

Índice de sentimento textual: uma análise empírica do impacto das notícias sobre risco sistemático

Sentiment text index: empirical analysis the impact on systematic risk

Índice de sentimiento en texto: análisis empírico del impacto de las noticias en riesgo sistemático

Maria Daniella de Oliveira Pereira da Silva

Doutora em Ciências Contábeis pelo Programa Multiinstitucional e Inter-Regional de Pós-Graduação em Ciências Contábeis (UnB/UFPB/UFRN)
Professora da Universidade Federal da Paraíba (UFPB)
Endereço: Cidade Universitária – Campos I
CEP: 58051-900 – João Pessoa/PB – Brasil
E-mail: mariadaniella75@hotmail.com
Telefone: (81) 99279 7525

Márcio André Veras Machado

Doutor em Administração pela Universidade de Brasília (UnB)
Professor do Programa de Pós-Graduação em Administração e em Ciências Contábeis da Universidade Federal da Paraíba (UFPB)
Endereço: Cidade Universitária - Campus I, Castelo Branco
CEP: 58059-900 - João Pessoa/PB - Brasil
E-mail: mavmachado@hotmail.com
Telefone: (83) 3216-7492

Artigo recebido em 28/04/2018. Revisado por pares em 15/09/2019. Reformulado em 15/12/2019. Recomendado para publicação em 23/12/2019 por Carlos Eduardo Facin Lavarda (Editor-Chefe). Publicado em 31/12/2019.

Resumo

A proposta do estudo foi estimar um índice de sentimento textual baseado em notícias e investigar o seu comportamento sobre o risco sistemático. Para isso, foi investigado se o risco sistemático das ações preferenciais da Vale (VALE5) sofreu influência do tom e do volume das notícias atreladas ao acidente ocasionado pela Samarco, nos períodos em que os investidores estavam mais vulneráveis ao risco, partindo da premissa de que as notícias contribuem com a atualização das crenças dos investidores sobre suas expectativas futuras, principalmente em períodos de maior incerteza. As relações entre o risco sistemático, volume de notícias e sentimento textual (tom) foram obtidas mediante regressão quantílica. As evidências empíricas encontradas entre o 5º e 9º *decil* levam a constatação de que o volume e o tom das notícias veiculadas na mídia influenciam o beta da ação, quando existe uma maior exposição ao risco, sugerindo indícios de que o risco sistemático apresenta conexão com as divulgações de notícias pela mídia, nos períodos de maior incerteza sobre os fluxos de caixa futuro dos ativos.

Palavras-chave: Notícias; Risco sistemático; Sentimento textual

Abstract

The purpose of the study was to estimate textual sentiment index of news and investigate its behavior on systematic risk. For this, it was investigated if the systematic risk of stock price Vale (VALE5) was influenced by the tone and volume of the news related the accident caused by Samarco, in periods when investors were more vulnerable to risk, based on premise that news contributes to updating investors beliefs about their future expectations, especially in periods of more uncertainty. The relationships between systematic risk, news volume and tone news were obtained through quantile regression. The empirical evidence found between the 5th and 9th deciles leads to realization that volume and tone of the news influences the beta of Vale when there is more exposure to risk, suggesting evidence that systematic risk is related to the news releases by the media in periods of increased uncertainty about the future cash flows of assets.

Keywords: News; Systematic risk; News tone

Resumen

El propósito del estudio fue estimar un índice de sentimiento textual basado en noticias e investigar su comportamiento ante el riesgo sistemático. Para esto, se investigó si el riesgo sistemático de las acciones preferenciales de Vale (VALE5) sufrió influencia del tono y del volumen de las noticias relacionadas con el accidente ocasionado por Samarco, en los períodos que los inversores eran más vulnerables al riesgo, partiendo de la premisa que las noticias contribuyen con la actualización de las creencias de los inversores sobre sus expectativas futuras, principalmente en períodos de mayor incertidumbre. Las relaciones entre el riesgo sistemático, volumen de noticias y sentimiento textual (tono) se obtuvieron mediante regresión cuantílica. Las evidencias empíricas encontradas entre el 5º y 9º decil llevan la constatación de que el volumen y el tono de las noticias influyen en la beta de la acción cuando existe una mayor exposición al riesgo, sugiriendo indicios de que el riesgo sistemático presenta conexión con la divulgación de noticias, en los períodos de mayor incertidumbre sobre los flujos de caja futura de los activos.

Palabras clave: Noticias; Riesgo sistemático; Sentimiento textual

1 Introdução

As discussões sobre a influência das questões socioambientais na performance financeira das organizações ganharam notoriedade a partir da década de 1960, diante da profusão de problemas sociais e ambientais oriundos do sistema de produção capitalista (COASE, 1960; FRIEDMAN, 1970; DAVIS, 1973; CARROLL, 1979). Problemas ambientais costumam causar

efeitos negativos sobre o desempenho financeiro das organizações, em função das incertezas geradas. Como por exemplo, o vazamento de material químico na Virgínia Ocidental, nos Estados Unidos, em 2014, que levou a empresa responsável pelo acidente, *Freedom Industries*, a pedir falência (LARSON, 2014) e o escândalo da Volkswagen, em 2015, com o anúncio de fraudes relacionada à emissão de poluentes, que provocou a queda das ações em mais de 20%, no dia seguinte à divulgação do escândalo (21 de setembro de 2015) (GRIFFIN; LONT, 2016).

Notícias relevantes, como eventos inesperados e catastróficos, afetam o mercado financeiro, por provocarem alterações nas expectativas dos fluxos de caixa futuro, considerando que o mercado é eficiente e que os preços refletem rapidamente as informações disponíveis, em função do ajuste das expectativas dos investidores quanto ao risco e retorno dos títulos (FAMA, 1970).

O risco de um ativo pode ser classificado em risco específico e risco sistemático. O risco específico está relacionado aos aspectos intrínsecos à organização, sendo possível eliminá-lo com estratégias de diversificação de investimento. Já o risco sistemático, também chamado de risco de mercado, corresponde à oscilação do ativo em relação à carteira de mercado, geralmente, modelado pelo beta dos modelos de precificação de ativos, como o beta estimado pelo *Capital Asset Pricing Model* (CAPM), desenvolvido por Sharpe (1964), Lintner (1965) e Mossin (1966).

O beta é uma importante medida de sensibilidade dos retornos dos ativos, utilizada para estimação e controle do risco sistemático (CAMPBELL; MEI, 1993). Devido à sua importância para o contexto financeiro, surgiram pesquisas, com o intuito de identificar elementos determinantes do risco sistemático (BEAVER; KETTLER; SCHOLLES, 1970; CAMPBELL; SHILLER, 1988; CAMPBELL; MEI, 1993; FOSTER; KASZNIK; SIDHU, 2012; WANG; LI; HUANG, 2013; MAR-MOLINERO; MENÉNDEZ-PLANS; ORGAZ-GUERRERO, 2017). Evidências empíricas mostram que o beta está vulnerável às notícias que apresentam implicações sobre os fluxos de caixa, taxas reais de juros e excesso de retorno (CAMPBELL; SHILLER, 1988; CAMPBELL; MEI, 1993).

Nas investigações sobre os fatores determinantes do risco sistemático, tradicionalmente, são utilizadas *proxies* de informações essencialmente quantitativas, como dados contábeis, econômicos e financeiros (CAMPBELL; MEI, 1993; AMORIM; LIMA; MURCIA, 2012; MAR-MOLINERO; MENÉNDEZ-PLANS; ORGAZ-GUERRERO, 2017). No entanto, vale salientar que os participantes do mercado também utilizam informações não numéricas para o alinhamento das suas expectativas sobre os ativos, como, por exemplo, informações textuais. Assim, é importante ressaltar que, para o mercado financeiro, os textos são *proxies* de notícias tão importantes quanto as *proxies* baseadas em medidas quantitativas (KEARNEY; LIU, 2014; HUANG; TEOH; ZHANG, 2014; LIU, 2014, ALLEN; MCALEER; SINGH, 2017).

O aperfeiçoamento da Linguagem de Processamento Natural (LPN) contribuiu com o desenvolvimento de instrumentos para análise dos conteúdos textuais, favorecendo investigações sobre as características linguísticas do processo de comunicação, como a análise do tom ou do sentimento presente nos textos. As investigações empíricas que buscam a análise do tom ou sentimento textual avaliam a semântica do processo de comunicação, por meio da extração de informações relacionadas ao posicionamento de otimismo ou pessimismo (BEATTIE, 2014; KEARNEY; LIU, 2014; HUANG; TEOH; ZHANG, 2014). Assim, a mensuração do sentimento textual está diretamente atrelada à análise do conteúdo linguístico dos textos, no qual são analisadas as palavras empregadas no processo de comunicação, visando capturar o posicionamento de otimismo ou pessimismo presente no texto.

Os estudos que analisaram o reflexo do sentimento textual de jornais, relatórios, redes sociais e *blogs* no mercado acionário apontam evidências de efeitos do sentimento textual sobre os preços das ações e sobre os volumes de negociações (TETLOCK, 2007; TETLOCK; SAAR-TSECHANSKY; MACSKASSY, 2008; KOTHARI; LI; SHORT, 2009; LI, 2010; LOUGHRAN; MCDONALD, 2011; GRIFFIN; HIRSCHHEY; KELLY, 2011; HUANG; TEOH; ZHANG, 2014;

HENDERSHOTT; LIVDAN; SCHÜRHOFF, 2015; FERGUSON; PHILIP; LAM; GUO, 2015; AHERN; SOSYURA, 2015; BIANCONI; HUA; TAN, 2015; STRAUß; VLIAGENTHART; VERHOEVEN, 2016; ALLEN; MCALEER; SINGH, 2017). No entanto, existe uma lacuna na literatura quanto à compreensão da influência do tom da notícia sobre o risco sistemático, visto que os estudos que procuraram investigar o impacto da informação sobre o risco sistemático, em geral, utilizaram como *proxies* de notícias informações provenientes de dados econômicos e financeiros (CAMPBELL; MEI, 1993; AMORIM; MAR-MOLINERO; MENÉNDEZ-PLANS; ORGAZ-GUERRERO, 2017), não sendo encontrados estudos que utilizaram índices de sentimento textual como *proxy* para notícia.

Admitindo que o risco sistemático é extremamente sensível às notícias, de um modo geral, e que, geralmente, eventos inesperados e catastróficos costumam afetar o comportamento do mercado financeiro, por provocarem alterações nas expectativas dos fluxos de caixa futuro, a presente pesquisa procurou investigar o efeito da surpresa da notícia de um evento inesperado no risco sistemático de um ativo. Para isso, a pesquisa restringiu-se a investigar a empresa Vale, por a mesma está envolvida no acidente da Samarco, considerado como a maior tragédia ambiental do Brasil e uma das maiores tragédias do mundo provocada por vazamento de minério, que ocorreu em 2015, com o rompimento de barragens na cidade de Mariana (BBC, 2019).

Nesse sentido, a presente pesquisa buscou examinar o impacto de notícias sobre um evento inesperado no risco sistemático, mediante a análise das notícias veiculadas em mídia financeira que retratavam sobre o acidente ambiental provocado pela empresa Samarco. O acidente ocorreu no dia 5 de novembro de 2015, com o rompimento de duas barragens de rejeitos de minério, que pertenciam ao complexo de mineração da empresa Samarco. O rompimento das barragens provocou inundações e espalhou dejetos nos rios, por mais de 400 quilômetros, sendo considerado um dos maiores desastres ambientais provocados por mineradoras no Brasil (BILLER; WILLIS, 2015). O acidente da Samarco também desencadeou queda no preço das ações da Vale e da BHP Billiton, em função de ser uma empresa de capital fechado gerida mediante uma *joint-venture* entre as empresas Vale S.A. e a BHP Billiton.

Diante da magnitude do acidente provocado pela Samarco, pressupõem-se a existência de uma maior incerteza sobre os fluxos de caixa futuro dos ativos pertencentes às empresas envolvidas no acidente. Assim, levando em consideração que as notícias textuais contribuem com a atualização das crenças dos investidores sobre as suas expectativas futuras, principalmente nos momentos de maior instabilidade econômica (GARCÍA, 2013), e admitindo que a chegada de informações relevantes pode exercer efeitos sobre o preço dos títulos, alterando as expectativas sobre o risco (BEAVER; KETTLER; SCHOLLES, 1970), o estudo se propôs a analisar se o risco sistemático das ações preferenciais da Vale (VALE5) sofreu influência do tom e do volume de notícias atreladas ao acidente ocasionado pela Samarco.

Desse modo, foi examinado, no estudo, a sensibilidade do risco sistemático em relação à divulgação de notícias, partindo do pressuposto de que os investidores estão mais vulneráveis ao conteúdo e ao volume das notícias provenientes de jornais em períodos que remetem a uma maior incerteza sobre os fluxos de caixa futuro. Para isso, procurou-se analisar o volume e o sentimento textual das notícias *online* divulgadas no Jornal Valor Econômico que faziam referência à empresa Samarco, relacionando o volume e o sentimento textual ao risco sistemático das ações preferenciais da Vale (VALE5), o qual foi modelado pelo beta do CAPM condicional, admitindo que a variância e a covariância oscilam no tempo, conforme a proposta de González-Rivera (2013).

O beta é uma medida sensível que capta o risco sistemático de um ativo mediante as oscilações do mercado. Assim, o ativo apresentará mais risco quando o beta assumir valor superior a 1 (um), sinalizando que o retorno esperado oscila mais do que a carteira teórica do mercado, podendo levar a uma valorização/desvalorização superior ao mercado. Por isso, ativos com betas superiores a 1 são considerados como ativos com elevado nível de risco. Dessa forma, pode-se dizer que os investidores estão mais expostos ao risco nos momentos em que os betas estão

elevados. Diante disso, o estudo examinou a relação entre a divulgação de notícias e o comportamento da distribuição condicional do beta, por meio de regressão quantílica, explorando os resultados dos maiores *decis*, por eles sintetizarem o comportamento dos períodos com maior exposição ao risco para os investidores.

A realização do estudo foi motivada pelo anseio de conhecer o poder informacional da mídia brasileira na disseminação de notícias no mercado e o seu reflexo sobre o risco sistemático. A escolha por estudar o comportamento acionário da Vale S.A. no período posterior ao desastre envolvendo a Samarco, deu-se em função da repercussão que o caso ganhou na mídia nacional e internacional, em decorrência dos problemas sociais, ambientais e financeiros provocados com o rompimento das barragens sob responsabilidade da mineradora Samarco.

O estudo contribui com literatura sobre sentimento textual, ao propor uma lista de palavras que remetem a um contexto de otimismo e pessimismo, visando a estimação do tom de textos com ênfase financeira e escritos na língua portuguesa, uma vez que a lista de termos mais utilizada nos estudos sobre sentimento textual é a lista de Loughran e McDonald (2011), que se encontra na língua inglesa. Nesse sentido, a estruturação da lista de termos em português fornece respaldo para a discussão sobre o papel da mídia financeira no mercado acionário, respeitando as características linguísticas do Brasil, pois existe um grande interesse na literatura por investigações sobre sentimento textual que levem em consideração a língua oficial dos países (TETLOCK, 2014; KEARNEY; LIU, 2014; LOUGHRAN; MCDONALD, 2016).

2 Evidências Empíricas e Hipótese de Pesquisa

No âmbito das investigações sobre o sentimento textual, uma das primeiras pesquisas a discorrer sobre a interação entre o tom das notícias e os movimentos do mercado acionário, utilizando análise de conteúdo automatizada, foi o estudo de Tetlock (2007), que analisou a coluna financeira *Abreast of the Market* presente no *Wall Street Journal* (WSJ), entre os anos de 1984 a 1999, com o intuito de verificar se o pessimismo da mídia poderia prever os movimentos no mercado. O *General Inquirer's Harvard IV-4 psychosocial* foi o dicionário adotado para classificação e mensuração da frequência das palavras, servindo de base para o levantamento do fator de pessimismo. Os resultados apontaram a existência de uma correlação entre o conteúdo da coluna diária do WSJ com as atividades do mercado, concluindo que níveis elevados de pessimismo na mídia preveem uma pressão sobre os preços de mercado, os quais reagem com quedas, que são revertidas rapidamente. Com isso, sugere que o sentimento de pessimismo da mídia poderia ser uma *proxy* para o sentimento dos investidores ou uma *proxy* para a aversão ao risco.

Partindo das evidências encontradas por Tetlock (2007), García (2013) buscou analisar o efeito do sentimento textual das notícias com a suposição de que os investidores seriam mais sensíveis às notícias dos jornais em períodos de recessão. Desse modo, investigou o poder informacional do sentimento das notícias no mercado acionário, ao longo dos ciclos econômicos, utilizando como *proxy* para notícia o sentimento textual de duas colunas financeiras do *New York Times*, adotando o dicionário de Loughran e McDonald (2011) para a categorização das palavras (positivas/negativas), e considerando que o sentimento das notícias seria uma relação entre a frequência de palavras positivas e negativas. Os resultados apontaram evidências de que, nos períodos de recessão, as notícias teriam um impacto maior sobre os retornos no índice de mercado analisado. Além disso, foram encontrados padrões de previsibilidade no retorno dos ativos financeiros atrelados ao conteúdo das notícias, principalmente nos períodos de recessão.

No Brasil, Faustino (2013) analisou o sentimento de notícias divulgadas em língua inglesa que faziam referência ao Brasil, confrontado com os movimentos acionários do mercado brasileiro, entre os anos de 2010 a 2012. Para o levantamento do tom das notícias, utilizou a frequência das palavras encontradas no texto, adotando o dicionário de Loughran e McDonald (2011) para a

categorização das mesmas. Os resultados apresentaram indícios de que um o aumento de notícias negativas estava atrelado a um aumento da volatilidade e do volume de negociações no mercado. No entanto, o tom das notícias não apresentou efeitos significativos sobre o retorno.

Ao analisar o sentimento textual de notícias específicas sobre empresas, Ferguson et al., (2015) procuraram examinar o poder informacional da combinação do tom (positivo/negativo) e do volume de notícias financeiras de empresas do Reino Unido sobre a capacidade de previsão dos retornos futuros das ações, restringindo-se à análise das notícias publicadas nos jornais *The Financial Times*, *The Times*, *The Guardian* e *Mirror*, no período de 1981 a 2010. No estudo, o tom das notícias foi obtido mediante a frequência das palavras positivas/negativas classificadas segundo o dicionário de Loughran e McDonald (2011), e as empresas foram segregadas em dois grupos, conforme o tamanho e valor do *book-to-market* (BM), visando a identificação das empresas com maior/menor visibilidade. Foram encontrados indícios de que as notícias positivas prediziam retornos futuros maiores, enquanto as notícias negativas indicavam retornos menores, concluindo que o tom e o volume das notícias específicas sobre as empresas apresentam informações valiosas que contribuam na predição dos retornos futuros.

Uma das primeiras investigações relacionadas ao efeito da notícia sobre o risco das empresas foi realizado por Bianconi, Hua e Tan (2015), ao mensurar o efeito potencial que as notícias poderiam provocar no risco sistêmico das instituições financeiras. Assim, foi investigado se o pessimismo previa o risco, não em termos do retorno médio condicionado ao pessimismo do investidor, mas em relação ao efeito do pessimismo sobre as caudas de distribuição do risco, visto que nelas era possível mensurar, essencialmente, o risco e as formas de contágios às outras instituições financeiras. Para isso, foram utilizados dois tipos de informações: (a) as de ampla circulação, relacionadas às matérias divulgadas pelos jornais, que seriam uma *proxy* para o sentimento de pessimismo dos investidores; e (b) as informações sobre as expectativas futuras da volatilidade do mercado, que seria uma *proxy* para o índice de medo dos investidores. Desse modo, foi analisado o mercado americano, no período de 1992 a 2006, utilizando para a mensuração do risco o valor do risco condicional (CoVar) e o valor em risco (VaR). Para a estimação das previsões, foram utilizados modelos Autoregressivos de Heterocedasticidade Condicional (ARCH-GARCH). Os resultados mostraram indícios de que o sentimento de pessimismo dos investidores era uma boa medida para a previsão dos retornos médios, mas não foi uma medida capaz de prever o risco extremo, concluindo que o pessimismo dos jornais não conseguiu prever o risco extremo das instituições financeiras.

Diante do exposto, observa-se que a ênfase das investigações empíricas de Tetlock (2007) e García (2013) centraram-se na análise do retorno e do volume de negociação. Já Faustino (2013), ao analisar o mercado brasileiro, investigou o retorno, volatilidade e volume de negociação, utilizando notícias sobre o Brasil divulgadas em língua inglesa. Ferguson et al. (2015), por sua vez, analisaram especificamente o retorno dos ativos, enquanto Biancone et al. (2015) examinaram o risco sistêmico de instituições financeiras. Dessa forma, constata-se que os estudos empíricos não exploraram o risco sistemático dos ativos (TETLOCK, 2007; GARCÍA, 2013; FAUSTINO, 2013; FERGUSON et al., 2015; BIANCONI; HUA; TAN, 2015), e, nas evidências sobre o mercado brasileiro, observa-se que a modelagem do sentimento textual ocorreu por meio de dicionário baseado em língua inglesa, não respeitando as características linguísticas do Brasil (FAUSTINO, 2013).

Nesse contexto, a presente pesquisa contribui com a literatura, ao investigar a influência do conteúdo e do volume de notícias sobre o risco sistemático dos ativos, mediante a utilização de notícias diárias, divulgadas em jornais, na língua portuguesa. Os estudos empíricos, na sua maioria, exploram os efeitos do conteúdo das notícias sobre o retorno das ações, e, portanto, este estudo é um dos primeiros a fornecer evidências empíricas sobre o efeito do conteúdo das notícias sobre o risco sistemático dos ativos. Por fim, este estudo corrobora com evidências empíricas sobre o poder informacional do sentimento da notícia para o mercado brasileiro, ao modelar o sentimento textual

por meio de um dicionário em língua portuguesa.

Diante das evidências sobre o poder informacional do sentimento textual capaz de impactar o retorno e o volume das negociações no mercado (TETLOCK, 2007; GARCÍA, 2013; FAUSTINO, 2013; FERGUSON et al., 2015), mediante a compreensão de que o mercado é mais sensível às notícias dos jornais em momentos de maior incerteza econômica, como recessões (GARCÍA, 2013), eventos inesperados e catastróficos, que são períodos em que os investidores estão mais expostos ao risco, devido às incertezas sobre os fluxos de caixa futuro, e diante da necessidade de melhor compreensão do poder informacional das notícias sobre o risco (BIANCONI, HUA; TAN 2015), tem-se a seguinte hipótese de pesquisa:

H₁: O risco sistemático das ações da Vale (VALE5) sofre a influência do volume e do tom de notícias nos períodos em que os investidores estão mais expostos ao risco.

3 Metodologia

Para a realização do estudo, foram utilizados o preço de fechamento diário das ações preferenciais da Vale S.A – VALE5, o índice diário do Ibovespa, como *proxy* para a carteira que melhor representa o mercado, e as notícias diárias sobre a mineradora Samarco. Os dados diários sobre o preço de fechamento das ações e sobre o índice Ibovespa foram coletados no *Thomson Reuters*, no período de 04 de novembro de 2015 a 16 de fevereiro de 2016, desconsiderando feriados e finais de semana.

Em relação às notícias, foram utilizadas as matérias publicadas *online* pelo Valor Econômico, que faziam referência à Samarco, sem fazer restrições quanto à coluna de publicação da notícia. Procurou-se analisar o comportamento diário do mercado ao longo dos três meses seguintes ao desastre envolvendo a Samarco, visto que não seria possível analisar um maior período, em decorrência do impacto que notícias de outras naturezas poderiam provocar sobre o comportamento do risco sistemático. Assim, o estudo delimitou-se à análise das notícias publicadas entre 05 de novembro de 2015 a 16 de fevereiro de 2016, totalizando 326 notícias.

A escolha do Jornal Valor Econômico foi dada mediante a importância que o mesmo apresenta em termos de conteúdo e propagação de informações de cunho econômico e político sobre o Brasil, e, por ter ênfase na apresentação das tendências e das principais movimentações do mercado financeiro brasileiro e internacional (VALOR, 2017). Desse modo, acredita-se que as matérias financeiras do jornal Valor Econômico sintetizam os acontecimentos relevantes para o mercado brasileiro. Nesse sentido, o estudo tratou as matérias do valor econômico como uma *proxy* para as notícias do mercado, considerando que, de algum modo, elas interferem no comportamento dos investidores, e isso é refletido no mercado.

A análise de conteúdo dos textos foi realizada mediante a aprendizagem baseada em dicionário, que corresponde a uma técnica para a extração de informações textuais, mediante a utilização de listas de palavras previamente definidas. Assim, para a extração do sentimento textual das notícias, foi estruturado um dicionário de termos positivos e negativos.

Para a elaboração do dicionário, foi realizado um levantamento dos principais jornais do Brasil, conforme os dados do Instituto Verificador de Comunicação (IVC), com o objetivo de gerar a lista de termos, baseadas nas principais palavras utilizadas na redação das matérias. Os jornais selecionados para a compor a amostra para estruturação do dicionário, foram: Estado de São Paulo, Gazeta Mercantil, O Globo, e Valor Econômico.

Após a da seleção dos principais jornais brasileiros, foram coletadas matérias desses jornais, divulgadas na Biblioteca Digital do Senado Federal do Brasil, e que faziam menção às discussões financeiras. Assim, foram coletados 1.118 textos do Jornal Estado de São Paulo, 869 textos da Gazeta Mercantil, 766 textos do Jornal O Globo, e 5.549 textos do Jornal Valor Econômico, totalizando uma amostra de 8.302 textos jornalísticos para a estruturação do dicionário.

As matérias selecionadas passaram por uma leitura automatizada, mediante um algoritmo escrito em linguagem R, para a contagem da frequência das palavras, totalizando 60.103 palavras. A lista de palavras gerais passou por um refinamento para a exclusão das palavras que não apresentavam significado sintático, como artigos, conjunções, pronomes, proposições, preposições e substantivos próprios, restando 36.388 palavras. Dentre as palavras restantes na lista geral, foi realizada uma segregação das palavras em dois grupos de termos, sendo o primeiro formado por palavras que remetem a um significado positivo (918 palavras), e o segundo grupo formado por palavras que reportam a um significado negativo (1.469 palavras). As palavras que não estavam relacionadas aos aspectos positivos ou negativos foram excluídas.

A primeira lista de palavras, proposta na língua portuguesa, foi apresentada na obra de Aguiar (2012), porém, a lista foi construída apenas com palavras oriundas dos relatórios anuais, gerando listas distribuídas entre as seguintes categorias de palavras: (a) positivas; (b) negativas; (c) litigiosas; (d) incerteza; e (e) verbos modais.

No presente estudo, foi construída uma nova lista de palavras em português, por meio da extração das palavras utilizadas em importantes fontes jornalísticas, com o intuito de possibilitar que as investigações sobre o sentimento textual possam ser realizadas em textos que remetam a aspectos contábeis, financeiros ou econômicos.

O processo de validação das listas de palavras positivas e negativas foi realizado mediante uma análise do significado e da origem da palavra, por meio de uma consulta manual a dicionários da língua portuguesa, sendo um dicionário tradicional da língua portuguesa e outro etimológico. Além disso, foi realizado um segundo procedimento de validação que consistiu em uma análise comparativa entre a lista de palavras positivas e negativas proposta e a lista de termos proposta por Aguiar (2012). O segundo procedimento de validação visou a verificação de possíveis erros no processo de classificação dos termos como positivos ou negativos. Em virtude da extensão do número de palavras do dicionário proposto, as listas de palavras positivas e negativas constam nos arquivos suplementares do presente estudo.

3.1 Variáveis

3.1.1 Índice de Sentimento Textual e Volume de Notícias

As notícias passaram por uma análise de conteúdo, levando em consideração a frequência das palavras presentes nos textos, por meio de uma Linguagem de Processamento Natural. Nesse contexto, a mensuração do sentimento textual foi obtida por meio da técnica do *vector space model*, na qual as palavras são reconhecidas como vetores e, após o levantamento da sua frequência, é realizada uma categorização das palavras positivas e negativas, para o posterior levantamento do sentimento textual da notícia.

Assim, o processo de mensuração do sentimento textual foi formado em três etapas: (1ª) leitura automatizada das matérias, extraíndo a frequência de palavras positivas e negativas; (2ª) levantamento do peso das palavras, em relação ao conjunto de textos analisados; e (3ª) cálculo do sentimento textual das notícias, com base no peso das palavras.

Para a categorização das palavras positivas e negativas, utilizou-se o dicionário proposto neste estudo. Já em relação ao levantamento do peso das palavras, utilizou-se a métrica proposta por Loughran e McDonald (2011), que faz uso da técnica do *vector space model*, obtida por meio de uma função logarítmica, conforme Equação 1:

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

Em que: $P_{i,j}$ - peso da palavra i no documento j ; $Tf_{i,j}$ - total de ocorrências de uma palavra i em um documento j ; a_j - média de palavras contadas em um documento j ; N - total de documentos na amostra; df_i - total de documentos com ao menos uma ocorrência da palavra i .

A média das palavras contadas em um documento foi obtida por média aritmética, conforme Equação 2:

$$a_j = \frac{1}{n} \sum_{i=1}^n FP_j \quad (2)$$

Em que: n corresponde ao número de palavras do dicionário que estão presentes no texto; e FP é a frequência das palavras que existem no texto j .

Ao analisar as palavras levando em consideração uma ponderação para a frequência obtida, conforme é apresentado na proposta de Loughran e McDonald (2011), é possível minimizar a discrepância existente entre as palavras que apresentam alta frequência daquelas que apresentam uma baixa frequência, minimizando, assim, os *outliers* e possibilitando uma análise mais precisa.

Mediante o levantamento do peso das palavras positivas e negativas encontradas nos textos, mensurou-se o índice do sentimento textual das notícias, Conforme Equação 3:

$$ST_j = \frac{\sum P_{p,j} - \sum P_{n,j}}{\sum P_{p,j} + \sum P_{n,j}} \quad (3)$$

Em que: ST_j - sentimento textual de um texto j ; $\sum P_{p,j}$ - somatório do peso das palavras positivas de um texto j ; e $\sum P_{n,j}$ - somatório das palavras negativas de um texto j . O sentimento textual da notícia (ST_j) será otimista, quanto mais próximo o índice estiver de 1(um), será neutro quando o resultado do índice for 0 (zero) e será pessimista quanto mais próximo estiver de -1 (menos um).

No estudo, a leitura mecânica das matérias foi realizada pelo site do jornal na área de assinante, por meio de um algoritmo de leitura em linguagem R. Nos dias em que foram publicadas mais de uma notícia, foi realizada uma acumulação dos pesos obtidos por texto e calculada uma média aritmética dos pesos, com o objetivo de levantar o sentimento textual diário. As matérias sobre a Samarco que foram divulgadas em feriados ou nos finais de semana também tiveram seus pesos acumulados e foram analisadas no conjunto de notícias do primeiro dia útil, por considerar que o mercado reage a essas informações no primeiro dia útil seguinte a sua publicação (García, 2013). Portanto, para cada dia analisado, foi levantado um índice do sentimento textual das notícias.

Além do sentimento textual das notícias, levantou-se a frequência diária de notícias publicadas durante o período do estudo, com a finalidade de investigar se o volume de notícias sobre o caso Samarco apresentava alguma relação com as oscilações do risco sistemático das ações preferenciais da Vale (VALE5), considerando como volume de notícias o somatório das matérias diárias que saíram sobre o caso Samarco, no Jornal Valor Econômico.

3.1.2 Risco sistemático

A modelagem do risco sistemático é dada pelo beta dos modelos de precificação de ativos. No CAPM, o retorno esperado de um ativo é uma função do risco existente, dada por uma relação de linearidade entre o retorno e o risco, conforme a Equação 4:

$$E(r_i) = r_f + \frac{cov(r_i r_m)}{var(r_m)} (E(r_m) - r_f) \quad (4)$$

Em que: $E(r_i)$ - é a esperança do retorno de um ativo i ; r_f - retorno do ativo livre de risco, $\frac{cov(r_i r_m)}{var(r_m)}$ - corresponde ao beta do ativo, $E(r_m)$ - esperança do retorno da carteira de mercado.

O beta - β_i é dado em função da $cov(r_i r_m)$ - covariância de um ativo i em relação à carteira de investimentos do mercado r_m dividida pela variância da carteira de investimento $var(r_m)$. Assim, é mensurado o risco agregado do ativo em relação à carteira que representa o mercado, sendo ponderada pela variância da carteira do mercado. Desse modo, o beta é uma medida sensível que capta o risco sistemático de um título mediante as oscilações do mercado.

Na visão tradicional do CAPM, a estimação dos parâmetros, como o beta, é constante no tempo. No entanto, ao considerar que a variância e a covariância oscilam no tempo, é possível estimar o CAPM condicional ou dinâmico, de modo que a esperança do retorno de um ativo está condicionada a um conjunto de informações, pressupondo, com isso, que os betas variam no tempo (González-Rivera, 2013). Com isso, o CAPM condicional é estimado conforme Equação 5.

$$E(r_{i,t}|I_{t-1}) = r_f + \frac{cov(r_{i,t} r_{m,t}|I_{t-1})}{var(r_{m,t}|I_{t-1})} (E(r_{m,t}|I_{t-1}) - r_f) \quad (5)$$

Ao assumir que a covariância e a variância oscilam no tempo, e que a covariância e a variância podem ser escritas em termos do desvio-padrão (σ), é possível estimar um beta que oscila no tempo dado um conjunto de informações, sendo chamado de beta dinâmico. A estimação do beta dinâmico, obtida conforme Equação 6, admite que o coeficiente de correlação do ativo em relação ao mercado - ρ_{im} é constante no tempo.

$$\beta_{i,t} = \frac{cov(r_{i,t} r_{m,t}|I_{t-1})}{var(r_{m,t}|I_{t-1})} = \frac{\rho_{im}(\sigma_{i,t|t-1})(\sigma_{m,t|t-1})}{\sigma_{m,t|t-1}^2} = \rho_{im} \frac{\sigma_{i,t|t-1}}{\sigma_{m,t|t-1}} \quad (6)$$

No presente estudo, para a estimação do risco sistemático, foi adotado o beta dinâmico, estimado com modelos Autoregressivos de Heterocedasticidade Condicional (ARCH-GARCH), em que foi testada a distribuição condicional do modelo GARCH, proposto por Bollerslev (1986), assumindo a distribuição com densidade normal e com distribuição *t-Student*. Assim, para a escolha do modelo, foram observados os critérios informacionais Akaike e Schwarz. Além disso, foi realizado o teste *ljung-box*, para testar a autocorrelação dos resíduos. Desse modo, foi utilizado para a estimação dos betas dinâmicos o GARCH (1,1), assumindo densidade normal.

3.2 Modelo Econométrico

A análise da relação entre as variáveis tom e o volume de notícias sobre o risco sistemático foi realizada mediante regressão quantílica, conforme equação 7, visando investigar o comportamento da distribuição condicional do beta dinâmico nos maiores *decis*, em função deles atribuírem um maior peso para os valores mais altos do beta dinâmico, que consiste no maior risco de um título. Assim, os maiores *decis* da distribuição do beta dinâmico representam os maiores riscos para os investidores.

$$Q\beta_{i,t}(\tau_i|F_t) = \alpha_0(\tau) + \alpha_1(\tau)ST_t + \alpha_2(\tau)VOL_t + \alpha_3(\tau)VOL_t * ST_t + \varepsilon_t \quad (7)$$

Em que: $Q\beta_{i,t}(\tau_i|F_t)$ - *decil* condicional de beta dinâmico ou risco sistemático; $\alpha_0(\tau)$ - intercepto no *decil* de ordem τ ; $\alpha_1(\tau)$ - coeficiente angular do sentimento textual no *decil* de ordem τ ; $\alpha_2(\tau)$ - coeficiente angular do volume de notícias no quantil de ordem τ ; $\alpha_3(\tau)$ - coeficiente angular da interação entre o volume de notícias e o sentimento textual; ST_t - sentimento

textual no período t ; VOL_t - quantidade de notícias sobre a Samarco no período t ; ε_t – termo de erro, o qual não recebeu nenhuma suposição a sua distribuição, conforme é orientado por Koenker (2005).

Assim, o coeficiente $\alpha_1(\tau)$ vai captar a influência que o sentimento das notícias exerce sobre o risco sistemático, o coeficiente $\alpha_2(\tau)$ vai capturar a relação que o volume de notícias exerce sobre a oscilação do risco e $\alpha_3(\tau)$ vai capturar o efeito da combinação entre o volume de notícia e o sentimento textual sobre o risco sistemático.

Os parâmetros da Equação 7 foram estimados por meio de regressão quantílica, em função da necessidade de examinar o comportamento das maiores distribuições do beta, visto que, nos extremos da distribuição, é encontrada a maior exposição ao risco. Além disso, foi utilizada a regressão quantílica em função da variável dependente (Beta) apresentar uma distribuição com grande dispersão nos seus valores, sugerindo a presença de *outliers* na amostra, conforme estatística descritiva apresentada na Tabela 1. Desse modo, optou-se pelo uso da regressão quantílica, tendo em vista a possibilidade de estimação dos coeficientes por *decis*, além de ser uma estimação robusta à presença de *outliers*.

O estudo parte da suposição de que os investidores são sensíveis às notícias provenientes dos jornais, em períodos que remetem a uma maior incerteza sobre os fluxos de caixa futuro (GARCÍA, 2013). Portanto, espera-se que o sentimento textual e o volume de notícias provoquem impactos sobre o risco sistemático nos maiores *decis*, por eles representarem os períodos de maior exposição ao risco.

4 Análise dos Resultados

4.1 Estatística descritiva

A Tabela 1 apresenta a estatística descritiva das variáveis utilizadas no estudo, relacionadas ao período de 05 de novembro de 2015 a 16 de fevereiro de 2016. Diante dos resultados levantados verifica-se que o risco sistemático da VALE5, estimado pelo Beta, apresentou um valor médio de 3,825, mediana de 3,420 e um desvio padrão de 1,906, indicando que, ao longo do período analisado, a ação VALE5 foi considerada um ativo de alto risco, por oscilar mais que a carteira teórica do mercado. Além disso, observa-se que na série temporal do risco sistemático existe a presença de valores extremos, em virtude da variável beta apresentar uma distribuição leptocúrtica (curtose= 7,059). Essas evidências sugerem que no período avaliado o ativo investigado (VALE5) apresentou, em média, uma alta exposição ao risco, resultado que, de certa forma, era esperado, em virtude da magnitude do acidente provocado pela Samarco e das incertezas, que esse acidente gerou sobre os fluxos de caixa futuro da Vale S.A, em função da empresa deter parte do controle da mineradora Samarco.

Em relação à variável sentimento textual (ST), observa-se que prevaleceu um tom de pessimismo nas notícias veiculadas na mídia sobre a empresa Samarco, em função da média (-0,561) e mediana (-0,613) serem negativas, o que era esperado, em decorrência da magnitude da tragédia provocada com o rompimento das barragens sob responsabilidade da Samarco. Isso indica que a mídia divulgou matérias sobre a Samarco, com a utilização de palavras que, no geral, exprimem um sentimento de pessimismo.

Quanto ao volume de notícias (Volume) sobre a mineradora Samarco, constata-se que o valor médio foi de 4.803 e a mediana foi de 3, evidenciando que, ao longo dos três meses seguintes ao acidente, a mídia chegou a divulgar mais de uma notícia por dia. É possível observar, também, que o volume de notícias apresentou grandes oscilações, chegando ao patamar de mais de 12 notícias por dia (percentil 95%=12,65), sugerindo que o aumento da cobertura da mídia pode estar associado à gravidade das notícias divulgadas, sobre as consequências do rompimento das barragens.

Tabela 1: Estatística descritiva das variáveis

Variável	Média	Percentil 5%	Mediana	Percentil 95%	Desvio padrão	Assimetria	Curtose
Beta	3,825	1,729	3,420	9,159	1,906	1,886	7,471
ST	-0,561	-1	-0,613	0,140	0,346	1,730	8,151
Volume	4,803	1	3	12,65	4,058	1,882	7,179

Nota: Amostra composta por 66 observações, referentes ao período de 05 de novembro de 2015 a 16 de fevereiro de 2016.

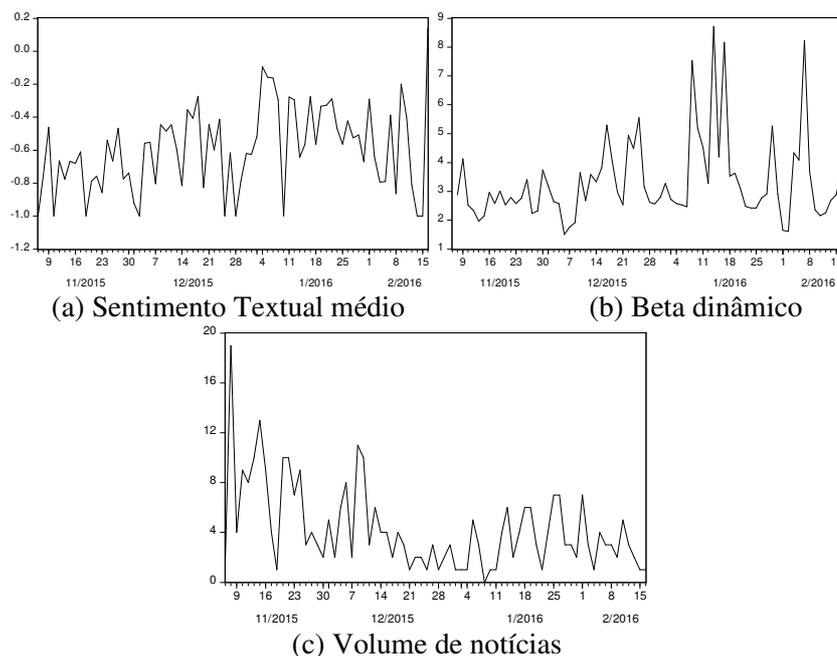
Fonte: dados da pesquisa

4.2 Análise gráfica do Volume de notícias e Sentimento Textual

As variáveis foram projetadas graficamente para a realização de uma análise comparativa dos comportamentos de maior oscilação das séries. Assim, o Beta estimado pelo CAPM condicional, corresponde a uma medida de mensuração do risco sistemático diário de um ativo, geralmente, influenciado por informações que provocam alterações no retorno das ações.

Conforme a Figura 1 (b), constata-se que, no mês de novembro/2015, as maiores oscilações do beta dinâmico da ação VALE5 ocorreram nos dias seguintes ao anúncio do desastre envolvendo a Mineradora Samarco, período marcado por uma ampla cobertura na divulgação de notícias, conforme pode ser observado na Figura 1 (c). Além disso, nesse mesmo período, a Figura 1 (a), apresenta o tom das notícias sobre a dimensão do acidente e sobre os primeiros impactos ambientais e financeiros, com uma alta predominância de um sentimento negativo nas notícias sobre a Samarco.

Figura 1: Série temporal do beta dinâmico do ativo VALE5, sentimento textual das matérias da Samarco e volume das notícias



Ao analisar o comportamento do beta dinâmico no mês de dezembro/2015, verifica-se que as maiores oscilações do beta da ação analisada (VALE5) ocorreram no período em que foram divulgadas notícias fazendo referência às consequências do acidente, como as matérias que trataram sobre o processo de indenização movido contra a Vale S.A., por conta das perdas causadas aos investidores, e em função dos problemas provocados no acidente envolvendo a Samarco.

No mês de janeiro/2016, o ativo (VALE5) apresentou uma maior exposição ao risco, em função da presença dos maiores betas da série temporal. Fazendo uma associação desse resultado com a divulgação de notícias, observa-se que os maiores betas ocorreram no período marcado pelo

aumento no volume de notícias, dentre as quais podem ser citadas as matérias sobre os riscos de novos acidentes com barragens sob responsabilidade da Samarco, e a divulgação da notícia de que a Polícia Federal havia indiciado a Vale e a Samarco como responsáveis pelo acidente, que culminou no vazamento de resíduos tóxicos.

Por fim, no mês de fevereiro/2016, os maiores betas estão concentrados no período em que surgiram notícias sobre a movimentação de processos contra a Samarco, responsabilizando-a pelas mortes caudas no acidente, bem como relatos de que o acidente teria ocorrido em função de negligência da empresa, visando uma maior lucratividade.

4.3 Resultados Econométricos

A relação entre o sentimento textual e o volume das notícias sobre a Samarco no risco sistemático das ações da Vale S.A. (VALE5) foi estimada mediante regressão quantílica, e as estimativas são apresentadas na Tabela 2. Assim, para verificar se o risco sistemático das ações da Vale (VALE5) sofre a influência do volume e do tom de notícias nos períodos de maior exposição ao risco, foi analisado o comportamento referente à distribuição do Beta entre o 5º ao 9º *decil*.

Os resultados apresentados na Tabela 2 evidenciam o impacto do sentimento textual e do volume das notícias sobre o risco sistemático. Ao analisar o 5º *decil* τ (0.5), que representa a mediana da amostra, observa-se que o risco sistemático não sofre a influência do sentimento textual e do volume das notícias sobre o acidente envolvendo a mineradora Samarco, no período de análise. Ao analisar o 6º τ (0.6) e o 7º τ (0.7) *decil*, constata-se um comportamento semelhante ao 5º *decil*, em que o volume e o sentimento textual não apresentam relações com o risco sistemático, visto que os coeficientes das variáveis do modelo não foram significativos. Isso indica que o comportamento do risco sistemático na mediana, assim como no 6º e no 7º *decil*, não é explicado pelo conteúdo ou pelo volume de notícias, o que pode estar relacionado a uma ausência de sensibilidade dos investidores às informações divulgadas na mídia em função de não ser uma situação de risco extremo.

Tabela 2: Estimativas dos parâmetros - Regressão Quantílica

Variáveis	Decis									
	0,5		0,6		0,7		0,8		0,9	
	Coef.	p-valor								
ST	-0,019	0,976	-0,072	0,914	-0,448	0,487	-1,001	0,165	-2,245	0,018**
Volume	0,094	0,597	0,187	0,376	0,280	0,317	0,676	0,026**	1,337	0,122
ST*Vol	0,042	0,837	0,140	0,533	0,215	0,428	0,627	0,053*	1,346	0,105
Pseudo R ²	0,023		0,026		0,045		0,085		0,134	

***significante a 1%, ** significante a 5% e * significante a 10%.

Fonte: dados da pesquisa

Nota: Amostra composta por 66 observações, referentes ao período de 05 de novembro de 2015 a 16 de fevereiro de 2016. Regressão nos quantis (τ): 0,5; 0,6; 0,7; 0,8; 0,9, admitindo que os resíduos não eram identicamente distribuídos, com base na matriz de *Huber Sandwich Standard*. A estatística R² foi o pseudo R², conforme proposta de Koenker e Machado (1999).

No 8º τ (0.8) *decil*, observa-se que o volume de notícias influencia o risco sistemático, de modo que o aumento no volume de notícias divulgadas sobre o acidente da Samarco provoca um aumento do beta em 0,67, ou seja, o volume de notícias sobre o acidente contribuiu com o aumento do risco, pois foi encontrada significância estatística para o coeficiente da referida variável (Volume *p-valor*=0,026). Já o sentimento textual, por si só, não provoca impactos sobre o beta da VALE5, porém, quando o sentimento textual é avaliado interagindo com o volume de notícias, observa-se que existem efeitos sobre o beta (ST*Vol *p-valor*=0.0537), com um aumento do beta

em 0,62, ou seja, períodos marcados por notícias que, em média, são pessimistas e com ampla cobertura tendem a influenciar o risco sistemático. Desse modo, pode-se inferir que o tom das notícias sobre a Samarco, atrelado a uma ampla divulgação da mídia contribuíram para aumento da exposição ao risco sistemático das ações VALE5 no período analisado, corroborando com as evidências empíricas apresentadas por Liu e McConnell (2013), de que os investidores são sensíveis ao sentimento das notícias que recebem uma ampla disseminação na mídia.

As evidências levantadas no 9º τ (0.9) *decil* mostram que, nos períodos com maior exposição ao risco sistemático, maiores betas, o tom das notícias, por si só, influencia o beta da ação (ST p -valor=0,018), provocando uma redução do beta em -2,24. A repercussão das notícias, estimada mediante o volume (p -valor=0,122), e a interação entre o sentimento textual e o volume de informações (p -valor=0,105), apesar de não apresentarem significâncias estatísticas, apresentaram uma significância marginal, nos resultados dos seus coeficientes, em decorrência dos valores serem próximos ao nível de significância adotado. Desse modo, os resultados encontrados para o 9º *decil* sugerem que o risco sistemático é influenciado pelo volume e pelo conteúdo das notícias nos momentos de maior exposição ao risco.

Ao analisar o comportamento do pseudo R^2 , constata-se que, em geral, os modelos apresentaram um baixo poder explicativo, o que está relacionado ao fato do modelo (Equação 7) incorporar apenas duas características das notícias, relacionadas ao conteúdo e ao volume, não sendo incorporado no modelo outros aspectos determinantes do risco sistemático, em função de divergência na frequência dos dados. Além disso, é possível observar que o poder explicativo do modelo melhora no 8º e 9º *decil*, fato que pode estar relacionado ao impacto que as informações textuais provocam sobre o risco sistemático nos momentos em que existe uma maior exposição ao risco.

Diante dos resultados encontrados na amostra analisada, é possível constatar que, na medida em que se aumenta a exposição ao risco sistemático, o tom e o volume das notícias tendem a influenciar, de algum modo, o beta da ação. O comportamento encontrado no estudo corrobora as evidências apontadas por García (2013), de que existe uma maior reação no mercado ao conteúdo das notícias da mídia em períodos que remetem a recessão ou crises, visto que, nos momentos de maior instabilidade econômica, os investidores estariam mais vulneráveis a serem movidos pelo sentimento das notícias da mídia.

As evidências empíricas levantadas no estudo sinalizam indícios de que os agentes são influenciados pelas notícias em períodos que remetem a uma maior incerteza sobre os fluxos de caixa futuro dos ativos. Desse modo, acredita-se que as notícias contribuem com a para atualização das crenças dos agentes, sobre fundamentos da organização, em períodos que remetem a uma maior incerteza.

Portanto, as evidências empíricas encontradas entre o 8º e 9º *decil* levam a constatação de que o volume e o tom das notícias veiculadas na mídia influenciam o beta da ação, nos momentos em que a empresa apresenta uma maior exposição ao risco. Com isso, não se pode rejeitar a hipótese da pesquisa, de que o risco sistemático da ação (VALE5) sofre a influência do volume e do tom de notícias nos períodos em que os investidores estão mais expostos ao risco.

5 Conclusão

O estudo examinou os efeitos que as informações provenientes de notícias divulgadas em jornais poderiam provocar sobre o risco sistemático das ações, partindo da suposição de que, em períodos de maior incerteza, o risco sistemático apresenta conexão com as divulgações de notícias pela mídia.

Assim, foram analisados o volume e o sentimento textual das matérias publicadas pelo Jornal Valor Econômico que faziam referência à mineradora Samarco, nos três meses subsequentes à tragédia, que ocorreu com o rompimento de barragens sob a responsabilidade da

Samarco. Desse modo, procurou-se investigar o choque das notícias sobre a Samarco, no risco sistemático das ações preferenciais da Vale (VALE5), partindo da suposição de que as divulgações de notícias sobre a Samarco tendem a afetar o risco das ações da Vale, tendo em vista que a Vale S.A. detém 50% do controle da Samarco.

As evidências empíricas mostraram interligações entre o risco sistemático, volume de notícias e o tom das matérias sobre a Samarco, nos períodos em que existiu uma maior oscilação da ação VALE5 em relação à carteira teórica do mercado. Desse modo, os resultados levaram à constatação de que o aumento na exposição do ativo ao risco sistemático direciona o sentimento textual e o volume de notícias a influenciar, de algum modo, o beta da ação. Portanto, na amostra analisada, pode-se concluir que o risco sistemático, estimando mediante o beta da ação, é influenciado pelo volume de notícias e/ou pelo sentimento textual das matérias veiculadas na mídia nos momentos em que o ativo avaliado (VALE5) apresentou uma maior exposição ao risco.

Com isso, pode-se dizer que a pesquisa contribuiu com o avanço da literatura sobre análise de sentimento textual, na medida em que propõem um índice de sentimento textual para a língua portuguesa, que permite a extração do posicionamento de otimismo ou pessimismo dos textos, possibilitando a exploração de *big data* de textos escritos em português. A pesquisa, também, corrobora com a literatura sobre risco sistemático, ao apresentar uma *proxy* alternativa para notícias baseada em informações textuais, visto que, nas investigações sobre risco sistemático, geralmente, são utilizadas informações puramente quantitativas, oriundas de dados financeiros e econômicos.

Além disso, ao levantar evidências empíricas da sensibilidade do risco sistemático a uma *proxy* de notícias (sentimento textual), o estudo amplia a compreensão do valor informacional do sentimento das notícias para o mercado acionário. Portanto, acredita-se que informações atreladas ao sentimento textual das notícias podem corroborar com os agentes do mercado, na estruturação de estratégias de investimento, visando a minimização da exposição ao risco, principalmente em períodos de maior incerteza econômica.

As evidências apresentadas neste estudo encontram-se limitadas à métrica de estimação do sentimento textual, que utilizou linguagem de processamento natural para a realização da leitura mecânica dos textos e para a extração do tom das notícias. Além disso, as conclusões do estudo estão limitadas à análise das matérias que retrataram sobre o acidente envolvendo a mineradora Samarco em relação ao comportamento do risco sistemático das ações preferenciais da Vale. Por fim, sugere-se que pesquisas futuras sejam realizadas com outras empresas e com a utilização de outros tipos de informações textuais, a fim de verificar se o risco sistemático é impulsionado pelo conteúdo da mídia.

Referências

AGUIAR, Marcelo Otone. **Sentiment analysis em relatórios da administração divulgados por firmas brasileiras**. 2012. 92 f. Dissertação (Mestrado em Administração de Empresas)-Fundação Instituto Capixaba de Pesquisas em Contabilidade, Economia e Finanças, Vitória, 2012.

AHERN, Kenneth R.; SOSYURA, Denis. Rumor has it: Sensationalism in financial media. **The Review of Financial Studies**, v. 28, n. 7, p. 2050-2093, 2015. DOI: <https://doi.org/10.1093/rfs/hhv006>

ALLEN, David E.; MCALEER, Michael; SINGH, Abhay K. An entropy-based analysis of the relationship between the DOW JONES Index and the TRNA Sentiment series. **Applied Economics**, v. 49, n. 7, p. 677-692, 2017. DOI: <https://doi.org/10.1080/00036846.2016.1203067>

AMORIM, A. L. G. C.; LIMA, I. S.; MURCIA, F. D. R. Análise da relação entre as informações

contábeis e o risco sistemático no mercado brasileiro. **Revista Contabilidade & Finanças-USP**, v. 23, n. 60, 2012. DOI: <http://dx.doi.org/10.1590/S1519-70772012000300005>

BEAVER, William; KETTLER, Paul; SCHOLLES, Myron. The association between market determined and accounting determined risk measures. **The Accounting Review**, v. 45, n. 4, p. 654-682, 1970. DOI: <https://www.jstor.org/stable/244204>.

BEATTIE, Vivien. Accounting narratives and the narrative turn in accounting research: Issues, theory, methodology, methods and a research framework. **The British Accounting Review**, v. 46, n. 2, p. 111-134, 2014. DOI: <https://doi.org/10.1016/j.bar.2014.05.001>

BBC BRASIL. Tragédia com barragem da Vale em Brumadinho pode ser a pior no mundo em 3 décadas. Disponível em: <https://www.bbc.com/portuguese/brasil-47034499>. Acesso: dez 2019.

BIBLIOTECA DIGITAL DO SENADO FEDERAL DO BRASIL. Acervo de jornais. **Portal O Senado**, 2017. Disponível em: <http://www2.senado.leg.br/bdsf/handle/id/5>. Acesso em: 30 jan. 2017.

BIANCONI, Marcelo; HUA, Xiabin; TAN, Chih Ming. Determinants of systemic risk and information dissemination. **International Review of Economics & Finance**, v. 38, p. 352-368, 2015. DOI: <https://doi.org/10.1016/j.iref.2015.03.010>

BILLER, D.; WILLIS, A. Samarco pode ter que pagar R\$ 1 bi por acidente em MG. **Bloomberg**, 2015. Disponível em: <https://www.bloomberg.com.br/blog/samarco-pode-ter-que-pagar-1-bi-por-acidente-em-mg/>. Acesso em: 30 jan. 2017.

BOLLERSLEV, Tim. Generalized autoregressive conditional heteroskedasticity. **Journal of econometrics**, v. 31, n. 3, p. 307-327, 1986. DOI: [https://doi.org/10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)

CAMPBELL, John Y.; MEI, Jianping. Where do betas come from? Asset price dynamics and the sources of systematic risk. **The Review of Financial Studies**, v. 6, n. 3, p. 567-592, 1993. DOI: <https://doi.org/10.1093/rfs/5.3.567>

CAMPBELL, John Y.; SHILLER, Robert J. The dividend-price ratio and expectations of future dividends and discount factors. **The Review of Financial Studies**, v. 1, n. 3, p. 195-228, 1988. DOI: <https://doi.org/10.1093/rfs/1.3.195>

COASE, R. H. The problem of social cost. **The Journal of Law and Economics**, v.3, 1960, p. 1-44.

CARROLL, Archie B. A three-dimensional conceptual model of corporate performance. **Academy of management review**, v. 4, n. 4, p. 497-505, 1979. DOI: <https://doi.org/10.5465/amr.1979.4498296>

DAVIS, Keith. The case for and against business assumption of social responsibilities. **Academy of Management journal**, v. 16, n. 2, p. 312-322, 1973. DOI: <https://doi.org/10.5465/255331>.

FAMA, Eugene F. Efficient capital markets: A review of theory and empirical work. **The journal of Finance**, v. 25, n. 2, p. 383-417, 1970. DOI: <https://doi.org/10.1111/j.1540-6261.1970.tb00518.x>

FAUSTINO, C. C. R. **O impacto das notícias no mercado financeiro brasileiro**. 2013, 72 f. Dissertação (Mestrado em Economia) Fundação Getúlio Vargas – FGV, São Paulo, 2012.

FRIEDMAN, Milton. The social responsibility of business is to increase its profits. **The New York Times Magazine**, 13, 1970.

FERGUSON, Nicky J. et al. Media content and stock returns: The predictive power of press. **Multinational Finance Journal**, v. 19, n. 1, p. 1-31, 2015.

FOSTER, George; KASZNIK, Ron; SIDHU, Baljit K. International equity valuation: the relative importance of country and industry factors versus company-specific financial reporting information. **Accounting & Finance**, v. 52, n. 3, p. 767-814, 2012. DOI: <https://doi.org/10.1111/j.1467-629X.2011.00423.x>

GARCIA, Diego. Sentiment during recessions. **The Journal of Finance**, v. 68, n. 3, p. 1267-1300, 2013. DOI: <https://doi.org/10.1111/jofi.12027>

GRIFFIN, John M.; HIRSCHHEY, Nicholas H.; KELLY, Patrick J. How important is the financial media in global markets?. **The Review of Financial Studies**, v. 24, n. 12, p. 3941-3992, 2011. DOI: <https://doi.org/10.1093/rfs/hhr099>

GRIFFIN, Paul A.; LONT, David H. **Game Changer? The Impact of the VW Emission Cheating Scandal on the Co-Integration of Large Automakers' Securities**. SSRN, 2016. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2838949. Acesso em: 31 jan. 2017.

GONZALEZ-RIVERA, Gloria. **Forecasting for economics and business**. Routledge, 2016.

HENDERSHOTT, Terrence; LIVDAN, Dmitry; SCHÜRHOFF, Norman. Are institutions informed about news?. **Journal of Financial Economics**, v. 117, n. 2, p. 249-287, 2015. DOI: <https://doi.org/10.1016/j.jfineco.2015.03.007>

HUANG, Xuan; TEOH, Siew Hong; ZHANG, Yinglei. Tone management. **The Accounting Review**, v. 89, n. 3, p. 1083-1113, 2013. DOI: <https://doi.org/10.2308/accr-50684>

INSTITUTO VERIFICADOR DE CIRCULAÇÃO - IVC. Dados oficiais do IVC relacionados a circulação de jornais impressos e digitais de Fevereiro 2017. Disponível em: <https://www.poder360.com.br/midia/jornais-e-revistas-continuam-avancando-em-2017-com-suas-edicoes-digitais/>. Acesso em: 19 set. 2017.

KOTHARI, Sabino P.; LI, Xu; SHORT, James E. The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. **The Accounting Review**, v. 84, n. 5, p. 1639-1670, 2009. DOI: <https://doi.org/10.2308/accr.2009.84.5.1639>

KOENKER, Roger. **Quantile regression**. Econometric Society Monographs. Cambridge University Press, n. 38, 2005.

KOENKER, Roger; MACHADO, Jose AF. Goodness of fit and related inference processes for

quantile regression. **Journal of the American Statistical Association**, v. 94, n. 448, p. 1296-1310, 1999. DOI: <http://dx.doi.org/10.1080/01621459.1999.10473882>.

LARSON, E. **Freedom Industries Files Bankruptcy After Elk River Spill**. Bloomberg, 2014. Disponível em: <https://www.bloomberg.com/news/articles/2014-01-17/freedom-industries-files-for-bankruptcy-in-west-virginia>. Acesso em: 31 jan. 2017.

LI, Feng. The information content of forward-looking statements in corporate filings—A naïve Bayesian machine learning approach. **Journal of Accounting Research**, v. 48, n. 5, p. 1049-1102, 2010. DOI: <https://doi.org/10.1111/j.1475-679X.2010.00382.x>

LINTNER, J. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. **The Review of Economics and Statistics**, 13-37, 1965. DOI: <https://doi.org/10.1016/B978-0-12-780850-5.50018-6>

LIU, Baixiao; MCCONNELL, John J. The role of the media in corporate governance: Do the media influence managers' capital allocation decisions? **Journal of Financial Economics**, v. 110, n. 1, p. 1-17, 2013. DOI: <https://doi.org/10.1016/j.jfineco.2013.06.003>

LOUGHRAN, Tim; MCDONALD, Bill. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. **The Journal of Finance**, v. 66, n. 1, p. 35-65, 2011. DOI: <https://doi.org/10.1111/j.1540-6261.2010.01625.x>

MAR-MOLINERO, Cecilio; MENÉNDEZ-PLANS, Carlota; ORGAZ-GUERRERO, Neus. Has the 2008 financial crisis changed the factors determining the systematic risk of shares in the “European Hospitality Industry”? (2003–2013). **Journal of Hospitality and Tourism Management**, v. 31, p. 59-69, 2017. DOI: <https://doi.org/10.1016/j.jhtm.2016.10.002>

MOSSIN, Jan. Equilibrium in a capital asset market. **Econometrica: Journal of the Econometric Society**, p. 768-783, 1966. Disponível em: <https://www.jstor.org/stable/1910098>

SHARPE, William F. Capital asset prices: A theory of market equilibrium under conditions of risk. **The journal of finance**, v. 19, n. 3, p. 425-442, 1964. DOI: <https://doi.org/10.1111/j.1540-6261.1964.tb02865.x>

STRAUß, Nadine; VLIAGENTHART, Rens; VERHOEVEN, Piet. Lagging behind? Emotions in newspaper articles and stock market prices in the Netherlands. **Public Relations Review**, v. 42, n. 4, p. 548-555, 2016. DOI: <https://doi.org/10.1016/j.pubrev.2016.03.010>

TETLOCK, Paul C. Giving content to investor sentiment: The role of media in the stock market. **The Journal of finance**, v. 62, n. 3, p. 1139-1168, 2007. DOI: <https://doi.org/10.1111/j.1540-6261.2007.01232.x>

TETLOCK, Paul C. Information transmission in finance. **Annu. Rev. Financ. Econ.**, v. 6, n. 1, p. 365-384, 2014. DOI: <https://doi.org/10.1146/annurev-financial-110613-034449>

TETLOCK, Paul C.; SAAR-TSECHANSKY, Maytal; MACSKASSY, Sofus. More than words: Quantifying language to measure firms' fundamentals. **The Journal of Finance**, v. 63, n. 3, p. 1437-1467, 2008. DOI: <https://doi.org/10.1111/j.1540-6261.2008.01362.x>

VALOR ECONÔMICO. Missão Editorial, 2017. Disponível em: <http://midiakit.valor.com.br/index.php?p=2>. Acesso em: 18 set. 2017.

WANG, Kent; LI, Jiawei; HUANG, Shicheng. Bad beta good beta, state-space news decomposition and the cross-section of stock returns. **Accounting & Finance**, v. 53, n. 2, p. 587-607, 2013. DOI: <https://doi.org/10.1111/j.1467-629X.2012.00470.x>

* Uma versão preprint do artigo foi apresentado no XI Congresso ANPCONT, 2017.