

UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS CONTÁBEIS  
CURSO DE MESTRADO EM CIÊNCIAS CONTÁBEIS

DYLIANE MOURÍ SILVA DE SOUZA

**EFEITO DO SENTIMENTO DO INVESTIDOR MANIFESTO VIA TWITTER  
SOBRE OS RETORNOS E O VOLUME NEGOCIADO NO MERCADO ACIONÁRIO  
BRASILEIRO**

JOÃO PESSOA- PB

2020

DYLIANE MOURÍ SILVA DE SOUZA

**EFEITO DO SENTIMENTO DO INVESTIDOR MANIFESTO VIA TWITTER  
SOBRE OS RETORNOS E O VOLUME NEGOCIADO NO MERCADO ACIONÁRIO  
BRASILEIRO**

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Ciências Contábeis da Universidade Federal da Paraíba, como requisito parcial à obtenção do grau de Mestre em Ciências Contábeis.

**Área de Concentração:** Informação Contábil

**Linha de Pesquisa:** Informação Contábil para Usuários Externos

**Orientador:** Dr. Orleans Silva Martins

JOÃO PESSOA - PB

2020

**Catálogo na publicação**  
**Seção de Catalogação e Classificação**

S729e Souza, Dyliane Mourí Silva de.

Efeito do Sentimento do Investidor Manifesto via  
Twitter sobre os Retornos e o Volume Negociado no  
Mercado Acionário Brasileiro / Dyliane Mourí Silva de  
Souza. - João Pessoa, 2020.

87 f. : il.

Orientação: Orleans Silva Martins.  
Dissertação (Mestrado) - UFPB/CCSA.

1. Análise de Sentimento. 2. Twitter. 3. Mercado  
Acionário. I. Martins, Orleans Silva. II. Título.

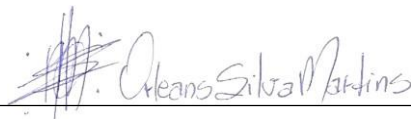
UFPB/BC

DYLIANE MOURÍ SILVA DE SOUZA

**EFEITO DO SENTIMENTO DO INVESTIDOR MANIFESTO VIA TWITTER  
SOBRE OS RETORNOS E O VOLUME NEGOCIADO NO MERCADO ACIONÁRIO  
BRASILEIRO**

Dissertação apresentada ao Curso de Mestrado do Programa de Pós-Graduação em Ciências Contábeis da Universidade Federal da Paraíba, como requisito parcial à obtenção do grau de Mestre em Ciências Contábeis.

**Banca Examinadora**



---

**Prof. Dr. Orleans Silva Martins**  
Universidade Federal da Paraíba  
Orientador



---

**Prof. Dr. Luiz Felipe de Araújo Pontes Girão**  
Universidade Federal da Paraíba  
Examinador Externo

---

**Prof. Dr. Moacir Manoel Rodrigues Júnior**  
Universidade Regional de Blumenau  
Examinador Externo

João Pessoa - PB

2020

## AGRADECIMENTOS

Agradeço primordialmente a Deus, por cuidar e olhar por mim em todos os momentos de minha vida.

Aos meus pais, Rita e Valmir, por todo amor e dedicação, além de todo o investimento a mim destinado, especialmente no âmbito educacional, para que eu possa atingir meus objetivos. Obrigada por entenderem minhas ausências durante esse período e por se dedicarem para melhorar a minha qualidade de vida, de modo que eu pudesse me dedicar de forma mais tranquila aos estudos. Sem vocês absolutamente nada seria possível. Amo vocês!

Ao meu namorado, Vinícius Matheus, por todo companheirismo, compreensão e encorajamento nos momentos difíceis. Por ter sido a pessoa que mais teve que lidar com minhas crises e inseguranças ao longo desse caminho, mas que sempre esteve ao meu lado, me mostrando o quanto capaz eu sou, me dando equilíbrio emocional e me ajudando em completamente tudo o que precisei, não medindo esforços para me ver bem. Meu amor e admiração por você crescem a cada dia que passa.

Ao meu irmão, Dayvisson, por ser meu amigo, companheiro de todas as horas, conselheiro e apoiador em todos os meus momentos. Também por ser a primeira pessoa da nossa família que fez um mestrado e que me serviu de incentivo para trilhar o mesmo caminho. Te amo!

Ao meu orientador, Dr. Orleans Silva Martins, por ter aceitado a ideia de fazer algo que não é tão comum na nossa área, e por toda confiança depositada. Obrigada por sempre ser tão acessível, comprometido e dedicado. Obrigada por todos os direcionamentos que me foram dados e por todo o conhecimento que fora partilhado.

Aos professores Dr. Luiz Felipe de Araújo Pontes Girão e Dr. Moacir Manoel Rodrigues Júnior, por aceitarem o convite para avaliar o trabalho, e por todas as valiosas críticas e sugestões para a realização dessa pesquisa.

Aos professores Rossana Guerra, Wenner Glaucio, Carla Janaína, e Tiago Echternacht, por terem visto potencial em mim e me incentivado, durante a graduação, a fazer o mestrado.

A todo o corpo docente do PPGCC, por todos os ensinamentos e conhecimentos compartilhados. Por terem me feito não apenas adquirir um maior nível de conhecimento, mas por me tornarem uma pessoa que possui um maior senso crítico, e por me fazerem crescer profissionalmente e academicamente.

Aos meus amigos, por terem entendido esse meu novo momento e as minhas ausências. Por me mandarem forças e sempre perguntarem como estavam as coisas, mesmo aqueles que nem entendem muito bem o que é que eu faço. Por me trazerem alegria e paz ao longo dos dias, mesmo que à distância. Agradeço ainda aos amigos que foram feitos na turma 04, os quais tornaram esses dois anos muito mais aprazíveis. Minha vida é muito mais feliz tendo vocês comigo.

## RESUMO

O objetivo desse estudo foi identificar, por meio das mensagens que são postadas no Twitter, como as informações que são divulgadas de forma *on-line* estão associadas aos movimentos que ocorrem no mercado brasileiro, especificamente no que tange aos retornos e ao volume de negócios. A coleta dos dados relacionados ao Twitter se deu por meio da biblioteca Tweepy. Já os dados financeiros foram obtidos mediante a base de dados Thomson Reuters. Para o desenvolvimento do índice de sentimento foi empregado um método netnográfico, no qual houve uma observação participante na rede social Twitter de modo a conhecer os termos que são utilizados a se referir ao mercado brasileiro e as ações que o compõem. A atribuição do sentimento se deu por meio de *machine learning*, por intermédio do *Google Cloud Natural Language API*, o qual possui uma ferramenta de análise de sentimentos. Para atender ao objetivo geral da pesquisa foram estimados sete modelos de regressão quantílica em cinco quantis, uma vez que os dados obtidos eram heterogêneos e não demonstravam normalidade, visto que esse tipo de regressão apresenta robustez para tais problemas, assim foi possível ter uma compreensão do assunto e interpretação dos dados. A análise e interpretação dos dados possibilitou perceber que, em geral, um sentimento otimista no tempo contemporâneo estará associado a um maior retorno contemporâneo, todavia essa relação se inverte com o passar dos dias, de modo que, um sentimento otimista no período atual estará associado a um menor retorno subsequente. Constatou-se ainda que existe associação entre o volume de mensagens que são postadas diariamente no Twitter e o volume de negócios do mercado acionário brasileiro. Ademais, viu-se que quão maior for o número de mensagens que são dotadas de um sentimento negativo, maior também será o volume negociado. Assim, os resultados são úteis para mostrar que existe relação entre as informações que são divulgadas na rede social Twitter e os movimentos do mercado acionário brasileiro, trazendo contribuições para a literatura, ao expandir o entendimento de como mercados emergentes estão associados às atividades que ocorrem no ambiente *on-line*. O estudo também traz contribuições práticas, uma vez que as atividades que ocorrem *on-line* no Twitter podem ser utilizadas como variáveis em estratégias de investimento, visto que essas estão associadas aos movimentos do mercado.

**Palavras-chave:** Análise de Sentimento; Twitter; Mercado Acionário.

## ABSTRACT

The aim of this paper was to identify, through the messages that are posted on Twitter, how the information that is disseminated online is associated with the movements that occur in the Brazilian market, specifically with regard to returns and traded volume. The collection of data related to Twitter took place through the Tweepy library. Financial data were obtained using the Thomson Reuters database. For the development of the sentiment index, a netnographic method was used, in which there was a participant observation in the social network Twitter in order to know the terms that are used to refer to the Brazilian market and the stocks that compose it. The sentiment was attributed through machine learning, by the Google Cloud Natural Language API, which has a sentiment analysis tool. To reach the general objective of the research, seven quantile regression models were estimated in five quantiles, since the data obtained were heterogeneous and did not demonstrate normality, since this type of regression is robust to such problems, so it was possible to have an understanding of the subject and interpretation of data. The analysis and interpretation of the data made it possible to perceive that, in general, an optimistic sentiment in contemporary time will be associated with a greater contemporary return, however this relationship is inverted over the days, so that, an optimistic sentiment in the current period will be associated with a subsequent lower return. It was also found that there is a significant association between the volume of messages that are posted daily on Twitter and traded volume of the Brazilian stock market. In addition, it was seen that the greater the number of messages that have a negative feeling, the greater the traded volume. Thus, it is believed that the results are useful to show that there is a relationship between the information that is released on the social network Twitter and the movements of the Brazilian stock market, bringing contributions to the literature, by expanding the understanding of how emerging markets are associated with activities that occur in the online environment. The study also has practical contributions, since activities that occur online on Twitter can be used as variables in investment strategies, since these are associated with market movements.

**Keywords:** Sentiment analysis; Twitter; Stock Market.

## LISTA DE FIGURAS

<b>Figura 1</b> – <i>Tweet</i> extraído do Twitter - 01/08/2017.....	37
<b>Figura 2</b> – <i>Tweet</i> extraído do Twitter - 02/06/2017.....	37



## LISTA DE TABELAS

<b>Tabela 1</b> – Estatísticas descritivas (01/01/2017 a 31/12/2018) .....	44
<b>Tabela 2</b> - Teste da HME utilizando <i>tweets</i> enquanto informação – IBOV.....	51
<b>Tabela 3</b> - Teste da HME controlado pelo volume de negócios –IBOV .....	54
<b>Tabela 4</b> -Relação entre o sentimento e o retorno – IBOV.....	57
<b>Tabela 5</b> -Associação entre o volume negociado e o volume de <i>tweets</i> .....	61
<b>Tabela 6</b> -Associação entre o volume negociado e os tipos de sentimento .....	63
<b>Tabela 7</b> -Teste da HME utilizando <i>tweets</i> enquanto informação – SMLL .....	65
<b>Tabela 8</b> - Relação entre o sentimento e o retorno – SMLL .....	67

## LISTA DE ABREVIATURAS E SIGLAS

API — *Application Programming Interface*

IPO — *Initial public offering*

LPA — Lucro por Ação

BM — *Book-to-Market*

ETF— *Exchange-traded fund*

PL— Preço por Lucro

PIB — Produto Interno Bruto

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>11</b>
1.2 Objetivos.....	13
1.2.1 Objetivo geral.....	13
1.2.2 Objetivos específicos.....	13
1.3 Justificativa e Contribuições.....	13
<b>2 REFERENCIAL TEÓRICO</b> .....	<b>16</b>
2.1 Hipótese do Mercado Eficiente e Racionalidade Limitada do Investidor.....	16
2.2 Novas Maneiras de Medir o Sentimento.....	24
2.3 Demais Fatores que Impactam o Retorno e o Volume Negociado .....	30
<b>3 METODOLOGIA</b> .....	<b>34</b>
3.1 Atribuição do sentimento .....	36
3.2 Modelos.....	37
3.3 Análise adicional - Small Caps.....	41
3.4 Limitações.....	42
<b>4 ANÁLISE DOS RESULTADOS</b> .....	<b>43</b>
4.1 Estatísticas Descritivas .....	43
4.2 Análise dos Gráficos .....	46
4.3 Associação entre o sentimento dos <i>tweets</i> e o Retorno do IBOVESPA .....	50
4.3.1 Teste da HME utilizando <i>tweets</i> enquanto informação - IBOV .....	50
4.3.2 Teste da HME controlado pelo volume de negócios .....	53
4.3.3 Relação entre o sentimento e o retorno controlada por demais variáveis - IBOV.....	55
4.4 Associação entre o volume de <i>tweets</i> e o volume negociado - IBOVESPA.....	60
4.4.1 Análise por Tipo de Sentimento dos <i>Tweets</i> .....	62
4.5 Associação entre o sentimento e o retorno do SMLL .....	64
4.5.1 Teste da HME utilizando <i>tweets</i> enquanto informação – SMLL .....	64
4.5.2 Relação entre o Sentimento e o Retorno controlada por demais variáveis .....	66
4.6 Resultados no Contexto Brasileiro.....	69
4.7 Testes Adicionais Não Reportados.....	70
<b>5 CONSIDERAÇÕES FINAIS</b> .....	<b>72</b>
<b>REFERÊNCIAS</b> .....	<b>75</b>
<b>APÊNDICE A- Tabelas da Associação entre o Sentimento Dos <i>Tweets</i> e o Retorno do IBOVESPA</b> .....	<b>81</b>

<b>APÊNDICE B- Tabelas da Associação entre o Volume De <i>Tweets</i> e o Volume Negociado</b> .....	<b>84</b>
<b>APÊNDICE C - Associação entre o sentimento e o retorno do SMLL</b> .....	<b>86</b>
<b>APÊNDICE D – Coleta de Dados e Atribuição do Sentimento</b> .....	<b>87</b>

## 1 INTRODUÇÃO

As oscilações que ocorrem no preço das ações presentes no mercado acionário por muito tempo estiveram associadas puramente a junção dos preços passados e do ambiente informacional, ideia que fora trazida pela Hipótese do Mercado Eficiente (HME), onde não seria possível prever o mercado (FAMA, 1970). Todavia, com o passar dos anos e o surgimento de anomalias, diversos autores, tais como Menkhoff (1998) e Barberis, Shleifer e Vishny (1998), passaram a questionar se o sentimento dos investidores também seria responsável por tais oscilações.

A relação entre o sentimento do investidor e os movimentos que ocorrem no mercado acionário tem sido objeto de uma série de estudos ao longo dos últimos anos. Não obstante, tais estudos demonstram que há uma forte ligação entre medidas de sentimento do investidor e os retornos no mercado (LEE; JIANG; INDRO, 2002; BROWN; CLIFF, 2004, 2005; BAKER; WURGLER, 2006, 2007). Assim, uma vez que as mais diversas pesquisas demonstram que o sentimento e os retornos estão associados, Baker e Wurgler (2007) afirmam que o foco de novas pesquisas deve ser em identificar qual a melhor maneira de medir esse sentimento, e como quantificar os seus efeitos.

Identificar novas maneiras de captar o sentimento se faz necessário porque a maioria dos estudos utilizam apenas duas fontes para auferir o sentimento, que são as pesquisas de opinião (BROWN; CLIFF, 2005) ou variáveis de mercado (BAKER; WURGLER, 2006). Nesse sentido, uma série de pesquisadores, assim como Wei, Mao e Wang (2016), Atkins, Niranjana e Gerding (2018), Das, Behera e Rath (2018), e Chen e Lo (2019), tem voltado sua atenção de modo a captar o sentimento dos investidores por meio de atividades *on-line*, especialmente porque nas últimas duas décadas, a forma de difundir informações no mercado de ações foi alterada drasticamente devido ao avanço da internet e desenvolvimento da Web 2.0 (FAN; GORDON, 2014).

Esse novo tipo de abordagem dá aos pesquisadores a possibilidade de captar o sentimento de forma instantânea, sendo essa uma solução para críticas como a de Klibanof, Lamont e Wizman (1998), que afirmam que grande parte dos estudos que consideram o comportamento do investidor são retrospectivos, e assim, determinantes desse comportamento, como o sentimento de investimento não são mais observáveis.

No que tange a captura do sentimento de forma instantânea, um dos principais meios para tal são as redes sociais, isso porque, a ascensão e popularidade delas faz como que essas sejam uma maneira rápida e fácil de postar, compartilhar e obter informação. Ademais, acredita-se que essa facilidade seja uma maneira de melhorar a eficiência do mercado, visto que haverá mais informações e que essas podem ser obtidas sem grandes esforços. Todavia, estudos demonstram que as informações divulgadas por meio das redes sociais podem ter efeitos sobre o sentimento dos investidores, de modo que eles atualizam as suas crenças acerca das potenciais gerações de fluxos de caixa futuros das ações, sendo responsáveis por oscilações que ocorrem no mercado. Assim, as redes sociais são tidas como uma importante fonte de dados, sendo uma forma de conhecer o sentimento dos investidores, e sendo possivelmente úteis na previsibilidade do mercado (FAN; GORDON, 2014; SHEN; LIU; ZHANG, 2018).

Embora muitos possam crer que essa previsibilidade do mercado, por meio de atividade *on-line*, estaria totalmente distante da eficiência do mercado, Renault (2017) acredita que esse seria apenas um modelo mais realista de eficiência de mercado, o qual seria respaldado por Grossman e Stiglitz (1980), apresentando um modelo da HME em sua forma semiforte. Isso porque, a previsibilidade, mesmo que a um baixo nível, e os retornos anormais obtidos seriam simplesmente um tipo de recompensa para investidores que gastam tempo e dinheiro de modo a monitorar uma ampla variedade de fontes de informação. Essa recompensa ajuda a entender o conflito entre a eficiência com que os mercados difundem suas informações e os incentivos para a aquisição delas.

Uma das formas mais usadas para obter informações e identificar supostas previsibilidades do mercado acionário é por meio da rede social Twitter, uma vez que, segundo Wei, Mao e Wang (2016), o Twitter é uma das redes sociais mais utilizadas pelo mercado, na qual muitos *traders*, investidores, analistas financeiros, entidades e perfis de notícias costumam postar mensagens acerca do mercado de ações. Ademais, estudos como o de Mao, Counts e Bollen (2011) comprovam que o sentimento extraído por meio do Twitter, bem como o número de *tweets*, podem ser usados para prever o retorno diário do mercado.

Assim, apesar de estudos como os de Mao, Counts e Bollen (2011) e Wei, Mao e Wang (2016) já terem comprovado que há relação entre o sentimento extraído por meio do Twitter e o comportamento do mercado acionário, Agarwal, Kumar e Goel (2019), ao fazerem um levantamento da literatura que utiliza redes sociais e atividades *on-line* como fontes para conhecer o sentimento do investidor e o comportamento do mercado, constataram que embora

haja um alto crescimento do número de usuários da internet em países emergentes, a maioria dos estudos na área de sentimento do investidor obtido de forma *on-line* tem por foco os países desenvolvidos. Assim, para os autores, há uma necessidade de que haja maior enfoque em como os mercados acionários de países emergentes são impactados pelas informações que são difundidas *on-line*.

Diante do que fora exposto, surgiu-se o interesse de entender como sentimento manifesto de forma *on-line* poderia impactar o comportamento de um mercado emergente, sendo nesse caso o mercado acionário brasileiro. Assim, chegou-se ao seguinte questionamento: **Qual é associação do sentimento do investidor, manifesto via Twitter, com o retorno e o volume negociado do IBOVESPA no mercado acionário brasileiro?**

## 1.2 Objetivos

### 1.2.1 Objetivo geral

Esse trabalho tem por objetivo geral avaliar a relação do sentimento dos investidores, identificado por meio de *tweets*, com os retornos e os volumes negociados do IBOVESPA.

### 1.2.2 Objetivos específicos

Em busca do alcance do objetivo geral, persegue-se neste estudo os seguintes objetivos específicos:

- a) Verificar a associação entre o sentimento do investidor por meio dos *tweets* e o retorno do IBOVESPA;
- b) Analisar a relação entre o volume de *tweets* e o volume negociado do IBOVESPA.

## 1.3 Justificativa e Contribuições

A busca por verificar a relação entre o sentimento dos investidores e o comportamento do mercado acionário tem apresentado uma série de trabalhos que confirmam essa associação, tais quais os estudos de Barberis, Shleifer, e Vishny (1998), Lee, Jiang e Indro (2002), Brown e Cliff (2004, 2005), e Baker e Wurgler (2006, 2007). Dessa maneira, para Baker e Wurgler (2007) a questão não está mais em verificar se essa relação existe, mas em identificar qual a melhor maneira de medir o sentimento e quais os efeitos desse sentimento sobre o mercado.

Assim, o presente estudo visa identificar uma nova maneira de medir o sentimento do investidor no mercado acionário brasileiro. Isso se faz necessário uma vez que normalmente as pesquisas que focam no comportamento dos investidores costumam utilizar variáveis de mercado como *proxy* para o auferir sentimento, tais como o retorno no primeiro dia de IPO ou o volume de negociações (BAKER; WURGLER, 2006). Há também o uso de alguns métodos ligadas à área comportamental, como questionários e entrevistas (BROWN; CLIFF, 2005).

De tal maneira, a diferença do estudo está no fato de que ele traz uma nova abordagem para o problema, de modo a utilizar mensagens que são postadas na rede social Twitter e extrair o sentimento dessas mensagens para conhecer o sentimento do investidor. O Twitter se configura como uma boa fonte de obtenção de dados, visto que ele é uma das redes sociais mais utilizadas pelo mercado (WEI; MAO; WANG, 2016), além de que ele tem muitos usuários, onde no final do ano de 2018, a rede social teve em média 321 milhões de usuários ativos mensalmente (TWITTER, 2018).

Ademais, buscou-se verificar se um índice composto dessa forma é viável para o contexto brasileiro, visto que existem diversos estudos que mostram a viabilidade de tal índice em mercados desenvolvidos, especialmente no que se refere ao mercado norte-americano (MAO *ET AL.*, 2012; WEI; MAO; WANG, 2016; OLIVEIRA; CORTEZ; AREAL, 2017, DAS; BEHERAA; KUMARB, RATHA, 2018; RUAN; DURRESI; ALFANTOUKH, 2018).

Assim, uma vez que a maior parte dos trabalhos verificou como o Twitter impacta o maior mercado desenvolvido das Américas, optou-se por estudar o maior mercado emergente das Américas, isso porque, conforme Agarwal, Kumar e Goel (2019), há uma limitação na literatura no que se refere ao estudo de como as atividades *on-line* impactam mercados acionários emergentes, assim, o trabalho vem como uma forma de preencher esse *gap* existente.

Outro ponto que ressalta a relevância da pesquisa é que utilizar atividades *on-line* como ferramenta para conhecer o sentimento dos investidores é algo pouco explorado na literatura nacional, havendo uma limitação no que se refere a como tais atividades afetam o mercado brasileiro. Nota-se ainda que, até onde se pôde observar, não existem outros estudos que busquem verificar como o sentimento manifesto por meio das redes sociais impacta o comportamento do mercado acionário brasileiro, não havendo indícios de pesquisas que tenham utilizado um índice de sentimento extraído por meio do Twitter, mostrando assim a originalidade da pesquisa no contexto nacional.



O presente estudo espera trazer contribuições para a literatura existente, aumentando a compreensão de como o sentimento manifesto por meio de redes sociais afetam certos movimentos do mercado brasileiro, como o retorno e o volume negociado. Para além, o estudo busca trazer contribuições na área de finanças comportamentais, uma vez que, caso confirmada a relação entre o sentimento dos investidores e os movimentos no mercado acionário brasileiro, isso poderá trazer indícios de que vieses cognitivos são capazes de causar ineficiências de mercado de curto prazo no mercado, assim como no estudo de Schnusenberg e Madura (2001).

Em um contexto de mercado, o estudo busca trazer contribuições práticas para os investidores, mostrando indícios de que o sentimento extraído por meio do Twitter pode ser útil na predição do retorno das ações. Assim, acredita-se que utilizar o sentimento dos investidores como estratégia de investimento pode ser uma ferramenta auxiliar na obtenção de retornos anormais, uma vez que Wei, Mao e Wang (2016) e Renault (2017) trouxeram evidências de que isso seria possível, ao menos em mercados desenvolvidos. Dessa maneira, busca-se mostrar que há potencial em se utilizar as atividades *on-line* para tentar obter retornos anormais no mercado acionário brasileiro.

Os resultados obtidos demonstram que as atividades que acontecem no ambiente *on-line* podem estar associadas aos movimentos que ocorrem no mercado acionário brasileiro, no que se refere aos seus retornos e ao volume de negociações. Especificamente, foi visto que há relação entre o sentimento dos investidores extraído por meio do Twitter e o retorno no mercado brasileiro, além de também haver relação entre o volume de *tweets* e o volume de negócios nesse mercado.

Outrossim, o estudo mostra como as informações, mesmo elas estando em um ambiente *on-line*, como é o caso das mensagens que são postadas no Twitter, têm papel importante no mercado, uma vez que essas podem ser responsáveis por causarem vieses cognitivos que estão associados a ineficiências temporárias no mercado. Ademais, o estudo contribui para a extensão da literatura, uma vez que demonstra como atividades *on-line* estão associados aos movimentos do mercado em um país emergente, como é o caso do Brasil, uma vez que a literatura acerca disso ainda é precária. Para além, o estudo traz indícios de que possivelmente o monitoramento do Twitter pode ser utilizado como ferramenta auxiliar em estratégias de investimento.

## 2 REFERENCIAL TEÓRICO

### 2.1 Hipótese do Mercado Eficiente e Racionalidade Limitada do Investidor

O anseio por entender o processo de formação do preço dos títulos presentes no mercado acionário fez com que diversos autores, ao longo dos anos, se debruçassem sobre investigações empíricas de modo a conhecer o que move os preços e tentar identificar padrões em suas oscilações. Tais investigações culminaram no que atualmente é tido como a Hipótese do Mercado Eficiente (HME).

Um dos primeiros autores a contribuir para a ideia de que o mercado seria eficiente foi o matemático Louis Bachelier (1900), o qual foi responsável por desenvolver o primeiro teste no comportamento dos preços de *commodities*. O autor trouxe à tona a questão da aleatoriedade presente nos preços, a qual é conceituada por Fama (1970) como *random walk*, onde as informações disponíveis refletem na mudança dos preços dos títulos, sendo as mudanças sucessivas independentes, e distribuídas de forma idêntica.

A Hipótese do Mercado Eficiente trouxe consigo a ideia de que os preços estão ligados às informações, de modo que a eficiência informacional de um mercado é quem definirá o preço de um ativo. Roberts (1967) foi pioneiro ao trazer sugestões sobre essa noção de eficiência informacional, todavia, Fama (1970) foi o responsável por a estruturar.

Fama (1970) assevera que para o mercado ser eficiente é preciso que as informações disponíveis nele reflitam em sua totalidade, a qualquer momento, o preço que se deve pagar pelos títulos que nele circulam. Nesse sentido, nenhum investidor seria capaz de obter retornos anormais, de modo que a única forma de obter um maior retorno seria ao investir em um ativo de maior risco.

Para Fama (1970), três condições básicas fariam com que o mercado fosse eficiente, sendo essas: não haver custos para transação de títulos negociáveis; todos os participantes do mercado possuírem, gratuitamente, toda a informação acerca do mercado, e essa estar disponível a todo o momento; e a concordância de todas as partes com o encadeamento existente entre a informação atual e preço atual, e com as distribuições de preços futuros para cada título. Essas condições são suficientes para que o mercado seja eficiente, todavia elas não são essenciais, de modo que essas condições não precisam ocorrer de forma totalmente exata para que o mercado seja eficiente.

Fama (1970) também dividiu essa eficiência em três categorias, sendo essas: a forma fraca, a forma semiforte e a forma forte. Na forma fraca, o conjunto de informações disponíveis

corresponde aos preços e retornos passados; na forma semiforte há a inclusão das informações publicamente disponíveis; e na forma forte, há a abrangência de informações públicas e informações privadas (também conhecidas como informações privilegiadas).

Com o passar dos anos, a eficiência de mercado tem sido amplamente testada. O próprio Fama, antes de lançar a sua revisão de literatura em 1970, foi um dos primeiros a desenvolver um trabalho sobre a Hipótese do Mercado Eficiente. Fama *et al.* (1969) tiveram por objetivo analisar a reação dos preços das ações em casos de eventos divididos. Os resultados mostraram que os preços das ações aumentavam frente às novas informações, confirmando a hipótese de que o mercado seria eficiente.

Bernard e Thomas (1990) também buscaram identificar se a eficiência do mercado era verídica. O estudo analisou 2626 empresas, verificando trimestralmente os anúncios financeiros, de 1974 a 1986. Os autores notaram que há correlação positiva dos preços no curto prazo, todavia, no longo prazo (para mais de três trimestres) a correlação dos preços se torna negativa. Nesse sentido, de maneira semelhante, Fama (1991), embora não invalide a Hipótese do Mercado Eficiente, afirma que o previsível componente dos retornos se dá por uma pequena variância diária, semanal e mensal. Mas há uma variância superior em retornos de longo prazo. Todavia, também existem evidências, em menor número, de que a previsibilidade de retornos passados vem de retornos de longo prazo.

No que tange aos testes do nível de eficiência dos mercados, conforme Jensen (1978), Camargos e Barbosa (2003), e Malkiel (2003), a literatura clarifica que o mercado não é capaz de ser totalmente eficiente, e que a forma semiforte seria aquela à qual se acha a maior quantidade de resultados favoráveis, trazendo assim um maior suporte teórico.

Visto que uma série de estudos surgiu de modo a questionar até que ponto o mercado seria eficiente, tais como os supracitados, um outro ponto que passou a ser amplamente debatido foi à questão da racionalidade do investidor, a qual é intrínseca a Hipótese do Mercado Eficiente. Nesse sentido, durante muito tempo, acreditou-se que em um mercado eficiente os investidores seriam totalmente racionais, todavia, estudos desenvolvidos na área da psicologia social, com enfoque na área de finanças comportamentais e na tomada de decisão, como o de Kahneman e Tversky (1979), mostram que os seres humanos são dotados de racionalidade limitada, de modo que aspectos comportamentais e psicológicos têm influência sobre a tomada de decisão, sendo esses capazes de influenciar os movimentos do mercado, tais como os retornos das ações, o volume de negociações etc.

O estudo de Kahneman e Tversky (1979) abre espaço para a afirmação de que há uma racionalidade limitada nos seres humanos, de modo que as decisões tomadas pelos participantes

do mercado não serão tomadas totalmente baseadas na razão, sendo também dotadas do viés do investidor. Todavia, isso não implica na ineficiência do mercado, mas na ideia de que ele não pode ser eficiente em sua totalidade, fato que é mostrado anos após no trabalho de Camargos e Barbosa (2003), onde os autores mostram que a forma fraca e a forma semiforte da eficiência de mercado seriam as que apresentam mais evidências. Assim, visto que há essa lacuna na HME, acredita-se então que certos investidores presentes no mercado podem se beneficiar desses aspectos comportamentais e psicológicos a fim de obterem retornos anormais.

O próprio Fama (1991), que em 1970 disseminou a ideia de que os investidores seriam totalmente racionais, afirma, anos depois, que existem bolhas irracionais no mercado, as quais podem ocorrer devido à reação exagerada frente a notícias sobre determinada empresa. Para o autor, não é possível separar as bolhas irracionais daquelas variações racionais esperadas nos preços das ações, de modo que esse tipo de bolha irracional aumenta a variância dos retornos no longo prazo. Assim, supõe-se que essa irracionalidade é um dos pontos que fazem com que os mais diversos modelos usados para captar o preço dos ativos não o consigam entender a sua formação em sua totalidade.

Em posição contrária, autores como Malkiel (2003) não acreditam que aspectos psicológicos tenham a capacidade de impactar o mercado, sendo ele mais eficiente e menos previsível do que muitos acreditam, onde o passado não define o futuro, ao menos no longo prazo. Para o autor, a ideia de eficiência está conectada à incapacidade dos investidores de obterem retornos acima da média sem possuírem um risco igualmente acima da média. Ele assevera ainda que, mesmo que diversos participantes do mercado sejam extremamente irracionais, ou ainda que os preços sejam bastante voláteis, isso não implica na ineficiência do mercado. Dessa maneira, para o autor, informações de curto prazo não são fontes confiáveis, visto que mesmo que no curto prazo pareça haver correlação entre os retornos das ações, no longo prazo, esses retornos se reverteriam.

Embora a questão de as ineficiências de mercado não serem importantes para autores como Malkiel (2003), visto que, em uma imagem maior, o mercado continuaria sendo eficiente, para alguns pesquisadores, conhecer o que causaria essas ineficiências é algo necessário. Nesse sentido, para muitos, essas ineficiências ocorrem devido a fatores psicológicos e vieses cognitivos, especialmente no curto prazo. Um dos estudos que buscaram identificar o que causaria ineficiências de curto prazo foi o de Schnusenber e Madura (2001), no qual os autores analisaram as reações *overreaction* e *underreaction* no mercado frente a novas informações. Segundo os autores, os vieses cognitivos dos investidores seriam os responsáveis pelo ajuste dos

preços as crenças sobre expectativas de fluxo de caixa futuro, fato que tem a capacidade de causar uma má precificação dos ativos no curto prazo.

Vieses cognitivos são tidos pela literatura como um dos principais causadores do distanciamento entre os preços dos ativos e seus valores intrínsecos, os quais também são responsáveis por darem margem a supostas previsibilidades. Esses vieses causam desvios na racionalidade dos investidores, e esse afastamento da racionalidade é definido por Brown e Cliff (2004) como o sentimento do investidor. Segundo os autores, o sentimento é o viés do especulador, sendo assim o excesso de otimismo ou pessimismo dos investidores.

Conforme Brown e Cliff (2004), a Teoria do Mercado Eficiente permaneceu intocável durante algum tempo, de modo que parecia não ser possível prever retornos, uma vez que eles refletiriam apenas as expectativas futuras dos resultados das entidades, as quais seriam atualizadas frente ao ambiente informacional, todavia, com o passar do tempo, e, mais precisamente a partir da década de 80, diversos autores começaram a apresentar evidências de que existem anomalias que fazem os ativos se distanciarem de seus valores intrínsecos, achando assim supostas previsibilidades.

Ainda na década de 80, Simon (1983) lança uma obra na área de psicologia cognitiva a qual buscou conhecer melhor a mente humana, trazendo problemas relacionados à tomada de decisão, a racionalidade e a racionalidade limitada. O termo “racionalidade limitada”, o qual aparece por diversas vezes nesse livro, fora cunhado pelo autor desde a década de 50, dando base para diversas pesquisas na área comportamental, e mais especificamente no que diz respeito ao processo de tomada de decisão.

No que se refere à limitação da racionalidade na tomada de decisão, Tversky e Kahneman (1974) expõem que os seres humanos possuem uma série de vieses cognitivos relacionados ao processo decisório, os quais provêm da confiança no uso de heurísticas. Heurísticas são simplificadoras do processo decisório que são usadas de forma voluntária pela mente humana. Essas simplificações podem ser úteis em certos casos, mas também têm a capacidade de levar à julgamentos e decisões enviesados e errados.

Uma vez que uma série de estudos apontam que o ser humano possui uma racionalidade limitada e que suas decisões são compostas por vieses, então, diversos pesquisadores passaram a questionar a ideia contida na Hipótese do Mercado Eficiente de que os investidores seriam totalmente racionais. Assim, ideias como as de Friedman e Friedman (1953), de que o estudo da não racionalidade não seria necessário, visto que investidores não racionais seriam expulsos do mercado por investidores racionais, passam a ser colocadas de lado, e surgem autores como Barbaries e Thaler (2003), os quais acreditam que, uma vez que desvios da racionalidade são

intrínsecos ao comportamento humano, então tais desvios devem ser incorporados à análise econômica, pois essas distorções têm importantes impactos econômicos.

No mercado acionário, por muito tempo fora deixado de lado o impacto causado pelos desvios cognitivos na tomada de decisão que fazem com que os valores dos ativos apresentem um espaçamento entre o preço e seu valor intrínseco, de modo que os modelos de precificação de ativos só buscavam conhecer o que era capaz de ser captado na esfera racional. Todavia, a partir do momento em que os pesquisadores passaram a considerar a parte que desvia da racionalidade, denominada sentimento do investidor (BARBERIS; SHLEIFER; VISHNY, 1998), constatou-se que grande parte das anomalias encontradas na eficiência do mercado são plenamente explicadas pelo sentimento (STAMBAUGH; YU; YUAN, 2014).

Ademais, estudos como o de Menkhoff (1998) surgiram para salientar que o comportamento não totalmente racional impacta o mercado como um todo, e que o sentimento tem influência tanto sobre aqueles participantes de mercado conhecidos como “*noise traders*”, quanto sobre aqueles considerados “arbitradores racionais”. “*Noise trader*” é um termo que fora introduzido por Black (1986), o qual, segundo o autor, se refere a um *trader* de ações que não tem informações privilegiadas e toma decisões irracionais de investimento.

Segundo Barberis, Shleifer e Vishny (1998), existem duas reações que comumente acontecem nos mercados acionários, sendo essas a *underreaction* e a *overreaction*. Conforme os autores, a *underreaction* está ligada a demora que o mercado apresenta para reagir ou a não reação à certas notícias, como a do anúncio de ganhos, por exemplo. Por outro lado, a *overreaction* é a reação exagerada a notícias, sejam elas boas ou más, que ocorrem mesmo que essas não estejam ligadas a geração de fluxo de caixa da entidade. Essas reações desafiam a Hipótese do Mercado Eficiente, pois elas fazem com que seja possível, por vezes, prever o mercado, de modo que certos investidores possam obter retornos anormais.

Conforme Barberis, Shleifer e Vishny (1998), a *overreaction* normalmente é conduzida pela heurística da representatividade, a qual é trazida por Tversky e Kahneman (1974) como a tendência a utilizar estereótipos no processo decisório, de modo a avaliar a probabilidade de que um evento ocorra pela simples similaridade com algum fato. Por exemplo, em finanças, um investidor pode optar por investir em determinada ação que obteve grandes retornos no período passado sem avaliar os fundamentos da entidade ou se ela possui capacidade de gerar esses mesmos retornos no futuro. Por outro lado, a *underreaction* costuma estar ligada a outro fenômeno psicológico conhecido como conservadorismo, onde os indivíduos costumam atualizar seus modelos de tomada de decisão de forma lenta, mesmo frente a novas evidências (EDWARDS, 1968).

Uma vez que há indícios de que o mercado não é totalmente racional e que reações individuais possam o afetar, Lee, Jiang e Indro (2002) desenvolveram um estudo de modo a identificar a relação entre a volatilidade do mercado, o excesso de retorno e o sentimento do investidor. Para isso, os autores utilizaram o índice Dow Jones Industrial Average (DJIA), o Standard and Poor's 500 (S&P500) e o NASDAQ, de 1973 a 1995. Conforme os resultados, o sentimento contemporâneo foi um fator importante para determinar os retornos excedentes e a volatilidade condicional em todos os índices. Notou-se ainda que as empresas menores (as quais compõem o índice NASDAQ) foram as mais impactadas pelo sentimento, trazendo indícios de que o tamanho da entidade pode ter efeito sobre a forma com que ela se movimenta no mercado.

As evidências do estudo apontam que a volatilidade aumenta (diminui) quando os investidores se tornam mais pessimistas (otimistas), de modo que mudanças no sentimento estão negativamente correlacionadas com a volatilidade do mercado. Já os retornos mais altos (baixos) estão associados à diminuição (aumento) da volatilidade, decorrente do aumento (diminuição) do sentimento positivo no mercado.

Ainda no que se refere a como o sentimento impacta o mercado, Brown e Cliff (2004) investigaram como sentimento age no processo de formação de preços para as carteiras de ações no mercado americano. Os autores revelam que há forte relação entre o sentimento do investidor e o retorno do mercado, todavia, no curto prazo, os retornos que seriam responsáveis por explicar parte da variação do sentimento. Adicionalmente, os autores afirmam que um alto nível de otimismo no tempo contemporâneo faz com que os preços atuais sejam elevados, e isso acarreta baixa nos retornos futuros subsequentes.

De maneira semelhante, Brown e Cliff (2005) mostram que há relação entre o sentimento e o retorno das ações no longo prazo, todavia no longo prazo o sentimento é tido como um preditor dos retornos futuros, tendo a capacidade de explicar os desvios do valor intrínseco dos ativos. Também foi possível perceber que quando os investidores estão otimistas, isso gera um valor de mercado excedente ao valor intrínseco para os ativos, e que esses erros causados pelo excesso de otimismo se revertem ao longo dos anos.

Baker e Wurgler (2006, 2007) encontraram que o sentimento impacta o mercado, e que em épocas de sentimento otimista, os retornos são altos, e os retornos subsequentes serão baixos, especialmente para empresas que possuem menor tamanho. Da mesma maneira, quanto menor for o sentimento, maior será o retorno subsequente. O trabalho de Baker e Wurgler (2006) se destaca por eles terem desenvolvido um índice, o qual é amplamente utilizado até os

dias atuais, para medir o sentimento do investidor baseado em seis *proxies* as quais eram comumente utilizadas pela literatura, sendo essas: o volume negociado; o prêmio de dividendos; o desconto de fundo fechado; o número e os retornos do primeiro dia de IPOs (*initial public offering*); e a participação acionária em novas emissões.

Embora os trabalhos de Brown e Cliff (2004, 2005) mostrem que o sentimento tem um maior efeito sobre o mercado no longo prazo, o estudo de Kling e Gao (2008) vem contradizer essa convicção, trazendo indícios de que a amostra e a forma como o sentimento é obtido são capazes de influenciarem os resultados. Kling e Gao (2008), ao estudarem o impacto do sentimento (auferido por meio de questionário) sobre o mercado chinês, constataram que no longo prazo a relação entre o sentimento e o retorno das ações não existe. Todavia, no curto prazo essa relação existe e é relevante, de modo que ela pode causar ineficiências momentâneas de mercado e supostas previsibilidades.

Segundo Verma e Verma (2008), o sentimento não é tão irracional quanto se imagina, pois, ao estudarem o mercado americano de 1988 a 2003, notou-se que em muitos casos o sentimento tende a ser conduzido por fatores de risco racionais, mesmo que esse continue a ser um viés do investidor, especialmente no caso de investidores institucionais, os quais costumam ser mais racionais do que investidores individuais. Assim, para os autores, a ligação do sentimento e dos preços corresponde a uma combinação de perspectivas racionais e expectativas que não podem ser plenamente justificadas pelas informações disponíveis no mercado.

Um exemplo de sentimento que afeta o mercado e não pode ser totalmente explicado pelas informações disponíveis está representado no estudo de Edmans, García e Norli (2007). Conforme os autores, existe relação entre os resultados dos jogos de futebol e o retorno das ações, em que um resultado negativo em uma partida de futebol acarreta retornos negativos, especialmente em pequenas ações e quando os jogos são mais importantes, como em uma copa do mundo.

O estudo Kaplanski e Levy (2010) também mostra como o mercado pode ser influenciado por sentimentos os quais não podem ser plenamente explicados. Nesse estudo, os autores evidenciam que nos dias em que ocorrem acidentes aéreos, as companhias aéreas tendem a perder grande parte do seu valor de mercado, sendo essa perda de em média \$ 60 bilhões por desastre aéreo, mesmo que os fundamentos das entidades não tenham mudado, ou que o acidente não seja grave e não vá causar grandes problemas na expectativa de fluxo de



caixa futuro da entidade envolvida no caso ou nas demais entidades do ramo. Notou-se ainda que no dia da ocorrência do acidente há um aumento significativo na volatilidade das ações, apresentando retornos negativos de forma exagerada, os quais tendem a se reverter em dois dias.

Quando se fala em sentimento impactando o mercado, pode-se perceber que muitos autores ainda buscam testar a eficiência do mercado e os pressupostos que a acompanham frente ao sentimento dos investidores. Nesse sentido, Kim, Ruy e Seo (2014) verificaram se o sentimento seria capaz de impactar a premissa da plena concordância. Os achados do estudo apontam que o sentimento causa discordância entre os retornos atuais e futuros no mercado acionário. Ademais, foi visto que essa discordância entre investidores pode e deve ser utilizada como estratégia de investimento em períodos de sentimento otimista, quando há um otimismo em excesso por parte dos investidores. Especificamente, constatou-se que investidores devem vender ações que estejam sob o enfoque de analistas, os quais “prometem” altas taxas de crescimento do lucro por ação (LPA) dessas entidades no longo prazo. Notou-se que essa estratégia trás retornos melhores do que a *buy-and-hold*.

Frente aos diversos estudos apresentados, as evidências confirmam que há relação entre o sentimento e os movimentos do mercado, todavia, com o passar do tempo, alguns autores passaram a questionar se o sentimento tem impacto sobre o mercado como um todo ou apenas sobre setores específicos. Dessa maneira, Uygur e Tas (2014) constataram que o sentimento impacta os setores presentes no mercado de Istambul de maneira distinta, de modo que mudanças no sentimento do investidor tem mais influência na volatilidade condicional de indústrias, bancos e do setor de alimentos e bebidas, quando comparado a outros setores, como o de varejo ou o de telecomunicações.

Debata, Dash e Mahakud (2018) demonstram como a liquidez é impactada pelo sentimento. Os autores estudaram doze países emergentes, de 2002 a 2015 sendo esses: Brasil, China, Índia, Indonésia, México, Filipinas, Polônia, África do Sul, Coreia do Sul, Rússia, Tailândia e Turquia. Para isso, eles utilizaram o Índice de Confiança do Consumidor (CCI) como *proxy* de sentimento específico do país. Os resultados demonstram que o sentimento do país tem impacto na liquidez do mercado nacional e pode ser o causador da sua variação. Também foi visto que o sentimento internacional, especialmente dos Estados Unidos e de países europeus podem impactar mercados emergentes. Os resultados ainda apontam que em períodos

de otimismo o volume negociado tende a crescer. Por outro lado, em períodos de pessimismo a liquidez tende a cair.

Visto que o sentimento fora tido como fator de risco por diversos autores, Bathia e Bredin (2018) constataram que, além de ser um fator de risco, o sentimento tem a capacidade de melhorar os modelos de precificação de ativos, sendo um fator que se mostrou relevante em todos os modelos analisados, os quais foram: o modelo CAPM, o modelo Fama-French (FF) de três fatores, o modelo FF incorporado com o fator de liquidez, o modelo FF incorporado com fator o *momentum* e modelo FF incorporado com a liquidez e o fator *momentum*.

Dentro dessa perspectiva de que o sentimento impacta o mercado, sendo ele um fator de risco, e diante dos diversos estudos acima apresentados, acredita-se que o retorno sofre influência do sentimento, bem como outros movimentos, tais como a liquidez e a volatilidade. Dessa maneira, têm-se a primeira hipótese da pesquisa, na qual se acredita que:

**Hipótese 1:** Quanto mais positivo é o sentimento do investidor, manifesto por meio do Twitter, maior é o retorno do IBOVESPA.

Assim, o ponto que deve ser debatido adiante não é mais se o sentimento impacta o mercado, mas sim qual a melhor maneira de captar esse sentimento, e como ele o faz.

## 2.2 Novas Maneiras de Medir o Sentimento

Desde a década de 50, quando Simon (1957) lançou o livro “*Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*”, a limitação da racionalidade humana passou a ser mais debatida. No entanto, fora próximo à década de 80 que surgiram estudos comportamentais com enfoque na área de finanças que utilizaram essa ideia, como o trabalho de Kahneman and Tversky (1979).

Os limites da racionalidade humana levaram a uma série de estudos acerca de como funciona o processo decisório e como essa limitação impacta o mercado acionário. Essa limitação, ou não racionalidade completa foi denominada de sentimento do investidor, o qual já foi provado pela literatura como sendo um fator determinante dos movimentos do mercado. Autores como Baker e Wurgler (2007) asseveram que já está claro que o sentimento afeta o mercado, de modo que o foco de novos estudos deveria estar em como medir o sentimento do investidor e quantificar os seus efeitos.

O próprio estudo de Baker e Wurgler (2007) mostra que existem diversas maneiras de medir o sentimento. Normalmente os comportamentalistas preferem pesquisas de opinião com os investidores. Já os pesquisadores da área de finanças clássicas acreditam que variáveis como o fluxo dos fundos mútuos, o volume de negociação, o prêmio de dividendos, o desconto em fundos fechados, a volatilidade implícita da ação, os retornos do primeiro dia de IPO, o volume do IPO, a participação em novas emissões e a negociação com informações privilegiadas seriam variáveis que mediriam de forma mais precisa o sentimento do investidor.

Muito embora se possa imaginar que a pesquisa de opinião com os investidores seja a melhor forma de captar o sentimento do investidor dentre todas as formas disponíveis, Baker e Wurgler (2007) afirmam que esse tipo de pesquisa não costuma ser bem vista, uma vez que não há garantias de que a forma com que o participante responde às perguntas é como ele realmente se comportaria na realidade. Assim, pesquisas de opinião usualmente são criticadas por pesquisadores da área de finanças clássicas, e variáveis de mercado são julgadas pelos comportamentalistas como não representativas para a captura do sentimento do investidor.

Visto que há uma necessidade de se encontrar novos métodos de captar o sentimento do investidor, e sabendo ainda que desde os anos 2000 há um crescimento do fenômeno da internet e das mídias sociais, acredita-se que o ambiente *on-line* pode ser uma importante fonte para a obtenção de dados a fim de conhecer o mercado e captar o sentimento dos investidores de forma direta e instantânea (FAN; GORDON, 2014). Nesse sentido, diversas pesquisas, bem como as de Wei, Mao e Wang (2016), Atkins, Niranjana e Gerding (2018), Das, Behera e Rath (2018), e Chen e Lo (2019), têm surgido com o intuito de captar o sentimento dos investidores por meio da internet. Ao revisar a literatura referente ao desenvolvimento de índices de sentimento por meio de atividades *on-line*, percebe-se que os autores têm buscado conhecer o sentimento dos investidores por meio de diversas fontes, sendo essas: *sites*, como o Yahoo! Finance ou por jornais que divulgam suas notícias *on-line*, tal qual o The Financial Times; por meio de redes sociais, tais como o Facebook e o Twitter; e por meio de ferramentas de pesquisa, assim como a ferramenta de busca do Google.

No que se refere à busca por sentimento por meio de *sites*, Kim e Kim (2014) desenvolveram um índice para medir o sentimento do investidor através de mensagens extraídas do Yahoo! Finance. Ao todo os autores trabalharam com 32 milhões de mensagens acerca de 91 empresas americanas, de 2005 a 2010. Para controlar os efeitos do sentimento, os autores adicionaram o tamanho das entidades e o *Book-to-Market* (BM). Os resultados mostram pouca evidência de que o sentimento do investidor prevê retornos futuros, mas foi visto que os retornos

presentes influenciam o sentimento futuro. Os achados também apontam que o sentimento não tem poder de prever a volatilidade e o volume de negociações.

Quanto ao uso de notícias retiradas de *sites*, Silva (2017), utilizando o IBOVESPA, e Galdi e Gonçalves (2018), lidando com o IBOVESPA e o IBRA, propuseram índices para medir o sentimento do mercado brasileiro, sendo ambos constituídos por meio do *site* do jornal Valor Econômico. Os resultados de Silva (2017) exibem que, referente ao retorno, o sentimento do dia, bem como o sentimento de quatro dias atrás se relaciona com o retorno contemporâneo. Assim, em períodos de otimismo exagerado por parte dos investidores ao reagirem às notícias, os retornos são maiores, mas eles tendem a se reverter com o tempo. Todavia, no que se refere à volatilidade, foi visto que o sentimento pessimista é capaz de reduzi-la. Já Galdi e Gonçalves (2018) demonstraram que o mercado brasileiro avalia com maior peso palavras negativas ou de incerteza divulgadas pela mídia do que palavras positivas, sendo a associação entre tais palavras e a rentabilidade negativa, e positiva no que se refere a associação com a volatilidade.

Em relação ao mercado americano, Atkins, Niranjana e Gerding (2018) o investigaram, de 09 de setembro de 2011 a 07 de setembro de 2012, utilizando notícias disponíveis no site da Reuters. Os achados comprovam que as notícias financeiras têm um maior poder de predição da volatilidade do mercado do que do preço de fechamento dos ativos, mostrando assim que as informações contidas nas notícias influenciam o sentimento dos investidores e afetam o mercado financeiro.

Ainda no que tange a captar como as atividades *on-line* impactam o mercado, Shen, Liu e Zhang (2018) estudaram um total de 26 índices internacionais. Para isso, os autores utilizaram dados de felicidade diária retirados do *site* da Hedonometer, o qual também utiliza dados de redes sociais como o Twitter, e constataram que quando há um maior nível de felicidade, há também um maior nível de assimetria nos retornos, provando assim haver uma estreita relação entre o desempenho do mercado acionário e os sentimentos expostos *on-line*.

Visto que as redes sociais têm sido utilizadas diariamente por grande parte da população mundial, uma série de autores focou sua atenção nas mesmas, tais como no Twitter, no Facebook e em redes sociais próprios para investidores, assim como a StockTwits. Nesse sentido, Mao *et al.* (2012) buscaram identificar a correlação das ações presentes no índice S & P 500 com os dados disponíveis no Twitter. Por meio da análise do volume diário de *tweets* que mencionavam o índice, os autores descobriram que ele se correlaciona significativamente com diversos indicadores financeiros das ações, tais como o volume de negociações, indicadores de

desempenho financeiro etc. Também foi visto que os *tweets* podem ser utilizados para prever os preços do mercado.

No que se refere ao uso do Twitter, Wei, Mao e Wang (2016) analisaram se haveria relação entre o pico de volume no Twitter e o preço das ações. Para isso, os autores coletaram dados no Twitter de 1 de agosto de 2013 a 6 de agosto de 2014. O pico de volume está ligado a um número de *tweets* relacionados a determinada ação em certo dia o qual é maior do que o valor médio nos últimos dias. Os achados indicam que a volatilidade implícita aumenta antes de um pico de volume no Twitter, e diminui rapidamente após a sua ocorrência. Notou-se ainda que logo após um pico de volume no Twitter há um tipo de *overreaction*, de modo que, quanto maior o número de *tweets* positivos sobre determinada ação, haverá uma maior procura pela mesma e o seu preço irá subir, fazendo com que as ações fiquem superfaturadas.

A grande diferença do estudo de Wei, Mao e Wang (2016) para os demais é que os autores propuseram uma estratégia de modo a aproveitar o sentimento de alta no mercado. Conforme a estratégia, mesmo em um cenário mais conservador, a carteira criada pelos autores obteve um retorno de 34,3% em um ano, enquanto o S & P 500 aumentou 12,8% no mesmo período, mostrando assim que o sentimento dos investidores pode ser utilizado como ferramenta de auxílio na obtenção de retornos anormais.

Nisar e Yeung (2018) analisaram se os *tweets* poderiam ser utilizados para medir o volume de negociações e o preço de fechamento. Assim, eles desenvolveram o estudo durante o período em que estava havendo eleições nacionais no Reino Unido. Os autores pegaram *tweets* que estavam relacionados às eleições, de 4 a 9 de maio (a eleição ocorreu no dia 5 de maio). Ao todo, a amostra contou 60.000 *tweets*. Os resultados apontam que o sentimento e o fechamento dos preços se relacionam. Também foi visto que existem correlações que podem dar resultados promissores entre o retorno e o sentimento, onde o sentimento do dia  $t$  irá impactar o mercado em  $t+3$ . Por outro lado, o volume de *tweets* não aparenta possuir relação com o volume de negócios ou o fechamento dos preços. Assim, os autores acreditam que o Twitter tem potencial como ferramenta na previsão dos retornos.

Ainda no que se refere a relação entre as atividades realizadas no Twitter e os movimentos das ações, Oliveira, Cortez e Areal (2017) levantaram 31 milhões de mensagens relacionadas a todas as ações negociadas nos mercados dos EUA (cerca de 3800 ações), de dezembro de 2012 a outubro de 2015. As ações foram alocadas em carteiras criadas por tamanho e por setor. Constatou-se que o índice de sentimento extraído do Twitter foi útil na

previsão dos retornos do índice S & P 500. Também foi visto que o índice tem maior poder explicativo sobre ações menores e sobre alguns setores, como os de energia, telecomunicações e de alta tecnologia. Por outro lado, foi visto pouco poder explicativo por parte do índice de sentimento na previsão do volume de negócios e da volatilidade.

Dessa maneira, embora no estudo de Nisar e Yeung (2018), o volume de *tweets* aparenta não se relacionar com outras variáveis para além dos retornos, fato que pode ser explicado pelo curto período da amostra, e no de Oliveira, Cortez e Areal (2017) ter sido visto pouco poder explicativo do sentimento na previsão do volume negociado, trabalhos como os de Mao *et al.* (2012) e Wei, Mao e Wang (2016), os quais foram aludidos anteriormente, mostram que o volume diário de *tweets* possui relação com diversas variáveis, como o volume de negociação do mercado, indicadores de desempenho financeiro etc. Destarte, emerge-se a segunda hipótese da pesquisa, onde supõe-se que:

**Hipótese 2:** O volume diário de *tweets* relacionados ao IBOVESPA possui associação positiva com volume negociado no mercado acionário brasileiro.

Visto que o sentimento extraído via Twitter é capaz de impactar o mercado e setores específicos, alguns autores analisaram se esse sentimento seria útil para a previsão de ações específicas. Para isso, Das, Behera e Rath (2018) analisaram as empresas: Google, Microsoft e Apple. Ao todo, os autores trabalharam com 5.060.000 *tweets*. Os resultados mostram que o índice de sentimento desenvolvido é útil na previsão do retorno das ações individuais. De maneira semelhante, Ruan, Durresi, Alfantoukh (2018) levantaram *tweets* das oito empresas com o maior número de *tweets* presentes no índice S & P 500, no período de 01/01/2015 a 31/08/2015. Os autores ponderaram os *tweets* pelo número de seguidores do autor. Assim, foi visto que há relação entre o sentimento e os retornos das ações, e que esse tipo de ponderação auxilia a ampliar a correlação entre o sentimento e os retornos, mostrando que esse método é melhor do que ao tratar os autores todos de forma igual.

No que se refere ao uso de *tweets* para a análise de sentimento, diversos autores também utilizaram *tweets* provenientes da rede social StockTwits, a qual é uma rede social direcionada para investidores. Foi o caso dos estudos de Renault (2017) e de Al-Nasseria e Ali (2018).

Renault (2017) levantou 59.598.856 mensagens de 239.996 usuários da StockTwits. As evidências demonstram que o sentimento do investidor auxilia na previsão diária dos retornos.

Especificamente, o estudo de Renault (2017) revela que a primeira mudança na primeira meia hora no sentimento do investidor tem a capacidade de prever o retorno da última meia hora do ETF do S & P 500, ou seja, o sentimento contemporâneo se relaciona com o retorno contemporâneo. O autor também elucida que o efeito do sentimento costuma ser impulsionado pelo sentimento de *traders* novatos, e que tal sentimento pode ser utilizado como estratégia para obtenção de retornos anormais. Ademais, foi visto que a alteração que o sentimento causa nos preços costuma se reverter no próximo dia em que há negociação.

Al-Nasseria e Ali (2018) construíram um indicador de sentimento que visava conhecer a divergência de opinião dos investidores. Os autores analisaram 289.443 mensagens. As evidências revelam que o impacto da divergência é negativo durante os períodos em que o mercado está otimista, e positivo quando o mercado está pessimista. Também foi visto que quanto maior o valor do índice de sentimento, maior o volume de negociações. Ademais, os autores mostraram que formar carteiras com base no índice por eles desenvolvido auxilia na aferição de ganhos anormais.

Ainda que diversos estudos tenham focado no Twitter e na StockTwits, alguns autores também buscaram captar o sentimento por meio de outras redes sociais. Nesse sentido, Siikanen *et al.* (2018) analisam a relação entre o uso do Facebook e a tomada de decisão. Para isso, eles escolheram a ação da empresa Nokia. Os autores coletaram postagens, comentários, curtidas e compartilhamentos do mural do Facebook da Nokia, de 2010 a 2016. Os achados demonstram que investidores menos sofisticados são mais propensos a terem suas cognições impactadas por essas notícias não reguladas, tendo o seu sentimento aumentado, de modo que eles tomam decisões com base nas notícias postadas nas redes sociais, fato que não ocorre com investidores mais sofisticados.

Muito embora existam diversos estudos com enfoque nas redes sociais, uma nova forma de medir sentimento dos investidores tem surgido, sendo essa a captura de sentimento por meio da ferramenta de pesquisa do Google. Como exemplos, têm-se os estudos de Qadan e Nama (2018), Chen e Lo (2019) e Kim *et al.* (2019).

Na investigação de Qadan e Nama (2018), apesar de os autores terem utilizado *proxies* conhecidas pela literatura para medir o sentimento do investidor, tais como o índice de sentimento desenvolvido por Baker e Wurgler (2006), o Índice de Confiança do Consumidor etc, eles optaram por desenvolver uma nova *proxy*, formada por dados de pesquisas no Google, extraídos do Google Trends. As evidências exibem que o sentimento obtido por meio do Google

não possui uma relação significativa com o retorno, todavia, ele possui uma relação positiva e significativa com a volatilidade. Assim, um aumento na atenção que os investidores dão ao mercado de petróleo causa uma maior volatilidade nos preços das ações.

Os demais estudos que consideram o volume de pesquisa no Google também não puderam encontrar relações entre essa *proxy* para o sentimento do mercado e o retorno das ações, todavia, Chen e Lo (2019) encontraram que no mercado chinês esse índice está positivamente relacionado com a volatilidade do retorno, o volume de negócios, e a taxa de rotatividade das ações. Semelhantemente, Kim *et al.* (2019) mostram que, no mercado norueguês, o sentimento extraído por meio do Google pode explicar e prever o volume de negociações, além de poder prever a volatilidade futura.

### **2.3 Demais Fatores que Impactam o Retorno e o Volume Negociado**

Estudos como os de Lee, Jiang e Indro (2002) e Bathia e Bredin (2018) revelam que o sentimento é um fator de risco, de modo que ele possui uma relação linear com o retorno do mercado, todavia, ele não é o único fator que exerce influência sobre ele. Uma série de autores tem tentado entender os retornos das ações ao longo dos anos e identificar fatores que sejam capazes de o explicar, especialmente a partir da década de 60, onde o CAPM fora originado por William Sharpe (1964), John Lintner (1965) e Mossin (1996), sendo esse modelo posteriormente adaptado por Black (1972). Com o passar do tempo, alguns autores encontraram fatores que poderiam aumentar a explicação da *cross-section* dos retornos médios fornecidos pelos  $\beta$ s de mercado, ou poderiam ter um papel tão importante quanto o próprio beta de mercado possui no modelo CAPM. Esses fatores seriam o tamanho das empresas, o valor de mercado, a alavancagem, o *book-to-market* e a relações de lucro/preço (E/P).

Fama e French (1992) investigaram se os fatores encontrados pela literatura eram verdadeiramente úteis na explicação dos retornos. Assim, os autores constataram que não é o  $\beta$  a melhor forma de explicar os retornos, mas sim o tamanho e o *book-to-market*. Ademais, a combinação dessas duas variáveis parece absorver os papéis de alavancagem e do E/P no retorno médio das ações. Frente aos resultados que foram encontrados, os autores desenvolveram um outro estudo no ano posterior, onde eles lançaram o modelo que ficou conhecido como modelo Fama-French de três fatores. Assim, Fama e French (1993) definiram que considerar o mercado, conforme o CAPM, o tamanho das entidades e o índice o BM é uma maneira de melhorar o poder explicativo do CAPM.



Subsequentemente, Carhart (1997) encontra que os mesmos fatores que explicam os retornos das ações são responsáveis por explicar o retorno dos fundos mútuos. Nesse trabalho o autor utiliza um modelo que fora produzido por ele, o qual possui quatro fatores, sendo eles os três fatores de Fama e French (1993) mais um fator adicional que captura a anomalia do efeito *momentum*, o qual fora descoberto por Jegadeesh e Titman (1993), que consiste em que ações vencedoras tendem a permanecer vencedoras, de modo que seguir uma estratégia de *momentum* em ações pode fazer com que se ganhe acima da média (CARHART, 1997). Esse modelo serviu como uma forma de atribuir desempenho e aumentar a capacidade de explicar os retornos das ações.

Após a adição do efeito *momentum* por Carhart (1997), outro fator que fora acoplado e mostrou aumentar o poder de explicação dos modelos de precificação de ativos foi o efeito liquidez, o qual fora adicionado por Keene e Peterson (2007), criando assim um modelo de cinco fatores. Tal modelo se constitui como o modelo que melhor explica o retorno das ações. Nesse sentido, Machado e Medeiros (2011) mostraram que no mercado brasileiro a liquidez é precificada. Segundo os autores, no contexto brasileiro o modelo de cinco fatores é o que apresenta maior poder explicativo.

Fama e French (2015) também desenvolveram um modelo de cinco fatores, adicionando a rentabilidade e o investimento ao seu modelo de três fatores. Todavia, Machado, Faff e Silva (2017), ao analisarem o mercado acionário brasileiro, demonstram que o investimento e a rentabilidade não foram significativos, de modo que o modelo de cinco fatores desenvolvido por Keene e Peterson (2007), que engloba os fatores de risco de mercado, tamanho, B/M, *momentum* e liquidez, possui o melhor desempenho, se mostrando superior aos demais modelos, ao menos no que se refere ao mercado acionário brasileiro. Ademais, conforme os autores, o B/M, o *momentum* e a liquidez são as variáveis que apresentam evidências mais fortes.

Para além dos fatores de risco encontrados pela literatura que possuem relação direta com o retorno, diversos autores apresentam evidências para o mercado brasileiro de outros aspectos que podem estar associados aos retornos das entidades. Nessa perspectiva, Callado *et al.* (2010) mostram que retornos de ações individuais podem estar ligados a fatores macroeconômicos, como a inflação e a taxa de câmbio. De maneira semelhante, Modro e Santos (2015) trazem evidências de que a inflação pode guiar o comportamento dos preços em certas entidades. Para Machado, Gartner e Machado (2017), ao estudarem a relação de longo prazo

entre variáveis macroeconômicas e o IBOVESPA, a inflação não possui um relacionamento com o retorno do mercado, todavia a oferta de moeda, a taxa de juros, o PIB (Produto Interno Bruto), e a taxa de câmbio estariam ligados ao índice de mercado.

Embora haja controversas acerca do efeito da inflação sobre o mercado, Portela e Santos (2018) optaram por verificar se a inflação poderia ser tida como um fator de risco. Para isso, os autores utilizaram o modelo de cinco fatores de Fama e French (2015), no qual a rentabilidade e o investimento são considerados. Os resultados mostram que a adesão da inflação ao modelo de cinco fatores aumenta o poder explicativo dele. Ademais, essa variável fora significativa em 19 das 32 carteiras construídas pelos autores.

Para além do retorno, movimentos do mercado relacionados ao volume negociado também costumam ser estudado. No contexto internacional, Chen (2012) apresenta evidências de que o retorno das ações tem capacidade de prever o volume negociado, todavia a relação entre volume e retorno é positiva apenas em mercado de alta. Quando o mercado está em baixa, a correlação dessas variáveis é negativa.

Souza, Barbedo e Araújo (2018) trazem indícios de que notícias são capazes de influenciar o volume negociado. Foi visto que o volume negociado é influenciado negativamente por notícias referentes a “dias bons” anteriores. Isso ocorre porque usualmente em dias bons a reação positiva ocorrerá no dia atual. Por outro lado, os autores não encontraram indícios de que notícias em dias de sentimento pessimista irão impactar o volume negociado. Todavia, notícias negativas em dias bons podem explicar parte da queda do volume de negócios do dia posterior.

Ainda no que tange ao retorno das ações e ao volume de negócios, existe uma série de estudos que trazem indícios da relação entre esses dois movimentos do mercado, e demonstram que na verdade o volume negociado é que seria um preditor dos retornos. Nesse sentido, Do *et al.* (2014) desenvolveram um estudo na área comportamental, de modo a verificar como investidores reagem a chegada de novas informações no mercado. Os autores demonstraram que o volume negociado tem influência sobre o retorno, especialmente nos momentos em que os retornos são maiores. Todavia, o sentido da relação entre eles pode ser positiva ou negativa, a depender do mercado em análise.

Em perspectiva semelhante a de Do *et al.*, Chen *et al.* (2015) afirmam que o volume de negócios e o retorno estão correlacionados, porém, essa correlação é positiva no mercado

chinês, e negativa no mercado americano, o que traz indícios de que a direção da relação existente entre o retorno e o volume negociado pode ser alterado frente a maturidade do mercado. Sheng, Brzeszczyński e Ibrahim (2017) mostram que as mudanças que ocorrem no retorno são sensíveis aos diferentes níveis de volume negociado. Por este ângulo, Kao, Chuang e Ku (2019) consideram que o volume de negócios pode ser visto como uma transição de informação, sendo ele capaz de causar mudanças nos preços. Assim, os autores demonstram que a relação entre o retorno e o volume negociado existe, e que o volume negociado pode ter efeitos sobre o retorno no tempo atual, e no período de um a três dias posteriores. Ademais, segundo os autores, um maior volume negociado normalmente acarreta um maior retorno, especialmente quando o volume de negociação é alto. Por outro lado, o volume negociado defasado influencia o retorno negativamente, quando o volume de negócios é baixo.

Assim, frente aos estudos expostos, acredita-se que, embora o sentimento possua influência sobre os retornos, é necessário que se considere as características específicas da empresa, os fatores relacionados ao mercado e variáveis macroeconômicas para que se tenha uma melhor visão do que compõe os retornos das ações no mercado acionário brasileiro. De maneira semelhante, é necessário que se considere outras variáveis para além do volume de *tweets* para entender as mudanças que ocorrem no volume negociado.

### 3 METODOLOGIA

O presente estudo teve como população as mensagens postadas no Twitter que estivessem relacionadas ao mercado acionário brasileiro. Todavia, frente a dificuldade de tal obtenção, optou-se por utilizar o índice IBOVESPA de modo a conhecer o mercado e consequentemente obter apenas *tweets* que estivessem relacionados a esse índice ou a alguma das ações que o compõe, sendo esses os *tweets* que fazem parte da amostra.

O índice IBOVESPA, o qual é representado no estudo pelo BOVA11, foi utilizado como *proxy* para conhecer o comportamento do mercado acionário brasileiro, pois ele engloba as ações mais representativas e líquidas desse mercado. O índice SMLL, representado pelo SMAL11, foi utilizado para representar as empresas de menor capitalização no mercado brasileiro, o qual fora utilizado em um teste adicional. Os dados referentes às séries históricas dos pontos e do volume de negociações do índice foram coletados por meio da base de dados Thomson Reuters. Já os *tweets* que fazem parte da amostra foram coletados por meio de uma API (*Application Programming Interface*) do Twitter escrita em *python*, chamada *Tweepy*. O estudo buscou analisar o efeito diário do sentimento sobre o mercado acionário. Assim, o período de análise foi de 1 de janeiro de 2017 à 31 de dezembro de 2018.

A escolha de fazer uma a análise no período de dois anos se deu devido ao tempo necessário para a obtenção dos dados e, especialmente, do custo para que houvesse a atribuição do sentimento por meio da ferramenta que fora utilizada. Assim, salienta-se que os anos de 2017 e 2018 foram escolhidos porque, além de serem os anos mais próximos disponíveis no momento da realização do estudo, esses anos foram escolhidos por causa da tensão política que os acompanha e por terem ocorrido fatos marcantes no mercado, como o “Joesley day”. Ademais, o número de pessoas físicas na bolsa teve aumento significativo nesses anos, o que pode estar correlacionado com o aumento do número de usuários no Twitter que fazem postagens acerca da bolsa de valores brasileira. Dessa maneira, o período escolhido é suficiente e o mais adequado para a proposta do estudo.

Para que a biblioteca *Tweepy* levantasse os *tweets* necessários à pesquisa, fez-se necessário que parâmetros fossem estabelecidos, tais como o período, a língua e os termos que precisavam constar nos *tweets*. Assim, optou-se por buscar *tweets* que estavam em português. Como o objetivo do estudo é conhecer o mercado acionário brasileiro como um todo, então, do mesmo modo que no estudo de Oliveira, Cortez e Areal (2017), rastreou-se todos os *tweets* que contivessem o *ticker* de alguma das ações presentes no índice estudado e que contivessem um

cifrão (\$) na frente, pois usuários do Twitter normalmente o utilizam quando estão se referindo a ações (Ex: \$MGLU).

Adicionalmente, utilizou-se o método netnográfico, onde houve, durante dois meses, uma observação participante de modo a identificar o comportamento dos indivíduos na rede social Twitter. Assim, foi visto que, diferente do que ocorre em países como os EUA, no contexto brasileiro, os usuários do Twitter, ao falarem acerca de ações, costumam utilizar uma *hashtag* (#) antes do *ticker* das ações (Ex: #MGLU). Para além, por meio do método netnográfico, foi possível detectar padrões e conhecer como os usuários costumam se referir a cada empresa que compõe o IBOVESPA, no que diz respeito exclusivamente à ação ou a dados que possam ter influência sobre a mesma, o que possibilitou o levantamento de termos que seriam utilizados no estudo. Assim, primeiramente se utilizou tais termos para encontrar os *tweets* que fariam parte da amostra.

No que se refere aos termos levantados, foram procurados os termos que representassem cada empresa que compõe o IBOVESPA, frente ao que fora visto por meio da observação participante no ambiente *on-line*. Assim, por exemplo, para levantar o sentimento diário dos investidores acerca da companhia Magazine Luiza S/A foram buscadas as seguintes palavras: #MGLU3, MGLU3, \$MGLU, #MGLU. Também foram buscados termos que remetessem ao mercado brasileiro como um todo, tais como: #BVMF3, \$BVMF, BVMF3, BVMF, B3SA3, BM&Fbovespa, B3, IBOV, IBOVESPA, Bolsa Brasileira, Brasil Bolsa Balcão, e Bolsa De São Paulo.

Uma vez que o IBOVESPA sofre atualizações a cada quatro meses, salienta-se que houve atualização das ações e termos que foram buscados frente as atualizações da carteira, de modo que, em cada dia foram buscadas exatamente as ações que compunham o índice IBOVESPA no respectivo dia.

Após a inserção de todos os termos, a biblioteca retornou todos os *tweets* do período em estudo, trazendo um total de 1.195.575 *tweets*. Esses *tweets* foram tratados, de modo a remover links, *tweets* que citam o Youtube, quando eles eram apenas curtidas de vídeos, e *tweets* que estavam repetidos, mas não eram *retweets* ou postados em datas distintas, ou por pessoas distintas. Após esse tratamento, o estudo chegou a uma amostra de 314.864 *tweets*, e se prosseguiu com a atribuição do sentimento.

Logo após a atribuição do sentimento, os valores levantados puderam ser utilizados nas estimações que estão aqui apresentadas na seção dos resultados. Optou-se por utilizar modelos de regressão quantílica em cinco quantis distintos, sendo esses: 0.1, 0.25, 0.50, 0.75, e 0.9. Salienta-se que esse tipo de modelo foi escolhido de modo a controlar a heterogeneidade dos dados, uma vez que o Brasil é um mercado emergente e menos maduro, então seus movimentos, como os do retorno e do volume negociado costumam ser mais instáveis.

O método de regressão quantílica foi escolhido pois além de sua robustez e capacidade para lidar com problemas de heterogeneidade, ele oferece a possibilidade de analisar as diferentes relações que o sentimento e o volume de *tweets* possuem com os movimentos do mercado, frente as variações de tais movimentos. Assim, é possível ver, por exemplo, se a relação que sentimento possui com os retornos é a mesma e possui a mesma direção quando o mercado, focando exclusivamente no retorno, está em alta ou em baixa.

### 3.1 Atribuição do sentimento

Para a atribuição do sentimento e conseqüente desenvolvimento do índice se utilizou *machine learning* na classificação da polaridade das palavras e, assim como no estudo de Bollen, Mao e Zeng (2011), optou-se por aplicar uma ferramenta de análise de sentimentos que é disponibilizada pelo Google. Foi empregado um processador de linguagem natural denominado de “*Google Cloud Natural Language API*”, o qual possui uma ferramenta de análise de sentimentos. Essa ferramenta possui os parâmetros de -1 a 1 na atribuição de sentimento, onde 1 é o sentimento extremamente otimista e -1 é o sentimento extremamente pessimista. Ademais, os *tweets* podem assumir qualquer valor dentro desse intervalo de -1 a 1, incluindo o valor 0, o qual é tido como um sentimento neutro.

Essencialmente o sentimento dos *tweets* ( $S_j$ ) é estimado pelo algoritmo com base no peso das palavras positivas e negativas, conforme a Equação 1.

$$S_j = \frac{\sum PPP - \sum PPN}{\sum PPPN} \quad (1)$$

Em que,

$S_j$  - Sentimento do *tweet* da empresa ou do índice  $j$  no dia  $t$ ;

$\sum PPP$  - Peso das palavras positivas;

$\sum PPN$  - Peso das palavras negativas;

$\Sigma$ PPP - Peso das palavras negativas e positivas somadas.

A figura 1 traz um exemplo de *tweet* em que o algoritmo atribuiu um sentimento positivo (0,5).

Figura 1- *Tweet* extraído do Twitter- 01/08/2017

ITUB4 hoje mostra que muito mais q um bom resultado é o management confiante, reiterando seu guidance. Vem 2S17 melhor pra frente

3:07 PM · 1 de ago de 2017 · Twitter for iPhone

5 Retweets 80 Curtidas



Fonte: Twitter (2020)

A figura a seguir traz um exemplo no qual o procedimento de análise atribuiu valor negativo (-0,5).

Figura 2- *Tweet* extraído do Twitter- 02/06/2017

Hoje o #Ibovespa parece uma escuna em mar revolto. Sobe, desce, sobe, desce. E em água suja, pq tem coisa feia subindo ou boiando!

11:37 AM · 2 de jun de 2017 · Twitter Web Client

5 Retweets 25 Curtidas



Fonte: Twitter (2020)

### 3.2 Modelos

Uma vez que se atribuiu sentimento a cada *tweet* da amostra, como a intenção do estudo é formar um índice agregado, acumulou-se o sentimento por dia, de modo a criar um índice diário de sentimento do investidor. Porém, visto que a B3 não funciona nos feriados ou finais de semana, então, o sentimento adquirido para os dias em que não há negociação foi acumulado e seus efeitos foram analisados no dia de negociação seguinte. O sentimento diário consiste na média do sentimento dos *tweets* do dia de negociação  $t$ .

$$SD_t = \frac{1}{n} \sum_{i=1}^n S_{j,t} \quad (2)$$

Em que,

$SDt$  - Sentimento médio dos *tweets* no dia de negociação  $t$ ;

$n$  - Número de *tweets* no dia de negociação;

$S_{j,t}$  - Sentimento do *tweet* da empresa ou do índice  $j$  no dia  $t$ ;

De modo a verificar o retorno do IBOVESPA e do SMLL, utilizou-se a Equação 3 para o cálculo dos retornos normais.

$$R_{p,t} = \ln\left(\frac{P_t}{P_{t-1}}\right) \quad (3)$$

Em que,

$R_{p,t}$  - Retorno observado do portfólio  $p$  (IBOVESPA ou SMLL);

$P_t$  - Pontuação do índice no dia  $t$ ;

$P_{t-1}$  - Pontuação do índice em  $t-1$ ;

Visto que o estudo tem por base a hipótese do mercado eficiente em sua forma semiforte (Fama, 1970,1991), acredita-se que o retorno passado das ações e as informações publicamente disponíveis compõem o retorno contemporâneo. Assim, uma vez que os *tweets* são tidos como um tipo de informação, a Equação 4 testou como o sentimento expresso por meio dessas mensagens se relaciona com o retorno do mercado acionário brasileiro (IBOVESPA).

$$Ret_t = \beta_0 + \beta_1 Sent_t + \gamma_1 \mathcal{L}_s(Ret_t) + \varepsilon_t \quad (4)$$

Em que,

$Ret_t$  - Log do retorno do IBOVESPA no dia  $t$ ;

$Sent_t$  - Sentimento do investidor no dia  $t$ ;

$\mathcal{L}_s$  - Operador de *lag*, onde  $s$  corresponde ao número de defasagens;

$\varepsilon_t$  - Termo de erro.

Visto que Kao, Chuang e Ku (2019) consideram o volume negociado uma fonte de informação, e que Do *et al.* (2014), Chen *et al.* (2015) e Sheng, Brzeszczyński e Ibrahim (2017) mostram que o volume negociado tem efeitos sobre o retorno das ações, optou-se por utilizar a Equação 4, a qual é um teste de eficiência do mercado, acrescida de uma variável referente ao volume negociado, e mais duas variáveis que incluem o volume de negócios, sendo essas: a



interação entre o sentimento e o volume negociado, e a interação entre o retorno passado e o volume negociado. Assim, têm-se a Equação 5.

$$Ret_t = \beta_0 + \beta_1 Sent_t + \gamma_1 \mathcal{L}_s(Ret_t) + \beta_2 VN_t + \beta_3 SentxVN_t + \gamma_2 \mathcal{L}_s(Ret_t) x VN_t + \varepsilon_t \quad (5)$$

Em que,

$VN_t$  – Volume (número) de negócios do Ibovespa no dia  $t$ ;

As demais variáveis estão conforme detalhadas anteriormente.

Embora se acredite que o mercado seja eficiente, ao menos em sua forma semiforte (MALKIEL, 2003), decidiu-se por utilizar um modelo adicional de modo a encontrar melhor explicação às mudanças nos retornos, controlando o efeito de outras variáveis que são tidas na literatura como sendo responsáveis por tais mudanças. Nesse sentido, utilizou-se no modelo a variável referente ao sentimento, a qual está ligada à parte informacional da HME; fatores tidos por Keene e Peterson (2007), sendo eles o fator mercado, o HML, o *momentum*, e a liquidez, excetuando o fator tamanho, visto que suas evidências não são fortes no mercado brasileiro (MACHADO; FAFF; SILVA, 2017); uma variável relacionada ao volume diário de *tweets*, e duas variáveis macroeconômicas, sendo essas a expectativa de inflação e a expectativa do PIB.

Bathia e Bredin (2018) observam que o uso do modelo de cinco fatores acoplado à variável de sentimento melhora a capacidade de explicação dos retornos. Ademais, no caso desta pesquisa, optou-se por reproduzir o modelo utilizado no trabalho de Galdi e Golçalves (2018), ao estudarem o mercado brasileiro, o qual utilizou apenas a taxa livre de juros no lugar do fator mercado, visto que o retorno do mercado é a variável dependente e está diretamente relacionada ao risco de mercado, pois, dessa forma, acredita-se que o modelo se torne mais robusto. Todavia, foram acrescentados ao modelo de Galdi e Golçalves (2018) outros aspectos referenciados na literatura que são tidos como úteis na explicação da *cross section* dos retornos e/ou no melhoramento do poder de explicação dos modelos de precificação de ativos, conforme Equação 6. Ressalta-se que todas as variáveis são diárias ou convertidas para essa escala.

$$Ret_t = \beta_0 + \beta_1 Sent_t + \beta_2 VDT_t + \beta_3 RF_t + \beta_4 HML_t + \beta_5 MOM_t + \beta_6 Liq_t + \beta_7 Infl_t + \beta_8 PIB_t + \varepsilon_t \quad (6)$$

Em que:

$VDT_t$  – Volume Diário de *Tweets*, representado pela quantidade diária de mensagens postadas acerca das ações presentes no mercado ou do mercado brasileiro;

$RF_t$  – Taxa livre de risco no dia  $t$ , representada pelo Swap DI x Pré 30;

$HML_t$  – *High minus Low* no dia  $t$ , calculado pela diferença do retorno entre portfólios de alto e baixo *book-to-market*;

$MOM_t$  – *momentum* no dia  $t$ , calculado pelo retorno do portfólio de ações vencedoras, menos o retorno do portfólio de ações perdedoras;

$Liq_t$  – Liquidez no dia  $t$ , calculada por meio do retorno da carteira de ações líquidas menos o retorno da carteira de ações ilíquidas.;

$Infl_t$  – Expectativa de inflação do mercado no dia  $t$ ;

$PIB_t$  – Expectativa de crescimento do PIB do mercado no dia  $t$ ;

As demais variáveis estão conforme detalhadas anteriormente.

Evidencia-se que os dados referentes aos fatores de risco (HML, MOM e LIQ) e a *riskfree*, a qual é medida pelo Swap DI x pré para 30 dias, foram obtidos por meio do *site* do Núcleo de Pesquisa em Economia Financeira (Nefin), o qual faz parte do Departamento de Economia da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo. Já as variáveis macroeconômicas de expectativas foram obtidas por meio do Relatório Focus, o qual é divulgado semanalmente pelo Banco Central. As variáveis de sentimento e do volume de *tweets* foram desenvolvidas neste estudo, conforme seção 3.1.

De modo a verificar se o volume de *tweets* diários têm relação com o volume negociado do mercado (número de transações), apresenta-se a Equação 7, a qual visa identificar a associação existente entre o número de mensagens postadas no Twitter e o volume negociado do mercado acionário brasileiro. Ademais, buscou-se averiguar a condição em que o volume de negócios (número de transações) pode estar associado ao sentimento. Ressalta-se que as variáveis utilizadas na Equação 7 foram as apresentadas na literatura como tendo influência sobre o volume negociado.

$$VN_t = \beta_0 + \beta_1 Sent_t + \beta_2 VDT_t + \gamma_1 \mathcal{L}_s(VN_t) + \varepsilon_t \quad (7)$$

Em que:

As variáveis estão conforme detalhadas anteriormente.

Uma vez que a amostra é composta por *tweets* positivos, negativos e neutros, optou-se por utilizar um novo modelo (Equação 8) de modo a verificar se havia diferença na forma em que o volume de *tweets* positivos e negativos se associavam ao volume negociado, ao excetuar os *tweets* neutros.

$$VN_t = \beta_0 + \beta_1 Sent_t + \beta_2 VTP_t + \beta_3 VTN_t \gamma_1 + \mathcal{L}_s(VN_t) + \varepsilon_t \quad (8)$$

Em que,

$VTP_t$  – Volume de *tweets* positivos;

$VTN_t$  – Volume de *tweets* negativos;

As demais variáveis estão conforme detalhadas anteriormente.

### 3.3 Análise adicional - Small Caps

De modo a verificar se o índice desenvolvido é robusto e tem comportamento diferente com a mudança de liquidez das ações, optou-se por analisar a associação do índice de sentimento com as empresas de menor capitalização (SMLL). Ressalta-se que o índice de sentimento utilizado é o que foi construído com base nas empresas mais representativas do mercado, o qual fora explicado anteriormente. Assim, o objetivo é verificar se um índice de sentimento desenvolvido com base em empresas de maior capitalização também terá efeitos sobre o retorno e o volume negociado de um índice que é composto por empresas de menor capitalização.

Dessa maneira, *a priori*, utilizou-se uma regressão de teste de eficiência de mercado, com base em um modelo igual à Equação 4, todavia, o retorno verificado foi o do ETF SMALL11, o qual é um fundo de investimento que busca refletir a performance do índice SMLL, sendo assim composto pelas empresas de menor capitalização presentes na B3.

Posteriormente, um outro modelo foi utilizado, visando controlar o efeito das variáveis que são tidas na literatura como tendo influência sobre as mudanças que ocorrem no retorno das entidades. O modelo é inspirado na Equação 5, porém, os fatores de risco relacionados às entidades (BM, MOM e Liq) não foram inseridos, visto que os dados obtidos pela pesquisa para

tais fatores foram calculados apenas para empresas com maior capitalização, não sendo viável seu uso ao estudar empresas de menor capitalização. Assim, tem-se a Equação 9.

$$Ret_t = \beta_0 + \beta_1 Sent_t + \beta_2 VTT_t + \beta_3 RF_t + \beta_4 Infl_t + \beta_5 PIB_t + \varepsilon_t \quad (9)$$

Em que,

$Ret_t$  – Retorno do SMLL no dia  $t$ ;

As demais variáveis estão conforme detalhadas anteriormente.

### 3.4 Limitações

O presente estudo se limitou a analisar os *tweets* referentes ao IBOVESPA que foram postados no período de 1 de janeiro de 2017 a 31 de dezembro de 2018. Assim, salienta-se que os resultados aqui encontrados são restritos a esse período.

Levando em consideração que o estudo analisou o sentimento dos investidores por meio de uma rede social, ressalta-se que o Twitter fora escolhido como fonte de dados devido à facilidade com que os usuários podem obter informações de maneira instantânea, além de poderem postar e compartilhar tais informações. Ademais, as APIs (*Application Programming Interface*) do Twitter são melhores para a obtenção de dados, comparando-se a outras redes sociais. Assim, as conclusões do estudo estão restritas aos *tweets* e ao sentimento que é expresso apenas por meio do Twitter.

Outra limitação do estudo é com relação ao índice de sentimento. O primeiro ponto é que o índice foi desenvolvido frente ao dicionário implantado na biblioteca, de modo que, visto a dificuldade do processo de linguagem, o algoritmo não tem a mesma eficácia que uma classificação humana, por exemplo. Nesse sentido, visto que a extração do sentimento foi obtida por meio da polaridade das palavras via programação, salienta-se que o objetivo não foi replicar o processo de leitura humana, sendo essa apenas uma ferramenta na extração de informações quantitativas por meio de dados textuais. Desta forma, como ao trabalhar com *Big Data* a classificação manual seria inexecutável, o programa se mostra eficiente.

## 4 ANÁLISE DOS RESULTADOS

De modo a responder o objetivo proposto pela pesquisa, no qual as hipóteses foram testadas por meio de regressão quantílica, essa seção traz os resultados aqui encontrados. Assim, há uma estatística descritiva das variáveis; uma análise dos gráficos das principais variáveis de interesse da pesquisa; e os resultados sobre a relação entre o sentimento obtido por meio do Twitter e o retorno do mercado, e da associação entre o volume de *tweets* e o volume de negócios do mercado brasileiro.

### 4.1 Estatísticas Descritivas

A Tabela 1 traz consigo as estatísticas descritivas das variáveis analisadas. Nota-se que, durante o período estudado, tanto o retorno do índice IBOVESPA, que representa o mercado como um todo, quanto o retorno do índice SMLL, que representa as empresas de menor capitalização na bolsa brasileira, obtiveram valores médios positivos, de 0,00085 e 0,0009, respectivamente, tendo as empresas menores um retorno levemente superior as demais, o que pode ser explicado por Fama e French (1991), uma vez que, segundo os autores, por vezes, ações de menores costumam apresentar maior expectativa de retorno.

Era esperado que o mercado brasileiro apresentasse valores positivos para o retorno, uma vez que Silva (2017), ao investigar o IBOVESPA de julho de 2011 a junho de 2017, encontrou um retorno diário positivo. Ademais, esse resultado também era esperado porque no trabalho de Shen, Liu e Zhang (2018), ao investigarem 26 índices internacionais, há indícios de que mercados com menor maturidade costumam trazer retornos maiores. O trabalho de Shen, Liu e Zhang (2018) também mostra que índices compostos por empresas de menor capitalização têm retornos maiores, se comparados a índices que representam o mercado como todo. Esse padrão foi seguido nos resultados aqui apresentados, visto que o SMLL obteve um retorno médio diário superior ao IBOVESPA.

Ademais, é possível verificar que o IBOVESPA e o índice SMLL se mostram atrativos para o mercado, uma vez que, no estudo de Shen, Liu e Zhang (2018), o valor máximo de retorno diário obtido entre os 26 índices internacionais foi de 0,0007, sendo esse valor inferior tanto ao retorno do IBOVESPA, quanto o do SMLL. Assim, há indícios de que a bolsa brasileira é atrativa não apenas nacionalmente, mas também para investidores internacionais.

Ao analisar o sentimento, nota-se que o sentimento diário dos *tweets* assumiu um valor de 0,0338, assim, percebe-se que o sentimento expresso por meio do Twitter foi positivo, o que pode estar associado ao aumento do IBOVESPA durante o período analisado, uma vez que quando o sentimento é maior, os investidores se tornam mais otimistas em relação a investirem no mercado. Esse resultado é diferente do esperado, pois, Galdi e Golçalves (2018), ao investigarem o mercado brasileiro, notaram que o sentimento das notícias costuma ser mais pessimistas. De forma semelhante, Silva (2017) encontrou, para o contexto brasileiro, um valor de -0,0577, mostrando que havia um sentimento mais pessimista em relação ao mercado brasileiro.

Essa diferença de resultados pode estar ligada a alguns fatores, como por exemplo, a diferença de amostra. Isso se dá porque Silva (2017) e Galdi e Gonçalves (2018) obtiveram o sentimento por meio de mídia especializada, enquanto o estudo obteve o sentimento por meio de *tweets*, de maneira que, pode ser que investidores sejam mais otimistas, enquanto que a mídia especializada tem um viés mais pessimista.

O período analisado também pode ser um fator que contribui para a diferença de resultados, visto que os demais estudos citados só analisaram até junho de 2017. Assim, os demais meses de 2017 e o ano de 2018 pode ter trazido um maior otimismo para os investidores, e um maior otimismo pode estar associado ao alto crescimento e aos retornos que o IBOVESPA obteve durante o período estudado.

**Tabela 1-** Estatísticas descritivas (01/01/2017 a 31/12/2018)

Variável	Média	Mediana	DP	Assimetria	Curtose	Obs.	Min	Máx.
<b>RetIbov</b>	0,00085	0,00093	0,0130	-0,5724	7,4521	491	-0,0880	0,0457
<b>RetSmall</b>	0,0009	0,0020	0,0125	-1,2621	11,5118	491	-0,0991	0,0451
<b>Sent</b>	0,0338	0,0381	0,0371	-1,3906	10,3707	491	-0,1778	0,1892
<b>VDT</b>	639,7495	575	357,0173	7,1576	78,6102	491	213	5281
<b>VTP</b>	243,3096	224	102,7091	2,6496	17,1238	491	86	1136
<b>VTN</b>	117,3686	89	175,3354	10,9021	139,108	491	32	2688
<b>VN</b>	3672363	3463024	1144582	1,84123	10,11979	491	833734	1,14. 10 <sup>+7</sup>
<b>HML</b>	0,00074	0,00071	0,0078	-,7398	8,2189	491	-0,0555	0,0215
<b>Mom</b>	0,0005	0,0008	0,0074	-1,1860	11,3603	491	-0,0581	0,0238
<b>Liq</b>	0,0006	0,0007	0,0078	0,1761	4,6693	491	-0,0309	0,0366
<b>Riskfree</b>	0,00030	0,00026	0,00007	0,9483	2,4112	491	0,0002	0,0005
<b>Infla</b>	0,0425	0,0417	0,0030	0,0026	1,9871	491	0,0369	0,0484
<b>PIB</b>	0,0122	0,0098	0,0095	-0,1896	5,3991	491	-0,0319	0,029

Fonte: Elaboração própria (2020)

Outro ponto importante a ser destacado é o volume de *tweets*, onde se nota que diariamente os usuários postaram no Twitter, em média, 640 mensagens que estavam ligadas ao mercado brasileiro, seja citando alguma ação específica, falando da bolsa de valores brasileira em si, ou do índice IBOVESPA. Todavia, essa variável é bem heterogênea, uma vez que o valor mínimo de postagens em um dia foi de 213, e o valor máximo de mensagens postadas acerca do mercado brasileiro em um mesmo dia foi de 5.281.

O volume diário de *tweets* positivos e negativos seguem a mesma tendência do volume de *tweets* sem segregação por tipo de sentimento, uma vez que o número médio diário de mensagens postadas com teor positivos foi de 243, obtendo valor mínimo de 86, e máximo de 1136, e a média diária de *tweets* negativos foi 117, tendo seu valor mínimo de 32, e seu valor máximo de 2.688. Assim, em média, diariamente há mais mensagens de cunho positivo do que negativo, todavia, houve período na amostra em que ocorreu um aumento abrupto na quantidade de *tweets* de cunho negativo.

No que tange ao volume de negócios, nota-se que o volume negociado diariamente no mercado brasileiro obteve valor médio de 3.672.363. Todavia, observa-se que essa variável é heterogênea, visto que há um alto valor de desvio padrão, e seus valores mínimo e máximo são bem distintos, pois essa variável teve valor mínimo de 833.734, e valor máximo de 11.400.000.

Embora o valor máximo obtido pelo volume de negócios seja alto se comparado a média, ainda assim é possível ver que o volume de negócios no mercado brasileiro ainda é baixo se comparado a mercados mais desenvolvidos, como o do Reino Unido, por exemplo, em que Nisar e Yeung (2018) encontram um volume negociado diário médio de 658.021.632. Ademais, é possível ver que há uma alta instabilidade do volume de negócios no mercado brasileiro, uma vez que em um mercado desenvolvido como o do Reino Unido, os valores de volume negociado desviam em 8% da média, enquanto no Brasil o desvio é de 31%.

Em relação às características específicas das entidades, percebe-se que o *High minus Low* (HML) assumiu valor médio diário de 0,0007 para as entidades brasileiras, o qual é considerado um baixo valor. Esse valor era esperado, pois Machado e Medeiros (2011) também encontraram valores baixos para esse fator no mercado brasileiro. Conforme os autores, esse baixo valor implica em oportunidades de crescimento para essas entidades.

O fator *momentum* obteve valor médio diário de 0,0005, e, uma vez que esse valor é positivo, acredita-se que as ações vencedoras, ou seja, as que tiveram altos retornos passados,

se sobressaíam sobre as ações perdedoras existentes no mercado acionário brasileiro, uma vez que o fator momento consiste na diferença entre o retorno do portfólio que teve altos retornos passados menos o portfólio com baixos retornos passados. A liquidez média diária para o período analisado foi de 0,0006, o que implica dizer que há uma maior liquidez no mercado brasileiro, uma vez que as ações que possuem alta liquidez se destacaram sobre as ações com baixa liquidez, o que resultou nesse valor médio diário positivo.

Com relação às variáveis macroeconômicas tidas na literatura como tendo influência sobre o retorno, nota-se que a taxa livre de risco diária foi de 0,0003, e que seu valor máximo foi de 0,0005, de modo que os retornos do IBOVESPA e do SMLL foram ambos superiores à taxa livre de risco, o que pode estar associado ao aumento do número de pessoas físicas na bolsa no período estudado.

A expectativa do mercado de inflação anual e a expectativa de crescimento do PIB anual, os quais foram tratados em bases diárias, obtiveram valor de 4,25% e 1,22%, respectivamente. Todavia, o PIB teve um alto nível de variação, visto que o seu valor mínimo foi de -3,19%, o que ocorreu no começo de 2017, pois o país ainda estava sentindo os efeitos da recessão econômica e da alta contração que houve no PIB em 2015, todavia, com o passar do tempo a economia voltou a melhorar, e a expectativa do PIB conseguiu se recuperar, atingindo em 2018 o valor máximo foi de 2,9%.

A Tabela 1 mostra que há uma grande dispersão nos dados, a qual pode ser observada por meio da discrepância dos valores máximos e mínimos das variáveis, os quais acarretam altos valores de desvio padrão, de modo que há uma grande heterogeneidade nos dados. Ademais, os valores encontrados de assimetria e curtose nas variáveis dão indícios de que elas não seguem uma distribuição normal, fato que é confirmado por meio de testes de *Jarque-Bera*, onde ao se analisar a distribuição das séries, constatou-se a ausência de normalidade. Assim, frente a heterogeneidade dos dados e a não normalidade deles, optou-se por utilizar regressão quantílica devido a sua robustez.

## 4.2 Análise dos Gráficos

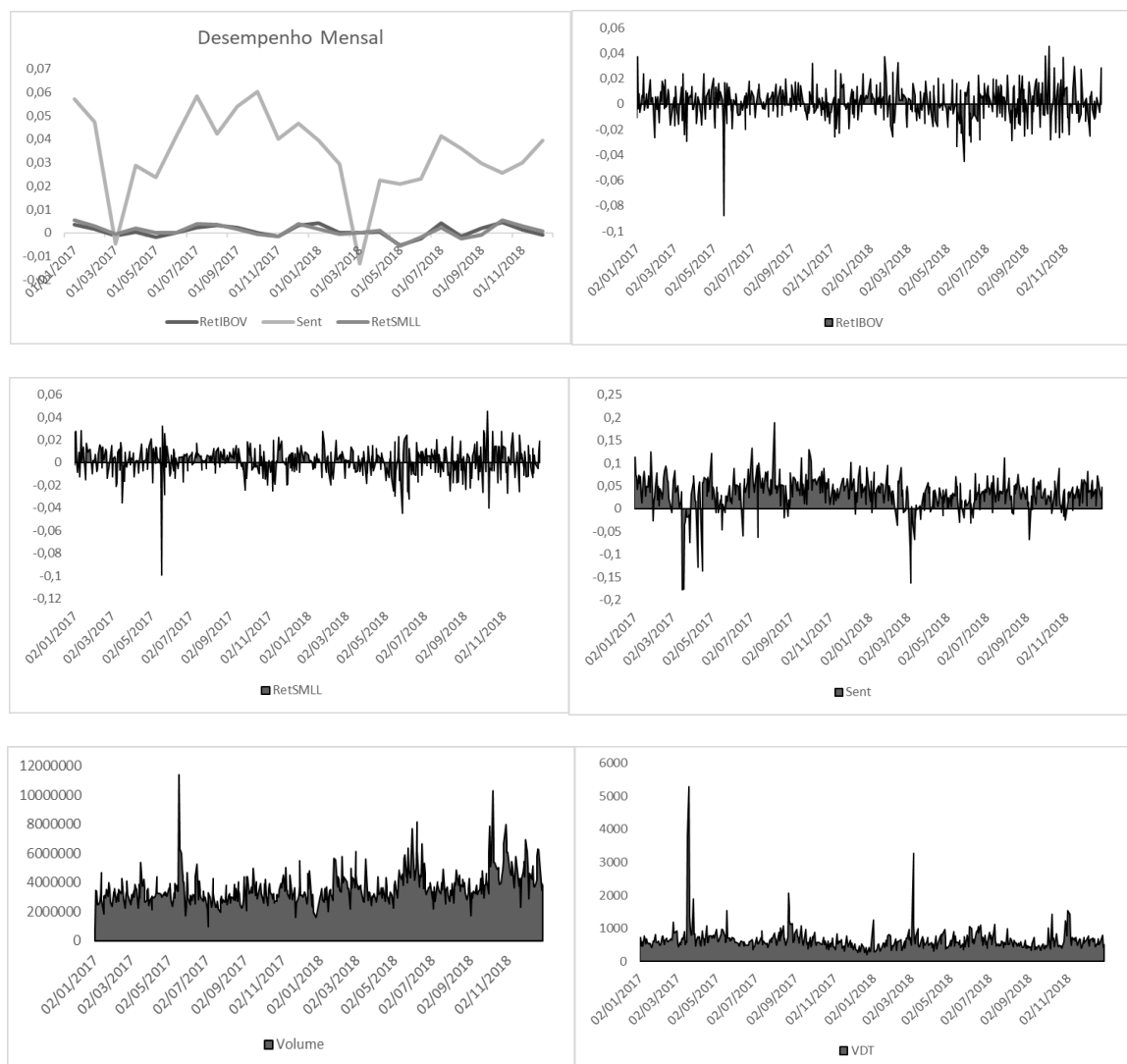
Os gráficos abaixo demonstram as oscilações que ocorreram nas principais variáveis de interesse do estudo durante o período da análise. É perceptível, por meio do gráfico do que traz



o desempenho mensal dos retornos do IBOVESPA, do SMLL, e do sentimento dos investidores, que essas variáveis parecem andar em sentidos próximos, especialmente os retornos das *small caps* e das empresas de maior capitalização, uma vez que por meio do gráfico é possível ver que as linhas dessas variáveis estão bem próximas uma da outra. Ademais, embora o sentimento pareça variar bem mais do que os retornos, o sentido em que ele caminha parece estar, em muitos pontos, próximo ao sentido em que os retornos caminham.

Os demais gráficos trazem os valores obtidos diariamente para cada uma das variáveis de interesse. Assim, quanto aos retornos, tanto do IBOVESPA quanto do SMLL, é notório que ambos sofrem fortes oscilações, o que pode ocasionar retornos maiores, fato que é comum em mercados emergentes, e que fora visto em estudos como o de Shen, Liu e Zhang (2018).

**Figura 3-** Distribuição das séries temporais em análise



**Fonte:** Dados da pesquisa (2020)

Nota-se que foi no dia 18/05/2017 que os retornos tiveram a maior oscilação, tanto na série referente ao IBOVESPA quanto a do SMLL, sendo esse movimento negativo. Isso ocorreu porque no dia 17/05/2017, o jornal “O Globo” trouxe à tona casos de corrupção que haviam sido relatados pelo executivo da empresa JBS S.A., fato que ficou conhecido pelo mercado como “Joesley *day*” e acabou por trazer impactos para todo o mercado brasileiro.

Percebe-se que em maio de 2018 os retornos também foram negativos próximo ao final desse mês, fato que é explicado devido à greve dos caminhoneiros, que teve início no dia 21 de maio de 2018, e durou até o dia 01 de junho de 2018. Nesse sentido, para além do evento ocorrido, a intervenção do estado, as concessões dadas, e a interferência no preço do diesel fizeram com que o mercado reagisse negativamente.

Um fato de destaque é que no dia 19/05/2017, enquanto o IBOVESPA permanecia em queda devido ao vazamento da JBS, o SMLL obtinha o seu maior retorno referente ao ano de 2017, o que pode oferecer indícios de que possivelmente investidores tenham migrados de empresas maiores para empresas menores de modo a se protegerem do impacto que a crise instaurada pela JBS poderia impor a outras empresas de grande porte, de modo que o aumento pela procura ocasionou um aumento no preço e conseqüentemente nos retornos das empresas menores. Todavia, esse movimento também pode ser plenamente explicado pelo fato de que as ações do IBOVESPA são mais líquidas, o que ocasiona maior variação.

Ainda no que tange os retornos, vê-se que, em ambos os índices, durante todo o período estudado, o dia 08/10/2018 foi o que trouxe os maiores retornos, o que dá indícios de que o mercado ficou satisfeito com o resultado do primeiro turno das eleições nacionais para presidente da república, e antecipou a vitória do candidato que era mais pró-mercado, o qual possuía um programa econômico mais liberal, sendo tais fatores bem vistos pelo mercado.

As oscilações que ocorreram no retorno do mercado à medida que novas informações vieram à tona, sejam elas devido a questões como o período eleitoral, escândalos de corrupção, ou a eventos que resultam em intervenção estatal, tal qual a greve dos caminhoneiros, estão em conformidade com hipótese de que os mercados seriam eficientes (FAMA, 1970, 1991), assim, o mercado acionário estaria precificando os ativos, em partes, com base em absorver as informações que são recebidas. Ademais, tais oscilações podem ter ocorrido por haver influências comportamentais e psicológicas nos preços das ações, as quais são possivelmente sensíveis a repostas racionais, mudanças em políticas econômicas, e a percepção de risco (MALKIEL, 2003).

No que tange ao sentimento, nota-se que ele começou a decair abruptamente a partir do dia 17/03/2017, chegando seu valor mínimo no dia 20/03/2017. Acredita-se que esse sentimento negativo esteja atrelado a Operação Carne Fraca, da Polícia Federal, a qual foi deflagrada na manhã do dia 17/03/2017, e que foi responsável por desmontar um sistema de venda de carnes impróprias para o consumo.

A Operação Carne Fraca parece ter sido uma surpresa para o mercado, pois, antes do dia 17/03/2017, os investidores vinham apresentando sentimento positivo diariamente durante a maior parte mês de março, todavia, após esse dia, o sentimento do mercado obteve valores negativos por dez dias seguidos, fato que não se repetiu em nenhum outro período da amostra, e foi nesse período que o sentimento obteve seu valor mínimo, de -0,1778, o qual fora anteriormente citado nas estatísticas descritivas.

Foi também no mês de março, porém do ano de 2018, no dia 06/03, que o sentimento apresentou um valor bem negativo (-0,1429), se comparado a média diária do valor do sentimento (0,0338). Acredita-se que isso também esteja relacionado a terceira etapa da operação carne fraca, que ocorreu no dia 05/03/2018, e fez com que a empresa BRF perdesse mais de R\$ 4 bilhões em valor de mercado nesse mesmo dia (05/03/2018).

A maior alta do sentimento foi no dia 07/08/2017, onde se crê que foi no período em que o IBOVESPA passou a voltar ao patamar anterior à crise trazida pela JBS. Acredita-se que alguns fatores levaram ao aumento desse sentimento, tais como o otimismo visto nas principais bolsas internacionais, a valorização do mineiro de ferro no mercado chinês, a certeza de que Temer continuaria no cargo de presidente da república, e o empenho do mesmo no final de semana anterior para avançar nas reformas (trabalhista e previdenciária). Para além, a queda da taxa de juros e sinais de recuperação econômica trouxeram otimismo aos investidores.

Já em 2018, o pico mais alto de sentimento foi no dia 30/07/2018, onde acredita-se que tenha ocorrido porque, finalmente, após dois meses seguidos de quedas, o IBOVESPA chegou próximo ao final do mês em alta, superando a faixa dos 80.000 pontos. Além disso, a pré-candidatura de Geraldo Alckmin, sendo apoiado pelo bloco de partidos conhecido como “centrão”, e o PIB dos EUA sendo abaixo do esperado pelo mercado contribuíram para que houvesse tal otimismo.

As mudanças ocorridas no sentimento frente a fatores externos e fatores políticos era esperada, uma vez que conforme Nisar e Yeung (2018) esses aspectos podem ter influências

sobre o sentimento dos investidores. Ademais, essas oscilações eram esperadas porque, segundo Malkiel (2003), essas questões psicológicas e comportamentais são sensíveis a percepção de risco.

O volume negociado teve seus maiores picos em momentos pontuais que ocorreram ao longo de 2017 e 2018. Nesse sentido, o maior pico de volume negociado foi no dia 18/05/2017, onde ele atingiu seu valor máximo de 11.388.420 (exposto anteriormente nas estatísticas descritivas), o que está possivelmente relacionado ao vazamento da delação da JBS. Esse aumento do volume pode estar atrelado a maior oscilação que o IBOVESPA teve em seu retorno, a qual fora supracitada, uma vez que, conforme Do *et al.* (2014), o volume de negócios tem influência sobre o retorno. Ademais, em 2018 o volume negociado também teve um aumento abrupto no dia 08/10, logo após o primeiro turno das eleições nacionais para presidente, onde o mercado parecia já adiantar quem sairia como vencedor no segundo turno.

Em relação ao volume diário de *tweets*, o maior número de *tweets* postados acerca do mercado brasileiro ocorreu no dia 20/03, o qual acredita-se estar ligado a Operação Carne Fraca. Essa mesma operação causou um outro pico de volume no Twitter, no dia 05/03/2018. Fora esse acontecimento, outros fatos que ocasionaram picos no volume de *tweets* foram o vazamento da delação da JBS, que teve impactos próximos ao dia 18/05/2018, e o período eleitoral, que causou picos no dia 08/10/2018 e 29/10/2018.

### **4.3 Associação entre o sentimento dos *tweets* e o Retorno do IBOVESPA**

#### **4.3.1 Teste da HME utilizando *tweets* enquanto informação - IBOV**

Uma vez realizadas as estatísticas descritivas, a análise gráfica das variáveis, e escolhido o método a ser utilizado, o estudo seguiu com a estimação dos modelos de regressão quantílica. No que concerne à estimação do efeito do sentimento extraído por meio do Twitter sobre os retornos e volume de negociações do mercado brasileiro, ela ocorreu sobre os quantis que representavam os menores (0.10, 0.25) e os maiores desempenhos do mercado (0.75 e 0.90), bem como o desempenho na mediana (0.50).

Ressalta-se que as variáveis apresentadas na literatura que mostraram que seus valores no tempo passado eram capazes de influenciar o retorno e o volume de negociações

contemporâneos foram trabalhadas com *lags*. Embora a quantidade de dias passados nos quais cada variável pode afetar o retorno ou o volume negociado seja distinta, optou-se por utilizar um *lag* padronizado de cinco dias para essas variáveis, visto que a literatura (SILVA, 2017; RENAULT, 2017; SOUZA; BARBEDO; ARAÚJO, 2018) traz indícios de que o sentimento pode afetar o retorno de um até cinco dias depois de a informação ter sido divulgada.

De modo a facilitar a compreensão, as tabelas no resultado, ao invés de tratarem o tempo como  $t$ , trarão um  $d$  no lugar, para representar o dia, visto que o estudo trabalha em bases diárias. Assim, visto que o estudo trabalha com o dia atual e com *lags* de até cinco dias, salienta-se que:  $d$  corresponde ao dia atual;  $d-1$  ao dia anterior;  $d-2$  a dois dias atrás;  $d-3$  a três dias atrás;  $d-4$  a quatro dias atrás e  $d-5$  a cinco dias atrás.

A Tabela 2 apresenta o teste da eficiência do mercado, em que tanto o sentimento, que está ligado à parte informacional dos *tweets*, quanto o retorno foram trabalhados com *lags* de até cinco dias. Nesse sentido, é possível notar que, majoritariamente há uma associação positiva e significativa entre o sentimento e o retorno do mercado, uma vez que foi encontrada tal relação em  $d$ ,  $d-2$ ,  $d-3$  e  $d-4$ , sendo essa relação negativamente significativa apenas em  $d-4$ , quando os retornos são altos.

Assim, é possível notar que geralmente, quão maior for o sentimento dos investidores obtido por meio do Twitter, maior será o retorno do mercado acionário brasileiro. Adicionalmente, conforme esperado, frente aos estudos de Silva (2017) e Renault (2017), essa relação é mais forte no tempo contemporâneo. Ou seja, o sentimento do dia atual se relaciona de maneira mais intensa com o retorno atual. Em relação aos retornos, nota-se que, semelhante ao que fora visto por Silva (2017) ao estudar o mercado brasileiro, os retornos passados de um dia, dois dias e cinco dias atrás se relacionam com os retornos contemporâneos. Especificamente, na maior parte dos casos essa relação entre os retornos passados e o retorno atual é negativa, todavia os retornos de cinco dias atrás se relacionam positivamente com o retorno contemporâneo, quando os retornos são baixos.

**Tabela 2** - Teste da HME utilizando *tweets* enquanto informação – IBOV

Variáveis	q.10	q.25	q.50	q.75	q.90
<b><math>D</math></b>					
<b><math>Sent_d</math></b>	0,1787*** (0,0484)	0,0823*** (0,0306)	0,0268 (0,0233)	0,0085 (0,0182)	-0,0071 (0,0168)
<b><math>Ret_{d-1}</math></b>	-0,1350 (0,1239)	-0,0424 (0,0525)	-0,1152** (0,0494)	-0,0973* (0,0560)	-0,0530 (0,1169)
<b><math>D_{-1}</math></b>					
<b><math>Sent_{d-1}</math></b>	0,0378	-0,0057	-0,0092	-0,0041	0,0122

	(0,0442)	(0,0192)	(0,0175)	(0,0168)	(0,0237)
<b>Ret<sub>d-1</sub></b>	-0,0474	0,0174	-0,1148**	-0,0728	-0,0399
	(0,1376)	(0,0594)	(0,0531)	(0,0442)	(0,0704)

Notas: Equação 4:  $Ret_t = \beta_0 + \beta_1 Sent_t + \gamma_1 \mathcal{L}_s(Ret_t) + \varepsilon_t$

A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

A Tabela traz as variáveis defasadas em até um período; a continuação da

Tabela mostrando defasagens de dois a cinco dias se encontra no apêndice A.

**Fonte:** Elaboração própria (2020)

Essas evidências demonstram que as informações passadas não estão totalmente incorporadas nos preços, e se acredita que essa mudança na direção com a qual os retornos passados e os retornos presentes estão associados seja devido a tendência que as variáveis têm de regressarem a média, e, visto que a média do retorno é positiva (0,00085), então a associação entre o retorno de cinco dias atrás e o retorno contemporâneo também é positiva. Essas reversões eram esperadas pois o gráfico do retorno do IBOVESPA, exposto anteriormente, já demonstrava tal tendência.

Dessa maneira, pode-se perceber que tanto os retornos quanto os sentimentos foram significativos. Assim, é possível afirmar que os resultados corroboram com a hipótese de que o mercado é eficiente (FAMA, 1970, 1991), em sua forma semiforte. Tais resultados clarificam que é possível testar a eficiência de mercado utilizando o Twitter como fonte de dados para entender oscilações que ocorrem no mercado brasileiro, uma vez que fora visto que os retornos passados e as informações publicamente disponíveis, que estão aqui representados pelos *tweets* e pelo sentimento que deles fora extraído, estão presentes na composição dos retornos contemporâneos.

Os achados reforçam a ideia de estudos como os de Brown e Cliff (2004, 2005) e Baker e Wurgler (2006, 2007), que mostra haver relação entre o sentimento e o retorno das entidades. Ademais, nota-se que há a mesma tendência encontrada nesses estudos, de que um sentimento otimista no tempo contemporâneo está associado a menores retornos em períodos futuros.

Em relação aos resultados relacionados ao sentimento, nota-se que usualmente o sentimento está positivamente relacionado ao retorno, de modo que quanto maior for o retorno, maior também tende a ser o sentimento (mais positivo). Assim, aceita-se a hipótese da pesquisa de que quanto mais positivo é o sentimento do investidor, manifesto por meio do Twitter, maior é o retorno do IBOVESPA. Todavia, destaca-se que, quando os retornos são muito altos, os investidores parecem perceber que houve um tipo de reação exagerada, e por isso sentimentos

muito altos de quatro dias atrás acabam por se associar negativamente ao retorno no tempo contemporâneo. Os resultados aqui encontrados são similares aos de Silva (2017) e de Renault (2017), que encontraram relação positiva entre o sentimento contemporâneo e o retorno contemporâneo, e que a direção da relação do sentimento com o retorno costuma se reverter com o passar dos dias.

Esses resultados estão em conformidade com a literatura, pois vão ao encontro de estudos como os de Barbaries e Thaler (2003), Schnusenbergl e Madura (2001), Kling e Gao (2008) e Silva (2017), que mostram que vieses cognitivos têm impactos econômicos e estão associados a má precificação no curto prazo. Assim, acredita-se que, *a priori*, os investidores possuem um sentimento otimista, o que está associado a um alto retorno, porém, com o passar dos dias, eles notam que houve uma *overreaction* e essa relação se inverte, de modo que um sentimento otimista no tempo presente leva a um menor retorno no tempo futuro.

#### 4.3.2 Teste da HME controlado pelo volume de negócios

Uma vez que a presente análise se configura em uma análise complementar, e frente a alta extensão dos resultados, optou-se por apresentar nas tabelas apenas aquelas variáveis que demonstraram significância estatística. Assim, ao realizar um teste de eficiência de mercado sendo controlado pelo efeito do volume negociado, é possível perceber, por meio da Tabela 3, que o nível de volume negociado diário influencia o sentimento, em geral, de forma negativa, visto que apenas em *d*, no quantil q.10, foi que o volume negociado apresentou um efeito moderador positivo sobre o sentimento, indicando um aumento no sentimento com o aumento do volume negociado.

Esses resultados mostram que quando o retorno contemporâneo é baixo, o volume negociado irá influenciar positivamente a relação entre o sentimento e o retorno, de modo que um maior sentimento estará associado a um maior retorno, confirmando assim, para esse caso específico a primeira hipótese da pesquisa, de que quanto mais positivo é o sentimento do investidor, manifesto por meio do Twitter, maior é o retorno do IBOVESPA. Todavia ela é refutada nos demais tempos, uma vez que o sentimento contemporâneo de quando os retornos são altos (q.75), o do dia anterior e o de até cinco dias atrás, ao serem controlados pelo volume

de negócios, mostram que um maior sentimento está associado a menores retornos contemporâneos.

Os achados referentes a interação do sentimento e do volume de negócios fazem sentido, pois, acredita-se que quando os retornos são altos e há um sentimento otimista, os investidores podem ficar receosos de entrarem em um mercado em alta, de modo que eles não estejam dispostos a pagar preços que desviem muito dos valores intrínsecos dos ativos, e isso ocasiona um menor retorno. E de forma semelhante, é esperado que um maior sentimento passado esteja associado a um menor retorno futuro, assim como fora encontrado por Brown e Cliff (2004, 2005) e Baker e Wurgler (2006, 2007), porque com o passar do tempo os excessos de reação dos investidores tendem a cessar, e os ativos retornam para mais próximo de seus reais valores.

**Tabela 3** - Teste da HME controlado pelo volume de negócios – IBOV

Variáveis	q.10	q.25	q.50	q.75	q.90
<b>D</b>					
<i>Sent<sub>d</sub></i>	-0,0944** (0,0404)	-0,0157 (0,0460)	0,0505 (0,0375)	0,0990** (0,0498)	0,1926 (0,1297)
<i>Ret<sub>d-1</sub></i>	-0,0133 (0,0223)	-0,0145 (0,0113)	-0,0137** (0,0057)	-0,0304** (0,0128)	-0,0453 (0,0290)
<i>VN<sub>d</sub></i>	-6,54.10 <sup>-9***</sup> (6,08.10 <sup>-10</sup> )	-3,98.10 <sup>-9***</sup> (7,29.10 <sup>-10</sup> )	-2,47.10 <sup>-9</sup> (5,70.10 <sup>-10</sup> )	2,99.10 <sup>-9***</sup> (6,73.10 <sup>-10</sup> )	4,04.10 <sup>-9***</sup> (2,13.10 <sup>-9</sup> )
<i>Sent x VN<sub>d</sub></i>	3,05.10 <sup>-8***</sup> (1,30.10 <sup>-8</sup> )	5,57.10 <sup>-9</sup> (1,41.10 <sup>-8</sup> )	-1,36.10 <sup>-8</sup> (1,04.10 <sup>-8</sup> )	-2,88.10 <sup>-8***</sup> (1,41.10 <sup>-8</sup> )	-6,00.10 <sup>-8</sup> (4,21.10 <sup>-8</sup> )
<i>Ret x VN<sub>d</sub></i>	2,24.10 <sup>-7***</sup> (1,02.10 <sup>-8</sup> )	2,32.10 <sup>-7***</sup> (8,32.10 <sup>-9</sup> )	2,44.10 <sup>-7***</sup> (6,39.10 <sup>-9</sup> )	2,41.10 <sup>-7***</sup> (7,98.10 <sup>-9</sup> )	2,29.10 <sup>-7***</sup> (2,00.10 <sup>-8</sup> )
<b>D<sub>-1</sub></b>					
<i>Sent<sub>d</sub></i>	0,2027 (0,1480)	0,0954 (0,0854)	0,1683* (0,0887)	0,1069 (0,0846)	0,0933 (0,1599)
<i>Sent x VN<sub>d-1</sub></i>	-4,67.10 <sup>-8</sup> (-4,67.10 <sup>-8</sup> )	-3,11.10 <sup>-8</sup> (-4,67.10 <sup>-8</sup> )	-5,14.10 <sup>-8**</sup> (-4,67.10 <sup>-8</sup> )	-3,17.10 <sup>-8</sup> (-4,67.10 <sup>-8</sup> )	-2,10.10 <sup>-8</sup> (-4,67.10 <sup>-8</sup> )

Notas: Equação 5:  $Ret_t = \beta_0 + \beta_1 Sent_t + \gamma_1 \mathcal{L}_s(Ret_t) + \beta_2 VN_t + \beta_3 SentxVN_t + \gamma_2 \mathcal{L}_s(Ret_t) x VN_t + \varepsilon_t$

A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

A Tabela traz as variáveis defasadas em até um período; a continuação da

Tabela mostrando defasagens de dois a cinco dias se encontra no apêndice A.

**Fonte:** Elaboração Própria (2020)

Em relação aos retornos é possível perceber que os retornos do dia anterior são positivamente influenciados pelo volume negociado, de modo que há um aumento no retorno com o aumento do volume negociado. Assim, constata-se que, quanto maior for retorno do dia anterior, controlado pelo volume de negócios, maior também será o retorno no tempo atual. Todavia, com o passar dos dias essa relação se inverte, o que era esperado, uma vez que para Malkiel (2003) sempre haverá, em algum ponto, uma reversão na correlação entre os retornos passados e o retorno contemporâneo. Dessa maneira, vê-se que os retornos de cinco dias atrás,



sofrem influência negativa do volume negociado, de modo que quanto maior for o retorno passado, menor será o retorno contemporâneo.

No tocante ao volume de negócios, conforme era esperado, frente aos trabalhos de Kao, Chuang e Ku (2019), Do *et al.* (2014), Chen *et al.* (2015) e Sheng, Brzeszczyński e Ibrahim (2017), o volume de negócios mostrou ter associação significativa com o retorno, principalmente no tempo contemporâneo, de modo que o volume de negócios do dia atual se mostrou fortemente associado ao retorno do dia atual. Todavia, para Kao, Chuang e Ku (2019), o volume negociado passado de até três dias atrás estaria associado ao retorno contemporâneo, porém, no caso do estudo, foi visto que o volume negociado de até cinco dias atrás pode estar associado ao retorno contemporâneo.

De maneira mais específica, nota-se que quando o retorno é alto, um maior volume de negócios está relacionado a um maior retorno. Todavia, nota-se que quando os retornos são baixos, o volume negociado contemporâneo, e o de dois, quatro e cinco dias atrás irão estar negativamente associados ao retorno. Assim, um maior volume negociado ocasionará um menor retorno, o que pode dar indícios de que, quando o retorno é baixo, os investidores irão vender suas ações, de modo que o volume negociado aumentará, porém o retorno irá diminuir.

Os resultados encontrados para a seção 4.3.2 podem ser vistos de forma prática por meio da análise dos gráficos, onde, no dia 18/05/2017, quando ocorreu o “Joesley *day*”, o aumento do volume negociado parece estar relacionado a diminuição do sentimento, e consequente diminuição dos retornos, de modo que, embora tenha ocorrido um aumento abrupto no volume de negócios nesse dia específico, foi também nesse dia que o IBOVESPA atingiu o seu menor retorno da série histórica em análise.

#### 4.3.3 Relação entre o sentimento e o retorno controlada por demais variáveis - IBOV

Após o teste da eficiência de mercado, utilizou-se outro modelo, o qual os resultados estão apresentados na Tabela 4, de modo a controlar o efeito, para além do sentimento e do retorno, de variáveis tidas na literatura como responsáveis pela composição dos retornos contemporâneos, sendo essas: o fator HML, o *momentum*, a liquidez, a taxa livre de risco, a expectativa da inflação, a expectativa do PIB e o volume diário de *tweets*. Ressalta-se que, uma

vez que o sentimento é a variável de interesse, então, nesse modelo, apenas ele foi trabalhado com *lags*, sendo as demais variáveis trabalhadas no tempo contemporâneo.

Nesse sentido, nota-se que o sentimento se relaciona positivamente com o retorno no tempo contemporâneo, de modo que, quanto maior for o sentimento no dia atual, maior será o retorno do mercado nesse mesmo dia. Todavia, percebe-se que, como passar dos dias, há uma reversão, e o sentimento de um dia atrás e o sentimento de quatro dias atrás se relacionam negativamente com o retorno, quando o retorno é baixo ou é alto, respectivamente. Assim, nota-se que o sentimento obteve alguns valores distintos se comparados aos resultados do modelo anterior. Ao verificar a relação entre o sentimento e o retorno após a inclusão de variáveis de controle, novamente é possível confirmar a primeira hipótese da pesquisa para o tempo *d*, de que quanto mais positivo é o sentimento do investidor, manifesto por meio do Twitter, maior é o retorno do IBOVESPA. Todavia, ressalta-se que essa hipótese não pôde ser confirmada ao lidar com os sentimentos defasados, uma vez que esses apresentaram uma relação negativa com o retorno.

Os resultados confirmam a relação entre sentimento e retorno, que fora vista em estudos como os de Mao *et al.* (2012), Oliveira, Cortez e Areal (2017), Shen, Liu e Zhang (2018), e Ruan, Durresi, Alfantoukh (2018). Assim, eles trazendo indícios de que o sentimento extraído por meio do Twitter pode ter utilidade na previsão do retorno das ações, uma vez que se sabe que um sentimento otimista contemporâneo está associado a um alto retorno no tempo contemporâneo e a um retorno inferior no futuro.

Frente a isso, investidores podem utilizar estratégias de modo a vender ações quando o sentimento estiver alto, auferindo assim retornos acima da média, o que já fora provado por estudos como os de Kim, Ruy e Seo (2014), Wei, Mao e Wang (2016), e Renault (2017), onde utilizar o sentimento pode ser uma estratégia que obtém retornos superiores a *buy-and-hold*. Observa-se, ainda, que há a mesma tendência vista por Renault (2017), de que o sentimento possui associação positiva com o retorno, mas que a partir do dia posterior essa relação passa a ser negativa, de modo a crer que houve um *overreaction*, fazendo com que os ativos se afastem dos seus reais valores intrínsecos.

No que tange ao volume de *tweets*, percebe-se que existe uma associação positiva entre o volume de *tweets* e os retornos em todos os tempos analisados, quando os retornos são altos (q.90), de modo que, quanto maior o volume de *tweets* acerca das ações que constam no IBOVESPA, maior também tende a ser o retorno do índice. Assim, diferentemente de Nisar e

Yeung (2018), que não encontraram relação entre o volume de *tweets* e o retorno, os resultados aqui encontrados corroboram com os achados de Wei, Mao e Wang (2016) e Al-Nasseria e Ali (2018), mostrando que essa relação existe e é forte.

A taxa livre de risco foi significativa em todos os tempos trabalhados, apresentando coeficientes positivos em  $d-2$ ,  $d-4$ ,  $d-5$ , quando os retornos são baixos. Essa relação positiva era esperada, pois Galdi e Gonçalves (2018) identificaram coeficientes positivos na relação entre a taxa livre de risco e o retorno, todavia, os autores não haviam encontrado significância estatística dessa variável no mercado brasileiro, o que foi encontrado aqui. Ela também apresentou coeficientes negativos e significativos, em todos os tempos analisados, quando os retornos são altos. Assim, quando há altos retornos, quanto menor for a taxa livre de risco, maior será o retorno.

**Tabela 4** – Relação entre o sentimento e o retorno– IBOV

Variáveis	q.10	q.25	q.50	q.75	q.90
<b><math>D_{-0}</math></b>					
<b><math>Sent_d</math></b>	0,0454** (0,0299)	0,0243 (0,0155)	0,0116 (0,0164)	-0,0158 (0,0171)	0,0232 (0,0204)
<b><math>VDT_d</math></b>	$3,02 \cdot 10^{-6}$ ( $4,28 \cdot 10^{-6}$ )	$1,50 \cdot 10^{-7}$ ( $2,83 \cdot 10^{-6}$ )	$1,29 \cdot 10^{-6}$ ( $2,51 \cdot 10^{-6}$ )	$-5,33 \cdot 10^{-7}$ ( $4,86 \cdot 10^{-6}$ )	$7,71 \cdot 10^{-6***}$ ( $4,63 \cdot 10^{-6}$ )
<b><math>Riskfree_d</math></b>	48,9670 (33,3669)	28,7116 (17,7539)	12,9191 (14,9481)	-40,5485 (28,2044)	-62,5591*** (14,3039)
<b><math>HML_d</math></b>	0,7863*** (0,1260)	0,8125*** (0,0793)	0,6944*** (0,0677)	0,7527*** (0,1092)	0,8469*** (0,1097)
<b><math>MOM_d</math></b>	0,4023* (0,2179)	0,2919* (0,1700)	0,1677** (0,0759)	0,1055 (0,1185)	0,0793 (0,1048)
<b><math>Liq_d</math></b>	-0,5252*** (0,1342)	-0,6674*** (0,0826)	-0,7606*** (0,0690)	-0,7078*** (0,0801)	-0,8165*** (0,1207)
<b><math>Infla_d</math></b>	-0,0132 (0,0084)	-0,0041 (0,0056)	-0,0009 (0,0036)	0,0091 (0,0063)	0,0133*** (0,0038)
<b><math>PIB_d</math></b>	-0,0002 (0,0009)	-0,0003 (0,0009)	-0,0006 (0,0006)	-0,0021* (0,0011)	-0,0019* (0,0018)
<b><math>D_{-1}</math></b>					
<b><math>Sent_{d-1}</math></b>	-0,0288* (0,0159)	-0,0057 (0,0158)	-0,0069 (0,0110)	0,0007 (0,0200)	0,0081 (0,0220)
<b><math>VDT_d</math></b>	$4,97 \cdot 10^{-7}$ ( $2,25 \cdot 10^{-6}$ )	$-6,00 \cdot 10^{-7}$ ( $2,25 \cdot 10^{-6}$ )	$4,10 \cdot 10^{-7}$ ( $2,44 \cdot 10^{-6}$ )	$2,11 \cdot 10^{-7}$ ( $4,25 \cdot 10^{-6}$ )	$8,59 \cdot 10^{-6**}$ ( $4,33 \cdot 10^{-6}$ )
<b><math>Riskfree_d</math></b>	45,2367 (32,9780)	31,0035 (20,2900)	12,6860 (14,3200)	-43,9491 (27,4500)	-65,3865*** (15,5803)
<b><math>HML_d</math></b>	0,8951*** (0,1038)	0,8164*** (0,0969)	0,7240*** (0,0892)	0,7366*** (0,1123)	0,9010*** (0,1300)
<b><math>MOM_d</math></b>	0,3515** (0,1717)	0,2720** (0,1296)	0,1382 (0,1014)	0,1319 (0,1214)	0,0958 (0,1060)
<b><math>Liq_d</math></b>	-0,4902*** (0,1149)	-0,6717*** (0,0819)	-0,7795*** (0,0970)	-0,7051*** (0,1337)	-0,7702*** (0,0862)
<b><math>Infla_d</math></b>	-0,0138 (0,0092)	-0,0060 (0,0041)	-0,0010 (0,0026)	0,0010* (0,0052)	0,0117*** (0,0041)
<b><math>PIB_d</math></b>	-0,0008 (0,0012)	-0,0010 (0,0012)	-0,0008 (0,0012)	-0,0023 (0,0014)	-0,0029* (0,0016)

Notas: Equação 6:  $Ret_t = \beta_0 + \beta_1 Sent_t + \beta_2 VDT_t + \beta_3 RF_t + \beta_4 HML_t +$

---


$$\beta_5 MOM_t + \beta_6 Liq_t + \beta_7 Infl_t + \beta_8 PIB_t + \varepsilon_t$$

A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

A Tabela traz as variáveis defasadas em até um período; a continuação da

Tabela mostrando defasagens de dois a cinco dias se encontra no apêndice A.

---

**Fonte:** Elaboração própria (2020)

O fator HML apresentou coeficiente positivo e foi significativo em todos os quantis, em todos os tempos analisados, mostrando que quanto maior for o HML do mercado, maior também tende a ser o retorno do IBOVESPA. Esse resultado era esperado, uma vez que o HML é a diferença entre a carteira formada por empresas com alto BM e a carteira formada por empresas com baixo BM. Assim, isso implica dizer que empresas com um alto BM possuem maiores retornos, e isso está associado ao aumento do retorno do mercado. Isso ocorre porque, segundo Fama e French (1993), essas empresas têm maiores oportunidades de crescimento e valorização.

Também foi visto que há o efeito *momentum* no mercado brasileiro, ou seja, ações vencedoras costumam permanecer vencedoras, e esse efeito faz parte da composição dos retornos, uma vez que essa variável apresentou coeficientes positivos em todos os quantis, sendo esses significantes em todos os tempos analisados, especialmente quando os retornos são baixos. Esses resultados são próximos aos encontrados por Machado e Medeiros (2011) e Machado, Faff e Silva (2017), ao estudarem o mercado brasileiro, onde foi encontrado uma associação positiva entre os retornos e o fator *momentum*, sendo o prêmio desse fator positivo.

A liquidez, assim como o HML, apresentou coeficiente significativo em todos os quantis e tempos analisados, porém, eles foram negativos. Esse resultado era esperado, uma vez que Machado e Medeiros (2011) e Machado, Faff e Silva (2017), ao estudarem o mercado brasileiro, identificaram que, ao contrário do que se possa imaginar, portfólios formados por ações menos líquidas obtiveram retornos superiores aos portfólios formados por ações mais líquidas. Assim, acredita-se que, no mercado brasileiro, quanto menos líquida for a ação, maior será o retorno.

Os resultados referentes aos fatores de risco foram conforme os previstos pela literatura, especificamente no tocante ao estudo de Machado, Faff e Silva (2017), ao utilizarem o modelo de Keene e Peterson (2007) para o mercado brasileiro. Assim sendo, os resultados aqui encontrados ratificam as evidências encontradas por esses autores, os quais mostraram que tais

fatores apresentam evidências no mercado nacional, especialmente no que se refere aos fatores BM (representado pelo HML), *momentum* e liquidez.

A expectativa de inflação, em todos os tempos trabalhados mostrou coeficientes positivos quando os retornos são altos. Todavia, a variável apresentou coeficiente negativo em  $d$ ,  $d-2$  e  $d-4$  quando os retornos são baixos ou estão sobre a mediana. Assim, nota-se que, usualmente, quando há uma maior expectativa de inflação, os retornos são maiores. Porém, quando os retornos são baixos ou estão na mediana, uma maior expectativa de inflação está associada a um menor retorno.

A expectativa de crescimento do PIB apresentou coeficiente negativo em todos os quantis em análise, sendo significativa em  $d$ ,  $d-1$ ,  $d-2$ ,  $d-4$  e  $d-5$ , quando os retornos são maiores, sendo que em  $d-4$  e  $d-5$  ela também se mostrou significativa quando os retornos eram baixos. Assim, nota-se que quanto maior for a expectativa de crescimento do PIB, menores são os retornos. Os resultados referentes a expectativa de inflação divergem em partes do que fora encontrado por Callado *et al.* (2010), onde foi vista relação negativa e significativa entre o retorno e a inflação, uma vez que, no presente estudo, para além dessa relação também foram encontradas associações positivas entre o retorno e a inflação, quando os retornos são altos.

Os achados referentes a relação entre a expectativa do PIB e os retornos do mercado reforçam os achados de Machado, Gartner e Machado (2017) e Portela e Santos (2018), onde foi encontrada uma relação negativa e significativa entre os retornos do mercado acionário brasileiro e o PIB. Nesse sentido, para Machado, Gartner e Machado (2017), isso ocorre porque possivelmente os investidores se sentem otimistas para investirem em mercados com alta expectativa de crescimento, de modo que eles fazem ofertas acima da média para os preços das ações, o que acarreta diminuição dos retornos, uma vez que os dividendos permanecerão os mesmos.

Nota-se que, na seção 4.3, é notório um padrão de que, *a priori*, o sentimento e o retorno andarão na mesma direção, e que o sentimento de dias anteriores se relacionam com o retorno do dia atual, mas, também é visto que, usualmente, quando os retornos são muito altos, o passar dos dias fará com que essa relação sofra reversão, e o mesmo ocorre se os retornos forem muito baixos.

As tendências encontradas eram esperadas e podem ser vistas, por exemplo, por meio da análise dos gráficos, onde em períodos que o sentimento foi bem negativo, como quando

houve a operação Carne Fraca, o retorno também foi negativo, porém, o retorno voltou a ser positivo pouco tempo após, enquanto o sentimento demorou um pouco mais para sofrer essa reversão, o que mostra que, a priori o retorno e sentimento andam na mesma direção, mas, com o passar dos dias essa relação se inverte.

Os achados também podem trazer indícios de que, por mais que, em um primeiro momento, os investidores terão um sentimento negativo (ou positivo) e isso fará com que eles tomem decisões de investimento com base nesse sentimento. Todavia, com o passar dos dias, embora o sentimento permaneça, os investidores começam a ponderar suas decisões de investimento baseados na razão, delineando como as notícias recebidas irão realmente impactar os seus investimentos, e não puramente baseados no sentimento.

#### **4.4 Associação entre o volume de *tweets* e o volume negociado - IBOVESPA**

Uma vez que fora vista a relação existente entre o sentimento e o retorno no mercado acionário brasileiro, a pesquisa buscou testar a hipótese de que o volume de *tweets* teria relação com o volume negociado do mercado acionário brasileiro. Ademais, de modo a controlar o modelo, foram incluídas as variáveis de sentimento e do volume negociado passado no mesmo, visto que essas variáveis são tidas na literatura como responsáveis pelas alterações no volume negociado. As variáveis presentes nos modelos referentes ao volume negociado também foram trabalhadas com *lags* de até cinco dias, visto que o sentimento é uma das variáveis de interesse e optou-se por padronizar o tempo por meio dela, uma vez que a literatura mostra que, no curto prazo ela pode ter influência em até cinco dias sobre os movimentos do mercado.

A Tabela 5 traz os resultados da relação entre o volume de *tweets* e o volume negociado no mercado acionário brasileiro. Assim, ao analisar as variáveis presentes no modelo, o sentimento mostrou possuir relação negativa e significativa com o volume negociado no tempo  $d$  e em  $d-2$ , quando o volume negociado é baixo ou alto. Isso implica dizer que quanto mais otimista for o sentimento contemporâneo e o sentimento de dois dias atrás, menor será o volume negociado. Os resultados referentes a relação entre o sentimento e o volume de negócios demonstram que, diferentemente do que fora encontrado por Kim e Kim (2014) e por Oliveira, Cortez e Areal (2017), o sentimento das mensagens postadas *on-line* têm sim relação com o volume de negociações, assim como no estudo de Al-Nasseria e Ali (2018), Chen e Lo (2019) e Kim *et al.* (2019). Esses resultados também estão em conformidade com os achados de Souza,

Barbedo e Araújo (2018), que encontraram uma relação negativa entre o sentimento e o volume negociado.

**Tabela 5** - Associação entre o volume negociado e o volume de *tweets*

Variáveis	q.10	q.25	q.50	q.75	q.90
<b>D</b>					
<i>Sent<sub>d</sub></i>	-2247542 (6706971)	-3202946*** (1136787)	-1932573 (1185367)	-3312553 (5,30. 10 <sup>+7</sup> )	-7924769* (4564934)
<i>VDT<sub>d</sub></i>	-27,2298 (309,3477)	-90,8856 (174,808)	-158,4177 (227,2466)	313,6504 (3220,03)	1405,109 (1094,967)
<i>Vol<sub>d-1</sub></i>	0,3864*** (0,1230)	0,4142*** (0,0321)	0,5440*** (0,0539)	0,6018*** (0,0745)	0,6918*** (0,0863)
<b>D-1</b>					
<i>Sent<sub>d-1</sub></i>	-2465460 (1866864)	-1173580 (860846)	-777277 (1037249)	613094 (1515157)	-383233 (2760839)
<i>VDT<sub>d-1</sub></i>	-44,4864 (154,1766)	89,5479 (94,9045)	11,6595 (126,3461)	-30,0617 (163,9141)	-209,9057 (150,0361)
<i>Vol<sub>d-1</sub></i>	0,3539*** (0,0681)	0,4154*** (0,0563)	0,5511*** (0,0612)	0,6498*** (0,0773)	0,6530*** (0,0827)

Notas: Equação 7:  $VN_t = \beta_0 + \beta_1 Sent_t + \beta_2 VDT_t + \gamma_1 L_s(VN_t) + \varepsilon_t$

A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

A Tabela traz as variáveis defasadas em até um período; a continuação da

Tabela mostrando defasagens de dois a cinco dias se encontra no apêndice B.

**Fonte:** Elaboração própria (2020)

Em relação ao volume diário de *tweets* (VDT), percebe-se que ele apresentou associação significativa e negativa com o volume negociado em alguns quantis dos tempos *d-2*, *d-4* e *d-5*. Assim, em síntese, especialmente quando o volume de negócios é alto, um maior volume diário de mensagens postadas *on-line* no tempo passado estará associado a um menor volume de negócios no tempo contemporâneo.

Essa relação entre o volume de *tweets* e o volume de negócios era esperada, embora se acreditasse que ela seria positiva, frente ao que era trazido pela literatura (MAO ET AL., 2012; WEI; MAO; WANG, 2016). Assim, a segunda hipótese da pesquisa, de que o volume diário de *tweets* relacionados ao IBOVESPA possui associação positiva com volume negociado no mercado acionário brasileiro não pôde ser confirmada, todavia, essa diferença pode estar atrelada a questões referentes a amostra, ou porque no presente estudo a amostra também possui *tweets* tidos como neutros, o que pode ser responsável por tais resultados. Desse modo, uma análise subsequente, que analisará os *tweets* frente aos seus tipos de sentimento poderá clarificar esses achados.

As evidências encontradas acerca da relação entre o volume de *tweets* e o volume negociado são diferentes das encontradas por Nisar e Yeung (2018), uma vez que os autores não encontraram associação entre as mesmas e o presente estudo encontrou relação significativa entre essas variáveis. Esses resultados distintos podem estar associados ao período estudado, uma vez que os autores trabalharam com apenas cinco dias, e o presente estudo trabalha com dois anos. Outro fator pode ser devido a amostra, visto que Nisar e Yeung (2018) trabalharam com um mercado mais maduro, que foi o do Reino Unido, e o presente estudo utilizou o mercado brasileiro, que tem menor maturidade, e frente a esse fato, pode ser que seus movimentos estejam mais suscetíveis a estarem associados a fatores como o volume de mensagens que são postadas *on-line*.

No que se refere a relação entre o volume negociado passado e o volume negociado no dia atual, nota-se que o volume negociado defasado obteve coeficientes positivos e significativos em quase todos os tempos e quantis, excetuando apenas o quantil  $q.75$ , do tempo  $d$ . Dessa maneira, em geral, quanto maior for o volume negociado no passado, maior também será o volume negociado contemporâneo.

#### 4.4.1 Análise por Tipo de Sentimento dos *Tweets*

Uma vez que há *tweets* que são tidos como neutros, optou-se por utilizar um novo modelo de modo a destrinchar o volume de *tweets* em negativos e positivos, e verificar se haveria diferença na forma com que o volume de *tweets* positivos e o volume de *tweets* negativos estariam associados ao volume negociado. Os resultados desse modelo constam na Tabela 6. Assim, no que tange aos resultados, nota-se que o sentimento está negativamente associado ao volume negociado em  $d$ , quando o volume negociado é baixo ou alto. Dessa maneira, quanto maior for o sentimento contemporâneo, menor será o volume negociado, ratificando, novamente, os indícios trazidos pelos estudos de Al-Nasseria e Ali (2018) e Souza, Barbedo e Araújo (2018).

No que tange ao volume de *tweets* positivos, ele se mostrou estatisticamente significativo em alguns quantis nos tempos que variam de  $d$  até  $d-5$ , de modo que a segunda hipótese da pesquisa, de que o volume diário de *tweets* relacionados ao IBOVSPA possui associação positiva com volume negociado no mercado acionário brasileiro não foi confirmada, pois notou-se que, especialmente quando o volume negociado é alto, os volumes de *tweets* de



um até cinco dias atrás se relacionam negativamente com o volume negociado contemporâneo, de modo que, quanto maior for o volume de *tweets* positivos, menor será o volume negociado. Isso pode dar indícios de que, como o mercado já está em alta, então uma maior quantidade de mensagens postadas *on-line* pode ser vista pelos investidores como exagero, e por isso o volume negociado tenderá a retornar para valores mais próximos de sua média.

**Tabela 6** - Associação entre o volume negociado e os tipos de sentimento

Variáveis	q.10	q.25	q.50	q.75	q.90
<b>D</b>					
<i>Sent<sub>d</sub></i>	-1586632 (1571567)	-3280104*** (1338365)	-1720730 (1,89.10 <sup>+7</sup> )	-4300311* (3412030)	-842931 (1,20.10 <sup>+7</sup> )
<i>VTP<sub>d</sub></i>	-194,1355 (605,9065)	163,3134 (447,6480)	-202,3108 (603,6381)	736,0237 (1046,333)	144,187 (2961,613)
<i>VTN<sub>d</sub></i>	48,28383 (563,0039)	-373,2366 (729,6275)	-186,8227 (1838,7970)	221,9430 (3467,391)	7277,507 (9344,4020)
<i>Vol<sub>d-1</sub></i>	0,4110 *** (0,0641)	0,4265*** (0,0189)	0,5490*** (0,2177)	0,6005*** (0,0474)	0,6602*** (0,3311)
<b>D<sub>-1</sub></b>					
<i>Sent<sub>d-1</sub></i>	-1999135 (2490658)	-1256662 (1409824)	-116087 (2,01.10 <sup>+7</sup> )	1996617 (5180558)	2791712 (4233094)
<i>VTP<sub>d-1</sub></i>	350,1698 (763,5038)	210,7409 (386,0635)	-696,2313 (669,7108)	-950,9526* (969,3100)	-1748,307 (1830,3930)
<i>VTN<sub>d-1</sub></i>	-147,1529 (1171,469)	87,53969 (635,3683)	238,4604 (1256,581)	375,6633 (2284,2030)	472,47 (1829,5370)
<i>Vol<sub>d-1</sub></i>	0,3732*** (0,0568)	0,4143*** (0,0515)	0,5424*** (0,2597)	0,6639*** (0,0746)	0,6250*** (0,1130)

Notas:  $VN_t = \beta_0 + \beta_1 Sent_t + \beta_2 VTP_t + \beta_3 VTN_t \gamma_1 + \mathcal{L}_s(VN_t) + \varepsilon_t$   
A primeira linha traz os coeficientes e a estatística *p*-valor;  
O erro padrão está entre parênteses.  
\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.  
A Tabela traz as variáveis defasadas em até um período; a continuação da Tabela mostrando defasagens de dois a cinco dias se encontra no apêndice B.

**Fonte:** Elaboração própria (2020)

Referente ao volume de *tweets* negativos, nota-se que ele possui associação positiva e significativa com o volume negociado contemporâneo em alguns quantis dos tempos *d-2* e *d-3*. Assim, isso implica dizer que quanto maior for o número de mensagens postadas no Twitter que são dotadas de um sentimento negativo, maior também será o volume negociado, o que confirma, para esse caso, a segunda hipótese de pesquisa, de que o volume diário de *tweets* relacionados ao IBOVESPA possui associação positiva com o volume negociado no mercado acionário brasileiro. Dessa forma, acredita-se que os investidores costumam negociar mais quando as informações recebidas pelo mercado são negativas.

Os resultados referentes a relação entre o volume de *tweets* negativos e positivos e o volume negociado novamente confirmam que há essa associação entre o número de mensagens

que são postadas *on-line* e o volume de negócios do mercado brasileiro, todavia, esses resultados além de ratificarem evidências de estudos que foram anteriormente citados (MAO *ET AL.*, 2012; WEI; MAO; WANG, 2016), eles corroboram com os achados de Galdi e Gonçalves (2018), uma vez que os sentimentos textuais de cunho negativo parecem estar associado a um aumento no volume de negócios do mercado.

O volume negociado passado apresentou coeficientes positivos em quase todos os resultados, não sendo significativa apenas em *d-5*, quando o volume negociado é mediano. Dessa maneira, nota-se que quanto maior o volume negociado passado, maior será o volume negociado contemporâneo.

Os achados da seção 4.4 demonstram uma tendência de que, quanto maior o volume de *tweets*, menor será o volume negociado. Todavia, quando o volume de *tweets* é dotado de cunho negativo, um maior volume de mensagens postadas diariamente estará associado a um maior volume negociado. Um exemplo disso pode ser visto na análise dos gráficos, quando houve o “Joesley *day*”, onde é possível ver que houve um aumento no volume de *tweets*, e, visto que esses *tweets* tinham um cunho negativo, também houve um aumento no volume negociado. Outro ponto de destaque é que o sentimento se relaciona negativamente com o volume negociado, e isso também pode ser visto no caso do “Joesley *day*”, onde um menor sentimento esteve associado a um maior volume de negócios.

#### **4.5 Associação entre o sentimento e o retorno do SMLL**

De modo a verificar a robustez do índice que fora desenvolvido, optou-se por verificar a associação entre o sentimento e o retorno das empresas de menor capitalização (*small caps*). Ressalta-se que, devido à alta extensão dos *outputs* da presente análise, e uma vez que essa se configura em uma análise complementar, então, optou-se por apresentar nas tabelas apenas os resultados das variáveis que apresentaram significância estatística.

##### **4.5.1 Teste da HME utilizando *tweets* enquanto informação – SMLL**

A Tabela 7 apresenta os resultados referentes a relação entre o retorno das *small caps* e o sentimento. Nesse sentido, percebe-se que o sentimento foi significativo e positivo em alguns

quantis de  $d$  e  $d-4$ . Assim, nota-se que quando o retorno é baixo ou mediano, um maior sentimento no dia atual e em quatro dias atrás por parte dos investidores estará associado a um maior no retorno. Ou seja, quanto mais positivo for o sentimento, maior será o retorno, confirmando, para o caso das *small caps*, a primeira hipótese da pesquisa, de que quanto mais positivo é o sentimento do investidor, manifesto por meio do Twitter, maior é o retorno do índice de mercado, sendo representado, nesse caso, pelo SMLL, no lugar do IBOVESPA. Isso pode ocorrer porque quando os investidores estão com um sentimento mais otimista, acredita-se que eles estarão dispostos a comprar mais e a pagar mais pelas ações, o que estaria associado ao aumento dos retornos.

Esses achados são semelhantes ao que fora encontrado por Silva (2017), ao estudar o mercado brasileiro, que identificou a influência do sentimento contemporâneo e do sentimento de quatro dias atrás sobre o retorno contemporâneo. Todavia, no caso de Silva (2017), e do que fora visto nos resultados do presente estudo ao se trabalhar com o IBOVESPA, *a priori*, a relação entre o sentimento e o retorno é positiva e essa direção mudaria ao longo do tempo. Todavia, no que tange ao retorno das *small caps*, o sentimento parece estar positivamente associado a ele, e continua assim com o passar dos dias. Dessa maneira, no caso das empresas de menor capitalização, a ideia de que um sentimento otimista estaria ligado a um baixo retorno futuro, o que fora visto por estudos como os de Brown e Cliff (2004, 2005), Baker e Wurgler (2006, 2007) e Renault (2017), não se aplica. Ao menos não ao se utilizar esse modelo e no período que fora estudado.

**Tabela 7** - Teste da HME utilizando *tweets* enquanto informação – SMLL

Variáveis	q,10	q,25	q,50	q,75	q,90
<b>D</b>					
<b>Sent<sub>d</sub></b>	0,1052*	0,0636	0,0276**	0,0054	-0,0127
	(0,0547)	(0,0436)	(0,0124)	(0,0175)	(0,0199)
<b>D<sub>-2</sub></b>					
<b>Ret<sub>d-2</sub></b>	0,0231	0,0479	0,0072	0,0002**	0,0075
	(0,0923)	(0,0653)	(0,0495)	(0,0531)	(0,0830)
<b>D<sub>-3</sub></b>					
<b>Ret<sub>d-3</sub></b>	0,1261	-0,0245	-0,0451	-0,1279*	-0,1110*
	(0,1003)	(0,0804)	(0,0380)	(0,0755)	(0,0629)
<b>D<sub>-4</sub></b>					
<b>Sent<sub>d-4</sub></b>	0,0680**	0,0317**	0,0103	-0,0048	-0,0119
	(0,0356)	(0,0153)	(0,0126)	(0,0202)	(0,0182)
<b>D<sub>-5</sub></b>					
<b>Ret<sub>d-5</sub></b>	0,0198	0,0414	0,0647**	0,0854	-0,0544
	(0,1171)	(0,0755)	(0,0267)	(0,0432)	(0,0834)

Notas: Equação 4:  $Ret_t = \beta_0 + \beta_1 Sent_t + \gamma_1 \mathcal{L}_s(Ret_t) + \varepsilon_t$

A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

---

A Tabela traz as variáveis que foram significativas no período contemporâneo e quando defasadas em até um até cinco períodos.

---

**Fonte:** Elaboração própria (2020)

Com relação aos retornos, nota-se que foram encontrados coeficientes significativos e positivos em  $d-2$  e em  $d-5$ , de modo que quanto maior for o retorno de dois e de cinco dias atrás, maior é o retorno contemporâneo. Também foi encontrada relação significativa e negativa entre o retorno defasado e o retorno contemporâneo em  $d-3$ , o que traz indícios de que, quando o retorno é alto, quanto maior for o retorno de três dias atrás, menor será o retorno no tempo atual.

Esses resultados demonstram que os retornos de dois, de três e de cinco dias atrás compõem o retorno atual, todavia, também se nota que as informações passadas não estão totalmente incorporadas nos preços.

#### 4.5.2 Relação entre o Sentimento e o Retorno controlada por demais variáveis

A Tabela 8 traz os resultados acerca de como o sentimento se relaciona com o retorno ao controlarmos o efeito de outras variáveis. Assim, ao verificarmos a relação entre o sentimento e o retorno das empresas de menor capitalização, nota-se que o sentimento contemporâneo (em  $d$ ) se relaciona positivamente com o retorno. Logo, confirma-se, para esse caso, a primeira hipótese da pesquisa, de que quanto mais positivo é o sentimento do investidor, manifesto por meio do Twitter, maior é o retorno no mercado acionário brasileiro. Todavia, ao trabalhar com defasagens a direção dessa relação é invertida, uma vez que foram vistas associações negativas entre o sentimento e o retorno do SMLL em  $d-3$  e em  $d-5$ .

Os achados mostram que, *a priori*, quanto maior for o sentimento, maior também será o retorno das *small caps*, porém, com o passar dos dias o sentido dessa relação é alterado, de modo que um maior sentimento passado acarreta menor retorno contemporâneo. Essa tendência de o sentimento contemporâneo ter efeitos positivos sobre os retornos contemporâneos e posteriormente haver uma reversão dessa direção está em conformidade com os achados de Renault (2017). Assim, acredita-se que essa reversão ocorra porque os investidores reagem de forma exagerada frente as notícias, elevando os preços das ações e conseqüentemente seu retorno, porém, com o passar dos dias o mercado nota que houve super-reação e os preços tendem a voltar ao normal. Essa tendência de reversão pode ser vista por meio da análise

gráfica, a qual fora tratada anteriormente (seção 4.2). Ademais, esse movimento também está ligado a ideia de Brown e Cliff (2004, 2005) e Baker e Wurgler (2006, 2007), de que um sentimento otimista no hoje acarreta baixos retornos posteriormente.

Embora encontrada relação entre o retorno do SMLL e o sentimento, nota-se que o sentimento foi mais significativo ao analisar a relação do sentimento com o retorno do IBOVESPA. Todavia, destaca-se que, na literatura anterior, estudos como os de Baker e Wurgler (2006, 2007), Lee, Jiang e Indro (2002) e Oliveira, Cortez e Areal (2017) apresentam indícios de que o sentimento tem maiores efeitos sobre empresas menores. Todavia, os resultados aqui encontrados podem ser justificados pelo fato de o índice desenvolvido ter sido construído dando maior enfoque a empresas de maior capitalização.

O volume de *tweets* se relaciona positivamente com o retorno em quase todos os tempos analisados, quando os retornos são altos, excetuando o tempo  $d-3$ , quando os retornos são baixos. Assim, frequentemente, quanto maior for a quantidade de *tweets*, maior será o retorno das ações de menor capitalização. Dessa maneira, acredita-se que o mercado absorva essas novas informações que estão chegando por meio do Twitter e essas informações façam parte da composição dos preços, o que pode explicar essa associação existente entre as mensagens postadas *on-line* e o retorno do SMLL. Esse resultado é semelhante ao encontrado para o IBOVESPA, mostrando mais uma vez que, diferentemente do que fora visto por Nisar e Yeung (2018), no contexto brasileiro, o retorno, tanto de empresas de grande capitalização quanto de empresas de menor capitalização possui relação com o volume de mensagens que são postadas *on-line*.

**Tabela 8** - Relação entre o sentimento e o retorno – SMLL

Variáveis	q,10	q,25	q,50	q,75	q,90
$T_{-0}$					
<i>Sent<sub>d</sub></i>	0,0827* (0,0623)	0,0327 (0,0368)	0,0255** (0,0184)	0,0112 (0,0148)	0,0190 (0,0226)
<i>Riskfree<sub>d</sub></i>	49,5435 (46,0136)	15,2345 (36,7060)	2,2152 (27,7199)	-33,0932 (30,0944)	-60,3222* (38,2762)
<i>Infla<sub>d</sub></i>	-0,0119 (0,0092)	-0,0041 (0,0059)	0,0013 (0,0055)	0,0075* (0,0051)	0,0089 (0,0060)
<i>PIB<sub>d</sub></i>	0,0003 (0,0021)	-0,0008 (0,0014)	-0,0002 (0,0012)	-0,0012 (0,0016)	-0,0027* (0,0019)
<i>VTT<sub>d</sub></i>	-2,48e.10 <sup>-6</sup>	-1,46e.10 <sup>-6</sup>	1,52.10 <sup>-6</sup>	2,18.10 <sup>-6</sup>	0,00001**
$T_{-1}$					
<i>Riskfree<sub>d</sub></i>	45,0111 (54,1907)	3,7962 (34,5986)	-0,1379 (34,1112)	-42,0101* (16,5409)	-73,7015** (44,5107)
<i>Infla<sub>d</sub></i>	-0,0114 (0,0103)	-0,0052 (0,0054)	0,0016 (0,0068)	0,0085 (0,0033)	0,0112* (0,0081)
<i>PIB<sub>d</sub></i>	-0,0001	-0,0019*	-0,0007	-0,0018***	-0,0035**

	(0,0020)	(0,0013)	(0,0013)	(0,0013)	(0,0021)
<b>VTT<sub>d</sub></b>	-8,05.10 <sup>-6</sup>	-7,92.10 <sup>-7</sup>	6,73.10 <sup>-7</sup>	5,00.10 <sup>-8</sup>	0,00001*
	(6,81.10 <sup>-6</sup> )	(4,62.10 <sup>-6</sup> )	(2,22.10 <sup>-6</sup> )	(4,20.10 <sup>-6</sup> )	(5,26.10 <sup>-6</sup> )

Nota: Equação 9:  $Ret_t = \beta_0 + \beta_1 Sent_t + \beta_2 VTT_t + \beta_3 RF_t + \beta_4 Infl_t + \beta_5 PIB_t + \varepsilon_t$

A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

A Tabela traz as variáveis defasadas em até um período; a continuação da

Tabela mostrando defasagens de dois a cinco dias se encontra no apêndice C.

**Fonte:** Elaboração própria (2020)

A respeito das variáveis macroeconômicas, a taxa livre de risco apresentou uma relação negativa e significativa com o retorno em *d*, *d-1* e em *d-4*. Assim, quando os retornos são altos, quanto maior for a taxa de juros, menores são os retornos. Esse resultado é esperado, uma vez que no estudo de Machado, Faff e Silva (2017), ao trabalharem com o fator mercado, que corresponde à diferença entre o retorno do mercado e a taxa livre de risco, esse fator foi positivo e significativo, de modo que quanto maior ele for, maior será o retorno. Assim, automaticamente, quando se tem uma taxa livre de risco menor, o fator mercado será maior, e o retorno do mercado também será maior. Ademais, os resultados diferem dos de Galdi e Gonçalves (2018), uma vez que eles haviam encontrado coeficientes positivos ao analisarem essa relação, apesar de que, no caso desses autores, a relação não foi estatisticamente significativa. Essa diferença pode estar associada as diferenças de período estudado, ou porque os autores trabalharam com empresas de alta capitalização.

Esse resultado também pode trazer indícios de que, quando a taxa livre de risco está baixa, mais investidores irão se voltar para o mercado, visto que se acredita que ele dará uma maior rentabilidade, fato que pode ser visto por meio da análise descritiva, em que o valor médio do retorno diário do mercado foi de 0,0009, enquanto que a taxa livre de risco foi de 0,0003. Essa expectativa de maior rentabilidade pode estar associada ao aumento do retorno das ações. A expectativa de inflação se mostrou positivamente associada aos retornos em *d* e *d-1*, quando os retornos são altos. Assim, quando há altos retornos, uma maior expectativa de inflação estará ligada a maiores retornos. Todavia, também houve resultados semelhantes aos de Callado *et al.* (2010), onde houve associação negativa entre a inflação e os retornos, sendo visto que, quando os retornos são baixos, quanto maior for a inflação, menores são os retornos.

A expectativa de crescimento do PIB apresentou coeficientes negativos e significativos em alguns quantis de *d*, *d-1*, *d-4*, e *d-5*, o que corrobora com os achados de Machado, Gartner e Machado (2017) e Portela e Santos (2018), de que o PIB teria relação com os retornos no mercado brasileiro. Desse modo, nota-se que quanto maior for a expectativa do PIB, menor será

o retorno. Nesse sentido, Machado, Gartner e Machado (2017) acreditam que essa relação ocorram porque a expectativa de crescimento do PIB causa otimismo nos investidores, que se sentem mais seguros para investir no mercado e acabam fazendo com que os preços se desviem de seus valores intrínsecos, o que acarreta diminuição dos retornos uma vez que os dividendos continuarão os mesmo e esses aumentos logo irão se reverter.

Os resultados encontrados na seção 4.5 apresentam uma tendência semelhante ao que fora visto na seção 4.3, ao se trabalhar com o índice de empresas de maior capitalização. Assim, nota-se que, no primeiro momento, o sentimento e o retorno se relacionam na mesma direção, porém, com o passar dos dias há uma reversão dessa relação. Essa relação pode ser vista, por exemplo, na análise dos gráficos, por meio do dia 08/10/2018, onde, após o primeiro turno da eleição para presidente da república, o mercado possuía um alto sentimento (otimista), e isso acarretou um alto retorno. Todavia, com o passar dos dias o mercado parece ter percebido que houve uma *overreaction*, e os retornos voltaram a patamares mais próximos do que estavam antes desse dia.

#### **4.6 Resultados no Contexto Brasileiro**

Uma vez que uma das justificativas da importância do estudo consiste em verificar se um índice de sentimento desenvolvido por meio de atividades *on-line* é viável em um contexto de mercado emergente, optou-se por trazer, de forma resumida, se os resultados aqui encontrados seguem a mesma linha dos que foram achados pela literatura no que tange aos países desenvolvidos, de modo a reforçar a validade do índice que aqui foram desenvolvido.

Os resultados aqui encontrados da relação entre o sentimento dos investidores e o retorno no mercado brasileiro vão ao encontro dos resultados encontradas ao estudarem países desenvolvidos, pois, a maioria dos estudos internacionais demonstram haver relação entre o sentimento obtidos por meio de atividades *on-line* e os retornos, e os resultados aqui encontrados apontam para essa mesma direção, assim como nos estudos de Mao *et al.* (2012), Kim e Kim (2014), Oliveira, Cortez e Areal (2017), Renault (2017), Das, Behera e Rath (2018) e Al-Nasseria e Ali (2018).

No que se refere a relação entre o volume de *tweets* e o volume de negócios, as evidências internacionais eram conflitantes, especialmente para países desenvolvidos, pois

estudos como os de Nisar e Yeung (2018) e Oliveira, Cortez e Areal (2017) não viam potencial nessa relação. Por outro lado, Mao *et al.* (2012) e Wei, Mao e Wang (2016) traziam indícios de que essa relação poderia existir. Assim, os resultados aqui encontrados corroboram com as evidências de Mao *et al.* (2012) e Wei, Mao e Wang (2016), visto que os achados mostram que há forte relação entre o volume de *tweets* e o volume de negócios no contexto brasileiro.

A literatura precedente, em um contexto internacional, também apresentava contradições no que tange a relação entre o sentimento e o volume de negócios. Todavia, a maioria dos estudos, tais como os de Mao *et al.* (2012), Al-Nasseria e Ali (2018), Chen e Lo (2019), e Kim *et al.* (2019) afirmam que tal relação existe, a qual também fora encontrada no presente estudo.

Frente a comparação dos resultados encontrados para o contexto brasileiro diante dos achados de estudos internacionais, especialmente no que tange ao estudo de países desenvolvidos, é notório que, em sua maioria, os achados para o mercado brasileiro são semelhantes aos resultados encontrados para países desenvolvidos. Assim, essa semelhança encontrada dá base para a afirmação de que o uso de um índice de sentimento obtido por meio de atividades *on-line* também é útil, viável, e aponta para as mesmas direções em um contexto de mercado emergente. Ademais, os achados dão suporte ao uso do índice desenvolvido.

#### **4.7 Testes Adicionais Não Reportados**

Foram realizados exames adicionais, onde, primeiramente, fora considerado o sentimento no período noturno, das 18h às 18h, de modo a verificar a relação desse sentimento com os movimentos do mercado, todavia, foram vista correlações muito fracas ou inexistentes entre o sentimento medido dessa maneira e os movimentos do mercado acionário brasileiro, de modo que, optou-se por trabalhar exclusivamente com o sentimento no período de 24h.

Adicionalmente, também foram realizados exames considerando os anos de 2017 e 2018 separadamente. Ademais, visto que a análise gráfica mostra que próximo ao período eleitoral e que nos meses de maio de 2017 e maio de 2018 ocorreram alguns eventos que poderiam trazer impactos para análise, optou-se por realizar exames em que o período eleitoral fora retirado de 2018, e também exames de modo a verificar como o sentimento foi significativo apenas no período eleitoral. Também foram realizados exames considerando a amostra sem os meses de



maio de 2017 e 2018, e exames apenas nesses meses, todavia, em todos esses casos citados, os resultados obtidos foram semelhantes aos que foram encontrados ao considerar a amostra como um todo, por isso optou-se por manter toda a amostra e por não reportar os resultados dessas análises adicionais.

## 5 CONSIDERAÇÕES FINAIS

O objetivo principal do estudo foi analisar a relação do sentimento dos investidores, identificado por meio de *tweets*, com os retornos e os volumes negociados no mercado acionário brasileiro. Diante disso, conjecturaram-se duas hipóteses de pesquisa, sendo essas que: quanto mais positivo é o sentimento do investidor, manifesto por meio do Twitter, maior é o retorno do IBOVESPA (H1); e que o volume diário de *tweets* relacionados ao IBOVESPA possui associação positiva com volume negociado no mercado acionário brasileiro (H2).

Frente as análises, é possível observar que a primeira hipótese da pesquisa pôde ser confirmada, a qual afirmava que quanto mais positivo é o sentimento do investidor, manifesto por meio do Twitter, maior é o retorno do IBOVESPA. Todavia, ressalta-se que foi visto um padrão nas análises, em que, normalmente a relação entre o sentimento e o retorno será positiva no tempo contemporâneo, de tal maneira que um sentimento otimista no dia atual estará associado a um maior retorno no dia atual, porém, com o passar do tempo essa relação costuma se reverter, de modo que quanto maior for o sentimento passado, menor será o retorno contemporâneo.

Acredita-se que o padrão identificado na relação entre o sentimento e o retorno ocorre porque, *a priori*, quando os investidores recebem boas notícias, eles podem deixar as suas cognições serem afetadas, e investirem em determinadas ações simplesmente por terem visto notícias positivas sobre a mesma, sem ao menos analisar qual seria a real expectativa de fluxo de caixa futuro da ação, tomando suas decisões com base em heurísticas, de maneira simples e sem que haja uma maior reflexão. Porém, como o passar dos dias o mercado parece se dar conta de que houve uma reação exagerada frente as notícias boas, e a relação entre o sentimento e o retorno irá se reverter, de modo que os preços voltem para valores mais próximos do patamar onde eles estavam antes de haver essa reação exagerada. Essa tendência de os preços dispararem e voltarem, com o passar dos dias, para próximo da média pôde ser claramente vista por meio dos gráficos dos retornos, apresentado nos resultados.

A segunda hipótese da pesquisa, a qual pressupunha que o volume diário de *tweets* relacionados ao IBOVESPA possui associação positiva com volume negociado no mercado acionário brasileiro só pôde ser confirmada ao fazer uma análise separando os tipos de volumes negociados em positivos e negativos, de maneira que, apenas quando as mensagens postadas diariamente *on-line* são dotadas de sentimento negativo é que, quanto maior for o volume de

*tweets*, maior será o volume negociado. Todavia, salienta-se que foram encontradas fortes associações entre o volume de *tweets* e o volume de negócios em todas as análises, porém essa relação era negativa.

Esse padrão faz sentido, pois, por meio da análise gráfica é possível perceber que o maior pico de volume negociado foi justamente quando ocorreu o “Joesley *day*”, o qual esteve acompanhado de uma série de notícias que eram tidas como negativas na visão do mercado. Adicionalmente, também foi visto que os volumes de *tweets* atingiram seu primeiro e segundo maior pico durante a operação Carne Fraca e durante o “Joesley *day*”, respectivamente. Assim, já havia indícios, por meio da análise gráfica, de que notícias ruins teriam maior efeito sobre o mercado brasileiro do que notícias boas. Adicionalmente, Galdi e Golçalves (2018) também já tinham visto, anteriormente, que o mercado brasileiro dá mais ênfase a coisas negativas.

Os resultados referentes ao volume negociado podem ser explicados por fatores psicológicos, pois, acredita-se, com base em estudos como o de Kahneman e Tversky (1979), que os investidores tenham uma percepção relacionada a perda 2,25x maior do que a percepção de ganho, de modo que quando há risco de perda, os investidores irão tentar se proteger. Isso pode ser visto por meio dos resultados obtidos, pois, parece que, um maior volume de notícias negativas irá causar uma percepção de risco aos investidores, e isso está associado ao aumento no volume de negócios que ocorre na bolsa, pois, essa seria uma maneira que os investidores veem de tentar se proteger frente a percepção de risco.

Destarte, os resultados confirmam que o sentimento e o volume de *tweets* possuem associação com os movimentos que ocorrem no mercado acionário brasileiro, trazendo indícios de que, possivelmente, os investidores brasileiros operam ao mesmo tempo que utilizam o Twitter. Assim, os achados contribuem de modo a mostrar que há potencial em utilizar atividades *on-line* como ferramenta auxiliar na previsibilidade das oscilações que ocorrem no mercado brasileiro, visto que o sentimento e o volume de *tweets* têm potencial para serem utilizados como variáveis em estratégias de investimento, como por exemplo, investidores podem vender suas ações em períodos em que há excesso de otimismo, visto que os preços estarão inflacionados.

Dessa maneira, o estudo serve para mostrar aos participantes do mercado que investidores capazes de obter e processar informações provenientes do Twitter de maneira rápida podem as utilizar em estratégias de modo a tentar bater o mercado. Assim, esses investidores podem aproveitar para vender ações que estejam com um sentimento otimista no

dia atual, de modo a obter retornos anormais, porém é preciso ter cuidado e não comprar ações que estavam atreladas a um sentimento otimista por parte dos demais investidores dias atrás, pois, normalmente, um sentimento otimista passado estará associado a menores retornos no futuro. Assim, a estratégia de utilizar o sentimento é mais aplicável ao utilizar o sentimento no tempo atual para obter retornos no dia atual.

Os resultados do estudo também trazem contribuições a literatura, de modo a ratificá-la, trazendo uma série de resultados semelhantes ao que foram encontrados anteriormente, e a expandi-la, uma vez que o estudo mostrou como as atividades *on-line*, e especificamente o sentimento obtido por meio da rede social Twitter, tem efeitos sobre um mercado emergente, como é o mercado brasileiro, uma vez que há limitação na literatura sobre como os mercados emergentes reagem ao sentimento expresso de forma *on-line*. Ademais, a pesquisa se destaca por trazer uma nova abordagem para o problema, visto que ela não fez uso de pesquisas de opinião ou de variáveis de mercado para auferir o sentimento dos investidores, mostrando que é possível utilizar esse novo tipo de método em um contexto nacional, especialmente porque também há limitação de literatura nacional que verifique como o sentimento de atividades *on-line* afeta o mercado brasileiro.

Por fim, recomenda-se que novas pesquisas que utilizem o Twitter para auferir sentimento possam averiguar se no contexto brasileiro, ao ponderar os *tweets* pelo número de seguidores de quem está postando, essa variável irá apresentar maior relação com o mercado, de modo a verificar se pessoas de grande visibilidade, como investidores institucionais, *digital influencers* etc, fariam com que o sentimento tivesse uma maior associação com o mercado. Também recomenda-se controlar *noise traders*, de modo a captar o sentimento com um viés mais próximo da racionalidade. Para além, recomenda-se a captura do sentimento por meio de outras redes sociais, como o Facebook, e de ferramentas de pesquisa, como a Google Trends, para que se possa verificar se seu impacto no mercado brasileiro é semelhante ao que ocorre com o Twitter.

## REFERÊNCIAS

- AGARWAL, S.; KUMAR, S.; GOEL, U. Stock market response to information diffusion through internet sources: A literature review. **International Journal of Information Management**, v. 45, p. 118-131, 2019.
- AL-NASSERI, A.; ALI, F. What Does Investors' Online Divergence of Opinion Tell Us About Stock Returns and Trading Volume. **Journal of Business Research**, v. 86, p. 166-178, 2018.
- ARAÚJO, T.; ELEUTÉRIO, S.; LOUÇÃ, F. Do sentiments influence market dynamics? A reconstruction of the Brazilian stock market and its mood. **Physica A: Statistical Mechanics and its Applications**, v. 505, p. 1139-1149, 2018.
- ATKINS, A.; NIRANJAN, M., GERDING, E. Financial news predicts stock market volatility better than close price. **The Journal of Finance and Data Science**, n. 4, p. 120-137, 2018.
- BACHELIER, L. Théorie de la spéculation. **Annales scientifiques de l'École normale supérieure**, v. 17, p. 21-86, 1900.
- BAKER, M.; WURGLER, J. Investor sentiment and the cross-section of stock returns. **Journal of Finance**, v. 61, p. 1645–1680, 2006.
- BAKER, M.; WURGLER, J. Investor Sentiment in the Stock Market. **Journal of Economic Perspectives**, v. 21, n. 2, p. 129–151, 2007.
- BARBERIS, N.; THALER, R. A Survey of Behavioral Finance. **Handbook of the Economics of Finance**, v. 1, p. 1053-1128, 2003.
- BARBERIS, N.; SHLEIFER, A.; VISHNY, R. A model of investor sentiment<sup>1</sup>. **Journal of financial economics**, v. 49, n. 3, p. 307-343, 1998.
- BATHIA, D.; BREDIN, D. Investor sentiment: Does it augment the performance of asset pricing models? **International Review of Financial Analysis**, v. 59, p. 290-303, 2018.
- BERNARD, V.; THOMAS, K. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. **Journal of Accounting and Economics**, v. 13, n. 4, p. 305-340, 1990.
- BLACK, F. Noise. **The journal of finance**, v. 41, n. 3, p. 528-543, 1986.
- BLACK, F. Capital market equilibrium with restricted borrowing. **The journal of business**, v. 45, n. 3, p. 444-455, 1972.
- BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. **Journal of Computational Science**, v. 2, n. 1, p. 1–8, 2011.
- BROWN, G.; CLIFF, M. Investor sentiment and asset valuation. **The Journal of Business**, v. 78, n. 2, p. 405–440, 2005.

BROWN, G.; CLIFF, M. Investor sentiment and the near-term stock market. **Journal of Empirical Finance**, v. 11, n. 1, p. 1–27, 2004.

CALLADO, A; CALLADO, A; MÖLLER, H.; LEITÃO, C. Relações entre os Retornos das Ações e Variáveis Macroeconômicas: um Estudo entre Empresas do Setor de Alimentos e Bebidas através de Modelos APT. **Sociedade, Contabilidade e Gestão**, v. 5, n. 1, 2010.

CAMARGOS, M.; BARBOSA. Teoria e evidência da eficiência informacional do mercado de capitais brasileiro. **REGE Revista de Gestão**, v. 10, n. 1, 2010.

CARHART, M. On persistence in mutual fund performance. **The Journal of finance**, v. 52, n. 1, p. 57-82, 1997.

CHEN, G.; QIU, T.; JIANG, X.; ZHONG, L.; WU, X. How trading volume responds to return in financial dynamics? **Physica A: Statistical Mechanics and Its Applications**, v. 424, p. 73–81, 2015.

CHEN, H.; LO, T. Online search activities and investor attention on financial markets. **Asia Pacific Management Review**, v. 24, n. 1, p. 21-26, 2019.

CHEN, L., ZHANG, L. A Better three-factor model that explains more anomalies. **The Journal of Finance**, v. 65, n. 2, 2010.

DAS, S; BEHERA, R.; RATH, S. Real-time sentiment analysis of Twitter streaming data for stock prediction. **Procedia computer science**, v. 132, p. 956-964, 2018.

DEBATA, B.; DASH, S.; MAHAKUD, J. Investor sentiment and emerging stock market liquidity. **Finance Research Letters**, v. 26, p. 15-31, 2018.

DO, H.; BROOKS, R.; TREEPONGKARUNA, S.; WU, E. How does trading volume affect financial return distributions? **International Review of Financial Analysis**, v. 35, p. 190–206, 2014.

EDMANS, A.; GARCIA, D.; NORLI, Ø. Sports sentiment and stock returns. **The Journal of Finance**, v. 62, n. 4, p. 1967-1998, 2007.

EDWARDS, W. Conservatism in human information processing. **Formal representation of human judgment**, 1968.

FAMA, E. Efficient markets: a review of theory and empirical work. **Journal of Finance**, v. 25, n. 2, p. 383-417, 1970.

FAMA, E. F. Efficient capital markets: II. **Journal of Finance**, v. 46, n. 5, p. 1575-1617, 1991.

FAMA, E. F.; FISHER, L; JENSEN, M.; ROLL, R. The adjustment of stock prices to new information. **International economic review**, v. 10, n. 1, p. 1-21, 1969.

- FAMA, E.; FRENCH, K. A five-factor asset pricing model. **Journal of financial economics**, v. 116, n. 1, p. 1-22, 2015.
- FAMA, E.; FRENCH, K. Common risk factors in the returns on stocks and bonds. **Journal of financial economics**, v. 33, n. 1, p. 3-56, 1993.
- FAMA, E.; FRENCH, K. The cross-section of expected stock returns. **The Journal of Finance**, v. 47, n. 2, p. 427-465, 1992.
- FAN, W.; GORDON, M. The power of social media analytics. **Communications of the ACM**, v. 57, n. 6, p. 74–81, 2014.
- FRIEDMAN, M.; FRIEDMAN, M. **Essays in positive economics**. University of Chicago Press, 1953.
- KAHNEMAN, D.; TVERSKY, A. Prospect Theory: an Analysis of Decision under Risk. **Econometrica**, v. 47, p. 263-291, 1979.
- KAO, Y.; CHUANG, H.; KU, Y. The Empirical Linkages among Market Returns, Return Volatility, and Trading Volume: Evidence from the S&P 500 VIX Futures. **The North American Journal of Economics and Finance**, *In press*, 2019.
- KAPLANSKI, G.; LEVY, H. Sentiment and stock prices: The case of aviation disasters. **Journal of Financial Economics**, v. 95, n. 2, p. 174-201, 2010.
- KEENE, M.; PETERSON, D. The importance of liquidity as a factor in asset pricing. **Journal of Financial Research**, v. 30, n. 1, p. 91-109, 2007.
- KIM, J.; RYU, D.; SEO, S. Investor sentiment and return predictability of disagreement. **Journal of Banking & Finance**, v. 42, p. 166-178, 2014.
- KIM, N; LUČIVJANSKÁ, K.; MOLNÁR, P.; VILLA, R. Google searches and stock market activity: Evidence from Norway. **Finance Research Letters**, v. 28, p. 208-220, 2019.
- KIM, S.; KIM, D. Investor sentiment from internet message postings and the predictability of stock returns. **Journal of Economic Behavior & Organization**, v. 107, p. 708-729, 2014.
- KLIBANOFF, P.; LAMONT, O.; WIZMAN, T. Investor reaction to salient news in closedend country funds. **Journal of Finance**, v. 53, n. 2, p. 673-699, 1998.
- KLING, G.; GAO, L. Chinese institutional investors' sentiment. **Journal of International Financial Markets, Institutions and Money**, v. 18, n. 4, p. 374-387, 2008.
- LEE, W.; JIANG, C.; INDRO, D. Stock market volatility, excess returns, and the role of investor sentiment. **Journal of banking & Finance**, v. 26, n. 12, p. 2277-2299, 2002.
- LI, Q.; LI, S.; XU, L. National elections and tail risk: International evidence. **Journal of Banking and Finance**, v. 88, p. 113–128, 2018.

LINTNER, J. Security prices, risk, and maximal gains from diversification. **The journal of finance**, v. 20, n. 4, p. 587-615, 1965.

MACHADO, M. A. V.; FAFF, R.; SILVA, S. C. Applicability of Investment and Profitability Effects in Asset Pricing Models, **RAC**, v.21, n.6, p. 851-874, 2017.

MACHADO, M.; MEDEIROS, O. Modelos de precificação de ativos e o efeito liquidez: evidências empíricas no mercado acionário brasileiro. **Revista Brasileira de Finanças**, v.9, p. 383-412, 2011.

MACHADO, M.; GARTNER, I.; MACHADO, L. Relação entre Ibovespa e Variáveis Macroeconômicas: Evidências a Partir de um Modelo Markov-Switching. **Revista Brasileira de Finanças**, v. 15, n. 3, p. 435-468, 2017.

MALKIEL, B. The Efficient markets hypothesis and its critics. **Journal of Economic Perspectives**, v. 17, n. 1, p. 59-82, 2003.

MAO, H.; COUNTS, S.; BOLLEN, J. Predicting financial markets: Comparing survey, news, twitter and search engine data. **arXiv preprint arXiv:1112.1051**, 2011.

MAO, Y.; WEI, W.; WANG, B., LIU; B. Correlating S&P 500 stocks with Twitter data. **In: Proceedings of the first ACM international workshop on hot topics on interdisciplinary social networks research**, p. 69-72, ACM, 2012.

MENKHOFF, L. The noise trading approach—questionnaire evidence from foreign exchange. **Journal of International Money and Finance**, v. 17, n. 3, p. 547-564, 1998.

MOSSIN, J. Equilibrium in a Capital Asset Market. **Econometrica**, v. 34, n. 768-783, 1966.

NISAR, T.; YEUNG, M. Twitter as a tool for forecasting stock market movements: A short-window event study. **The Journal of Finance and Data Science**, v. 4, p. 101-119, 2018.

OLIVEIRA, N.; CORTEZ, P.; AREAL, N. The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. **Expert Systems with Applications**, v. 73, p. 125-144, 2017.

PORTELA, D.; SANTOS, J. Risco macroeconômico e o modelo de cinco fatores no mercado acionário brasileiro. **REAd. Revista Eletrônica de Administração**, v. 24, n. 3, p. 269-293, 2018.

QADAN, M.; NAMA, H. Investor sentiment and the price of oil. **Energy Economics**, v. 69, p. 42-58, 2018.

RENAULT, T. Intraday online investor sentiment and return patterns in the US stock market. **Journal of Banking & Finance**, v. 84, p. 25-40, 2017.

ROBERTS, H. Statistical versus clinical prediction of the stock market. **In Conference of Securities Price Analysis**, Chicago, 1967.



RUAN, Y.; DURRESI, A.; ALFANTOUKH, L. Using Twitter trust network for stock market analysis. **Knowledge-Based Systems**, v. 145, p. 207-218, 2018.

SCHNUSENBERG, O.; MADURA, J. Do US stock market indexes over-or underreact?. **Journal of Financial Research**, v. 24, n. 2, p. 179-204, 2001.

SHARPE, W. Capital asset prices: A theory of market equilibrium under conditions of risk. **The journal of finance**, v. 19, n. 3, p. 425-442, 1964.

SHEN, D.; LIU, L.; ZHANG, Y. Quantifying the cross-sectional relationship between online sentiment and the skewness of stock returns. **Physica A**, v. 490, p. 928-934, 2018.

SIKANEN, M.; BALTAKYS, K.; KANNIAINEN, J.; VATRAPU, R.; MUKKAMALA, R.; HUSSAIN, A. Facebook drives behavior of passive households in stock markets. **Finance Research Letters**, v. 27, p. 208-213, 2018.

SILVA, M. **O efeito do sentimento das notícias sobre o comportamento dos preços no mercado acionário brasileiro**. 2017. Tese (Doutorado em Ciências Contábeis). 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. João Pessoa, 2017.

SIMON, H. A. **Models of man; social and rational**. 1957.

SIMON, H.A. **Reason in Human Affairs**. Stanford: Stanford University Press, 1983.

STAMBAUGH, R.; YU, J.; YUAN, Y. The long of it: Odds that investor sentiment spuriously predicts anomaly returns. **Journal of Financial Economics**, v. 114, n. 3, p. 613-619, 2014.

TVERSKY, A.; KAHNEMAN, D. "Judgment under Uncertainty: Heuristics and Biases," **Science**, v. 185, p. 1124-1131, 1974.

TWITTER. **Twitter reports last quarter 2018 results**. Disponível em: <http://d18rn0p25nwr6d.cloudfront.net/CIK-0001418091/b353de9c-2664-4e81-8acb-52e9031e4836.pdf>. Acesso em: 07 maio 2015.

UYGUR, U.; TAŞ, O. The impacts of investor sentiment on different economic sectors: evidence from Istanbul stock exchange. **Borsa Istanbul Review**, v. 14, n. 4, p. 236-241, 2014.

VERMA, R.; VERMA, P. Are survey forecasts of individual and institutional investor sentiments rational?. **International Review of Financial Analysis**, v. 17, n. 5, p. 1139-1155, 2008.

WEI, W.; MAO, Y.; WANG, B. Twitter volume spikes and stock options pricing. **Computer Communications**, n. 73, p. 271-281, 2016.

SHENG, X.; BRZESZCZYŃSKI, J.; IBRAHIM, B. International stock return co-movements and trading activity. **Finance Research Letters**, n. 23, p. 12-18, 2017.

SHILLER, R. From efficient markets theory to behavioral finance. **Journal of economic perspectives**, v. 17, n. 1, p. 83-104, 2003.

SLOVIC, P. Psychological study of human judgment: Implications for investment decision making. **The Journal of Finance**, v. 27, n. 4, p. 779-799, 1972.

**APÊNDICE A- Tabelas da Associação entre o Sentimento Dos *Tweets* e o Retorno do IBOVESPA**

**Tabela 2** - Teste da HME utilizando *tweets* enquanto informação – IBOV (Continuação)

<b>D<sub>-2</sub></b>					
<b>Sent<sub>d-2</sub></b>	0,0635*	0,0224	0,0227*	-0,0066	0,0015
	(0,0348)	(0,0210)	(0,0128)	(0,0223)	(0,0294)
<b>Ret<sub>d-2</sub></b>	0,0242	-0,0062	-0,1025	-0,1184**	-0,1018
	(0,0962)	(0,0545)	(0,0499)	(0,0556)	(0,0630)
<b>D<sub>-3</sub></b>					
<b>Sent<sub>d-3</sub></b>	0,0191	0,0030	0,0171	0,0135	0,0403**
	(0,0239)	(0,0228)	(0,0241)	(0,0115)	(0,0198)
<b>Ret<sub>d-3</sub></b>	0,0720	0,0169	-0,0096	-0,0676	0,0337
	(0,0925)	(0,0757)	(0,0628)	(0,0437)	(0,0555)
<b>D<sub>-4</sub></b>					
<b>Sent<sub>d-4</sub></b>	0,0715*	0,0194	-0,0014	-0,0189	-0,0474*
	(0,0401)	(0,0188)	(0,0138)	(0,0174)	(0,0267)
<b>Ret<sub>d-4</sub></b>	0,0177	0,0219	-0,0442	-0,0707	-0,0142
	(0,1482)	(0,0543)	(0,0608)	(0,0621)	(0,0671)
<b>D<sub>-5</sub></b>					
<b>Sent<sub>d-5</sub></b>	0,0006	-0,0067	0,0043	0,0064	0,0171
	(0,0307)	(0,0180)	(0,0162)	(0,0121)	(0,0215)
<b>Ret<sub>d-5</sub></b>	0,1227	0,1137***	0,0113	-0,0207	0,0143
	(0,1209)	(0,0417)	(0,0623)	(0,0871)	(0,0864)

Notas: Equação 4:  $Ret_t = \beta_0 + \beta_1 Sent_t + \gamma_1 \mathcal{L}_s(Ret_t) + \varepsilon_t$

A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

**Fonte:** Elaboração Própria (2020)

**Tabela 3** - Teste da HME controlado pelo volume de negócios– IBOV (Continuação)

<b>D<sub>-2</sub></b>					
<b>VN<sub>d-2</sub></b>	-3,76.10 <sup>-9</sup> ***	-2,41.10 <sup>-9</sup> **	-1,45.10 <sup>-9</sup>	-1,33.10 <sup>-9</sup>	2,76.10 <sup>-10</sup>
	(1,25. 10 <sup>-9</sup> )	(1,33.10 <sup>-9</sup> )	(1,21.10 <sup>-9</sup> )	(1,49.10 <sup>-9</sup> )	(2,21.10 <sup>-9</sup> )
<b>D<sub>-4</sub></b>					
<b>VN<sub>d-4</sub></b>	-2,96.10 <sup>-9</sup> **	-8,60.10 <sup>-10</sup>	-2,13.10 <sup>-10</sup>	-2,99.10 <sup>-10</sup>	4,05.10 <sup>-10</sup>
	(1,21.10 <sup>-9</sup> )	(7,01.10 <sup>-10</sup> )	(7,84.10 <sup>-10</sup> )	(9,70.10 <sup>-10</sup> )	(2,33.10 <sup>-9</sup> )
<b>D<sub>-5</sub></b>					
<b>Sent<sub>d-5</sub></b>	0,1715***	0,0326	0,0619	-0,0137	-0,0733
	(0,0501)	(0,0507)	(0,0611)	(0,0927)	(0,0816)
<b>Ret<sub>d-5</sub></b>	0,4261***	0,26588***	0,0641	-0,1073	-0,1895
	(0,1452)	(0,0974)	(0,1235)	(0,2086)	(0,2711)
<b>VN<sub>d-5</sub></b>	-1,79.10 <sup>-9</sup> ***	-2,06.10 <sup>-9</sup> ***	-1,10.10 <sup>-11</sup>	-1,10.10 <sup>-11</sup>	5,39.10 <sup>-10</sup>
	(6,77.10 <sup>-10</sup> )	(7,84.10 <sup>-10</sup> )	(9,71.10 <sup>-10</sup> )	(1,15.10 <sup>-9</sup> )	(1,62.10 <sup>-9</sup> )
<b>Sent x VN<sub>d-5</sub></b>	-5,60.10 <sup>-8</sup> ***	-1,52.10 <sup>-8</sup>	-2,11.10 <sup>-8</sup>	5,22. 10 <sup>-9</sup>	2,31.10 <sup>-8</sup>
	(1,61.10 <sup>-8</sup> )	(1,59.10 <sup>-8</sup> )	(1,76.10 <sup>-8</sup> )	(3,02.10 <sup>-8</sup> )	(2,78.10 <sup>-8</sup> )
<b>Ret x VN<sub>d-5</sub></b>	-4,27.10 <sup>-8</sup>	-3,80.10 <sup>-8</sup> *	-4,48.10 <sup>-9</sup>	1,63.10 <sup>-8</sup>	2,84.10 <sup>-8</sup>
	(3,90.10 <sup>-8</sup> )	(2,16.10 <sup>-8</sup> )	(2,64.10 <sup>-8</sup> )	(4,65.10 <sup>-8</sup> )	(6,43.10 <sup>-8</sup> )

Notas: Equação 5:  $Ret_t = \beta_0 + \beta_1 Sent_t + \gamma_1 \mathcal{L}_s(Ret_t) + \beta_2 VN_t +$

$\beta_3 SentxVN_t + \gamma_2 \mathcal{L}_s(Ret_t) x VN_t + \varepsilon_t$

A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

**Fonte:** Elaboração Própria (2020)

Tabela 4 – Relação entre o sentimento e o retorno – IBOV (Continuação)

$D_{-2}$					
$Sent_{d-2}$	-0,0183 (0,0389)	-0,0068 (0,0179)	0,0030 (0,0070)	0,0082 (0,0166)	-0,0326 (0,0220)
$VDT_d$	$3,09 \cdot 10^{-7}$ ( $1,59 \cdot 10^{-6}$ )	$-5,70 \cdot 10^{-7}$ ( $1,97 \cdot 10^{-6}$ )	$7,23 \cdot 10^{-7}$ ( $1,61 \cdot 10^{-6}$ )	$1,65 \cdot 10^{-7}$ ( $3,31 \cdot 10^{-6}$ )	$6,53 \cdot 10^{-6**}$ ( $3,29 \cdot 10^{-6}$ )
$Riskfree_d$	43,0413* (24,4131)	33,7043** (16,4292)	11,4855 (15,6443)	-35,4575* (20,2579)	-60,5679*** (14,5414)
$HML_d$	0,8478*** (0,1980)	0,8198*** (0,1343)	0,6931*** (0,0722)	0,7541*** (0,0977)	0,7908*** (0,1340)
$MOM_d$	0,3358* (0,1980)	0,2661* (0,1601)	0,1505 (0,0722)	0,1152 (0,1341)	0,0457 (0,1088)
$Liq_d$	-0,4513*** (0,1542)	-0,6475*** (0,1047)	-0,7662*** (0,0529)	-0,6863*** (0,0841)	-0,7768*** (0,1058)
$Infla_d$	-0,0143*** (0,0059)	-0,0067 (0,0042)	-0,0004 (0,0040)	0,0095** (0,0043)	0,0121*** (0,0031)
$PIB_d$	-0,0009 (0,0012)	-0,0011 (0,0009)	-0,0006 (0,0009)	-0,0014 (0,0009)	-0,0029* (0,0017)
$D_{-3}$					
$Sent_{d-3}$	-0,0071 (0,0224)	-0,0149 (0,0153)	-0,0073 (0,0129)	0,0158 (0,0145)	0,0208 (0,0255)
$VDT_d$	$2,03 \cdot 10^{-7}$ ( $2,45 \cdot 10^{-6}$ )	$3,87 \cdot 10^{-7}$ ( $2,18 \cdot 10^{-6}$ )	$7,38 \cdot 10^{-7}$ ( $2,34 \cdot 10^{-6}$ )	$2,39 \cdot 10^{-7}$ ( $3,44 \cdot 10^{-6}$ )	$9,10 \cdot 10^{-6**}$ ( $4,75 \cdot 10^{-6}$ )
$Riskfree_d$	47,6101 (34,4515)	28,1758 (21,1950)	10,1117 (14,0849)	-31,2278 (29,8849)	-64,4963*** (19,3817)
$HML_d$	0,8938*** (0,1593)	0,7930*** (0,0890)	0,7040*** (0,0802)	0,7468*** (0,1231)	0,8174*** (0,1689)
$MOM_d$	0,3383* (0,1593)	0,2776 (0,1840)	0,1633* (0,0845)	0,1095 (0,1078)	0,0545 (0,1030)
$Liq_d$	-0,4671*** (0,1061)	-0,6906*** (0,0834)	-0,7626*** (0,0677)	-0,6610*** (0,0926)	-0,7394*** (0,1162)
$Infla_d$	-0,0130 (0,0090)	-0,0054 (0,0050)	-0,0002 (0,0039)	0,0090 (0,0059)	0,0123** (0,0051)
$PIB_d$	-0,0007 (0,0011)	-0,0011 (0,0008)	-0,0008 (0,0009)	-0,0011 (0,0015)	-0,0020 (0,0015)
$D_{-4}$					
$Sent_{d-4}$	-0,0217 (0,0358)	-0,0057 (0,0304)	-0,0074 (0,0134)	-0,0327** (0,0129)	-0,0239 (0,0185)
$VDT_d$	$4,20 \cdot 10^{-8}$ ( $2,01 \cdot 10^{-6}$ )	$-6,46 \cdot 10^{-7}$ ( $2,00 \cdot 10^{-6}$ )	$7,84 \cdot 10^{-7}$ ( $2,09 \cdot 10^{-6}$ )	$1,48 \cdot 10^{-7}$ ( $3,50 \cdot 10^{-6}$ )	$8,55 \cdot 10^{-6*}$ ( $3,98 \cdot 10^{-6}$ )
$Riskfree_d$	44,7338** (22,8012)	34,0394** (16,8927)	3,7658 (18,1940)	-42,8483* (25,4857)	-52,9551** (20,5613)
$HML_d$	0,8288*** (0,1512)	0,8349*** (0,0996)	0,7195 (0,0795)	0,7530*** (0,1242)	0,8665*** (0,1185)
$MOM_d$	0,3268 (0,2000)	0,3268** (0,1427)	0,1729 (0,1108)	0,1089 (0,1538)	0,0198 (0,1365)
$Liq_d$	-0,5210*** (0,1832)	-0,6597*** (0,1205)	-0,7531*** (0,0935)	-0,6364*** (0,1003)	-0,7224*** (0,1051)
$Infla_d$	-0,0149** (0,0057)	-0,0069 (0,0050)	0,0002 (0,0041)	0,0089 (0,0057)	0,0094* (0,0055)
$PIB_d$	-0,0013 (0,0015)	-0,0013** (0,0006)	-0,0012 (0,0010)	-0,0023** (0,0011)	-0,0022* (0,0012)
$D_{-5}$					
$Sent_{d-5}$	-0,0235 (0,0216)	-0,0096 (0,0177)	0,0004 (0,0143)	-0,0164 (0,0151)	-0,0040 (0,0173)
$VDT_d$	$1,71 \cdot 10^{-7}$	$-5,78 \cdot 10^{-7}$	$7,81 \cdot 10^{-7}$	$1,47 \cdot 10^{-7}$	$8,41 \cdot 10^{-6**}$

	(1,11.10 <sup>-6</sup> )	(2,24.10 <sup>-6</sup> )	(2,69.10 <sup>-6</sup> )	(3,73.10 <sup>-6</sup> )	(5,49.10 <sup>-6</sup> )
<b>Riskfree<sub>d</sub></b>	48,8094	33,4450*	8,2729	-33,4504	-58,4077***
	(33,4178)	(19,2066)	(20,4073)	(24,2053)	(17,1358)
<b>HML<sub>d</sub></b>	0,8285***	0,8473***	0,7010***	0,7989***	0,8561***
	(0,0841)	(0,1272)	(0,0912)	(0,1104)	(0,1547)
<b>MOM<sub>d</sub></b>	0,3304**	0,2331*	0,1631	0,0817	0,0438
	(0,1505)	(0,1364)	(0,1032)	(0,1332)	(0,1075)
<b>Liq<sub>d</sub></b>	-0,4367***	-0,6657***	-0,7541***	-0,6383***	-0,7561***
	(0,1554)	(0,1295)	(0,1345)	(0,1397)	(0,1283)
<b>Infla<sub>d</sub></b>	-0,0153**	-0,0056	-0,0001	-0,0094*	0,0112**
	(0,0075)	(0,0042)	(0,0047)	(0,0053)	(0,0049)
<b>PIB<sub>d</sub></b>	-0,0011	-0,0014*	-0,0009	-0,0014	-0,0019
	(0,0016)	(0,0008)	(0,0010)	(0,0013)	(0,0020)

Notas: Equação 6:  $Ret_t = \beta_0 + \beta_1 Sent_t + \beta_2 VDT_t + \beta_3 RF_t + \beta_4 HML_t + \beta_5 MOM_t + \beta_6 Liq_t + \beta_7 Infl_t + \beta_8 PIB_t + \varepsilon_t$

A primeira linha traz os coeficientes e a estatística p-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

**Fonte:** Elaboração Própria (2020)

**APÊNDICE B- Tabelas da Associação entre o Volume De Tweets e o Volume Negociado**

**Tabela 5** - Associação entre o volume negociado e o volume de tweets (Continuação)

<i>D</i> <sub>-2</sub>					
<i>Sent</i> <sub><i>d</i>-2</sub>	-735842 (1269088)	-1614550* (867001)	-959817 (1213213)	-1434241 (1413107)	-7159330* (3779801)
<i>VDT</i> <sub><i>d</i>-2</sub>	-6,4800 (196,911)	-42,2228 (150,8997)	-4,0378 (158,1749)	-44,2065 (81,0947)	-477,2957** (257,5766)
<i>Vol</i> <sub><i>d</i>-2</sub>	0,2337*** (0,0574)	0,2963*** (0,0361)	0,3335*** (0,0539)	0,4736*** (0,0592)	0,5610*** (0,0982)
<i>D</i> <sub>-3</sub>					
<i>Sent</i> <sub><i>d</i>-3</sub>	1819355 (5112795)	506713 (1158304)	-185660 (4081478)	809181 (4247968)	-187098 (4473947)
<i>VDT</i> <sub><i>d</i>-3</sub>	70,3131 (1071,833)	-94,1820 (179,5137)	1,0775 (197,0636)	-91,2646 (218,9271)	-315,3132 (216,1127)
<i>Vol</i> <sub><i>d</i>-3</sub>	0,2981*** (0,2401)	0,2613*** (0,0340)	0,3156*** (0,0488)	0,4388*** (0,0659)	0,7340*** (0,1401)
<i>D</i> <sub>-4</sub>					
<i>Sent</i> <sub><i>d</i>-4</sub>	-151798 (1292008)	567442 (1197182)	-873136 (967672)	-2328215 (2854203)	-4426665 (3054792)
<i>VDT</i> <sub><i>d</i>-4</sub>	-104,6936 (187,0162)	-124,3154 (188,637)	-225,3841** (112,6008)	-261,3857 (276,6482)	-541,1465 (513,8758)
<i>Vol</i> <sub><i>d</i>-4</sub>	0,2630*** (0,0391)	0,2977*** (0,0415)	0,3537*** (0,0240)	0,4151*** (0,0899)	0,6072*** (0,1379)
<i>D</i> <sub>-5</sub>					
<i>Sent</i> <sub><i>d</i>-5</sub>	-469451 (826777)	-1424083 (971814)	-1687813 (6106168)	-2840694 (8177978)	-1888089 (3778152)
<i>VDT</i> <sub><i>d</i>-5</sub>	-233,0119 (160,8544)	-397,4199*** (111,9268)	-360,3521 (238,9109)	-443,3836 (476,559)	-566,1729** (229,5984)
<i>Vol</i> <sub><i>d</i>-5</sub>	0,2117*** (0,0499)	0,2874*** (0,0474)	0,3794*** (0,0494)	0,4387 (0,3030)	0,7236*** (0,0377)

Notas: Equação 7:  $VN_t = \beta_0 + \beta_1 Sent_t + \beta_2 VDT_t + \gamma_1 L_s(VN_t) + \varepsilon_t$

A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

**Fonte:** Elaboração Própria (2020)

Tabela 6 - Associação entre o volume negociado e os tipos de sentimento (Continuação)

<b>D<sub>-2</sub></b>					
<b>Sent<sub>d-2</sub></b>	56884,25 (1402149)	-594389 (1494372)	-634972 (1194758)	-1151133 (1,00.10 <sup>+7</sup> )	-1894480 (1,24.10 <sup>+7</sup> )
<b>VTP<sub>d-2</sub></b>	-690,4037 (702,328)	-513,1944 (499,8424)	-82,28989 (706,2039)	-257,4085 (1517,621)	-1874,344** (7861,94)
<b>VTN<sub>d-2</sub></b>	641,0169* (462,1879)	319,5961 (448,4488)	228,1378 (501,2522)	33,61192 (893,5159)	218,455 (5629,652)
<b>Vol<sub>d-2</sub></b>	0,2464*** (0,0796)	0,3066*** (0,0504)	0,3513*** (0,0563)	0,4809*** (0,0487)	0,5912*** (0,2926)
<b>D<sub>-3</sub></b>					
<b>Sent<sub>d-3</sub></b>	3114658 (2447782)	2342393 (1959189)	1345721 (1,72.10 <sup>+7</sup> )	1711663 (9629631)	1885268 (4480674)
<b>VTP<sub>d-3</sub></b>	-496,1669 (832,6587)	-1138,294 (826,4602)	-888,3375* (2353,72)	-1498,199** (864,8611)	-2693,676** (1067,419)
<b>VTN<sub>d-3</sub></b>	758,6855 (896,4747)	767,7264** (618,1122)	429,7631 (2607,6250)	430,4617 (1471,8160)	590,5022 (2455,7700)
<b>Vol<sub>d-3</sub></b>	0,2922*** (0,0698)	0,2681*** (0,0498)	0,3289*** (0,0548)	0,4377*** (0,1252)	0,6874*** (0,1592)
<b>D<sub>-4</sub></b>					
<b>Sent<sub>d-4</sub></b>	90437 (2637417)	-1416803 (6304715)	389427 (1,94.10 <sup>+7</sup> )	-3082225 (3372716)	-7503150 (8953238)
<b>VTP<sub>d-4</sub></b>	68,1395 (573,1010)	-284,7384 (3727,3700)	-1312,158*** (1663,8400)	-844,2344 (1080,9620)	321,9207 (2381,6650)
<b>VTN<sub>d-4</sub></b>	-43,31399 (1197,0040)	-213,9444 (4824,8130)	153,6309 (2253,9640)	-276,4457 (3197,9720)	-1225,99 (6252,9540)
<b>Vol<sub>d-4</sub></b>	0,2684 *** (0,0548)	0,2857*** (0,0468)	0,3626*** (0,0746)	0,4176*** (0,0966)	0,6051*** (0,1160)
<b>D<sub>-5</sub></b>					
<b>Sent<sub>d-5</sub></b>	-471026 (1,23.10 <sup>+7</sup> )	-84881 (1320638)	-1437952 (1,29.10 <sup>+7</sup> )	-542605,1 (2967534)	-22622,05 (1,45.10 <sup>+7</sup> )
<b>VTP<sub>d-5</sub></b>	-227,2302 (2272,5460)	-566,7637 (843,5531)	-762,033 (1932,4580)	-956,1496 (601,1182)	-2017,801** (3141,8050)
<b>VTN<sub>d-5</sub></b>	-183,3852 (1398,0550)	-392,6263 (539,5218)	-369,023 (2096,5680)	-249,1003 (750,6068)	-174,9469 (2301,8450)
<b>Vol<sub>d-5</sub></b>	0,1948*** (0,5258)	0,2981*** (0,0464)	0,3715 (0,1085)	0,5177*** (0,1171)	0,7375*** (0,1311)

Notas: Equação 8:  $VN_t = \beta_0 + \beta_1 Sent_t + \beta_2 VTP_t + \beta_3 VTN_t \gamma_1 + \mathcal{L}_s(VN_t) + \varepsilon_t$

A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.

## APÊNDICE C - Associação entre o sentimento e o retorno do SMLL

**Tabela 8** - Relação entre o sentimento e o retorno – SMLL (Continuação)

$D_{-2}$					
<b>Infla<sub>d</sub></b>	-0,0145*	-0,0039	0,0011	0,0086	0,0084
	(0,0077)	(0,0062)	(0,0043)	(0,0041)	(0,0079)
<b>VTT<sub>d</sub></b>	$-8,20e.10^{-6}$	$5,76.10^{-8}$	$4,41.10^{-7}$	$3,66.10^{-7}$	$0,0001^{***}$
	$(6,77.10^{-6})$	$(4,53.10^{-6})$	$(1,89.10^{-6})$	$(2,58.10^{-6})$	$(5,65.10^{-6})$
$D_{-3}$					
<b>Sent<sub>d-3</sub></b>	-0,0431	-0,0369**	0,0063	0,0189	-0,0014
	(0,0311)	(0,0212)	(0,0146)	(0,0187)	(0,0306)
<b>VTT<sub>d</sub></b>	$-8,54.10^{-6***}$	$-7,30.10^{-7}$	$4,48.10^{-7}$	$2,32.10^{-6}$	$9,97.10^{-6**}$
	$(3,74.10^{-6})$	$(4,28.10^{-6})$	$(2,19.10^{-6})$	$(3,34.10^{-6})$	$(6,91.10^{-6})$
$D_{-4}$					
<b>Riskfree<sub>d</sub></b>	75,4794*	5,3414	1,7893	-48,8284**	-43,1222
	(50,5097)	(34,5474)	(21,5043)	(19,2177)	(36,8463)
<b>Infla<sub>d</sub></b>	-0,0142*	-0,0046	0,0014	0,0105	0,0069
	(0,0092)	(0,0055)	(0,0055)	(0,0043)	(0,0070)
<b>PIB<sub>d</sub></b>	0,0012	-0,0017	-0,0004	-0,0018*	-0,0014
	(0,0019)	(0,0015)	(0,0011)	(0,0010)	(0,0019)
$D_{-5}$					
<b>Sent<sub>d-5</sub></b>	-0,0360*	-0,0194	-0,0027	-0,0013	0,0080
	(0,0245)	(0,0142)	(0,0132)	(0,0188)	(0,0374)
<b>Infla<sub>d</sub></b>	-0,0154*	-0,0051	0,0011	0,0087	0,0086
	(0,0096)	(0,0066)	(0,0059)	(0,0048)	(0,0097)
<b>PIB<sub>d</sub></b>	0,0001	-0,0025*	-0,0003	-0,0012	-0,0010
	(0,0018)	(0,0015)	(0,0010)	(0,0010)	(0,0015)
<b>VTT<sub>d</sub></b>	$-8,15.10^{-6}$	$-3,07.10^{-6}$	$4,07.10^{-7}$	$2,03.10^{-6}$	$9,45.10^{-6**}$
	$(4,53.10^{-6})$	$(4,04.10^{-6})$	$(1,32.10^{-6})$	$(1,97.10^{-6})$	$(5,16.10^{-6})$

Nota: Equação 9:  $Ret_t = \beta_0 + \beta_1 Sent_t + \beta_2 VTT_t + \beta_3 RF_t + \beta_4 Infl_t + \beta_5 PIB_t + \varepsilon_t$   
 A primeira linha traz os coeficientes e a estatística *p*-valor;

O erro padrão está entre parênteses.

\* Significante a 10%, \*\* a 5% e \*\*\* a 1%.



## APÊNDICE D – Coleta de Dados e Atribuição do Sentimento

Os códigos utilizados para adquirir os *tweets* que fizeram parte da amostra e para atribuir o sentimento aos textos se encontram disponíveis por meio do seguinte link: <https://github.com/dyliane/twitter-sentiment-stock-analysis>.