadas; limitado; limitados; limitam; limitando; limitar; limitaram; limitava; limite; limites; limitou; litígio; litígios; litigiosidade; máfia; máfias; máfiosa; máfioso; máfiosos; maldade; maldades; maldição; maldita; maldito; malefício; malefícios; malefício; males; malfeita; malfeito; malfeitores; malfeitorias; malfeitos; maliciosa; maltratadas; maltratava; manipula; manipulação; manipulações; manipulada; manipuladas; manipulado; manipulador; manipuladores; manipulados; manipulam; manipulando; manipular; massacrada; massacrado; massacrados; massacrar; massacre; massacres; matam; matança; matanças; matando; matar; mataram; matavam; mau; medíocre; medo; medos; mensala; Mensala; mensalões; mentir; mentira; mentiras; mentirosa; mentirosas; mentiroso; mentirosos; minimamente; minimiza; miseráveis; miserável; miséria; miserias; morosa; morosidade; moroso; morra; morram; morre; morrem; morrendo; morrer; morreria; morreram; morreria; morrerem; morreria; morresse; morreu; morria; morriam; morrido; mortalidade; mortas; morte; mortes; morto; mortos; multa; multada; multadas; multado; multados; multar; multas; nega; negação; negada; negadas; negado; negados; negam; negamos; negando; negar; negara; negaram; negarem; negativa; negativamente; negativas; negativo; negativos; negava; negavam; negligência; negligenciadas; negligenciar; negligente; obito; obitos; obrigações; obsessão; obsessiva; obsessivamente; obsessivo; obsessões; obsolescência; obsoleta; obsoletas; obsoleto; obsoletos; obstáculo; obstáculos; obstante; obstrua; obstruações; obstruída; obstruído; obstruindo; obstruir; obstruíram; obstruíu; oculta;

ocultacao; ocultado; ocultados; ocultam; ocultando; ocultar; ocultas; ocultava; oculto; ocultos; odio; odios; ofende; ofendem; ofendendo; ofender; ofendeu; ofendida; ofendido; ofendidos; ofensa; ofensas; ofensiva; ofensivas; ofensivo; ofensivos; ofuscadas; ofuscaram; omissao; omissas; omissos; omissoes; omissos; omite; omitem; omitida; omitido; omitidos; omitindo; omitir; omitiram; omitiu; onera; oneradas; onerado; onerados; oneram; onerando; onerar; onerosa; onerosas; oneroso; onerosos; opoe; opoem; opondo; oponente; oponentes; opor; oportunismo; oportunista; oportunistas; oposicoes; opositor; opositora; opositoras; opositores; oposta; opostas; oposto; opostos; opressao; opressiva; opressor; opressores; oprime; oprimida; oprimidadas; oprimidados; oprimir; opunha; opunham; opuseram; paga; pagado; pagam; pagar; panico; paralisa; paralisação; paralisações; paralisada; paralisadas; paralisado; paralisados; paralisam; paralisando; paralisante; paralisar; paralisaram; peculato; penalidade; penalidades; perda; perdas; perde; perdedor; perdedora; perdedoras; perdedores; perdem; perdemos; perdendo; perder; perdera; perderam; perderao; perderem; perderemos; perderia; perderiam; perdermos; perdesse; perdeu; perdia; perdiam; perdida; perdidas; perdido; perdidos; perigo; perigos; perigosa; perigosamente; perigosas; perigosissimo; perigoso; perigosos; perturba; perturbacao; perturbacoes; perturbado; perturbador; perturbadora; perturbadoras; perturbadores; perturbar; perversa; perversas; perversidade; perversidades; perverso; perversos; pesando; pesar; pesaram; pessima; pessimas; pessimismo; pessimista; pessimistas; pessimo; pessimos; pior; piora; piorado; pioram; piorando; piorar; pioraram; piorarem; piore; piores; piorou; pobre; pobres; pobreza; poluem; poluente; poluentes; poluicao; poluida; poluidas; poluido; poluidor; poluidora; poluidoras; poluidores; poluidos; poluindo; poluir; precaria; precarias; precariedade; precario; precarios; precipitacao; precipitadas; precipitado; preconceito; preconceitos; predatoria; predatorio; prejudica; prejudicada; prejudicadas; prejudicado; prejudicados; prejudicam; prejudicando; prejudicar; prejudicara; prejudicaram; prejudicaria; prejudicariam; prejudicasse; prejudicava; prejudiciais; prejudicial; prejudicou; prejudique; prejudiquem; prejuizo; prejuizos; preocupa; preocupacao; preocupacoes; preocupada; preocupadas; preocupado; preocupados; preocupam; preocupando; preocupante; preocupantes; preocupar; preocupou; preso; presos; pressao; pressionada; pressionado; Pressionado; pressionados; pressionam; pressionando; pressionar; pressionaram; pressionou; pressoes; prevaricacao; prevaricar; prisao; prisoneira; prisoneiras; prisoneiro; prisoneiros; prisoes; priva; privacao; privacoes; privam; privando; privar; privilegiadas; privilegiados; privilegiar; privilegios; problema; problemas; problematica; problematicas; problematico; problematicos; processado; proibe; proibem; proibida; proibicao; proibicoes; proibida; proibidas; proibido; proibidos; proibindo; proibir; proibira; proibiram; proibiu; propina; propinas; propinoduto; protecionismo; protecionista; protecionistas; protestos; provocacao; provocacoes; provocada; provocadas; provocado; provocados; provocam; provocando; provocar; provocar; provocaram; provocarem; provocaria; provocariam; provocou; punem; punicao; punicoes; punido; punindo; punir; quebra; quebrada; quebradas; quebradeira; quebrado; quebrados; quebram; quebrando; quebrar; quebraram; quebraria; quebras; quebras; quebrasse; quebrava; quebravam; quebre; quebrem; quebrou; queda; quedas; queixa; queixado; queixam; queixando; queixar; queixaram; queixas; queixava; queixavam; queixo; queixosos; queixou; racionalizacao; racionalizar; racionamento; rebaixa; rebaixada; rebaixadas; rebaixado; rebaixados; rebaixam; rebaixamento; rebaixamentos; rebaixando; rebaixar; rebaixou; recaem; recaí; recaia; recaída; recaídas; recair; recaira; recaíram; recaíu; receio; receios; recessao; recessivo; recesso; recessoes; reclama; reclamacao; reclamacoes; reclamada; reclamadas; reclamado; reclamados; reclamam; reclamando; reclamar; reclamaram; reclamarem; reclamava; reclamavam; reclamou; recuado; recuando; recuar; recuaram; recuo; recuos; recuou; recusa; recusada; recusadas; recusado; recusados; recusam; recusamos; recusando; recusar; recusara; recusaram; recusarem; recusasse; recusava; recusavam; recusou; reducao; reducoes; reduz; reduza; reduzam; reduzem; reduzi; reduzida; reduzidas; reduzido; reduzidos; reduzimos; reduzindo; reduzir; reduzira; reduziram; reduzirrem; reduziria; reduziu; refazer; regressiva; rejeicao; rejeita; rejeitada; rejeitadas; rejeitado; rejeitados; rejeitam; rejeitando; rejeitar; rejeitara; rejeitaram; rejeitarem; rejeitasse; rejeitava; rejeitavam; rejeite; rejeitem; rejeito; rejeitos; rejeitou; removendo; remover; removeu; removida; removidas; removido; removidos; renunciar; renunciaram; renuncias; repressao; repressivo; reprime; reprimi; reprimida; reprimidas; reprimido; reprimidos; reprimindo; reprimir; reprimiu; reprovacao; reprovado; reprovados; repudia; repudiada; repudiado; repudiam; repudiamos; repudiando; repudiar; repudiaram; repudio; repudiou; repugnancia; repugnante; repulsa; repulsivo; restricao;

restricoes; restringe; restringem; restringi; restringia; restringido; restringindo; restringir; restringira; restringiram; restringiria; restringiu; restrinja; restrinjam; retaliacao; retaliar; retardar; retencao; retira; retirada; retiradas; retirado; retirados; tiram; retirando; retirar; retiraram; retirarem; retiraria; retirasse; retirassem; retirava; retiravam; retracao; retroceder; retrocederam; retrocedeu; retrocesso; retrocessos; revolta; revoltada; revoltado; revoltados; revoltante; revoltaram; revoltas; revoltosos; revoltou; ridicularizada; ridicularizados; ridicularizam; ridiculo; ridiculos; risco; riscos; rivais; rival; rivalidade; rivalidades; rombo; rombos; rompe; rompem; rompendo; romper; rompera; romperam; romperia; rompeu; rompida; rompidas; rompido; rompidos; rompimento; rompimentos; rouba; roubada; roubadas; roubado; roubados; roubalheira; roubam; roubando; roubar; roubaram; roubava; roubo; roubos; roubou; ruído; ruídos; ruidosa; ruidoso; ruidosos; ruim; ruína; ruínas; ruins; ruir; ruiram; ruiu; ruptura; rupturas; sabotadores; sabotagem; sabotagens; sabotar; sabotou; sacrifica; sacrificada; sacrificadas; sacrificado; sacrificados; sacrificam; sacrificando; sacrificar; sacrificaram; sacrificio; sacrificios; sacrificou; sancao; sancoes; segregacao; severamente; severo; sobrecarga; sobretaxa; sobretaxas; socorre; socorrendo; socorrer; socorreram; socorreu; socorrida; socorridas; socorrido; socorridos; socorro; socorros; sofra; sofre; sofrem; sofremos; sofrendo; sofrer; sofrera; sofreram; sofrerao; sofrerem; sofreu; sofra; sofrida; sofridas; sofrido; sofridos; sofrimento; sofrível; sonega; sonegacao; sonegado; sonegador; sonegadores; sonegados; sonegam; sonegar; sonegou; stress; subdesenvolvido; subdesenvolvidos; subdesenvolvimento; subestimada; subestimado; subestimados; subestimar; subordinacao; subordinada; subordinado; subordinados; subornado; subornar; suborno; subornos; subtrair; subutilizadas; sumico; sumido; sumindo; sumir; sumiram; sumiu; superfaturada; superfaturadas; superfaturado; superfaturados; superfaturamento; superfaturamentos; superfaturar; suspeita; suspeitam; suspeitar; suspeitaram; suspeitas; suspeitava; suspeitavam; suspeito; suspeitos; suspeitou; suspende; suspendem; suspendendo; suspender; suspenderam; suspenderia; suspendeu; suspenso; suspensa; suspensao; suspensas; suspenso; suspensoes; suspensos; susto; sustos; teme; temem; temendo; temer; temeram; temerem; temerosa; temeroso; temerosos; temeu; temia; temiam; temida; temidas; temido; temidos; temor; temores; tensa; tensao; tensas; tenso; tensoes; tensos; termina; terminada; terminadas; terminado; terminados; terminam; terminar; termino; terriveis; terrível; terrivelmente; terror; terrorismo; terrorista; terroristas; tirar; tiraram; tirarem; tiraria; tirava; tragedia; tragedias; tragica; tragicamente; tragicas; tragico; tragicos; traicao; traicoes; traida; traído; traidor; traidores; traídos; traindo; trair; traíram; traiu; transgressao; transgressoes; trauma; traumas; traumática; traumáticas; traumático; traumatizada; traumatizado; triste; tristemente; tristes; tristeza; turbulencia; turbulencias; urgencia; urgencias; vandalismo; vandalos; vergonha; vergonhas; vergonhosa; vergonhosas; vergonhoso; vergonhosos; veta; vetada; vetadas; vetado; vetados; vetam; vetando; vetar; vetaram; veto; vetos; vetou; vingado; vingança; vingar; vingou; viola; violacao; violacoes; violada; violadas; violado; violador; violadores; violados; violam; violando; violar; violaram; violarem; violaria; violencia; violencias; violenta; violentamente; violentas; violento; violentos; vulateis; volátil; volatilidade; volatilidades; zera; zerada; zeradas; zerado; zerados; zerar; zero; zeros; zerou

beneficiariao; beneficiarem; beneficiaria; beneficiarias; beneficiario; beneficiava; beneficiem; beneficio; beneficios; beneficiou; benefico; beneficos; bons; brilhante; capaz; cedeu; celebra; celebrao; celebraoes; celebrada; celebradas; celebrado; celebrados; celebram; celebramos; celebrando; celebrar; celebraram; celebrarao; celebraremos; celebrou; certeza; certezas; clareza; colabora; colaboracao; colaboracoes; colaborado; colaborador; colaboradora; colaboradoras; colaboradores; colaboram; colaborando; colaborar; colaboraram; colaborarem; colaborativa; colaborativas; colaborativo; colaborem; colaborou; combatendo; combater; combateram; combatida; combatido; comemoracao; comemorar; competente; competentes; competentissimos; comprometidos; comprometimento; compromissadas; compromisso; compromissos; conceder; concederam; concedeu; concedida; concedidas; concedido; concedidos; concessao; concessoes; concordancia; concretamente; concretas; concretizacao; concretizada; concretizadas; concretizado; concretizar; concretizaram; concretize; concretizem; concretizou; concreto; concretos; confia; confiabilidade; confiado; confiados; confiam; confianca; confiancas; confiando; confiante; confiantes; confiar; confiaveis; confiavel; confirmar; confirmaram; confirmarem; confirmou; confortavel; confortavelmente; conforto; confortos; conquista; conquistada; conquistadas; conquistado; conquistados; conquistam; conquistamos; conquistando; conquistar; conquistara; conquistaram; conquistarmos; conquistas; conquistassem; conquiste; conquistem; conquistou; consegue; conseguido; conseguindo; conseguir; conseguiram; conseguira; conseguirem; conseguiria; conseguiriam; conseguirmos; conseguisse; conseguiu; consistencia; consistente; consistentes; consolida; consolidacao; consolidacoes; consolidada; consolidadas; consolidado; consolidador; consolidadora; consolidadoras; consolidadores; consolidados; consolidam; consolidamos; consolidando; consolidar; consolidara; consolidaram; consolidaremos; consolidarmos; consolidasse; consolidava; consolidem; consolidou; construímos; construindo; construir; construiram; construiu; construtiva; construtivo; contente; contentes; contingenciamento; contingenciamentos; contratacao; contratacoes; contratar; contribua; contribuam; contribuem; contribui; contribuia; contribuiam; contribuicao; contribuicoes; contribuido; contribuimos; contribuindo; contribuir; contribuiria; contribuiriam; contribuirmos; contribuissem; contribuiu; cooperacao; correta; corretamente; corretas; correto; corrigida; corrigidas; corrigido; corrigidos; corrigir; corrigiram; credibilidade; cresca; crescam; cresce; crescem; crescemos; crescendo; crescente; crescentemente; crescentes; crescer; crescerá; cresceram; crescerao; crescerem; cresceremos; cresceria; crescermos; crescesse; crescessem; cresceu; crescido; crescimento; crescimentos; cumprir; cumpriram; cumpriu; defende; defendem; defendemos; defendendo; defender; defenderam; defendeu; defendido; defendidos; defesa; defesas; deflacao; desejado; desejavel; desempenha; desempenhadas; desempenhado; desempenham; desempenhando; desempenhar; desempenharam; desempenho; desempenhos; desenvolva; desenvolvam; desenvolve; desenvolvedora; desenvolvedores; desenvolvem; desenvolvemos; desenvolvendo; desenvolver; desenvolvera; desenvolveram; desenvolverao; desenvolverem; desenvolveremos; desenvolvermos; desenvolvessem; desenvolveu; desenvolviam; desenvolvida; desenvolvidas; desenvolvido; desenvolvidos; desenvolvimentismo; desenvolvimento; desenvolvimentos; desonera; desoneracao; desoneracoes; desonerar; diferenciada; diferenciadas; diferenciado; diferenciados; diferenciais; diferencial; dispor; disposicao; dividendos; economiza; economizada; economizado; economizador; economizados; economizam; economizamos; economizando; economizar; economizara; economizou; educar; educativa; educativas; educativo; educativos; eficacia; eficaz; eficazes; eficazmente; eficiencia; eficiente; eficientemente; eficientes; elevacao; elevacoes; elevada; elevadas; elevadissima; elevadissimas; elevadissimo; elevadissimos; elevado; elevados; elevam; elevar; elevaria; elevaram; elevarao; elevarem; elevaria; elevariam; elevasse; elevava; elevem; elogio; elogiada; elogiadas; elogiado; elogiados; elogiam; elogiando; elogiar; elogiaram; elogiava; elogio; elogios; elogiou; emerge; emergem; empenha; empenhada; empenhadas; empenhado; empenhados; empenham; empenhamento; empenhamos; empenhando; empenhar; empenharam; empenhasse; empenho; empenhos; empenhou; emprega; empregam; empregando; empregar; empregaram; empregarao; empregarem; empregavel; enorme; enormemente; enriquece; enriquecem; enriquecendo; enriquecer; enriqueceu; enriquecida; enriquecido; enriquecimento; entusiasmado; entusiasmo; entusiasta; equilibrio; erario; erradicacao; erradicar; esforca; esforcado; esforcam; esforcando; esforcar; esforcaram; esforcava; esforco; esforcos; esperanca; espetacular; espetaculares; espetacularmente; estabilidade; estabilidades; estabiliza;

estabilizaco; estabilizada; estabilizadas; estabilizado; estabilizados; estabilizamos; estabilizando; estabilizar; estabilizaram; estabilize; estabilizem; estabilizou; estaveis; estavel; estimulada; estimuladas; estimulado; estimulados; estimulam; estimulamos; estimulando; estimulante; estimulantes; estimular; estimularam; estimulava; estimulou; evolua; evolucao; evolucoes; evolui; evoluído; evoluídos; evoluimos; evoluindo; evoluir; evoluíram; evoluírao; evoluírem; evoluiria; evoluirmos; evoluii; evolutiva; evolutivo; evolutivos; excelencia; excelente; excelentes; excepcional; excepcionalmente; exito; exitos; expandida; expandidas; expandido; expandimos; expandindo; expandir; expandira; expandiram; expandirao; expandirem; expandirmos; expandisse; expandiu; expansao; expansionista; expansionistas; expectativa; expectativas; facil; facilmente; fantastico; favoraveis; favoravel; favoravelmente; felicidade; feliz; fidelidade; fideliza; fidelizaco; fidelizado; fidelizam; fidelizar; flexibilidade; flexibilizaco; flexibilizar; fortaleca; fortalecam; fortalece; fortalecem; fortalecemos; fortalecendo; fortalecer; fortalecera; fortaleceram; fortalecerao; fortalecerem; fortaleceremos; fortaleceu; fortalecida; fortalecidas; fortalecido; fortalecidos; fortalecimento; ganha; ganhado; ganhador; ganhadora; ganhadores; ganham; ganhamos; ganhando; ganhar; ganhara; ganharam; ganharao; ganharem; ganharemos; ganharia; ganhariam; ganharmos; ganhas; ganhasse; ganhassem; ganhava; ganhavam; ganhe; ganhem; ganho; ganhos; ganhou; garantia; garantias; garantido; garantidos; garantir; honestidade; honra; honrada; honrado; honrar; honraria; honrarias; honras; honrosa; honrosas; honroso; honrou; igualdade; impressionante; impulsiona; impulsionada; impulsionadas; impulsionado; impulsionados; impulsionam; impulsionando; impulsionar; impulsionaram; impulsionou; incentiva; incentivada; incentivadas; incentivado; incentivadores; incentivados; incentivam; incentivando; Incentivar; incentivar; incentivar; incentive; incentivem; incentivo; incentivos; incentivou; inovacao; inovacoes; inovador; inovadora; inovadoras; inovadores; inovam; inovar; Integracao; integrou; investe; investem; investiamos; investida; investidas; investido; investidos; Investimento; investimento; investimentos; Investimentos; investimos; investindo; investir; investira; investiram; investirao; investirem; investiremos; investiria; investirmos; investiu; lidera; liderada; lideradas; liderado; liderados; lideram; lideramos; liderando; liderar; louvavel; lucra; lucrado; lucram; lucrando; lucrar; lucraram; lucrativa; lucrativas; lucratividade; lucrativo; lucrativos; lucro; lucros; lucrou; maravilhas; maxima; maximizaco; maximizar; maximo; melhorado; melhorando; melhorar; melhoraram; melhorarem; melhoras; melhorassem; melhores; melhora; melhorias; melhorou; moralidade; oportunidade; oportunidades; otima; otimas; otimismo; otimista; otimistas; otimiza; otimizaco; otimizacoes; otimizada; otimizado; otimizam; otimizando; otimizar; otimizou; otimo; otimos; perfeito; positiva; positivamente; positivas; positividade; positivismo; positivo; positivos; potencial; potencialidade; potencialidades; potencializa; potencializada; potencializado; potencializando; potencializar; potencializou; potencialmente; poupa; poupadores; poupando; poupar; premiacao; premiacoes; premiada; premiadas; premiado; premiados; premiando; premiar; premiara; premiaram; premio; premios; preservacao; preservado; preservados; preservando; preservar; preservaram; prestigia; prestigiada; prestigiadas; prestigiado; prestigiados; prestigiar; prestigio; prestigiou; prevencao; prevencoes; preventiva; preventivas; preventivo; preventivos; produtivas; produtivo; produtivos; progredido; progredimos; progredindo; progredir; progredirem; progrediu; progressao; progressivas; progressividade; progressivo; progressivos; progresso; progressos; progride; progridem; promover; promoveram; promoveu; promovida; promovidas; promovido; promovidos; prospera; prosperam; prosperando; prosperar; prosperaram; prosperas; prospere; prosperem; prosperidade; prospero; prosperos; prosperou; prudencia; prudenciais; prudencial; prudente; prudentes; reagia; reagiam; reagido; reagindo; reagir; reagiram; realizacao; realizacoes; realizado; realizando; realizar; reaquecendo; reaquecer; reaqueceu; reaquecimento; receita; receitas; recompensa; reconhecimento; recupera; recuperacao; recuperacoes; recuperada; recuperadas; recuperado; recuperados; recuperam; recuperamos; recuperando; recuperar; recuperara; recuperaram; recuperarao; recuperarem; recuperaria; recuperarmos; recuperasse; recuperava; recuperaveis; recuperavel; recupere; recuperem; recuperou; reestruturacao; reestruturar; reforca; reforcado; reforcando; reforcar; reforco; reformas; regularizar; relevancia; relevante; relevantes; remunera; remuneracao; remuneracoes; rende; rendem; rendendo; render; renderam; renderao; renderia; rendeu; rendimento; rendimentos; renovavel; rentabilidade; rentabiliza; rentabilizar; rentavel; reparacao; reparacoes; reparado; resolvido; resolvidos; respeitado; respeitados; respeitam; respeitar; respeitaveis; respeitavel; restauracao; retomada; retomadas; retomado;

retomados; retomam; retomar; reverter; reverteram; revertida; revertidas; revertido; riqueza; riquezas; salto; saltos; saltou; salvar; salvaram; sanada; sanadas; sanado; sanados; sanar; satisfatoriamente; satisfeita; satisfeitas; satisfeito; satisfeitos; saudavel; seguro; seriedade; significativa; significativamente; significativo; significativos; sobem; solidariedade; solidez; subida; subir; subiram; subirao; subirem; subiria; subiriam; subisse; subissem; subsidios; sucedida; sucedidas; sucedido; sucedidos; sucesso; sucessos; supera; superacao; superada; superadas; superado; superam; superando; superar; superaram; superarem; superaria; superasse; superava; superavam; superavit; superavitaria; superavits; supere; superem; superior; superiores; superioridade; superou; suprimento; sustentabilidade; sustentaveis; sustentavel; transformacao; transformacoes; transformados; transformam; transformando; transformar; transformou; transparencia; transparentes; triunfar; triunfo; triunfos; triunfou; valiosa; valiosas; valioso; valiosos; valoriza; valorizacao; valorizacoes; valorizada; valorizadas; valorizado; valorizados; valorizam; valorizamos; valorizando; valorizar; valorizaram; valorizasse; valorizassem; valorize; valorizem; valorizou; vantagem; vantagens; vantajosa; vantajosas; vantajoso; vantajosos; vence; vencedor; vencedora; vencedoras; vencedores; vencem; vencendo; vencer; vencera; venceram; venceu

ANEXO A – Exemplo da estratificação do banco de dados usados na proposta do dicionário

(Continua)

J O R N A L	I D	Sítio	A U T O R	M A T É R I A
O G L O B O	6 7 8 9 7	www2.senado.leg.br/bdsf/bitstream/handle/id/415507/noticia.htm?sequence=1	9 F R E A N C I S E C I O L E A	Matéria O dono Marka foi condenado pela Justica brasileira por rombo nos cofres publicos Michel denunciou ontem pressoes governo brasileiro para extradicao criticou recomendacao Justica Monaco Corte Apelacao nao tomou uma decisao Ela prestou favor governo brasileiro disse advogado agencia AFP Brasil pediu extradicao Cacciola outubro ministro Justica Tarso Genro comemorou Para ele retorno Cacciola indica existencia pais novo padrao combate corrupcao importante ponto vista simbolico porque sinaliza para sociedade brasileira que existe disposicao pais lutar contra impunidade todos terrenos simbolo importante para sociedade brasileira Segundo Michel Constituicao Principado permite contestar ato administrativo junto Supremo Tribunal instancia maxima composta por cinco juristas todos franceses professores universitarios Eles reuinem algumas vezes por ano Como decisao principe sobre extradicao ato administrativo advogados veem uma chance atrasar extradicao banqueiro tempo que Cacciola passar atras das grades Monaco sera descontado sua pena Brasil banqueiro que insiste dizer que sua prisao Brasil politica esta melhor instalado rico principado com vista mar Para Genro Albert devera decidir ate mes mantem extradicao Cacciola Mas segundo ministro governo brasileiro esta pronto para buscar banqueiro dois dias Pelo plano tracado pela Policia Federal dois delegados pelo menos dois agentes deverao ser destacados para escoltar Cacciola banqueiro seria levado aviao carreira Monaco Paris iria para Rio onde deve come car cumprir pena Cacciola foi condenado anos prisao por gestao fraudulenta corrupcao passiva peculato Mas estava fora Brasil desde gracias habeas corpus que livrou prisao momento que tivermos decisao principe horas estamos prontos para fazer extradicao disse ministro Justica Marka quebrou com desvalorizacao cambial Cacciola pediu ajuda Banco Central que alegando querer evitar que quebra dos bancos elevasse nervosismo mercado decidiu injetar bilhao Marka FonteCindam Cacciola chegou ficar preso Brasil mas obteve habeas corpus viajou para Italia Ele foi preso setembro luxuoso hotel Fairmont Monaco onde passava ferias com namorada registrado com seu nome verdadeiro Correspondente DO SOCORRO FINANCEIRO PRISAO LUXO Salvatore Alberto Cacciola que era dono falido Banco Marka entrou para cobertura policial dos jornais logo apos maxidesvalorizacao real janeiro Comprometido mercado futuro entregar bilhao Marka que contramao mercado apostava manutencao politica cambial ficou sem dinheiro para quitar divida Com alegacao nao deixar quebrar sistema financeiro Banco Central vendeu dolares instituicao cotacao inferior mercado fez mesmo com FonteCindam causando prejuizos aos cofres publicos estimados bilhao Acusado gestao fraudulenta peculato desvio dinheiro publico Cacciola passou dias carceragem Ponto Zero Benfica mas conseguiu habeas corpus foi para Italia foi condenado anos prisao Protegido pela dupla cidadania italiano naturalizado brasileiro morava Roma sem risco extradicao Cacciola foi preso Monaco setembro Interpol Italia descobriu que ele estava indo para Monaco com namorada avisou policia principado que prendeu Cacciola num hotel cassino Desde entao banqueiro esta casa detencao monegasca proxima palacio real Catedral Sao Nicolas Nas celas lavabo cobertores verao inverno refrigerador com controle remoto Cacciola divide uma cela com dois internos presidio com vista para mar oferece ainda pista para caminhadas longo dos corredores uma biblioteca depender Corte Apelacoes Monaco que chegou adiar tres vezes sua posicao quanto pedido brasileiro extradicao ele tera voltar realidade das prisoes brasileiras Jornal GLOBO Autor Deborah Berlinck Jailton Carvalho Editoria Economia Tamanho palavras Edicao Pagina Coluna Secao Caderno Primeiro Caderno

(Continua)

J O R N A L D	Sítio	D A T A	A U T O R	Matéria
V A L O R	8 4 2 8 1 www2.senado.leg.br/bdsf/bitstream/handle/id/522178/noticia.html?sequence=1	16 D E D E Z M A R B R O	M A R	<p>Por Marli Olmos argentinos estao torcida para que nova equipe governo brasileiro consiga manter taxa cambio nos niveis atuais algum sinal por menor que seja que Produto Interno Bruto PIB vai parar contrair partir proximo ano Para setor produtivo argentino essas variaveis vizinho sao tao importantes quanto seu proprio pais Delas dependem competitividade grande parte atividade Argentina Cada ponto contracao PIB brasileiro golpeia atividade economica pais vizinho que destina mercado brasileiro sua exportacao industrial real continuasse desvalorizar exportacoes argentinas perderiam cada vez mais Talvez por isso Diego Burzaco economista chefe Inversor Global consultoria financas sediada Buenos Aires refira ascensao presidente interino Michel Temer como alivio provisorio questao politica principal parceiro comercial Argentina nao esta completamente resolvida mas lado economico surgiram sinais preciosos politica cambial deles Quando dolar chegou surgiu grande temor Argentina principalmente pelas perspectivas que poderia chegar ate fim ano afirma Burzaco sinais intencao governo Temer nao deixar cotacao muito alem dos tranquiliza argentinos principalmente porque nao momento indicacoes desvalorizacao peso Embora cambio esteja defasado analistas consideram que autoridades monetarias argentinas deverao continuar intervir nos proximos meses para impedir desvalorizacao moeda assim evitar mais pressao inflacionaria Com economia brasileira debilitada atividade argentina sofre duas frentes destaca Burzaco lado porque suas exportacoes sobretudo produtos manufaturados diminuem por falta compradores outro lado fronteira outro queda demanda Brasil tende levar industria brasileira pensar mercado argentino como destino natural para estoque que encalha economista Dante Sica diretor consultoria Abeceb diz que entre intercambio comercial entre dois paises caiu bilhoes ano passado relacao deteriorou mais ainda nos ultimos meses Argentina viu importacoes crescerem destaca Nos quatro primeiros meses deficit bilateral triplicou relacao igual periodo ano passado atingindo bilhao governo argentino manteve posicao cautela relacao crise politica Brasil mesmo depois decisao Senado pelo afastamento presidente Dilma Rousseff cargo Desejamos que essa discussao politica complexa resolva melhor maneira possivel disse ministro chefe gabinete Marcos Pena Brasil nosso vizinho pais irmao principal aliado estrategico nas questoes politica economica cultural tudo que acontece tem uma transcendencia importante para todos nos destacou Ministerio das Relacoes Exteriores Argentina divulgou uma nota pouco depois concluida votacao Senado por meio qual destacou que respeita processo institucional Brasil confia que desenlace situacao consolide democracia brasileira nota chancelaria destacou ainda intencao governo argentino continuar dialogando com autoridades constituídas para continuar avancando com processo integracao bilateral regional Para Burzaco entanto governo Temer tera pouco avancar tanto Mercosul quanto nas negociacoes bloco sul americano com Uniao Europeia Brasil ainda passa por uma situacao institucional delicada afirma escolha Henrique Meirelles para Ministerio Fazenda foi bem recebida pelos analistas pela forte reputacao que ele tem mercado Trata profissional que consegue consenso que parece ter claro sacrificios necessarios para ajuste inevitavel diz Burzaco Para Sica regioa espera que Temer consiga fazer com que crise economica chegue limite Seu maior desafio agora conseguir apoio Poder Legislativo destaca economista afirma ainda que maior expectativa volta agora sinalizacao medidas que indiquem que havera uma recuperacao economica brasileira mudanca politica Brasil entusiasmou argentinos assunto foi principal tema mostrado pela imprensa durante todo dia Todos concordaram que nao pode esperar por uma rapida recuperacao Mas diante uma crise tao profunda sinal que coisas podem melhorar alento Orgaos relacionados Senado Federal</p>

(conclusão)

JORNAL	ID	Sítio	DATA	AUTOR	Matéria
ESTADÃO	45787	www2.senado.leg.br/bdsf/bitstream/handle/id/314245/noticia.htm?sequence=1	3 DE FEVEREIRO DE 2017	LEONARDO	Ao longo dos meses deste ano governo retirou mercado papeis divida atrelada variacao preco dolar total bilhoes divida cambial apenas total foram refinanciados pelo governo resgate desse tipo divida foi bilhoes com uma taxa rolagem refinanciamento dos vencimentos bem maior Goldenstein atribuiu queda divida cambial tres principais fatores menor volatilidade taxa cambio longo ano percepcao dos agentes economicos que nao pressoes sobre balanco pagamentos externo reducao divida externa setor privado reducao divida externa setor privado diminuiu necessidade demanda por hedge protecao cambial previsao governo que divida publica cambial feche ano cerca total Segundo chefe Demab estrategia reducao exposicao cambial divida continuara Ate momento nao mudanca estrategia por parte Goldenstein entanto fez questao ressaltar que essa estrategia nao tem nenhuma vinculacao com formacao taxa cambio Nao estamos mirando nenhum objetivo taxa cambio acordo com Goldenstein devera haver ligeiro aumento exposicao cambial dezembro entre outros fatores por causa desvalorizacao real relacao dolar Ele explicou que estatisticas fechadas ano sobre resgate divida cambial foram fechadas porque nao fez rolagens dezembro Desde agosto nao faz rolagens dos vencimentos titulos contratos swap cambial PREFIXADOS Alem reducao exposicao cambial governo conseguiu tambem elevar participacao titulos prefixados bilhoes para bilhoes total divida novembro inicio governo Lula papeis prefixados que trazem menor risco financiamento para governo representavam apenas total divida ponto positivo que novo perfil divida reduziu risco Pais qualquer variacao mercado avaliou coordenador Divida Publica Tesouro Nacional Paulo Valle volume titulos que vencem meses divida curto prazo caiu novembro para total divida mobiliaria federal interna equivalente bilhoes outubro divida curto prazo correspondia Esse outro dado positivo apresentado nas estatisticas divida publica prazo medio estoque divida entanto teve ligeira queda outubro para novembro passando meses para meses

Nota: Devido ao banco de dados do dicionário ser formado por 121.435 notícias, foi apresentado apenas uma notícia de cada jornal para fins de exemplificação.

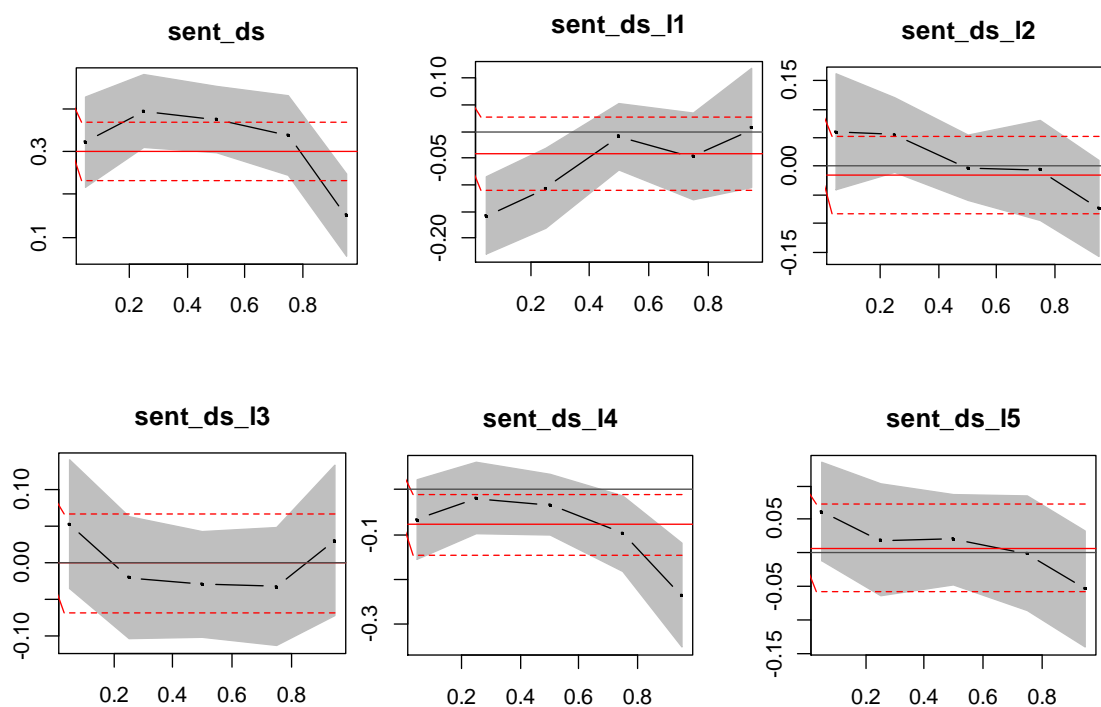
ANEXO B – Correlograma da Variância Condicional do IBOVESPA

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
*****	*****	1	0.974	0.974	1397.0	0.000
*****		2	0.949	0.010	2724.3	0.000
*****	*	3	0.920	-0.090	3972.6	0.000
*****		4	0.890	-0.030	5142.8	0.000
*****		5	0.863	0.034	6243.1	0.000
*****		6	0.837	0.016	7279.0	0.000
*****		7	0.813	0.017	8257.1	0.000
*****		8	0.792	0.032	9184.9	0.000
*****		9	0.772	0.016	10067.	0.000
*****		10	0.752	-0.017	10905.	0.000
*****		11	0.732	-0.019	11698.	0.000
*****		12	0.711	-0.010	12449.	0.000
*****		13	0.691	0.010	13159.	0.000
*****		14	0.673	0.024	13833.	0.000
*****		15	0.656	0.012	14473.	0.000
*****		16	0.640	-0.003	15083.	0.000
*****		17	0.626	0.032	15666.	0.000
*****		18	0.611	-0.010	16223.	0.000
*****		19	0.597	-0.010	16754.	0.000
*****		20	0.583	0.004	17262.	0.000
*****		21	0.567	-0.042	17743.	0.000
*****		22	0.551	-0.022	18196.	0.000
*****		23	0.533	-0.021	18621.	0.000
*****		24	0.516	0.010	19020.	0.000
*****		25	0.501	0.019	19395.	0.000
*****		26	0.485	-0.005	19749.	0.000
*****		27	0.472	0.018	20083.	0.000
*****		28	0.460	0.024	20401.	0.000
*****		29	0.449	-0.004	20704.	0.000
*****		30	0.439	0.004	20993.	0.000
*****		31	0.428	-0.013	21268.	0.000
*****		32	0.418	0.022	21531.	0.000
*****		33	0.405	-0.055	21778.	0.000
*****		34	0.394	0.015	22012.	0.000
*****		35	0.382	0.004	22232.	0.000
*****		36	0.372	0.010	22441.	0.000

ANEXO C – Estimativas Sobre o Sentimento Textual

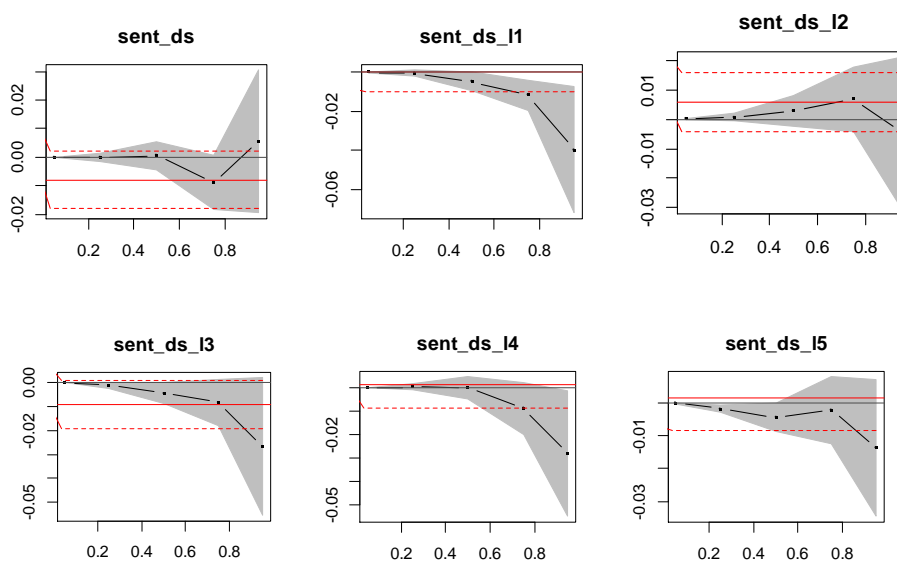
As figuras apresentam as estimativas dos parâmetros das variáveis relacionadas ao sentimento textual obtidos por regressão quantílica, fazendo comparação a estimativa realizada por um modelo de regressão linear (OLS). As imagens possibilitam que sejam observados o comportamento do sentimento (eixo Y), ao longo dos quantis de distribuição condicional do desempenho do mercado acionário brasileiro (eixo X). Admitindo que, as linhas vermelhas tracejadas mostram o intervalo de confiança da estimativa por OLS, e as áreas hachuradas apresentam o intervalo de confiança da estimação por regressão quantílica, admitindo, em ambos, intervalos com 95% de confiança.

Figura 8 - Estimativas dos coeficientes do sentimento textual ao longo da distribuição condicional do retorno do IBOVESPA



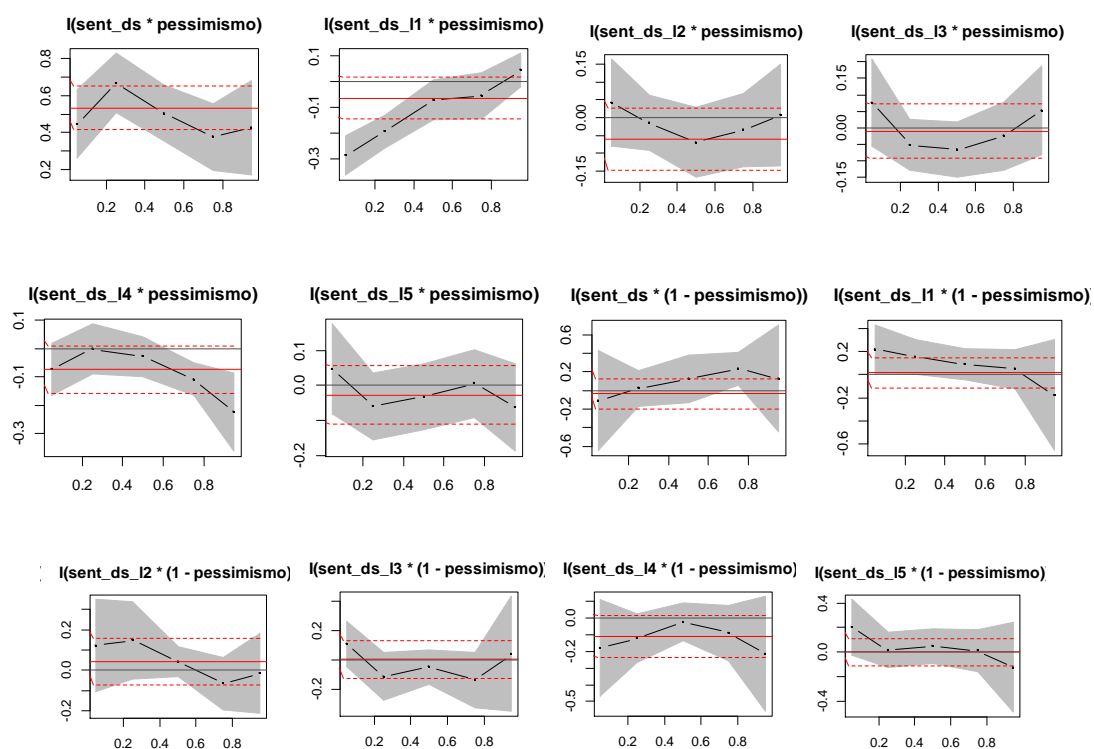
Nota: linha contínua vermelha corresponde a estimação por OLS, linha preta contínua estimação por regressão quantílica. Linhas tracejadas vermelhas – intervalo de confiança do OLS, área hachurada intervalo de confiança da estimação por regressão quantílica.

Figura 9 - Estimativas dos coeficientes do sentimento textual ao longo da distribuição condicional da volatilidade do IBOVESPA



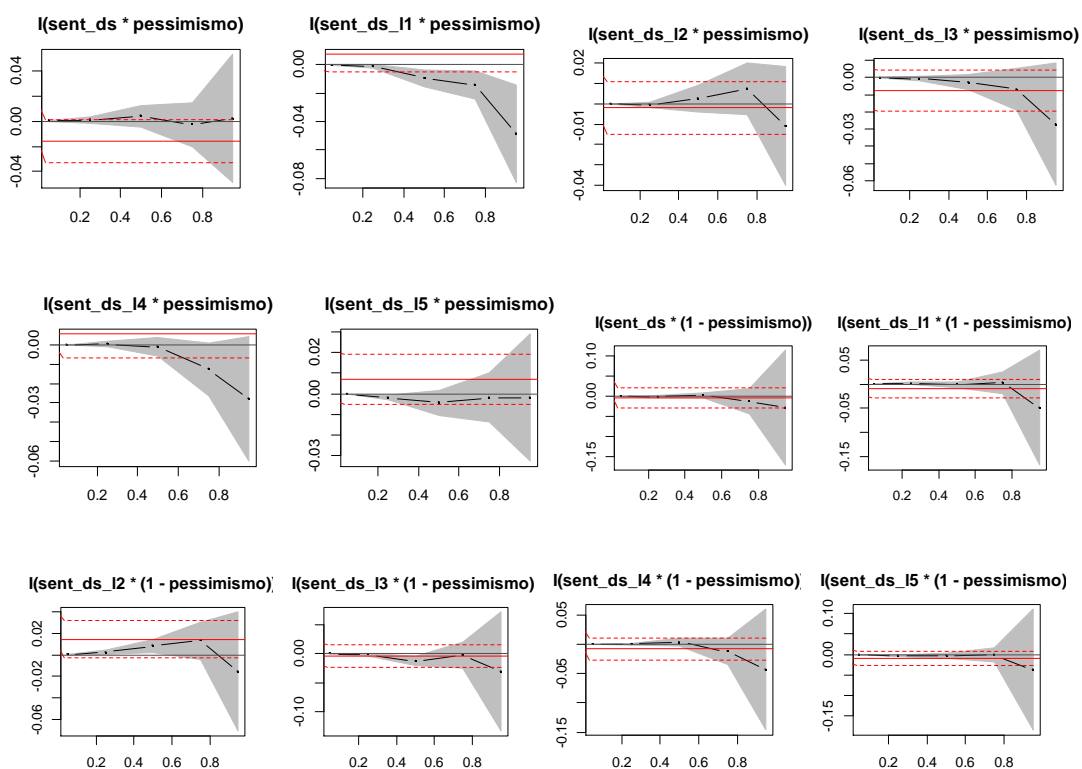
Nota: linha contínua vermelha corresponde a estimação por OLS, linha preta contínua estimação por regressão quantílica. Linhas tracejadas vermelhas – intervalo de confiança do OLS, área hachurada intervalo de confiança da estimação por regressão quantílica.

Figura 10 - Estimativas dos coeficientes do sentimento de pessimismo e otimismo ao longo da distribuição condicional do Retorno do IBOVESPA



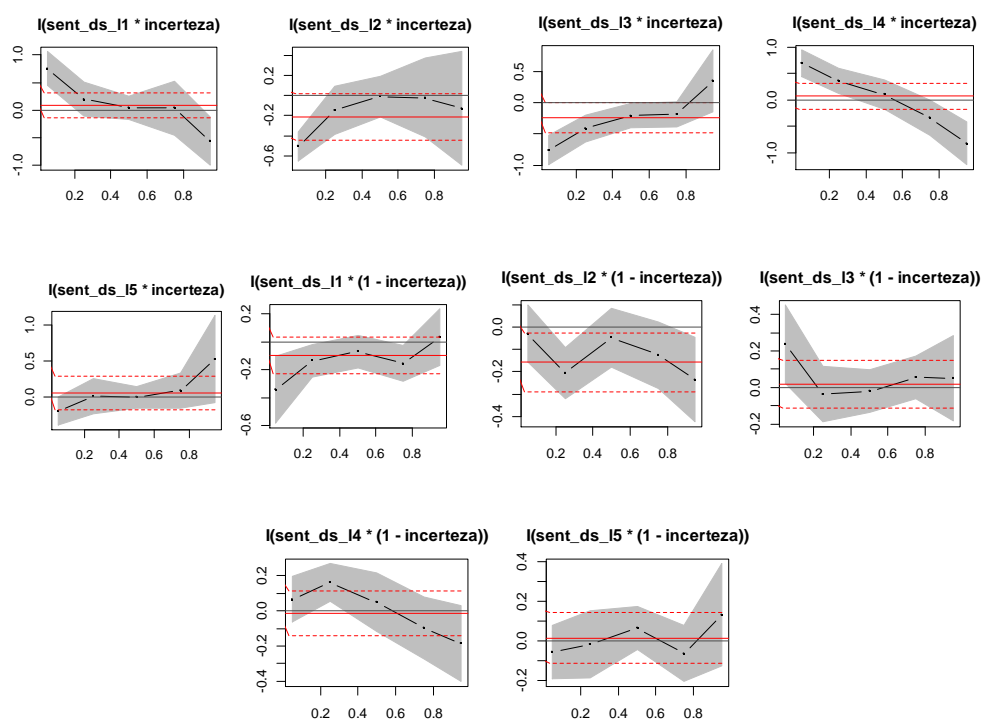
Nota: linha contínua vermelha corresponde a estimação por OLS, linha preta contínua estimação por regressão quantílica. Linhas tracejadas vermelhas – intervalo de confiança do OLS, área hachurada intervalo de confiança da estimação por regressão quantílica.

Figura 11 - Estimativas dos coeficientes do sentimento de pessimismo e otimismo ao longo da distribuição condicional da Volatilidade do IBOVESPA



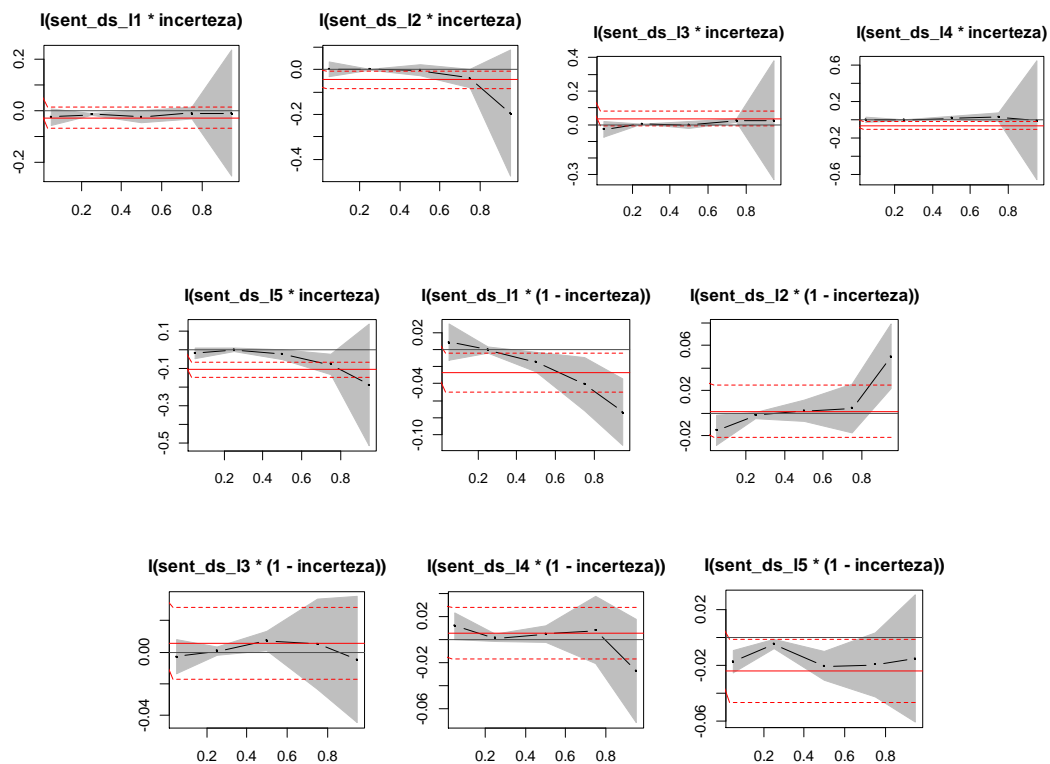
Nota: linha contínua vermelha corresponde a estimação por OLS, linha preta contínua estimação por regressão quantílica. Linhas tracejadas vermelhas – intervalo de confiança do OLS, área hachurada intervalo de confiança da estimação por regressão quantílica.

Figura 12 - Estimativas dos coeficientes do sentimento de pessimismo em dias de incerteza ao longo da distribuição condicional da retorno do IBOVESPA



Nota: linha contínua vermelha corresponde a estimação por OLS, linha preta contínua estimação por regressão quantílica. Linhas tracejadas vermelhas – intervalo de confiança do OLS, área hachurada intervalo de confiança da estimação por regressão quantílica.

Figura 13 - Estimativas dos coeficientes do sentimento de pessimismo em dias de incerteza ao longo da distribuição condicional da volatilidade do IBOVESPA



Nota: linha contínua vermelha corresponde a estimação por OLS, linha preta contínua estimação por regressão quantílica. Linhas tracejadas vermelhas – intervalo de confiança do OLS, área hachurada intervalo de confiança da estimação por regressão quantílica

