

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E CONTABILIDADE  
DEPARTAMENTO DE ADMINISTRAÇÃO  
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO

CLÁUDIO GOMES

**Revisitando a relação entre reputação e desempenho financeiro: uma análise  
baseada em mineração de texto e aprendizado de máquina**

São Paulo

2020

Prof. Dr. Vahan Agopyan  
Reitor da Universidade de São Paulo

Prof. Dr. Fábio Frezatti  
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Moacir de Miranda Oliveira Júnior  
Chefe de Departamento de Administração

Prof. Dr. Eduardo Kazuo Kayo  
Coordenador do Programa de Pós-Graduação em Administração

CLÁUDIO GOMES

**Revisitando a relação entre reputação e desempenho financeiro: uma análise baseada em mineração de texto e aprendizado de máquina**

Dissertação apresentada ao Programa de Pós-Graduação em Administração do Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, como requisito parcial para a obtenção do título de Mestre em Ciências.

Área de Concentração: Métodos Quantitativos e Informática

Orientadora: Profa. Dra. Adriana Backx Noronha Viana

Versão Corrigida

(versão original disponível na Biblioteca da Faculdade de Economia, Administração e Contabilidade)

São Paulo  
2020

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação (CIP)  
Ficha Catalográfica com dados inseridos pelo autor

Gomes, Claudio.

Revisitando a relação entre reputação e desempenho financeiro: uma análise baseada em mineração de texto e aprendizado de máquina / Claudio Gomes. - São Paulo, 2020.  
185 p.

Dissertação (Mestrado) - Universidade de São Paulo, 2020.  
Orientador: Adriana Backx Noronha Viana.

1. Reputação corporativa. 2. Desempenho financeiro. 3. Mídia. 4. Notícias on-line. 5. Mineração de texto. I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

**Nome:** Gomes, Cláudio

**Título:** Revisitando a relação entre reputação e desempenho financeiro: uma análise baseada em mineração de texto e aprendizado de máquina

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Administração do Departamento de Administração da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito para a obtenção do título de Mestre em Ciências.

Aprovado em:

Banca examinadora

Prof. Dra. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_



## RESUMO

Gomes, C. (2020). Revisitando a relação entre reputação e desempenho financeiro: uma análise baseada em mineração de texto e aprendizado de máquina (Dissertação de Mestrado). Faculdade de Economia, Administração e Contabilidade. Universidade de São Paulo, São Paulo.

A reputação corporativa é um ativo intangível valioso, raro, difícil de imitar e não substituível, o qual conduz a uma vantagem competitiva sustentável. Pesquisas empíricas anteriores confirmam que existe uma relação positiva entre reputação corporativa e desempenho financeiro. Tendo como base a teoria da Visão Baseada em Recursos e na teoria da Sinalização, este estudo explora os efeitos da reputação corporativa, mensurada por meio de notícias na mídia *on-line*, sobre a lucratividade e o valor de mercado das empresas de capital aberto listadas na B3 – Brasil, Bolsa, Balcão. A mídia atua como uma intermediária no processo de sinalização entre as empresas e os *stakeholders*, o que mitiga a assimetria de informações. Além desses aspectos, esta pesquisa examina, também, a interação entre as dimensões da reputação corporativa, o que pode reforçar suas consequências no desempenho financeiro. As técnicas de regressão linear e de regressão através do algoritmo de aprendizado de máquina *Gradient Boosting Machines* são usadas para testar empiricamente um modelo teórico que vincula a reputação corporativa ao desempenho financeiro de maneira multidimensional. O conjunto de dados da pesquisa consiste em indicadores financeiros e textos de notícias *on-line* de empresas listadas na B3. Os resultados da pesquisa demonstraram que a reputação corporativa influencia a lucratividade e o valor de mercado das empresas, em conformidade com o que preconiza a teoria da Visão Baseada em Recursos, uma vez que ela é um ativo estratégico socialmente construído a partir das percepções dos *stakeholders*. Este estudo contribui para a literatura sobre a relação entre reputação e desempenho financeiro, mediante a mensuração da reputação por meio de informações da mídia jornalística. Em termos de contribuições gerenciais, os resultados indicam que as empresas devem considerar a implementação de ações de monitoramento da mídia para gestão da reputação, uma vez que a reputação influencia o desempenho financeiro. Finalmente, as sugestões para futuras pesquisas são discutidas à luz das limitações metodológicas apresentadas.

Palavras-chave: Reputação corporativa. Desempenho financeiro. Mídia. Notícias *on-line*.  
Mineração de texto.

## ABSTRACT

Gomes, C. (2020). Revisiting the relationship between reputation and financial performance: an analysis based on text mining and machine learning (Dissertação de Mestrado). Faculdade de Economia, Administração e Contabilidade. Universidade de São Paulo, São Paulo.

Corporate reputation is a valuable, rare, difficult to imitate and non-replaceable intangible asset, which leads to a sustainable competitive advantage. Previous empirical research confirms that there is a positive relationship between corporate reputation and financial performance. Based on the theory of Resource Based View and Signaling theory, this study explores the effects of corporate reputation, measured through news in the online media, on the profitability and market value of listed public companies on B3 – Brasil, Bolsa, Balcão. The media acts as an intermediary in the signaling process between companies and stakeholders, which mitigates information asymmetry. In addition to these aspects, this research also examines the interaction between the dimensions of corporate reputation, which can reinforce its consequences on financial performance. The techniques of linear regression and regression through the machine learning algorithm Gradient Boosting Machines are used to empirically test a theoretical model that links corporate reputation to financial performance in a multidimensional way. The research data set consists of financial indicators and online news texts from companies listed on B3. The results of the research showed that corporate reputation influences the profitability and market value of companies, in accordance with the Resource-Based View theory, since it is a strategic resource socially built from the perceptions of stakeholders. This study contributes to the literature on the relationship between reputation and financial performance, by measuring reputation through information from the news media. In terms of managerial contributions, the results indicate that companies should consider implementing media monitoring actions for reputation management, since reputation influences financial performance. Finally, suggestions for future research are discussed in the light of the methodological limitations presented.

Keywords: Corporate reputation. Financial performance. Media. Online news. Text mining.



## LISTA DE FIGURAS

Figura 1 – Estrutura do trabalho.....	25
Figura 2 – Aspectos do referencial teórico.....	27
Figura 3 – Representação tridimensional da reputação corporativa.....	37
Figura 4 – Etapas da metodologia .....	49
Figura 5 – Validação cruzada k-fold (exemplo para k = 10).....	77
Figura 6 – Modelo preditivo.....	78
Figura 7 – Diagrama de fluxo da revisão sistemática .....	86
Figura 8 – Distribuição de estudos primários por anos (de 2009 até 2018).....	87
Figura 9 – Áreas temáticas dos periódicos.....	88
Figura 10 – Países das empresas da amostra dos estudos .....	89
Figura 11 – Embasamento teórico dos estudos primários.....	89
Figura 12 – Abordagem para medição da reputação corporativa.....	90
Figura 13 – Indicadores de desempenho financeiro dos estudos primários.....	91
Figura 14 – Variáveis de controle analisadas nos estudos primários .....	91
Figura 15 – Metodologia de análise quantitativa .....	92
Figura 16 – Dados em painel – heterogeneidade entre trimestres.....	96
Figura 17 – Dados em painel – heterogeneidade entre empresas .....	97
Figura 18 – Setores das empresas da amostra .....	98
Figura 19 – Presença no índice Bovespa das empresas da amostra .....	99
Figura 20 – Correlação cruzada entre variáveis ROA e ROE e variáveis independentes.....	102
Figura 21 – Correlação cruzada entre variáveis MVA® e Q de Tobin (QTB) e variáveis independentes .....	103



## LISTA DE QUADROS

Quadro 1 – Perspectiva teórica dos estudos sobre reputação corporativa.....	33
Quadro 2 – Tipo de indicador de desempenho financeiro associado à reputação corporativa	41
Quadro 3 – Abordagens de medição convencionais da reputação corporativa.....	44
Quadro 4 – Bases de dados bibliográficas .....	51
Quadro 5 – Conceitos Mapeados .....	52
Quadro 6 – Strings de pesquisa .....	53
Quadro 7 – Estudos envolvendo Reputação Corporativa e Mídia.....	65
Quadro 8 – Descrição das variáveis da pesquisa.....	71
Quadro 9 – Testes para definição do melhor modelo para análise de dados em painel.....	74
Quadro 10 – Testes dos pressupostos da análise de regressão linear .....	75
Quadro 11 – Matriz de amarração metodológica .....	82



## LISTA DE TABELAS

Tabela 1 – Estudos identificados .....	84
Tabela 2 – Aderência da classificação automática de teor com a manual .....	93
Tabela 3 – Estatísticas descritivas .....	94
Tabela 4 – Coeficientes de correlação de Pearson entre as variáveis.....	100
Tabela 5 – Testes para determinação do modelo de dados em painel mais apropriado .....	104
Tabela 6 – Testes dos pressupostos da regressão linear (modelo com termo de interação)...	104
Tabela 7 – Testes dos pressupostos da regressão linear (modelo com variáveis independentes) .....	104
Tabela 8 – Análise de dados em painel do indicador de desempenho ROA .....	105
Tabela 9 – Análise de dados em painel do indicador de desempenho ROE .....	107
Tabela 10 – Análise de dados em painel do indicador de desempenho MVA® .....	108
Tabela 11 – Análise de dados em painel do indicador de desempenho Q de Tobin (QTB)...	109
Tabela 12 – Coeficiente de determinação (R <sup>2</sup> ) obtidos com validação cruzada (k = 20).....	114
Tabela 13 – Resultados do teste T para comparação de médias - Coeficiente de determinação R <sup>2</sup> . .....	114
Tabela 14 – Erro quadrático médio (RMSE) obtidos com validação cruzada (k = 20) .....	115
Tabela 15 – Resultados do teste T para comparação de médias – RMSE.....	115
Tabela 16 – Erro Absoluto Médio (MAE) obtido com validação cruzada (k = 20).....	116
Tabela 17 – Resultados do teste T para comparação de médias – MAE.....	116



## LISTA DE SIGLAS

AMAC	<i>America's Most Admired Companies</i>
B3	Brasil, Bolsa, Balcão
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
CAR	<i>Cumulative Abnormal Return</i>
CSV	<i>Comma Separated Values</i>
CVM	Comissão de Valores Mobiliários
DPO	<i>Days Payable Outstanding</i>
GBM	<i>Gradient Boosting Machines</i>
IPO	<i>Initial Public Offering</i>
MAE	<i>Mean Absolute Error</i>
MQO	Mínimos Quadrados Ordinários
MVA <sup>®</sup>	<i>Market Value Added</i>
POLS	<i>Pooled Ordinary Least Square</i>
RI	Relações com Investidores
RMSE	<i>Root Mean Square Error</i>
ROA	<i>Return On Assets</i>
ROE	<i>Return On Equity</i>
RSC	Responsabilidade Social Corporativa
SQL	<i>Structured Query Language</i>
SWOT	<i>Strengths, Weaknesses, Opportunities and Threats</i>
VBR	Visão Baseada em Recursos
VIF	<i>Variance Inflation Factor</i>
VRIN	<i>Valuable, Rare, Imperfectly Imitable, Non-Substitutable</i>
WMAC	<i>World's Most Admired Companies</i>





## SUMÁRIO

1	<b>INTRODUÇÃO</b> .....	17
1.1	Problematização/Problema de Pesquisa .....	20
1.2	Objetivos (Geral e Específicos) .....	21
1.2.1	Objetivo Geral .....	21
1.2.2	Objetivos específicos .....	21
1.3	Contribuições Esperadas (acadêmica e prática/gerencial e metodológica) .....	22
1.4	Impacto .....	23
1.5	Estrutura da dissertação .....	24
2	<b>REFERENCIAL TEÓRICO</b> .....	27
2.1	Perspectivas teóricas sobre reputação corporativa .....	27
2.1.1	Teoria Institucional .....	28
2.1.2	Teoria da Sinalização .....	29
2.2	O conceito da reputação corporativa .....	33
2.2.1	Ser conhecido – Visibilidade .....	37
2.2.2	Ser conhecido por algo .....	38
2.2.3	Favorabilidade generalizada - Favorabilidade .....	38
2.3	Reputação corporativa e desempenho financeiro .....	39
2.4	Abordagens de medição da reputação corporativa .....	41
2.4.1	Índice America’s Most Admired Companies (AMAC) .....	42
2.4.2	Índice Reputation Quotient .....	43
2.5	Reputação corporativa e mídia .....	44
3	<b>MÉTODO</b> .....	49
3.1	Revisão sistemática da literatura .....	49
3.1.1	Protocolo de revisão .....	51
3.1.2	Processo de busca .....	51
3.1.3	Critérios de inclusão .....	54
3.1.4	Critérios de exclusão .....	54
3.1.5	Processo de seleção dos estudos primários .....	55
3.1.6	Avaliação da qualidade dos estudos primários .....	55
3.1.7	Coleta de dados dos estudos primários selecionados .....	56
3.1.8	Análise de dados dos estudos primários selecionados .....	57
3.2	Definição das variáveis .....	58
3.2.1	Variáveis dependentes – Desempenho financeiro .....	58

3.2.1.1	ROA .....	58
3.2.1.2	ROE.....	59
3.2.1.3	Q de Tobin .....	59
3.2.1.4	MVA® .....	60
3.2.2	Variáveis independentes .....	61
3.2.2.1	Favorabilidade da mídia.....	61
3.2.2.2	Visibilidade da mídia .....	64
3.2.3	Variáveis de controle .....	66
3.2.3.1	Alavancagem financeira.....	66
3.2.3.2	Idade.....	66
3.2.3.3	Indústria - Setor.....	67
3.2.3.4	Tamanho da empresa .....	67
3.2.3.5	Governança corporativa .....	68
3.2.3.6	Inverse Mills Ratio – IMR .....	69
3.3	Modelo e análise de dados .....	73
3.4	Coleta de dados .....	79
4	<b>RESULTADOS</b> .....	83
4.1	Revisão sistemática da literatura.....	83
4.2	Estatísticas descritivas.....	93
4.3	Análise de regressão.....	103
4.4	Análise de regressão com algoritmo GBM e comparação com a regressão linear	113
5	<b>CONCLUSÕES</b> .....	117
	<b>REFERÊNCIAS</b> .....	121
	<b>APÊNDICE A - Estudos primários selecionados na revisão sistemática da literatura</b> ..	133
	<b>APÊNDICE B - Avaliação da qualidade dos estudos primários</b> .....	139
	<b>APÊNDICE C – Dados das empresas da amostra da pesquisa</b> .....	141
	<b>APÊNDICE D – Listagens dos programas implementados</b> .....	145

## 1 INTRODUÇÃO

A noção de reputação corporativa, e alguns conceitos relacionados a ela, como os de identidade, imagem e marca, tem recebido atenção considerável nos âmbitos da teoria e da prática da administração (Money, Saraeva, Garnelo-Gomez, Pain, & Hillenbrand, 2017). Diferentes pesquisas mencionam que, por se tratar de um ativo intangível estratégico que conduz a vantagem competitiva (Deephouse, 2000), esse tema vem sendo alvo de crescente interesse na área. Além disso, um volume cada vez maior de estudos argumenta que a reputação corporativa, definida como a percepção dos *stakeholders* sobre a capacidade de a empresa criar valor e superar seus competidores, tem valor estratégico para as corporações e tem recebido atenção considerável dos acadêmicos que estudam organizações (Rindova, Williamson, Petkova, & Sever, 2005).

Acrescente-se, ainda, que a reputação é um dos recursos intangíveis mais importantes, o qual diferencia empresas de suas competidoras, reduz a assimetria informacional e tem como consequência o desempenho superior, tanto financeiro como não financeiro. Dessa forma, esse elemento parece ser o fator fundamental para entender porque algumas organizações superam outras, o que muitos consideram o âmago do campo da gestão estratégica (Boyd, Bergh, & Ketchen, 2010). A importância dos recursos intangíveis para o sucesso de uma organização pode ser percebida pelo fato de que 75% do valor de mercado de uma empresa é determinado por esse tipo de ativo (Kaplan & Norton, 2006). Nesse sentido, atualmente *stakeholders* buscam presumir a qualidade de um produto ofertado a partir da qualidade do ofertante, considerando que uma empresa com “boa” reputação irá ofertar produtos de qualidade (Kaur & Singh, 2018). Apoiando essas observações, executivos do Reino Unido indicaram a reputação como o principal ativo intangível das empresas (Hall, 1993) que facilita a criação de valor (Deephouse, 2000) e que permite a elas melhorarem seus resultados financeiros, colocando-as em um ciclo virtuoso de reconhecimento e de criação de valor econômico (Orozco, Vargas, & Galindo-Dorado, 2018). Assim, levando em conta o que estabelece a teoria da Visão Baseada em Recursos (Barney, 1991), as empresas que possuem “boa” reputação têm um desempenho financeiro superior, uma vez que algumas das vantagens de possuir uma reputação positiva são associadas a um melhor desempenho financeiro. Entre essas vantagens se destacam: a atração e a fidelização de clientes, a redução dos custos, a contratação de funcionários mais qualificados e a criação de barreiras competitivas que aumentam a estabilidade das receitas.

Com respeito à importância desse tipo de desempenho, é notório que resultados financeiros satisfatórios são necessários para garantir a continuidade das empresas em uma

economia capitalista (Grinyer & Norburn, 1975) e são fundamentais, também, para a sobrevivência quando essas empresas estão sujeitas a um ambiente de negócios competitivo e incerto (Kyoony Yoo & Ah Park, 2007). Adicionalmente, o desempenho financeiro é necessário, ainda, para satisfazer os interesses dos acionistas e dos credores e para garantir empregos (Doyle, 1998).

A importância da reputação corporativa e sua relação com o desempenho financeiro tornam-se ainda mais evidentes diante de episódios recentes envolvendo grandes empresas. Após o escândalo dos testes de emissão de poluentes por carros a diesel, batizado de “*Dieseldgate*”, ocorrido em setembro de 2015, o preço da ação do grupo Volkswagen caiu cerca de 24% entre setembro de 2015 e agosto de 2016 (Bachmann, Ehrlich, & Ruzic, 2017) e as vendas nos EUA declinaram, em novembro de 2015, 24,7% em relação a novembro de 2014 (Welch & Hull, 2015). Somado a isso, em julho de 2016, a empresa teve que pagar 15,3 bilhões de dólares de indenização em um acordo fechado com a justiça federal americana (Flore, Kolaric, & Schiereck, 2017). Outro caso emblemático envolveu a empresa petrolífera britânica BP após a explosão da plataforma de produção de petróleo em águas profundas *Deepwater Horizon*, no Golfo do México, em 2010, que causou o maior vazamento acidental de petróleo da história e a morte de 11 pessoas (Walumbwa, Maidique, & Atamanik, 2014). Em decorrência desse acidente, a empresa sofreu elevadas perdas (Hargreaves, 2010), além de ter que pagar uma multa de 4 bilhões de dólares em 2012 e uma indenização adicional de 18,5 bilhões de dólares em 2015 (Flore et al., 2017).

No que se refere ao seu processo de construção, a reputação corporativa resulta, ou das percepções dos *stakeholders* a respeito da capacidade que a empresa tem de criar valor (Rindova et al., 2005), ou das impressões desses mesmos agentes relativas à disposição da empresa em se comportar de uma determinada maneira (Basdeo, Smith, Grimm, Rindova, & Derfus, 2006). Adicionalmente, a reputação atua como um sinal que ajuda a reduzir a assimetria informacional entre os *stakeholders* e as empresas, conforme preconizado pela teoria da Sinalização (Spence, 1973). No que se refere à diminuição dessa assimetria, a mídia atua como uma intermediária no processo de sinalização envolvendo empresas e *stakeholders*. Desse modo, a cobertura e as avaliações realizadas pelos meios de comunicação têm um papel relevante para a construção das percepções dos *stakeholders* a respeito das empresas e, por conseguinte, na construção da reputação corporativa (Van den Bogaerd & Aerts, 2015).

Com respeito à definição do constructo reputação corporativa, este estudo emprega a conceituação multidimensional proposta por Lange, Lee, e Dai (2011), a qual é composta pelas três dimensões a seguir: “ser conhecido”, ou seja, ter visibilidade, dimensão que representa a

importância, o destaque ou a proeminência da empresa na percepção coletiva; “ser conhecido por algo”, dimensão que indica as crenças sobre as características e/ou comportamentos diferenciados de uma organização; e a “favorabilidade generalizada”, dimensão relacionada às percepções ou julgamentos dos *stakeholders* sobre vários atributos organizacionais da empresa de forma agregada, isto é, a avaliação geral da organização como boa, atraente e apropriada.

Apesar de a teoria da Visão Baseada em Recursos indicar que a reputação produz efeitos positivos no desempenho financeiro, as descobertas de estudos empíricos prévios mostraram resultados mistos (Agnihotri, 2014; Kaur & Singh, 2018; J. Lee & Roh, 2012). A pesquisa existente vê a reputação como um ativo corporativo importante que, de fato, proporciona vantagem competitiva para as empresas (Hall, 1992), mas os resultados relacionados às consequências que uma “boa” reputação tem sobre o desempenho forneceram evidências mistas, de modo que alguns pesquisadores identificaram uma relação positiva, enquanto outros constataram uma relação ambígua entre esses conceitos (Sabate & Puente, 2003). Fernández Sánchez, Luna Sotorrío, e Baraibar Díez (2012) apontam que todas as pesquisas analisadas por eles que relataram uma relação positiva entre reputação e desempenho financeiro usaram o índice *America’s Most Admired Companies*, da revista *Fortune*, para medir a reputação. Entretanto, essa operacionalização é bastante criticada na literatura, devido à alta correlação com desempenho financeiro (forte efeito *halo*) e por capturar apenas as percepções de altos executivos e analistas do mercado financeiro e não de todos os *stakeholders* das empresas que compõem o *ranking*.

Esse conjunto contraditório de resultados motiva a busca de novas perspectivas metodológicas, diferentes daquelas bem estabelecidas e tradicionalmente utilizadas, com o objetivo de lançar luz sobre a controvérsia. Neste estudo a reputação é avaliada por meio da reputação da mídia, que é mensurada mediante técnicas de mineração de texto em notícias *online*, ao invés de rankings de reputação de revistas ou de escalas de medição, os quais foram utilizados em estudos anteriores sobre os impactos da reputação corporativa no desempenho financeiro (Cao, Myers, Myers, & Omer, 2015; Cardoso, De Luca, Lima, & Vasconcelos, 2013; Domingos & Moura, 2013; Fernández Sánchez et al., 2012; Fernández-Gámez, Gil-Corral, & Galán-Valdivieso, 2016; Filbeck, Gorman, & Zhao, 2013; Ghosh, 2017; J. Lee & Roh, 2012; Y. M. Lee & Hu, 2018; Lopes, De Luca, Gois, & de Vasconcelos, 2017; Orozco et al., 2018; Roberts & Dowling, 2002; Rose & Thomsen, 2004; Soleimani, Schneper, & Newbury, 2014; Stuebs & Sun, 2010; Tischer & Hildebrandt, 2014; Y. Wang & Berens, 2015). Cabe mencionar que os métodos convencionais para medir a reputação que foram empregados nessas pesquisas

possuem limitações que foram analisadas por Sarstedt, Wilczynski, e Melewar (2013) e que são discutidas no referencial teórico.

Índices de sentimento produzidos a partir de notícias da internet, de mídias sociais ou de mecanismos de busca na internet, similares ao utilizado nesta pesquisa, são capazes de, mediante a aplicação de algoritmos adequados, fornecer informações úteis e tempestivas a respeito do sentimento dos consumidores e, dessa forma, poderiam servir como potenciais indicadores de alerta precoce (Lanquillon, 2013) para serem usados pelos gestores das organizações.

Finalmente, na área de métodos quantitativos e informática do curso de administração, pesquisas relacionadas a novas técnicas de análise de dados que possam ser aplicadas a fontes de dados não estruturados de qualquer tamanho são cada vez mais necessárias e oportunas.

### 1.1 Problematização/Problema de Pesquisa

Além das pesquisas já mencionadas a respeito da associação entre reputação corporativa e desempenho financeiro, alguns estudos, que revisaram a literatura a respeito do tema, observaram resultados empíricos inconclusivos (Castilla-Polo, Gallardo-Vázquez, Sánchez-Hernández, & Ruiz-Rodríguez, 2018; J. Lee & Roh, 2012). Outros pesquisadores sugerem, também, que poderia existir sinergia entre as dimensões da reputação corporativa, o que produziria um efeito multiplicativo no desempenho financeiro, isto é, a integração das dimensões da reputação cria mais valor do que quando cada uma é considerada separadamente, de modo que o efeito conjunto das dimensões da reputação no desempenho financeiro é maior que a soma do efeito de cada uma das dimensões isoladamente (Boyd et al., 2010; Lange et al., 2011; Martinez, Russell, Maher, Brandon-Lai, & Ferris, 2017). Dessa forma, a despeito da grande quantidade de trabalhos empíricos que se debruçaram sobre essa relação, seguem existindo lacunas a serem preenchidas, no tocante aos efeitos sinérgicos das dimensões da reputação corporativa no desempenho financeiro e diante dos resultados divergentes dos estudos empíricos anteriores acerca da associação entre a reputação e o desempenho financeiro.

Assim, este trabalho explora a relação unidirecional entre reputação corporativa e desempenho financeiro com base nos conceitos da teoria da Visão Baseada em Recursos e da teoria da Sinalização. A reputação corporativa é modelada pelas dimensões denominadas visibilidade e favorabilidade, da conceituação de Lange et al., (2011), previamente descritas na contextualização, as quais são medidas por meio de notícias publicadas na mídia. Vale salientar que a literatura pesquisada não apresenta procedimentos para medição da dimensão “ser

conhecido por algo” por meio de técnicas de mineração de texto. Deste modo, considerando a viabilidade da pesquisa, não é escopo elaborar tais procedimentos. Isso posto, este estudo que visa responder a seguinte questão de pesquisa:

**Qual a associação entre reputação corporativa (i.e., favorabilidade e visibilidade) e desempenho financeiro?**

## 1.2 Objetivos (Geral e Específicos)

### 1.2.1 Objetivo Geral

O objetivo deste estudo é examinar o efeito da reputação corporativa (i.e., visibilidade e favorabilidade) sobre o desempenho financeiro das empresas de capital aberto listadas na B3 – Brasil, Bolsa, Balcão. Além disso, o trabalho investiga o efeito moderador da visibilidade da mídia sobre a associação entre a favorabilidade da mídia e o desempenho financeiro.

### 1.2.2 Objetivos específicos

Para a concretização do objetivo geral devem ser cumpridos os seguintes objetivos específicos:

- Identificar a variação da reputação corporativa (i.e., favorabilidade e visibilidade) das empresas listadas na B3 no período do primeiro trimestre de 2012 ao quarto trimestre de 2018.
- Identificar a variação do desempenho financeiro baseado em indicadores contábeis e do desempenho financeiro baseado em indicadores de mercado das empresas listadas na B3 no período do primeiro trimestre de 2012 ao quarto trimestre de 2018.
- Verificar como a reputação corporativa (i.e., favorabilidade e visibilidade) afeta o desempenho financeiro das empresas, o que é feito por meio de duas técnicas de regressão: linear com dados em painel e por intermédio de algoritmo de aprendizado de máquina supervisionado.
- Comparar o modelo preditivo, construído empregando técnicas de aprendizado de máquina, com um modelo construído utilizando técnicas de regressão linear.

### 1.3 Contribuições Esperadas (acadêmica e prática/gerencial e metodológica)

Este trabalho contribui para a pesquisa sobre a reputação corporativa de várias maneiras. Em primeiro lugar, ele traz novos elementos, em relação as pesquisas anteriores sobre reputação corporativa e suas consequências, mediante a medição desse conceito por meio da avaliação das empresas pela mídia (Agnihotri, 2014; Deephouse, 2000; Kioussis, Popescu, & Mitrook, 2007; Pollock & Rindova, 2003; Strauß & van der Meer, 2017; Van den Bogaerd & Aerts, 2014, 2015). Em segundo lugar, neste trabalho é explorada a visão multidimensional do constructo reputação, o que é feito por meio da representação bidimensional contendo as dimensões “ser conhecido” (visibilidade da mídia) e “favorabilidade generalizada” (favorabilidade da mídia). Desse modo, atende-se, ao chamado de Lange et al. (2011) para que sejam estudadas as implicações nas consequências da reputação de dimensões que não são tão óbvias e/ou mais facilmente mensuráveis: “*We suspect that the dimensions of generalized favorability and being known both have interesting but greatly understudied implications for organizational outcomes.*” (p. 174)<sup>1</sup>. Em terceiro lugar, tanto quanto é do meu conhecimento, este é o primeiro estudo no Brasil que analisa empiricamente os efeitos da reputação medida através da avaliação das empresas por meio de sua apresentação na mídia. Além disso, os poucos estudos explorando a relação entre a reputação da mídia e o desempenho foram realizados em países desenvolvidos (Agnihotri, 2014). Com respeito a esta contribuição, vale destacar também que há evidências na literatura de que a associação entre a reputação corporativa e suas consequências varia de país para país (Ali, Lynch, Melewar, & Jin, 2015), assim, a verificação dessa relação no Brasil mostra-se conveniente e relevante. Finalmente, este trabalho investiga o efeito moderador da visibilidade da mídia na relação entre a favorabilidade da mídia e o desempenho financeiro. Nesse sentido, esta análise endereça, ainda, outra solicitação de Lange et al. (2011) para que os pesquisadores e praticantes investiguem como uma mudança em uma dada dimensão pode afetar as consequências de outra dimensão. Assim, com respeito à dimensão “ser conhecido”, que é mapeada, neste estudo, na visibilidade da mídia, Lange et al. (2011) argumentam que ela “*considerably amplifies determinants of changes in the other dimensions since familiar firms may be especially salient and memorable targets*

---

<sup>1</sup> “Suspeitamos que as dimensões favorabilidade generalizada e “ser conhecido” tenham implicações interessantes, mas muito pouco estudadas, nos resultados organizacionais.” (Tradução do autor).



*for the kinds of disparaging and laudatory claims, and negative and positive observations”* (p. 168)<sup>2</sup>.

Com respeito aos aspectos metodológicos, na pesquisa bibliográfica realizada foram encontrados resultados mistos sobre as consequências da reputação corporativa no desempenho financeiro. Dessa forma, é esperado que as novas perspectivas metodológicas utilizadas neste estudo, as quais são diferentes daquelas tradicionalmente empregadas, possam lançar alguma luz sobre a controvérsia. Além disso, há, também, a expectativa de constatar que novas técnicas de análise de dados, tais como o aprendizado de máquina e a mineração de texto, podem ser empregadas em pesquisas da área das ciências sociais aplicadas como um complemento às técnicas tradicionalmente utilizadas nessa área. Por fim, vale destacar o uso, em pesquisa de ciência social aplicada, de fontes de dados não estruturados em formato textual, que tiveram um crescimento exponencial devido à ascensão da internet e das mídias sociais nos últimos anos.

#### 1.4 Impacto

A influência da reputação corporativa no desempenho financeiro, mais especificamente as consequências das dimensões de favorabilidade e visibilidade relacionadas com as informações publicadas na mídia, sugere que as organizações deveriam tanto adotar ações com a finalidade de monitorar as diversas mídias, como atuar com transparência e tempestividade em casos de informações negativas, condutas que têm como propósito de melhorar as percepções de seus stakeholders, as quais formam a reputação da empresa. Cabe salientar, a esse respeito, que existem evidências empíricas de que as notícias negativas têm efeito mais pronunciado na reputação do que as notícias positivas (Jonkman, Boukes, Vliegenthart, & Verhoeven, 2019). Quanto aos investidores do mercado de ações, esses agentes poderiam considerar a inclusão de informações de notícias da mídia em seus sistemas automáticos de negociação, com a finalidade de fazer previsões sobre o comportamento das ações incorporando esse aspecto informacional.

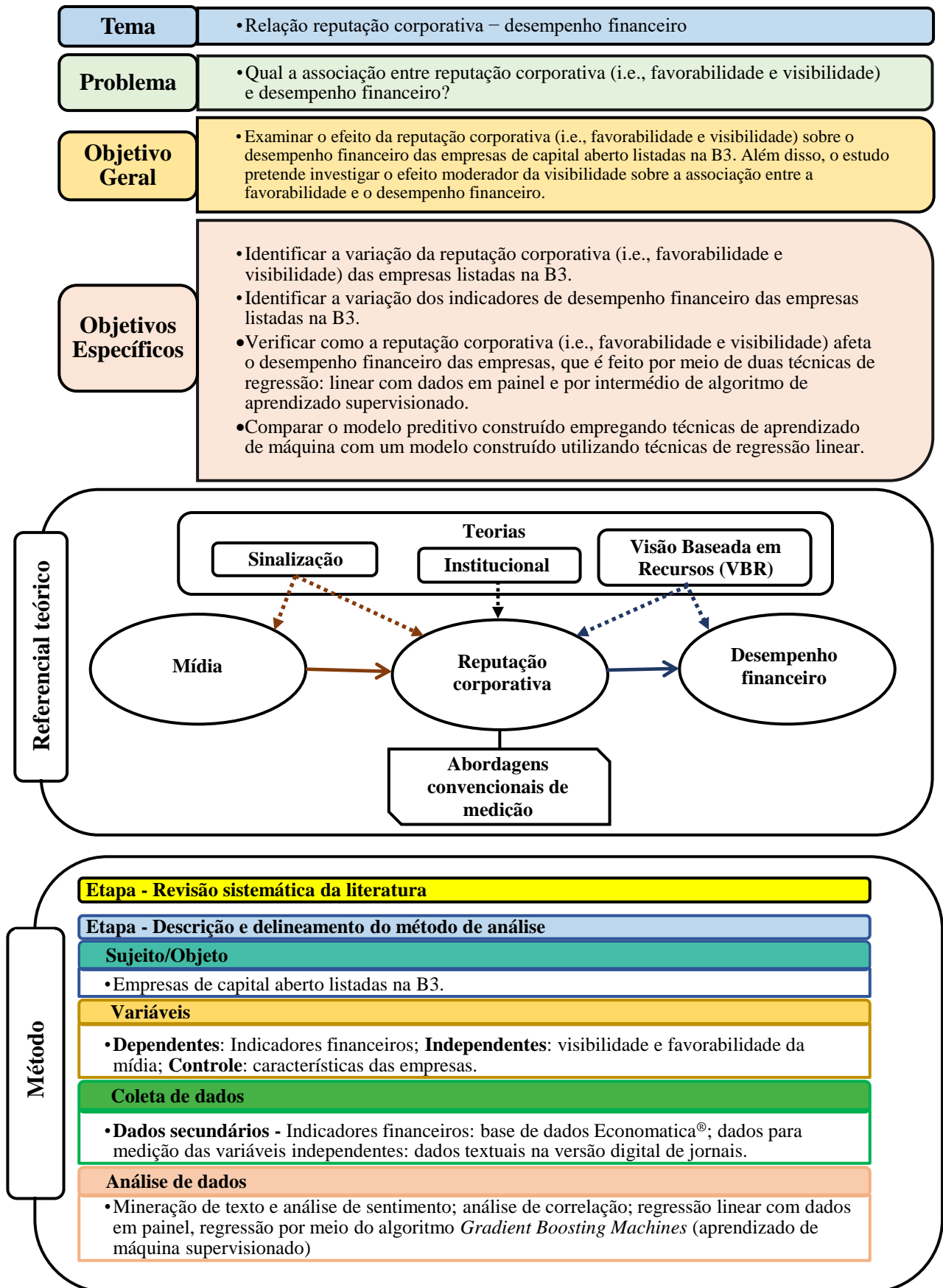
---

<sup>2</sup> “amplia consideravelmente os efeitos de mudanças nas outras dimensões, uma vez que as empresas conhecidas podem ser alvos notórios para comentários depreciativos e elogiosos, e observações negativas e positivas.” (Tradução do autor).

## 1.5 Estrutura da dissertação

A presente dissertação está dividida em cinco partes, sendo a primeira composta pela introdução e as outras quatro organizadas em diferentes capítulos. Dessa forma, enquanto a introdução traz um panorama do trabalho, as demais seções se compõem da seguinte maneira: o capítulo 2 apresenta o referencial teórico que embasa a pesquisa; o capítulo 3 expõe o método adotado para conduzir o estudo, detalhando passo a passo todos os procedimentos realizados; a análise dos resultados é feita no capítulo 4; finalmente, no capítulo 5, são apresentadas as considerações finais. A estrutura geral do trabalho pode ser vista na Figura 1.

Figura 1 – Estrutura do trabalho



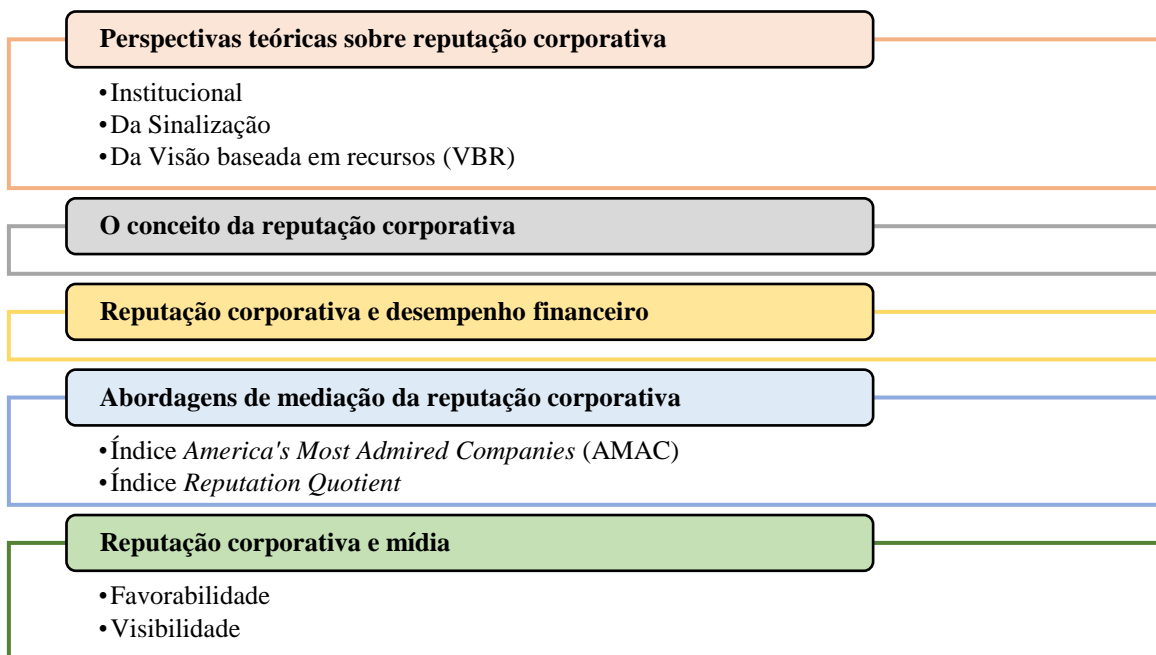
Fonte: Elaborado pelo autor.



## 2 REFERENCIAL TEÓRICO

O presente capítulo apresenta, de forma concisa, uma revisão de literatura. Assim, ao longo desta seção, serão tratados, respectivamente, os seguintes pontos: inicialmente, as principais perspectivas teóricas no estudo da reputação corporativa; em seguida, o conceito da reputação corporativa; posteriormente, as abordagens de medição da reputação corporativa mais proeminentes na literatura; e, finalmente, a conexão entre mídia e reputação corporativa, além do desenvolvimento das hipóteses do trabalho. A Figura 2 destaca os aspectos que são abordados neste referencial teórico.

Figura 2 – Aspectos do referencial teórico



Fonte: Elaborado pelo autor.

### 2.1 Perspectivas teóricas sobre reputação corporativa

Como consequência de pesquisas interdisciplinares realizadas sob óticas distintas, que vão desde a psicologia até a administração, a reputação corporativa possui diferentes conceituações (Tischer & Hildebrandt, 2014). Em sua revisão sistemática de literatura, que selecionou artigos publicados entre 1980 e 2007, Walker (2010) identificou que as três perspectivas teóricas mais utilizadas para embasar os estudos de reputação corporativa em sua amostra eram: a Institucional, a da Sinalização e a da Visão Baseada em Recursos (VBR), resultados que estão em consonância com o estudo teórico de Bergh, Ketchen, Boyd, e Berg

(2010), o qual indica que a relação reputação–desempenho pode ser explicada pelas teorias da Sinalização e da VBR. Assim, as três teorias indicadas por Walker (2010) são explicadas a seguir.

### 2.1.1 Teoria Institucional

Em uma revisão de literatura, Money et al. (2017) identificaram vários estudos sobre reputação embasados na teoria Institucional. A teoria Institucional é empregada para explorar como a reputação é construída (Walker, 2010). Além disso, essa teoria possibilita a identificação dos fatores no ambiente organizacional que permitem a construção da reputação corporativa (Ghosh, 2017) e é utilizada para entender tanto a relação entre a legitimidade e a reputação quanto a diferença entre esses conceitos, os quais possuem várias similaridades (Deephouse & Carter, 2005). A legitimidade consiste no grau de apoio cultural recebido por uma empresa (Meyer & Scott, 1992) e, para obtê-la, as corporações devem atuar dentro de seus contextos institucionais, já que os esforços para construir uma reputação sem considerar o contexto institucional podem ser infrutíferos (Walker, 2010). A reputação, por seu turno, é uma consequência do processo de legitimação, de modo que tanto a reputação como a legitimidade são aspectos complementares da criação da identidade organizacional (H. Rao, 1994). De acordo com Walker (2010), a identidade organizacional é um dos componentes da reputação corporativa que reflete a avaliação dos *stakeholders* internos. Com relação à legitimidade, existem dois tipos: a sociopolítica (regulatória e normativa), que consiste no endosso à organização por órgãos governamentais e autoridades, e a cognitiva, que implica que a empresa é apropriada dentro de um conjunto compartilhado de normas e valores (Scott & Søren, 1995). Cabe destacar que Rindova et al. (2005) empregam a teoria Institucional para explicar porque a proeminência (visibilidade ou reconhecimento generalizado) é uma das dimensões da reputação corporativa.

Além do que é evidenciado nos trabalhos mencionados, a teoria Institucional é empregada, ainda, nos estudos sobre a reputação corporativa, para justificar por qual motivo o país no qual foi realizada determinada pesquisa é um potencial moderador da relação entre a reputação corporativa e seus antecedentes e consequências, uma vez que diferentes países possuem regras e regulações distintas (Ali et al., 2015). Corroborando essa visão, Deephouse, Newburry e Soleimani (2016) argumentam que pouco se sabe sobre o quanto e por que as reputações diferem entre países e, diante dessa questão, sugerem que a teoria Institucional deve ser utilizada para embasar estudos transnacionais sobre reputação corporativa.

Entretanto, o uso da teoria Institucional como perspectiva teórica no estudo da reputação é também motivo de críticas, pois as pressões do ambiente institucional (isomorfismo institucional) levariam as empresas, no decorrer do tempo, a tornarem-se cada vez mais semelhantes, enquanto que a pesquisa sobre reputação corporativa tem como foco o exame da diferenciação organizacional (Bergh et al., 2010).

### 2.1.2 Teoria da Sinalização

A teoria da Sinalização surgiu no campo da economia como uma proposta para contornar o problema da assimetria informacional. Spence (1973), em seu trabalho seminal, desenvolveu esta teoria para mostrar que o nível educacional dos candidatos a um emprego sinaliza sua competência ao empregador que possui informação incompleta sobre eles.

O processo de sinalização envolve dois atores: o sinalizador e o receptor. O sinalizador determina se envia ou não as informações (sinais) e o receptor decide como interpretar e responder ao sinal recebido (OuYang, Xu, Wei, & Liu, 2017). No contexto organizacional, as empresas (sinalizadoras) enviam sinais aos *stakeholders* (receptores) com a finalidade de reduzir a assimetria informacional, o que é feito porque corporações possuem informações que *stakeholders* externos podem considerar úteis (Alon & Vidovic, 2015).

Spence (1973) definiu um sinal como um indicador observável de uma qualidade. Para o autor, esse indicador é, pelo menos, parcialmente alterável e, portanto, sujeito a manipulação, além disso, tem um custo ou dificuldade para ser obtido inversamente proporcional ao nível de qualidade sinalizado. Sinais transmitem informações a respeito de características de agentes econômicos, tais como empresas, consumidores e candidatos a um emprego. A eficácia da sinalização é parcialmente determinada pela observabilidade do sinal, assim, sinais menos observáveis tornam a comunicação menos eficaz, um problema que, no entanto, pode ser atenuado por intermediários (OuYang et al., 2017), como a mídia, um dos mais relevantes nesse âmbito (Deephouse, 2000). Dessa forma, a mídia dá uma importante contribuição para reduzir a incerteza dos *stakeholders* em relação às qualidades e aos problemas de uma empresa (OuYang et al., 2017). A participação da mídia no processo de construção da reputação será tratada com mais detalhes em tópico específico a seguir.

A confiabilidade dos sinais define seu valor informacional para o receptor e está associada com duas características: a correlação que o sinal (informação pública) tem com a qualidade não observável do sinalizador (informação privada) e a honestidade do sinalizador (Connelly, Certo, Ireland, & Reutzel, 2010).

Levando isso em conta, o processo de sinalização pode ser representado como uma linha cronológica: o sinalizador, ator interno à organização, tem informações positivas e negativas sobre a empresa, as quais *stakeholders* externos poderiam considerar úteis. Os sinalizadores, então, comunicam a informação aos receptores (*stakeholders* externos), os quais operam com informação incompleta sobre a organização (OuYang et al., 2017). A sinalização deve ter um efeito estratégico, isto é, o sinalizador deve se beneficiar de alguma ação tomada pelo receptor, a qual não ocorreria caso o sinal não fosse emitido (Connelly et al., 2010). Em meio a essas ações, o receptor também pode enviar um feedback ao sinalizador, na forma de contraindicações, pois a assimetria de informações é bidirecional: tanto os receptores desejam informações dos sinalizadores como os sinalizadores desejam informações dos receptores. A partir do *feedback*, os sinalizadores podem, então, ajustar sinais futuros, de forma a obter uma sinalização mais eficaz (Connelly et al., 2010).

Em sua revisão sistemática da literatura, Walker (2010) argumenta que a teoria da Sinalização explica como empresas constroem, mantêm e defendem sua reputação mediante a formação de imagens organizacionais por meio do envio de sinais estratégicos aos *stakeholders* externos. Nesse sentido, a formação da reputação pode ser amplamente entendida como um processo de sinalização, no qual as empresas fazem escolhas estratégicas sobre que sinais enviar aos receptores (*stakeholders* externos), os quais podem usar a informação recebida para formar sua impressão sobre a empresa (Basdeo et al., 2006). Esses sinais moldam a reputação da empresa, pois permitem que os *stakeholders* externos formem impressões e opiniões sobre a capacidade de a empresa criar valor (Rindova & Fombrun, 1999). Assim, a reputação reduz a incerteza, através da sinalização da qualidade de produtos (Rindova et al., 2005), e é, também, um sinal poderoso que se desenvolve como resultado de numerosos sinais de suporte relacionados à empresa e a seus competidores (Basdeo et al., 2006).

A teoria da Sinalização foi empregada, também, como perspectiva teórica em estudos empíricos que examinam a reputação corporativa. Ela é aplicada nesses estudos para analisar aspectos como: o impacto das atividades de uma empresa em sua reputação (Basdeo et al. 2006); o papel moderador de variáveis na relação entre a reputação corporativa e o comportamento de apoio das pessoas (Newbury, 2010); o efeito moderador da confiança pública na relação entre os constructos referentes ao desempenho social, à reputação corporativa e ao desempenho financeiro (Gardberg, Symeou, & Zyglidopoulos, 2015); o impacto da filantropia na reputação corporativa (Gardberg, Zyglidopoulos, Symeou, & Schepers, 2017); investigar os efeitos da reputação baseada no consumidor no risco percebido pelos clientes e na confiança dos clientes (Walsh, Schaarschmidt, & Ivens, 2017); a influência



do desempenho sustentável e a da certificação externa da sustentabilidade na reputação corporativa (Alon & Vidovic, 2015); e a influência da reputação da mídia na redução da assimetria informacional e na reação dos investidores em crises (OuYang et al., 2017).

### 2.1.3. Teoria da Visão Baseada em Recursos (VBR)

A VBR é um modelo que assume que os recursos estratégicos são a chave para um desempenho superior de uma empresa e, além disso, examina a ligação entre as características internas da organização e seu desempenho. Nesse sentido, organizações obtêm vantagens competitivas por intermédio da implementação de estratégias para uso de recursos internos à empresa. Tais recursos devem ser, valiosos, raros, difíceis de imitar e não substituíveis (Barney, 1991, 1995), esse conjunto de atributos é conhecido como *VRIN framework*. Assim, os recursos são valiosos se eles ajudam a organização a aumentar o valor ofertado aos clientes, o que pode ser feito por meio da diferenciação dos produtos e/ou da redução de custos; são raros se somente podem ser obtidos por uma ou algumas empresas; e são difíceis de imitar ou substituir caso uma organização rival não consiga replicá-los perfeitamente ou utilize outros recursos para obter os mesmos resultados (Barney, 1991).

Recursos raros e valiosos não garantem uma vantagem competitiva de longo prazo, uma vez que os competidores podem imitá-los. Para obter esse tipo de vantagem competitiva, os recursos deveriam, ou ser difíceis de imitar, ou ser imperfeitamente imitáveis. De acordo com a VBR, as razões que tornam recursos difíceis de imitar ou imperfeitamente imitáveis são: condições históricas únicas, ambiguidade causal e complexidade social (Barney, 1991). A reputação corporativa é um fenômeno social complexo (Hall, 1992), portanto, a habilidade de outras organizações em imitá-la é muito reduzida, além disso, é possível especificar como a reputação adiciona valor a uma empresa, de modo que não existe ambiguidade causal a qual ocorre quando a ligação entre o recurso e a vantagem competitiva não é entendida ou é entendida parcialmente (Barney, 1991).

A propriedade de recursos valiosos, raros e difíceis de imitar confere apenas um potencial de vantagem competitiva para a empresa, a organização deve possuir estratégias para tirar vantagem e capturar valor desses recursos a fim de obter vantagem competitiva sustentável (Barney, 1995). Nesse sentido, com base na análise de pontos fortes, pontos fracos, oportunidades e ameaças (análise SWOT), Barney (1995) sugere, ainda, que empresas que tiram proveito de suas forças internas na exploração de oportunidades e na neutralização de

ameaças, enquanto evitam fraquezas internas, têm maior probabilidade para obter vantagens competitivas do que outros tipos de empresas.

Os recursos empregados pelas empresas são classificados em dois tipos: tangíveis, que são concretos e têm existência física, e intangíveis, ou seja, aqueles que não têm existência física, os quais incluem bens como patentes, marcas registradas, direitos autorais, cultura organizacional, etc. (Hall, 1992). A reputação corporativa é um exemplo de recurso intangível das organizações. Uma pesquisa realizada por Hall (1993) mostra que a reputação é um ativo intangível chave para muitas empresas, o qual, além de contribuir substancialmente para o sucesso do negócio, pode servir como uma “reserva de boa vontade”, protegendo a empresa durante crises (Jones, Jones, & Little, 2000). A reputação é difícil de imitar devido a sua construção social complexa, que envolve as percepções dos *stakeholders* (Bergh et al., 2010); não é substituível, porque não é um ativo que as organizações possam comprar ou vender, mais sim cultivar (J. Lee & Roh, 2012). Um recurso tão raro, difícil de imitar e não substituível que permite que a empresa que o possui obtenha vantagem competitiva a qual conduz a desempenho financeiro superior (J. Lee & Roh, 2012) e que explica o melhor desempenho financeiro relativo aos competidores (López & Iglesias, 2010).

Como a teoria da Sinalização, a VBR também foi utilizada como perspectiva teórica em estudos empíricos que analisaram a reputação corporativa, o que é feito, em geral, para explicar suas consequências. A partir desses trabalhos, pesquisadores concluíram que a reputação é um recurso estratégico que conduz a vantagem competitiva sustentável (Walker, 2010). A mesma teoria serviu como embasamento, ainda, em estudos que analisaram os efeitos da reputação sobre vários indicadores de desempenho financeiro e não financeiro, tais como: lucratividade (ROA) (Ang & Wight, 2009; Deephouse, 2000; Roberts & Dowling, 2002), valor de mercado da empresa (Fernández-Gámez et al., 2016), medidas de desempenho financeiro contábil e de mercado (Bandeira, Góis, De Luca, & Vasconcelos, 2015; Koh, Lee, & Boo, 2009; J. Lee & Roh, 2012; Zhang & Rezaee, 2009), desempenho anterior e futuro (López & Iglesias, 2010), retorno cumulativo anormal – CAR (Pfarrer, Pollock, & Rindova, 2010; Tischer & Hildebrandt, 2014), desempenho social (Ghosh, 2017), proeminência no mercado (Boyd et al., 2010), retornos no mercado de ações (Raithel & Schwaiger, 2015; Soleimani et al., 2014), *Market-to-Book ratio* (Eberl & Schwaiger, 2005; Raithel, Wilczynski, Schloderer, & Schwaiger, 2010; Rose & Thomsen, 2004), prêmio no preço (Rindova et al., 2005), medidas de desempenho baseadas no mercado (Góis, 2015) e desempenho social e desempenho econômico (Nardi, 2015).

Ao final da descrição das três perspectivas teóricas mais empregadas nos estudos da reputação corporativa, de acordo com a revisão sistemática de literatura realizada por Walker (2010), nota-se que a teoria Institucional tem sido usada para examinar quais fatores, dentro do contexto institucional, levam à construção da reputação, isto é, tem sido usada para explicar seus antecedentes. Da mesma forma, a teoria da sinalização também tem sido utilizada, nas pesquisas, para explicar os antecedentes da reputação, uma vez que essa teoria examina tanto os sinais estratégicos enviados pelas empresas como de que maneira os *stakeholders* os interpretam para construir imagens e impressões a respeito das corporações, as quais são responsáveis pela construção da reputação. Quanto à VBR, ela tem sido aplicada, por outro lado, para explicar os efeitos ou consequências da reputação corporativa. O Quadro 1 apresenta um conjunto de pesquisas que, ao examinar a reputação corporativa, seus antecedentes e resultados, empregam as teorias abordadas nessa seção como perspectiva teórica.

Quadro 1 – Perspectiva teórica dos estudos sobre reputação corporativa

Teoria	Aplicação da teoria nas pesquisas	Artigos
Institucional	Examinar quais fatores no contexto institucional conduzem à construção da reputação corporativa.	Ali et al. (2015); Bergh et al. (2010); Deephouse e Carter (2005); Deephouse et al. (2016); Ghosh (2017); Rao, (1994); Rindova et al. (2005).
Sinalização	Explicar antecedentes da reputação corporativa e sua manutenção.	Alon e Vidovic (2015); Basdeo et al. (2006); Gardberg et al. (2015); Gardberg et al. (2017); Newburry (2010); OuYang et al. (2017); Walsh et al. (2017).
VBR	Explicar as consequências da reputação corporativa.	Ang e Wight (2009); Bandeira et al. (2015); Deephouse (2000); Eberl e Schwaiger (2005); Fernández-Gámez et al. (2016); Góis (2015); Koh et al. (2009); López e Iglesias (2010); J. Lee e Roh (2012); Nardi (2015); Pfarrer et al. (2010); Raithel e Schwaiger (2015); Rindova et al. (2005); Schloderer et al. (2010); Roberts e Dowling (2002); Rose e Thomsen (2004); Soleimani et al. (2014); Tischer e Hildebrandt (2014); Zhang e Rezaee (2009).

Fonte: Elaborado pelo autor.

A seção seguinte apresenta detalhadamente o constructo reputação corporativa.

## 2.2 O conceito da reputação corporativa

A reputação corporativa tornou-se uma preocupação relevante para gestores e *stakeholders*, o que conduziu à proliferação de rankings de reputação em todo o mundo (Fombrun & Gardberg, 2000). Em resposta a uma pesquisa realizada no Reino Unido, executivos indicaram a reputação corporativa como o principal ativo intangível que contribui para o sucesso das empresas, de modo que, para eles esse recurso deveria ser protegido e

desenvolvido considerando sua fundamental contribuição para obter vantagem competitiva sustentável (Hall, 1993). A reputação é resultado de anos de competência superior comprovada e, além disso, é um ativo frágil, que leva tempo para formar, não pode ser comprado em mercado de fatores e pode ser arruinado facilmente (Hall, 1992). Como mencionado na seção em que foram expostas as perspectivas teóricas, a VBR indica que ativos intangíveis como a reputação corporativa produzem desempenho superior quando são raros, valiosos, difíceis de imitar e não substituíveis (Barney, 1991).

O conceito reputação corporativa é intuitivo e simples no seu uso habitual, porém é extremamente complicado quando utilizado e analisado em pesquisas do campo da administração (Lange et al., 2011). Na literatura existe um entendimento geral a respeito da importância da reputação corporativa (Einwiller, Carroll, & Korn, 2010). Entretanto, a definição e a medição do constructo são apontadas como problemas fundamentais da área (Dowling, 2016; Lange et al., 2011; Walker, 2010). Os problemas surgiram na década de 2001 a 2010 quando houve um aumento expressivo da atenção dos pesquisadores em administração ao tema (Lange et al., 2011). A definição de um constructo é fundamental para sua medição (Dowling, 2016; Wartick, 2002) e para estabelecer a maneira pela qual as pesquisas serão realizadas, isto é, definir quais as questões de pesquisa, quais os métodos e como os resultados são analisados (Dowling, 2016). Dessa maneira, constructos mal definidos geram desconfiças nos resultados empíricos, dificultando o desenvolvimento teórico de conhecimento (Dowling, 2016).

Uma barreira apontada na literatura para a criação de uma conceituação única para a reputação corporativa reside na confusão existente no uso dos conceitos de identidade, imagem e reputação (Barnett, Jermier, & Lafferty, 2006), apesar de distintos, esses três conceitos ainda são frequentemente usados como sinônimos (Wartick, 2002). Chun (2005), por sua vez, vê identidade, imagem e reputação como rótulos de conceitos diferentes, ainda que parceiros.

Barnett et al. (2006) visualizam uma organização como: “a collection of material and behavioral symbols and corresponding systems of beliefs and values and basic, underlying assumptions.” (p. 34)<sup>3</sup>. Com base nessa definição, a identidade é conceituada como uma coleção de símbolos que representam a organização (Barnett et al., 2006), mas o termo também é definido como as características da organização que são centrais e duradoras na visão de seus

---

<sup>3</sup> “uma coleção de símbolos materiais e comportamentais, e correspondentes sistemas de crenças e valores, além de suposições básicas subjacentes.” (Tradução do autor).

funcionários (Fombrun, 1996). A identidade é baseada nas percepções reais de *stakeholders* internos as quais podem ser positivas ou negativas (Walker, 2010).

Quanto ao conceito de imagem, ele constitui-se, por outro lado, em uma projeção da “figura” interna para os *stakeholders* externos, ou seja, naquilo que a empresa deseja que eles pensem/saibam sobre ela (Walker, 2010) ou, ainda, constitui-se na imagem mental da empresa percebida por seus *stakeholders* (Gray & Balmer, 1998). A formação da imagem é feita, desse modo, através da comunicação corporativa, que consiste em mensagens que uma empresa transmite para seus *stakeholders* (Gray & Balmer, 1998). Assim, a imagem pode ser moldada, porém não controlada pela empresa, pois outros atores, tais como a mídia e o governo, influenciam a percepção dos *stakeholders* externos a respeito da empresa (Barnett et al., 2006). Walker (2010) indica que a imagem é a percepção positiva desejada que emana do interior da empresa, a qual é avaliada por *stakeholders* externos. O tempo para a formação é uma importante distinção entre a imagem e a reputação (Walker, 2010). Assim, a reputação leva tempo para ser construída e tem um significado histórico acumulado. Em contrapartida, a imagem é formada mais rapidamente. Outra distinção é a resiliência, com respeito a isso, a reputação é, em geral, estável e duradora, enquanto a imagem é mais volátil (Walker, 2010).

Os constructos identidade e imagem são considerados componentes da reputação na perspectiva integrativa de Fombrun e van Riel (1997). Desse ponto de vista, a identidade é composta pelas percepções que empregados e gerentes têm sobre sua empresa e a imagem, por outro lado, é a percepção de *stakeholders* externos a respeito dela (Barnett et al., 2006). De acordo com Fombrun & Van Riel (1997), a reputação corporativa é definida como: “a collective representation of a firm’s past actions and results that describes the firm’s ability to deliver valued outcomes to multiple *stakeholders*. It gauges a firm’s relative standing both internally with employees and externally with its *stakeholders*, in both its competitive and institutional environment.” (p. 10)<sup>4</sup>.

Em relação à ideia de reputação corporativa, na literatura é constatada a existência de uma profusão de definições. Em sua revisão sistemática de literatura, Walker (2010) analisa 13 definições de reputação corporativa, enquanto, Dowling (2016), em um trabalho semelhante, elenca 50 definições. Dentre elas, a definição proposta por Charles Fombrun é a mais citada na literatura sobre reputação corporativa (Dowling, 2016; Walker, 2010). Para ele, a reputação é

---

<sup>4</sup> “uma representação coletiva das ações e resultados passados de uma empresa que descrevem a sua habilidade de entregar valor para vários *stakeholders*. Ela mede a posição relativa da empresa tanto internamente, entre seus funcionários, como externamente entre seus *stakeholders*, o que faz no ambiente competitivo e, também, no institucional.” (Tradução do autor).

concebida como “a perceptual representation of a company’s past actions and future prospects that describes the firm’s overall appeal to all of its key constituents when compared with other leading rivals” (Fombrun, 1996, p. 72)<sup>5</sup>. De acordo com Walker (2010), a definição de Fombrun (1996) inclui três características: medida subjetiva baseada em percepções, agregada de todos os *stakeholders* e inerentemente comparativa. Adicionalmente, Walker (2010) identificou mais duas características nos principais artigos da literatura acadêmica por ele examinados: as ideias de que reputação pode ser positiva ou negativa e de que ela é estável e perene. Considerando essas cinco características, Walker (2010) propôs uma nova definição da reputação corporativa, qual seja: “A relatively stable, issue specific aggregate perceptual representation of a company’s past actions and future prospects compared against some standard.” (p. 370)<sup>6</sup>.

Finalmente, Lange et al. (2011) propuseram que a reputação consiste na familiaridade com a organização, nas crenças sobre o que esperar do futuro e nas impressões sobre quão favorável é a organização. Esta conceituação é composta de três dimensões:

- “ser conhecido” – que se refere à visibilidade, familiaridade, saliência ou proeminência da organização na percepção coletiva;
- “ser conhecido por algo” – relativa às crenças sobre a previsibilidade dos resultados e sobre características e/ou comportamentos relevantes para os *stakeholders*;
- “favorabilidade generalizada” – relacionada a uma avaliação geral, que diz respeito a ser apropriado, atraente e bom.

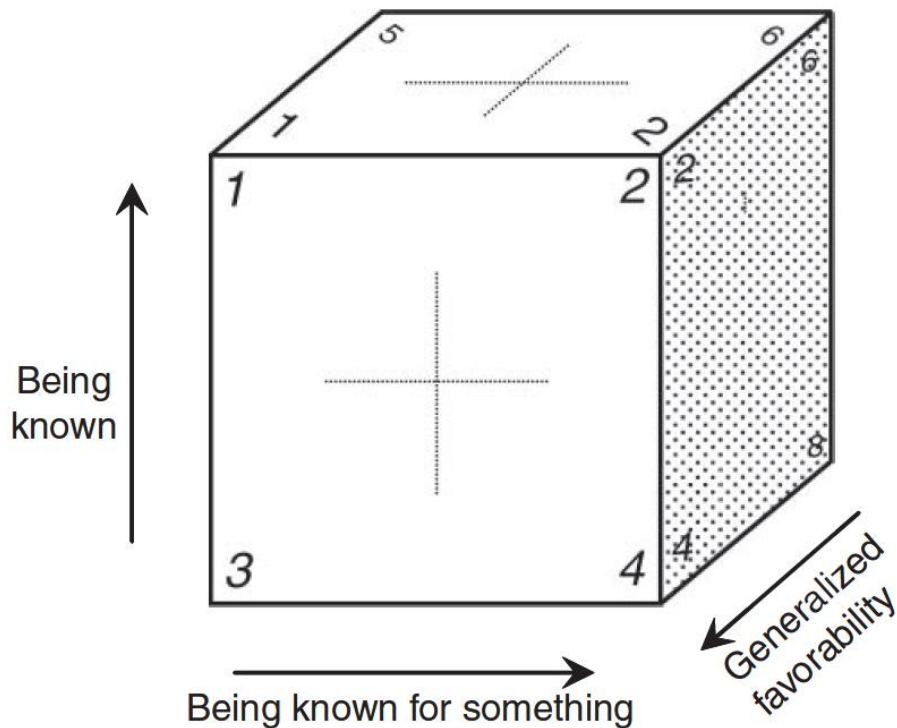
Este estudo baseia-se na conceituação de reputação corporativa de Lange et al. (2011), que está ilustrada graficamente, por meio de uma representação tridimensional, na Figura 3. A seguir, cada dimensão é detalhada.

---

<sup>5</sup> “uma representação perceptual das ações passadas e das perspectivas futuras de uma empresa que descrevem a atratividade da empresa para todos os seus públicos-chaves em comparação com os principais concorrentes.” (Tradução do autor).

<sup>6</sup> “uma representação perceptual relativamente estável, específica a uma característica, das ações passadas e das perspectivas futuras de uma empresa comparada com algum padrão” (Tradução do autor).

Figura 3 – Representação tridimensional da reputação corporativa



Nota: Os extremos das três dimensões são numerados como regiões. A região 7 está escondida.  
 Fonte: Lange et al., 2011.

### 2.2.1 Ser conhecido – Visibilidade

A dimensão “ser conhecido” é evidente em vários estudos no campo da administração e representa a extensão do reconhecimento de uma organização (Lange et al., 2011). Barnett et al. (2006) referem-se a esta dimensão como o reconhecimento geral da organização dado pelos *stakeholders* e pelos observadores, sem que seja feita nenhuma avaliação sobre ela. Outra designação dessa dimensão é a ideia de proeminência, que consiste no nível de reconhecimento da organização entre os *stakeholders* (Rindova et al., 2005). Entretanto, alguns autores não acham que a proeminência seja uma dimensão do constructo reputação corporativa: Boyd et al. (2010) consideram-na como outro conceito que media a relação entre reputação corporativa e o desempenho, e Ali et al., (2015) entendem-na como um antecedente da reputação corporativa.

Lange et al. (2011), por sua vez, retratam a dimensão “ser conhecido” através das respostas às seguintes questões: “(a) What are the shared perceptual representations of the focal

organization within the social set of perceivers? and (b) Within that social set, how commonly held are those shared understandings?” (p. 164)<sup>7</sup>.

### 2.2.2 Ser conhecido por algo

A dimensão “ser conhecido por algo” envolve percepções de que a empresa tem uma característica de interesse ou de valor para um observador (Lange et al., 2011). Esta dimensão está presente no trabalho de Rindova et al. (2005) e é designada como uma qualidade percebida, a qual se constitui no grau em que os *stakeholders* avaliam alguma característica da organização. Em contraste, com a dimensão “ser conhecido”, aqui o julgamento dos *stakeholders* é um fator fundamental.

Ademais, Lange et al. (2011) retratam a dimensão “ser conhecido por algo” a partir das respostas às seguintes questões: “(a) What are the strongly held expectations for an organization’s consequences and tangible outputs that exist within the set of perceivers? and (b) Within that social set, how strongly and commonly held are predictions that the focal organization will or will not meet those expectations?” (p. 165)<sup>8</sup>.

### 2.2.3 Favorabilidade generalizada - Favorabilidade

A dimensão “favorabilidade generalizada” consiste em diferentes aspectos: na avaliação geral sobre quão boa, atraente e apropriada é a organização (Lange et al., 2011), na avaliação geral da organização pela mídia (Deephouse, 2000), no nível de legitimidade da organização (Deephouse & Carter, 2005) e na extensão com a qual *stakeholders* externos consideram a organização como adequada e satisfatória (Roberts & Dowling, 2002). Esta dimensão resulta de avaliações da organização feitas de modo agregado, ao invés da avaliação de alguma capacidade específica em atender expectativas de observadores (Lange et al., 2011). Lange et al. (2011) descrevem a dimensão favorabilidade generalizada por meio das respostas às seguintes questões: “(a) What evaluative, generalized opinions are held about the organization

---

<sup>7</sup> “a) Quais são as representações perceptuais da organização em foco que são compartilhadas pelo conjunto de observadores? e b) Dentro do conjunto social, quão comumente essas percepções são compartilhadas?” (Tradução do autor).

<sup>8</sup> “a) Quais são as expectativas fortemente mantidas pelo conjunto de observadores para as consequências e resultados tangíveis de uma organização? e b) Dentro desse conjunto de observadores, quão fortemente e comumente são as previsões que a organização em foco atenderá ou não às expectativas?” (Tradução do autor).



within the set of perceivers? and (b) Where those options are held within that social set, how strongly and consistently are they held?” (p. 165)<sup>9</sup>.

### 2.3 Reputação corporativa e desempenho financeiro

A suposição de que a reputação corporativa gera efeitos positivos para as empresas está de acordo com as teorias da administração, como a VBR (Bergh et al., 2010; Roberts & Dowling, 2002). A considerável atenção dada a esse tema é demonstrada pelos vários estudos empíricos que tematizam as consequências da reputação corporativa, publicados em periódicos de alto impacto durante a década de 2001 a 2010 (Lange et al., 2011). Considerando o caso específico do desempenho financeiro, algumas pesquisas mostram que esse aspecto pode ser tanto um antecedente da reputação (desempenho financeiro anterior) como uma consequência dela (desempenho financeiro futuro) (Ali et al., 2015; López & Iglesias, 2010). Assim, a relação entre a reputação corporativa e o desempenho financeiro pode operar em ambas as direções, ou seja, tem natureza bidirecional (Sabate & Puente, 2003). Roberts e Dowling (2002) encontraram uma relação positiva entre o desempenho financeiro anterior (reputação financeira) e a reputação corporativa, bem como, uma relação positiva entre a reputação e a lucratividade futura de uma empresa representada pelo indicador de retorno sobre os ativos (ROA), enquanto que Fombrun e Shanley (1990) concluíram que medidas contábeis de risco e retorno são antecedentes da reputação de uma empresa medida através do *ranking America's Most Admired Companies* da revista Fortune.

Cabe recapitular que a reputação corporativa é um ativo intangível, valioso, difícil de replicar, socialmente construído e de natureza complexa, o qual beneficia as empresas de “boa” reputação, sinalizando a qualidade dos produtos e dos serviços oferecidos por elas. Essa característica funcionaria como uma barreira para empresas concorrentes de reputação inferior que, sem a ajuda da sinalização relativa à “boa” reputação, teriam dificuldade para demonstrar com rapidez suficiente a qualidade de seus produtos e serviços (Roberts & Dowling, 2002).

Como mencionado, *stakeholders* externos não têm acesso a toda informação sobre uma empresa para formar sua percepção a respeito dela, de modo que o desempenho financeiro passado atua como um sinal para o observador externo, o qual indica que a empresa é “boa” (Roberts & Dowling, 2002).

---

<sup>9</sup> “a) Que opiniões avaliativas e generalizadas são mantidas pelo conjunto de observadores sobre a organização? e b) Onde estas opiniões são mantidas dentro desse conjunto social e quão forte e consistentemente elas são mantidas?” (Tradução do autor).

A literatura descreve consequências positivas para empresas com “boa” reputação, como atração e fidelização de clientes, os quais são compelidos a pagar preços mais altos (Eberl & Schwaiger, 2005; Rindova et al., 2005), redução de custos, uma vez que fornecedores enxergam riscos menores ao realizar transações com empresas de “boa” reputação, contratação de funcionários mais qualificados, que, em geral, são mais eficientes ou trabalham por uma remuneração menor (Roberts & Dowling, 2002) e criação de barreiras competitivas, que permitem aumentar a estabilidade das receitas (Fernández Sánchez et al., 2012).

Estudos empíricos anteriores mensuraram o desempenho financeiro tanto por meio de medidas de desempenho baseadas em indicadores de mercado, que refletem o valor presente dos fluxos futuros de renda (Agnihotri, 2014; Cardoso et al., 2013; Lopes et al., 2017; OuYang et al., 2017; Pfarrer et al., 2010; Y. Wang & Berens, 2015), como por meio de medidas de desempenho baseadas em indicadores contábeis, que refletem o desempenho passado das empresas (Deephouse, 2000; Martinez et al., 2017; Orozco et al., 2018; Roberts & Dowling, 2002). Outras pesquisas empregaram tanto medidas de desempenho baseadas em indicadores de mercado como medidas de desempenho baseadas em indicadores contábeis (Bandeira et al., 2015; Koh et al., 2009; J. Lee & Roh, 2012; Y. M. Lee & Hu, 2018). O Quadro 2 apresenta a consolidação desses trabalhos, de acordo com o tipo de indicador financeiro impactado pela reputação corporativa. Aqueles pesquisadores que utilizaram medidas de desempenho financeiro contábil se basearam na associação, bem estabelecida na literatura, entre reputação corporativa e desempenho contábil, porém essas medidas de desempenho podem não refletir os benefícios de uma “boa” reputação, os quais geralmente se materializam a longo prazo, assim, a utilização de medidas de desempenho financeiro baseado em indicadores de mercado que refletem fluxos futuros de renda poderiam evidenciar uma relação mais significativa entre reputação corporativa e desempenho financeiro (Raithel & Schwaiger, 2015).

Uma “boa” reputação pode favorecer as empresas quando alguma informação negativa é divulgada. Nesses casos, os *stakeholders* podem conceder o “benefício da dúvida”, dando pouca importância ou desconsiderando essa informação, de modo que suas percepções sobre a empresa permanecem inalteradas (Lange et al., 2011). Por outro lado, há estudos que revelam uma faceta negativa de uma “boa” reputação, situação em que as empresas são demandadas a atender expectativas maiores (Lange et al., 2011). Assim, uma “boa” reputação pela qualidade dos produtos pode resultar em perdas financeiras maiores após *recalls* de produtos (Rhee & Haunschild, 2003), o que justificaria a reação negativa dos consumidores, como no caso envolvendo a Toyota Motor Corporation, relatado por Lange et al. (2011).

Quadro 2 – Tipo de indicador de desempenho financeiro associado à reputação corporativa

<b>Tipo de indicador de desempenho financeiro</b>	<b>Artigos</b>	<b>Indicadores</b>
Desempenho financeiro baseado em indicadores de mercado	Pfarrer et al. (2010)	Retorno Anormal Cumulativo (CAR).
	Cardoso et al. (2013)	Q de Tobin.
	Agnihotri (2014)	Q de Tobin.
	Y. Wang e Berens (2015)	Q de Tobin.
	Lopes et al. (2017)	Valor Econômico Agregado (EVA), Market Value Added® (MVA®).
	OuYang et al. (2017)	Retorno Anormal Cumulativo (CAR).
Desempenho financeiro baseado em indicadores contábeis	Deephouse (2000)	Retorno sobre os Ativos (ROA).
	Roberts e Dowling (2002)	Retorno sobre os Ativos (ROA).
	Martinez et al. (2017)	Retorno sobre o Capital Investido (ROIC).
	Orozco et al. (2018)	Retorno sobre os Ativos (ROA), Retorno sobre o Patrimônio Líquido (ROE) e Média das vendas líquidas.
Desempenho financeiro baseado em indicadores de mercado e contábeis.	Koh et al. (2009)	Retorno sobre os Ativos (ROA) e Q de Tobin.
	J. Lee e Roh (2012)	Retorno sobre os Ativos (ROA), Retorno sobre o Patrimônio líquido (ROE), Crescimento de Vendas e Q de Tobin.
	Bandeira et al. (2015)	Retorno sobre o Patrimônio líquido (ROE) e <i>Market Value Added</i> ® (MVA®).
	Y. M. Lee e Hu (2018)	Retorno sobre os Ativos (ROA) e Q de Tobin.

Fonte: Elaborado pelo autor.

#### 2.4 Abordagens de medição da reputação corporativa

A reputação corporativa é um dos ativos intangíveis mais importantes no que se refere a manter e a melhorar o desempenho das empresas. A fim de compreender quais são seus antecedentes e suas consequências, pesquisadores e praticantes têm empregado várias abordagens, as quais visam medir esse constructo (Ali et al., 2015; Sarstedt et al., 2013). Berens e van Riel (2004), em sua revisão da literatura, identificaram três correntes conceituais utilizadas para explorar e medir a reputação corporativa: a primeira é baseada na noção de expectativa social que os *stakeholders* têm com relação ao comportamento das empresas, a segunda é baseada na ideia de personalidade corporativa, que consiste nos traços de “personalidade” que os *stakeholders* atribuem às empresas, e a terceira usa a noção de confiança, que é a percepção dos *stakeholders* a respeito da honestidade, da confiabilidade e da benevolência da empresa. A primeira delas, relacionada às expectativas sociais, é a mais proeminente na literatura e as abordagens de medição desta corrente mais empregadas em estudos sobre reputação são: o índice *America's Most Admired Companies*, da revista Fortune, e o índice *Reputation Quotient*, do Reputation Institute (Sarstedt et al., 2013).

A seguir, é apresentada uma breve descrição dessas abordagens de medição, bem como as principais limitações, apresentadas na literatura, de cada uma.

#### 2.4.1 Índice *America's Most Admired Companies* (AMAC)

O índice AMAC foi o primeiro ranking de reputação de empresas dos EUA. Ele vem sendo publicado, anualmente, desde 1984, pela revista *Fortune* e, mais tarde, tendo sido estendido para abranger empresas de todo mundo, tornou-se o índice *World's Most Admired Companies* (WMAC) (Sarstedt et al., 2013). A medida AMAC/WMAC é a mais comumente usada em pesquisas acadêmicas sobre reputação corporativa (Deephouse, 2000; Kaur & Singh, 2018; Walker, 2010), o que ocorre, provavelmente, devido à conveniência e facilidade para obter dados de longo prazo, que permitem realizar análises longitudinais das empresas que o compõem.

Resultado da associação entre a empresa de consultoria Hay Group e a revista *Fortune*, esse índice foi desenvolvido para medir a reputação de empresas por meio de nove atributos, na versão WMAC: (1) solidez financeira; (2) valor de investimento a longo prazo; (3) uso inteligente dos ativos corporativos; (4) inovação; (5) capacidade para atrair, desenvolver, e manter pessoas talentosas; (6) qualidade de produtos ou serviços; (7) qualidade de gestão; (8) responsabilidade social e ambiental; e (9) eficácia em fazer negócios globalmente (Ali et al., 2015; Sarstedt et al., 2013). Os dados no qual se baseia são coletados por intermédio de pesquisa entre altos executivos e analistas do mercado financeiro, a partir disso, cada atributo é avaliado usando um único item e o índice final é calculado por meio de uma média simples (Sarstedt et al., 2013).

Apesar de seu uso frequente, vários autores criticam o índice AMAC/WMAC. As conclusões tiradas dos resultados de pesquisas empíricas que o utilizam como medida de reputação são limitadas devido a suas fraquezas metodológicas e conceituais (Deephouse, 2000), as quais são apresentadas a seguir. A primeira se refere ao fato de que, o índice é altamente correlacionado com o desempenho financeiro das empresas, o que gera um forte efeito halo do desempenho financeiro (Ali et al., 2015; Deephouse, 2000; Sarstedt et al., 2013); a segunda diz respeito a que o grupo de *stakeholders* que responde a pesquisa é limitado a altos executivos e analistas do mercado financeiro, os quais forçosamente não representam todos os grupos de *stakeholders* (Ali et al., 2015; Deephouse, 2000; Sarstedt et al., 2013); a terceira fraqueza é que, a pesquisa somente inclui grandes organizações com base nas receitas obtidas, assim, resultados empíricos obtidos a partir desse índice, provavelmente não podem ser generalizados para empresas menores (Ali et al., 2015; Deephouse, 2000); a quarta está relacionada com a falta de definição formal do constructo (Dowling, 2016); e finalmente, a quinta fraqueza que corresponde a que o índice AMAC/WMAC utiliza itens simples para

medição dos atributos, quando, para obter resultados confiáveis na medição de constructos complexos, como a reputação corporativa, é recomendado o uso de escalas com múltiplos itens (Ali et al., 2015; Sarstedt et al., 2013).

#### 2.4.2 Índice Reputation Quotient

O *Reputation Quotient* foi desenvolvido tanto com a finalidade de levar em conta as percepções de múltiplos grupos de *stakeholders* como com a de instituir a multidimensionalidade do constructo reputação corporativa (Fombrun, Gardberg, & Sever, 2000). Em 2011, foi publicada uma atualização deste índice denominada *RepTrak™ Pulse* (Fombrun, Ponzi, & Newburry, 2015). O índice consiste em uma escala de 20 atributos, os quais são agrupados em seis dimensões: (1) apelo emocional, (2) produtos e serviços, (3) desempenho financeiro, (4) responsabilidade social e ambiental, (5) visão e liderança e (6) ambiente de trabalho (Wartick, 2002). No que diz respeito a sua consistência, os resultados obtidos na validação dessa escala são sólidos (Fombrun et al., 2000; Wartick, 2002). Fombrun et al. (2000) argumentam que essa definição do constructo reputação combina dois fatores: emocional (ligado à dimensão apelo emocional) e cognitivo (vinculado às demais 5 dimensões).

Outra característica deste índice é a de que, no seu desenvolvimento, foram levadas em conta algumas das críticas a metodologia do índice AMAC/WMAC, tais como: incorporar múltiplas dimensões, avaliar as dimensões por meio de múltiplos itens na escala e levar em conta a percepção de vários grupos de *stakeholders*, ao invés de somente altos executivos e analistas do mercado (Ali et al., 2015).

Com relação às suas limitações, ainda que o *Reputation Quotient* seja embasado por uma definição formal, em contraste com o índice AMAC/WMAC, sua medição não espelha precisamente sua definição (Dowling, 2016). Outra crítica está relacionada com a dimensão de apelo emocional, que avalia aspectos de confiança, admiração e respeito, a qual deveria ser considerada um antecedente ou uma consequência da reputação ao invés de uma dimensão (Sarstedt et al., 2013).

O Quadro 3 sumariza as duas abordagens convencionais de medição da reputação corporativa que foram descritas, destacando as críticas a cada uma delas.

Quadro 3 – Abordagens de medição convencionais da reputação corporativa

Abordagem de medição	Críticas	Artigos
Índice <i>America's Most Admired Companies</i> (AMAC) / Índice <i>World's Most Admired Companies</i> (WMAC)	Altamente correlacionado com o desempenho financeiro.	Ali et al. (2015); Deephouse (2000); Sarstedt et al. (2013).
	Questionário respondido apenas por altos executivos e analistas do mercado. Respondentes não representam todos os grupos de <i>stakeholders</i> .	Ali et al. (2015); Deephouse (2000); Sarstedt et al. (2013).
	Somente inclui grandes empresas.	Ali et al. (2015); Deephouse (2000).
	Não possui definição formal.	Dowling (2016).
	Escala utiliza item simples para medição de cada atributo.	Ali et al. (2015); Sarstedt et al. (2013).
Índice <i>Reputation Quotient / RepTrak™ Pulse</i>	Sua medição não espelha precisamente sua definição.	Dowling, 2016.
	A dimensão apelo emocional é um antecedente ou um a consequência da reputação corporativa.	Sarstedt et al. (2013).

Fonte: Elaborado pelo autor.

## 2.5 Reputação corporativa e mídia

Como discutido anteriormente, esta pesquisa utiliza o modelo conceitual de reputação corporativa de Lange et al. (2011). Cabe destacar, no entanto, que a literatura não apresenta métodos para mensuração da dimensão “ser conhecido por algo” por intermédio de técnicas de mineração de texto. Assim, considerando o princípio da viabilidade, não é escopo deste trabalho desenvolver tais métodos. Trabalhos prévios que abordam a reputação corporativa mensuram essa dimensão por meio de escalas (Carter, 2006; Deephouse & Carter, 2005; Rindova et al., 2005). Isto posto, as duas dimensões, cujos efeitos nos resultados das empresas são explorados nesta pesquisa, são tratadas a seguir.

Devido às limitações das medidas de reputação descritas, houve a necessidade de desenvolver uma versão mais bem formada do conceito abstrato e complexo da reputação, o que foi feito utilizando a apresentação das empresas na mídia (Deephouse, 2000). Essa abordagem decorre da convicção de que a mídia é um importante avaliador, o qual tanto reflete como influencia a opinião pública a respeito das organizações (Carroll & McCombs, 2003; Deephouse & Heugens, 2009). Nesse sentido, além de a informação veiculada na mídia oferecer uma oportunidade para os *stakeholders* tomarem conhecimento das estratégias e das ações das empresas, e o teor das notícias também ajuda a moldar as opiniões com relação a elas (Agnihotri, 2014). Ainda com respeito a essa metodologia, o uso da mídia para medir a reputação corporativa implica que não são capturadas percepções de um grupo específico de

*stakeholders* e sim uma percepção consolidada dos diferentes *stakeholders* (Pollock, Rindova, & Maggitti, 2008).

A avaliação geral da empresa apresentada pela mídia, denominada por Deephouse (2000) como reputação da mídia, é resultado das matérias veiculadas em diversos meios de comunicação, as quais procedem de diversas fontes como, por exemplo: departamentos de relações públicas das empresas, *stakeholders*, órgãos governamentais, agências de classificação de risco e jornalistas.

A mídia atua como uma intermediária institucional no processo de sinalização entre as empresas e os *stakeholders* (Basdeo et al., 2006; Deephouse, 2000; OuYang et al., 2017; Pollock & Rindova, 2003), mitigando o problema da assimetria informacional (Deephouse, 2000; OuYang et al., 2017). No caso de sinais difíceis de serem observados diretamente, os *stakeholders* recebem essa informação por meio de intermediários, ou “infomediários”, como a mídia (Basdeo et al., 2006; OuYang et al., 2017). Além disso, a mídia influencia a percepção dos *stakeholders*, devido à sua habilidade de focar a atenção em empresas e em assuntos que ela seleciona para cobrir (Pollock & Rindova, 2003; Rindova et al., 2005). Em suma, a mídia reduz a incerteza dos *stakeholders* a respeito das características de uma empresa e atua como parte do processo de sinalização envolvido na construção da reputação (Deephouse, 2000; OuYang et al., 2017). Vale destacar, ainda, que os jornalistas, por divulgarem suas opiniões, além de relatar fatos, também são vistos como fontes confiáveis de informação e possuem grande potencial para afetar a reputação de uma empresa (Rindova, Petkova, & Kotha, 2007).

Notícias publicadas, em meios de comunicação em massa, como os jornais, ajudam, também, os investidores a moldar suas percepções sobre as empresas. Nesse sentido, notícias favoráveis (positivas) aumentam a reputação percebida pelos *stakeholders*, enquanto notícias desfavoráveis (negativas) diminuem a percepção desse aspecto por eles (Agnihotri, 2014).

Em seu estudo, Deephouse (2000) conclui que a reputação da mídia tem as propriedades de um recurso valioso, raro, difícil de imitar e não substituível (recurso VRIN). Desse modo, de acordo com a VBR, ela proporciona a uma empresa oportunidades para desenvolver vantagem competitiva sustentável (Barney, 1991), a qual conduz a um desempenho financeiro superior sustentável (Roberts & Dowling, 2002). A reputação da mídia corresponde à dimensão “favorabilidade generalizada” da conceituação de reputação proposta por Lange et al. (2011). Ela consiste em uma avaliação global da favorabilidade da organização e pode ser quantificada pelo coeficiente de desequilíbrio de Janis-Fadner, índice que foi utilizado em pesquisas anteriores de avaliação do teor da cobertura da mídia (Agnihotri, 2014; Deephouse, 2000; Janis & Fadner, 1943; OuYang et al., 2017; Pollock & Rindova, 2003; Rindova et al., 2005; Van den

Bogaerd & Aerts, 2014, 2015), e que calcula a proporção entre a diferença do número de notícias favoráveis (positivas) e de notícias desfavoráveis (negativas) e o número total de notícias relativas a uma dada empresa (Deephouse, 2000).

Levando em conta o exposto, como primeira hipótese deste trabalho, admite-se que a favorabilidade de uma empresa na mídia esteja relacionada com seu desempenho financeiro, fornecendo suporte para os argumentos que ela represente um sinal válido para os *stakeholders* e que afete o desempenho financeiro de acordo com a VBR. Deste modo, a seguinte hipótese é considerada:

**Hipótese 1 (H1): A favorabilidade da mídia de uma empresa é positivamente relacionada com seu desempenho financeiro.**

A visibilidade de uma empresa é seu grau de exposição na mídia e é, também, função direta do volume de cobertura que a empresa recebe (Pollock & Rindova, 2003). Além disso, é um motivo para o aumento da proeminência ou do destaque da organização aos olhos dos *stakeholders* (OuYang et al., 2017). A proeminência é uma das dimensões do constructo reputação corporativa, o qual captura o grau em que uma empresa é amplamente reconhecida pelos *stakeholders* em seu campo organizacional (Rindova et al., 2005). Visibilidade, proeminência e conscientização generalizada da empresa na percepção coletiva são conceitos sinônimos da dimensão “ser conhecido” do modelo conceitual de reputação corporativa proposto por Lange et al. (2011).

A visibilidade da mídia constitui-se um sinal institucional que contribui para a redução da assimetria informacional entre empresas e *stakeholders* (Fombrun & Shanley, 1990; OuYang et al., 2017). Fombrun e Shanley (1990), no entanto, encontraram que uma acentuada visibilidade da mídia, estimada por meio do número de notícias relacionadas a uma empresa, tem um efeito negativo sobre a reputação. Os achados desse estudo podem ser explicados pelos resultados da pesquisa de Jonkman, Boukes, Vliegthart, e Verhoeven (2019), que identificou que o efeito de notícias desfavoráveis sobre a avaliação geral de uma empresa é três vezes maior que o efeito de notícias favoráveis e, adicionalmente, esses autores observaram que uma maior visibilidade de uma empresa na mídia, na maioria dos casos, significa uma exposição à notícias com teor desfavorável. Assim, de acordo com o preconizado na VBR, a visibilidade estaria negativamente relacionada com o desempenho financeiro. Outros autores, entretanto, como Philippe e Durand (2011), encontraram uma associação positiva marginalmente significativa entre a cobertura da mídia e a reputação de empresas mensurada por *rankings*. Já, OuYang et al. (2017) não encontraram, em seu estudo, suporte para a associação entre o desempenho



financeiro e a visibilidade da mídia, também mensurada por meio do número de notícias sobre cada empresa da amostra. Finalmente, Pfarrer et al. (2010) concluem que a visibilidade da mídia, isoladamente, pode não ser positiva para uma empresa quando não existe um amplo reconhecimento sobre sua capacidade de criar valor. Dessa maneira, é proposto que o nível de visibilidade da mídia esteja relacionado com o desempenho financeiro, mas sem expectativa em relação ao sinal dessa relação.

**Hipótese 2 (H2): A visibilidade da mídia de uma empresa é associada a seu desempenho financeiro.**

O efeito positivo da favorabilidade de uma empresa pode ser maior quando a organização é mais conhecida, uma vez que o resultado do processo de sinalização pode ser mais pronunciado quando um maior número de sinais observáveis é enviado (OuYang et al., 2017). O volume de exposição na mídia aumenta a familiaridade com a empresa, o que propicia a construção de percepções favoráveis a respeito dela (Pollock & Rindova, 2003). Lange et al., (2011) consideram que a dimensão “ser conhecido” pode amplificar os efeitos produzidos por mudanças nas dimensões “ser conhecido por algo” e “favorabilidade generalizada”, já que empresas com grande visibilidade são alvos preferenciais de disputas e de observações positivas e negativas.

Em seu estudo, OuYang et al. (2017) observaram que a reputação da mídia de uma empresa está positivamente associada com o indicador Retornos Anormais Ajustados (CAR) após uma crise e que esta relação é moderada pela exposição da empresa na mídia.

Dessa maneira, é proposto que o nível de visibilidade modera a relação entre a favorabilidade da mídia e o desempenho financeiro de uma empresa, isto é, essa relação é fortalecida quando o nível de visibilidade é alto e vice-versa. Deste modo, segue a seguinte hipótese:

**Hipótese 3 (H3): A relação entre a favorabilidade da mídia de uma empresa e seu desempenho financeiro é reforçada por sua visibilidade da mídia.**

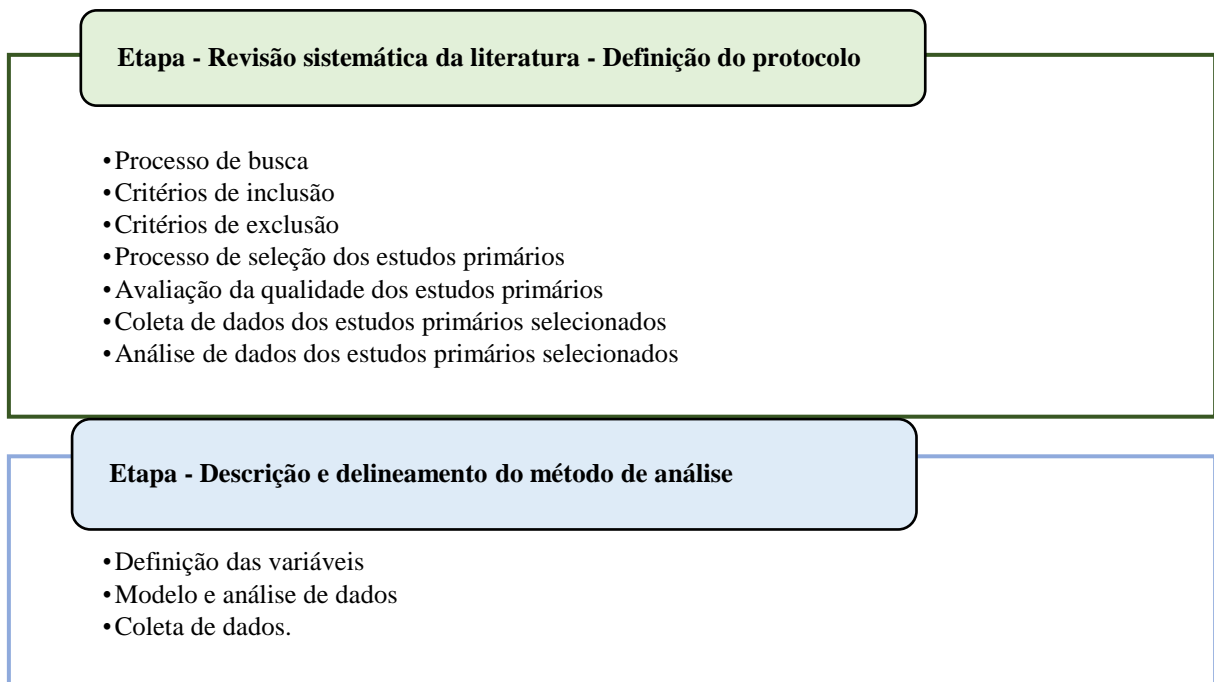
Assim, o teste dessas três hipóteses possibilita que seja examinado o efeito das dimensões visibilidade e favorabilidade da reputação sobre o desempenho financeiro das empresas, bem como, o efeito moderador da dimensão visibilidade da reputação sobre a associação entre a dimensão favorabilidade da reputação e o desempenho financeiro.



### 3 MÉTODO

A pesquisa é desenvolvida em duas etapas. A primeira consiste na revisão de literatura, o que é feito empregando-se a metodologia de revisão sistemática (Kitchenham & Charters, 2007), e a segunda compreende a descrição e o delineamento do método de análise dos dados do estudo. A Figura 4 exhibe as atividades que compõem cada etapa.

Figura 4 – Etapas da metodologia



Fonte: Elaborado pelo autor.

#### 3.1 Revisão sistemática da literatura

A fim de analisar a literatura acadêmica investigando a relação entre reputação corporativa e desempenho financeiro, esta pesquisa empregou o processo de revisão sistemática (Khan, Kunz, Kleijnen, & Antes, 2003; Kitchenham & Charters, 2007), o que foi feito tanto para identificar lacunas na literatura como para explorar os métodos empíricos quantitativos utilizados nos estudos sobre o tema. Uma revisão sistemática é um processo que visa identificar, analisar e interpretar todas as evidências importantes que atendam a critérios de elegibilidade pré-estabelecidos, com o objetivo de responder a questões específicas por meio de métodos documentados e sistemáticos, os quais buscam mitigar o viés do pesquisador na identificação, seleção, síntese e resumo dos estudos (Moher et al., 2015).

As questões que devem ser respondidas com esta revisão sistemática, a fim de atingir os propósitos mencionados, são as seguintes:

Questão de revisão 1. Quais são as pesquisas recentes (período de 2009 a 2018) que têm examinado as consequências da reputação corporativa no desempenho financeiro?

Questão de revisão 2. Quais são as teorias empregadas para explicar a relação entre reputação corporativa e desempenho financeiro?

Questão de revisão 3. Quais são as medidas de reputação corporativa?

Questão de revisão 4. Quais são os indicadores de desempenho financeiro que são relacionados com a reputação corporativa?

Questão de revisão 5. Quais são as variáveis de controle consideradas?

Questão de revisão 6. Quais são os métodos quantitativos utilizados para avaliar a relação entre reputação corporativa e desempenho financeiro?

Como mencionado, parte da motivação desta revisão sistemática é explorar estudos quantitativos acerca do tema, assim, as questões de 3 a 6 se vinculam diretamente a esse objetivo.

De acordo com Kitchenham e Charters (2007), em seu relatório técnico, uma revisão sistemática envolve três fases: planejamento, execução e relato da revisão. Similarmente, Khan et al., (2003), em seu artigo, descrevem cinco passos para realizar uma revisão sistemática: definição das questões de pesquisa, identificação das pesquisas relevantes, avaliação da qualidade das pesquisas, sumarização das evidências e interpretação dos resultados. Kitchenham e Charters (2007) consideram que o processo de revisão sistemática pressupõe o envolvimento de um grupo com vários pesquisadores, entretanto, eles sugerem que uma versão reduzida do processo pode ser realizada por um único pesquisador, como no caso de um estudante de pós-graduação. Dessa forma, este trabalho, executa o processo sugerido por Kitchenham e Charters (2007) para um único pesquisador. Cada fase da revisão sistemática é composta por certas atividades, as quais são listadas a seguir: (1) definir as questões da revisão, (2) desenvolver o protocolo de revisão, (3) especificar os critérios de inclusão e de exclusão que permitem endereçar o problema do viés da seleção, no caso de um único pesquisador, (4) definir a estratégia de busca e o processo de seleção dos estudos, (5) realizar o processo de avaliação de qualidade, e (6) extrair e sintetizar os dados. Os detalhes de cada atividade são desenvolvidos na sequência.

### 3.1.1 Protocolo de revisão

O protocolo é o procedimento metódico que é utilizado na revisão sistemática, cuja finalidade é mitigar o viés do pesquisador, já que, sem um procedimento documentado previamente, a seleção dos estudos primários poderia ser influenciada pelas expectativas de quem realiza a investigação (Kitchenham & Charters, 2007). A seguir, são apresentados os demais itens que compõem o protocolo de revisão sugerido por Kitchenham e Charters (2007).

### 3.1.2 Processo de busca

O processo de busca foi realizado por meio de mecanismos de pesquisa em 10 bases de dados bibliográficas, as quais foram identificadas como relevantes na área de administração e nas quais artigos significativos relacionados ao tema da pesquisa já haviam sido encontrados em uma revisão narrativa de literatura preliminar. Assim, as bases de dados pesquisadas foram: Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), EBSCO (base de dados: Academic Search Premier e Business Source Complete), Emerald eJournals (Emerald), Sage Journals, Scielo, Scopus, SpringerLink, Taylor Francis, Web of Science e Wiley Online Library. Com relação a esses repositórios, é importante notar que eles permitem pesquisar em diferentes campos, como título, autor, palavras-chave e resumo, bem como selecionar áreas de pesquisa, anos de publicação e tipos de documentos. Quanto à maneira de realizar a pesquisa, ela depende de cada base de dados, já que elas contam com distintos operadores e campos de busca, além disso, o formato de exportação dos resultados também é variado. Todas as bases de dados bibliográficas escolhidas podem ser acessadas pela funcionalidade de “Buscar base”, presente no portal de periódicos da Capes (<http://www.periodicos.capes.gov.br/>). No Quadro 4, são exibidas algumas características de cada base de dados, tais como: campos de pesquisa e formato para exportar resultados. Assim, as *strings* de busca devem ser executadas nos campos de pesquisa exibidos a seguir.

Quadro 4 – Bases de dados bibliográficas

Base de dados	Campos de pesquisa	Outros filtros	Formato para exportar resultados
Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)	Todos os campos	Ano de defesa: 2009 a 2018.	Valores separados por vírgulas (CSV).
EBSCO (Academic Search Premier e Business Source Complete)	Texto completo	Data de publicação: Jan-2009 a Dez-2018; Idioma: Pt, Eng.	Bibtex.
Emerald eJournals (Emerald)	Texto completo	<i>Publication date: From</i> Jan-2009 <i>To</i> Dec-2018.	Dados extraídos da página de resultados da pesquisa.

Base de dados	Campos de pesquisa	Outros filtros	Formato para exportar resultados
Sage Journals	Texto completo	<i>Publication date: Custom range: 2009 To 2018.</i>	Dados extraídos da página de resultados da pesquisa.
Scielo	Todos os índices	Ano de publicação: 2009 a 2018; Idioma: Inglês, português.	Bibtex.
Scopus	Título, resumo e palavras chave do autor	<i>Published: 2009 to 2018</i>	Valores separados por vírgulas (CSV).
SpringerLink	Texto completo	<i>Published between 2009 and 2018</i>	Valores separados por vírgulas (CSV).
Taylor Francis	Texto completo	<i>Date range: Jan-2009 to Dec-2018.</i>	Dados extraídos da página de resultados da pesquisa.
Web of Science (Coleção principal)	Título, resumo e palavras chave	Tempo estipulado: de 2009 até 2018.	Bibtex.
Wiley Online Library	Texto completo	<i>Publication Date – Custom range: 01–2009 to 12–2018.</i>	Dados extraídos da página de resultados da pesquisa.

Fonte: Elaborado pelo autor.

As *strings* de busca são definidas através dos conceitos principais presentes nas questões de pesquisa, como também pelos seus sinônimos e variações. Nessas *strings* de busca, o operador booleano de união (OR) é usado entre os sinônimos e variações dos conceitos principais, enquanto o operador de intersecção (AND) é utilizado entre dois ou mais conceitos. Para bases de dados que suportam busca no texto completo dos artigos, a *string* teve de ser elaborada tanto em português como em inglês. Por outro lado, para bases de dados que não suportam busca no texto dos artigos, uma *string* simplificada foi desenvolvida para permitir que resultados fossem retornados. Todas as bases bibliográficas empregadas na pesquisa permitem filtrar os resultados pelo período de publicação. Os quadros 5 e 6 mostram os conceitos mapeados e as *strings* de pesquisa empregadas na busca.

Quadro 5 – Conceitos Mapeados

Conceito	Expressão booleana do conceito
	<b>Inglês</b>
<i>Corporate reputation</i>	<i>("corporate reputation" OR "firm reputation" OR "organizational reputation" OR "media reputation" OR "business reputation" OR "company reputation" OR "brand reputation")</i>
<i>Measurement</i>	<i>(measure* OR mensuration OR quantif*)</i>
<i>Financial Performance</i>	<i>("corporate performance" OR "financial performance" OR "economic performance" OR "firm performance" OR "organizational performance" OR "company performance" OR "firm value" OR "stakeholder value" OR "shareholder value" OR "credit risk" OR "social performance" OR "market value")</i>
<i>Quantitative Method</i>	<i>("generalized regression" OR "generalized linear model" OR "regression model" OR "regression analysis" OR "linear regression" OR "multiple regression" OR "probabilistic learning" OR "machine learning" OR "statistical learning" OR "neural network" OR "deep learning" OR "dynamic system" OR "complex system" OR "dynamical system" OR "dynamic simulation" OR "structural equation modeling" OR SEM OR "partial least squares" OR PLS OR "partial least squares path modeling" OR PLS-PM)</i>

Conceito	Expressão booleana do conceito
<b>Português</b>	
Reputação corporativa	("reputação corporativa" OR "reputação da empresa" OR "reputação organizacional" OR "reputação na mídia" OR "reputação do negócio" OR "reputação da companhia" OR "reputação da marca")
Medição	(medição OR medida OR quantificação OR mensuração)
Desempenho financeiro	("desempenho corporativo" OR "desempenho financeiro" OR "desempenho econômico" OR "desempenho da empresa" OR "desempenho organizacional" OR "desempenho da companhia" OR "valor da empresa" OR "valor do acionista" OR "valor das partes interessadas" OR "risco de crédito" OR "desempenho social" OR "valor de mercado")
Relação	(relacionamento OR conexão OR relação OR associação OR ligação OR correlação OR correspondência OR elo OR interrelação OR interconexão)
Método quantitativo	("regressão generalizada" OR "modelo linear generalizado" OR "modelo de regressão" OR "análise de regressão" OR "regressão linear" OR "regressão múltipla" OR "aprendizado probabilístico" OR "aprendizado de máquina" OR "aprendizado estatístico" OR "rede neural" OR "deep learning" OR "sistema dinâmico" OR "sistema complexo" OR "simulação dinâmica" OR "modelagem de equações estruturais" OR MEE OR "mínimos quadrados parciais" OR MQP OR PLS OR "mínimos quadrados parciais análise do caminho" OR "PLS-PM")

Fonte: Elaborado pelo autor.

Quadro 6 – Strings de pesquisa

Uso	Expressão booleana aplicada ao campo da base de dados
<i>String</i> de pesquisa para bases de dados que suportam busca no texto completo dos artigos – idioma: inglês.	("corporate reputation" OR "firm reputation" OR "organizational reputation" OR "media reputation" OR "business reputation" OR "company reputation" OR "brand reputation") AND (measure* OR mensuration OR quantif*) AND ("corporate performance" OR "financial performance" OR "economic performance" OR "firm performance" OR "organizational performance" OR "company performance" OR "firm value" OR "stakeholder value" OR "shareholder value" OR "credit risk" OR "social performance" OR "market value") AND (relationship OR connection OR relation OR association OR link OR correlation OR correspondence OR bond OR interrelation OR interconnection) AND ("generalized regression" OR "generalized linear model" OR "regression model" OR "regression analysis" OR "linear regression" OR "multiple regression" OR "probabilistic learning" OR "machine learning" OR "statistical learning" OR "neural network" OR "deep learning" OR "dynamic system" OR "complex system" OR "dynamical system" OR "dynamic simulation" OR "structural equation modeling" OR SEM OR "partial least squares" OR PLS OR "partial least squares path modeling" OR PLS-PM)
<i>String</i> de pesquisa para bases de dados que não suportam busca no texto completo dos artigos – idioma: inglês	("corporate reputation" OR "firm reputation" OR "organizational reputation" OR "media reputation" OR "business reputation" OR "company reputation" OR "brand reputation") AND ("corporate performance" OR "financial performance" OR "economic performance" OR "firm performance" OR "organizational performance" OR "company performance" OR "firm value" OR "stakeholder value" OR "shareholder value" OR "credit risk" OR "social performance" OR "market value") AND (relationship OR connection OR relation OR association OR link OR correlation OR correspondence OR bond OR interrelation OR interconnection)

Uso	Expressão booleana aplicada ao campo da base de dados
<p><i>String</i> de pesquisa para bases de dados que suportam busca no texto completo dos artigos – idioma português</p>	<p>("reputação corporativa" OR "reputação da empresa" OR "reputação organizacional" OR "reputação na mídia" OR "reputação do negócio" OR "reputação da companhia" OR "reputação da marca") AND (medição OR medida OR quantificação OR mensuração) AND ("desempenho corporativo" OR "desempenho financeiro" OR "desempenho econômico" OR "desempenho da empresa" OR "desempenho organizacional" OR "desempenho da companhia" OR "valor da empresa" OR "valor do acionista" OR "valor das partes interessadas" OR "risco de crédito" OR "desempenho social") AND (relacionamento OR conexão OR relação OR associação OR ligação OR correlação OR correspondência OR elo OR interrelação OR interconexão) AND ("regressão generalizada" OR "modelo linear generalizado" OR "modelo de regressão" OR "análise de regressão" OR "regressão linear" OR "regressão múltipla" OR "aprendizado probabilístico" OR "aprendizado de máquina" OR "aprendizado estatístico" OR "rede neural" OR "deep learning" OR "sistema dinâmico" OR "sistema complexo" OR "simulação dinâmica" OR "modelagem de equações estruturais" OR MEE OR "mínimos quadrados parciais" OR MQP OR PLS OR "mínimos quadrados parciais análise do caminho" OR PLS-PM) NOT (rsc OR "responsabilidade social corporativa" OR "sustentabilidade corporativa" OR "sustentabilidade ambiental" OR "responsabilidade ambiental")</p>

Fonte: Elaborado pelo autor.

### 3.1.3 Critérios de inclusão

- Artigos de periódicos e de conferências;
- Artigos em inglês e português;
- Dissertações e teses em português;
- Pesquisas quantitativas;
- Estudos disponíveis em texto completo, com acesso gratuito através do portal de periódicos da Capes, ou em busca na web, através do Google Scholar (<https://scholar.google.com.br/>).
- Estudos publicados entre Jan–2009 e Dez–2018 ou entre 2009 e 2018 (no caso de bases de dados bibliográficas que não suportam o mês no filtro de data de publicação).
- Estudos que analisam a relação reputação–desempenho e possuem pelo menos uma hipótese que relaciona a reputação corporativa com o desempenho financeiro.

### 3.1.4 Critérios de exclusão

- Estudos que tratam o constructo reputação como mediador ou moderador da relação entre desempenho e outro constructo, como, por exemplo, a responsabilidade social corporativa (RSC).



- Estudos que tratam da relação entre imagem corporativa e desempenho financeiro.

### 3.1.5 Processo de seleção dos estudos primários

O processo de seleção dos estudos primários foi realizado em quatro etapas (Younas, Jawawi, Ghani, Fries, & Kazmi, 2018):

Etapa 1. As *strings* de busca foram executadas em cada base de dados bibliográfica e os resultados foram carregados em um banco de dados Microsoft® SQL Server. A seguir, foram gravados, nesse banco de dados, os artigos que haviam sido localizados, a partir de outras fontes de informação, no processo de revisão narrativa preliminar já descrito. Neste ponto, artigos duplicados, ou seja, os que possuíam mesmo título e autor, foram removidos. Finalmente, os registros foram carregados em uma planilha eletrônica Microsoft® Excel, a qual é a ferramenta utilizada nesta revisão.

Etapa 2. Seguindo trabalhos anteriores (Inkinen, 2015; Younas et al., 2018), os estudos são analisados, inicialmente, com base no título, de modo que aqueles que não podem ser excluídos pela avaliação do título passam para a etapa seguinte.

Etapa 3. Seguindo trabalhos anteriores (Inkinen, 2015; Younas et al., 2018), estudos são avaliados, agora, com base no resumo. De modo análogo, aqueles que não podem ser excluídos com base nesse critério passam para a próxima etapa.

Etapa 4. Leitura do texto completo do estudo, a fim de assegurar que ele é relevante para esta pesquisa. Os critérios de inclusão e de exclusão também são usados nesta etapa.

### 3.1.6 Avaliação da qualidade dos estudos primários

O processo de avaliação da qualidade de cada estudo é realizado por meio de quatro questões, as quais são apresentadas a seguir (Busalim & Hussin, 2016):

Critério 1: O tema endereçado no estudo é relevante para esta revisão sistemática da literatura?

Critério 2: A metodologia de pesquisa está descrita no estudo?

Critério 3: O método de coleta de dados está descrito no estudo?

Critério 4: As etapas de análise de dados são descritas no estudo?

Inicialmente, é atribuída uma nota a cada questão, o que é feito utilizando-se a seguinte escala: estudos que cumprem plenamente o critério recebem nota 2, aqueles que atendem parcialmente o critério obtêm nota 1, e os estudos que não satisfazem o critério recebem nota 0. A qualidade de cada estudo é avaliada pela nota final, a qual é igual à soma da nota de cada

questão. Consistente com Busalim & Hussin (2016), esta pesquisa avaliou a qualidade de cada estudo como alta, média ou baixa, dependendo de sua nota final. Assim, os estudos que obtiverem pontuação igual ou superior a 5 foram considerados com qualidade alta, enquanto os que alcançaram pontuação igual a quatro, foram considerados com qualidade média e, se abaixo de quatro, foram considerados de qualidade baixa. A partir dessa classificação, foram selecionados os estudos com qualidade alta ou média.

### 3.1.7 Coleta de dados dos estudos primários selecionados

O formulário para registrar as informações coletadas por meio da leitura de cada estudo foi implementado em planilha eletrônica Microsoft® Excel. Nesta revisão, foram preenchidos os seguintes campos com os dados extraídos de cada estudo:

- Título do estudo.
- Palavras-chave.
- Resumo.
- Autores: nomes de todos os autores.
- URL de acesso.
- Nome da base de dados bibliográfica.
- Data de publicação: ano de publicação do trabalho.
- Tipo: artigo de periódico, artigo de congresso, dissertação ou tese.
- Periódico: nome do periódico (ou campo vazio nos casos em que o estudo não tenha sido publicado em um periódico).
- Área temática do periódico: nome da área (ou campo vazio nos casos em que o estudo não tenha sido publicado em um periódico).
- Teorias: teorias empregadas no embasamento teórico do estudo.
- País do estudo: países nos quais estão localizadas as empresas da amostra do estudo.
- Indústria: setor de atividade das empresas da amostra do estudo.
- Período: intervalo de tempo, para amostras longitudinais, ou data, para amostras transversais.
- Medidas de desempenho: medidas de desempenho associadas com a reputação corporativa.
- Medida de reputação: operacionalização do constructo reputação corporativa.
- Variáveis de controle: outras variáveis que poderiam afetar a relação reputação corporativa – desempenho.

- Método quantitativo: técnica quantitativa para analisar a relação reputação corporativa – desempenho.

### 3.1.8 Análise de dados dos estudos primários selecionados

Para esta parte do processo, primeiramente os dados devem ser tabulados, a fim de se obter uma lista que contenha a informação completa de cada estudo. Em seguida, um identificador único é atribuído a cada item nessa relação. A lista, então, é explorada para responder às questões de pesquisa e para identificar quaisquer tendências ou limitações interessantes nos estudos. Assim, para responder as perguntas de revisão devem ser considerados os pontos a seguir:

Questão de revisão 1. Quais são as pesquisas recentes (período de 2009 a 2018) que têm examinado as consequências da reputação corporativa no desempenho financeiro? – Para responder essa pergunta, deve ser verificada a evolução do número de estudos primários no período designado.

Questão de revisão 2. Quais são as teorias empregadas para explicar a relação entre reputação corporativa e desempenho financeiro? – Com relação a esta pergunta, deve-se investigar se alguma teoria é predominantemente empregada como embasamento nos estudos selecionados.

Questão de revisão 3. Quais são as medidas de reputação corporativa? – Aqui, deve-se buscar identificar quais medidas são empregadas nos estudos selecionados.

Questão de revisão 4. Quais são os indicadores de desempenho financeiro que são relacionados com a reputação corporativa? – No caso dessa questão, devem ser identificados quais indicadores financeiros são considerados no estudo das consequências que a reputação corporativa tem para o desempenho financeiro.

Questão de revisão 5. Quais são as variáveis de controle consideradas? – No que se refere a esta questão, devem ser identificadas quais variáveis de controle são consideradas no estudo da relação entre reputação corporativa e desempenho financeiro.

Questão de revisão 6. Quais são os métodos quantitativos utilizados para avaliar a relação entre reputação corporativa e desempenho financeiro? – Para responder a esta pergunta, deve-se investigar se algum tipo de método quantitativo é predominante nos estudos selecionados.

### 3.2 Definição das variáveis

Nesta seção são descritas as variáveis presentes no modelo empregado na pesquisa.

#### 3.2.1 Variáveis dependentes – Desempenho financeiro

Este estudo explora o desempenho financeiro das empresas como uma variável dependente e, segue a pesquisa de J. Lee e Roh (2012) ao evitar o uso de apenas um indicador e ao empregar mensuração de desempenho baseada tanto em indicadores contábeis como em indicadores de mercado. Para medir o desempenho financeiro das empresas, o presente trabalho escolheu quatro indicadores, comumente utilizados em pesquisas empíricas anteriores que examinaram as consequências da reputação corporativa. No que se refere ao desempenho baseado em indicadores contábeis, duas medidas foram selecionadas: ROA (*return on assets* – retorno sobre os ativos) e ROE (*return on equity* – retorno sobre o patrimônio líquido). Com relação ao desempenho baseado em indicadores de mercado, por sua vez, as medidas adotadas foram: Q de Tobin e MVA<sup>®</sup> (*Market Value Added* – Valor de Mercado Adicionado).

##### 3.2.1.1 ROA

O ROA é uma medida de desempenho contábil calculada por meio do quociente entre o lucro líquido e o ativo total de uma empresa. Sendo um indicador consagrado e usado com frequência em estudos da área de negócios (Koh et al., 2009), ele é, também, uma medida muito comum da lucratividade de uma organização (Deephouse, 2000; Roberts & Dowling, 2002) e mede, ainda, a eficiência das empresas em gerar lucro com seus ativos totais (L. Liu, Li, Xu, & Zhang, 2014). Além disso, o ROA dá aos investidores a noção de como a empresa é eficaz na conversão do dinheiro investido em lucros, sem incluir a alavancagem financeira como bônus ou dedução (Y. M. Lee & Hu, 2018)

Trabalhos prévios apontam que ativos intangíveis, tais como reputação corporativa e capital intelectual, têm impacto positivo no ROA de uma empresa (Deephouse, 2000; Roberts & Dowling, 2002). Dessa forma, seguindo diversos estudos empíricos anteriores (Blajer-Gołębiewska, 2014; Deephouse, 2000; Domingos & Moura, 2013; Koh et al., 2009; J. Lee & Roh, 2012; Y. M. Lee & Hu, 2018; L. Liu et al., 2014; Orozco et al., 2018; Roberts & Dowling, 2002; Zhang & Rezaee, 2009), este estudo examina os efeitos da reputação corporativa na lucratividade de uma empresa, mensurada pelo retorno sobre os ativos (ROA).

### 3.2.1.2 ROE

O ROE é uma medida de desempenho contábil que permite aos acionistas constatar a lucratividade e examinar a gestão dos recursos próprios e de terceiros de uma empresa (Bandeira et al., 2015). Calculado por meio do quociente entre o lucro líquido e o patrimônio líquido (Gabriel, Assaf Neto, & Corrar, 2005), esse indicador expressa como uma empresa pode lucrar através de seu capital próprio. Para elevar o ROE, a empresa necessita lucrar mais ou obter recursos de terceiros por meio de endividamento (alavancagem financeira favorável), porém, avaliando somente o ROE, não é possível afirmar se uma empresa é eficiente para ganhar dinheiro ou para pedir dinheiro emprestado (Y. M. Lee & Hu, 2018).

O principal fator que diferencia o ROE e o ROA é a alavancagem financeira ou o endividamento. A equação fundamental do balanço mostra como isso é verdade:  $\text{ativos} = \text{passivo} + \text{patrimônio líquido}$ . Essa equação nos diz que, se uma empresa não possui dívidas, seu patrimônio líquido e seus ativos totais serão os mesmos. Segue-se, então, que o ROE e o ROA também serão os mesmos. Se, no entanto, essa empresa assumir dívidas, seu patrimônio líquido diminuir e o ROE subiria acima do ROA, desde que a dívida tomada tenha taxa atrativa frente ao retorno dos investimentos na forma de lucro.

Apesar das limitações do ROE, acompanhando várias pesquisas empíricas prévias que abordaram os efeitos da reputação corporativa (Bandeira et al., 2015; Blajer-Gołębiowska, 2014; Domingos & Moura, 2013; J. Lee & Roh, 2012; Orozco et al., 2018), este estudo examina o impacto da reputação corporativa na lucratividade de uma empresa, mensurada pelo retorno sobre o patrimônio líquido (ROE).

### 3.2.1.3 Q de Tobin

O indicador Q de Tobin é uma medida de desempenho financeiro baseada no mercado. Definido como a razão entre o valor de mercado de uma empresa e o custo de reposição de seus ativos (Y. M. Lee & Hu, 2018), ele relaciona dados de mercado e dados contábeis, possibilitando uma mensuração mais precisa do desempenho financeiro das corporações (Cardoso et al., 2013). Wang e Berens (2015) observam que o indicador Q de Tobin é melhor quando comparado a outros indicadores de desempenho financeiro, uma vez que ele revela o quadro completo da lucratividade de uma empresa, capturando seu potencial de crescimento. Ademais, esse indicador pode medir ativos intangíveis, como a reputação corporativa. Assim, se o Q de Tobin for maior que 1, o valor de mercado será maior que o valor dos ativos

registrados na empresa, o que sugere que esse indicador reflete alguma fonte de valor não mensurada ou não registrada, a qual pode ser atribuída aos ativos intangíveis (V. R. Rao, Agarwal, & Dahlhoff, 2004).

Nesta pesquisa, o indicador Q de Tobin é calculado por meio da fórmula aproximada proposta por Chung e Pruitt (1994), que é exibida na Equação 1.

$$Q \text{ de Tobin} = (VO + VP + DÍVIDA) / AT \quad (1)$$

Onde:

VO = valor da ação ordinária x número de ações ordinárias emitidas;

VP = valor da ação preferencial x número de ações preferenciais emitidas;

DÍVIDA = valor do passivo circulante subtraído do valor do ativo circulante somado ao valor do passivo não circulante;

AT = valor do ativo total.

Agnihotri (2014) argumenta que uma empresa com boa reputação é capaz de aumentar seu valor de mercado mais que uma empresa de menor reputação, de modo que a reputação corporativa pode ter um impacto positivo em medidas de desempenho baseado no mercado, como é o caso do indicador Q de Tobin. Assim, seguindo pesquisas empíricas anteriores (Agnihotri, 2014; Cardoso et al., 2013; Góis, De Luca, & Souza, 2017; Koh et al., 2009; J. Lee & Roh, 2012; Y. M. Lee & Hu, 2018; Y. Wang & Berens, 2015), este estudo examina, também, as consequências da reputação corporativa no indicador de desempenho financeiro Q de Tobin.

#### 3.2.1.4 MVA<sup>®</sup>

O MVA<sup>®</sup> (Market Value Added – Valor de Mercado Adicionado) é outra medida de desempenho baseada no mercado que demonstra que a administração agrega valor, isto é, eleva o valor da empresa para seus acionistas (Bandeira et al. 2015). Marca registrada da consultoria Stern Stewart & Co., criadora desse indicador (Góis, 2015), o MVA<sup>®</sup> é calculado subtraindo-se o valor de mercado do patrimônio líquido da empresa (Lópes et al. 2013). Quando os investidores querem saber qual o desempenho de uma empresa para seus acionistas, eles olham primeiro para o MVA<sup>®</sup>, um alto MVA<sup>®</sup> é evidência de gestão efetiva e de fortes capacidades operacionais, um baixo MVA<sup>®</sup>, por outro lado, pode significar que a eficácia da atuação dos gestores e dos investimentos realizados é menor que o valor do capital investido pelos acionistas (Góis, 2015).

Assim, seguindo pesquisas empíricas anteriores que analisaram a relação entre reputação corporativa e desempenho financeiro (Bandeira et al., 2015; Góis, 2015; Lopes et al.,

2017), este estudo examina, igualmente, as consequências da reputação corporativa no valor de mercado de uma empresa, mensurado pelo indicador MVA<sup>®</sup>. Visto que esse indicador de desempenho financeiro é uma medida absoluta, sua magnitude pode afetar o desempenho dos algoritmos de regressão, assim, foi aplicada a transformação logaritmo natural sobre o valor do indicador somado ao valor negativo máximo dessa série de dados acrescido de 1.

### 3.2.2 Variáveis independentes

As variáveis independentes da pesquisa são apresentadas a seguir.

#### 3.2.2.1 Favorabilidade da mídia

Os artigos sobre as empresas listadas na B3, cujo texto constitui a informação empregada nesta pesquisa para medir a favorabilidade, foram coletados por meio das páginas de consulta de notícias nos sítios de internet dos seguintes jornais: *Folha de São Paulo*, em sua editoria de mercado, *O Estado de São Paulo*, em sua editoria de economia, e *Valor Econômico*, em sua editoria de empresas. Essas matérias jornalísticas, relativas ao período deste estudo, que vai de 01-01-2012 a 31-12-2018, foram localizadas utilizando-se os mecanismos de busca de cada um desses sítios, os quais permitem o uso de um conjunto de palavras-chave, como o nome completo da empresa e a forma abreviada do nome da empresa.

Para garantir o processamento de notícias relevantes relativas a cada empresa, este estudo empregou parte do procedimento descrito por Tetlock, Saar-Tsechansky, e Macskassy (2008), no qual cada notícia, para ser selecionada deve atender aos seguintes critérios: mencionar o nome da empresa pelo menos uma vez nas primeiras 25 palavras, incluindo o título, e no mínimo duas vezes em todo o texto; e conter ao menos 50 palavras no total. Esses critérios de seleção foram impostos para eliminar notícias que contenham apenas tabelas ou listas com nomes de empresas e informações quantitativas; para excluir aquelas nas quais o nome da empresa apareça em função secundária; e para limitar a influência de *outliers* na medida do teor.

Seguindo a prática habitual em pesquisas empíricas anteriores, cada notícia foi classificada, por meio da análise de seu conteúdo, como favorável, neutra, ou desfavorável (Agnihotri, 2014; Deephouse, 2000; OuYang et al., 2017; Pfarrer et al., 2010; Pollock & Rindova, 2003; Van den Bogaerd & Aerts, 2014, 2015). Para isso, o teor de cada notícia selecionada deve ser determinado através do uso da técnica de mineração de texto denominada análise de sentimento baseada em léxico. No caso desta pesquisa, optou-se por testar os quatro

léxicos de sentimento indicados a seguir: OpLexicon (M. Souza, Vieira, Buseti, Chishman, & Alves, 2011, 2014): dicionário que possui cerca de 32.000 termos em português do Brasil (a cada termo é atribuída a polaridade: -1, 0 ou 1) e foi utilizado em pesquisas prévias que empregaram a análise de sentimento baseada em léxico (Avanço & Nunes, 2014; M. Souza & Vieira, 2012); Sentilex-PT (Carvalho & Silva, 2017; Silva, Carvalho, & Sarmiento, 2012): léxico como cerca de 82.000 termos em português de Portugal (a cada termo é atribuída a polaridade: -1, 0 ou 1), os quais foram ajustados para a ortografia do Brasil por meio das informações disponibilizadas no sítio “Portal da Língua Portuguesa” (ILTEC - Instituto de Linguística Teórica e Computacional, 2019). Esse léxico de sentimento foi utilizado em uma pesquisa que empregou análise de sentimento de *tweets* originados no Brasil (A. Souza, Figueredo, Cacho, Araújo, & Prolo, 2016); o léxico de sentimento baseado nos termos em inglês, com polaridades negativas e positivas, do dicionário proposto por Loughran e McDonald (Loughran & McDonald, 2011, 2018), que foi empregado em pesquisas anteriores de análise de sentimento nas áreas de negócios e finanças (Myskova & Hajek, 2016; Picault & Renault, 2017); e o léxico de sentimentos em inglês de M. Hu e Liu (M. Hu & Liu, 2004, 2019), que contém cerca de 6.800 termos e foi empregado em pesquisas prévias de análise de sentimentos em diferentes áreas (Almestekawy & AbdulSalam, 2019; Mostafa, 2018). Os dois léxicos de sentimentos em inglês empregados na pesquisa foram traduzidos para o português por meio dos recursos de tradução automática da Google Cloud<sup>10</sup>. A tradução de léxicos de sentimentos em inglês para outras línguas está entre as principais estratégias utilizadas por pesquisadores a fim de usar os vastos recursos e ferramentas disponíveis em inglês para análise de sentimentos em outras línguas (B. Liu, 2012).

Inicialmente, cada notícia relevante selecionada foi pré-processada, etapa em que foram realizados procedimentos de remoção de pontuação; remoção de números; conversão das letras no texto para minúsculas; remoção de palavras comuns (*stopwords*), empregando o dicionário de palavras irrelevantes em português, do projeto *Snowball* (Cavalcanti, Cavalcanti, Pires, Costa, & Cavalcanti, 2011); e remoção de espaços em branco. Após o pré-processamento, foi aplicado o processo de classificação de sentimentos conhecido como abordagem baseada em léxico, método no qual não é considerada nem a ordem dos termos dentro de um documento (*bag-of-words*), nem suas relações léxico-sintáticas (Becker & Tuminan, 2013), mas em que, basta identificar e contabilizar as palavras de sentimento presentes. Assim, o processo de definição do sentimento expresso em cada notícia é bastante simples: utilizando a matriz das

---

<sup>10</sup> API de tradução do Google Cloud <https://translation.googleapis.com/language/translate/v2>



palavras do texto do artigo e do léxico de sentimentos, é verificado quantas palavras negativas e positivas estão presentes. Dessa forma, o cálculo do sentimento da notícia pode ser feito como segue:

$$\text{Sentimento} = (P - N) / (P + N) \quad (2)$$

Onde:

P é a quantidade de palavras positivas;

N é a quantidade de palavras negativas.

Seguindo a pesquisa de Pfarrer et al. (2010), uma notícia é codificada como favorável se o sentimento positivo é de pelo menos 20% e como desfavorável se o sentimento negativo é maior ou igual a 20%. Notícias com proporções que ficam entre esses dois limiares são codificadas como neutras. Ademais, para cada artigo relevante foi produzida uma quintupla: empresa, identificador da notícia, data da notícia, teor da notícia (-1, 0, 1) e dicionário de sentimento.

A fim de determinar qual o dicionário de sentimento é mais apropriado, uma amostra aleatória das notícias relevantes foi extraída do conjunto de dados e cada notícia nessa amostra foi classificada manualmente pelo pesquisador como favorável, neutra ou desfavorável. Para isso, o pesquisador utilizou as seguintes diretrizes de codificação: notícias são favoráveis sempre que a empresa é apreciada pela mídia devido a sua estratégia, atuação e resultados; notícias são desfavoráveis quando a empresa é criticada por sua atuação e resultados e, finalmente, notícias são classificadas como neutras quando não se enquadram nas definições anteriores ou quando o nome da empresa aparece em função secundária. Os resultados obtidos na classificação manual foram, então, comparados com os resultados obtidos pela classificação automática. Assim, o léxico de sentimento que produziu resultados com maior aderência com a classificação manual foi empregado no cálculo da variável independente favorabilidade (FAV).

A partir do conjunto de notícias, calcula-se a favorabilidade da mídia usando o coeficiente de desequilíbrio de Janis-Fader (Deephouse, 2000, Janis & Fader, 1943), o que é feito por intermédio da fórmula a seguir:

$$\text{Favorabilidade} = (f^2 - fd) / t^2, \text{ se } f > d; 0, \text{ se } f = d; (fd - d^2) / t^2, \text{ se } d > f \quad (3)$$

Onde:

f = número de notícias favoráveis (positivas) sobre uma empresa em um dado trimestre;

d = número de notícias desfavoráveis (negativas) sobre uma empresa em um dado trimestre;

$t$  = número total de notícias, incluindo artigos que receberam classificação neutra, sobre uma empresa em um dado trimestre.

A faixa desse coeficiente varia de  $-1$  a  $1$ , onde  $-1$  indica cobertura totalmente desfavorável,  $1$  indica cobertura completamente favorável e  $0$  uma cobertura equilibrada. Por fim, vale enfatizar que, no cálculo do coeficiente de desequilíbrio de Janis-Fader, é atribuído um peso unitário a todas as notícias relevantes não sendo feita nenhuma ponderação baseada na intensidade do sentimento medido por meio da técnica de análise de conteúdo ou mesmo baseada no maior efeito que as notícias negativas poderiam ter sobre a reputação corporativa (Jonkman et al., 2019).

Cumprindo mencionar que Van den Bogaerd e Aerts (2015) argumentam que o cálculo da favorabilidade requer um número mínimo de notícias e, nesse sentido, descrevem que empresas com menos de 30 notícias coletadas durante um período de cinco anos são eliminadas da amostra em seu estudo. De maneira análoga, este trabalho, exclui empresas com menos de 140 notícias relevantes coletadas durante o período da pesquisa. Deste modo, o truncamento do conjunto de dados pode introduzir um viés de seleção que é tratado através da inclusão de uma variável de controle de acordo com o procedimento de correção de Heckman (Heckman, 1979).

### 3.2.2.2 Visibilidade da mídia

Seguindo estudos prévios (Fombrun & Shanley, 1990; OuYang et al., 2017; Pfarrer et al., 2010; Pollock & Rindova, 2003; Strauß & van der Meer, 2017; Van den Bogaerd & Aerts, 2014, 2015), este trabalho operacionaliza a visibilidade da mídia como sendo o grau de cobertura da mídia, o qual é entendido como o número de artigos relacionados a uma empresa específica que tenham sido recuperados por intermédio do procedimento já descrito, em um dado trimestre. A extensão da cobertura de uma empresa pela mídia afeta a dimensão de visibilidade da mídia de sua reputação corporativa, nesse sentido, uma empresa necessita ser conhecida a fim de criar um contexto no qual uma avaliação pública generalizável da empresa possa surgir (Van den Bogaerd & Aerts, 2015).

A transformação através do logaritmo natural dessa variável é aplicada de acordo com a pesquisa de OuYang et al. (2017), que argumentam que a finalidade é transformar uma variável com distribuição assimétrica em uma que seja mais aproximada da normal. O Quadro 7 apresenta um conjunto de pesquisas que examinaram a reputação corporativa e a mídia.

Quadro 7 – Estudos envolvendo Reputação Corporativa e Mídia

<b>Trabalho</b>	<b>Título</b>	<b>Operacionalização da medida de favorabilidade</b>	<b>Método de classificação do teor dos artigos</b>	<b>Operacionalização da medida de visibilidade</b>
Agnihotri (2014)	Mass-Media-based Corporate Reputation and Firms' Market Valuation – Evidence from Emerging Markets	Coeficiente de Janis-Fadner (Janis & Fadner, 1943)	Leitura dos artigos pelos pesquisadores.	Não usa.
Deephouse (2000)	Media Reputation as a Strategic Resource: An Integration of Mass Communication and Resource-Base Theories	Coeficiente de Janis-Fadner (Janis & Fadner, 1943)	Leitura dos artigos pelos pesquisadores.	Não usa.
Pfarrer et al. (2010)	A TALE OF TWO ASSETS: THE EFFECTS OF FIRM REPUTATION AND CELEBRITY ON EARNINGS SURPRISES AND INVESTORS' REACTIONS	Coeficiente de Janis-Fadner (Janis & Fadner, 1943)	Análise de sentimentos dos artigos baseada em léxico.	Número de artigos de jornais sobre as empresas.
Pollock e Rindova (2003)	MEDIA LEGITIMATION EFFECTS IN THE MARKET FOR INITIAL PUBLIC OFFERINGS	Coeficiente de Janis-Fadner (Janis & Fadner, 1943)	Leitura dos artigos pelos pesquisadores.	Número de artigos de jornais sobre as empresas.
Ouyang et al. (2017)	Information Asymmetry and Investor Reaction to Corporate Crisis: Media Reputation as a Stock Market Signal	Coeficiente de Janis-Fadner (Janis & Fadner, 1943)	Leitura dos artigos pelos pesquisadores.	Logaritmo natural do número de artigos de jornais sobre as empresas.
Strauß e Van der Meer (2017)	News media coverage and initial public offerings in Germany: explaining flotation performance	Sentimento: positivo, negativo e neutro	Leitura dos artigos pelos pesquisadores.	Número de artigos de jornais sobre as empresas.
Van den Bogaerd e Aerts (2014)	Media Reputation of a Firm and Extent of Trade Credit Supply	Coeficiente de Janis-Fadner (Janis & Fadner, 1943)	Leitura de parte dos artigos pelos pesquisadores (dados de treinamento) e aprendizado de máquina supervisionado para classificação.	Número de artigos de jornais sobre as empresas.
Van den Bogaerd e Aerts (2015)	Does media reputation affect properties of accounts payable?	Coeficiente de Janis-Fadner (Janis & Fadner, 1943)	Leitura de parte dos artigos pelos pesquisadores (dados de treinamento) e aprendizado de máquina supervisionado para classificação.	Número de artigos de jornais sobre as empresas.

Fonte: Elaborado pelo autor.

### 3.2.3 Variáveis de controle

Uma vez que tanto o desempenho financeiro como as variáveis independentes que medem a reputação podem sofrer influências de diversos fatores, considerou-se necessária a inclusão de variáveis de controle para separar seus efeitos das variáveis independentes e mitigar o viés de variáveis omitidas na análise. Com base em pesquisas empíricas anteriores que examinaram reputação corporativa e desempenho financeiro, as variáveis de controle neste estudo incluem a alavancagem financeira, a idade, o setor, o tamanho da empresa, o nível de governança corporativa e o fator de correção do viés de seleção que pode ser introduzido em decorrência do procedimento de cálculo da variável independente favorabilidade. A apresentação individual de cada variável está a seguir.

#### 3.2.3.1 Alavancagem financeira

A alavancagem financeira afeta os custos de agência e, portanto, influencia sistematicamente o desempenho financeiro de uma empresa (H. Wang, Choi, & Li, 2008). Consistente com pesquisas empíricas anteriores que analisaram reputação corporativa e desempenho financeiro (Cahan, Cahan, Lee, & Nguyen, 2017; Dell'Atti, Trotta, Iannuzzi, & Demaria, 2017; Góis et al., 2017; Kabir & Thai, 2017; Koh et al., 2009; J. Lee & Roh, 2012; OuYang et al., 2017; Sainty & Dunn, 2009; H. Wang et al., 2008; Y. Wang & Berens, 2015; Zhang & Rezaee, 2009), este estudo inclui a variável de controle alavancagem financeira a fim de monitorar seus efeitos no desempenho financeiro. A alavancagem financeira é operacionalizada pelo índice de endividamento geral (*debt ratio*), que resulta da razão entre o passivo total e os ativos totais (Y. Wang & Berens, 2015; Zhang & Rezaee, 2009).

#### 3.2.3.2 Idade

Y. M. Lee e Hu (2018) observam que empresas estabelecidas há mais tempo obtém vantagem competitiva sobre empresas mais jovens, uma vez que a experiência na área de atuação é percebida como um fator que contribui para a melhoria do desempenho das empresas. Além disso, empresas mais antigas podem ser mais conhecidas publicamente e, também, mais dignas de destaque na mídia quando comparadas a empresas mais jovens (Strauß & van der Meer, 2017). Destarte, o modelo do estudo inclui uma variável de controle para a idade da empresa. Seguindo pesquisas empíricas anteriores sobre a relação entre reputação corporativa

e desempenho, a idade da empresa é medida como o logaritmo natural do número de dias contados desde sua constituição até a data-base do trimestre (Bandeira et al., 2015; Van den Bogaerd & Aerts, 2014, 2015).

### 3.2.3.3 Indústria - Setor

Flanagan, O'Shaughnessy, e Palmer (2011) proporcionaram evidências empíricas de que a indústria a qual pertence uma empresa tem influência importante na reputação corporativa, adicionalmente, em cada estágio de um ciclo econômico o desempenho financeiro de cada indústria varia (Pfarrer et al., 2010). Winn, MacDonald, & Zietsma (2008) mostraram que a reputação da indústria, definida como o julgamento coletivo de uma indústria, tanto por *stakeholders* como pelo público em geral, pode influenciar a reputação individual das empresas. Levando em conta que o conjunto de empresas listadas na B3 abrange uma ampla gama de setores e seguindo pesquisas empíricas anteriores que examinaram reputação corporativa e desempenho financeiro (Bandeira et al., 2015; Eberl & Schwaiger, 2005; Fernández Sánchez et al., 2012; Fernández-Gámez et al., 2016; Flanagan et al., 2011; Y. M. Lee & Hu, 2018; Martinez et al., 2017; Pfarrer et al., 2010; Roberts & Dowling, 2002; Strauß & van der Meer, 2017; Van den Bogaerd & Aerts, 2014, 2015; Y. Wang & Berens, 2015; Zhang & Rezaee, 2009), o modelo deste estudo controla os efeitos da indústria (setor) da empresa por meio da definição de duas variáveis *dummy*, a fim de codificar as três classes de indústria propostas na pesquisa de Wang e Berens (2015): a de manufatura, a financeira e a de comércio e serviços.

### 3.2.3.4 Tamanho da empresa

Roberts e Dowling (2002) argumentam que o tamanho da empresa reforça sua capacidade para manter uma vantagem competitiva, seja por meio de economias de escala, de economias de escopo ou de efeitos de aprendizagem. Em geral, empresas maiores têm melhor gerenciamentos de riscos o que, em teoria, mitiga as ameaças financeiras (Chen, Gao, Li, & Wu, 2016). Adicionalmente, OuYang et al. (2017) indicam que empresas maiores atraem mais atenção dos *stakeholders* e podem ter uma cobertura mais positiva da mídia.

Assim, seguindo inúmeras pesquisas empíricas anteriores examinando reputação corporativa e desempenho financeiro (Agnihotri, 2014; Bandeira et al., 2015; Cardoso et al., 2013; Góis et al., 2017; J. Lee & Roh, 2012; Y. M. Lee & Hu, 2018; Lopes et al., 2017; OuYang et al., 2017; Pfarrer et al., 2010; Raithel et al., 2010; Van den Bogaerd & Aerts, 2014, 2015; Y. Wang & Berens, 2015; Zhang & Rezaee, 2009), o modelo deste estudo inclui uma variável de

controle para o tamanho da empresa, que visa monitorar seus efeitos. Nas pesquisas prévias, o tamanho da empresa é operacionalizado através das seguintes medidas: logaritmo natural das vendas (OuYang et al., 2017; J. Lee & Roh, 2012), ativo total (Agnihotri, 2014), valor de mercado (Zhang & Rezaee, 2009), logaritmo natural do número de empregados (Strauß & van der Meer, 2017), vendas (Roberts & Dowling, 2002; Eberl & Schwaiger, 2005) e logaritmo natural do ativo total (Bandeira et al., 2015; Cardoso et al., 2013; Góis, De Luca, & Souza, 2017; Y. M. Lee & Hu, 2018; Lopes et al., 2017; Pfarrer et al., 2010; Raithel et al., 2010; Van den Bogaerd & Aerts, 2014, 2015; Y. Wang & Berens, 2015). Assim, seguindo a operacionalização mais comum dessa variável em pesquisas anteriores, o tamanho é mensurado pelo logaritmo natural do ativo total da empresa.

### 3.2.3.5 Governança corporativa

A atenção sobre a governança corporativa foi desencadeada por grandes escândalos corporativos, como os relacionados às empresas: Enron, Worldcom, Tyco, Poly Peck, Maxwell, entre outros (Kurniati, 2019). A alta administração dessas corporações manipulou informações contábeis e gerenciais, lesou direitos de acionistas e não forneceu informações relevantes para os investidores (Pereira & Martins, 2015). Esses eventos trouxeram prejuízos para os *stakeholders*, principalmente, para os acionistas, e motivaram a busca de mecanismos para eliminar falhas na gestão corporativa, o que levou ao surgimento do conceito de governança corporativa (Bayrakdaroglu, Ersoy, & Citak, 2012). Dessa forma, a governança corporativa consiste em boas práticas de gerenciamento, as quais consideram os interesses dos *stakeholders* ao máximo e se baseiam nos princípios de transparência, justiça, prestação de contas, e responsabilidade (Bayrakdaroglu et al., 2012), de modo a permitir limitar os conflitos de interesses entre acionistas e administradores e, especialmente, a minimizar os custos gerados por tais conflitos (Berthelot, 2010). A governança corporativa compõe-se de mecanismos capazes de controlar e monitorar o problema da agência descrito por Jensen e Meckling (1976), no qual a divergência de interesses entre acionistas e administradores ou entre acionistas e credores ou entre acionistas e *stakeholders* resulte em conflitos de agência. Pereira e Martins (2015) argumentam que essas circunstâncias afetam o valor e o desempenho econômico-financeiro das empresas.

Assim, uma empresa com melhor governança, por ter custos de agência mais baixos, deve demonstrar melhor desempenho financeiro (López-Quesada, Camacho-Miñano, & Idowu, 2018). Bayrakdaroglu et al. (2012) observam que a implementação dos princípios de

governança corporativa melhora o desempenho e aumentam o valor das empresas. Ademais, empresas com boa governança corporativa têm melhor credibilidade, o que contribui para obter empréstimos de instituições financeiras com taxas de juros mais baixas do que empresas com governança corporativa pior (Mishra & Mohanty, 2014).

Apesar de não terem sido selecionados na revisão sistemática de literatura estudos examinando a relação reputação–desempenho financeiro que incluíssem, também, a governança corporativa como uma variável de controle, foram localizados vários estudos que consideram esse aspecto como um antecedente do desempenho financeiro (Al-ahdal, Alsamhi, Tabash, & Farhan, 2020; Bayrakdaroglu et al., 2012; Berthelot, 2010; Hossain, Alamgir, & Alam, 2016; Kurniati, 2019; López-Quesada et al., 2018; Mishra & Mohanty, 2014; Pereira & Martins, 2015; Shan & McIver, 2011; Silveira, Barros, & Famá, 2003).

Com relação a esta pesquisa, seu modelo controla os efeitos da governança corporativa por meio da inclusão de variáveis *dummy* para codificar os seguintes níveis de governança diferenciada da B3 (B3, 2019b): Nível 1 (N1), Nível 2 (N2) e Novo Mercado (NM), sendo que o Nível 1 tem um menor grau de exigência e o Novo Mercado tem maior exigência. Essas variáveis *dummy* foram utilizadas em estudos prévios (Dalmácio, Lopes, Rezende, & Sarlo Neto, 2013; Sarlo Neto, 2009), como uma proxy para a governança corporativa. Cabe mencionar que não foram definidas variáveis *dummy* para codificar os segmentos especiais de listagem: Bovespa Mais e Bovespa Mais Nível 2.

### 3.2.3.6 *Inverse Mills Ratio* – IMR

Conforme mencionado na descrição do procedimento de cálculo da variável independente favorabilidade, o conjunto de dados não inclui as empresas com menos de 140 notícias relevantes coletadas durante o período do estudo. Dessa forma, os resultados obtidos podem ser afetados pelo viés de seleção da amostra (Heckman, 1979). Wooldridge (2012) descreve um procedimento estatístico de duas etapas, conhecido como correção de Heckman (Heckman, 1979). A etapa inicial consiste em calcular um termo, o qual é obtido pela modelagem do mecanismo de seleção, que deve ser incluído como variável adicional da regressão (conhecido como *Inverse Mills ratio* - razão inversa de Mill). O procedimento para o cálculo deste parâmetro é descrito por Wooldridge (2012) e consiste em realizar uma regressão *Probit*, que modela o mecanismo de seleção, conforme a Equação 4:

$$Pri,t = \beta_0 + \beta_1 * TAM_{i,t} + \beta_2 * IDA_{i,t} + \beta_3 * IBOV_{i,t} + \varepsilon_{i,t} \quad (4)$$

Onde:  $i$  e  $t$  representam as empresas e os trimestres, respectivamente, e  $Pr$  representa a probabilidade de que a empresa supere o limiar estabelecido de notícias relevantes no período da pesquisa, isto é, seu valor é igual a 1 para empresas que superaram o limiar (empresas incluídas na amostra) e igual a 0 caso contrário (empresas não incluídas na amostra). O estudo segue os trabalhos de Van den Bogaerd (2014, 2015) e Agnihotri (2014), os quais mediram a reputação por meio de notícias e truncaram a amostra com base na quantidade de notícias das empresas. Esses trabalhos incluíram na modelagem do mecanismo de seleção as variáveis tamanho (TAM) e idade (IDA). Ademais, considerando que as empresas presentes na carteira teórica do índice Bovespa têm mais cobertura na mídia do que outras empresas, foi incluída no modelo do mecanismo de seleção a variável *dummy* IBOV, que é igual a 1 se a empresa está no índice Bovespa, e 0, caso não esteja.

A seguir, o resultado da regressão *Probit*, o qual consiste em um objeto da classe *generalized linear models* do pacote *stats* do *software* R (R Core Team, 2019), é usado como entrada da função *invMillsRatio*, do pacote *sampleSelection*, integrante do *software* R (Toomet & Henningsen, 2008), a qual retorna o fator *Inverse Mills Ratio*, que, na segunda etapa do procedimento, deve ser incorporado ao modelo de regressão a fim de corrigir o viés de seleção da amostra, uma vez que somente são selecionadas empresas que superaram o limiar de notícias relevantes definido. A verificação da possível influência do viés de seleção amostral nos resultados é feita por meio da análise da significância do coeficiente da variável *Inverse Mills Ratio*, se esse coeficiente for significativo, o procedimento de Heckman não foi capaz de corrigir o viés introduzido pelo truncamento do conjunto de dados da pesquisa (Wooldridge, 2012).

Um sumário das variáveis, incluindo a descrição, a operacionalização e o embasamento teórico, é apresentado no Quadro 8, a seguir.



Quadro 8 – Descrição das variáveis da pesquisa

Variável	Tipo	Operacionalização	Fonte de dados	Embasamento teórico
Retorno sobre os ativos – ROA	Dependente	Lucro líquido / Ativo total	Economática®	Blajer-Gołębiewska (2014); Deephouse (2000); Domingos e Moura, (2013); Koh et al. (2009); J. Lee e Roh (2012); Y. M. Lee e Hu (2018); L. Liu et al. (2014); Orozco et al. (2018); Roberts & Dowling (2002); Zhang e Rezaee (2009).
Retorno sobre o patrimônio líquido – ROE	Dependente	Lucro líquido / Patrimônio líquido	Economática®	Bandeira et al. (2015); Blajer-Gołębiewska (2014); Domingos e Moura (2013); J. Lee e Roh (2012); Orozco et al. (2018).
Q de Tobin – QTB	Dependente	Equação 1.	Economática®	Agnihotri (2014); Cardoso et al. (2013); Góis et al. (2017); Koh et al. (2009); J. Lee e Roh (2012); Y. M. Lee e Hu (2018); Y. Wang e Berens (2015).
Market Value Added® - MVA® a	Dependente	Logaritmo natural (Valor de mercado do Patrimônio líquido / Patrimônio líquido)	Economática®	Bandeira et al. (2015); Góis (2015); Lopes et al. (2017).
Favorabilidade da mídia (FAV)	Independente	Equação 3.	Sítios de notícias <i>on-line</i>	Agnihotri (2014); Deephouse (2000); OuYang et al. (2017); Pfarrer et al. (2010); Pollock e Rindova (2003); Van den Bogaerd e Aerts (2014, 2015).
Visibilidade da mídia (VIS)	Independente	Logaritmo natural do número de notícias sobre cada empresa no trimestre.	Sítios de notícias <i>on-line</i>	Fombrun e Shanley (1990); OuYang et al. (2017); Pfarrer et al. (2010); Pollock e Rindova (2003); Strauß e Van der Meer (2017); Van den Bogaerd & Aerts (2014, 2015).
Alavancagem financeira (ALV)	Controle	Passivo total / Ativo total	Economática®	Cahan et al. (2017); Dell'Atti et al. (2017); Góis et al. (2017); Kabir e Thai (2017); Koh et al. (2009); J. Lee e Roh (2012); OuYang et al. (2017); Sainty e Dunn (2009); H. Wang et al. (2008); Y. Wang e Berens (2015); Zhang e Rezaee (2009).
Idade da empresa (IDA)	Controle	Logaritmo natural do número de dias desde a constituição até a data-base do trimestre.	(CVM)	Bandeira et al. (2015); Van den Bogaerd e Aerts (2014).
Indústria (IND1 e IND2)	Controle	Variáveis <i>dummy</i> para as seguintes classes de indústrias: comércio e serviços (IND1= 0 e IND2 = 0), manufatura (IND1= 1, IND2= 0) e financeira (IND1= 0 e IND2= 1).	Classificação Setorial – B3	Bandeira et al. (2015); Eberl e Schwaiger (2005); Fernández Sánchez et al. (2012); Fernández-Gámez et al. (2016); Flanagan et al. (2011); Y. M. Lee e Hu (2018); Martinez et al. (2017); Pfarrer et al. (2010); Roberts e Dowling (2002); Strauß & Van der Meer, (2017); Van den Bogaerd e Aerts (2014, 2015); Y. Wang e Berens (2015); Zhang e Rezaee (2009).

Variável	Tipo	Operacionalização	Fonte de dados	Embasamento teórico
Tamanho da empresa (TAM)	Controle	Logaritmo natural do ativo total.	Economatica®	Agnihotri (2014); Bandeira et al. (2015); Cardoso et al. (2013); Góis et al. (2017); J. Lee e Roh (2012); Y. M. Lee e Hu (2018); Lopes et al. (2017); OuYang et al. (2017); Pfarrer et al. (2010); Raithel et al. (2010); Van den Bogaerd e Aerts (2014, 2015); Y. Wang e Berens (2015); Zhang e Rezaee (2009).
Governança corporativa (NM, N2 e N1) <sup>b</sup>	Controle	Variáveis <i>dummy</i> : Empresa listada no Novo Mercado (NM=1, N2=0 e N1=0), empresa listada no Nível 2 de governança (NM=0, N2=1 e N2=0), empresa é listada no Nível 1 de governança (NM=0, N2=0 e N1=1) e empresa não listada em um dos níveis anteriores (NM=0, N2=0, N1=0).	B3	Dalmácio et al. (2013); Sarlo Neto (2009).
<i>Inverse Mills Ratio</i> – IMR	Controle	Item 3.2.3.6 e Apêndice D.	B3, CVM, Economatica®	Agnihotri (2014), Deephouse (2000); Van den Bogaerd e Aerts (2014, 2015).

Nota: <sup>(a)</sup> Valor calculado deslocado conforme descrito na seção 3.2.1.4. <sup>(b)</sup> Empresas não pertencentes a nenhum desses níveis diferenciados representam o grupo base da codificação das variáveis *dummy* relacionadas com governança corporativa.

Fonte: Elaborado pelo autor.

### 3.3 Modelo e análise de dados

Para testar a associação entre o desempenho financeiro de uma empresa e sua favorabilidade (FAV) e visibilidade da mídia (VIS), este estudo utiliza o modelo empírico básico a seguir:

$$IFI_{i,t} = \beta_0 + \beta_1 * FAV_{i,t} + \beta_2 * VIS_{i,t} + \beta_3 * FAVxVIS_{i,t} + \beta_4 * ALV_{i,t} + \beta_5 * IDA_{i,t} + \beta_6 * TAM_{i,t} + \beta_7 * IND1_{i,t} + \beta_8 * IND2_{i,t} + \beta_9 * NM_{i,t} + \beta_{10} * N1_{i,t} + \beta_{11} * N2_{i,t} + \beta_{12} * IMR_{i,t} + \varepsilon_{i,t} \quad (5)$$

Na Equação 5, os subscritos i e t denotam empresa e trimestre, respectivamente. O modelo empírico inclui variáveis de controle consistentes com a literatura prévia sobre determinantes do desempenho financeiro, além do termo  $\varepsilon$ , que representa o erro. O termo IFI indica cada uma das medidas de desempenho financeiro cuja associação com a reputação é verificada neste estudo: ROA, ROE, MVA<sup>®</sup> e Q de Tobin.

A fim de testar a hipótese H3, da influência da variável visibilidade (VIS) na magnitude da associação entre a variável favorabilidade (FAV) e o desempenho financeiro, deve-se avaliar se o coeficiente do termo de interação FAVxVIS é significativo e diferente de zero (Hayes, 2013), não sendo necessário que existam conjuntamente efeitos significativos, seja para a variável preditora, seja para a variável moderadora (Baron & Kenny, 1986).

Vale salientar que a heterocedasticidade e a presença de valores extremos (*outliers*) são características comuns de dados financeiros (Duarte, Girão, & Paulo, 2017). O método de regressão aplicado na pesquisa, se mantém não tendencioso e consistente mesmo quando o pressuposto de homocedasticidade não é satisfeito, entretanto, o erro-padrão dos coeficientes estimados é viesado de modo que a estimativa da significância de cada coeficiente fica prejudicada. Assim, surgem questionamentos sobre como os pesquisadores devem lidar com os valores extremos do conjunto de dados, alguns autores sugerem a remoção dos valores extremos (Salama, 2005) e outros sua suavização através da técnica de *winsorizing* (Duarte, Girão, & Paulo, 2017). O emprego de erros padrão robustos é uma das técnicas aplicadas para lidar com heterocedasticidade. Dentre eles, os estimadores do tipo HAC (*Heteroskedasticity and Autocorrelation Consistent covariance matrix*) podem superar tanto o problema de heterocedasticidade como o problema de auto correlação. Cumpre destacar que Arellano (1987) propõe o cálculo de erros padrão consistentes com heterocedasticidade e auto correlação para regressão em painel. Newey e West (1987) também definiram um estimador que lida com a auto correlação (também chamada correlação serial) dos resíduos de regressões aplicadas a

dados de séries temporais. Compre mencionar que a técnica de suavização (*winsorizing*) das variáveis é empregada neste estudo. Além disso, a pesquisa deve utilizar estimadores robustos caso a análise dos resíduos da regressão indiquem violação das premissas de homocedasticidade e da não existência de auto correlação.

O conjunto de dados do estudo constitui-se em um painel composto por empresas (variável de seção transversal), de modo que cada uma teve até 28 observações, com periodicidade trimestral, no período com início no primeiro trimestre de 2012 e fim no último trimestre de 2018. A análise de dados em painel possui três abordagens mais comuns: POLS (*Pooled Ordinary Least Square*), efeitos fixos e efeitos aleatórios (Fávero, Belfiore, Silva, & Chan, 2009). O Quadro 9 apresenta os testes que devem ser executados para determinar qual o modelo de dados em painel deve ser empregado. A regressão com dados em painel, bem como, os testes relativos a qual modelo de dados em painel empregar foram realizados por meio das funções do pacote *plm* do *software R* (Croissant & Millo, 2008). Quanto ao processo de definição da abordagem a ser utilizada no estudo, ele começa com o teste de Hausman (Hausman, 1978), se o resultado desse teste indicar efeitos fixos, deve ser executado o teste de Chow (Wooldridge, 2012) para decidir entre efeitos fixos e POLS; caso contrário deve ser executado o teste LM (Lagrange Multiplier) Breusch – Pagan (Breusch & Pagan, 1980) para decidir entre efeitos aleatórios e POLS.

Quadro 9 – Testes para definição do melhor modelo para análise de dados em painel

Testes	Modelos de dados em painel	Hipóteses	Software R pacote::função
Hausman	Efeitos aleatórios x efeitos fixos.	H0: modelo de efeitos aleatórios é melhor. Ha: modelo de efeitos fixos é melhor.	<i>plm::phtest</i>
Chow	POLS x efeitos fixos.	H0: modelo POLS é melhor. Ha: modelo de efeitos fixos é melhor.	<i>plm::pooltest</i>
LM Breusch-Pagan	POLS x efeitos aleatórios.	H0: modelo POLS é melhor. Ha: modelo de efeitos aleatórios é melhor.	<i>plm::plmtest</i>

Fonte: Elaborado pelo autor com base em (Fávero et al., 2009).

No que se refere aos testes para as hipóteses deste estudo, foi aplicada a análise de regressão linear com dados em painel. O procedimento de análise de dados começa com o exame de estatísticas descritivas e, posteriormente, se faz a verificação das correlações de Pearson entre as medidas, a fim de averiguar se existe associação entre a variável dependente (desempenho financeiro) e as variáveis independentes e de controle e, também, a fim de constatar que não existem correlações relevantes entre essas variáveis. Adicionalmente, deve ser testada a correlação cruzada para estimar a necessidade e o valor da defasagem (*lag*) entre cada variável independente e a variável dependente.

Em seguida, são executados os modelos de regressão com dados em painel empregando a abordagem definida a partir do resultado dos testes, que são apresentados no Quadro 9. Primeiramente, é executado um modelo composto somente das variáveis de controle (modelo de controle). A seguir, um modelo com ambas variáveis independentes (favorabilidade e visibilidade), a fim de testar as hipóteses H1 e H2 por intermédio do valor e da significância dos coeficientes  $\beta_1$  e  $\beta_2$  e, finalmente, é executado o terceiro modelo, com inclusão do termo de interação, para testar a hipótese H3 por intermédio do valor e da significância do coeficiente  $\beta_3$ . Com relação à verificação dos pressupostos da regressão linear, o teste de Breusch-Pagan, implementado por meio da função *bptest* do pacote *lmtest* (Zeileis & Hothorn, 2002), deve ser executado, com objetivo de verificar se a premissa de homocedasticidade é satisfeita. Outra premissa da regressão linear, relacionada com dados longitudinais, é a ausência de auto correlação, assim, o teste de Breusch-Godfrey, implementado por meio da função *bpptest* do pacote *plm* (Croissant & Millo, 2008), deve ser efetuado. A seguir, é verificada a premissa de normalidade dos resíduos, o que é feito por meio do teste de Kolmogorov-Smirnov, implementado por intermédio da função *ks.test* do pacote *stats* (R Core Team, 2019). Por fim, é analisada a multicolinearidade entre as variáveis independentes e de controle por intermédio do fator de inflação da variância (*Variance Inflation Factor – VIF*), implementado por meio da função *icmdiag* do pacote *mctest* (Imdadullah, Aslam, & Altaf, 2016). O poder preditivo dos modelos é avaliado pelo coeficiente de determinação ( $R^2$ ). O Quadro 10 mostra os testes dos pressupostos da regressão linear que devem ser realizados por meio de pacotes do *software* estatístico R.

Quadro 10 – Testes dos pressupostos da análise de regressão linear

Testes	Pressuposto	Hipóteses	Pacote::função do software R
Breusch-Pagan	Homocedasticidade	H0: Homocedasticidade. H1: Heterocedasticidade.	<i>lmtest::bptest</i>
Breusch-Godfrey	Auto correlação (correlação serial)	H0: Ausência de correlação serial. H1: Presença de correlação serial.	<i>plm::bpptest</i>
Kolmogorov-Smirnov	Distribuição normal dos resíduos	H0: Resíduos são normalmente distribuídos. H1: Resíduos não são normalmente distribuídos.	<i>stats::ks.test</i>
VIF <sup>a</sup>	Multicolinearidade	Valor do indicador inferior a 5 descarta a existência de multicolinearidade.	<i>mctest:icmdiag</i>

Nota: <sup>(a)</sup> Variance Inflation Factor.

Fonte: Elaborado pelo autor com base em Fávero et al. (2009).

A seguir, empregando a técnica de regressão por meio de aprendizado de máquina supervisionado, deve ser feita a análise do modelo completo, que inclui as variáveis de controle, as variáveis independentes e o termo de interação. Este estudo utiliza o algoritmo *Gradient*

*boosting machines* (GBM), que produz um modelo preditivo por meio da combinação de outros modelos preditivos com menor acurácia (*ensemble method*), tipicamente árvores de decisão. Esse algoritmo demonstrou sucesso em várias aplicações práticas e é altamente configurável para as necessidades da aplicação, incluindo diferentes funções de perda (Natekin & Knoll, 2013). No processo iterativo desse algoritmo é minimizada uma função de perda (resíduo). Neste estudo é utilizada a função de perda quantílica, que tem melhor comportamento na presença de *outliers*, os quais são comuns em dados financeiros. Além disso, o modelo GBM possui duas categorias de hiperparâmetros que devem ser ajustados: os parâmetros *boosting* e os parâmetros específicos de árvore. Os dois principais hiperparâmetros *boosting* são:

Número de árvores: refere-se à quantidade total de árvores a serem ajustadas que equivale ao número de iterações do algoritmo. O algoritmo GBM, em geral, exige muitas árvores (não é incomum ter milhares delas), entretanto, com um número muito grande de árvores a minimização da função de perda pode produzir *over fitting* (sobreajuste), situação em que, o modelo descreve bem o comportamento do conjunto de dados de treinamento, mas não é capaz de generalizar para novos dados. Assim, é necessário estabelecer um número apropriado de árvores que não acarrete sobreajuste.

Taxa de aprendizado: também é chamado de encolhimento (*shrinkage*), esse hiperparâmetro determina a contribuição de cada árvore no resultado e controla a rapidez com que o algoritmo converge para o valor mínimo da função perda. Seus valores variam de 0 a 1, de modo que valores menores tornam o modelo robusto às características específicas de cada árvore, propiciando que o modelo produzido não seja viesado e, além disso, facilitam a parada antes que o sobreajuste ocorra. Geralmente, quanto menor esse valor, mais preciso será o modelo resultante, de outro lado, porém, mais iterações serão necessárias.

Os dois principais hiperparâmetros específicos de árvore em um modelo GBM simples são:

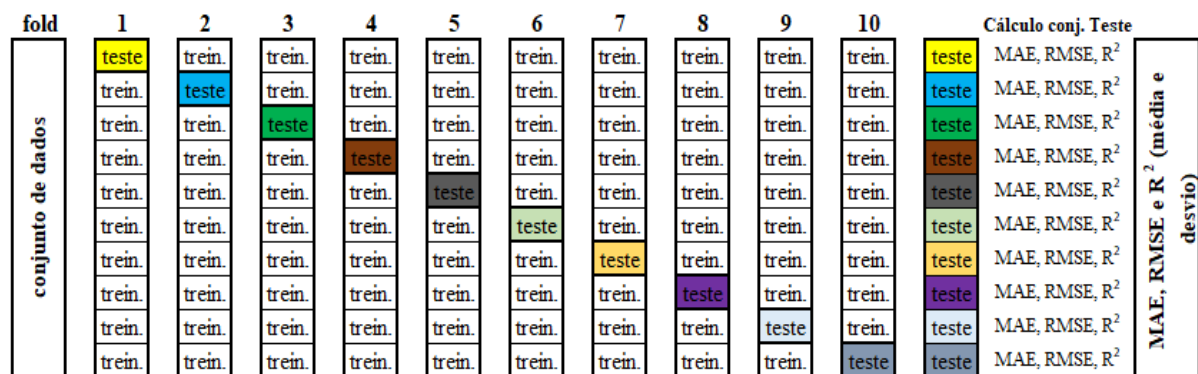
Profundidade da árvore: parâmetro que controla a profundidade das árvores individuais, isto é, o número de nós folhas que cada árvore pode ter. Os valores típicos de profundidade variam de 3 a 8, de maneira que árvores de menor profundidade são computacionalmente eficientes (mas requerem mais árvores), enquanto árvores de maior profundidade permitem que o algoritmo capture comportamentos particulares, mas também aumentam o risco de sobreajuste.

Número mínimo de observações nos nós terminais: parâmetro que controla a complexidade de cada árvore. Seus valores típicos variam de 5 a 15, onde valores mais altos

ajudam a impedir o sobreajuste e valores menores podem ajudar a detectar classes em conjunto de dados desbalanceados em problemas de classificação.

Após o ajuste dos hiperparâmetros, deve ser empregada a técnica de validação cruzada do modelo preditivo, o que é feito através do método *k-fold*, com *k* igual a 20. A técnica de validação cruzada consiste em, primeiramente, particionar o conjunto de dados em subconjuntos mutuamente exclusivos e, posteriormente, utilizar um dos subconjuntos para a estimação dos parâmetros do modelo (dados de treinamento) e o outro subconjunto (dados de validação ou de teste) para a validação do modelo. As métricas de qualidade de previsão em cada um dos conjuntos de teste calculadas são: coeficiente de determinação ( $R^2$ ), erro absoluto médio (MAE) e erro quadrático médio (RMSE). Adicionalmente, a média e o desvio das métricas de qualidade das 20 amostras são computadas. A Figura 5 exibe o processo para um valor de  $k=10$ .

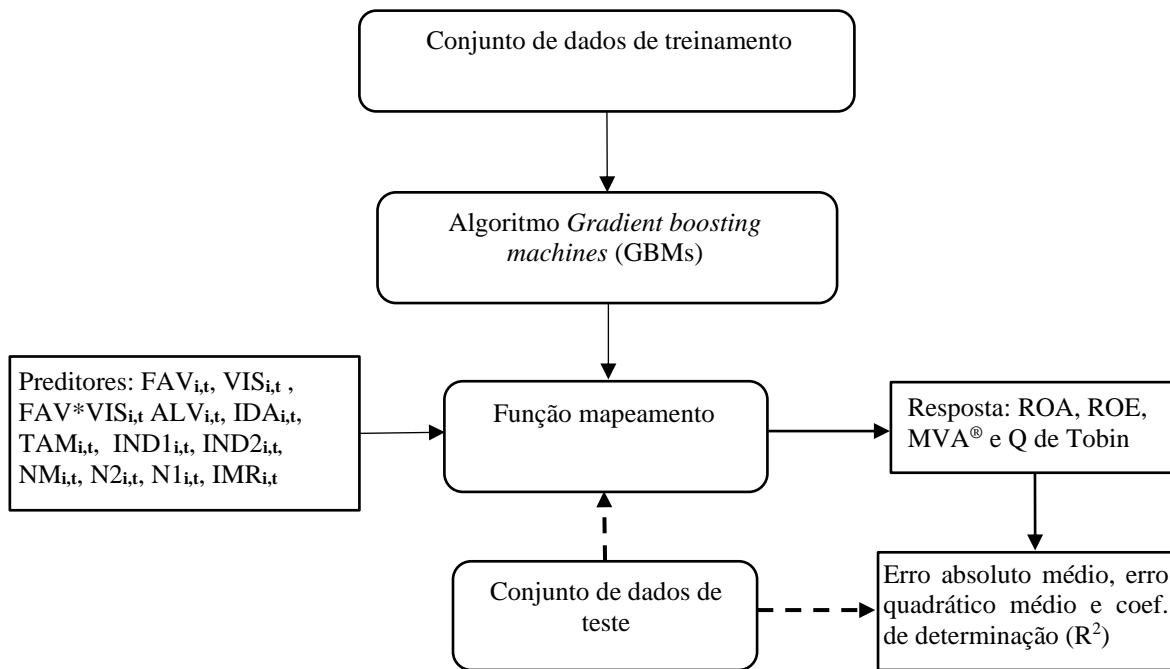
Figura 5 – Validação cruzada *k-fold* (exemplo para  $k = 10$ )



Fonte: Elaborado pelo autor.

Esta pesquisa emprega o *software* R (R Core Team, 2019) para fazer a análise de regressão por meio do algoritmo GBM. Quanto à validação cruzada para obter um modelo menos enviesado, ela é realizada por meio dos seguintes pacotes: *gbm* (Greenwell, Boehmke, Cunningham, & Gbm, 2019), que é a implementação mais popular em R do algoritmo GBM de Friedman (Friedman, 2002), e *caret* (Kuhn, 2008) para a validação cruzada. A Figura 6 descreve o processo de construção do modelo preditivo por meio de aprendizado de máquina supervisionado.

Figura 6 – Modelo preditivo



Fonte: Elaborado pelo autor.

Finalmente, para comparar o modelo preditivo construído empregando a técnica de aprendizado de máquina supervisionado com aquele construído utilizando a técnica de regressão linear com dados em painel, o modelo de regressão linear completo, que inclui as variáveis de controle, independentes e termo de interação, deve ser executado utilizando o método de validação cruzada *k-fold*, com *k* igual a 20. O erro médio absoluto e o erro médio quadrático, bem como os desvios padrão associados devem ser computados. A análise deve ser feita por meio de um teste de hipótese (teste T para médias) com a finalidade de constatar se existem evidências estatísticas para afirmar, com grau de confiança de pelo menos 95%, que a regressão por meio do algoritmo de aprendizado de máquina utilizado tem maior exatidão para ambas as medidas de erro (MAE, RMSE) e maior poder explicativo ( $R^2$ ). As hipóteses estatísticas, relativas a esses procedimentos, estão enunciadas nas equações 6 a 8, nas quais os subscritos *rl* e *ap* denotam regressão linear e aprendizado de máquina, respectivamente. Nas equações 6 e 7, a rejeição da hipótese nula indica que existem evidências estatísticas para afirmar que o método de aprendizado de máquina é mais preciso. Na equação 8, a rejeição da hipótese nula indica que existem evidências estatísticas para afirmar que o método de aprendizado de máquina conduz a um resultado com maior poder explicativo ( $R^2$ ).

$$\begin{cases} H_0: MAE_{rl} - MAE_{ap} \leq 0 \\ H_a: MAE_{rl} - MAE_{ap} > 0 \end{cases} \quad (6)$$

$$\begin{cases} H_0: RMSE_{rl} - RMSE_{ap} \leq 0 \\ H_a: RMSE_{rl} - RMSE_{ap} > 0 \end{cases} \quad (7)$$



$$\begin{cases} H_0: R_{rl}^2 - R_{ap}^2 \geq 0 \\ H_a: R_{rl}^2 - R_{ap}^2 < 0 \end{cases} \quad (8)$$

### 3.4 Coleta de dados

A pesquisa utiliza o método de levantamento de dados secundários, que consiste em empregar dados pré-existentes, os quais são coletados de diversas fontes. A partir da interpretação dessas informações é possível produzir conhecimento. Dessa forma, os dados secundários são coletados de duas fontes distintas: a base de dados Economatica<sup>®</sup> e em diferentes sítios de notícias *on-line*. A população da pesquisa é composta pelas empresas de capital aberto listadas na B3 (B3, 2019a) em dezembro de 2018. Da população de 423 empresas, três delas foram suprimidas: a Bradespar que administra a participação acionária do Banco Bradesco na Vale S.A.<sup>11</sup>, a Itausa, que é o conglomerado que controla o Banco Itaú Unibanco S.A.<sup>12</sup>, e a BNDESPar<sup>13</sup>, que administra as participações em empresas detidas pelo Banco Nacional do Desenvolvimento Social (BNDES). Com respeito a fusões e aquisições, uma pesquisa manual – feita em julho de 2019 no sítio de RI das empresas que compõem a população do estudo – encontrou os seguintes casos: a fusão da empresa FIBRIA com a empresa SUZANO S.A., ocorrida em janeiro de 2019, que não afeta a população do estudo; a fusão da empresa ALIANSCE SHOPPING CENTERS S.A. com a empresa SONAE SIERRA BRASIL S.A., realizada em junho de 2019, que não afeta a população do estudo; a aquisição da UNIDAS pela LOCAMERICA, em janeiro de 2018, caso em que ambas as empresas são consideradas na população da pesquisa. Com relação às mudanças de razão social, a consulta manual mencionada encontrou os seguintes casos: a empresa MPX Energia teve a razão social alterada para ENEVA em outubro de 2013, de modo que o nome MPX também deve ser usado na coleta de notícias relevantes da empresa ENEVA; a empresa TRACTEBEL S.A. teve a razão social alterada para ENGIE S.A. em julho de 2016, de modo que o nome TRACTEBEL também deve ser usado na coleta de notícias relevantes da ENGIE S.A.; e a empresa ALL – América

---

<sup>11</sup> Informação no sítio da Bradespar (<https://www.bradespar.com.br/SiteBradespar/Inicio/A-Bradespar/Investidas>)

<sup>12</sup> Informação no sítio da Itausa (<http://www.itusa.com.br/pt/conheca-a-itausa/quem-somos?>)

<sup>13</sup> Informação no sítio do BNDESPar (<https://www.bndes.gov.br/wps/portal/site/home/transparencia/consulta-operacoes-bndes/carteira-acionaria>).

Latina Logística S.A. teve a razão social alterada para RUMO S.A. em fevereiro de 2015, assim, o nome ALL também deve ser usado na coleta de notícias relevantes da empresa RUMO S.A.

O primeiro conjunto de dados foi coletado na base de dados Economatica<sup>®</sup> e consiste em métricas e indicadores financeiros das empresas listadas na B3 com periodicidade trimestral, considerando o período de sete anos do estudo, que vai desde o primeiro trimestre de 2012 até o quarto trimestre de 2018. O uso de empresas listadas em bolsa de valores nas pesquisas sobre reputação corporativa é habitual, pois os dados financeiros são abertos e, além disso, essas empresas são mais visíveis na mídia do que as empresas de capital fechado (OuYang et al., 2017). A esse conjunto de dados são incluídas duas variáveis: a data de constituição da empresa, que é obtida nos “Dados Cadastrais” no sítio da CVM (CVM), e a presença nos níveis diferenciados de governança corporativa da B3, que é obtida a partir das informações da cotação das ações para empresas negociadas (B3, 2019c) e, também, no sítio de Relações com Investidores (RI) das empresas que não são negociadas na bolsa.

O segundo conjunto de dados consiste em notícias *on-line*. Estudos prévios indicam que notícias de jornais são um dos meios primários pelos quais os *stakeholders* conhecem as empresas e suas atividades (Einwiller et al., 2010) e são, também, a melhor fonte de informações para a formação das percepções dos *stakeholders* sobre as empresas (Deephouse, 2000; OuYang et al., 2017). Em sua revisão sistemática de literatura, Nassirtoussi, Aghabozorgi, Ying Wah, e Ngo (2014) revelaram que a maioria dos sistemas que aplicam mineração de texto para previsão do mercado utilizam notícias financeiras e que, por outro lado, as mídias sociais, tais como Twitter e postagens em *blogs*, são menos utilizadas. Além disso estudos mostram que a natureza ruidosa dos dados provenientes das mídias sociais oferece sérios desafios para sua utilização (Chatterjee, Deng, Liu, Shan, & Jiao, 2018; T.-r. Hu, Luo, Kautz, & Sadilek, 2016; Sarker & Gonzalez, 2016) e que as mídias sociais, apesar de terem a capacidade de transmitir notícias em grande escala e de possibilitarem a proliferação de conteúdo gerado pelos usuários, veiculam, em grande parte, informações cuja veracidade é questionável (Conroy, Rubin, & Chen, 2015), o que pode criar e perpetuar a desinformação (Conroy et al., 2015). A disseminação de *fake news*, ou seja, notícias de baixa qualidade com informações intencionalmente falsas, tem potencial de causar impactos extremamente negativos para as pessoas e a sociedade. Assim, o uso de dados provenientes de mídias sociais deveria ser acompanhado da aplicação de métodos de detecção automática de *fake news* para garantir a confiabilidade da informação (Shu, Sliva, Wang, Tang, & Liu, 2017).

Portanto, acompanhando pesquisas empíricas anteriores que mediram a reputação corporativa por meio de informações da mídia empregando notícias publicadas em jornais

(Agnihotri, 2014; Deephouse, 2000; Kioussis et al., 2007; OuYang et al., 2017; Pfarrer et al., 2010; Pollock & Rindova, 2003; Strauß & van der Meer, 2017; Van den Bogaerd & Aerts, 2014, 2015) e considerando os problemas de confiabilidade das informações de mídias sociais, este estudo utiliza notícias de jornais *on-line* para medir duas dimensões da reputação corporativa: a visibilidade da mídia e a favorabilidade da mídia. As versões digitais de dois dos principais jornais de referência do Brasil, a *Folha de São Paulo* e *O Estado de São Paulo*, foram escolhidas, uma vez que estes jornais estão entre os quatro primeiros com maior número de assinantes digitais em 2017 (ABI, 2018). Além desses veículos de informação, também foi eleita para a pesquisa a edição digital do jornal *Valor Econômico*, que é o maior jornal de economia e negócios do país (Folha Uol, 2016). As edições digitais desses jornais são assinadas pelo pesquisador.

Dessa forma, nas edições digitais desses três jornais, foram coletadas informações das notícias *on-line* sobre as empresas listadas na B3. Para isso, foi utilizado um processo composto das seguintes etapas: extração de dados *web* (*web scraping*), que realiza a leitura do texto do artigo (Harrison, 2019; Wickham, 2019), aplicação do filtro de relevância e cômputo do teor de cada notícia. Assim, de cada notícia relevante lida foram armazenadas somente as seguintes informações: título, data de publicação, *hiperlink*, índice de sentimento e léxico de sentimento. Além disso, notícias duplicadas foram excluídas, restando apenas um registro da notícia quando vários itens com um título idêntico para uma mesma empresa foram identificados.

O Quadro 11, apresenta a Matriz de Amarração Metodológica adaptada à presente pesquisa, instrumento que foi sugerido por Mazzon em 1981<sup>14</sup>, o qual “constitui uma representação matricial em que se apresentam as conexões e os vínculos entre modelo, objetivos, questões e/ou hipóteses de pesquisa e procedimentos e técnicas de análise de dados” (Telles, 2001, p. 61).

---

<sup>14</sup> MAZZON, José Afonso. Análise do programa de alimentação do trabalhador sob o conceito de marketing social. 1981. (Tese de Doutorado) – Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo, São Paulo, SP, Brasil.

Quadro 11 – Matriz de amarração metodológica

<b>Questão de pesquisa:</b> Qual a associação entre reputação corporativa (i.e., favorabilidade e visibilidade) e desempenho financeiro?			
<b>Objetivo geral:</b> Examinar o efeito da reputação corporativa (i.e., favorabilidade e visibilidade) sobre o desempenho financeiro das empresas de capital aberto listadas na B3. Além disso, investigar o efeito moderador da visibilidade da mídia sobre a associação entre a favorabilidade da mídia e o desempenho financeiro.			
<b>Objetivos específicos</b>	<b>Hipóteses de pesquisa</b>	<b>Coleta de dados</b>	<b>Técnica de análise dos dados</b>
Identificar a variação da reputação corporativa (i.e., favorabilidade e visibilidade) das empresas listadas na B3 no período do primeiro trimestre de 2012 ao quarto trimestre de 2018.	–	Levantamento de dados secundários: D1) Empresas listadas na B3 (sítios: B3 – Empresas listadas (B3, 2019a) , B3 – Histórico de Cotações (B3, 2019c) - CVM – Dados Cadastrais – CIAS ABERTAS (CVM). D2) Levantamento de dados secundários: Texto de notícias <i>on-line</i> por meio de <i>web scraping</i> em sítios dos jornais <i>Folha de São Paulo</i> , <i>O Estado de São Paulo</i> e <i>Valor Econômico</i> (para detalhes ver seção 3.2.2.1).	Mineração de texto e análise de sentimento. Análise estatística descritiva dos dados.
Identificar a variação do desempenho financeiro baseado em indicadores contábeis e do desempenho financeiro baseado em indicadores de mercado das empresas listadas na B3 no período do primeiro trimestre de 2012 ao quarto trimestre de 2018.	–	Levantamento de dados secundários: D3) Métricas financeiras na base dados Economatica® (para detalhes ver seções 3.2.2 e 3.2.3).	Análise estatística descritiva dos dados.
Verificar como a reputação corporativa (i.e., favorabilidade e visibilidade) afeta o desempenho financeiro das empresas, o que é feito por meio de duas técnicas de regressão: linear com dados em painel e por intermédio de algoritmo de aprendizado de máquina supervisionado.	H1, H2, H3	Dados: D1, D2 e D3.	Análise de correlação, regressão linear com dados em painel e regressão por meio do algoritmo <i>Gradient Boosting Machines</i> .
Comparar o modelo preditivo, construído empregando técnicas de aprendizado de máquina, com um modelo construído utilizando técnicas de regressão linear.	–	Dados: D1, D2 e D3.	Teste T para médias, comparação do erro absoluto médio, erro quadrático médio e coeficiente de determinação obtido em cada modelo.

Fonte: Elaborado pelo autor com base em Telles (2001).

## 4 RESULTADOS

Este capítulo destina-se à apresentação dos resultados, analisando-os em comparação com a literatura. Ele está estruturado em quatro seções, a saber:

- Revisão sistemática da literatura: contém uma investigação focada em questões bem definidas, com a finalidade de identificar, selecionar, analisar a literatura sobre o tema deste estudo;
- Estatísticas descritivas: contempla as análises descritivas que visam descrever e sumarizar a amostra de dados deste trabalho;
- Análise da regressão: realiza o teste das hipóteses da pesquisa por meio da análise da significância dos coeficientes dos modelos de regressão linear com dados em painel e a avaliação dos resultados à luz da literatura pesquisada;
- Análise de regressão com algoritmo GBM e comparação com a regressão linear com dados em painel: efetua a comparação entre os modelos de regressão por intermédio do algoritmo GBM e os modelos de regressão linear com dados em painel, com respeito à acurácia e ao poder preditivo.

### 4.1 Revisão sistemática da literatura

A revisão sistemática da literatura foi executada de acordo com o protocolo apresentado no capítulo anterior. Os registros foram coletados em buscas realizadas em dois momentos: a primeira, em 1º de novembro de 2018, que coletou artigos que foram publicados entre janeiro de 2009 e outubro de 2018, e a segunda, em 23 de setembro de 2019, sendo que esta última teve a finalidade de coletar os artigos que foram publicados em novembro e dezembro de 2018. Os resultados obtidos em cada base de dados bibliográfica foram exportados para arquivos nos formatos descritos no Quadro 4 e os arquivos produzidos foram carregados em um banco de dados Microsoft® SQL Server, através de funcionalidades do *software* R. Os resultados obtidos após a execução das *strings* de busca em cada base de dados são mostrados na Tabela 1. Em seguida, 29 estudos obtidos em outras fontes bibliográficas foram gravados manualmente também nesse banco de dados.

Tabela 1 – Estudos identificados

<b>Base de dados</b>	<b>Número de registros</b>
Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)	6
EBSCO (Academic Search Premier e Business Source Complete)	1408
Emerald eJournals	690
Sage Journals	264
Scielo	3
Scopus	139
SpringerLink	566
Taylor Francis	406
Web of Science (Coleção principal)	172
Wiley Online Library	505
<b>TOTAL</b>	<b>4159</b>

Fonte: Elaborado pelo autor.

O processo selecionou 54 estudos primários, publicados entre janeiro de 2009 e dezembro de 2018, consistindo em artigos de conferências, artigos de periódicos, dissertações e teses. A seguir são apresentados os resultados intermediários de cada etapa.

Na etapa 1, 4.159 estudos foram coletados através da execução das *strings* de busca nas bases de dados bibliográficas selecionadas. Após a remoção dos estudos duplicados, esse número foi reduzido para 3.469. Adicionando outros 29 registros, obtidos em outras fontes de informação e que não haviam sido selecionados através das *strings* de busca, o resultado chegou a 3.498 estudos, os quais foram gravados em uma planilha eletrônica Microsoft® Excel.

Na etapa 2, a seleção foi realizada por meio da análise dos títulos de cada estudo, o que ocasionou a exclusão de 2.980 registros, reduzindo-os a 518. Considerando os critérios de inclusão e de exclusão, as remoções se deram pelas seguintes razões:

- (2.616) estudos não tratavam da relação reputação e desempenho;
- (311) estudos tratavam de responsabilidade social corporativa;
- (23) estudos tratavam apenas de reputação corporativa;
- (19) estudos não eram quantitativos (revisões de literatura);
- (9) estudos eram o sobre responsabilidade ambiental corporativa;
- (2) estudos não eram artigos de periódico ou de conferências.

Na etapa 3, a seleção foi realizada por meio da leitura do resumo, o que gerou a exclusão de 416 estudos, reduzindo a lista para 102 resultados. Considerando os critérios de inclusão e exclusão, os textos foram excluídos pelas seguintes razões:

- (164) estudos não tratavam da relação reputação e desempenho;

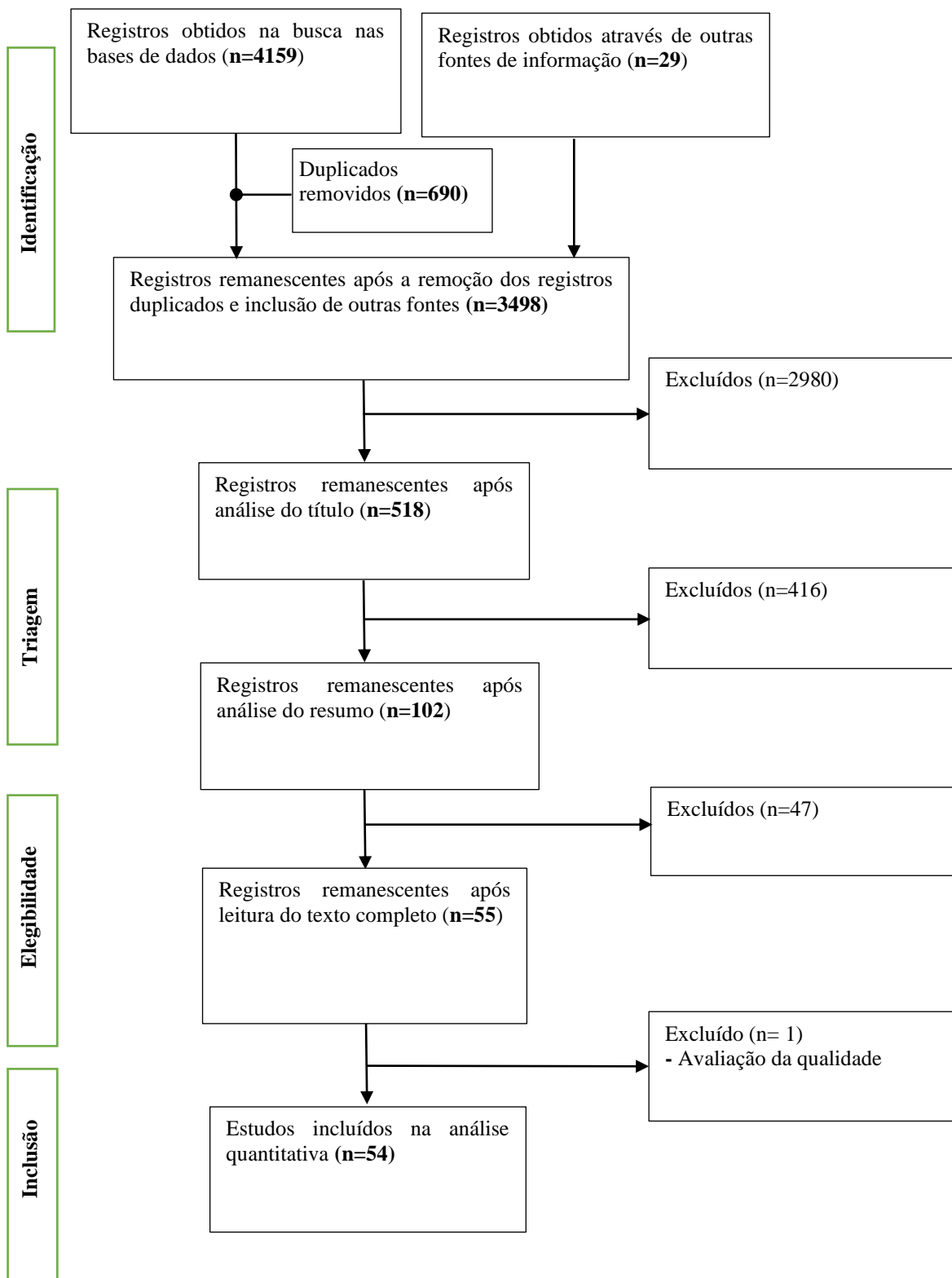
- (74) estudos tratavam da relação responsabilidade social corporativa e desempenho financeiro;
- (69) estudos tratavam apenas de responsabilidade social corporativa;
- (33) estudos tratavam apenas de reputação corporativa;
- (31) estudos tratavam da relação responsabilidade social corporativa e desempenho;
- (16) estudos tratavam da relação responsabilidade social corporativa e reputação;
- (13) estudos tratavam da relação sustentabilidade e desempenho;
- (3) estudos tratavam apenas de responsabilidade ambiental corporativa;
- (6) estudos não eram quantitativos (revisão de literatura, estudo de caso ou trabalho conceitual);
- (7) estudos não estavam disponíveis em texto completo para acesso.

Na etapa 4, a seleção foi feita por meio da análise do texto completo, o que teve como consequência a exclusão de 47 estudos. Dessa forma, os textos restantes totalizaram 55. Levando em conta os critérios de inclusão e exclusão, a recusa desses registros se deu pelas seguintes razões:

- (18) estudos não tratavam da relação reputação e desempenho;
- (8) estudos tratavam da relação da reputação com outro constructo;
- (5) estudos tratavam da relação do desempenho com outro constructo;
- (7) estudos não eram quantitativos;
- (7) estudos tratavam apenas da reputação corporativa;
- (2) estudos não estavam escritos em português ou inglês.

Finalmente, aplicou-se a avaliação da qualidade descrita no protocolo desta revisão sistemática. Nesta etapa, apenas um estudo foi excluído, uma vez que obteve nota inferior a quatro, de acordo com a definição de qualidade mínima do protocolo da revisão. A Figura 7 mostra as quatro etapas da seleção dos estudos primários, bem como, a etapa de verificação da qualidade dos estudos (Moher et al., 2015). A lista de estudos primários identificados está apresentada no Apêndice A e os resultados da avaliação da qualidade estão expostos no Apêndice B.

Figura 7 – Diagrama de fluxo da revisão sistemática



Fonte: Elaborado pelo autor com base em Moher et al. (2015).

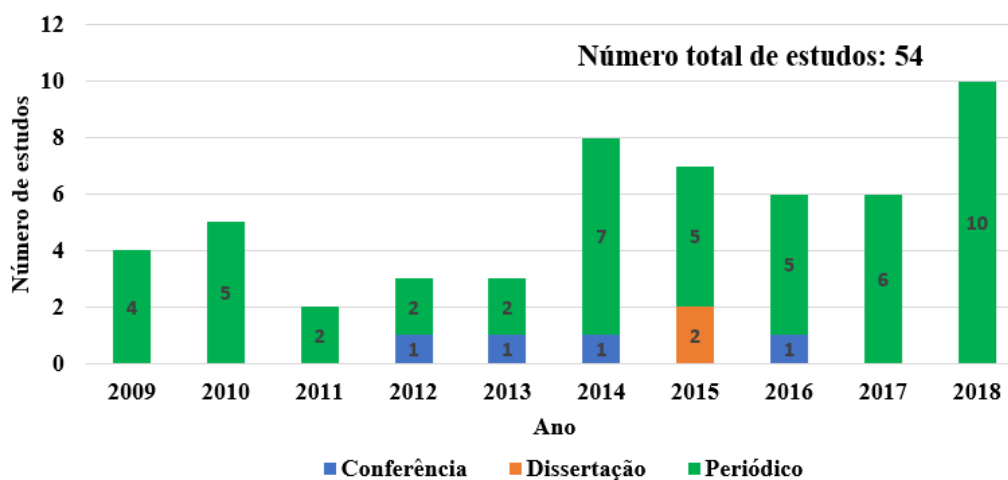
A seguir são apresentados os resultados da revisão sistemática com respeito às questões da revisão.



**Questão de revisão 1.** Quais são as pesquisas recentes (período de 2009 a 2018) que têm examinado as consequências da reputação corporativa no desempenho financeiro?

A demografia e a tendência de publicação dos estudos primários são mostradas na Figura 8. Dos 54 estudos primários selecionados: 88,9% são artigos de periódicos, 7,4% são artigos de conferência e 3,7% são dissertações de mestrado. No quinquênio 2014–2018, o número médio de publicações (igual a 7,4) dobrou em relação ao quinquênio 2009–2013 (igual a 3,4).

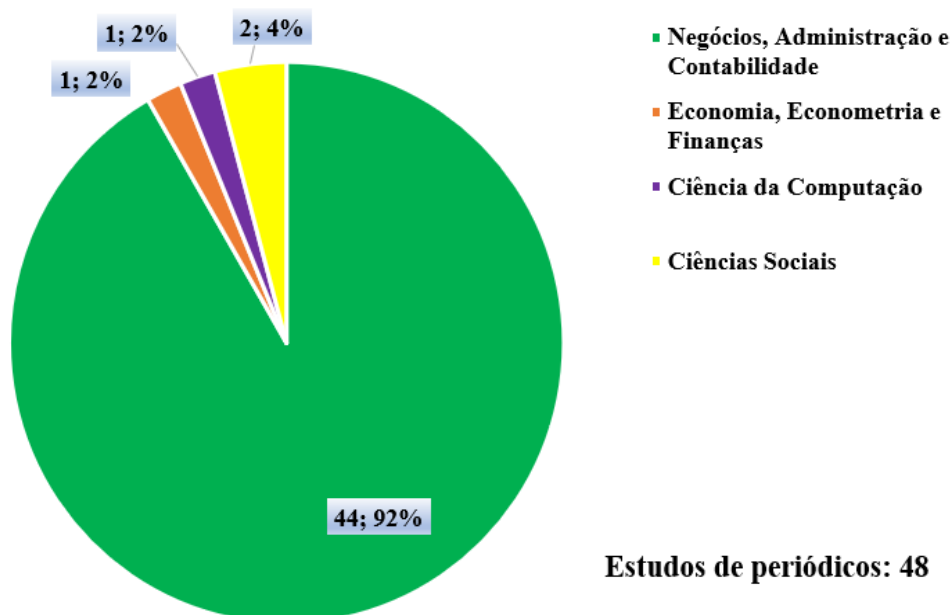
Figura 8 – Distribuição de estudos primários por anos (de 2009 até 2018)



Fonte: Elaborado pelo autor.

A maioria dos estudos primários selecionados teve revisão por pares (88,9%), uma vez que foram publicados em periódicos, predominantemente, da área temática de negócios, administração e contabilidade (92%), conforme exibido na Figura 9. Vale destacar que não foi definido nenhum critério de inclusão ou de exclusão relacionado com a área temática do periódico no qual o estudo foi publicado, assim foram selecionados: um estudo da área da ciência da computação; um estudo da área de economia, econometria e finanças; e dois estudos da área de ciências sociais.

Figura 9 – Áreas temáticas dos periódicos

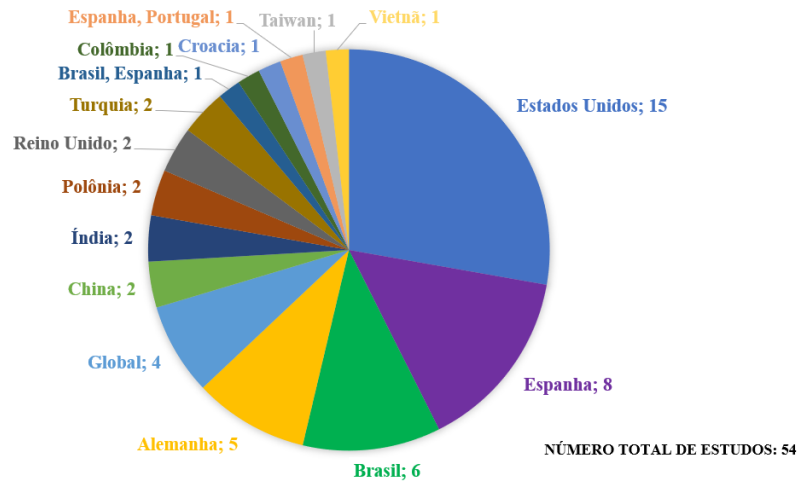


Fonte: Elaborado pelo autor.

Com relação à distribuição geográfica das empresas analisadas nos estudos primários, há predominância de empresas localizadas em países da OCDE<sup>15</sup> (Estados Unidos, Espanha, Alemanha, Reino Unido, Polônia, Turquia e Portugal), dos quais são originários 35 estudos (64,8%). Dentre as nações que não figuram como membros da OCDE, 6 estudos (11,1%) investigam empresas do Brasil e 4 (7,4%), que foram agrupados sob a etiqueta “Global”, analisaram empresas de vários países. Quando aos que tratam de organizações nacionais, essa quantidade de estudos foi influenciada pelas premissas desta revisão, uma vez que foram incluídas 2 dissertações de mestrado e 4 artigos publicados no país, os quais foram encontrados na revisão narrativa da literatura preliminar. Os resultados são exibidos na Figura 10.

<sup>15</sup> A *Organisation for Economic Co-operation and Development* (OCDE) possui 36 países membros (<http://www.oecd.org/about/membersandpartners/>).

Figura 10 – Países das empresas da amostra dos estudos

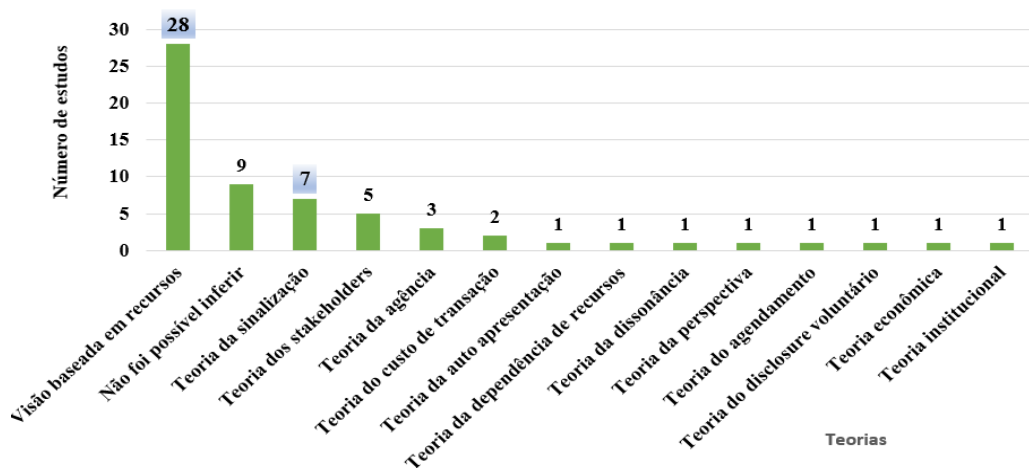


Fonte: Elaborado pelo autor.

**Questão de revisão 2.** Quais são as teorias empregadas para explicar a relação entre reputação corporativa e desempenho financeiro?

As teorias mais comumente usadas para explicar a relação reputação–desempenho são a da VBR e da Sinalização, o que está em linha com os resultados obtidos por Walker (2010). Com relação aos resultados obtidos, em 9 dos estudos selecionados não foi possível inferir qual teoria havia sido empregada pelos autores, pois os trabalhos não traziam a definição explícita da perspectiva teórica empregada. Quanto aos demais estudos, 8 utilizaram duas teorias como embasamento teórico. A Figura 11 mostra todas as teorias usadas como perspectiva dos estudos primários selecionados.

Figura 11 – Embasamento teórico dos estudos primários



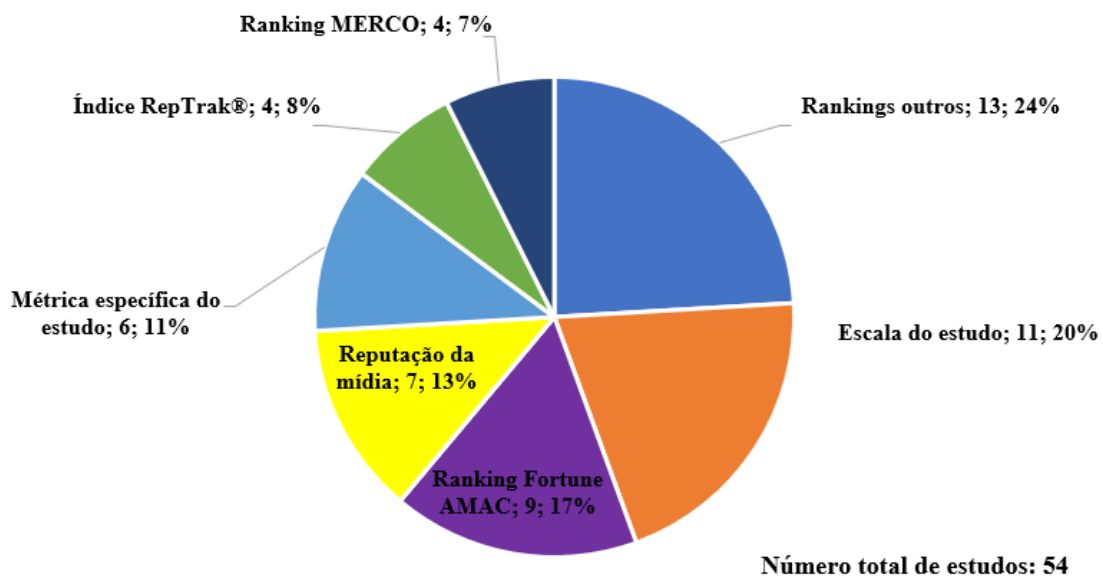
Número de estudos: 54 - 8 estudos usam 2 teorias como perspectiva os demais apenas 1 teoria.

Fonte: Elaborado pelo autor.

**Questão de revisão 3.** Quais são as medidas de reputação corporativa?

A maioria dos estudos primários selecionados, cerca de 48,1%, mede a reputação através de *rankings*, incluindo: *ranking* MERCO (*Monitor Empresarial de Reputación Corporativa*), *ranking* Fortune AMAC e outros *rankings* de reputação. O uso do *ranking* Fortune AMAC não foi tão frequente como o relatado nas revisões de literatura de Dowling (2016) e de Walker (2010). Quanto à medição da reputação corporativa por meio de informações da mídia, ela foi utilizada em 7 estudos primários (13% da amostra), todos publicados no quinquênio iniciado em 2014, o que poderia indicar uma tendência. As frequências relativas a cada abordagem para mensuração da reputação corporativa são exibidas na Figura 12.

Figura 12 – Abordagem para medição da reputação corporativa



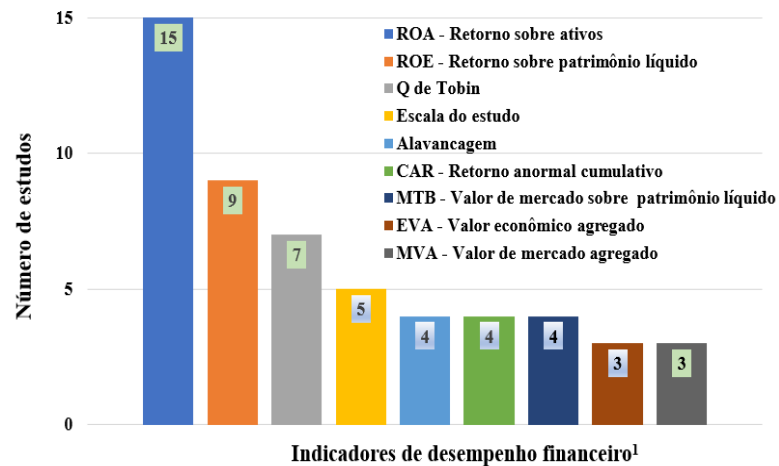
Fonte: Elaborado pelo autor.

**Questão de revisão 4.** Quais são os indicadores de desempenho financeiro que são relacionados com a reputação corporativa?

A Figura 13 exhibe os indicadores usados em 3 ou mais estudos primários para mensuração do desempenho financeiro das empresas. Cada estudo analisa um ou mais indicadores. Duas medidas de desempenho contábil são as mais utilizadas: o ROA (retorno sobre os ativos), que figura em 15 estudos (27,8%), e o ROE (retorno sobre patrimônio líquido), que aparece em 9 estudos (16,7%). A terceira medida de desempenho mais utilizada foi o indicador baseado em mercado Q de Tobin, presente em 7 estudos (13%). Vale destacar que a

alavancagem financeira aparece tanto como um efeito da reputação corporativa (variável dependente) quanto como uma variável de controle.

Figura 13 – Indicadores de desempenho financeiro dos estudos primários

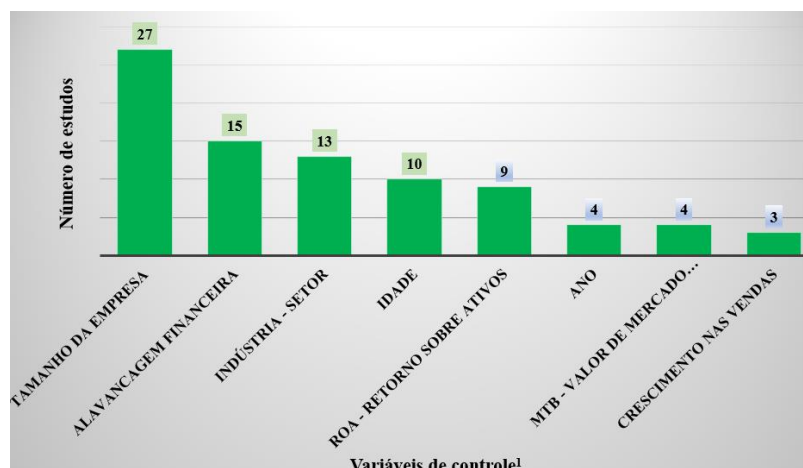


Nota: <sup>1)</sup> Não são exibidos indicadores empregados em menos de 3 estudos.  
Fonte: Elaborado pelo autor.

#### Questão de revisão 5. Quais são as variáveis de controle consideradas?

Apesar da possível existência de outros fatores que podem interferir na relação entre a reputação corporativa e desempenho financeiro, 15 estudos primários (29%) não controlam os efeitos de outras variáveis e 37 estudos primários (71%) que analisaram os efeitos de uma ou mais variáveis de controle. A Figura 14 exhibe as variáveis de controle que foram empregadas em 3 ou mais estudos da amostra. Tamanho, alavancagem financeira, indústria (setor) e idade são características das empresas que os estudos consideraram ter mais influência sobre a relação reputação–desempenho.

Figura 14 – Variáveis de controle analisadas nos estudos primários

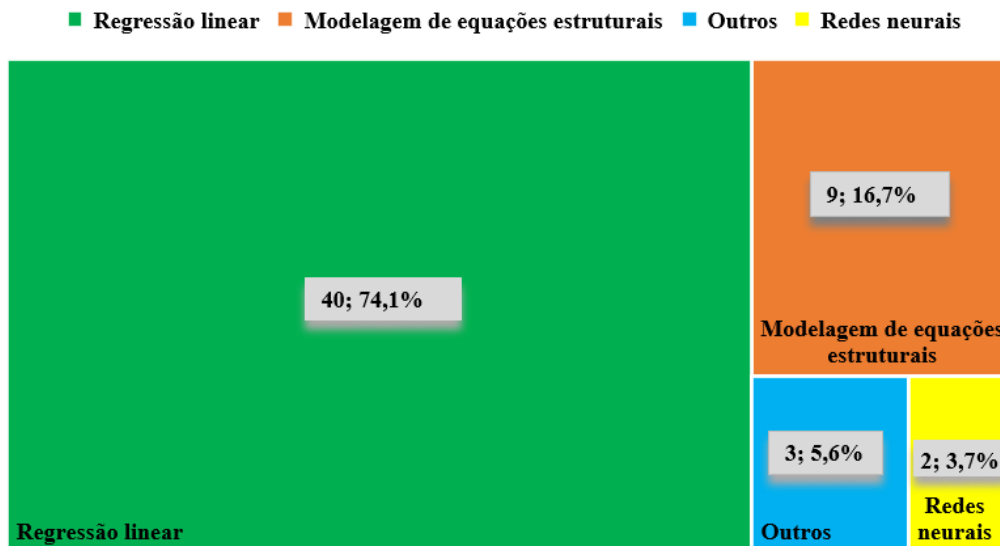


Nota: <sup>1)</sup> Não são exibidas as variáveis de controle usadas em menos de 3 estudos.  
 Fonte: Elaborado pelo autor.

**Questão de revisão 6.** Quais são os métodos quantitativos utilizados para avaliar a relação entre reputação corporativa e desempenho financeiro?

O método da regressão linear (regressão linear múltipla através de mínimos quadrados parciais) é predominante na amostra, sendo usado por 38 estudos primários (74,1%), em seguida, 9 estudos (16,7%) utilizam a modelagem de equações estruturais. Outros métodos representam 3 estudos (5,6%), entre os quais a metodologia de estudo de eventos, o agrupamento *K-means* e a análise de correlação. Finalmente, 2 estudos (3,7%) empregam regressão através de redes neurais. Com relação a eles, ambos os estudos argumentam que os resultados obtidos por meio dessa técnica de regressão são mais robustos que aqueles obtidos por meio de regressão linear (Casado, Yanez, & Pelaez, 2016; Fernández-Gámez et al., 2016). O uso predominante do método da regressão linear é decorrente do emprego de dados secundários tanto para mensuração da reputação como para a medição do desempenho financeiro. O mapa da Figura 15 representa o emprego dos métodos quantitativos nos estudos primários.

Figura 15 – Metodologia de análise quantitativa



Fonte: Elaborado pelo autor.

## 4.2 Estatísticas descritivas

Seguindo o procedimento descrito no capítulo de metodologia, foram selecionadas as notícias relevantes relativas às 420 empresas de capital aberto listadas na B3, as quais constituem a população da pesquisa, e foi calculado o teor de cada uma utilizando os quatro dicionários de sentimentos analisados no estudo: Oplexicon (M. Souza, Vieira, Buseti, Chishman, & Alves, 2011), Sentilex-PT (Silva, Carvalho, & Sarmiento, 2012), Loughran e McDonald (Loughran & McDonald, 2011) e o léxico de sentimentos de Bing Liu (M. Hu & Liu, 2004). Esse procedimento produziu um conjunto de dados com 51.597 notícias relevantes. Desse conjunto de dados foi extraída uma amostra aleatória de 510 notícias, as quais foram acessadas novamente, para leitura dos textos e para classificação pelo pesquisador como favoráveis, neutras e desfavoráveis. A dimensão da amostra foi obtida por intermédio da tabela de tamanho da amostra para proporções multinominais apresentada por Thompson (1987) para uma população que possui 3 categorias (favorável, neutra e desfavorável) com grau de confiança de 95% e margem de erro de 5%. A Tabela 2 mostra que a classificação automática de teor da amostra de 510 notícias, produzida a partir do emprego do léxico de sentimentos de Bing Liu, obteve maior concordância em relação à classificação manual, desse modo, o cômputo da variável independente favorabilidade (FAV) foi feito com o teor calculado utilizando esse léxico.

Tabela 2 – Aderência da classificação automática de teor com a manual

<b>Léxico de Sentimentos</b>	<b>Acurácia da classificação automática (%)</b>
Bing Liu	44,1
Loughran e McDonald	38,6
Oplexicon 3.0	36,3
Sentilex-PT 2.0	35,3

Fonte: Elaborado pelo autor.

Cabe salientar que o processo de seleção, leitura e cálculo de teor das notícias relevantes foi realizado entre 4 de agosto e 9 de setembro de 2019. O código em R apresentado no Apêndice D para realizar esse procedimento está ajustado para a versão dos sítios dos jornais nesse período.

Van den Bogaerd e Aerts (2015) argumentam que o cálculo da favorabilidade requer um número mínimo de notícias e descrevem, também, que empresas com menos de 30 notícias coletadas durante um período de cinco anos são eliminadas da amostra em seu estudo, ou seja,

para esses autores, a fim de que a amostra seja relevante para o cálculo, são necessárias ao menos 6 matérias jornalísticas por ano. De maneira similar, este estudo excluiu da amostra empresas com menos de 140 notícias relevantes coletadas durante 28 trimestres, ou seja, cerca de 5 notícias por trimestre, ou 20 notícias por ano. Após aplicação do critério de seleção, o conjunto de dados resultante continha 67 empresas e um total de 44.417 notícias relevantes (662 notícias por empresa, em média). Após o cálculo das variáveis favorabilidade (FAV) e visibilidade (VIS), conforme exposto na seção 3.2.2, da junção dessa coleção de dados com a coleção de dados dos indicadores financeiros coletados, e da remoção das observações com patrimônio líquido negativo, o conjunto de dados obtido continha observações de 65 empresas durante o período de 28 trimestres, resultando em 1.597 observações (empresa-trimestres). Cabe destacar, no entanto, que duas empresas não possuíam valor de mercado em todos os trimestres na base de dados Economatica<sup>®</sup> e, por isso, não foram consideradas na análise, devido à impossibilidade de cálculo dos indicadores de desempenho de mercado: MVA<sup>®</sup> e Q de Tobin. Uma característica, ainda, sobre o conjunto de dados da pesquisa é que ele representa um painel desbalanceado, uma vez que cada empresa tem um número diferente de observações (entre 5 e 28) e, também, que esse painel é classificado como curto (Cameron & Trivedi, 2009). A lista das empresas da amostra é apresentada no Apêndice C.

Antes de prosseguir para os resultados da regressão linear, que visa atender o objetivo principal desta pesquisa, realizou-se uma análise estatística descritiva das variáveis empregadas no modelo empírico do estudo. Isto posto, a Tabela 3 apresenta o resultado dessas estatísticas descritivas, não sendo exibidas as variáveis *dummy*: IND1, IND2, NM, N2 e N1.

Tabela 3 – Estatísticas descritivas

Variável	N	Mínimo	Máximo	Média	Mediana	Desvio Padrão	Coefficiente de Variação	Assimetria
<b>FAV</b>	1597	-1,00	1,00	0,10	0,07	0,32	3,27	-0,34
<b>VIS</b>	1597	0,00	6,54	2,46	2,40	1,13	0,46	0,32
<b>FAVxVIS</b>	1597	-1,39	2,47	0,26	0,15	0,55	2,15	0,54
<b>ROA (%)</b>	1597	-6,91	7,16	0,83	0,70	2,03	2,45	-0,38
<b>ROE (%)</b>	1594	-21,99	21,40	2,28	2,33	6,68	2,93	-0,60
<b>MVA<sup>®</sup></b>	1597	25,28	26,75	25,98	25,96	0,16	0,01	1,08
<b>QTB</b>	1597	0,08	4,24	1,00	0,80	0,75	0,75	2,29
<b>ALV</b>	1597	0,21	0,95	0,64	0,65	0,18	0,28	-0,19
<b>IDA</b>	1597	7,36	10,66	9,27	9,33	0,79	0,08	-0,07
<b>TAM</b>	1597	20,50	28,13	24,02	23,85	1,50	0,06	0,92

Nota: As variáveis ROA, MVA<sup>®</sup>, QTB e ALV foram suavizadas (*winsorizing*) ao nível de 1% e a variável ROE ao nível de 2%.

Fonte: Elaborado pelo autor.



De acordo com os dados na Tabela 3, a alavancagem financeira (ALV) das empresas da amostra tem um valor médio de 64%, sendo que 50% dessas empresas tem alavancagem superior a 65%. Tal resultado sugere que essas empresas preferem compartilhar o risco dos investimentos com terceiros a fim de acelerar seu crescimento, mesmo levando em consideração o elevado patamar das taxas de juros praticadas no Brasil.

Com relação à idade das empresas (IDA), ela apresenta uma distribuição aproximadamente simétrica, com média (9,27) bem próxima da mediana (9,23) e coeficiente de assimetria próximo a zero (-0,07), o que se explica devido à transformação por meio da função logaritmo natural que foi aplicada.

Quanto ao tamanho (TAM), a transformação logaritmo natural que foi utilizada tornou a distribuição dessa variável aproximadamente simétrica, com média (24,02) bem próxima da mediana (23,85) e coeficiente de variação próximo a zero (0,06).

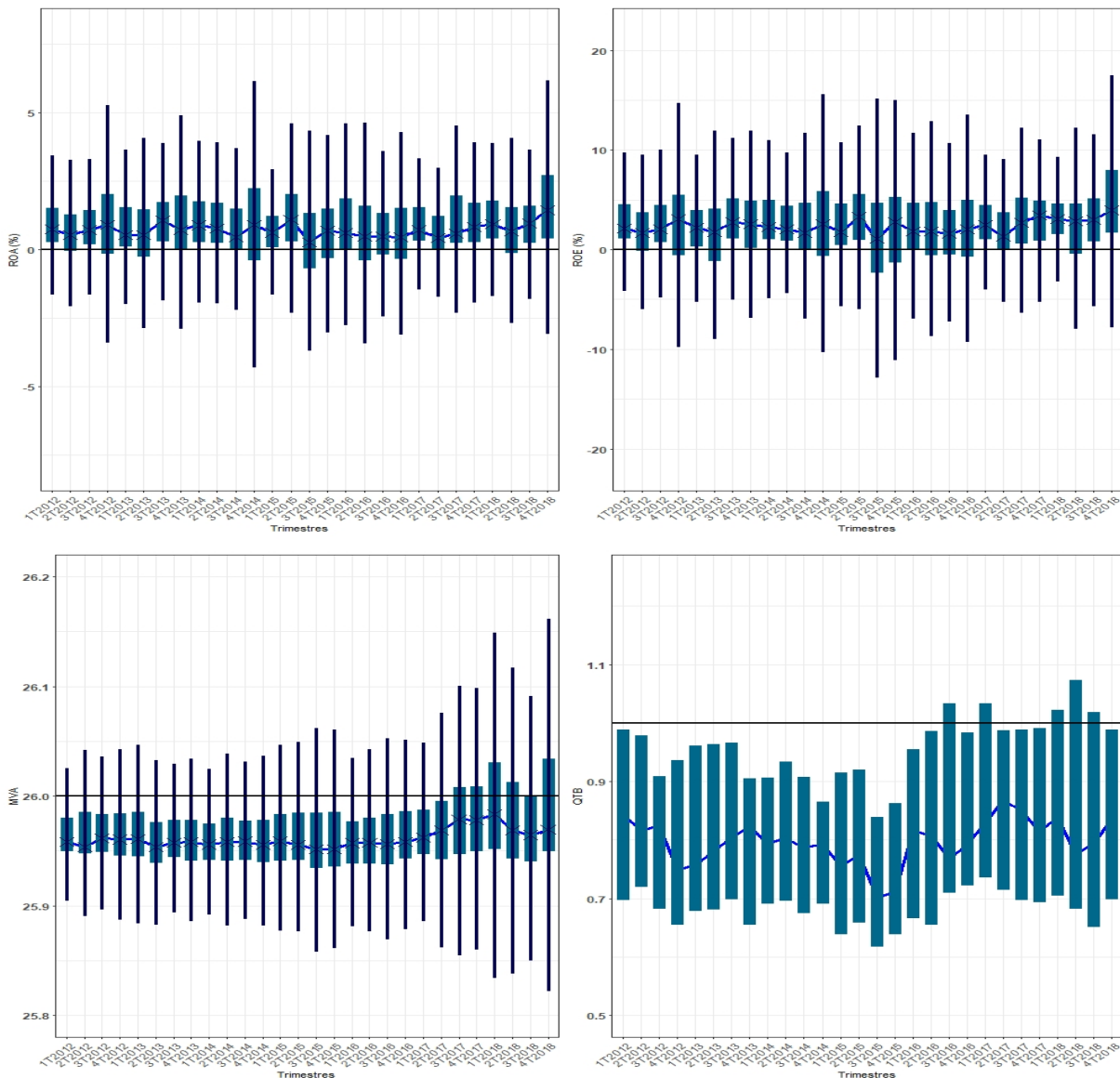
No que se refere às variáveis independentes, pode-se perceber que a média da favorabilidade (FAV) é de 0,10, indicando que as notícias relevantes têm um teor predominantemente positivo. Quanto ao valor mínimo, de -1,00, ele apenas aponta que em certos trimestres todas as notícias relevantes a respeito de alguma empresa da amostra foram desfavoráveis. De maneira análoga, o valor máximo, de 1,00, assinala que em certos trimestres todas as notícias relevantes sobre alguma empresa da amostra foram favoráveis. Com respeito à visibilidade (VIS), ela se apresentou como aproximadamente simétrica (assimetria igual a 0,32), conforme o esperado, uma vez que foi aplicada a transformação por meio do logaritmo natural a essa variável. Seu valor mínimo de 0,00 indica que, ainda que a amostra contenha uma média de 24 notícias por empresa por trimestre, em certos trimestres a favorabilidade (FAV) de alguma empresa foi calculada com apenas uma notícia relevante, fato que não poderia ser evitado, dado o limiar de pelo menos 140 notícias relevantes por empresa durante os 28 trimestres da pesquisa.

Ainda analisando os dados, vê-se que os indicadores de desempenho financeiro, Q de Tobin (QTB), ROA e ROE, que são usados como variável dependente do estudo, apresentaram grande dispersão, com coeficientes de variação de 75%, 245% e 293%, respectivamente, o que permite afirmar que as 65 empresas da amostra apresentam um desempenho financeiro bastante heterogêneo. Cumpre salientar que foi aplicado tratamento de suavização (*winsorizing*) ao nível de 1% nas variáveis dependentes ROA e Q de Tobin e ao nível de 2% na variável dependente ROE, pois ela apresentava alta dispersão com um coeficiente de variação de 688% antes do tratamento dos valores extremos. No que se refere ao indicador Q de Tobin (QTB), seu valor médio, igual a 1,00, indica que o mercado valoriza de maneira justa os ativos das empresas da

amostra, isto é, o valor de mercado dessas empresas é aproximadamente igual ao custo de reposição de seus ativos tangíveis, em média. Quanto ao valor médio dos indicadores de desempenho contábeis ROA (0,83%) e ROE (2,28%), eles mostram que, em média, as empresas da amostra mantiveram sua lucratividade positiva, mesmo num ambiente econômico adverso, como o que havia no Brasil durante o período do estudo, o qual foi marcado pela recessão econômica, nos anos 2015 e 2016, e pelo baixo crescimento, nos anos de 2017 e 2018.

A heterogeneidade entre trimestres do desempenho financeiro das empresas da amostra pode ser verificada na Figura 16, que exhibe os indicadores de desempenho financeiro por meio do *boxplot* em cada trimestre.

Figura 16 – Dados em painel – heterogeneidade entre trimestres

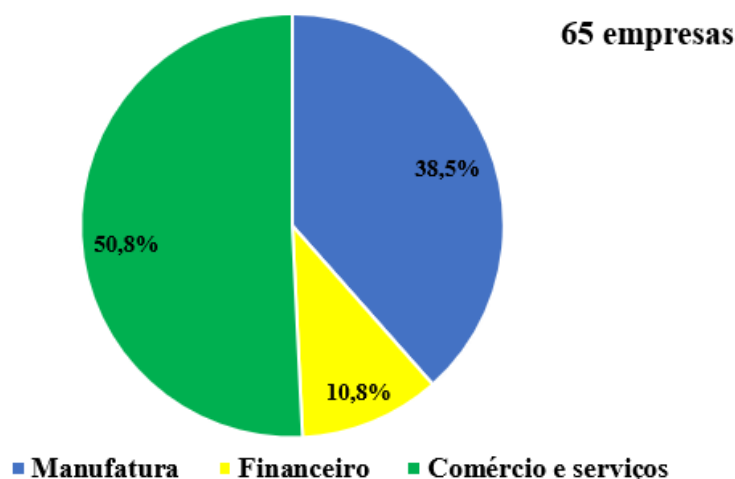


Fonte: Elaborado pelo autor.



Com relação às variáveis *dummy* para a indústria (setor) da empresa, não apresentadas na Tabela 3, os valores das variáveis IND1 e IND2 indicam que: 50,8% das empresas da amostra pertencem ao setor de manufatura (IND1 = 1 e IND2 = 0), 10,8% das empresas pertencem ao setor financeiro (IND1 = 0 e IND2 = 1) e 38,5% das empresas, ao setor de comércio e serviços (IND1 = 0 e IND2 = 0). A Figura 18 exibe esses resultados. A amostra possui uma menor proporção de empresas da indústria financeira em relação à proporção de 20,6% na população de todas as empresas listadas na B3, o que mostra que essas empresas não têm tanta proeminência na mídia. Vale mencionar que o setor financeiro na B3 é composto por securitizadoras, por empresas do ramo imobiliário e por intermediários financeiros, como bancos, financeiras, empresas de arrendamento mercantil, *leasing* e consórcios.

Figura 18 – Setores das empresas da amostra



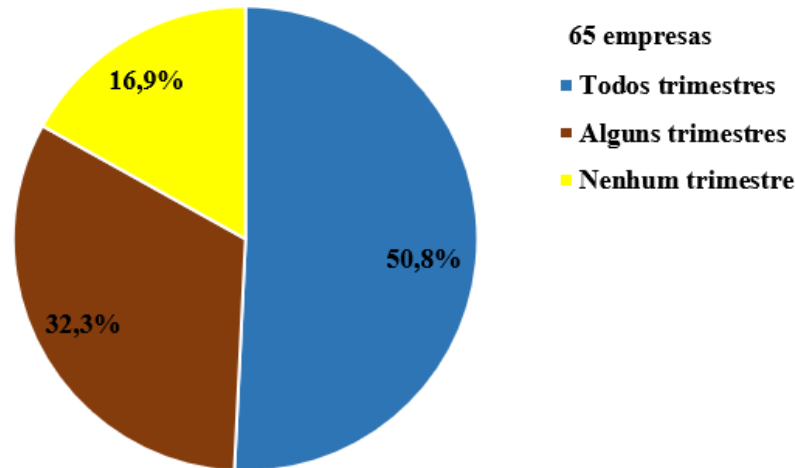
Fonte: Elaborado pelo autor.

Por fim, com relação aos níveis diferenciados de governança corporativa da B3 relativos a dezembro de 2018, cerca de 61,5% das empresas da amostra pertencem ao nível Novo Mercado de governança (NM = 1), 6,2% pertencem ao Nível 2 de governança (N2 = 1), 20,0% pertencem ao Nível 1 de governança (N1 = 1) e 12,3% das empresas pertencem ao mercado tradicional. Compre mencionar que a amostra não continha nenhuma empresa dos segmentos Bovespa Mais e do Bovespa Mais Nível 2.

Vale salientar que a maioria das empresas da amostra fazem, ou fizeram parte, da carteira teórica do índice Bovespa (Ibovespa) durante o período de sete anos do estudo, que vai do primeiro trimestre de 2012 até o quarto trimestre de 2018, de modo que, ao longo desse período, 33 empresas permaneceram, ininterruptamente, nessa carteira, 21 permaneceram pelo

menos um trimestre nela e apenas 11 empresas não participaram da carteira teórica do Ibovespa em nenhum trimestre. Essa constatação corrobora o mecanismo de modelagem do viés de seleção da amostra, apresentado na Equação 4. Tais informações estão mostradas na Figura 19.

Figura 19 – Presença no índice Bovespa das empresas da amostra



Fonte: Elaborado pelo autor.

Uma vez apresentada a análise estatística descritiva dos dados, passou-se à análise de correlação entre as variáveis da pesquisa, objetivando verificar se existe associação linear entre elas. Para isso, foram utilizados os coeficientes de correlação de Pearson. Os resultados obtidos são exibidos na Tabela 4. Ao observar os dados, vê-se que a favorabilidade (FAV) está correlacionada positivamente com todos indicadores de desempenho financeiro ROA, ROE, MVA<sup>®</sup> e Q de Tobin (QTB), com níveis de significância de 1%. Essa relação positiva sugere que quanto maior a favorabilidade, maior o desempenho financeiro. Quanto à visibilidade (VIS), ela está correlacionada negativamente com todos os indicadores de desempenho financeiro, com níveis de significância de 1%. Nesse caso a relação negativa sugere que quando maior a visibilidade, menor o desempenho financeiro medido por meio desses indicadores. A amostra não apresenta nenhum par de variáveis de controle com valores de correlação forte (de 0,7 a 0,89) ou muito forte (de 0,9 a 1,0) segundo a classificação de magnitude de Pett, Lackey, e Sullivan (2003). Cumpre salientar que a correlação fortemente significativa de 0,88 entre ROA e ROE não consiste um problema, uma vez que ambas são variáveis dependentes que não são incluídas em um mesmo modelo de regressão. Mesmo não havendo indicação de multicolinearidade, conforme mencionado no capítulo de metodologia, essa situação deve ser analisada por meio do coeficiente VIF (*Variance Inflation Factor*).

Tabela 4 – Coeficientes de correlação de Pearson entre as variáveis

	ROA	ROE	MVA <sup>®</sup>	QTB	FAV	VIS	FAVxVIS	ALV	IDA	TAM	IMR	IND1	IND2	NM	N1
<b>ROA</b>															
<b>ROE</b>	0,88**														
<b>MVA<sup>®</sup></b>	0,24**	0,21**													
<b>QTB</b>	0,49**	0,39**	0,44**												
<b>FAV</b>	0,14**	0,12**	0,12**	0,19**											
<b>VIS</b>	-0,15**	-0,11**	-0,13***	-0,13**	0,05*										
<b>FAVxVIS</b>	0,15**	0,14**	0,21**	0,19**	0,77**	0,11**									
<b>ALV</b>	-0,28**	-0,05	0,04	-0,20**	-0,02	0,06*	0,04								
<b>IDA</b>	-0,10**	0,04	-0,11**	-0,22**	-0,04	0,19**	-0,05	0,30**							
<b>TAM</b>	-0,18**	-0,06*	0,07**	-0,33**	0,00	0,54**	0,07**	0,28**	0,37**						
<b>IMR</b>	0,11**	0,04	-0,09**	0,18**	0,02	-0,40***	0,01	-0,03	-0,07**	-0,70**					
<b>IND1</b>	-0,11**	-0,09**	0,00	-0,07**	0,00	0,22**	0,01	-0,06*	0,09**	0,08**	-0,23**				
<b>IND2</b>	-0,05*	0,05*	0,20**	-0,20***	0,06*	0,14**	0,12**	0,21**	0,07**	0,55**	-0,27**	-0,28**			
<b>NM</b>	0,08**	0,05*	-0,03	0,14**	0,01	-0,27**	-0,02	-0,10**	-0,50**	-0,38**	0,13**	-0,06*	-0,06*		
<b>N1</b>	-0,11**	-0,05	-0,02	-0,24**	-0,02	0,23**	-0,01	-0,03	0,41**	0,31**	-0,18**	0,12**	0,06*	-0,65**	
<b>N2</b>	-0,09**	-0,09**	-0,03	-0,07**	0,04	0,04	0,03	0,18**	0,11**	-0,02	0,09*	-0,08**	0,00	-0,29**	-0,15**

Nota: (\*\*\*)  $p < 0,01$ ; (\*\*)  $p < 0,05$  (bicaudal). As variáveis ROA, MVA<sup>®</sup>, QTB e ALV foram suavizadas (*winsorizing*) ao nível de 1% e a variável ROE ao nível de 2%.

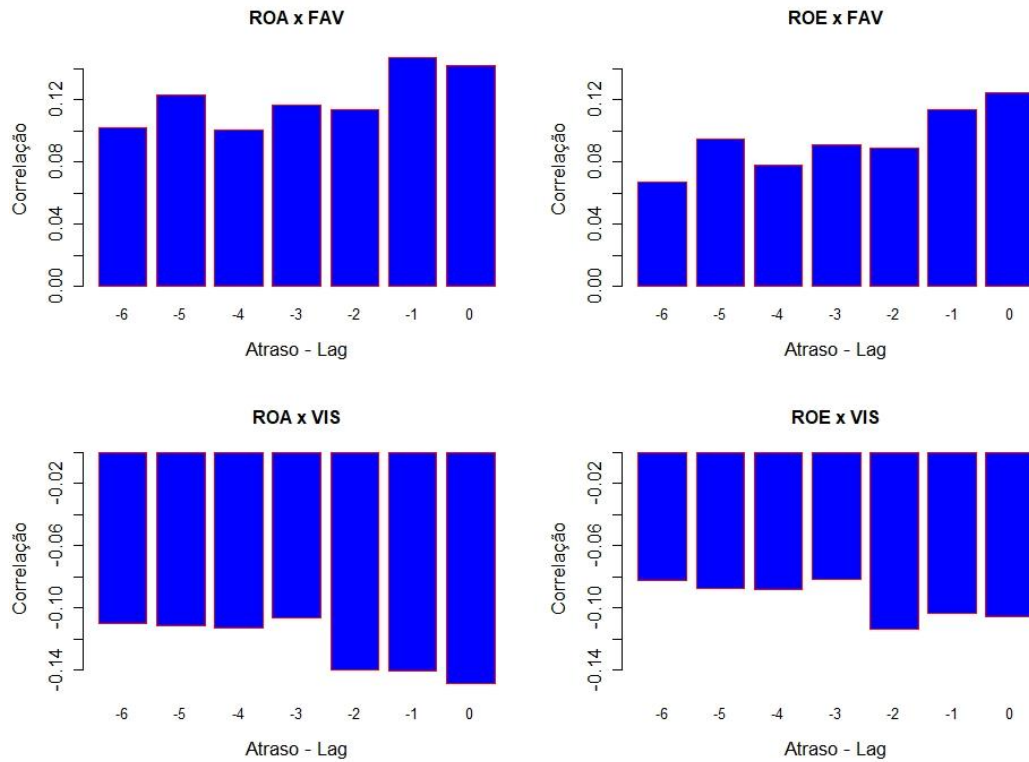
Fonte: Elaborado pelo autor.

Os resultados da correlação entre a variável dependente e as variáveis independentes, mostrados na Tabela 4, apesar de revelarem uma correlação fraca e muito significativa, estão em linha com os resultados de estudos anteriores que examinaram reputação corporativa e mídia. Deephouse (2000), por exemplo, obteve uma correlação entre favorabilidade e ROA de 0,14 ( $p\text{-value} < 5\%$ ) e o estudo de Agnihotri (2014), no mesmo sentido, apresentou uma correlação entre favorabilidade e Q de Tobin de 0,36 ( $p\text{-value} < 10\%$ ). Outros estudos, que não empregaram indicadores de desempenho financeiro similares aos utilizados nesta pesquisa, também revelaram uma relação fraca entre as variáveis independentes e dependentes. Nessa direção, Pfarrer et al. (2010) encontraram uma correlação entre favorabilidade e o retorno anormal cumulativo (CAR) de 0,05 ( $p\text{-value} < 5\%$ ) e OuYang et al. (2017) obtiveram uma correlação entre CAR e favorabilidade de 0,287 ( $p\text{-value} < 1\%$ ) e entre CAR e visibilidade de 0,147 (não significativa).

Em seguida, uma análise de correlação cruzada foi efetuada para entender se existe uma defasagem temporal entre os indicadores de desempenho financeiro e as variáveis independentes (FAV e VIS). A função de correlação cruzada *ccf*, presente no pacote *stats* do software R (R Core Team, 2019), foi empregada nessa tarefa. Nas análises, foram considerados valores de defasagem de até seis trimestres, o que foi feito em decorrência da dimensão temporal limitada da amostra e, também, para excluir achados que não eram devidos a uma correlação real entre o par de variáveis, mas que, provavelmente, representavam um resultado aleatório.

Não foram encontradas evidências nas correlações cruzadas das variáveis de desempenho financeiro contábil, presentes na Figura 20, que possam justificar a necessidade de incluir atraso nas variáveis independentes, seja no modelo que possui como variável dependente o indicador ROA, seja no modelo que possui como variável dependente o indicador ROE.

Figura 20 – Correlação cruzada entre variáveis ROA e ROE e variáveis independentes

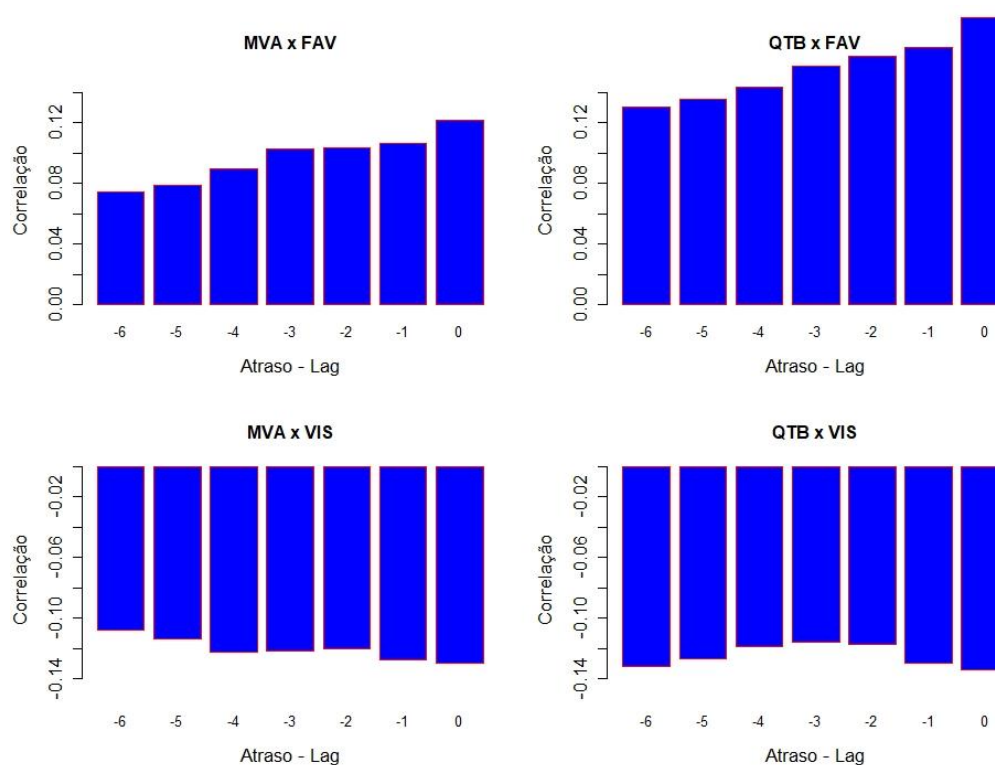


Fonte: Elaborado pelo autor.

Da mesma maneira, como se vê na Figura 21, não foram encontrados indícios nas correlações cruzadas das variáveis de desempenho financeiro baseado em indicadores de mercado que possam justificar a necessidade de incluir atraso nas variáveis independentes, seja no modelo que possui como variável dependente o indicador MVA<sup>®</sup>, seja no modelo que possui como variável dependente o indicador Q de Tobin (QTB).



Figura 21 – Correlação cruzada entre variáveis MVA<sup>®</sup> e Q de Tobin (QTB) e variáveis independentes



Fonte: Elaborado pelo autor.

#### 4.3 Análise de regressão

Como descrito na metodologia, a etapa seguinte na análise de dados refere-se à realização de testes para determinar qual o modelo de dados em painel é mais adequado. Com relação a isso, os resultados apresentados na Tabela 5 indicam que o modelo de efeitos fixos é o melhor para os indicadores de mercado (MVA<sup>®</sup> e Q de Tobin), uma vez que tanto a hipótese nula do teste de Hausman (efeitos aleatórios x efeitos fixos) foi rejeitada, indicando que o modelo de efeitos fixos é mais adequado, quanto a hipótese nula do teste de Chow (POLS x efeitos fixos) foi rejeitada, confirmando que o modelo de efeitos fixos é o mais adequado. Por outro lado, o modelo POLS é o melhor para os indicadores contábeis (ROA e ROE), uma vez que a hipótese nula do teste de Hausman (efeitos aleatórios x efeitos fixos) foi rejeitada, indicando que o modelo de efeitos fixos é mais adequado, porém a hipótese nula do teste de Chow (POLS x efeitos fixos) não foi rejeitada, determinando que o modelo POLS é o mais apropriado. Vale destacar que não foi necessário executar o teste LM – Breusch-Pagan, já que o teste de Hausman descartou a necessidade de emprego de um modelo com efeitos aleatórios.

Tabela 5 – Testes para determinação do modelo de dados em painel mais apropriado

Variável dependente	Teste de Hausman			Teste de Chow		
	<i>p-value</i>	Rejeita H0 (Aleatório)	Modelo	<i>p-value</i>	Rejeita H0 (POLS)	Modelo
ROA	< 0,001	Sim	Efeitos Fixos	0,927	Não	POLS
ROE	< 0,001	Sim	Efeitos Fixos	0,465	Não	POLS
MVA <sup>®</sup>	< 0,001	Sim	Efeitos Fixos	< 0,001	Sim	Efeitos Fixos
QTB	< 0,001	Sim	Efeitos Fixos	< 0,001	Sim	Efeitos Fixos

Fonte: Elaborado pelo autor.

Posteriormente, foram analisados os pressupostos da regressão linear. A Tabela 6 exhibe o resultado dos testes para verificação dessas premissas para os modelos completos, que incluem tanto as variáveis independentes FAV e VIS como o termo de interação FAVxVIS.

Tabela 6 – Testes dos pressupostos da regressão linear (modelo com termo de interação)

Variável	Teste de Breusch-Pagan		Teste de Breusch-Godfrey		Teste K-S (Dcrítico= 0,04)		Maior valor de VIF
	<i>p-value</i>	Rejeita hipótese H0 - Homocedasticidade	<i>p-value</i>	Rejeita hipótese H0 - Sem Auto correlação	Estat. D	<i>p-value</i>	
ROA	< 0,001	Sim	< 0,001	Sim	0,080	< 0,001	4,922
ROE	< 0,001	Sim	< 0,001	Sim	0,284	< 0,001	4,922
MVA <sup>®</sup>	< 0,001	Não	< 0,001	Sim	0,384	< 0,001	4,922
QTB	< 0,001	Sim	< 0,001	Sim	0,190	< 0,001	4,922

Fonte: Elaborado pelo autor.

A Tabela 7 exhibe o resultado dos testes para verificação das premissas da regressão linear dos modelos que incluem as variáveis independentes FAV e VIS.

Tabela 7 – Testes dos pressupostos da regressão linear (modelo com variáveis independentes)

Variável	Teste de Breusch-Pagan		Teste de Breusch-Godfrey		Teste K-S (Dcrítico= 0,04)		Maior valor de VIF
	<i>p-value</i>	Rejeita hipótese H0 - Homocedasticidade	<i>p-value</i>	Rejeita hipótese H0 - Sem Auto correlação	Estat. D	<i>p-value</i>	
ROA	< 0,001	Sim	< 0,001	Sim	0,078	< 0,001	4,914
ROE	< 0,001	Sim	< 0,001	Sim	0,287	< 0,001	4,914
MVA <sup>®</sup>	< 0,001	Não	< 0,001	Sim	0,382	< 0,001	4,914
QTB	< 0,001	Sim	< 0,001	Sim	0,190	< 0,001	4,914

Fonte: Elaborado pelo autor.

De acordo com as Tabelas 6 e 7, verifica-se que o requisito de distribuição normal dos resíduos não é satisfeito para todos os modelos testados, uma vez que os valores da estatística D do teste de Kolmogorov-Smirnov são superiores ao valor crítico. Entretanto, a violação desse

pressuposto não tem efeitos relevantes nos resultados, quando o número de observações na amostra por variável é maior do que 10 (Schmidt & Finan, 2018). Além disso, a hipótese nula do teste de Breusch-Pagan de que os resíduos possuem variância constante foi rejeitada para todos os modelos, assinalando problemas de heterocedasticidade. Quanto à correlação serial (auto correlação), a hipótese nula do teste de Breusch-Godfrey de ausência de auto correlação dos resíduos também foi rejeitada para todos os modelos. Com base nessas evidências, foram empregados estimadores robustos para ajustar os resultados das regressões. Finalmente, o maior valor do indicador VIF obtido para todos os modelos foi inferior a 5 o que indica que o pressuposto de não haver colinearidade entre os preditores foi atendido (Fávero et al., 2009).

Após a verificação das premissas da regressão linear e da determinação do melhor modelo de dados em painel, realizou-se a análise de regressão, cujo propósito foi testar as hipóteses da pesquisa. Vale salientar que foi feita a correção dos modelos empregando estimadores robustos HAC (*Heteroskedasticity and Autocorrelation Consistent covariance matrix*) devido aos problemas de heterocedasticidade e de correlação serial constatados.

A Tabela 8 mostra os resultados da análise de dados em painel POLS para a variável dependente ROA.

Tabela 8 – Análise de dados em painel do indicador de desempenho ROA

Variáveis	Sinal Esperado	Modelo 1		Modelo 2		Modelo 3	
		Coef.	p-value	Coef.	p-value	Coef.	p-value
<b>Intercepto</b>	+ / -	<b>6,960 *</b>	<b>0,037</b>	5,590	0,107	5,658	0,089
<b>IDA</b>	+	0,262	0,243	0,260	0,231	0,287	0,171
<b>TAM</b>	+	-0,235	0,140	-0,165	0,344	-0,177	0,294
<b>ALV</b>	+ / -	<b>-3,191 ***</b>	<b>&lt; 0,001</b>	<b>-3,207 ***</b>	<b>&lt; 0,001</b>	<b>-3,276</b>	<b>&lt; 0,001</b>
<b>IND1</b>	+ / -	-0,432	0,161	-0,415	0,158	-0,435	0,132
<b>IND2</b>	+ / -	0,384	0,443	0,228	0,622	0,167	0,703
<b>NM</b>	+	-0,555	0,236	-0,560	0,218	-0,538	0,225
<b>N1</b>	+	<b>-1,028 *</b>	<b>0,025</b>	<b>-1,005 *</b>	<b>0,021</b>	<b>-0,994 *</b>	<b>0,019</b>
<b>N2</b>	+	<b>-1,047 *</b>	<b>0,026</b>	<b>-1,038 *</b>	<b>0,016</b>	<b>-1,025 *</b>	<b>0,016</b>
<b>IMR</b>	+ / -	-0,283	0,410	-0,286	0,401	-0,325	0,337
<b>FAV</b>	+			<b>0,784 ***</b>	<b>&lt; 0,001</b>	0,015	0,955
<b>VIS</b>	+ / -			-0,139	0,103	-0,156	0,065
<b>FAVxVIS</b>	+					<b>0,582 **</b>	<b>0,002</b>
Painel não balanceado		n = 65, T = 5-28, N = 1573					
R <sup>2</sup>		0,133		0,151		0,161	
Estatística F		<b>26,57 ***</b>	<b>&lt; 0,001</b>	<b>25,25 ***</b>	<b>&lt; 0,001</b>	<b>24,99 ***</b>	<b>&lt; 0,001</b>

Notas: Significância (\*\*\*) p < 0,001; (\*\*) p < 0,01; (\*) p < 0,05. As variáveis ROA, MVA<sup>®</sup>, QTB e ALV foram suavizadas (*winsorizing*) ao nível de 1% e a variável ROE ao nível de 2%. Equação:  $ROA_{i,t} = \beta_0 + \beta_1 * FAV_{i,t} +$

$$\beta_2 * VIS_{i,t} + \beta_3 * FAVxVIS_{i,t} + \beta_4 * ALV_{i,t} + \beta_5 * IDA_{i,t} + \beta_6 * TAM_{i,t} + \beta_7 * IND1_{i,t} + \beta_8 * IND2_{i,t} + \beta_9 * NM_{i,t} + \beta_{10} * NI_{i,t} + \beta_{11} * N2_{i,t} + \beta_{12} * IMR_{i,t} + \varepsilon_{i,t}$$

Fonte: Elaborado pelo autor.

Inicialmente, convém reapresentar as três hipóteses do estudo: a hipótese 1 (H1) que prevê a associação positiva entre a favorabilidade (FAV) e o desempenho financeiro, a hipótese (H2) que prediz associação entre a visibilidade (VIS) e o desempenho financeiro, e a hipótese 3 (H3) que descreve o efeito moderador da visibilidade (VIS) na associação entre a favorabilidade (FAV) e o desempenho financeiro.

O resultado da estatística F dos três modelos, mostrados na Tabela 8, os quais têm como variável dependente o Retorno sobre os Ativos (ROA), indica que esses modelos são significantes ( $p\text{-value} < 0,001$ ), o que sugere que alguma das variáveis preditoras influencia a variável dependente. A análise dos coeficientes do modelo 2, que inclui as variáveis independentes, indica que a favorabilidade (FAV) possui uma relação positiva e estatisticamente significativa, ao nível de 0,1% ( $\beta = 0,784$ ;  $p\text{-value} < 0,001$ ), com a lucratividade das empresas mensurada por intermédio do indicador ROA. Quanto à visibilidade (VIS), observa-se que o coeficiente desse termo não é significativo ( $p\text{-value} = 0,103$ ), que indica que não há associação entre a visibilidade e a lucratividade das empresas mensurada por intermédio do indicador ROA. O parâmetro  $R^2$  do modelo 2 aumenta em relação ao modelo 1 (modelo de controle), apontando que a inclusão das variáveis independentes melhora significativamente o ajuste do modelo. Esses resultados apoiam a hipótese H1 para o indicador contábil ROA, quando considerada a amostra analisada.

Com respeito ao modelo 3, que inclui o termo de interação (FAVxVIS), o coeficiente desse termo é positivo e significativo ( $\beta = 0,582$ ;  $p\text{-value} = 0,002$ ), o que indica que a associação positiva entre a favorabilidade (FAV) e o indicador de desempenho financeiro ROA é fortalecida com níveis crescentes de visibilidade quando considerada a amostra estudada, o que sustenta a hipótese H3 da pesquisa. Além disso, a inclusão do termo de interação entre a visibilidade e a favorabilidade melhora significativamente o ajuste do modelo.

Finalmente, o coeficiente do termo *inverse Mills Ratio* (IMR) não é significativo em todos modelos testados, o que evidencia que o viés de seleção da amostra não está afetando os resultados obtidos.

A Tabela 9 mostra os resultados da análise de dados em painel POLS para a variável dependente ROE.

Tabela 9 – Análise de dados em painel do indicador de desempenho ROE

Variáveis	Sinal Esperado	Modelo 1		Modelo 2		Modelo 3	
		Coef.	<i>p-value</i>	Coef.	<i>p-value</i>	Coef.	<i>p-value</i>
<b>Intercepto</b>	+ / -	14,649	0,247	10,234	0,467	10,484	0,448
<b>IDA</b>	+	<b>1,641 *</b>	<b>0,038</b>	<b>1,633 *</b>	<b>0,035</b>	<b>1,712 *</b>	<b>0,025</b>
<b>TAM</b>	+	-0,990	0,100	-0,764	0,270	-0,802	0,238
<b>ALV</b>	+ / -	-2,602	0,425	-2,675	0,421	-2,877	0,377
<b>IND1</b>	+ / -	-1,060	0,307	-1,002	0,315	-1,063	0,273
<b>IND2</b>	+ / -	2,677	0,106	2,190	0,162	2,004	0,176
<b>NM</b>	+	-0,737	0,569	-0,751	0,544	-0,685	0,572
<b>N1</b>	+	-2,171	0,107	-2,103	0,101	-2,062	0,098
<b>N2</b>	+	-3,709	0,060	<b>-3,672 *</b>	<b>0,041</b>	<b>-3,631 *</b>	<b>0,042</b>
<b>IMR</b>	+ / -	-1,718	0,124	-1,723	0,125	-1,838	0,100
<b>FAV</b>	+			<b>2,332 ***</b>	<b>0,001</b>	-0,050	0,958
<b>VIS</b>	+ / -			-0,445	0,194	-0,495	0,135
<b>FAVxVIS</b>	+					<b>1,781 **</b>	<b>0,008</b>
Painel não balanceado	n = 65, T = 5-28, N = 1570						
R <sup>2</sup>		0,053		0,069		0,077	
Estatística F		<b>9,78 ***</b>	<b>&lt; 0,001</b>	<b>10,45 ***</b>	<b>&lt; 0,001</b>	<b>10,87 ***</b>	<b>&lt; 0,001</b>

Notas: Significância (\*\*\*)  $p < 0,001$ ; (\*\*)  $p < 0,01$ ; (\*)  $p < 0,05$ . As variáveis ROA, MVA®, QTB e ALV foram suavizadas (*winsorizing*) ao nível de 1% e a variável ROE ao nível de 2%. Equação:  $ROE_{i,t} = \beta_0 + \beta_1 * FAV_{i,t} + \beta_2 * VIS_{i,t} + \beta_3 * FAVxVIS_{i,t} + \beta_4 * ALV_{i,t} + \beta_5 * IDA_{i,t} + \beta_6 * TAM_{i,t} + \beta_7 * IND1_{i,t} + \beta_8 * IND2_{i,t} + \beta_9 * NM_{i,t} + \beta_{10} * N1_{i,t} + \beta_{11} * N2_{i,t} + \beta_{12} * IMR_{i,t} + \varepsilon_{i,t}$ .

Fonte: Elaborado pelo autor.

O resultado da estatística F dos três modelos, mostrados na Tabela 9, os quais têm como variável dependente o Retorno sobre o Patrimônio Líquido (ROE), indica que esses modelos são significantes ( $p\text{-value} < 0,001$ ), o que sugere que alguma das variáveis preditoras influenciam a variável dependente. A análise dos coeficientes do modelo 2, que inclui as variáveis independentes, indica que a favorabilidade (FAV) possui uma relação positiva e estatisticamente significativa, ao nível de 0,1% ( $\beta = 2,332$ ;  $p\text{-value} = 0,001$ ), com a lucratividade das empresas mensurada por intermédio do indicador ROE. Quanto à visibilidade (VIS), observa-se que o coeficiente desse termo não é significativo ( $p\text{-value} = 0,194$ ), que indica que não há associação entre a visibilidade e a lucratividade das empresas mensurada por intermédio do indicador ROE. O parâmetro R<sup>2</sup> do modelo 2 aumenta em relação ao modelo 1 (modelo de controle), apontando que a inclusão das variáveis independentes melhora significativamente o ajuste do modelo. Esses resultados apoiam a hipótese H1 para o indicador contábil ROE, quando considerada a amostra analisada.

Com respeito ao modelo 3, que inclui o termo de interação (FAVxVIS), o coeficiente desse termo é positivo e significativo ( $\beta = 1,781$ ;  $p\text{-value} = 0,008$ ), que indica que a associação

positiva entre a favorabilidade (FAV) e o indicador de desempenho financeiro ROA é fortalecida com níveis crescentes de visibilidade quando considerada a amostra estudada, o que sustenta a hipótese H3 da pesquisa. Além disso, a inclusão do termo de interação entre a visibilidade e a favorabilidade melhora significativamente o ajuste do modelo.

Por fim, o coeficiente do termo *inverse Mills Ratio* (IMR) não é significativo em todos modelos testados, o que indica que o viés de seleção da amostra não está afetando os resultados obtidos.

A Tabela 10 apresenta os resultados da análise de dados em painel com efeitos fixos para a variável dependente MVA<sup>®</sup>.

Tabela 10 – Análise de dados em painel do indicador de desempenho MVA<sup>®</sup>

Variáveis	Sinal Esperado	Modelo 1		Modelo 2		Modelo 3	
		Coef.	p-value	Coef.	p-value	Coef.	p-value
IDA	+	-0,035	0,358	-0,036	0,345	-0,032	0,375
TAM	+	-0,024	0,514	-0,006	0,827	-0,009	0,744
ALV	+ / -	0,072	0,517	0,059	0,568	0,049	0,623
IND1	+ / -	0,027	0,486	0,034	0,366	0,031	0,392
IND2	+ / -	0,149	0,129	0,121	0,145	0,114	0,147
NM	+	-0,084	0,448	-0,080	0,460	-0,078	0,464
N1	+	-0,041	0,717	-0,036	0,745	-0,034	0,748
N2	+	-0,072	0,380	-0,060	0,464	-0,059	0,465
IMR	+ / -	-0,037	0,327	-0,034	0,305	-0,040	0,228
FAV	+			<b>0,058 **</b>	<b>0,002</b>	-0,045	0,076
VIS	+ / -			-0,032	0,074	-0,033	0,059
FAVxVIS	+					<b>0,079 **</b>	<b>0,004</b>
Painel não balanceado		n = 65, T = 5-28, N = 1573					
R <sup>2</sup>		0,095		0,135		0,162	
Estatística F		<b>17,81 ***</b>	<b>&lt; 0,001</b>	<b>21,67 ***</b>	<b>&lt; 0,001</b>	<b>24,73 ***</b>	<b>&lt; 0,001</b>

Notas: Significância (\*\*\*) p < 0,001; (\*\*) p < 0,01; (\*) p < 0,05. As variáveis ROA, MVA<sup>®</sup>, QTB e ALV foram suavizadas (*winsorizing*) ao nível de 1% e a variável ROE ao nível de 2%. Equação:  $MVA_{i,t} = \beta_0 + \beta_1*FAV_{i,t} + \beta_2*VIS_{i,t} + \beta_3*FAVxVIS_{i,t} + \beta_4*ALV_{i,t} + \beta_5*IDA_{i,t} + \beta_6*TAM_{i,t} + \beta_7*IND1_{i,t} + \beta_8*IND2_{i,t} + \beta_9*NM_{i,t} + \beta_{10}*N1_{i,t} + \beta_{11}*N2_{i,t} + \beta_{12}*IMR_{i,t} + \varepsilon_{i,t}$ .

Fonte: Elaborado pelo autor.

O resultado da estatística F dos três modelos, mostrados na Tabela 10, os quais têm como variável dependente o *Market Value Added* (MVA<sup>®</sup>), indica que esses modelos são significantes ( $p\text{-value} < 0,001$ ), o que sugere que alguma das variáveis preditoras influenciam a variável dependente. A análise dos coeficientes do modelo 2, que inclui as variáveis independentes, indica que a favorabilidade (FAV) possui uma relação positiva e estatisticamente significativa, ao nível de 1% ( $\beta = 0,058$ ;  $p\text{-value} = 0,002$ ), com o desempenho financeiro das empresas mensurado por intermédio do indicador MVA<sup>®</sup>. Com relação à

visibilidade (VIS), observa-se que o coeficiente desse termo não é significativo ( $p\text{-value} = 0,074$ ), o que assinala que não há associação entre a visibilidade e o desempenho financeiro mensurado por intermédio do indicador MVA<sup>®</sup>. O parâmetro R<sup>2</sup> do modelo 2 aumenta em relação ao modelo 1 (modelo de controle), apontando que tanto a inclusão da favorabilidade como da visibilidade melhoram significativamente o ajuste do modelo. Esses resultados apoiam a hipótese H1 para o indicador de desempenho de mercado MVA<sup>®</sup>, quando considerada a amostra analisada.

Com respeito ao modelo 3, que inclui o termo de interação (FAVxVIS), o coeficiente desse termo é positivo e significativo ( $\beta = 0,079$ ;  $p\text{-value} = 0,004$ ), o que indica que a associação positiva entre a favorabilidade (FAV) e o indicador de desempenho financeiro MVA<sup>®</sup> é fortalecida com níveis crescentes de visibilidade quando considerada a amostra estudada, o que apoia a hipótese H3 da pesquisa. Além disso, a inclusão do termo de interação entre a visibilidade e a favorabilidade melhora significativamente o ajuste do modelo.

Finalmente, o coeficiente do termo *inverse Mills Ratio* (IMR) não é significativo em todos modelos analisados, o que sugere que o viés de seleção não está afetando os resultados obtidos.

A Tabela 11 mostra os resultados da análise de dados em painel com efeitos fixos para a variável dependente Q de Tobin (QTB).

Tabela 11 – Análise de dados em painel do indicador de desempenho Q de Tobin (QTB)

Variáveis	Sinal Esperado	Modelo 1		Modelo 2		Modelo 3	
		Coef.	<i>p-value</i>	Coef.	<i>p-value</i>	Coef.	<i>p-value</i>
IDA	+	-0,014	0,935	-0,009	0,956	0,001	0,993
TAM	+	-0,170	0,068	-0,194	0,055	<b>-0,202 *</b>	<b>0,040</b>
ALV	+ / -	-0,476	0,268	-0,430	0,318	-0,458	0,288
IND1	+ / -	-0,111	0,558	-0,131	0,471	-0,140	0,432
IND2	+ / -	-0,093	0,640	-0,095	0,643	-0,115	0,570
NM	+	-0,423	0,300	-0,437	0,277	-0,430	0,276
N1	+	-0,597	0,085	-0,607	0,075	-0,602	0,072
N2	+	-0,553	0,065	<b>-0,601 *</b>	<b>0,040</b>	<b>-0,597 *</b>	<b>0,036</b>
IMR	+ / -	-0,204	0,170	-0,219	0,131	-0,236	0,102
FAV	+			<b>0,384 **</b>	<b>0,005</b>	0,073	0,537
VIS	+ / -			0,035	0,375	0,032	0,403
FAVxVIS	+					<b>0,239 *</b>	<b>0,024</b>
Painel não balanceado		n = 65, T = 5-28, N = 1573					
R <sup>2</sup>		0,185		0,214		0,226	
Estatística F		<b>38,75 ***</b>	<b>&lt; 0,001</b>	<b>37,87 ***</b>	<b>&lt; 0,001</b>	<b>37,29 ***</b>	<b>&lt; 0,001</b>

Notas: Regressão com dados em painel e efeitos fixos temporais. Significância (\*\*\*)  $p < 0,001$ ; (\*\*)  $p < 0,01$ ; (\*)  $p < 0,05$ . As variáveis ROA, MVA<sup>®</sup>, QTB e ALV foram suavizadas (*winsorizing*) ao nível de 1% e a variável ROE ao nível de 2%. Equação:  $QTB_{i,t} = \beta_0 + \beta_1*FAV_{i,t} + \beta_2*VIS_{i,t} + \beta_3*FAVxVIS_{i,t} + \beta_4*ALV_{i,t} + \beta_5*IDA_{i,t} + \beta_6*TAM_{i,t} + \beta_7*IND1_{i,t} + \beta_8*IND2_{i,t} + \beta_9*NM_{i,t} + \beta_{10}*N1_{i,t} + \beta_{11}*N2_{i,t} + \beta_{12}*IMR_{i,t} + \varepsilon_{i,t}$ .

Fonte: Elaborado pelo autor.

O resultado da estatística F dos três modelos, mostrados na Tabela 11, os quais têm como variável dependente o Q de Tobin (QTB), indica que esses modelos são significantes ( $p\text{-value} < 0,001$ ), o que sugere que alguma das variáveis preditoras influenciam a variável dependente. A análise dos coeficientes do modelo 2, que inclui as variáveis independentes, indica que a favorabilidade (FAV) possui uma relação positiva e estatisticamente significativa, ao nível de 1% ( $\beta = 0,384$ ;  $p\text{-value} = 0,005$ ), com o desempenho financeiro das empresas mensurado por intermédio do indicador QTB. Quanto à visibilidade (VIS), observa-se que o coeficiente desse termo não é significativo ( $p\text{-value} = 0,375$ ), que evidencia que não há associação entre a visibilidade e o desempenho financeiro mensurado por intermédio do indicador QTB. O parâmetro  $R^2$  do modelo 2 aumenta em relação ao modelo 1 (modelo de controle), apontando que tanto a inclusão das variáveis independentes melhora significativamente o ajuste do modelo. Esses resultados apoiam a hipótese H1 para o indicador de desempenho de mercado Q de Tobin, quando considerada a amostra analisada.

Com relação ao modelo 3, que inclui o termo de interação (FAVxVIS), o coeficiente desse termo é positivo e significativo ( $\beta = 0,239$ ;  $p\text{-value} = 0,024$ ), o que indica que a associação positiva entre a favorabilidade (FAV) e o indicador de desempenho financeiro QTB é fortalecida com níveis crescentes de visibilidade quando considerada a amostra estudada, o que apoia a hipótese H3 da pesquisa. Além disso, a inclusão do termo de interação entre a visibilidade e a favorabilidade melhora significativamente o ajuste do modelo.

Por fim, o coeficiente do termo *inverse Mills Ratio* (IMR) não é significativo em todos os modelos, o que indica que o viés de seleção da amostra não está afetando os resultados obtidos.

A seguir, são analisados conjuntamente os resultados obtidos para os quatro indicadores de desempenho financeiro considerados. Primeiramente, observa-se que o coeficiente de determinação ( $R^2$ ) é baixo em todos os modelos analisados, de modo que eles explicam uma parte limitada da variância da variável dependente. Entretanto, os resultados revelam a existência de preditores estatisticamente significativos, os quais permitem tirar conclusões importantes sobre como as alterações nos seus valores estão associadas com alterações no valor da variável dependente. Vale destacar que um coeficiente de determinação baixo seria problemático se o objetivo da pesquisa fosse fazer previsões a respeito do comportamento da



variável dependente e, embora valores baixos de  $R^2$  apresentem limitações, o nível de poder explicativo obtido não é incomum em pesquisas de gestão estratégica (Muller & Kräussl, 2011).

Dando sequência à análise, com respeito à avaliação das variáveis de controle que apresentam significância estatística, nota-se que a variável alavancagem (ALV) tem uma associação negativa e significativa apenas com o indicador de desempenho ROA, ao nível de 0,1%, nos modelos 1, 2 e 3. Esse resultado pode sugerir que empresas com alto índice de alavancagem, tendem a ter despesas financeiras elevadas, o que geraria a redução do lucro líquido, e, por consequência uma lucratividade mais baixa. Esses resultados estão em conformidade com os trabalhos de: Wang e Berens (2014) que encontraram uma associação negativa entre o indicador Q de Tobin e a alavancagem financeira para empresas presentes no índice AMAC da revista Fortune; J. Lee e Roh (2012) que observaram uma relação negativa entre os indicadores ROA, ROE e Q de Tobin e a alavancagem financeira para empresas do setor de alta tecnologia que figuram no índice AMAC da revista Fortune; Koh, Lee e Boo (2009) que relataram uma associação negativa entre o indicador Q de Tobin e a alavancagem financeira para empresas multinacionais americanas do setor de restaurantes e Bandeira et al. (2015) que identificaram uma correspondência negativa entre o indicador ROE e a alavancagem financeira para empresas listadas na BM&FBovespa.

Com relação à variável idade (IDA), percebe-se uma relação positiva e significativa somente com o indicador de desempenho contábil ROE, ao nível de 5%, nos modelos 1, 2 e 3. Esse resultado pode ser explicado pelo fato de que empresas estabelecidas há mais tempo obtém vantagem competitiva sobre empresas mais jovens, uma vez que a experiência na área de atuação é percebida como um fator que contribui para a melhoria do desempenho das empresas (M. Lee & Hu 2018). Tal resultado concorda com os achados de Bandeira et al. (2015) para o indicador ROE.

Concluindo a avaliação das variáveis de controle, a variável *dummy* que indica a presença das empresas no Nível 2 de governança corporativa da B3 (N2) apresenta uma relação negativa com o ROA, ao nível de 5%, nos modelos 1, 2 e 3; com o ROE, ao nível de 5%, nos modelos 2 e 3; e com o Q de Tobin, ao nível de 5%, nos modelos 2 e 3. De maneira similar, a variável *dummy* que indica a presença das empresas no Nível 1 de governança corporativa da B3 (N1) também apresenta uma relação negativa com o ROA, ao nível de 5%, nos modelos 1, 2 e 3. Assim, depreende-se que o desempenho de empresas presentes no Nível 2 e no Nível 1 de governança corporativa da B3, mensurado pelos indicadores mencionados, é menor do que das empresas listadas no mercado tradicional na amostra analisada. Os resultados obtidos não

estão de acordo com as considerações apresentadas no capítulo de metodologia, as quais sugerem que empresas com melhor governança corporativa tem melhor desempenho financeiro, pois possuem custos de agência mais baixos (López-Quesada et al., 2018) e conseguem levantar empréstimos com taxas de juros menores que aqueles obtidos por empresas com governança corporativa pior (Mishra & Mohanty, 2014). Esses resultados poderiam ter sido decorrentes do efeito de *outliers* que não foram removidos através do processo de suavização (*winsorizing*) empregado.

Concluindo a análise das regressões, é feita, a seguir, a comparação dos resultados obtidos com os achados de pesquisas prévias que examinaram a relação entre reputação e desempenho financeiro e que empregaram informações da mídia para mensurar o constructo reputação.

Os resultados obtidos referentes à associação da favorabilidade com todos os indicadores de desempenho financeiro avaliados (ROA, ROE, MVA<sup>®</sup> e Q de Tobin) corroboram com os achados dos estudos de Deephouse (2000), Agnihotri (2014) e OuYang et al. (2017). Entretanto, vale salientar que cada trabalho apresenta suas particularidades, sejam de amostra, sejam de método de determinação do teor de cada notícia utilizado no cálculo da favorabilidade ou do indicador de desempenho financeiro avaliado. O estudo de Deephouse (2000), realizado com bancos comerciais de duas grandes cidades norte-americanas, identificou uma relação positiva entre a favorabilidade e o indicador contábil ROA; o trabalho de Agnihotri (2014), cuja amostra continha empresas indianas, encontrou uma associação positiva entre favorabilidade e o indicador de mercado Q de Tobin; e a pesquisa de OuYang et al. (2017), realizada com empresas chinesas listadas em bolsa de valores, observou uma relação positiva entre a favorabilidade e o indicador retorno anormal cumulativo (CAR). Em comum, essas pesquisas analisaram somente um indicador financeiro e determinaram o teor de cada notícia, empregado no cálculo da favorabilidade, por meio de classificação manual que consistia na leitura dessas notícias. Esta pesquisa, por sua vez, estende os estudos citados, o que faz por meio do emprego de classificação automática do teor de cada notícia, por intermédio de análise de sentimento baseada em léxico, e pela análise de quatro indicadores de desempenho financeiro: dois indicadores contábeis e dois indicadores de mercado.

Os resultados obtidos relativos à associação da visibilidade com todos os indicadores de desempenho financeiro avaliados (ROA, ROE, MVA<sup>®</sup> e Q de Tobin) estão em linha com os achados do estudo de OuYang et al. (2017) que não observou qualquer relação entre a visibilidade, mensurada por meio do número de notícias, e o desempenho financeiro de empresas.

Finalmente, com relação ao efeito moderador da visibilidade na associação entre a favorabilidade e o desempenho financeiro, os resultados obtidos corroboram com o estudo de OuYang et al. (2017), já mencionado, que constatou que a relação entre favorabilidade e o indicador retornos anormais ajustados (CAR) é fortalecida pela exposição da empresa na mídia.

#### 4.4 Análise de regressão com algoritmo GBM e comparação com a regressão linear

Completando a análise de dados, foi feita, por intermédio de teste de hipóteses, a comparação entre os resultados dos modelos de regressão linear com dados em painel e os resultados dos modelos de regressão por meio do algoritmo de aprendizado de máquina *Gradient Boosting Machines* (GBM). Os procedimentos de estimação intervalar e de teste de hipóteses para duas médias populacionais com os desvios padrão dessas populações desconhecidos são robustos e podem ser empregados quando o tamanho da amostra é pequeno (Sweeney, Williams, & Anderson, 2017). Adicionalmente, Sweeney et al. (2017) argumentam, ainda, que, na maioria dos casos, os resultados obtidos são muito bons, mesmo que as populações não apresentem distribuição normal, quando o tamanho total da amostra ( $n_1+n_2$ ) é maior ou igual a 20. Assim, os modelos do estudo foram executados com validação cruzada *k-fold*, com  $k=20$ , de modo que o tamanho total da amostra é igual a 40.

Como mencionado na metodologia, foi feito o ajuste dos hiperparâmetros do algoritmo GBM com função de perda quantílica, uma vez que esse algoritmo pode ter alta variabilidade na exatidão, dependendo das configurações dos hiperparâmetros (Probst, Boulesteix, & Bischl, 2019). A otimização dos parâmetros foi feita concomitantemente com a validação cruzada, já que o pacote *caret* do software R provê uma funcionalidade que permite especificar um vetor com os valores possíveis para cada hiperparâmetro e que, também, determina qual a combinação desses valores minimiza a métrica de desempenho fixada como o erro quadrático médio. O hiperparâmetro taxa de aprendizado (denominada de *shrinkage* no pacote *gbm* do R) foi fixado para um valor pequeno, de 0,01, que permite obter um modelo mais preciso e que possibilita a parada do algoritmo antes que o sobreajuste ocorra. Assim, três hiperparâmetros do algoritmo foram otimizados: número de árvores (*n.trees* – variando de 1000 a 4000, com incrementos de 200), profundidade da árvore (*interaction.depth* – valores: 5, 15, 25, 35) e número mínimo de observações nos nós terminais (*n.minobsinnode* – valores: 3, 5, 7, 10). Em seguida, foi feita a validação cruzada dos modelos de regressão com dados em painel.

A Tabela 12 compara os dois métodos de regressão analisados com respeito ao coeficiente de determinação ( $R^2$ ). Os resultados mostram que a regressão por meio do algoritmo GBM produz resultados com maior poder preditivo para todas as variáveis dependentes.

Tabela 12 – Coeficiente de determinação ( $R^2$ ) obtidos com validação cruzada ( $k = 20$ )

Variável dependente	Estatística	Regressão Linear com dados em painel - modelo de efeitos fixos	Regressão Aprendizado de Máquina - algoritmo GBM
ROA	Média	0,158	0,369
	Desvio padrão	0,072	0,105
ROE	Média	0,082	0,335
	Desvio padrão	0,068	0,125
MVA <sup>®</sup>	Média	0,148	0,894
	Desvio padrão	0,073	0,100
QTB	Média	0,210	0,841
	Desvio padrão	0,071	0,043

Fonte: Elaborado pelo autor.

Os resultados do teste T para comparação de médias para populações com variâncias desconhecidas e diferentes, exibidos na Tabela 13, mostram que a hipótese estatística  $H_0$  do teste de poder preditivo das regressões foi rejeitada para todas as variáveis dependentes. Assim existem evidências estatísticas para afirmar que a regressão por meio do algoritmo GBM produz resultados com maior capacidade explanatória do que a regressão linear com dados em painel, quando considerada a amostra da pesquisa.

Tabela 13 – Resultados do teste T para comparação de médias - Coeficiente de determinação  $R^2$ .

Variável dependente	Estatística T	<i>p-value</i>	Rejeita $H_0^a$
ROA	7,465	< 0,001	Sim
ROE	7,964	< 0,001	Sim
MVA <sup>®</sup>	27,001	< 0,001	Sim
QTB	34,040	< 0,001	Sim

Notas: <sup>(a)</sup> Hipótese  $H_0$ :  $R^2$  é maior na regressão linear com dados em painel. Populações com variâncias desconhecidas e diferentes. Tamanho da primeira amostra = 20. Tamanho da segunda amostra = 20.

Fonte: Elaborado pelo autor.

A Tabela 14 compara os dois métodos de regressão analisados com respeito ao erro quadrático médio (RMSE). Os resultados mostram que a regressão por meio do algoritmo GBM produz resultados com menor erro quadrático médio (RMSE) para todas as variáveis dependentes.

Tabela 14 – Erro quadrático médio (RMSE) obtidos com validação cruzada (k = 20)

Variável dependente	Estatística	Regressão Linear com dados em painel – modelo de efeitos fixos	Regressão Aprendizado de Máquina – algoritmo GBM
ROA	Média	1,799	1,579
	Desvio padrão	0,269	0,236
ROE	Média	6,267	5,323
	Desvio padrão	0,840	0,845
MVA <sup>®</sup>	Média	0,147	0,049
	Desvio padrão	0,029	0,014
QTB	Média	0,635	0,293
	Desvio padrão	0,127	0,041

Fonte: Elaborado pelo autor.

Os resultados do teste T para comparação de médias para populações com variâncias desconhecidas e diferentes, exibidos na Tabela 15, mostram que a hipótese estatística H0 do teste de acurácia das regressões, mensurada pela métrica RMSE, foi rejeitada para todas as variáveis dependentes. Assim, existem evidências estatísticas para afirmar que a regressão por meio do algoritmo GBM produz resultados com maior exatidão que a regressão linear com dados em painel, quando considerada a amostra da pesquisa.

Tabela 15 – Resultados do teste T para comparação de médias – RMSE

Variável dependente	Estatística T	<i>p-value</i>	Rejeita H0 <sup>a</sup>
ROA	2,753	0,005	Sim
ROE	3,543	0,001	Sim
MVA <sup>®</sup>	13,502	< 0,001	Sim
QTB	11,451	< 0,001	Sim

Notas: <sup>(a)</sup> Hipótese H0: RMSE é menor na regressão linear. Populações com variâncias desconhecidas e diferentes. Tamanho da primeira amostra = 20. Tamanho da segunda amostra = 20.

Fonte: Elaborado pelo autor.

A Tabela 16 compara os dois métodos de regressão analisados com respeito ao erro absoluto médio (MAE). Os resultados mostram que a regressão por meio do algoritmo GBM produz resultados com menor erro absoluto médio (MAE) para todas as variáveis dependentes.

Tabela 16 – Erro Absoluto Médio (MAE) obtido com validação cruzada (k = 20)

Variável dependente	Estatística	Regressão linear com dados em painel - modelo de efeitos fixos	Regressão Aprendizado de Máquina - algoritmo GBM
ROA	Média	1,236	0,977
	Desvio padrão	0,166	0,131
ROE	Média	4,086	3,206
	Desvio padrão	0,575	0,398
MVA <sup>®</sup>	Média	0,089	0,026
	Desvio padrão	0,012	0,005
QTB	Média	0,439	0,167
	Desvio padrão	0,067	0,017

Fonte: Elaborado pelo autor.

Os resultados do teste T para comparação de médias para populações com variâncias desconhecidas e diferentes, exibidos na Tabela 17, mostram que a hipótese estatística H0 do teste de exatidão das regressões foi rejeitada para todas as variáveis dependentes. Dessa forma, existem evidências estatísticas para afirmar que a regressão por meio do algoritmo GBM produz resultados com maior exatidão que a regressão linear com dados em painel, quando considerada a amostra da pesquisa.

Tabela 17 – Resultados do teste T para comparação de médias – MAE

Variável dependente	Estatística T	p-value	Rejeita H0 <sup>a</sup>
ROA	5,498	< 0,001	Sim
ROE	5,627	< 0,001	Sim
MVA <sup>®</sup>	21,967	< 0,001	Sim
QTB	17,458	< 0,001	Sim

Notas: <sup>(a)</sup> Hipótese H0: MAE é menor na regressão linear. Populações com variâncias desconhecidas e diferentes. Tamanho da primeira amostra = 20. Tamanho da segunda amostra = 20.

Fonte: Elaborado pelo autor.

Os resultados da comparação dos métodos de regressão estão em consonância com os achados das pesquisas de Fernández-Gámez et al. (2016) e de Casado et al. (2016), que analisaram a associação entre a reputação corporativa e o valor de mercado e o risco de falência, respectivamente, e que compararam a regressão com redes neurais com a regressão linear simples. Cabe destacar que esses estudos não incluíram na amostra empresas do setor financeiro, entretanto, os autores não analisaram os problemas associados com o viés de seleção da amostra. Esta pesquisa estende os estudos citados por meio de uma comparação aperfeiçoada pelo uso de teste T para médias e, também, pelo emprego do algoritmo de aprendizado de máquina GBM ao invés de redes neurais artificiais.

## 5 CONCLUSÕES

O objetivo principal desta pesquisa foi verificar a associação entre a reputação corporativa e o desempenho financeiro das empresas listadas na B3 (Brasil, Bolsa, Balcão) no período do primeiro trimestre de 2012 ao quarto trimestre de 2018. Duas dimensões do constructo reputação corporativa, definido por Lange et al. (2011), foram mensuradas trimestralmente, no período do estudo, por meio de notícias das edições digitais dos jornais *Folha de São Paulo*, *O Estado de São Paulo* e *Valor Econômico*. Com respeito ao desempenho financeiro, foram empregados dois indicadores de desempenho contábil, o ROA e o ROE, e dois indicadores de desempenho de mercado, o MVA<sup>®</sup> e o Q de Tobin.

De acordo com a metodologia de mensuração da reputação por meio de informações da mídia, empresas com um número de notícias relevantes abaixo de 140, durante o período da pesquisa, foram excluídas da amostra. Apesar do uso de um conjunto de dados truncado, não houve indicação de que o viés de seleção amostral tenha afetado os resultados obtidos. Ademais, observou-se que o desempenho financeiro das empresas da amostra da pesquisa é bastante heterogêneo. Com respeito ao setor de atuação, constatou-se que o percentual de empresas da área financeira na amostra correspondeu à metade da participação desse setor quando consideradas todas as companhias listadas na B3. Além disso, a maioria das empresas fazia, ou fazem, parte da carteira teórica do índice Bovespa, o que indica que tais companhias têm maior cobertura na mídia jornalística.

Os resultados das regressões com dados em painel apresentados indicam que o teor das notícias relativas às empresas na mídia se reflete no desempenho financeiro dessas empresas, uma vez que foi constatada uma relação significativa e positiva entre a dimensão favorabilidade da reputação e todos os indicadores de desempenho analisados. A reputação é um ativo formado pelas percepções e avaliações dos stakeholders a respeito das organizações. Assim, a cobertura da mídia contribui para a construção das percepções dos *stakeholders* acerca das empresas, as quais formam o ativo estratégico reputação, que influencia o desempenho financeiro dessas companhias. Além disso, conforme o estabelecido pela Teoria da Sinalização, a mídia atua como intermediária do processo de sinalização entre as organizações e os *stakeholders*, mitigando a assimetria informacional. Com relação ao volume de cobertura da mídia, não foi observada associação significativa dessa variável com nenhum dos indicadores de desempenho estudados. Por fim, foi possível comprovar a hipótese do efeito moderador da visibilidade na associação entre a favorabilidade e o desempenho financeiro, uma vez que houve suporte para

essa suposição, nos modelos de todos os indicadores de desempenho analisados. Quanto a esse aspecto, foi observada, novamente, a participação da cobertura da mídia no processo de sinalização envolvido na construção da reputação corporativa.

Constatou-se, considerando a dimensão favorabilidade, que a reputação corporativa – neste estudo representada pela conceptualização de Lange et al. (2011) – está associada com o desempenho financeiro, podendo-se depreender que esse ativo intangível estratégico pode conduzir a melhor desempenho, em conformidade com o que prescreve a teoria da VBR e corroborando os achados de pesquisas anteriores que analisaram a relação entre reputação e desempenho financeiro (Agnihotri, 2014; Bandeira et al., 2015, Deephouse, 2000; Fernández-Gámez et al., 2016; Gois et al., 2017, J. Lee & Roh, 2012; Roberts & Dowling, 2002; OuYang et al. , 2017). Assim sendo, percebe-se que a reputação corporativa é um ativo socialmente construído que consiste na percepção dos *stakeholders* a respeito das organizações. A “boa” reputação das empresas sinaliza a qualidade dos produtos e dos serviços oferecidos, o que pode manter e melhorar seus resultados financeiros, fato que obteve suporte para todos os indicadores de desempenho analisados neste estudo. De maneira análoga, se a reputação corporativa é “ruim”, as empresas podem ter seu desempenho empresarial comprometido.

Com respeito às hipóteses estatísticas sobre as diferenças entre o poder explicativo e a exatidão da regressão linear com dados em painel e da regressão pelo algoritmo GBM, constatou-se que o modelo obtido por meio da técnica de aprendizado de máquina tem melhor poder explicativo para todas as variáveis de desempenho estudadas, o que pode sugerir que esse método é mais eficiente do que as técnicas lineares tradicionais na modelagem de funções complexas. Ademais, acerca da exatidão, foi possível verificar estatisticamente a maior acurácia da regressão pelo algoritmo GBM, uma vez que houve suporte tanto para a métrica de erro quadrático médio (RMSE) como para a métrica de erro absoluto médio (MAE) para todos os indicadores de desempenho examinados. Deste modo, os resultados obtidos estão de acordo com a expectativa de que a análise preditiva de dados, empregando técnicas de aprendizado de máquina, produz resultados melhores quando se deseja fazer previsões sobre o comportamento da variável dependente do modelo e o foco da análise não é determinar a existência de variáveis preditoras estatisticamente significativas.

Em termos de contribuições acadêmicas, este estudo trata de duas lacunas da literatura indicadas por Lange et al. (2011): a análise das consequências de duas dimensões pouco estudadas da reputação corporativa, a favorabilidade generalizada e a visibilidade, e a investigação de como a mudança de uma dada dimensão da reputação afeta as consequências de outra dimensão. Além disso, esta pesquisa estende a literatura sobre a associação entre a



reputação e o desempenho financeiro mediante a mensuração da reputação por meio de informações da mídia jornalística. Finalmente, considerando que há evidências na literatura de que a associação entre a reputação corporativa e suas consequências pode variar de país para país (Ali et al., 2015), este estudo se junta a pesquisas prévias (Bandeira et al., 2015; Domingos & Moura, 2013; Gois, 2015; Gois et al., 2017; Lopes et al. 2017) que analisaram as consequências da reputação de empresas brasileiras.

Com relação às contribuições metodológicas, estudos anteriores, examinando a relação entre reputação corporativa e desempenho financeiro, usaram preponderantemente *rankings* ou escalas para mensuração da reputação, como demonstrou a revisão sistemática de literatura realizada. Este trabalho, de outra forma, aplica técnicas de processamento de linguagem natural para análise quantitativa de dados textuais longitudinais massivos a fim de mensurar esse ativo intangível estratégico. Além disso, o estudo demonstra o uso de um conjunto de ferramentas de código aberto (*Open Source*) para processar informações disponíveis em sítios de internet. Por fim, cabe novamente enfatizar o uso de fontes de dados não estruturados em formato textual, que tiveram um crescimento exponencial devido à ascensão da internet e das mídias sociais nos últimos anos, em uma pesquisa de ciência social aplicada.

Na perspectiva gerencial, ressalta-se a relevância dos achados desta pesquisa, na medida em que, a partir dela, se permite depreender que as empresas devem levar em consideração a implementação de ações para gestão do ativo reputação corporativa e para monitoramento da mídia tanto nos canais tradicionais como nas novas mídias digitais, já que a reputação influi no desempenho financeiro, fator fundamental para a sobrevivência das empresas em um ambiente de negócios competitivo.

Como limitações desta pesquisa, vale salientar: o emprego da técnica de análise de sentimentos baseada em léxico, que modela cada notícia como um conjunto de palavras (*unigramas*), mas que não leva em consideração a ordem das palavras dentro de cada notícia, nem as relações léxico-sintáticas entre as palavras; a granularidade da avaliação do sentimento, que é feita por notícia relevante, isto é, foi considerado que cada notícia expressa somente uma opinião favorável, neutra ou desfavorável a respeito de uma empresa; o método de seleção de notícias relevantes de uma dada empresa, que consistia de um algoritmo simples de contagem de palavras; a amostra, que continha apenas empresas listadas na bolsa de valores brasileira; e a não inclusão da dimensão “ser conhecido por algo”, da conceptualização de Lange et al. (2011), na análise. Assim, recomenda-se, para estudos futuros: o uso de técnicas de análise de sentimento mais avançadas, que envolvam uso de *n-gramas*, ao invés de *unigramas*; a

classificação do sentimento das notícias em uma granularidade mais fina, isto é, de forma que se possa determinar o sentimento de cada sentença que compõe uma notícia e agregar os resultados obtidos; o uso de técnicas de classificação de sentimento supervisionadas, ao invés da técnica de classificação não supervisionada utilizada nesta pesquisa, as quais demandariam a obtenção de *corpus* de notícias previamente classificadas a respeito das empresas; a utilização de uma amostra que englobe empresas listadas nas bolsas de valores de outros países; e, finalmente, a elaboração de métodos, utilizando processamento de linguagem natural, para mensurar as características das empresas relacionadas com a dimensão da reputação corporativa “ser conhecido por algo”, tais como: qualidade de produtos e serviços, responsabilidade socioambiental, ambiente de trabalho, desempenho e qualidade do gerenciamento.

## REFERÊNCIAS

- ABI. (2018). Em três anos, jornais perdem 520 mil exemplares no Brasil. Retrieved from <http://www.abi.org.br/em-3-anos-jornais-perdem-520-mil-exemplares-no-brasil/>
- Agnihotri, A. (2014). Mass-Media-based Corporate Reputation and Firms' Market Valuation – Evidence from Emerging Markets. *Corporate Reputation Review*, 17(3), 206-218. doi:10.1057/crr.2014.10
- Al-ahtal, W. M., Alsamhi, M. H., Tabash, M. I., & Farhan, N. H. S. (2020). The impact of corporate governance on financial performance of Indian and GCC listed firms: An empirical investigation. *Research in International Business and Finance*, 51, 101083. doi:<https://doi.org/10.1016/j.ribaf.2019.101083>
- Ali, R., Lynch, R., Melewar, T. C., & Jin, Z. (2015). The moderating influences on the relationship of corporate reputation with its antecedents and consequences: A meta-analytic review. *Journal of Business Research*, 68(5), 1105-1117. doi:<https://doi.org/10.1016/j.jbusres.2014.10.013>
- Almestekawy, A., & AbdulSalam, M. (2019). Sentiment Analysis of Product Reviews Using Bag of Words and Bag of Concepts. *International Journal of Electronics*, 11, 49-60. doi:10.6636/IJEIE.201912
- Alon, A., & Vidovic, M. (2015). Sustainability Performance and Assurance: Influence on Reputation. *Corporate Reputation Review*, 18(4), 337-352. doi:10.1057/crr.2015.17
- Ang, S. H., & Wight, A.-M. (2009). Building Intangible Resources: The Stickiness of Reputation. *Corporate Reputation Review*, 12(1), 21-32. doi:10.1057/crr.2009.3
- Arellano, M. (1987). Computing Robust Standard Errors for Within-Groups Estimators. *Oxford Bulletin of Economics and Statistics*, 49(4), 431-434.
- Avanço, L. V., & Nunes, M. d. G. V. (2014, 18-22 Oct. 2014). *Lexicon-Based Sentiment Analysis for Reviews of Products in Brazilian Portuguese*. Paper presented at the 2014 Brazilian Conference on Intelligent Systems.
- B3. (2019a). Empresas Listadas. Retrieved from [http://www.b3.com.br/pt\\_br/produtos-e-servicos/negociacao/renda-variavel/empresas-listadas.htm](http://www.b3.com.br/pt_br/produtos-e-servicos/negociacao/renda-variavel/empresas-listadas.htm)
- B3. (2019b). Segmentos de Listagem. Retrieved from [http://www.b3.com.br/pt\\_br/produtos-e-servicos/solucoes-para-emissores/segmentos-de-listagem/sobre-segmentos-de-listagem/](http://www.b3.com.br/pt_br/produtos-e-servicos/solucoes-para-emissores/segmentos-de-listagem/sobre-segmentos-de-listagem/)
- B3. (2019c). Séries históricas. Retrieved from [http://www.b3.com.br/pt\\_br/market-data-e-indices/servicos-de-dados/market-data/historico/mercado-a-vista/series-historicas/](http://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/historico/mercado-a-vista/series-historicas/)
- Bachmann, R., Ehrlich, G., & Ruzic, D. (2017). *Firms and Collective Reputation: The Volkswagen Emission Scandal as a Case Study*. Retrieved from Germany: <https://ssrn.com/abstract=3124125>
- Bandeira, M., Góis, A., De Luca, M., & Vasconcelos, A. (2015). REPUTAÇÃO CORPORATIVA NEGATIVA E O DESEMPENHO EMPRESARIAL. *Revista De Contabilidade E Organizações*, 9(24), 71-83. doi:10.11606/rco.v9i24.88647
- Barnett, M. L., Jermier, J. M., & Lafferty, B. A. (2006). Corporate Reputation: The Definitional Landscape. *Corporate Reputation Review*, 9(1), 26-38. doi:10.1057/palgrave.crr.1550012
- Barney, J. B. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management*, 17(1), 99-120. doi:10.1177/014920639101700108
- Barney, J. B. (1995). Looking inside for Competitive Advantage. *The Academy of Management Executive (1993-2005)*, 9(4), 49-61.

- Baron, R., & Kenny, D. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173-1182. doi:10.1037//0022-3514.51.6.1173
- Basdeo, D. K., Smith, K. G., Grimm, C. M., Rindova, V. P., & Derfus, P. J. (2006). The impact of market actions on firm reputation. *Strategic Management Journal*, *27*(12), 1205-1219. doi:10.1002/smj.556
- Bayrakdaroglu, A., Ersoy, E., & Citak, L. (2012). Is There a Relationship Between Corporate Governance and Value-based Financial Performance Measures? A Study of Turkey as an Emerging Market. *Asia-Pacific Journal of Financial Studies*, *41*(2), 224-239. doi:10.1111/j.2041-6156.2012.01071.x
- Becker, K., & Tumitan, D. (2013). *Introdução a Mineração de Opiniões: Conceitos, Aplicações e Desafios*. Paper presented at the 28º Simpósio Brasileiro de Banco de Dados, Recife.
- Berens, G., & van Riel, C. B. M. (2004). Corporate Associations in the Academic Literature: Three Main Streams of Thought in the Reputation Measurement Literature. *Corporate Reputation Review*, *7*(2), 161-178. doi:10.1057/palgrave.crr.1540218
- Bergh, D. D., Ketchen, D. J., Boyd, B. K., & Bergh, J. (2010). New Frontiers of the Reputation—Performance Relationship: Insights From Multiple Theories. *Journal of Management*, *36*(3), 620-632. doi:10.1177/0149206309355320
- Berthelot, S. (2010). Corporate governance rating and financial performance: a Canadian study. *Corporate Governance: The international journal of business in society*, *10*(5), 635-646. doi:10.1108/14720701011085599
- Blajer-Gołębiewska, A. (2014). CORPORATE REPUTATION AND ECONOMIC PERFORMANCE: THE EVIDENCE FROM POLAND. *Economics & Sociology*, *7*(3), 194-2017. doi:10.14254/2071-789X.2014/7-3/15
- Boyd, B. K., Bergh, D. D., & Ketchen, D. J. (2010). Reconsidering the Reputation—Performance Relationship: A Resource-Based View. *Journal of Management*, *36*(3), 588-609. doi:10.1177/0149206308328507
- Breusch, T. S., & Pagan, A. R. (1980). The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics. *The Review of Economic Studies*, *47*(1), 239-253. doi:10.2307/2297111
- Busalim, A. H., & Hussin, A. R. C. (2016). Understanding social commerce: A systematic literature review and directions for further research. *International Journal of Information Management*, *36*(6, Part A), 1075-1088. doi:<https://doi.org/10.1016/j.ijinfomgt.2016.06.005>
- Cahan, R. H., Cahan, S. F., Lee, T., & Nguyen, N. H. (2017). Media Content, Accounting Quality, and Liquidity Volatility. *European Accounting Review*, *26*(1), 1-25. doi:10.1080/09638180.2015.1087866
- Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics Using Stata*. Texas: Stata Press.
- Cao, Y., Myers, J. N., Myers, L. A., & Omer, T. C. (2015). Company reputation and the cost of equity capital. *Review of Accounting Studies*, *20*(1), 42-81. doi:10.1007/s11142-014-9292-9
- Cardoso, V. I. C., De Luca, M. M. M., Lima, G. A. S. F., & Vasconcelos, A. C. (2013). Reputação corporativa nas empresas brasileiras: uma questão relevante para o desempenho empresarial?. *Revista De Contabilidade e Organizações*, *10*(21), 115-136.
- Carroll, C. E., & McCombs, M. (2003). Agenda-setting Effects of Business News on the Public's Images and Opinions about Major Corporations. *Corporate Reputation Review*, *6*(1), 36-46. doi:10.1057/palgrave.crr.1540188

- Carter, S. M. (2006). The Interaction of Top Management Group, Stakeholder, and Situational Factors on Certain Corporate Reputation Management Activities\*. *Journal of Management Studies*, 43(5), 1145-1176. doi:10.1111/j.1467-6486.2006.00632.x
- Carvalho, P., & Silva, M. J. (2017). *SentiLex-PT 02*. Retrieved from: <http://doi.org/10.23728/b2share.93ab120efdaa4662baec6adee8e7585f>
- Casado, A. M., Yanez, E. R., & Pelaez, A. (2016). *The value of corporate reputation in the bankruptcy risk*. Paper presented at the WMSCI 2016 - 20th World Multi-Conference on Systemics, Cybernetics and Informatics, Proceedings.
- Castilla-Polo, F., Gallardo-Vázquez, D., Sánchez-Hernández, M. I., & Ruiz-Rodríguez, M. C. (2018). An empirical approach to analyse the reputation-performance linkage in agrifood cooperatives. *Journal of Cleaner Production*, 195, 163-175. doi:<https://doi.org/10.1016/j.jclepro.2018.05.210>
- Cavalcanti, E. R., Cavalcanti, E. P., Pires, C. E., Costa, R. A., & Cavalcanti, C. R. (2011). Detecção e Avaliação de Cola em Provas Escolares Utilizando Mineração de Texto: um Estudo de Caso. *Brazilian Journal of Computers in Education; Vol 19, No 02 (2011)*. doi:10.5753/rbie.2011.19.02.56
- Chatterjee, S., Deng, S., Liu, J., Shan, R., & Jiao, W. (2018). Classifying facts and opinions in Twitter messages: a deep learning-based approach. *Journal of Business Analytics*, 1(1), 29-39. doi:10.1080/2573234X.2018.1506687
- Chen, Z., Gao, L., Li, X., & Wu, P. (2016). Managing reputation loss in China: in-depth analyses of financial restatements. *Chinese Management Studies*, 10(2), 312-345. doi:10.1108/CMS-12-2015-0275
- Chun, R. (2005). Corporate reputation: Meaning and measurement. *International Journal of Management Reviews*, 7(2), 91-109. doi:doi:10.1111/j.1468-2370.2005.00109.x
- Chung, K. H., & Pruitt, S. W. (1994). A Simple Approximation of Tobin's q. *Financial Management*, 23(3), 70-74. doi:10.2307/3665623
- Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2010). Signaling Theory: A Review and Assessment. *Journal of Management*, 37(1), 39-67. doi:10.1177/0149206310388419
- Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). *Automatic deception detection: methods for finding fake news*. Paper presented at the Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community, St. Louis, Missouri.
- Croissant, Y., & Millo, G. (2008). Panel Data Econometrics in R: The plm Package. *Journal of Statistical Software*, 27, 1-43. doi:10.18637/jss.v027.i02
- CVM. Dados Cadastrais. Retrieved from <http://sistemas.cvm.gov.br/port/cadastro/ftp.asp>
- Dalmácio, F. Z., Lopes, A. B., Rezende, A. J., & Sarlo Neto, A. (2013). Uma análise da relação entre governança corporativa e acurácia das previsões dos analistas do mercado brasileiro. *RAM. Revista de Administração Mackenzie*, 14, 104-139.
- Deephouse, D. L. (2000). Media Reputation as a Strategic Resource: An Integration of Mass Communication and Resource-Based Theories. *Journal of Management*, 26(6), 1091-1112. doi:10.1177/014920630002600602
- Deephouse, D. L., & Carter, S. M. (2005). An Examination of Differences Between Organizational Legitimacy and Organizational Reputation. *Journal of Management Studies*, 42(2), 329-360. doi:10.1111/j.1467-6486.2005.00499.x
- Deephouse, D. L., & Heugens, P. P. M. A. R. (2009). Linking Social Issues to Organizational Impact: The Role of Infomediaries and the Infomediary Process. *Journal of Business Ethics*, 86(4), 541-553. doi:10.1007/s10551-008-9864-3

- Deephouse, D. L., Newbury, W., & Soleimani, A. (2016). The effects of institutional development and national culture on cross-national differences in corporate reputation. *Journal of World Business*, 51(3), 463-473. doi:<https://doi.org/10.1016/j.jwb.2015.12.005>
- Dell'Atti, S., Trotta, A., Iannuzzi, A. P., & Demaria, F. (2017). Corporate Social Responsibility Engagement as a Determinant of Bank Reputation: An Empirical Analysis. *Corporate Social Responsibility and Environmental Management*, 24(6), 589-605. doi:10.1002/csr.1430
- Domingos, S. R. M., & Moura, A. A. F. (2013). *Reputação Corporativa e Desempenho: uma análise nas maiores companhias abertas do Brasil*. Paper presented at the 13<sup>o</sup> Congresso USP de Controladoria e Contabilidade, São Paulo.
- Dowling, G. R. (2016). Defining and Measuring Corporate Reputations. *European Management Review*, 13(3), 207-223. doi:10.1111/emre.12081
- Doyle, P. (1998). Radical strategies for profitable growth. *European Management Journal*, 16(3), 253-261. doi:[https://doi.org/10.1016/S0263-2373\(98\)00002-4](https://doi.org/10.1016/S0263-2373(98)00002-4)
- Duarte, F. C. d. L., Girão, L. F. d. A. P., & Paulo, E. (2017). Avaliando Modelos Lineares de Value Relevance: Eles Captam o que Deveriam Captar? *Revista de Administração Contemporânea*, 21, 110-134.
- Eberl, M., & Schwaiger, M. (2005). Corporate reputation: disentangling the effects on financial performance. *European Journal of Marketing*, 39(7-6), 838-854. doi:10.1108/03090560510601798
- Einwiller, S. A., Carroll, C. E., & Korn, K. (2010). Under What Conditions Do the News Media Influence Corporate Reputation? The Roles of Media Dependency and Need for Orientation. *Corporate Reputation Review*, 12(4), 299-315. doi:10.1057/crr.2009.28
- Fernández Sánchez, J. L., Luna Sotorrió, L., & Baraibar Díez, E. (2012). Can Corporate Reputation Protect Companies' Value? Spanish Evidence of the 2007 Financial Crash. *Corporate Reputation Review*, 15(4), 228-239. doi:10.1057/crr.2012.13
- Fernández-Gámez, M. A., Gil-Corral, A. M., & Galán-Valdivieso, F. (2016). Corporate reputation and market value: Evidence with generalized regression neural networks. *Expert Systems with Applications*, 46, 69-76. doi:<https://doi.org/10.1016/j.eswa.2015.10.028>
- Filbeck, G., Gorman, R., & Zhao, X. (2013). Are the best of the best better than the rest? The effect of multiple rankings on company value. *Review of Quantitative Finance and Accounting*, 41(4), 695-722. doi:10.1007/s11156-012-0329-5
- Flanagan, D. J., O'Shaughnessy, K. C., & Palmer, T. B. (2011). Re-Assessing the Relationship between the Fortune Reputation Data and Financial Performance: Overwhelming Influence or Just a Part of the Puzzle? *Corporate Reputation Review*, 14(1), 3-14. doi:10.1057/crr.2011.4
- Flore, C., Kolaric, S., & Schiereck, D. (2017). Settlement agreement types of federal corporate prosecution in the U.S. and their impact on shareholder wealth. *Journal of Business Research*, 76, 145-158. doi:<https://doi.org/10.1016/j.jbusres.2017.03.015>
- Folha Uol. (2016). Grupo Globo adquire controle do jornal 'Valor Econômico'. *Folha de São Paulo*. Retrieved from <https://www1.folha.uol.com.br/mercado/2016/09/1812854-grupo-globo-adquire-controle-do-jornal-valor-economico.shtml>
- Fombrun, C. J. (1996). *Reputation: Realizing Value from the Corporate Image*. Boston, MA: Harvard Business School Press.
- Fombrun, C. J., & Gardberg, N. (2000). Who's Tops in Corporate Reputation? *Corporate Reputation Review*, 3(1), 13-17. doi:10.1057/palgrave.crr.1540095



- Fombrun, C. J., Gardberg, N. A., & Sever, J. M. (2000). The Reputation Quotient<sup>SM</sup>: A multi-stakeholder measure of corporate reputation. *Journal of Brand Management*, 7(4), 241-255. doi:10.1057/bm.2000.10
- Fombrun, C. J., Ponzi, L. J., & Newbury, W. (2015). Stakeholder Tracking and Analysis: The RepTrak (R) System for Measuring Corporate Reputation. *Corporate Reputation Review*, 18(1), 3-24. doi:10.1057/crr.2014.21
- Fombrun, C. J., & Shanley, M. (1990). What's in a Name? Reputation Building and Corporate Strategy. *Academy of Management Journal*, 33(2), 233-258. doi:10.5465/256324
- Fombrun, C. J., & van Riel, C. B. M. (1997). The Reputational Landscape. *Corporate Reputation Review*, 1(1), 5-13. doi:10.1057/palgrave.crr.1540008
- Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4), 367-378. doi:[https://doi.org/10.1016/S0167-9473\(01\)00065-2](https://doi.org/10.1016/S0167-9473(01)00065-2)
- Fávero, L. P. L., Belfiore, P. P., Silva, F. L. d., & Chan, B. L. (2009). *Análise de dados: modelagem multivariada para tomada de decisões*. Rio de Janeiro: Elsevier.
- Gabriel, F., Assaf Neto, A., & Corrar, L. J. (2005). O impacto do fim da correção monetária no retorno sobre o patrimônio líquido dos bancos no Brasil. *Revista de Administração - RAUSP*, 40(1), 44-54.
- Gardberg, N. A., Symeou, P. C., & Zyglidopoulos, S. C. (2015). Public Trust's Duality in the CSP–Reputation–Financial Performance Relationship across Countries. *Academy of Management Proceedings*, 2015(1), 15705. doi:10.5465/ambpp.2015.190
- Gardberg, N. A., Zyglidopoulos, S. C., Symeou, P. C., & Schepers, D. H. (2017). The Impact of Corporate Philanthropy on Reputation for Corporate Social Performance. *Business & Society*, 0007650317694856. doi:10.1177/0007650317694856
- Ghosh, K. (2017). Corporate reputation, social performance, and organizational variability in an emerging country perspective. *Journal of Management & Organization*, 23(4), 545-565. doi:10.1017/jmo.2016.25
- Gray, E. R., & Balmer, J. M. T. (1998). Managing Corporate Image and Corporate Reputation. *Long Range Planning*, 31(5), 695-702. doi:[https://doi.org/10.1016/S0024-6301\(98\)00074-0](https://doi.org/10.1016/S0024-6301(98)00074-0)
- Greenwell, B., Boehmke, B., Cunningham, J., & Gbm, D. (2019). gbm: Generalized Boosted Regression Models. Retrieved from <https://cran.r-project.org/web/packages/gbm/gbm.pdf>
- Grinyer, P. H., & Norburn, D. (1975). Planning for Existing Markets: Perceptions of Executives and Financial Performance. *Journal of the Royal Statistical Society. Series A (General)*, 138(1), 70-97. doi:10.2307/2345251
- Góis, A. D. (2015). *Reputação corporativa: constructos e implicações para a criação de valor*. (Master). Universidade Federal do Ceará, Fortaleza-Brazil. Retrieved from <http://www.repositorio.ufc.br/handle/riufc/15094>
- Góis, A. D., De Luca, M. M. M., & Souza, N. M. (2017). Reputação Corporativa e Oportunidade de Crescimento. *Revista Organizações em Contexto*, 13(25), 300-322.
- Hall, R. (1992). THE STRATEGIC ANALYSIS OF INTANGIBLE RESOURCES. *Strategic Management Journal*, 13(2), 135-144. doi:10.1002/smj.4250130205
- Hall, R. (1993). A Framework Linking Intangible Resources and Capabilities to Sustainable Competitive Advantage. *Strategic Management Journal*, 14(8), 607-618.
- Hargreaves, S. (2010). BP's \$70 billion whipping. Retrieved from [http://money.cnn.com/2010/06/01/news/companies/BP\\_analysts/](http://money.cnn.com/2010/06/01/news/companies/BP_analysts/)
- Harrison, J. (2019). RSelenium: R Bindings for 'Selenium WebDriver'. Retrieved from <https://CRAN.R-project.org/package=RSelenium>

- Hausman, J. A. (1978). Specification Tests in Econometrics. *Econometrica*, 46(6), 1251-1271. doi:10.2307/1913827
- Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis*.: The Guilford Press.
- Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica*, 47(1), 153-161. doi:10.2307/1912352
- Hossain, M. M., Alamgir, M., & Alam, M. (2016). The Mediating Role of Corporate Governance and Corporate Image on the CSR-FP Link: Evidence from a developing country. *Journal of General Management*, 41(3), 33-51. doi:10.1177/030630701604100303
- Hu, M., & Liu, B. (2004). *Mining and summarizing customer reviews*. Paper presented at the Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, Seattle, WA, USA.
- Hu, M., & Liu, B. (2019). *Sentiment Lexicon*. Retrieved from: <http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar>
- Hu, T.-r., Luo, J.-b., Kautz, H., & Sadilek, A. (2016). Home location inference from sparse and noisy data: models and applications. *Frontiers of Information Technology & Electronic Engineering*, 17(5), 389-402. doi:10.1631/FITEE.1500385
- ILTEC - Instituto de Linguística Teórica e Computacional. (2019). Portal da Língua Portuguesa. Retrieved from <http://www.portaldalinguaportuguesa.org/novoacordo.php?action=novoacordo&act=list>
- Imdadullah, M., Aslam, M., & Altaf, S. (2016). mctest: An R Package for Detection of Collinearity among Regressors. *The R Journal*, 8(2), 495-505. doi:10.32614/RJ-2016-062
- Inkinen, H. (2015). Review of empirical research on intellectual capital and firm performance. *Journal of Intellectual Capital*, 16(3), 518-565. doi:doi:10.1108/JIC-01-2015-0002
- Janis, I. L., & Fadner, R. H. (1943). A COEFFICIENT OF IMBALANCE FOR CONTENT ANALYSIS. *Psychometrika*, 8(2), 105-119. doi:10.1007/bf02288695
- Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4), 305-360. doi:[https://doi.org/10.1016/0304-405X\(76\)90026-X](https://doi.org/10.1016/0304-405X(76)90026-X)
- Jones, G. H., Jones, B. H., & Little, P. (2000). *Reputation as Reservoir: Buffering Against Loss in Times of Economic Crisis* (Vol. 3).
- Jonkman, J. G. F., Boukes, M., Vliegenthart, R., & Verhoeven, P. (2019). Buffering Negative News: Individual-level Effects of Company Visibility, Tone, and Pre-existing Attitudes on Corporate Reputation. *Mass Communication and Society*, 1-25. doi:10.1080/15205436.2019.1694155
- Kabir, R., & Thai, H. M. (2017). Does corporate governance shape the relationship between corporate social responsibility and financial performance? *Pacific Accounting Review*, 29(2), 227-258. doi:10.1108/PAR-10-2016-0091
- Kaplan, R. S., & Norton, D. P. (2006). *Alignment: Using the balanced scorecard to create corporate synergies*. Boston, MA: Harvard Business Press.
- Kaur, A., & Singh, B. (2018). Re-examining Reputation–Performance Liaison in Indian Context. *Business Perspectives and Research*, 6(2), 100-112. doi:10.1177/2278533718764501
- Khan, K. S., Kunz, R., Kleijnen, J., & Antes, G. (2003). Five steps to conducting a systematic review. *Journal of the Royal Society of Medicine*, 96(3), 118-121.
- Kiouis, S., Popescu, C., & Mitrook, M. (2007). Understanding Influence on Corporate Reputation: An Examination of Public Relations Efforts, Media Coverage, Public



- Opinion, and Financial Performance From an Agenda-Building and Agenda-Setting Perspective. *Journal of Public Relations Research*, 19(2), 147-165. doi:10.1080/10627260701290661
- Kitchenham, B., & Charters, S. (2007). *Guidelines for Performing Systematic Literature Reviews in Software Engineering*. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.117.471>
- Koh, Y., Lee, S., & Boo, S. (2009). Impact of brand recognition and brand reputation on firm performance: U.S.-based multinational restaurant companies' perspective. *International Journal of Hospitality Management*, 28(4), 620-630. doi:<https://doi.org/10.1016/j.ijhm.2009.04.002>
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software; Vol 1, Issue 5 (2008)*. doi:10.18637/jss.v028.i05
- Kurniati, S. (2019). Stock returns and financial performance as mediation variables in the influence of good corporate governance on corporate value. *Corporate Governance: The International Journal of Business in Society*, 19(6), 1289-1309. doi:10.1108/CG-10-2018-0308
- Kyoon Yoo, D., & Ah Park, J. (2007). Perceived service quality: Analyzing relationships among employees, customers, and financial performance. *International Journal of Quality & Reliability Management*, 24(9), 908-926. doi:10.1108/02656710710826180
- Lange, D., Lee, P. M., & Dai, Y. (2011). Organizational Reputation: A Review. *Journal of Management*, 37(1), 153-184. doi:10.1177/0149206310390963
- Lanquillon, C. (2013). *Listening to the Voice of the Customers: An Early Warning System Based on Sentiment*. In C. Moewes & A. Nürnberger (Eds.), *Computational Intelligence in Intelligent Data Analysis*. (Vol. 445, pp. 223-235). doi:[https://doi.org/10.1007/978-3-642-32378-2\\_15](https://doi.org/10.1007/978-3-642-32378-2_15)
- Lee, J., & Roh, J. J. (2012). Revisiting corporate reputation and firm performance link. *Benchmarking: An International Journal*, 19(4/5), 649-664. doi:10.1108/14635771211258061
- Lee, Y. M., & Hu, J. L. (2018). Integrated approaches for business sustainability: The perspective of corporate social responsibility. *Sustainability (Switzerland)*, 10(7). doi:10.3390/su10072318
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167. doi:10.2200/S00416ED1V01Y201204HLT016
- Liu, L., Li, Q., Xu, Y., & Zhang, Y. (2014). *Firm celebrity, reputation and performance: A social media perspective*. Paper presented at the Proceedings - Pacific Asia Conference on Information Systems, PACIS 2014.
- Lopes, A. C., De Luca, M. M. M., Gois, A. D., & de Vasconcelos, A. C. (2017). SOCIOENVIRONMENTAL DISCLOSURE, CORPORATE REPUTATION AND VALUE CREATION IN FIRMS TRADED ON BM&FBOVESPA. *Revista Ambiente Contabil*, 9(1), 364-382.
- Loughran, T., & McDonald, B. (2011). When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65. doi:10.1111/j.1540-6261.2010.01625.x
- Loughran, T., & McDonald, B. (2018). *Master Dictionary*. Retrieved from: [https://drive.google.com/file/d/12ECPJMxV2wSalXG8ykMmkpa1fq\\_ur0Rf/view](https://drive.google.com/file/d/12ECPJMxV2wSalXG8ykMmkpa1fq_ur0Rf/view)
- López, V. A., & Iglesias, S. (2010). A Reputational--Performance Framework in an SME Context: Some Empirical Evidence from Spain. *Irish Journal of Management*, 29(2), 35-66.

- López-Quesada, E., Camacho-Miñano, M.-d.-M., & Idowu, S. O. (2018). Corporate governance practices and comprehensive income. *Corporate Governance: The International Journal of Business in Society*, 18(3), 462-477. doi:10.1108/CG-01-2017-0011
- Martinez, A. D., Russell, Z. A., Maher, L. P., Brandon-Lai, S. A., & Ferris, G. R. (2017). The Sociopolitical Implications of Firm Reputation: Firm Financial Reputation  $\times$  Social Reputation Interaction on Firm Financial Performance. *Journal of Leadership & Organizational Studies*, 24(1), 55-64. doi:10.1177/1548051816656005
- Meyer, J. W., & Scott, W. R. (1992). *Organizational environments: Ritual and rationality* (Updated ed. ed.). Newbury Park, Calif. :: Sage Publications.
- Mishra, S., & Mohanty, P. (2014). Corporate governance as a value driver for firm performance: evidence from India. *Corporate Governance*, 14(2), 265-280. doi:10.1108/CG-12-2012-0089
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Group, P.-P. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1. doi:10.1186/2046-4053-4-1
- Money, K., Saraeva, A., Garnelo-Gomez, I., Pain, S., & Hillenbrand, C. (2017). Corporate Reputation Past and Future: A Review and Integration of Existing Literature and a Framework for Future Research. *Corporate Reputation Review*, 20(3-4), 193-211. doi:10.1057/s41299-017-0034-3
- Mostafa, M. M. (2018). Clustering halal food consumers: A Twitter sentiment analysis. *International Journal of Market Research*, 61(3), 320-337. doi:10.1177/1470785318771451
- Muller, A., & Kräussl, R. (2011). Doing good deeds in times of need: a strategic perspective on corporate disaster donations. *Strategic Management Journal*, 32(9), 911-929. doi:10.1002/smj.917
- Myskova, R., & Hajek, P. (2016). The Effect of Managerial Sentiment on Market-to-Book Ratio. *Transformations in Business and Economics*, 15(2), 498-513. doi:<http://www.transformations.khf.vu.lt/34>
- Nardi, V. A. M. (2015). *Reputação corporativa: efeitos sobre o desempenho socioeconômico e a atividade exportadora das organizações*. (Master). Universidade do Vale do Rio dos Sinos, São Leopoldo - Brazil. Retrieved from <http://www.repositorio.jesuita.org.br/handle/UNISINOS/3033?show=full>
- Nassirtoussi, A. K., Aghabozorgi, S., Ying Wah, T., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, 41(16), 7653-7670. doi:<https://doi.org/10.1016/j.eswa.2014.06.009>
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, 7, 21-21. doi:10.3389/fnbot.2013.00021
- Newbury, W. (2010). Reputation and Supportive Behavior: Moderating Impacts of Foreignness, Industry and Local Exposure. *Corporate Reputation Review*, 12(4), 388-405. doi:10.1057/crr.2009.27
- Newey, W. K., & West, K. D. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), 703-708. doi:10.2307/1913610
- Orozco, L. A., Vargas, J., & Galindo-Dorado, R. (2018). Trends on the relationship between board size and financial and reputational corporate performance: The Colombian case. *European Journal of Management and Business Economics*, 27(2), 183-197. doi:10.1108/EJMBE-02-2018-0029

- OuYang, Z., Xu, J., Wei, J., & Liu, Y. (2017). Information Asymmetry and Investor Reaction to Corporate Crisis: Media Reputation as a Stock Market Signal. *Journal of Media Economics*, 30(2), 82-95. doi:10.1080/08997764.2017.1364256
- Pereira, L. H. M., & Martins, O. S. (2015). Rating de crédito, governança corporativa e desempenho das empresas listadas na BM&FBOVESPA. *REGE - Revista de Gestão*, 22(2), 205-221. doi:<https://doi.org/10.5700/rege559>
- Pett, M., Lackey, N., & J. Sullivan, J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*.
- Pfarrer, M. D., Pollock, T. G., & Rindova, V. P. (2010). A Tale of Two Assets: The Effects of Firm Reputation and Celebrity on Earnings Surprises and Investors' Reactions. *Academy of Management Journal*, 53(5), 1131-1152. doi:10.5465/amj.2010.54533222
- Philippe, D., & Durand, R. (2011). The impact of norm-conforming behaviors on firm reputation. *Strategic Management Journal*, 32(9), 969-993. doi:10.1002/smj.919
- Picault, M., & Renault, T. (2017). Words are not all created equal: A new measure of ECB communication. *Journal of International Money and Finance*, 79, 136-156. doi:<https://doi.org/10.1016/j.jimonfin.2017.09.005>
- Pollock, T. G., & Rindova, V. P. (2003). Media Legitimation Effects in the Market for Initial Public Offerings. *Academy of Management Journal*, 46(5), 631-642. doi:10.5465/30040654
- Pollock, T. G., Rindova, V. P., & Maggitti, P. G. (2008). Market Watch: Information and Availability Cascades Among the Media and Investors in the U.S. IPO Market. *Academy of Management Journal*, 51(2), 335-358. doi:10.5465/amj.2008.31767275
- Probst, P., Boulesteix, A.-L., & Bischl, B. (2019). Tunability: Importance of Hyperparameters of Machine Learning Algorithms. *Journal of Machine Learning Research*, 20(53), 1-32.
- R Core Team. (2019). R: A Language and Environment for Statistical Computing. In *R Foundation for Statistical Computing*. Vienna, Austria.
- Raithel, S., & Schwaiger, M. (2015). The effects of corporate reputation perceptions of the general public on shareholder value. *Strategic Management Journal*, 36(6), 945-956. doi:10.1002/smj.2248
- Raithel, S., Wilczynski, P., Schloderer, M. P., & Schwaiger, M. (2010). The value-relevance of corporate reputation during the financial crisis. *Journal of Product & Brand Management*, 19(6), 389-400. doi:10.1108/10610421011085703
- Rao, H. (1994). The Social Construction of Reputation: Certification Contests, Legitimation, and the Survival of Organizations in the American Automobile Industry: 1895–1912. *Strategic Management Journal*, 15(S1), 29-44. doi:10.1002/smj.4250150904
- Rao, V. R., Agarwal, M. K., & Dahlhoff, D. (2004). How Is Manifest Branding Strategy Related to the Intangible Value of a Corporation? *Journal of Marketing*, 68(4), 126-141.
- Rhee, M., & Haunschild, P. R. (2003). THE LIABILITY OF GOOD REPUTATION: A STUDY OF PRODUCT RECALLS IN THE U.S. AUTOMOBILE INDUSTRY. *Academy of Management Proceedings*, 2003(1), B1-B6. doi:10.5465/ambpp.2003.13792432
- Rindova, V. P., & Fombrun, C. J. (1999). Constructing Competitive Advantage: The Role of Firm-Constituent Interactions. *Strategic Management Journal*, 20(8), 691-710.
- Rindova, V. P., Petkova, A. P., & Kotha, S. (2007). Standing out: how new firms in emerging markets build reputation. *Strategic Organization*, 5(1), 31-70. doi:10.1177/1476127006074389

- Rindova, V. P., Williamson, I. O., Petkova, A. P., & Sever, J. M. (2005). Being Good or Being Known: An Empirical Examination of the Dimensions, Antecedents, and Consequences of Organizational Reputation. *Academy of Management Journal*, 48(6), 1033-1049. doi:10.5465/amj.2005.19573108
- Roberts, P. W., & Dowling, G. R. (2002). Corporate reputation and sustained superior financial performance. *Strategic Management Journal*, 23(12), 1077-1093. doi:doi:10.1002/smj.274
- Rose, C., & Thomsen, S. (2004). The Impact of Corporate Reputation on Performance:: Some Danish Evidence. *European Management Journal*, 22(2), 201-210. doi:<https://doi.org/10.1016/j.emj.2004.01.012>
- Sabate, J. M. d. I. F., & Puente, E. d. Q. (2003). Empirical Analysis of the Relationship Between Corporate Reputation and Financial Performance: A Survey of the Literature. *Corporate Reputation Review*, 6(2), 161-177. doi:10.1057/palgrave.crr.1540197
- Sainty, B., & Dunn, P. (2009). The relationship among board of director characteristics, corporate social performance and corporate financial performance. *International Journal of Managerial Finance*, 5(4), 407-423. doi:10.1108/17439130910987558
- Salama, A. (2005). A note on the impact of environmental performance on financial performance. *Structural Change and Economic Dynamics*, 16(3), 413-421. doi:<https://doi.org/10.1016/j.strueco.2004.04.005>
- Sarker, A., & Gonzalez, G. (2016). *Data, tools and resources for mining social media drug chatter*.
- Sarlo Neto, A. (2009). *Relação entre a estrutura de propriedade e a informatividade dos lucros contábeis no mercado brasileiro*. (Doutorado). Universidade de São Paulo, São Paulo. Retrieved from [https://teses.usp.br/teses/disponiveis/12/12136/tde-22042009-143539/publico/Tese\\_Alfredo\\_Sarlo\\_Neto.pdf](https://teses.usp.br/teses/disponiveis/12/12136/tde-22042009-143539/publico/Tese_Alfredo_Sarlo_Neto.pdf)
- Sarstedt, M., Wilczynski, P., & Melewar, T. C. (2013). Measuring reputation in global markets—A comparison of reputation measures' convergent and criterion validities. *Journal of World Business*, 48(3), 329-339. doi:<https://doi.org/10.1016/j.jwb.2012.07.017>
- Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology*, 98, 146-151. doi:<https://doi.org/10.1016/j.jclinepi.2017.12.006>
- Scott, W. R., & Søren, C. (1995). *The institutional construction of organizations : international and longitudinal studies* (S. Publications. Ed.). Twelve Oaks, CA: Sage Publications.
- Shan, Y. G., & McIver, R. P. (2011). Corporate governance mechanisms and financial performance in China: panel data evidence on listed non financial companies. *Asia Pacific Business Review*, 17(3), 301-324. doi:10.1080/13602380903522325
- Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake News Detection on Social Media: A Data Mining Perspective. *SIGKDD Explor. Newsl.*, 19(1), 22-36. doi:10.1145/3137597.3137600
- Silva, M. J., Carvalho, P., & Sarmiento, L. (2012, 2012//). *Building a Sentiment Lexicon for Social Judgement Mining*. Paper presented at the Computational Processing of the Portuguese Language, Berlin, Heidelberg.
- Silveira, A. D. M. d., Barros, L. A. B. d. C., & Famá, R. (2003). Estrutura de governança e valor das companhias abertas brasileiras. *Revista de Administração de Empresas*, 43, 50-64.
- Soleimani, A., Schneper, W. D., & Newburry, W. (2014). The Impact of Stakeholder Power on Corporate Reputation: A Cross-Country Corporate Governance Perspective. *Organization Science*, 25(4), 991-1008. doi:10.1287/orsc.2013.0889



- Souza, A., Figueredo, M., Cacho, N., Araújo, D., & Prolo, C. A. (2016). Using Big Data and Real-Time Analytics to Support Smart City Initiatives. *IFAC-PapersOnLine*, 49(30), 257-262. doi:<https://doi.org/10.1016/j.ifacol.2016.11.121>
- Souza, M., & Vieira, R. (2012). *Sentiment Analysis on Twitter Data for Portuguese Language*, Berlin, Heidelberg.
- Souza, M., Vieira, R., Buseti, D., Chishman, R. L. d. O., & Alves, I. M. d. R. (2011). *Construction of a Portuguese Opinion Lexicon from multiple resources*. Paper presented at the STIL.
- Souza, M., Vieira, R., Buseti, D., Chishman, R. L. d. O., & Alves, I. M. d. R. (2014). *OpLexicon*. Retrieved from: [http://ontolp.inf.pucrs.br/Recursos/downloads/OpLexicon/oplexicon\\_v3.0.zip](http://ontolp.inf.pucrs.br/Recursos/downloads/OpLexicon/oplexicon_v3.0.zip)
- Spence, M. (1973). Job Market Signaling\*. *The Quarterly Journal of Economics*, 87(3), 355-374. doi:10.2307/1882010
- Strauß, N., & van der Meer, T. G. L. A. (2017). News media coverage and initial public offerings in Germany: explaining flotation performance. *Corporate Communications: An International Journal*, 22(4), 523-541. doi:10.1108/CCIJ-04-2017-0028
- Stuebs, M., & Sun, L. (2010). Business Reputation and Labor Efficiency, Productivity, and Cost. *Journal of Business Ethics*, 96(2), 265-283. doi:10.1007/s10551-010-0464-7
- Sweeney, D. J., Williams, T. A., & Anderson, D. R. (2017). *Estatística aplicada à administração e economia* (3rd. ed.). São Paulo: Cengage Learning.
- Telles, R. (2001). A efetividade da matriz de amarração de Mazzon nas pesquisas em Administração. *Revista de Administração*, 36, 64-72.
- Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance*, 63(3), 1437-1467. doi:10.1111/j.1540-6261.2008.01362.x
- Thompson, S. K. (1987). Sample Size for Estimating Multinomial Proportions. *The American Statistician*, 41(1), 42-46. doi:10.2307/2684318
- Tischer, S., & Hildebrandt, L. (2014). Linking corporate reputation and shareholder value using the publication of reputation rankings. *Journal of Business Research*, 67(5), 1007-1017. doi:<https://doi.org/10.1016/j.jbusres.2013.08.007>
- Toomet, O., & Henningsen, A. (2008). Sample Selection Models in R: Package sampleSelection. *Journal of Statistical Software; Vol 1, Issue 7* (2008).
- Van den Bogaerd, M., & Aerts, W. (2014). Media Reputation of a Firm and Extent of Trade Credit Supply. *Corporate Reputation Review*, 17(1), 28-45. doi:10.1057/crr.2013.24
- Van den Bogaerd, M., & Aerts, W. (2015). Does media reputation affect properties of accounts payable? *European Management Journal*, 33(1), 19-29. doi:<https://doi.org/10.1016/j.emj.2014.05.002>
- Walker, K. (2010). A Systematic Review of the Corporate Reputation Literature: Definition, Measurement, and Theory. *Corporate Reputation Review*, 12(6), 357-387. doi:<https://doi.org/10.1057/crr.2009.26>
- Walsh, G., Schaarschmidt, M., & Ivens, S. (2017). Effects of customer-based corporate reputation on perceived risk and relational outcomes: empirical evidence from gender moderation in fashion retailing. *Journal of Product & Brand Management*, 26(3), 227-238. doi:10.1108/JPBM-07-2016-1267
- Walumbwa, F. O., Maitique, M. A., & Atamanik, C. (2014). Decision-making in a crisis: What every leader needs to know. *Organizational Dynamics*, 43(4), 284-293. doi:<https://doi.org/10.1016/j.orgdyn.2014.09.005>

- Wang, H., Choi, J., & Li, J. (2008). Too Little or Too Much? Untangling the Relationship between Corporate Philanthropy and Firm Financial Performance. *Organization Science*, *19*(1), 143-159.
- Wang, Y., & Berens, G. (2015). The Impact of Four Types of Corporate Social Performance on Reputation and Financial Performance. *Journal of Business Ethics*, *131*(2), 337-359. doi:10.1007/s10551-014-2280-y
- Wartick, S. L. (2002). Measuring Corporate Reputation: Definition and Data. *Business & Society*, *41*(4), 371-392. doi:10.1177/0007650302238774
- Welch, D., & Hull, D. (2015). Volkswagen of America November Vehicle Sales Down 24.7%. Retrieved from <https://www.bloomberg.com/news/articles/2015-12-01/vw-brand-u-s-sales-tumble-25-in-november-amid-diesel-scandal>
- Wickham, H. (2019). rvest: Easily Harvest (Scrape) Web Pages. Retrieved from <https://CRAN.R-project.org/package=rvest>
- Winn, M. I., MacDonald, P., & Zietsma, C. (2008). Managing Industry Reputation: The Dynamic Tension Between Collective And Competitive Reputation Management Strategies. *Corporate Reputation Review*, *11*(1), 35-55. doi:10.1057/crr.2008.4
- Wooldridge, J. M. (2012). *Introductory Econometrics A Modern Approach*. Mason, OH: South-Western, Cengage Learning.
- Younas, M., Jawawi, D. N. A., Ghani, I., Fries, T., & Kazmi, R. (2018). Agile development in the cloud computing environment: A systematic review. *Information and Software Technology*, *103*, 142-158. doi:<https://doi.org/10.1016/j.infsof.2018.06.014>
- Zeileis, A., & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. *Zeileis*, *2*(3), 7-10.
- Zhang, R., & Rezaee, Z. (2009). Do Credible Firms Perform Better in Emerging Markets? Evidence from China. *Journal of Business Ethics*, *90*(2), 221. doi:10.1007/s10551-009-0038-8

### **APÊNDICE A - Estudos primários selecionados na revisão sistemática da literatura**

[P1] Agnihotri, A. (2014). Mass-Media-based Corporate Reputation and Firms' Market Valuation – Evidence from Emerging Markets. *Corporate Reputation Review*, 17(3), 206-218.

[P2] Ang, S. H., & Wight, A.-M. (2009). Building Intangible Resources: The Stickiness of Reputation. *Corporate Reputation Review*, 12(1), 21-32.

[P3] Arnaldo, C., & Graca, C. A. M. (2016). The role of corporate reputation on co-operants behavior and organizational performance. *Journal of Management Development*, 35(1), 17-37.

[P4] Bandeira, M., Góis, A., De Luca, M., & Vasconcelos, A. (2015). REPUTAÇÃO CORPORATIVA NEGATIVA E O DESEMPENHO EMPRESARIAL. *Revista De Contabilidade E Organizações*, 9(24), 71-83.

[P5] Blajer-Golebiewska, A. (2014). CORPORATE REPUTATION AND ECONOMIC PERFORMANCE: THE EVIDENCE FROM POLAND. *Economics & Sociology*, 7(3), 194-2017.

[P6] Boyd, B. K., Bergh, D. D., & Ketchen, D. J. (2010). Reconsidering the Reputation—Performance Relationship: A Resource-Based View. *Journal of Management*, 36(3), 588-609.

[P7] Bravo, F. (2016). Forward-looking disclosure and corporate reputation as mechanisms to reduce stock return volatility. *Revista de Contabilidad*, 19(1), 122-131.

[P8] Cahan, R. H., Cahan, S. F., Lee, T., & Nguyen, N. H. (2017). Media Content, Accounting Quality, and Liquidity Volatility. *European Accounting Review*, 26(1), 1-25.

[P9] Cao, Y., Myers, J. N., Myers, L. A., & Omer, T. C. (2015). Company reputation and the cost of equity capital. *Review of Accounting Studies*, 20(1), 42-81.

[P10] Cardoso, V. I. C., De Luca, M. M. M., Lima, G. A. S. F., & Vasconcelos, A. C. (2013). Reputação corporativa nas empresas brasileiras: uma questão relevante para o desempenho empresarial?. *Revista De Contabilidade e Organizações*, 10(21), 115-136.

[P11] Casado, A. M., Yanez, E. R., & Pelaez, A. (2016). The value of corporate reputation in the bankruptcy risk. Paper presented at the WMSCI 2016 - 20th World Multi-Conference on Systemics, Cybernetics and Informatics, Proceedings.

[P12] Castilla-Polo, F., Gallardo-Vázquez, D., Sánchez-Hernández, M. I., & Ruiz-Rodríguez, M. C. (2018). An empirical approach to analyse the reputation-performance linkage in agrifood cooperatives. *Journal of Cleaner Production*, 195, 163-175.

[P13] Chen, T.-J. (2016). Corporate Reputation and Financial Performance of Life Insurers. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 41(3), 378-397.

[P14] Dell'Atti, S., Trotta, A., Iannuzzi, A. P., & Demaria, F. (2017). Corporate Social Responsibility Engagement as a Determinant of Bank Reputation: An Empirical Analysis. *Corporate Social Responsibility and Environmental Management*, 24(6), 589-605.

[P15] Domingos, S. R. M., & Moura, A. A. F. (2013). Reputação Corporativa e Desempenho: uma análise nas maiores companhias abertas do Brasil. Paper presented at the 13<sup>o</sup> Congresso USP de Controladoria e Contabilidade, São Paulo.

[P16] Fernández Sánchez, J. L., Luna Sotorrío, L., & Baraibar Díez, E. (2012). Can Corporate Reputation Protect Companies' Value? Spanish Evidence of the 2007 Financial Crash. *Corporate Reputation Review*, 15(4), 228-239.

[P17] Fernández-Gámez, M. A., Gil-Corral, A. M., & Galán-Valdivieso, F. (2016). Corporate reputation and market value: Evidence with generalized regression neural networks. *Expert Systems with Applications*, 46, 69-76.

[P18] Filbeck, G., Gorman, R., & Zhao, X. (2013). Are the best of the best better than the rest? The effect of multiple rankings on company value. *Review of Quantitative Finance and Accounting*, 41(4), 695-722.

[P19] Flanagan, D. J., O'Shaughnessy, K. C., & Palmer, T. B. (2011). Re-Assessing the Relationship between the Fortune Reputation Data and Financial Performance: Overwhelming Influence or Just a Part of the Puzzle? *Corporate Reputation Review*, 14(1), 3-14.

[P20] Góis, A. D. (2015). Reputação corporativa: constructos e implicações para a criação de valor. (Master), Universidade Federal do Ceará, Fortaleza-Brazil.

[P21] Góis, A. D., De Luca, M. M. M., & Souza, N. M. (2017). Reputação Corporativa e Oportunidade de Crescimento. *Revista Organizações em Contexto*, 13(25), 300-322.

[P22] Gök, O., & Özkaya, H. (2011). Does Corporate Reputation Improve Stock Performance in an Emerging Economy? Evidence From Turkey. *Corporate Reputation Review*, 14(1), 53-61.

[P23] González Sánchez, M., & Morales de Vega, M. E. (2018). Corporate reputation and firms' performance: Evidence from Spain. *Corporate Social Responsibility and Environmental Management*, 25(6), 1231-1245.

[P24] Huynh, Q. L. (2018). Earnings quality with reputation and performance. *Asian Economic and Financial Review*, 8(2), 269-278.

[P25] Kaur, A., & Singh, B. (2018). Re-examining Reputation–Performance Liaison in Indian Context. *Business Perspectives and Research*, 6(2), 100-112.



[P26] Koh, Y., Lee, S., & Boo, S. (2009). Impact of brand recognition and brand reputation on firm performance: U.S.-based multinational restaurant companies' perspective. *International Journal of Hospitality Management*, 28(4), 620-630.

[P27] Kucharska, W., & Kowalczyk, R. (2018). How to achieve sustainability?—Employee's point of view on company's culture and CSR practice. *Corporate Social Responsibility and Environmental Management*, 0(0).

[P28] Lee, J., & Roh, J. J. (2012). Revisiting corporate reputation and firm performance link. *Benchmarking: An International Journal*, 19(4/5), 649-664.

[P29] Lee, Y. M., & Hu, J. L. (2018). Integrated approaches for business sustainability: The perspective of corporate social responsibility. *Sustainability (Switzerland)*, 10(7).

[P30] Liu, L., Li, Q., Xu, Y., & Zhang, Y. (2014). Firm celebrity, reputation and performance: A social media perspective. Paper presented at the Proceedings - Pacific Asia Conference on Information Systems, PACIS 2014.

[P31] Lopes, A. C., De Luca, M. M. M., Gois, A. D., & de Vasconcelos, A. C. (2017). SOCIOENVIRONMENTAL DISCLOSURE, CORPORATE REPUTATION AND VALUE CREATION IN FIRMS TRADED ON BM&FBOVESPA. *Revista Ambiente Contabil*, 9(1), 364-382.

[P32] López, V. A., & Iglesias, S. (2010). A Reputational--Performance Framework in an SME Context: Some Empirical Evidence from Spain. *Irish Journal of Management*, 29(2), 35-66.

[P33] Martinez, A. D., Russell, Z. A., Maher, L. P., Brandon-Lai, S. A., & Ferris, G. R. (2016). The Sociopolitical Implications of Firm Reputation: Firm Financial Reputation × Social Reputation Interaction on Firm Financial Performance. *Journal of Leadership & Organizational Studies*, 24(1), 55-64.

[P34] Nardi, V. A. M. (2015). Reputação corporativa: efeitos sobre o desempenho socioeconômico e a atividade exportadora das organizações. (Master), Universidade do Vale do Rio dos Sinos, São Leopoldo - Brazil.

[P35] Olmedo-Cifuentes, I., & Martinez-Leon, I. (2012). Organizational Learning and Corporate Reputation: Relations and Effects on Financial Performance. *Proceedings of the 13th European Conference on Knowledge Management*, Vols 1 and 2, 889-898.

[P36] Olmedo-Cifuentes, I., Martinez-Leon, I. M., & Davies, G. (2014). Managing internal stakeholders' views of corporate reputation. *Service Business*, 8(1), 83-111.

[P37] Orozco, L. A., Vargas, J., & Galindo-Dorado, R. (2018). Trends on the relationship between board size and financial and reputational corporate performance: The Colombian case. *European Journal of Management and Business Economics*, 27(2), 183-197.

[P38] OuYang, Z., Xu, J., Wei, J., & Liu, Y. (2017). Information Asymmetry and Investor Reaction to Corporate Crisis: Media Reputation as a Stock Market Signal. *Journal of Media Economics*, 30(2), 82-95.

[P39] Özbay, D. (2018). THE IMPACT OF CORPORATE REPUTATION ON MARKET VALUE, EMPIRICAL EVIDENCE FROM TURKEY. *Financial and credit activity: problems of theory and practice*; Vol 4, No 27 (2018).

[P40] Pfarrer, M. D., Pollock, T. G., & Rindova, V. P. (2010). A Tale of Two Assets: The Effects of Firm Reputation and Celebrity on Earnings Surprises and Investors' Reactions. *Academy of Management Journal*, 53(5), 1131-1152.

[P41] Raithel, S., & Schwaiger, M. (2015). The effects of corporate reputation perceptions of the general public on shareholder value. *Strategic Management Journal*, 36(6), 945-956.

[P42] Raithel, S., Wilczynski, P., Schloderer, M. P., & Schwaiger, M. (2010). The value-relevance of corporate reputation during the financial crisis. *Journal of Product & Brand Management*, 19(6), 389-400.

[P43] Rowe, W. G., Harris, I. C., Cannella, A. A., & Francolini, T. (2009). In Search of Meaning: Does the Fortune Reputation Survey Alter Performance Expectations? *Canadian Journal of Administrative Sciences / Revue Canadienne des Sciences de l'Administration*, 20(3), 187-195.

[P44] Silvija, V., Ksenija, D., & Igor, K. (2017). The Impact of Reputation on Corporate Financial Performance: Median Regression Approach. *Business Systems Research Journal*, 8(2), 40-58.

[P45] Soleimani, A., Schnepfer, W. D., & Newbury, W. (2014). The Impact of Stakeholder Power on Corporate Reputation: A Cross-Country Corporate Governance Perspective. *Organization Science*, 25(4), 991-1008.

[P46] Strauß, N., & van der Meer, T. G. L. A. (2017). News media coverage and initial public offerings in Germany: explaining flotation performance. *Corporate Communications: An International Journal*, 22(4), 523-541.

[P47] Stuebs, M., & Sun, L. (2010). Business Reputation and Labor Efficiency, Productivity, and Cost. *Journal of Business Ethics*, 96(2), 265-283.

[P48] Thomaz, J. C., & Brito, A. P. Z. (2010). Reputação corporativa: construtos formativos e implicações para a gestão. *Revista de Administração Contemporânea*, 14, 229-250.

[P49] Tischer, S., & Hildebrandt, L. (2014). Linking corporate reputation and shareholder value using the publication of reputation rankings. *Journal of Business Research*, 67(5), 1007-1017.

[P50] Van den Bogaerd, M., & Aerts, W. (2014). Media Reputation of a Firm and Extent of Trade Credit Supply. *Corporate Reputation Review*, 17(1), 28-45. doi:10.1057/crr.2013.24

[P51] Van den Bogaerd, M., & Aerts, W. (2015). Does media reputation affect properties of accounts payable? *European Management Journal*, 33(1), 19-29.

[P52] Walsh, G., Albrecht, A. K., Kunz, W., & Hofacker, C. F. (2016). Relationship between Online Retailers' Reputation and Product Returns. *British Journal of Management*, 27(1), 3-20.

[P53] Wang, Y., & Berens, G. (2015). The Impact of Four Types of Corporate Social Performance on Reputation and Financial Performance. *Journal of Business Ethics*, 131(2), 337-359.

[P54] Xiaoman, W., Tongying, L., & Xiaolong, L. (2018). Corporate Reputation and Financial Performance: The Moderating Effects of Industry competitiveness and Environmental uncertainty. *IOP Conference Series: Materials Science and Engineering*, 439, 032055.

[P55] Zhang, R., & Rezaee, Z. (2009). Do Credible Firms Perform Better in Emerging Markets? Evidence from China. *Journal of Business Ethics*, 90(2), 221.



### APÊNDICE B - Avaliação da qualidade dos estudos primários

Identificador	Critério 1	Critério 2	Critério 3	Critério 4	Total
P1	2	2	2	2	8
P2	1	2	2	1	6
P3	0	2	2	2	6
P4	2	2	2	2	8
P5	2	2	2	2	8
P6	1	2	2	2	7
P7	0	2	2	2	6
P8	1	2	2	2	7
P9	1	2	2	2	7
P10	2	2	2	1	7
P11	1	2	1	1	5
P12	0	2	2	2	6
P13	1	2	2	2	7
P14	1	2	2	2	7
P15	2	2	2	1	7
P16	2	2	2	2	8
P17	2	2	2	2	8
P18	1	2	2	2	7
P19	1	1	2	2	6
P20	2	2	2	2	8
P21	2	2	2	2	8
P22	0	2	2	1	5
P23	1	2	2	1	6
P24	1	1	1	1	4
P25	1	1	1	1	4
P26	2	2	2	2	8
P27	0	2	2	2	6
P28	2	2	2	2	8
P29	2	2	2	2	8
P30	1	1	2	2	6
P31	2	1	2	1	6
P32	1	1	2	1	5
P33	2	1	2	2	7
P34	1	1	2	2	6
P35	0	2	1	1	4
P36	0	2	1	1	4
P37	2	2	2	1	7
P38	2	2	2	2	8
P39	1	1	1	2	5
P40	2	2	2	2	8
P41	1	2	2	2	7
P42	2	2	2	2	8
P43	0	1	2	2	5
P44	0	1	2	2	5
P45	1	2	2	2	7
P46	2	2	2	2	8
P47	1	1	2	2	6
P48	0	2	1	1	4
P49	2	2	2	2	8
P50	2	2	2	2	8
P51	2	2	2	2	8
P52	1	1	2	2	6
P53	2	2	2	2	8
<b>P54</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>3</b>
P55	2	2	2	2	8



**APÊNDICE C – Dados das empresas da amostra da pesquisa**

<b>Nome pregão</b>	<b>Data da constituição</b>	<b>Data de listagem na B3</b>	<b>Indústria</b>	<b>Setor B3</b>	<b>Segmento diferenciado de governança<sup>a</sup></b>
AMBEV S/A	08/07/2005	04/04/2000	Manufatura	Consumo não cíclico / Bebidas	Tradicional
B3	14/12/2007	20/08/2008	Financeiro	Financeiro / Serviços Financeiros Diversos	NM
BRADESCO	02/03/1943	22/03/2004	Financeiro	Financeiro / Intermediários	N1
BANCO DO BRASIL	30/12/1905	26/01/2004	Financeiro	Financeiro / Intermediários	NM
BRASKEM	18/06/1974	16/05/2005	Manufatura	Financeiros Materiais Básicos / Químicos	N1
BRF SA	23/11/1945	21/06/2000	Manufatura	Consumo não cíclico / Alimentos Processados	NM
CCR SA	23/09/1998	01/02/2002	Comércio e serviços	Bens Industriais / Transporte	NM
CEMIG	22/05/1952	04/06/2007	Comércio e serviços	Utilidade Pública / Energia Elétrica	N1
COPEL	26/10/1954	06/08/2007	Comércio e serviços	Utilidade Pública / Energia Elétrica	N1
COSAN	01/02/2000	18/11/2005	Manufatura	Petróleo. Gás e Biocombustíveis	NM
CYRELA REALT	01/01/1973	28/10/1996	Manufatura	Consumo Cíclico / Construção Civil	NM
EMBRAER	02/09/2005	05/06/2006	Manufatura	Bens Industriais / Material de Transporte	NM
FIBRIA	23/06/1986	01/01/2009	Manufatura	Materiais Básicos / Madeira e Papel	NM
GERDAU	20/11/1961	02/05/2003	Manufatura	Materiais Básicos / Siderurgia e Metalurgia	N1
ITAUUNIBANCO	01/01/1943	20/10/2004	Financeiro	Financeiro / Intermediários	N1
JBS	01/01/1953	29/03/2007	Manufatura	Financeiros Consumo não cíclico / Alimentos Processados	NM
KLABIN S/A	08/11/1978	02/01/2002	Manufatura	Materiais Básicos / Madeira e Papel	N2
LOJAS AMERIC	27/05/1929	28/08/2007	Comércio e serviços	Consumo Cíclico / Comércio	N1
LOJAS RENNERT	10/06/1965	30/06/2005	Comércio e serviços	Consumo Cíclico / Comércio	NM
NATURA	28/08/1969	26/05/2004	Manufatura	Consumo não cíclico / Produtos de Uso Pessoal e de Limpeza	NM
P.ACUCAR-CBD	11/11/1981	03/09/2007	Manufatura	Consumo não cíclico / Comércio e Distribuição	N1
PETROBRAS	03/10/1953	23/06/2000	Manufatura	Petróleo. Gás e Biocombustíveis	N2

Nome pregão	Data da constituição	Data de listagem na B3	Indústria	Setor B3	Segmento diferenciado de governança <sup>a</sup>
RUMO S.A.	30/09/1997	17/05/2004	Comércio e serviços	Bens Industriais / Transporte	NM
SABESP	06/09/1973	04/06/2007	Comércio e serviços	Utilidade Pública / Água e Saneamento	NM
SID NACIONAL	09/04/1941	31/05/2004	Manufatura	Materiais Básicos / Siderurgia e Metalurgia	Tradicional
TELEF BRASIL	22/05/1998	27/06/2005	Comércio e serviços	Telecomunicações / Telecomunicações	Tradicional
TIM PART S/A	22/05/1998	03/07/2007	Comércio e serviços	Telecomunicações / Telecomunicações	NM
USIMINAS	25/04/1956	17/06/1997	Manufatura	Materiais Básicos / Siderurgia e Metalurgia	N1
VALE	01/06/1942	11/04/1996	Manufatura	Materiais Básicos / Mineração	NM
BR MALLS PAR	26/05/2004	20/03/2007	Financeiro	Financeiro / Exploração de Imóveis	NM
ECORODOVIAS	07/11/2000	01/04/2010	Comércio e serviços	Bens Industriais / Transporte	NM
ELETROBRAS	25/04/1961	20/08/2007	Comércio e serviços	Utilidade Pública / Energia Elétrica	N1
ENERGIAS BR	24/06/2000	13/07/2005	Comércio e serviços	Utilidade Pública / Energia Elétrica	NM
ENGIE BRASIL	29/01/1998	23/05/2005	Comércio e serviços	Utilidade Pública / Energia Elétrica	NM
EQUATORIAL	16/06/1999	03/04/2006	Comércio e serviços	Utilidade Pública / Energia Elétrica	NM
YDUQS PART	31/03/2007	30/07/2007	Comércio e serviços	Consumo Cíclico / Diversos	NM
GOL	12/03/2004	24/06/2004	Comércio e serviços	Bens Industriais / Transporte	N2
HYPERA	01/01/2002	18/04/2008	Comércio e serviços	Saúde / Comércio e Distribuição	NM
KROTON	26/08/1998	08/06/2007	Comércio e serviços	Consumo Cíclico / Diversos	NM
MAGAZ LUIZA	30/10/1957	02/05/2011	Comércio e serviços	Consumo Cíclico / Comércio	NM
MARFRIG	06/04/2000	28/06/2007	Manufatura	Consumo não cíclico / Alimentos Processados	NM
MRV	31/05/2006	23/07/2007	Manufatura	Consumo Cíclico / Construção Civil	NM
SANTANDER BR	09/08/1985	01/11/2007	Financeiro	Financeiro / Intermediários Financeiros	Tradicional
SMILES	22/01/2003	13/10/2017	Comércio e serviços	Consumo Cíclico / Diversos	NM
SUZANO S.A.	08/12/1987	05/08/2004	Manufatura	Materiais Básicos / Madeira e Papel	NM
VIAVAREJO	12/07/1946	09/05/1996	Comércio e serviços	Consumo Cíclico / Comércio	NM
ALPARGATAS	03/04/1907	09/06/2005	Manufatura	Consumo Cíclico / Tecidos. Vestuário e Calçados	N1
AGPART	31/07/2000	19/10/2006	Financeiro	Financeiro / Holdings Diversificadas	Tradicional
CARREFOUR BR	13/08/1980	04/07/2017	Manufatura	Consumo não cíclico / Comércio e Distribuição	NM



Nome pregão	Data da constituição	Data de listagem na B3	Indústria	Setor B3	Segmento diferenciado de governança a
AZUL	03/01/2008	21/03/2017	Comércio e serviços	Bens Industriais / Transporte	N2
BTGP BANCO	18/01/1979	26/04/2012	Financeiro	Financeiro / Intermediários Financeiros	N2
CELPA	31/08/1960	27/04/2007	Comércio e serviços	Utilidade Pública / Energia Elétrica	Tradicional
CESP	06/12/1966	03/09/2007	Comércio e serviços	Utilidade Pública / Energia Elétrica	N1
COMGAS	11/11/1968	02/10/2007	Comércio e serviços	Utilidade Pública / Gás	Tradicional
CPFL ENERGIA	20/03/1998	02/06/2003	Comércio e serviços	Utilidade Pública / Energia Elétrica	NM
ELETROPAULO	07/04/1902	01/04/2008	Comércio e serviços	Utilidade Pública / Energia Elétrica	NM
ENEVA	25/04/2001	14/12/2007	Comércio e serviços	Utilidade Pública / Energia Elétrica	NM
GAFISA	12/11/1996	26/01/2006	Manufatura	Consumo Cíclico / Construção Civil	NM
LIGHT S/A	27/07/1999	03/12/2007	Comércio e serviços	Utilidade Pública / Energia Elétrica	NM
MMX MINER	26/08/1998	11/04/2006	Manufatura	Materiais Básicos / Mineração	NM
OI	28/02/2000	14/05/2007	Comércio e serviços	Telecomunicações / Telecomunicações	N1
OSX BRASIL	03/07/2007	22/03/2010	Manufatura	Petróleo. Gás e Biocombustíveis	NM
PDG REALT	01/01/2003	26/01/2007	Manufatura	Consumo Cíclico / Construção Civil	NM
REDE ENERGIA	02/01/1929	21/06/2006	Comércio e serviços	Utilidade Pública / Energia Elétrica	Tradicional
SANTOS BRP	25/08/1998	21/11/2007	Comércio e serviços	Bens Industriais / Transporte	NM
SER EDUCA	19/03/2002	08/10/2013	Comércio e serviços	Consumo Cíclico / Diversos	NM
TRIUNFO PART	11/01/1999	18/06/2007	Comércio e serviços	Bens Industriais / Transporte	NM

Nota: (a) Nível diferenciado de governança corporativa da B3 no quarto trimestre de 2018.



## APÊNDICE D – Listagens dos programas implementados

### #Cálculo das variáveis independentes: Favorabilidade (FAV) e Visibilidade (VIS)

```
require("odbc")
require("DBI")
require("RODBC")
require("textreadr")
require("stringr")

dics = data.frame (tipo= c ("teor_lough", "teor_oplex", "teor_stlex", "teor_bing"),
                  dicionario= c ("loughran", "oplexicon", "sentilex", "bingliu"))

REPUTACAO_TB = "reputacao_12"
NOTICIAS_RELEV_TB = "[dbo].[noticias_relevantes]"
NOTICIAS_RELEV_DISTINTAS_TB = "[dbo].[noticias_relevantes_distintas_r]"

calculaReputacao <- function ()
{
  options (stringsAsFactors = FALSE)
  setwd ("C:/Users/clgms/Documents/RevisaoSistematica/noticias")

  #Conexão banco de dados SQL Server
  conexao_escrita <- dbConnect (odbc (), .connection_string= CONEXAO_STR)
  conexao_leitura <- odbcDriverConnect (connection=CONEXAO_STR)

  query <- paste0 ("SELECT DISTINCT empresa_id, empresa FROM ", NOTICIAS_RELEV_DISTINTAS_TB,
                  " ORDER BY empresa_id")
  empresas <- sqlQuery(conexao_leitura, query, as.is=TRUE)

  for (k in 1:nrow (empresas)) {

    carregaReputacao (empresas$empresa_id[k], empresas$empresa[k], conexao_escrita, conexao_leitura)

  } #for

  #odbc
  dbDisconnect (conexao_escrita)
  #RODBC
  odbcClose (conexao_leitura)

} #calculaReputacao

carregaReputacao <- function (eid, empresa, conexao, con)
{
  reputacao = data.frame (empresa_id=c (0), trimestre=c (""), data_base=c (Sys.Date ()),
                          favorabilidade= c (0.0), visibilidade=c (0), ln_visibilidade=c (0.0),
                          dicionario= c (""))

  for (j in 1:nrow(dics))
  {
    tipo_dic= dics$tipo[j]
    dicio= dics$dicionar[j]

    #Visibilidade
    query <- paste0 ("SELECT empresa_id, data_base, trimestre, COUNT (titulo) as visibilidade,
                    LOG (COUNT (titulo)) as ln_visibilidade FROM ", NOTICIAS_RELEV_DISTINTAS_TB,
                    " WHERE empresa_id = ", eid,
                    " GROUP BY trimestre, data_base, empresa_id ORDER BY data_base")
  }
}
```

```

visibilidade <- sqlQuery (con, query, as.is=TRUE)

#Calculo de favorabilidade
query <- paste0 ("SELECT notd.empresa_id, notd.data_base, notd.trimestre,
CASE WHEN notr.",tipo_dic, " = -1 THEN 'desfavoravel' WHEN notr.", tipo_dic ,
" = 0 THEN 'neutra' ELSE 'favoravel' END AS teor",
" , COUNT (notd.titulo) as quantidade FROM ", NOTICIAS_RELEV_TB,
" AS notr INNER JOIN ", NOTICIAS_RELEV_DISTINTAS_TB,
" AS notd ON notr.noticia_id = notd.noticia_id WHERE notd.empresa_id=", eid,
" GROUP BY notd.trimestre, notd.data_base, notd.empresa_id, notr.",
tipo_dic, " ORDER BY notd.data_base")

teor <- sqlQuery (con, query, as.is=TRUE)

bloco = data.frame ()
for (i in 1:nrow(visibilidade))
{
vis = visibilidade[i, ]
trm = subset (teor, (trimestre == vis$trimestre))
fav = subset (trm, teor == "favoravel")$quantidade
desfav = subset (trm, teor == "desfavoravel")$quantidade

#Equação 3
#-----
if (length(fav) == 0){fav = 0}
if (length(desfav) == 0) {desfav = 0}
if (fav > desfav) {
favorabilidade= (fav^2 - fav*desfav) / vis$visibilidade^2
} else if (fav < desfav) {
favorabilidade = (fav*desfav - desfav^2) / vis$visibilidade^2
} else {
favorabilidade = 0
}
#-----

reputacao$empresa_id = vis$empresa_id
reputacao$trimestre =vis$trimestre
reputacao$data_base = vis$data_base
reputacao$favorabilidade = favorabilidade
reputacao$visibilidade= vis$visibilidade
reputacao$ln_visibilidade = vis$ln_visibilidade
reputacao$dicionario = dicio
bloco <- rbind (bloco, reputacao)
} #for i

#grava na tabela
retorno = dbWriteTable (conexao, REPUTACAO_TB, bloco, append=TRUE)
rm (bloco)

} #for j

} #carregaReputacao

#Cálculo do Inverse Mills Ratio

require("textreadr")
require("sampleSelection")

dadosImr <- read.csv2 ("dados_dissertacao_imr_cgomes.csv", header =TRUE)

```

```
#LIMT=1 se empresa tem mais que 140 notícias relevantes durante o período do estudo, e LIMT=0 caso contrário
formula <- LIMT ~ TAM + IDA + IBOV
controle <- glm.control (epsilon = 1e-9, maxit = 500, trace = FALSE)
modeloProbit <- stats::glm (formula , control= controle,
                           family= binomial (link = "probit"), data=dadosImr)
summary (modeloProbit)
```

```
#Inverse Mills Ratio
dadosImr$IMR <- invMillsRatio(modeloProbit)$IMR1
```

### #Funções comuns – arquivo comum.R

```
DIRETORIO_TRABALHO = "diretorio_de_trabalho_utilizado"
CONEXAO_STR = "string_de_conexao_ao_db_utilizada"
```

```
ultimaTentativa <- function (codigo, lk)
{
  ret <- tryCatch(
    codigo,
    error=function (e)
    {
      message(paste("Ultima tentativa", "Leitura da pagina", lk, sep= " : "))
      Sys.sleep(3)
      return ("NOK")
    }
  )
  return (ret)
} #ultimaTentativa
```

```
acessarPagina <- function (codigo, lk)
{
  ret <- tryCatch(
    codigo,
    error=function (e)
    {
      Message (paste ("Primeira tentativa", "Leitura da página", lk, sep= " : "))
      Sys.sleep(3)
      ret <- ultimaTentativa (codigo, lk) #nao entra em loop
    }
  )
  return (ret)
} #acessarPagina
```

```
removerCaracteres <- function (texto)
{
  return (str_replace_all (str_replace_all (str_replace_all (str_replace_all (str_replace_all(
    texto, "\uFFFFD", ""), "\u200B", ""), "\u0082", ""), "\u2029", ""), "\200", ""))
} #removerCaracteres
```

```
ajustarParagrafo <- function (texto)
{
  texto_ajustado = str_replace_all (str_replace_all (stri_replace_first_fixed (stri_replace_last_fixed (
    texto, ".\n", ". "), "\n", ""), "[\n]", " "), "[\n]", "")
  return (texto_ajustado)
} #ajustaParagrafo
```

### #Coleta dados Folha de São Paulo

```
require("textreadr")
require("rvest")
require("lubridate")
require("odbc")
require("stringr")
require("stringi")
require("DBI")
require("RODBC")
require("httr")
require("purrr")
require("RSelenium")
```

```
NOTICIAS_RELEVANTES_TB = "noticias_relevantes"
```

```
#Notícias cujo tempo de carregamento for maior que o tempo de espera
#não serão consideradas na análise de dados
TIMEOUT_PAGE_LOAD = 20000
```

```
EMAIL_USUARIO = "email_do_usuario"
SENHA_USUARIO = "senha_do_usuario"
```

```
source(file="comum.R")
source(file="mtexto.R")
```

```
#Conexao Browser
if (exists ("servChrome") == TRUE)
{
  servChrome$server$stop()
  rm (servChrome)
}
```

```
if (exists ("driver") == TRUE)
{
  driver$close ()
  rm (driver)
}
```

```
servChrome <- rsDriver (browser=c ("chrome"), port = 8882L, chromever= "79.0.3945.36")
driver <- servChrome$client
driver$setTimeout (type = "page load", milliseconds = TIMEOUT_PAGE_LOAD)
```

```
#Pagina de login
driver$navigate ("https://login.folha.com.br/login")
```

```
element <- driver$findElement (using = "css", 'input[name="email"]')
element$sendKeysToElement (list (EMAIL_USUARIO))
element <- driver$findElement (using = "css", 'input[name="password"]')
element$sendKeysToElement (list (SENHA_USUARIO))
element <- driver$findElement (using = "css", '.btn.btn-block.btn-primary')
element$clickElement ()
```

```
coletarDadosFsp <- function(inicio, termino, consulta, empresa, empresa_id, completo)
{
  nr_noticias_relv = 0
  nr_noticias = 0
  falha_elemento = 0
  options(stringsAsFactors = FALSE)
  setwd (DIRETORIO_TRABALHO)
  REGISTROS = 25
```

```

retorno <- tryCatch ({resultado <- driver$getSessions()},
  error= {function(e) {
    message (paste0("eeee> driver$getSession() - Sem Sessao"))
    return (0)}})

if (is.numeric (retorno) == TRUE)
{
  message ("Sem sessão aberta no Browser")
  return ("NOK")
}

linha = data.frame (empresa=c (""),empresa_id=c (0),titulo=c (""), descricao=c (""), link=c (""),
  link_redir= c (""), data=c (Sys.Date ()), aux=c (""), busca=consulta,
  fonte= c ("Folha de São Paulo"), data_extracao=c (Sys.Date ()))

registro = data.frame (empresa=c (""), empresa_id=c (0),titulo=c (""), link=c (""), link_redir=c (""),
  data=c (Sys.Date ()), trimestre=c (""), data_base=c (Sys.Date ()),
  s_lough=c (0.0), teor_lough=c (0), s_oplex=c (0.0), teor_oplex=c (0),
  s_stlex=c (0.0), teor_stlex=c (0), s_bing=c (0.0), teor_bing=c (0),
  busca=c (""), fonte=c ("Folha de São Paulo"), data_leitura=c (Sys.Date ()))

final = dmy(termino)
if (completo == 1)
{
  fim = termino
} else {
  intervalo = 1
  if (abs (month (final) – month (dmy(inicio))) %% 2 == 1)
  {
    intervalo = 2
  }
  fim = format (as.Date (inicio, "%d/%m/%Y") + months (intervalo) – days (1), "%d/%m/%Y")
}

inicioExecucao = Sys.time ()
conexao <- dbConnect (odbc (),.connection_string= CONEXAO_STR)

consultaOriginal = consulta
if (str_detect (consulta, "[+]") == FALSE)
{
  consulta = paste0("%22", consulta, "%22")
}
consulta = str_replace_all (consulta," ", "%20")

while (as.Date (fim, "%d/%m/%Y") <= final)
{
  #Editoria Mercado
  busca = paste0("http://search.folha.uol.com.br/search?q=", str_replace_all(consulta, "[ ]", "+"),
"&periodo=personalizado&sd=", inicio, "&ed=", fim, "&site=sitefolha&site[]=online/dinheiro")

  #Le lista de noticias
  noticias= read_html (busca, trim=TRUE, encoding="UTF-8")
  nosDom= html_nodes (noticias,".c-pagination__item a")
  paginas= html_attr (nosDom,"href")
  nosDom= html_nodes(noticias,".c-search__result")
  resultados= str_wrap(html_text (nosDom))
  nrResultados= as.numeric (str_split (resultados, "[ ]")[[1]][1])
}

```

```

print (paste0("Numero de noticias: ", nrResultados))

if (nrResultados != 0)
{
  for (pagina in 1: (nrResultados %% REGISTROS + 1))
  {
    nosDom = html_nodes (noticias,".c-headline__content")
    conteudos = html_text (nosDom)
    nosDom = html_nodes (noticias,".c-headline__title")
    titulos = html_text (nosDom)
    nosDom = html_nodes(noticias,".c-headline__standfirst")
    descricoes = html_text (nosDom)
    nosDom = html_nodes(noticias,".c-headline__dateline")
    datas = html_text (nosDom)
    nosDom = html_nodes(noticias,".c-headline__content a")
    links = html_attr(nosDom,"href")

    #iterar pelos titulos
    for (noticia in 1:length(titulos))
    {
      aux = ""
      linha$aux = ""

      linha$empresa = empresa
      linha$empresa_id = empresa_id

      ttl = str_replace_all (str_wrap(titulos[noticia]), "\u20AC", "EUR")
      linha$titulo = str_replace_all(ttl, "[\r\n]", "")
      linha$titulo = iconv(linha$titulo, "UTF-8", "latin1")
      Encoding(linha$titulo) <- "latin1"

      linha$descricao = iconv(str_replace_all(str_wrap(removerCaracteres(descricoes[noticia])), "[\n]", " "),
"UTF-8", "latin1")
      Encoding(linha$descricao) <- "latin1"

      linha$link = str_replace_all (str_wrap(links[noticia]), "[\r\n]", "")
      linha$link_redir = ""

      linha$data = as.Date (str_remove (str_split (str_wrap(datas[noticia]), "[ ]")[1][1], ""), "%d.%b.%Y")
      if (is.na (linha$data))
      {
        linha$data = '1970-01-01'
      }

      linha$aux = ""
      lnk = linha$link
      iniExec = Sys.time ()

      retorno <- tryCatch ({driver$navigate (lnk)},
        warning= function (w) print (w),
        error= function(e) {
          warning (paste0("eeee> driver$navigate (lnk) - Leitura do texto da noticia: link= ",
lnk, " - Tempo: ", (Sys.time () - iniExec)))
          return (0)
        })

      if (!is.null (retorno)){
        print (paste0("eeee> - driver$navigate (lnk) - Falha na leitura do texto da noticia: link= ", lnk))
      }
    }
  }
}

```



```

    auxs=".content p:not (.text):not (.content-text),.c-news__body p:not ([class]),.n--noticia__content p:not ([class])"

    txto <- tryCatch (
      {driver$findElements(using = "css selector",auxs)},
      warning= function (w) print (w),
      error=function(e) {
        warning (paste0("eeee> driver$findElement (using = 'css','seletores') Leitura do texto da noticia: link= ",
link))
        return(NA)}
    )

    if (is.list (txto) == FALSE){
      print (paste0("eeee> driver$findElement (using = css selector,'seletores') - Falha no elemento da noticia:
link= ", link))
      falha_elemento = falha_elemento + 1
      if (falha_elemento > 2)
      {
        driver$close ()
        message (paste0 ("EEEE> ERRO FATAL, fechando Sessao empresa_id: ",empresa_id,
          "- empresa: ", empresa, " - data_noticia: ", linha$data ))
        return ("NOK")
      } else {
        next
      }

    } else {
      falha_elemento = 0
    }

    txts <- sapply(txto, function(x) x$getElementText ())

    linha$link = driver$getCurrentUrl()[[1]]

    auxt = ""
    for (tt in txts){
      auxt = paste0(auxt, tt)
      if (substring (auxt, str_length(auxt)) != ".") auxt = paste0(auxt, ". ")
      else auxt = paste0(auxt, " ")
    }
    gsub ("..", ". ", auxt, fixed = TRUE)
    auxt = ajustarParagrafo(auxt)
    auxt = removerCaracteres(auxt)
    auxt = str_replace_all(auxt, "[\r\n]" , "")

    linha$auxt = iconv(auxt, "UTF-8", "latin1")
    Encoding(linha$auxt) <- "latin1"
    if (is.na (linha$auxt) == TRUE)
    {
      linha$auxt = auxt
      Encoding(linha$auxt) <- "UTF-8"
    }

    if (is.na (linha$auxt) == TRUE || (linha$auxt == "") == TRUE)
    {
      print (paste0("eeee> A seguinte URL produziu texto vazio, nulo ou NA e foi desconsiderada: ",
link$link))
      next
    } #if texto vazio

```

```

#Contagem de noticias lidas
nr_noticias = nr_noticias + 1

retorno = noticiasRelevantes (linha)
if (retorno$resultado == 0)
{
  print (paste0("Noticia não relevante Empresa ", linha$empresa, " Titulo ", linha$titulo))
  aux = ""
  linha$aux = ""
  next
} #if retorno$resultado

rdata = dataBase (linha$data)

registro$empresa = linha$empresa
registro$empresa_id = linha$empresa_id
registro$titulo = linha$titulo
registro$link = linha$link
registro$link_redir = linha$link_redir
registro$data = linha$data
registro$trimestre = rdata$trimestre
registro$data_base = rdata$data_base
registro$busca = linha$busca
registro$fonte = linha$fonte

#===== Dicionario Loughran
rsent = sentim (retorno$ret_texto$aux, lexico_loughran)
registro$s_lough = rsent$sentimento
registro$teor_lough = rsent$teor

#===== Dicionario Oplexicon 3.0
rsent = sentim (retorno$ret_texto$aux[[1]], lexico_oplexicon)
registro$s_oplex = rsent$sentimento
registro$teor_oplex = rsent$teor

#===== SentilexPT 2.0
rsent = sentim (retorno$ret_texto$aux, lexico_sentilex)
registro$s_stlex = rsent$sentimento
registro$teor_stlex = rsent$teor

#===== Lexico de sentimento Bing Liu
rsent = sentim (retorno$ret_texto$aux, lexico_bing_liu)
registro$s_bing = rsent$sentimento
registro$teor_bing = rsent$teor

ret = dbWriteTable (conexao, NOTICIAS_RELEVANTES_TB, registro , append=TRUE)

#Contagem de noticias relevantes
nr_noticias_relv = nr_noticias_relv + 1

} #for noticia

if ((nrResultados - (pagina)*REGISTROS) > 0)
{
  outraPagina = str_replace (paginas[2], "personalizado&sr=26",
paste0 ("personalizado&sr=", ((pagina)*REGISTROS + 1)))
  acessarPagina (noticias <- read_html(outraPagina, trim=TRUE, encoding="UTF-8"), outraPagina)

  if (typeof (retorno) == "character")

```

```

    {
    if (retorno == "NOK")
    {
        message (paste0("eeee> A seguinte URL gerou um erro não recuperável: ", outraPagina))
        print (Sys.time () - inicioExecucao)
        dbDisconnect (conexao)
        return(retorno)
    }
    }
} else {
print ("----- Finalizado -----")
break
}
} #for pagina

print (Sys.time () - inicioExecucao)

} #nrResultados !=0

if (completo != 1)
{
    inicio = format ((as.Date (fim, "%d/%m/%Y") + days(1)), "%d/%m/%Y")
    fim=format (as.Date (inicio, "%d/%m/%Y") + months(intervalo) - days(1), "%d/%m/%Y")
} else
{
    fim = format (final + days (1), "%d/%m/%Y")
}

Sys.sleep(3)

} #while

print (paste0 ("Tempo processamento= ", (Sys.time () - inicioExecucao)) )
print (paste0 ("Noticias Lidas= ", nr_noticias, " - Noticias Relevantes= ", nr_noticias_relv))

dbDisconnect (conexao)
return ("OK")

} #coletarDadosFsp

#Coleta dados Valor Econômico

require("textreadr")
require("rvest")
require("httr")
require("lubridate")
require("odbc")
require("stringr")
require("stringi")
require("DBI")
require("RODBC")
require("purrr")
require("rjson")
require("readr")
require("XML")
require ("RSelenium")

NOTICIAS_RELEVANTES_TB = "noticias_relevantes"

```

```

#Notícias cujo tempo de carregamento for maior que o tempo de espera
#não serão consideradas na análise de dados
TIMEOUT_PAGE_LOAD = 20000

EMAIL_USUARIO = "email_do_usuario"
SENHA_USUARIO = "senha_do_usuario"

source(file="comum.R")
source(file="mtexto.R")

#Conexao Browser
if (exists("servChrome") == TRUE)
{
  servChrome$server$stop()
  rm(servChrome)
}

if (exists("driver") == TRUE)
{
  driver$close()
  rm(driver)
}

servChrome <- rsDriver(browser=c("chrome"), port = 8882L, chromever= "79.0.3945.36")
driver <- servChrome$client
driver$setTimeout(type = "page load", milliseconds = TIMEOUT_PAGE_LOAD)

driver$navigate("https://login.globo.com/login/6668?url=&tam=WIDGET")

element <- driver$findElement(using = "css", 'input[name="login"]')
element$sendKeysToElement(list(EMAIL_USUARIO))
element <- driver$findElement(using = "css", 'input[name="password"]')
element$sendKeysToElement(list(SENHA_USUARIO))
element <- driver$findElement(using = "css", '.button.ng-scope')
element$clickElement()

coletarDadosVlr <- function(inicio, termino, consulta, empresa, empresa_id, completo, categoria)
{
  nr_noticias_relv = 0
  nr_noticias = 0
  falha_elemento = 0
  options(stringsAsFactors = FALSE)
  setwd(DIRETORIO_TRABALHO)
  REGISTROS = 10

  retorno <- tryCatch({resultado <- driver$getSessions()},
    error= {function(e) {
      message(paste0("eeee> driver$getSession() - Sem Sessao"))
      return(0)}})

  if (is.numeric(retorno) == TRUE)
  {
    message("Sem sessão aberta no Browser")
    return("NOK")
  }

  linha = data.frame(empresa=c(""), empresa_id=c(0), titulo=c(""), descricao=c(""), link=c(""),
    link_redir=c(""), data=c(Sys.Date()), aux=c(""), busca=consulta,
    fonte=c("Valor Economico"), data_extracao=c(Sys.Date()))

```

```

registro = data.frame (empresa=c (""), empresa_id=c (0),titulo=c (""), link=c (""), link_redir=c (""),
                      data=c (Sys.Date ()), trimestre=c (""), data_base=c (Sys.Date ()),
                      s_lough=c (0.0), teor_lough=c (0), s_oplex=c (0.0), teor_oplex=c (0),
                      s_stlex=c (0.0), teor_stlex=c (0), s_bing=c (0.0), teor_bing=c (0),
                      busca=c (""), fonte=c ("Valor Economico"), data_leitura=c (Sys.Date ()))

final = dmy (termino)
if (completo == 1)
{
  fim = termino
} else
{
  intervalo = 1
  if (abs (month (final) - month (dmy (inicio))) %% 2 == 1)
  {
    intervalo = 2
  }
  fim = format (as.Date (inicio, "%d/%m/%Y") + months (intervalo) - days (1), "%d/%m/%Y")
}

if (categoria == "Finanças" || categoria == "finanças")
{
  categoria= "Finan%25C3%25A7as"
} else if (categoria == "empresas")
{
  categoria = "Empresas"
}

inicioExecucao = Sys.time ()
conexao <- dbConnect (odbc (), .connection_string=CONEXAO_STR)

if (str_detect (consulta, "[+]") == TRUE)
{
  consulta = str_replace_all (consulta, "[+]", "%25252B")
}
consulta = str_replace_all (consulta, " ", "%2B")

while (as.Date (fim, "%d/%m/%Y") <= final)
{
  #Categoria Financas ou Empresas
  busca= paste0 ("https://www.valor.com.br/busca/", consulta, "?Tipos=Not%25C3%25ADcia&datainicial=",
                str_replace_all (inicio, "/", "-"), "&datafinal=", str_replace_all (fim, "/", "-"), "&Categorias=",
                categoria, "&method=ajax")

  #Le lista de noticias
  tryCatch ({ dados <- rjson::fromJSON (file=busca)},
            warning= function (w) print (w),
            error= {function (e) {message (paste0 ("eeee> dados = rjson::fromJSON (file=busca) - Lista noticias:
link= ", busca))}})

  if (exists ("dados") == FALSE) {
    Sys.sleep (10)
    message (paste0 ("eeee> dados <- rjson::fromJSON (file=busca): ", busca))
    print (Sys.time () - inicioExecucao)
    dbDisconnect (conexao)
    return ("NOK")
  }
}

```

```

write (dados$result, "lista.html")
noticias= read_html ("lista.html")

#proxima iteracao
rm (dados)

nos = html_nodes (noticias, ".search-result-term:contains('Mostrando')")
resultados = str_split (str_wrap (html_text (nos)), " de ")
if (length(resultados) != 0)
{
  if (str_detect (resultados[[1]][2], "resultado") == TRUE){

    aux= resultados[[1]][2]
    nrResultados = as.numeric (str_split (aux, " ")[1][1])

    if (nrResultados < 1 && completo == 1)
    {
      print (paste0("eeee> Nao foi possivel obter resultados no periodo especificado: ", busca))
      print (Sys.time () - inicioExecucao)
      dbDisconnect (conexao)
      return("NOK")
    }
  }
} else {
  print (paste0("eeee> Nao foi possivel obter resultados no periodo especificado: ", busca))
  print (Sys.time () - inicioExecucao)
  dbDisconnect (conexao)
  return("NOK")
}

print (paste0("====> Numero de noticias: ", nrResultados))

if (nrResultados != 0)
{
  for (pagina in 1: ((nrResultados %/% REGISTROS) + 1))
  {
    nosDom = html_nodes (noticias, ".search-result-title a")
    titulos = html_text (nosDom)

    links = html_attr(nosDom, "href")

    nosDom= html_nodes (noticias, ".search-result-item p")
    descricoes= html_text (nosDom)

    nosDom = html_nodes (noticias, ".date")
    datas = html_text (nosDom)

    publicados = 1:length (datas)
    class (publicados) <- "Date"
    for (i in 1:length (datas))
    {
      partes = str_split (datas[i], "[ às ]")
      publicados[i] = as.Date (partes[[1]][1], format = '%d/%m/%Y')
    }

    #iterar pelos titulos
    for (noticia in 1:length(titulos))
    {
      aux = ""
      linha$aux = ""
    }
  }
}

```

```

lnk_redir = ""

linha$link = str_replace_all (str_wrap(links[noticia]),"[\r\n]", "")
linha$empresa = empresa
linha$empresa_id = empresa_id

linha$titulo = iconv(str_wrap(titulos[noticia]), "UTF-8", "latin1")
Encoding(linha$titulo) <- "latin1"

linha$descricao = iconv(str_replace_all(str_wrap(removerCaracteres(descricoes[noticia])), "[\n]", " "),
"UTF-8", "latin1")
Encoding(linha$descricao) <- "latin1"

linha$data= publicados[noticia]

lnk = linha$link
if (str_detect (lnk, "br/video/") == TRUE || str_detect (lnk, "/curtas") == TRUE ||
    str_detect (lnk, "/destaque") == TRUE)
{
  print (paste0("iiii> Vídeo, curtas e destaque: link= ", lnk))
  next
}

iniExec = Sys.time ()
retorno <- tryCatch ({driver$navigate (lnk)},
  warning= function (w) print (w),
  error= function(e) {
    warning (paste0("eeee> driver$navigate (lnk) - Leitura do texto da noticia: link= ",
      lnk, " - Tempo: ", (Sys.time () - iniExec)))
    return (0)
  })

if (!is.null (retorno)){
  print (paste0("eeee> - driver$navigate (lnk) - Falha na leitura do texto da noticia: link= ", lnk))
}

txto <- tryCatch (
  {driver$findElement (using = "css selector", ".n-content")},
  warning= function (w) print (w),
  error=function(e) {
    warning (paste0("eeee> driver$findElement (using = 'css', '.n-content') Leitura do texto da noticia: link=
", lnk))
    return(NA)}
)

if (is.object (txto) == FALSE){
  print (paste0("eeee> driver$findElement (using = css selector, .n-content) - Falha na obtencao do
elemento da noticia: link= ", lnk))
  falha_elemento = falha_elemento + 1
  if (falha_elemento > 2)
  {
    driver$close ()
    message (paste0 ("EEEE> ERRO FATAL, fechando Sessao empresa_id: ",
      empresa_id, " - empresa: ", empresa, " - data_noticia: ", linha$data ))
    return ("NOK")
  } else {
    next
  }
}

```

```

} else {
  falha_elemento = 0
}

tryCatch ({txts <- txto$getElementText ()},
  warning= function (w) print (w),
  error= {function(e) {warning (paste0("eeee> txts <- txto$getElementText () Leitura do texto da
noticia: link= ", lnk))}})

if (typeof (txts[[1]]) == "character")
{
  texto = ajustarParagrafo(txts)
} else {
  print (paste0("eeee> Falha na leitura do texto da noticia: link= ", lnk))
  next
}

lnk_redir = driver$getCurrentUrl()[[1]]

if (typeof (lnk_redir) == "character")
{
  if (str_detect (lnk_redir, str_replace (lnk, "/u/", "/empresas/")) == TRUE)
    linha$link_redir = lnk_redir
}

linha$auxt = iconv(auxt, "UTF-8", "latin1")
Encoding(linha$auxt) <- "latin1"

if (is.na (linha$auxt) == TRUE || (linha$auxt == "") == TRUE)
{
  print (paste0("eeee> A seguinte URL produziu texto vazio, nulo ou NA e foi desconsiderada: ", lnk))
  next
}

#Contagem de noticias lidas
nr_noticias = nr_noticias + 1

retorno = noticiasRelevantes (linha)
if (retorno$resultado == 0)
{
  print (paste0("Noticia não relevante da Empresa ", linha$empresa, " Titulo: ", linha$titulo, " Link ",
linha$link, " Link redir: ", linha$link_redir))
  Sys.sleep (1.0)
  next
} #if retorno$resultado

#calculo de data_base e trimestre
rdata = dataBase (linha$data)

registro$empresa= linha$empresa
registro$empresa_id= linha$empresa_id
registro$titulo= linha$titulo
registro$link= linha$link
registro$link_redir= linha$link_redir
registro$data= linha$data
registro$trimestre= rdata$trimestre
registro$data_base= rdata$data_base
registro$busca= linha$busca
registro$fonte= linha$fonte

```



```

#===== Dicionario Loughran
rsent= sentim (retorno$ret_texto$auxt, lexico_loughran)
registro$s_lough= rsent$sentimento
registro$teor_lough = rsent$teor

#===== Dicionario Oplexicon 3.0
rsent= sentim (retorno$ret_texto$auxt[[1]], lexico_oplexicon)
registro$s_oplex= rsent$sentimento
registro$teor_oplex= rsent$teor

#===== SentilexPT 2.0
rsent= sentim (retorno$ret_texto$auxt, lexico_sentilex)
registro$s_stlex= rsent$sentimento
registro$teor_stlex= rsent$teor

#===== Lexico de sentimento Bing Liu
rsent= sentim (retorno$ret_texto$auxt, lexico_bing_liu)
registro$s_bing= rsent$sentimento
registro$teor_bing = rsent$teor

ret = dbWriteTable (conexao, NOTICIAS_RELEVANTES_TB , registro , append=TRUE)

#Contagem de noticias relevantes
nr_noticias_relv= nr_noticias_relv + 1

} #for noticia

if ((nrResultados - (pagina)*REGISTROS) > 0)
{
  #Categoria Finanças Empresas
  outraPagina= paste0 ("https://www.valor.com.br/busca/", consulta, "?page=", (pagina),
"?Tipos=Not%25C3%25ADcia&datainicial=", str_replace_all(inicio, "/", "-"),
"&datafinal=", str_replace_all(fim, "/", "-"), "&Categorias=", categoria, "&method=ajax")

  rm (publicados)

  tryCatch ({dados = rjson::fromJSON(file=outraPagina)},
    warning= function (w) print (w),
    error= {function(e) {warning (paste0("eeee> dados = rjson::fromJSON(file=outraPagina) - Lista
noticias: link= ", outraPagina))}})

  if (exists ("dados") == FALSE)
  {
    Sys.sleep(10)
    message (paste0("eeee> dados <- rjson::fromJSON(file=busca): ", busca))
    print (Sys.time () - inicioExecucao)
    dbDisconnect (conexao)
    return("NOK")
  }

  if (str_length(dados$result) == 0)
  {
    message (paste0("eeee> A seguinte URL gerou um erro não recuperável: ", outraPagina))
    print (Sys.time () - inicioExecucao)
    dbDisconnect (conexao)
    return("NOK")
  }

  write (dados$result, "lista.html")

```

```

    noticias = read_html ("lista.html")

    rm (dados)

    } else
    {
        print ("----- Finalizado Periodo -----")
        break
    }
} #for pagina

print (Sys.time () - inicioExecucao)
} #nrResultados !=0

if (completo != 1)
{
    inicio = format ((as.Date (fim, "%d/%m/%Y") + days(1)), "%d/%m/%Y")
    fim=format (as.Date (inicio, "%d/%m/%Y") + months(intervalo) - days(1), "%d/%m/%Y")
} else
{
    fim = format (final + days (1), "%d/%m/%Y")
}

Sys.sleep(3) #Fim do Periodo

} #while

print (Sys.time () - inicioExecucao)
print (paste0 ("Noticias Lidas= ", nr_noticias, " - Noticias Relevantes= ", nr_noticias_relv))
dbDisconnect (conexao)
return("OK")

} #coletarDadosVlr

#Coleta dados O Estado de São Paulo

require("textreadr")
require("rvest")
require("lubridate")
require("odbc")
require("stringr")
require("stringi")
require("DBI")
require("RODBC")
require("httr")
require("purrr")
require("RSelenium")

NOTICIAS_RELEVANTES_TB = "noticias_relevantes"

#Notícias cujo tempo de carregamento for maior que o tempo de espera
#não serão consideradas na análise de dados
TIMEOUT_PAGE_LOAD = 20000

EMAIL_USUARIO = "email_do_usuario"
SENHA_USUARIO = "senha_do_usuario"

source(file="comum.R")
source(file="mtexto.R")

```

```

#Conexao Browser
if (exists ("servChrome") == TRUE)
{
  servChrome$server$stop()
  rm (servChrome)
}

if (exists ("driver") == TRUE)
{
  driver$close ()
  rm (driver)
}

#Variavel global
servChrome <- rsDriver (browser=c ("chrome"), port = 8882L, chromever= "79.0.3945.36")
driver <- servChrome$client
driver$setTimeout (type = "page load", milliseconds = TIMEOUT_PAGE_LOAD)

driver$navigate ("https://acesso.estadao.com.br/login/")

element <- driver$findElement (using = "css", 'input[name="email_login"]')
element$sendKeysToElement (list (EMAIL_USUARIO))
element <- driver$findElement (using = "css", 'input[name="senha"]')
element$sendKeysToElement (list (SENHA_USUARIO))

element <- driver$findElement (using = "css", 'input[type="submit"]')
element$clickElement ()

coletarDadosEsp <- function (inicio, termino, consulta, empresa, empresa_id, completo)
{

  nr_noticias_relv = 0
  nr_noticias = 0
  falha_elemento = 0
  options(stringsAsFactors = FALSE)
  REGISTROS = 10
  setwd (DIRETORIO_TRABALHO)

  retorno <- tryCatch ({resultado <- driver$getSessions()},
    error= {function(e) {
      message (paste0("eeee> driver$getSession() - Sem Sessao"))
      return (0)}})

  if (is.numeric (retorno) == TRUE)
  {
    message ("Sem sessão aberta no Browser")
    return ("NOK")
  }

  linha = data.frame (empresa=c (""), empresa_id=c (0), titulo=c (""), descricao=c (""), link=c (""),
    link_redir=c (""), data=c (Sys.Date ()), auxt=c (""), busca=consulta,
    fonte=c ("O Estado de São Paulo"), data_extracao=c (Sys.Date ()))

  registro = data.frame (empresa=c (""), empresa_id=c (0), titulo=c (""), link=c (""), link_redir=c (""),
    data=c (Sys.Date ()), trimestre=c (""), data_base=c (Sys.Date ()),
    s_lough=c (0.0), teor_lough= c (0), s_oplex=c (0.0), teor_oplex=c (0),
    s_stlex= c (0.0), teor_stlex=c (0), s_bing=c (0.0), teor_bing=c (0),
    busca=c (""), fonte= c ("Estado de São Paulo"), data_leitura=c (Sys.Date ()))

```

```

final = dmy(termino)
if (completo == 1)
{
  fim = termino
} else
{
  intervalo = 1
  if (abs (month (final) – month (dmy (inicio))) %% 2 == 1)
  {
    intervalo = 2
  }
  fim = format (as.Date (inicio, "%d/%m/%Y") + months (intervalo) – days (1), "%d/%m/%Y")
}

inicioExecucao = Sys.time ()
conexao <- dbConnect (odbc (),connection_string=CONEXAO_STR)

#Vários termos entre aspas
consultaOriginal = consulta
if (str_detect (consulta, "[+]" ) == FALSE)
{
  consulta = paste0("%22", consulta, "%22")
}
consulta = str_replace_all (consulta, " ", "%20")

while (as.Date (fim, "%d/%m/%Y") <= final)
{

  busca = sprintf ("https://busca.estadao.com.br/?tipo_contenido=Notícias&quando=%s-
%s&q=%s&editoria[]=Economia", inicio, fim, consulta)
  periodo = paste0 ("quando%3D", str_sub (inicio, 1, 2),"%252F", str_sub (inicio, 4, 5), "%252F",str_sub
(inicio, 7, 10)
    ,"-", str_sub (fim, 1, 2),"%252F", str_sub (fim, 4, 5),"%252F", str_sub (fim, 7, 10))

  #Le lista de noticias
  noticias = read_html (busca, trim=TRUE, encoding="UTF-8")
  nosDom = html_nodes (noticias, ".flt-registros")
  resultados = str_wrap (html_text (nosDom))
  nrResultados = as.numeric (str_split (resultados, "[ ]")[[1]][3])

  print (paste0("Numero de registros: ", nrResultados))

  if (nrResultados != 0)
  {
    for (pagina in 1: ((nrResultados %% REGISTROS) + 1))
    {
      nosDom = html_nodes (noticias, "h3")
      titulos = html_text (nosDom)

      nosDom = html_nodes(noticias,"a p")
      descricoes = html_text (nosDom)

      nosDom = html_nodes(noticias,".data-posts")
      datas = html_text (nosDom)

      publicados = 1:length (datas)
      class (publicados) <- "Date"
      for (i in 1:length (datas))
      {
        partes = str_split (datas[i], "[ ]")

```

```

publicados[i]= as.Date (paste0(partes[[1]][1], partes[[1]][3], partes[[1]][5]), format = '%d%b%Y')
}

nosDom = html_nodes (noticias, ".link-title")
links = html_attr(nosDom,"href")

#Iterar pelos titulos
for (noticia in 1:length(titulos))
{
  aux1 = ""
  linha$aux1 = ""

  linha$link = str_replace_all (str_wrap(links[noticia]),"[\r\n]", "")
  if (str_detect (linha$link, "blogs.estadao.com.br") == TRUE
      || str_detect (linha$link, "economia.estadao.com.br/blogs/") == TRUE)
  {
    print (paste0("iiii> Desconsiderando Blog: ", linha$link))
    next
  }

  linha$link_redir = ""
  linha$empresa = empresa
  linha$empresa_id = empresa_id

  ttl = str_replace_all (str_wrap(titulos[noticia]), "\u20AC", "EUR")
  linha$titulo = str_replace_all(ttl, "[\r\n]", "")
  linha$titulo = iconv(linha$titulo, "UTF-8", "latin1")
  Encoding(linha$titulo) <- "latin1"

  linha$descricao = iconv(str_replace_all(str_wrap(removerCaracteres(descricoes[noticia])),
                                          "[\n]", " "),"UTF-8", "latin1")
  Encoding(linha$descricao) <- "latin1"

  linha$data= publicados[noticia]

  lnk = linha$link
  iniExec = Sys.time ()

  retorno <- tryCatch ({driver$navigate (lnk)},
                      warning= function (w) print (w),
                      error= function(e) {
                        warning (paste0("eeee> driver$navigate (lnk) - Leitura do texto da noticia: link= ",
                                          lnk, " - Tempo: ", (Sys.time () - iniExec )))
                        return (0)
                      })

  if (!is.null (retorno)){
    print (paste0("eeee> - driver$navigate (lnk) - Falha na leitura do texto da noticia: link= ", lnk))
  }

  aux2 = ".n--noticia__content p:not ([class]),.content-text"

  txto <- tryCatch (
    {driver$findElements(using = "css selector", aux2)},
    warning= function (w) print (w),
    error=function(e) {
      warning (paste0("eeee> driver$findElement (using = 'css','Seletores') Leitura do texto da noticia: link=
", lnk))
      return (NA)}

```

```

)

if (is.list (txto) == FALSE){
  print (paste0("eeee> driver$findElement (using = css selector,'Seletores') - Falha elementos da noticia:
link= ", lnk))
  falha_elemento = falha_elemento + 1
  if (falha_elemento > 2)
  {
    driver$close ()
    message (paste0 ("EEEE> ERRO FATAL, fechando Sessao empresa_id: ",empresa_id,
                    "- empresa: ", empresa, " - data_noticia: ", linha$data))
    return ("NOK")
  } else {
    next
  }

} else {
  falha_elemento = 0
}

txts <- sapply(txto, function(x) x$getElementText ())

linha$link = driver$getCurrentUrl()[[1]]
auxt = ""
for (tt in txts){
  auxt = paste0(auxt, tt)
}
auxt = removerCaracteres(auxt)
auxt = str_replace_all (auxt, "[\r\n]" , "")
linha$auxt = iconv (auxt, "UTF-8", "latin1")
Encoding (linha$auxt) <- "latin1"

if (is.na (linha$auxt) == TRUE || (linha$auxt == "") == TRUE || str_detect (linha$auxt, "html, body ") ==
TRUE)
{
  print (paste0("eeee> A seguinte URL produziu texto vazio, nulo, com TAGS HTML ou NA e foi
desconsiderada: ", linha$link))
  next
} #texto vazio

#Contagem de noticias lidas
nr_noticias = nr_noticias + 1

retorno = noticiasRelevantes (linha)
if (retorno$resultado == 0)
{
  auxt = ""
  linha$auxt = ""
  next
} #if retorno$resultado

#calculo de data_base e trimestre
rdata = dataBase (linha$data)

registro$empresa = linha$empresa
registro$empresa_id = linha$empresa_id
registro$titulo = linha$titulo
registro$link = linha$link
registro$link_redir = linha$link_redir
registro$data = linha$data

```

```

registro$trimestre = rdata$trimestre
registro$data_base = rdata$data_base
registro$busca = linha$busca
registro$fonte = linha$fonte

#===== Dicionario Loughran
rsent = sentim (retorno$ret_texto$auxt, lexico_loughran)
registro$s_lough = rsent$sentimento
registro$teor_lough = rsent$teor

#===== Dicionario Oplexicon 3.0
rsent = sentim (retorno$ret_texto$auxt[[1]], lexico_oplexicon)
registro$s_oplex = rsent$sentimento
registro$teor_oplex = rsent$teor

#===== SentilexPT 2.0
rsent = sentim (retorno$ret_texto$auxt, lexico_sentilex)
registro$s_stlex = rsent$sentimento
registro$teor_stlex = rsent$teor

#===== Lexico de sentimento Bing Liu
rsent = sentim (retorno$ret_texto$auxt, lexico_bing_liu)
registro$s_bing = rsent$sentimento
registro$teor_bing = rsent$teor

ret = dbWriteTable (conexao, NOTICIAS_RELEVANTES_TB , registro , append=TRUE)

#Contagem de noticias relevantes
nr_noticias_relv = nr_noticias_relv + 1

} #for noticia

if ((nrResultados - (pagina)*REGISTROS) > 0)
{
    outraPagina= paste0("https://busca.estadao.com.br/modulos/busca-resultado?modulo=busca-
resultado&config%5Bbusca%5D%5Bpage%5D=", (pagina + 1),
"&config%5Bbusca%5D%5Bparams%5D=tipo_conteudo%3DNot%C3%ADcias%26",
perido, "%26q%3D", consulta, "%26editoria%255B%255D%3DEconomia&ajax=1")

    rm (publicados)

    retorno <- acessarPagina (noticias <- read_html(outraPagina, trim=TRUE, encoding="UTF-8"),
outraPagina)

    if (typeof (retorno) == "character")
    {
        if (retorno == "NOK")
        {
            message (paste0("eeee> A seguinte URL gerou um erro não recuperável: ", outraPagina))
            print (Sys.time () - inicioExecucao)
            dbDisconnect (conexao)
            return(retorno)
        } #NOK
    } #OK

} else
{
    print ("----- Finalizado Periodo -----")
}

```

```

        break
    }
} #for pagina

print (Sys.time () - inicioExecucao)

} #nrResultados !=0

if (completo != 1)
{
    inicio = format ((as.Date (fim, "%d/%m/% Y") + days(1)), "%d/%m/% Y")
    fim=format (as.Date (inicio, "%d/%m/% Y") + months(intervalo) - days(1), "%d/%m/% Y")
} else {
    fim = format (final + days (1), "%d/%m/% Y")
}

Sys.sleep(3) #Fim do Periodo

} #while

print (paste0("Tempo processamento= ", (Sys.time () - inicioExecucao)))
print (paste0 ("Noticias Lidas= ", nr_noticias, " - Noticias Relevantes= ", nr_noticias_relv))
dbDisconnect (conexao)
return("OK")

} #coletarDadosEsp

```

### **#Mineração de texto e análise de sentimento – arquivo mtexto.R**

```

require ("odbc")
require ("DBI")
require ("RODBC")
require ("textreadr")
require ("stringr")
require ("lubridate")

#Mineracao de texto
require ("tm")
require ("SnowballC") #stopwords
require ("quanteda") #dictionary
require ("dplyr") #pull
require ("tokenizers")
require ("tidyr")

NR_OCORRENCIAS_TEXTO = 2
NR_PALAVRAS_INICIAIS = 25
NR_MINIMO_PALAVRAS = 50
LIMIAR = 0.2
TRI = 3

trimestres= data.frame (mes= (1:12),
                        trimestre=c (rep ("1T", TRI), rep ("2T", TRI), rep ("3T", TRI), rep ("4T", TRI)) ,
                        data_base=c (rep ("-03-31", TRI), rep ("-06-30", TRI), rep ("-09-30", TRI), rep ("-12-
31", TRI)))

if(!exists("acessoPagina", mode="function")) source ("comum.R")

options(stringsAsFactors = FALSE)
setwd (DIRETORIO_TRABALHO)

```



```

#Carrega lexicos de sentimento

#===== Oplexicon 3.0 - UTF-8
dic_oplexicon = read.csv ("oplexicon_v3.0/lexico_v3.0_header.csv", header = TRUE,
                          as.is = TRUE, sep=";", fileEncoding = "utf-8")
positiva = unique (subset (dic_oplexicon, (polaridade == 1)))
negativa = unique (subset (dic_oplexicon, (polaridade == -1)))

#Transforma em um array de palavras dplyr
pst = pull(positiva, palavra)
ngt = pull(negativa, palavra)
lexico_oplexicon = quanteda::dictionary (list (positiva = pst, negativa= ngt))

#===== SentilexPt
dic_sentilex = read.csv ("SentiLex-PT02/SentiLex-flex-PT02_ansi_br.csv", header = TRUE,
                          as.is = TRUE, sep=";")
positiva = unique (subset (dic_sentilex, (polaridade == 1)))
negativa = unique (subset (dic_sentilex, (polaridade == -1)))

pst = pull(positiva, palavra)
ngt = pull(negativa, palavra)
lexico_sentilex = quanteda::dictionary (list (positiva = pst, negativa= ngt))

#=====Loughran
dic_loughran = read.csv ("LoughranMcDonald/dicionario_loughran_mcdonald_2018.csv", header = TRUE,
                          as.is = TRUE, sep=";")
positiva <- unique (subset (dic_loughran, (polaridade == 1)))
negativa <- unique (subset (dic_loughran, (polaridade == -1)))
pst <- pull (positiva, palavra)
ngt <- pull (negativa, palavra)
lexico_loughran <- quanteda::dictionary (list (positiva = pst, negativa= ngt))

#===== Bing Liu
dic_bing_liu = read.csv ("opinion_lexicon_bing_liu/lexico_bing_liu.csv", header = TRUE,
                          as.is = TRUE, sep=";")
positiva <- unique (subset (dic_bing_liu, (polaridade == 1)))
negativa <- unique (subset (dic_bing_liu, (polaridade == -1)))
pst <- pull (positiva, palavra)
ngt <- pull (negativa, palavra)
lexico_bing_liu <- quanteda::dictionary (list (positiva = pst, negativa= ngt))

dataBase <- function (dt)
{
  mes = as.numeric (format (dt, "%m"))
  ano = as.numeric (format (dt, "%Y"))
  dbase = as.Date (paste0(ano, trimestres[mes, ]$data_base))
  tr = paste0 (trimestres[mes, ]$trimestre , ano)
  return (list (trimestre=tr, data_base=dbase))
} #dataBase

preprocessamento <- function (titulo_texto)
{
  documento <- VCorpus(VectorSource (titulo_texto))

  #Transforma texto em minusculas
  documento = tm_map(documento, content_transformer(tolower))

  #Ajusta nomes compostos
  documento[[1]]$content <- str_replace_all(documento[[1]]$content,

```

```

        "banco do brasil", "banco_do_brasil")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "lojas americanas", "lojas_americanas")
documento[[1]]$content <- str_replace_all(documento[[1]]$content, "
        pão de açúcar", "pão_de_açúcar")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "br malls", "br_malls")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "energias do brasil", "energias_do_brasil")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "magazine luiza", "magazine_luiza")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "cvc brasil", "cvc_brasil")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "via varejo", "via_varejo")
#Não presentes no Ibovespa
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "aes tiete", "aes_tiete")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "alfa holdings", "alfa_holdings")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "andrade gutierrez", "andrade_gutierrez")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "banco mercantil de investimentos", "banco_mercantil_de_investimentos")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "banco mercantil do brasil", "banco_mercantil_do_brasil")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "banco do nordeste", "banco_do_nordeste")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "bk brasil", "bk_brasil")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "gp investments", "gp_investments")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "porto seguro", "porto_seguro")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "rb capital", "rb_capital")
documento[[1]]$content <- str_replace_all(documento[[1]]$content,
        "eco securitizadora", "eco_securitizadora")

#remoção de números
documento = tm_map (documento, removeNumbers)

#remoção de pontuação
documento = tm_map (documento, removePunctuation)

#remoção de palavras irrelevantes
documento = tm_map (documento, removeWords, stopwords("portuguese"))

#remoção de espaços em branco adicionais
documento = tm_map (documento, stripWhitespace)

return (list (doc=documento, aux=documento[[1]]$content))
} #preprocessamento

sentim <- function (texto, lexico)
{
  qp = 0
  qn = 0
  termos <- Boost_tokenizer (texto)

```

```

for (termo in termos)
{
  tp = sum (lexico[[1]] == termo)
  tn = sum (lexico[[2]] == termo)
  qp = qp + tp
  qn = qn + tn
}

sentimento = 0
if ((qp + qn) != 0)
{
  sentimento = (qp - qn) / (qp + qn)
}

if (sentimento >= LIMIAR)
{
  teor = 1
} else if (sentimento < LIMIAR && sentimento > -LIMIAR){
  teor = 0
} else {
  teor = -1
}

return (list (sentimento=sentimento, teor=teor, qp=qp, qn=qn))

} #sentim

noticiasRelevantes <- function (noticia)
{
  options(stringsAsFactors = FALSE)
  retorno <- list (resultado=0, ret_texto=NULL)

  #verificação relevancia titulo e noticia
  titulo_texto <- paste0 (noticia$titulo, " . ", noticia$auxt)
  ret_titulo_texto <- preprocessamento (titulo_texto)
  ret_texto <- preprocessamento (noticia$auxt)

  termos <- Boost_tokenizer (ret_titulo_texto$auxt)
  termos_texto <- Boost_tokenizer (ret_texto$auxt)

  if (length(termos_texto) < NR_MINIMO_PALAVRAS)
  {
    return (retorno)
  }

  #verificar ocorrência do nome da empresa nas primeiras NR_MINIMO_PALAVRAS palavras
  #(titulo + texto)
  busca = str_replace_all (tolower (noticia$busca), "[._]", "")
  empresa = str_replace_all (tolower (noticia$empresa), "[._]", "")

  r1 = str_detect (termos[1:NR_PALAVRAS_INICIAIS] , busca)

  #Texto original
  busca_empresa = tolower (noticia$busca) != tolower(noticia$empresa)
  if (busca_empresa)
  {
    r2 = str_detect (termos[1:NR_PALAVRAS_INICIAIS] , empresa)
    p25 = sum(r1) + sum (r2)
  }
}

```

```

} else {
  p25 = sum(r1)
}

if (p25 >= 1) #nome da empresa nas primeiras 25 palavras
{
  p1 = str_detect (termos_texto, busca)
  if (busca_empresa)
  {
    p2 = str_detect (termos_texto , empresa)
    ptotal = sum(p1) + sum (p2)
  } else {
    ptotal = sum(p1)
  }

  if (ptotal < NR_OCORRENCIAS_TEXTO)
  {
    return (retorno)
  } else {
    texto_vcorpus <- ret_texto
    retorno$resultado <- 1
    retorno$ret_texto <- ret_texto
    return (retorno)
  }
} else {
  return (retorno)
}
} #noticiasRelevantes

#Teste modelos de dados em painel e premissas do método MQO
if(!require("xlsx")){install.packages("xlsx")}
if(!require("plm")){install.packages("plm")}
if(!require("lmtest")){install.packages("lmtest")}
if(!require("DescTools")){install.packages("DescTools")}
if(!require("mctest")){install.packages("mctest")}
if(!require("sqldf")){install.packages("sqldf")}
if(!require("sandwich")){install.packages("sandwich")}
if(!require("lm.beta")){install.packages("lm.beta")}
if(!require("moments")){install.packages("moments")} #skewness
if(!require("xtable")){install.packages("xtable")} #corrplot
if(!require("Hmisc")){install.packages("Hmisc")}

#x matrix com o dados
#method: metodo da correlação: "pearson" or "spearman"
#removeTriangle: não exibe triangulo inferior ou superior
#results: tipo de resultado "html", "latex"
#Adaptado de: www.sthda.com/english/wiki/elegant-correlation-table-using-xtable-r-package
corstars <- function(x, method= c ("pearson", "spearman"), removeTriangle= c ("upper", "lower"),
  result= c ("none", "html", "latex"))
{

  x <- as.matrix (x)
  correlation_matrix<-rcorr (x, type=method[1])
  R <- correlation_matrix$r
  p <- correlation_matrix$p

  mystars <- ifelse (p < .001, "*** ", ifelse (p < .01, "** ", ifelse (p < .05, "* ", " ")))

  R <- format ( round (cbind( rep(-1.11, ncol(x)), R), 2))[-1]

```

```

Rnew <- matrix (paste (R, mystars, sep=""), ncol=ncol(x))
diag(Rnew) <- paste (diag (R), " ", sep="")
rownames (Rnew) <- colnames (x)
colnames (Rnew) <- paste (colnames(x), "", sep="")

if(removeTriangle[1]=="upper"){
  Rnew <- as.matrix (Rnew)
  Rnew[upper.tri (Rnew, diag = TRUE)] <- ""
  Rnew <- as.data.frame (Rnew)
}

else if (removeTriangle[1]=="lower"){
  Rnew <- as.matrix (Rnew)
  Rnew[lower.tri (Rnew, diag = TRUE)] <- ""
  Rnew <- as.data.frame (Rnew)
}

Rnew <- cbind (Rnew[1:length(Rnew)-1])
if (result[1]=="none") return(Rnew)
else{
  if (result[1]=="html") print(xtable(Rnew), type="html")
  else print(xtable(Rnew), type="latex")
}
} #corstars

#Painel sem intercepto
vif_painel <- function (modelo, todos_diagnostics = FALSE, grafico = FALSE)
{

  y <- as.numeric (modelo$model[ , 1])
  x <- modelo$model[ , -1]

  if (grafico == TRUE) mc.plot ( x , y )

  if (todos_diagnostics == FALSE) resultado <- imcdiag (x,y, method='VIF')$idiags[,1]
  if (todos_diagnostics == TRUE) resultado <- imcdiag (x,y)
  return (resultado)
} #vif_painel

VIF <- function (modelo_rq, sem_intercepto = FALSE, todos_diagnostics = FALSE, grafico = FALSE)
{
  y <- modelo_rq$model[ , 1]

  if (sem_intercepto == FALSE) x <- modelo_rq$model[ , -1]
  if (sem_intercepto == TRUE) x <- modelo_rq$model
  if (grafico == TRUE ) mc.plot ( x , y )

  if(todos_diagnostics == FALSE) resultado <- imcdiag (x,y, method='VIF')$idiags[,1]
  if(todos_diagnostics == TRUE) resultado <- imcdiag (x,y )
  return ( resultado )
} #VIF

source(file="comum.R")
setwd(DIRETORIO_TRABALHO)

dados <- readRDS ("dados_dissertacao_cgomes_defesa.rds")

#Winsorizing

```

```

var_winsorizadas= c("ROA", "ROE", "MVA", "QTB", "ALV")
for (var in var_winsorizadas)
{
  if (var == "ROE")
    dados[[var]] <- Winsorize ( dados[[var]], probs = c(0.02, 0.98), na.rm = TRUE, type = 5 )
  else
    dados[[var]] <- Winsorize ( dados[[var]], probs = c(0.01, 0.99), na.rm = TRUE, type = 5 )
}

var_descr = c ("ROA", "ROE", "MVA", "QTB", "FAV", "VIS", "FAVxVIS",
              "ALV", "IDA", "TAM", "IMR", "IND1", "IND2", "NM", "N1", "N2")

#=====
#DESCRITIVAS
#=====
dados_desc= dados [ , var_descr]

sink (paste0 ("Descritivas_Dissertacao_Defesa.txt"))

for (variavel in var_descr)
{
  print ( paste0( variavel,
                "#",sum (!is.na(dados[[variavel]])),
                "#",min(dados[ ,variavel], na.rm=TRUE),
                "#",max(dados[ ,variavel], na.rm=TRUE),
                "#",mean(dados[ ,variavel], na.rm=TRUE),
                "#",median(dados[ ,variavel], na.rm=TRUE),
                "#",sd(dados[ ,variavel], na.rm=TRUE),
                "#",sd(dados[ ,variavel], na.rm=TRUE)/mean(dados[ ,variavel], na.rm=TRUE),
                "#",skewness(dados[ ,variavel], na.rm=TRUE)))
}

sink()

sink( paste0 ("Correlacoes_Dissertacao_Defesa.html") )
corstars (dados_desc, method = "pearson", removeTriangle= "upper", result= "html")
sink ()

rm (dados_desc)

#Testes modelos de dados em painel
testes_painel <- function (form, dados)
{
  print ("=====")
  print (form)
  print ("=====")

  tx2 <- dplyr::count (dados, EMP)
  dx1 <- merge (dados, subset (tx2, tx2$n > 17))
  dx1$n <- NULL
  dados <- dx1

  polsi <- plm (formula= form, data= dados, index= c ("EMP", "TRI"),
              model= "pooling", na.action= "na.exclude")
  rst = plmtest (polsi, type= c ("bp"))
  print (paste0 ("Teste LM Breusch-Pagan - POLS (H0) x Aleatório (Ha) - Estatística: ",
              round (rst$statistic, 4), " p-value: ", round (rst$p.value, 4)) )

  aleatorio <- plm (formula= form, data= dados, index= c ("EMP", "TRI"),
                  model= "random", na.action= "na.exclude")

```

```

fixo <- plm (formula= form, data=dados, index= c ("EMP", "TRI"),
            model= "within", na.action= "na.exclude")

rst = phtest (fixo, aleatorio)
print (paste0 ("Teste Hausman - Aleatório (H0) x Fixo (Ha) - Estatística: ",
              round (rst$statistic, 4), " p-value: ", round (rst$p.value, 4)))

rst <- tryCatch ( {pooltest (as.formula (form), data=dados, index= c ("EMP", "TRI"),
                           model= "within", na.action= "na.exclude")},
                 warning= function (w) print(w), error= function(e) {return (0)})
if (identical (rst, 0) != TRUE)
{
  print (paste0 ("Teste Chow - POLS (H0) x FE (Ha) - Estatística: ",
                round (rst$statistic, 4), " p-value: ", round (rst$p.value, 4)))
} else {
  print (paste0 ("Teste Chow - POLS (H0) x FE (Ha) - Não foi possível testar: ND"))
}

print ("=====")

} #testes_painel

#Testes premissas da regressão linear
testes_mqo <- function (mod)
{
  KMSMIR = 1.62762
  print ("=====")
  print (mod$call)
  print (mod$formula)
  print ("=====")

  rst = lmtest::bptest ( mod )
  print (paste0 ("Teste Breusch-Pagan - Homocedasticidade (H0), Heterocedasticidade (Ha) Estatística: ",
                round (rst$statistic, 4), " p-value: ", round (rst$p.value, 4)))

  rst = plm::pbgttest( mod )
  print (paste0 ("Teste Breusch-Godfrey - Sem Autocorrelação (H0), Autocorrelacao (Ha) Estatística: ",
                round (rst$statistic, 4), " p-value: ", round (rst$p.value, 4)))

  rst = DescTools::JarqueBeraTest ( mod$residuals, robust=FALSE )
  print (paste0 ("Normalidade Jarque-Bera - Normalidade (H0), Não Normalidade (Ha) - Estatística: ",
                round ( rst$statistic, 4), " p-value: ", round (rst$p.value, 4)))

  rst = ks.test (mod$residuals, "pnorm")

  print (paste0 ("Teste Kolmogorov-Smirnov - Normalidade - Estatística D: ",
                round (rst$statistic, 4), " # p-value: ", round (rst$p.value, 4), " # D-Critico: ",
                round (KMSMIR/sqrt (length (mod$residuals)), 3)))
  print ("===== Variance Inflation Factor")

  print ( vif_painel (mod))
  print ("=====")
}#testes_mqo

sink ("testes_painel_defesa.txt")
#=====
#ROA

```

```

#=====
form = ROA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
testes_painel (form, dados)
form = ROA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
testes_painel (form, dados)

#=====
#ROE
#=====
form = ROE ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
testes_painel (form, dados)
form = ROE ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
testes_painel (form, dados)

#=====
#MVA
#=====
form = MVA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
testes_painel (form, dados)
form = MVA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
testes_painel (form, dados)

#=====
#QTB
#=====
form = QTB ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
testes_painel (form, dados)
form = QTB ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
testes_painel (form, dados)

sink()

sink ("testes_MQO_dissertacao_defesa.txt")
#=====
#ROA
#=====
form = ROA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
mod <- plm (formula= form, data=dados, index= c ("EMP", "TRI"), model= "pooling",
           na.action= "na.exclude")
testes_mqo (mod)

form = ROA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
mod <- plm (formula= form, data=dados, index= c ("EMP", "TRI"), model= "pooling",
           na.action= "na.exclude")
testes_mqo (mod)

#=====
#ROE
#=====
form = ROE ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
mod <- plm (formula= form, data=dados, index= c ("EMP", "TRI"), model= "pooling",
           na.action= "na.exclude")
testes_mqo (mod)

form = ROE ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
mod <- plm (formula= form, data=dados, index= c ("EMP", "TRI"), model= "pooling",
           na.action= "na.exclude")
testes_mqo (mod)

```



```

#=====
#MVA
#=====

form = MVA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
mod <- plm (formula= form, data=dados, index= c ("EMP", "TRI"), model= "within",
           effect= "time", na.action= "na.exclude")
testes_mqo (mod)

form = MVA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
mod <- plm (formula= form, data=dados, index= c ("EMP", "TRI"), model= "within",
           effect= "time", na.action= "na.exclude")
testes_mqo (mod)

#=====
#QTB
#=====

form = QTB ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
mod <- plm (formula= form, data=dados, index= c ("EMP", "TRI"), model= "within",
           effect= "time", na.action= "na.exclude")
testes_mqo (mod)

form = QTB ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
mod <- plm (formula= form, data=dados, index= c ("EMP", "TRI"), model= "within",
           effect= "time", na.action= "na.exclude")
testes_mqo (mod)

sink()

#Análise de regressão com dados em painel e efeitos fixos e algoritmo GBM

if(!require("plm")){install.packages("plm")}
if(!require("lmtest")){install.packages("lmtest")}
if(!require("DescTools")){install.packages("DescTools")}
if(!require("stats")){install.packages("stats")}
if(!require("gbm")){install.packages("gbm")}
if(!require("caret")){install.packages("caret")}
if(!require("sqldf")){install.packages("sqldf")}
if(!require("plyr")){install.packages("plyr")}
if(!require("xlsx")){install.packages("xlsx")}
if(!require("robustHD")){install.packages("robustHD")}

source(file="comum.R")
setwd(DIRETORIO_TRABALHO)

dados <- readRDS ("dados_dissertacao_cgomes_defesa.rds")

#Winsorizing
var_winsorizadas= c ("ROA", "ROE", "MVA", "QTB", "ALV")
for (var in var_winsorizadas)
{
  if (var == "ROE")
    dados[[var]] <- Winsorize (dados[[var]], probs = c(0.02, 0.98), na.rm = TRUE, type = 5)
  else
    dados[[var]] <- Winsorize (dados[[var]], probs = c(0.01, 0.99), na.rm = TRUE, type = 5)
}

estimador_robusto <- function (modelo)

```

```

{
  return ( lmtest::coefest(mod, vcov.=plm::vcovHC(mod, method="arellano", type="HC0")) )
}# estimador_robusto

sink ("regressao_resultados_defesa.txt")

#-----#
#ROA - POLS
#-----#
print ("===== ROA =====")
#=====
#CONTROLE
#=====

form = ROA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR
mod <- plm(formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
          model= "pooling")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

#=====
#INDEPENDENTES
#=====

form = ROA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
mod <- plm (formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
          model= "pooling")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

#=====
#INTERAÇÃO
#=====

form = ROA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
mod <- plm (formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
          model= "pooling")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

#-----#
#ROE - POLS
#-----#
print ("===== ROE =====")
#=====
#CONTROLE
#=====

form = ROE ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR
mod <- plm (formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
          model= "pooling")
summary (mod)

robusta <- estimador_robusto(mod)

```

```

print (robusta)

#=====
#INDEPENDENTES
#=====

form = ROE ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
mod <- plm (formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
           model= "pooling")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

#=====
#INTERAÇÃO
#=====

form = ROE ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS

mod <- plm (formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
           model= "pooling")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

#-----#
#MVA - FE
#-----#
print ("===== MVA =====")
#=====
#CONTROLE
#=====

form = MVA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR
mod <- plm (formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
           model= "within", effect = "time")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

#=====
#INDEPENDENTES
#=====

form = MVA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
mod <- plm (formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
           model= "within", effect = "time")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

#=====
#INTERAÇÃO
#=====

```

```

form = MVA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
mod <- plm (formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
           model= "within", effect = "time")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

#-----#
#QTB FE
#-----#
print ("===== QTB =====")
#=====
#CONTROLE
#=====

form = QTB ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR
mod <- plm (formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
           model= "within", effect = "time")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

#=====
#INDEPENDENTES
#=====

form = QTB ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS
mod <- plm (formula=form, data=dados, index=c("EMP", "TRI"), na.action="na.exclude",
           model= "within", effect = "time")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

#=====
#INTERAÇÃO
#=====

form = QTB ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS
mod <- plm (formula=form, data=dados, index=c ("EMP", "TRI"), na.action="na.exclude",
           model= "within", effect = "time")
summary (mod)

robusta <- estimador_robusto(mod)
print (robusta)

sink()

#=====
#VALIDAÇÃO CRUZADA
#=====

NR_FOLDS= 20
particoes_ids <- c (1:NR_FOLDS)

particao_teste <- function (particoes_ids)
{
  set.seed(9876)

```

```

selec <- NULL
part_ids <- particoes_ids

for (k in 1:(NR_FOLDS-1))
{
  teste_aleatorio <- sample (part_ids, 1, replace=FALSE)
  selec <- c (selec, teste_aleatorio)
  part_ids <- particoes_ids [-selec]
  if (length(part_ids) == 1) {
    return (c (selec, part_ids))
  }
} #for k
} #particao_teste

previsao_plm_ef <- function (mod, conjunto_teste)
{
  interceptos <- fixef (mod)
  coeficientes <- mod$coefficients
  previsao <- c()

  for (i in 1:nrow (conjunto_teste))
  {
    val = 0
    for (j in 1:length (coeficientes)){
      produto <- coeficientes [[j]]*conjunto_teste[[attr (coeficientes, "names", 1)[j]][i]
      val = val + produto
    }

    if (mod$args$effect == "individual"){
      prv <- interceptos[[ toString (conjunto_teste$EMP[i])]] + val
    } else if (mod$args$effect == "time") {
      prv <- interceptos[[ toString (conjunto_teste$TRI[i])]] + val
    } else {
      prv <- interceptos[[ toString (conjunto_teste$EMP[i])]] + val
    }
    previsao <- c (previsao, prv)
  } #for i

  return (previsao)
} #previsao_plm_ef

previsao_plm_re_pols <- function (mod, conjunto_teste)
{
  coeficientes <- mod$coefficients
  previsao <- c()

  for (i in 1:nrow (conjunto_teste))
  {
    val = coeficientes [[1]]
    for (j in 2:length (coeficientes)){
      produto <- coeficientes [[j]]*conjunto_teste[[attr (coeficientes, "names", 1)[j]][i]
      val = val + produto
    }
    previsao <- c (previsao, val)
  } #for i

  return (previsao)
} #previsao_re_pols

```

```

validacao_cruzada <- function (dependente, formula, dados, efeito, modelo, teste_ids)
{
  #reprodutibilidade
  set.seed (9876)

  particoes <- caret::createFolds(dados$EMP, k = NR_FOLDS)

  kfold_plm <- data.frame ()

  for (t in teste_ids)
  {
    treina <- particoes_ids [-t]
    treinamento_ids <- NULL

    for (i in treina){
      part <- sprintf("Fold%02d", i)
      treinamento_ids <- c (treinamento_ids, particoes[i][[part]])
    } #i in treina

    treinamento <- dados [treinamento_ids, ]
    part <- sprintf("Fold%02d", t)
    teste <- dados [particoes[t][[part]], ]
    teste <- teste[complete.cases(teste), ]

    #painel
    if ( modelo == "within")
    {
      mod <- plm(formula=as.formula(formula), data=treinamento, index=c("EMP", "TRI"),
        na.action="na.exclude", model = modelo, effect = efeito)
      previsoes <- previsao_plm_ef (mod, teste)
    } else {
      mod <- plm(formula=as.formula(formula), data=treinamento, index=c("EMP", "TRI"),
        na.action="na.exclude", model = modelo)
      previsoes <- previsao_plm_re_pols (mod, teste)
    }

    dtr <- data.frame( variavel = dependente,
      regressao = "plm::plm",
      R2 = caret::R2(previsoes, teste[[dependente]]),
      RMSE = caret::RMSE(previsoes, teste[[dependente]]),
      MAE = caret::MAE(previsoes, teste[[dependente]]))

    kfold_plm <- rbind (kfold_plm, dtr)

  } #for t

  return (kfold_plm)

} #validacao_cruzada

#=====
# Regressão Painel PLM
#=====
sink ("validacao_cruzada_defesa.txt")
#=====
#ROA - FE individual
#=====

```

```

teste_ids <- particao_teste (particoes_ids)

dependente = "ROA"
formula <- "ROA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS"
rst <- validacao_cruzada (dependente, formula, dados, "", "pooling", teste_ids)

desvio_r2 <- round (sd(rst$R2), 4)
media_r2 <- round (mean (rst$R2), 4)
desvio_rmse = round (sd(rst$RMSE), 4)
media_rmse = round (mean (rst$RMSE), 4)
desvio_mae = round (sd(rst$MAE), 4)
media_mae = round (mean(rst$MAE), 4)

print (paste0("plm;", dependente, ";", "R2;", media_r2, ";", desvio_r2))
print (paste0("plm;", dependente, ";", "RMSE;", media_rmse, ";", desvio_rmse))
print (paste0("plm;", dependente, ";", "MAE;", media_mae, ";", desvio_mae))

#=====
#ROE - FE time
#=====

teste_ids <- particao_teste (particoes_ids)

dependente = "ROE"
formula = "ROE ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS"
rst <- validacao_cruzada (dependente, formula, dados, "", "pooling", teste_ids)

desvio_r2 <- round (sd(rst$R2), 4)
media_r2 <- round (mean (rst$R2), 4)
desvio_rmse = round (sd(rst$RMSE), 4)
media_rmse = round (mean (rst$RMSE), 4)
desvio_mae = round (sd(rst$MAE), 4)
media_mae = round (mean(rst$MAE), 4)

print (paste0 ("plm;", dependente, ";", "R2;", media_r2, ";", desvio_r2))
print (paste0 ("plm;", dependente, ";", "RMSE;", media_rmse, ";", desvio_rmse))
print (paste0 ("plm;", dependente, ";", "MAE;", media_mae, ";", desvio_mae))

#=====
#MVA - FE time
#=====

teste_ids <- particao_teste (particoes_ids)

dependente = "MVA"
formula = "MVA ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS"
rst <- validacao_cruzada (dependente, formula, dados, "time", "within", teste_ids)

desvio_r2 <- round (sd(rst$R2), 4)
media_r2 <- round (mean (rst$R2), 4)
desvio_rmse = round (sd(rst$RMSE), 4)
media_rmse = round (mean (rst$RMSE), 4)
desvio_mae = round (sd(rst$MAE), 4)
media_mae = round (mean(rst$MAE), 4)

print (paste0 ("plm;", dependente, ";", "R2;", media_r2, ";", desvio_r2))
print (paste0 ("plm;", dependente, ";", "RMSE;", media_rmse, ";", desvio_rmse))
print (paste0 ("plm;", dependente, ";", "MAE;", media_mae, ";", desvio_mae))

```

```

#=====
#QTB - FE time
#=====

teste_ids <- particao_teste (particoes_ids)

dependente = "QTB"
formula = "QTB ~ IDA + TAM + ALV + IND1 + IND2 + NM + N1 + N2 + IMR + FAV + VIS + FAVxVIS"

rst <- validacao_cruzada (dependente, formula, dados, "time", "within", teste_ids)
desvio_r2 <- round (sd(rst$R2), 4)
media_r2 <- round (mean (rst$R2), 4)
desvio_rmse = round (sd(rst$RMSE), 4)
media_rmse = round (mean (rst$RMSE), 4)
desvio_mae = round (sd(rst$MAE), 4)
media_mae = round (mean(rst$MAE), 4)

print (paste0 ("plm;", dependente, ";", "R2;", media_r2, ";", desvio_r2))
print (paste0 ("plm;", dependente, ";", "RMSE;", media_rmse, ";", desvio_rmse))
print (paste0 ("plm;", dependente, ";", "MAE;", media_mae, ";", desvio_mae))

sink()

#=====
# Regressão algoritmo ML GBM
#=====

varc <- c ("IDA", "ALV", "TAM", "IND1", "IND2", "NM", "N1", "N2", "IMR")
VISIB = "VIS"
var_ind2 <- c("FAVt_2", paste0(VISIB,"t_2"), paste0("FAVx", VISIB, "t_2"))
var_ind <- c("FAV", VISIB, paste0("FAVx", VISIB))

var_roa <- c ("ROA", varc, var_ind)
var_roe <- c ("ROE", varc, var_ind)
var_mva <- c ("MVA", varc, var_ind)
var_qtb <- c ("QTB", varc, var_ind)

coefics_gbm <- data.frame()
#Parametros Caret
data_ctrl <- trainControl(method = "cv", number = NR_FOLDS)

#Parametros GBM para otimizacao
grade_caret <- expand.grid (interaction.depth= c (5, 15, 25, 35), n.trees= (5:20)*200,
                           shrinkage= c (0.01), n.minobsinnode=c (3, 5, 7, 10))

resultados_xval <- function (modelo_caret, regressao)
{
  par_original <- par()

  par (las=2)
  par(mar= c(5, 10, 1, 2) + 0.2)
  summary (modelo_caret)
  par (par_original)

  desvio_r2 <- round (sd(modelo_caret$resample$Rsquared), 4)
  media_r2 <- round (mean (modelo_caret$resample$Rsquared), 4)
  desvio_rmse = round (sd(modelo_caret$resample$RMSE), 4)
  media_rmse = round (mean (modelo_caret$resample$RMSE), 4)
  desvio_mae = round (sd(modelo_caret$resample$MAE), 4)
}

```



```

media_mae = round (mean(modelo_caret$resample$MAE), 4)

print (paste0 (regressao, ";", dependente, ";R2;", media_r2, ";", desvio_r2))
print (paste0 (regressao, ";", dependente, ";RMSE;", media_rmse, ";", desvio_rmse))
print (paste0 (regressao, ";", dependente, ";MAE;", media_mae, ";", desvio_mae))
} #resultados_xval

coefs_gbm <- function (dependente, perda, grade, modelo_caret, coef)
{
  coeficientes_gbm <- data.frame()
  coef <- data.frame (dependente= c (""), variavel= c (""), influencia = 0.0, perda= c (""),
    arvores = 0, profundidade = 0, aprendizado= 0.0 , nos=0)

  coef$dependente = dependente
  coef$perda= perda
  coef$arvores = grade$n.trees
  coef$profundidade = grade$interaction.depth
  coef$aprendizado = grade$shrinkage
  coef$nos = grade$n.minobsinnode

  cgbm <- summary (modelo_caret)
  for (i in 1:nrow (cgbm))
  {
    coef$variavel = cgbm$var[i]
    coef$influencia = round ((cgbm$rel.inf[i] / 100.00) , 4)
    coeficientes_gbm <- rbind (coeficientes_gbm, coef)
  } #for i

  return (coeficientes_gbm)
} #coefs_gbm

#=====
# ROA
#=====

dependente = "ROA"
formula = "ROA ~ ."
dvar <- dados [ ,var_roa ]

#=====
# VALIDAÇÃO CRUZADA E AJUSTE DE HIPER PARÂMETROS
#=====

#reprodutibilidade
set.seed(9876)

inicioExecucao = Sys.time()
modelo_caret <- train( as.formula (formula),
  data= dvar,
  na.action= "na.exclude",
  distribution= list (name= "quantile", alpha = 0.5),
  trControl= data_ctrl,
  method= "gbm",
  metric= "RMSE",
  tuneGrid= grade_caret,
  verbose= FALSE
)
print (Sys.time() - inicioExecucao)

```

```

print (modelo_caret)
print (summary (modelo_caret))
resultados_xval( modelo_caret, "gbm-quantile")

ret <- coefs_gbm ("ROA", "quantile", grade_quant, modelo_caret)
coefics_gbm <- rbind (coefics_gbm, ret)

#=====
# ROE
#=====
dependente = "ROE"
formula = "ROE ~ ."
dvar <- dados [ , var_roe]
dvar <- dvar[complete.cases(dvar), ]

#=====
# VALIDAÇÃO CRUZADA E AJUSTE DE HIPER PARÂMETROS
#=====

#reprodutibilidade
set.seed(9876)

inicioExecucao = Sys.time()
modelo_caret <- train( as.formula (formula),
                      data= dvar,
                      na.action= "na.exclude",
                      distribution= list (name= "quantile", alpha = 0.5),
                      trControl= data_ctrl,
                      method= "gbm",
                      metric= "RMSE",
                      tuneGrid= grade_caret,
                      verbose= FALSE
)
print (Sys.time() - inicioExecucao)

print (modelo_caret)
print (summary (modelo_caret))
resultados_xval (modelo_caret, "gbm-quantile")
ret <- coefs_gbm ("ROE", "quantile", grade_quant, modelo_caret)
coefics_gbm <- rbind (coefics_gbm, ret)

#=====
# MVA
#=====

dependente = "MVA"
formula = "MVA ~ ."
dvar <- dados [ , var_mva ]

#=====
# VALIDAÇÃO CRUZADA E AJUSTE DE HIPER PARÂMETROS
#=====

#reprodutibilidade
set.seed(9876)

inicioExecucao = Sys.time()
modelo_caret <- train( as.formula ( formula ),
                      data= dvar,
                      na.action= "na.exclude",

```

```

distribution= list (name= "quantile", alpha = 0.5),
trControl= data_ctrl,
method= "gbm",
metric= "RMSE",
tuneGrid= grade_caret,
verbose= FALSE
)
print (Sys.time() - inicioExecucao)

print (modelo_caret)
print (summary (modelo_caret))
resultados_xval(modelo_caret, "gbm-quantile")
ret <- coefs_gbm ("MVA", "quantile", grade_quant, modelo_caret)
coefics_gbm <- rbind (coefics_gbm, ret)

#=====
# QTB
#=====

dependente = "QTB"
formula = "QTB ~ ."
dvar <- dados [ , var_qtb ]
str ( dvar )

#=====
# VALIDAÇÃO CRUZADA E AJUSTE DE HIPER PARÂMETROS
#=====

#reprodutibilidade
set.seed(9876)

inicioExecucao = Sys.time()
modelo_caret <- train( as.formula ( formula ),
data = dvar,
na.action = "na.exclude",
distribution = list (name= "quantile", alpha = 0.5),
trControl = data_ctrl,
method = "gbm",
metric = "RMSE",
tuneGrid = grade_caret,
verbose = FALSE
)
print (Sys.time() - inicioExecucao)
print (modelo_caret)
print (summary (modelo_caret))
resultados_xval( modelo_caret, "gbm-quantile")

ret <- coefs_gbm ( "QTB", "quantile", grade_quant, modelo_caret )
coefics_gbm <- rbind ( coefics_gbm, ret )

#=====

write.xlsx (coefics_gbm, file= "gbm_coef_caret_k_20_Defesa.xlsx",
sheetName="coeficientes", row.names=FALSE, showNA=FALSE)

```