

**UNIVERSIDADE METODISTA DE PIRACICABA  
FACULDADE DE GESTÃO E NEGÓCIOS  
DOUTORADO EM ADMINISTRAÇÃO**

**JOSÉ ERASMO SILVA**

**CRIAÇÃO DE PORTFÓLIOS DE INVESTIMENTO EM AÇÕES A PARTIR DE  
MODELOS HIERÁRQUICOS NÃO LINEARES DE REGRESSÃO**

**PIRACICABA**

**2020**

**JOSÉ ERASMO SILVA**

**CRIAÇÃO DE PORTFÓLIOS DE INVESTIMENTO EM AÇÕES A PARTIR DE  
MODELOS HIERÁRQUICOS NÃO LINEARES DE REGRESSÃO**

Tese apresentada ao Curso de Doutorado em Administração da Faculdade de Gestão e Negócios da Universidade Metodista de Piracicaba, como parte dos requisitos para obtenção do título de Doutor em Administração.

Campo de conhecimento: Finanças

Orientadora: Prof. Dra. Maria Imaculada de Lima Montebello

**PIRACICABA**

**2020**

Ficha Catalográfica elaborada pelo Sistema de Bibliotecas da UNIMEP

Bibliotecário: Fábio Henrique dos Santos Corrêa – CRB: 8/10150

S586c Silva, José Erasmo  
Criação de portfólios de investimento em ações a partir de modelos hierárquicos não lineares de regressão / José Erasmo Silva. – 2020.  
126 fls.; il.; 30 cm.

Orientador (a): Prof. Dra. Maria Imaculada de Lima Montebello.  
Tese (Doutorado) – Universidade Metodista de Piracicaba, Programa de Pós-Graduação em Administração, Piracicaba, 2020.

1. Carteiras de investimento. 2. Long and short. 3. Modelos hierárquicos. I. Montebello, Maria Imaculada de Lima. II. Título.

CDD – 332.6

**JOSÉ ERASMO SILVA**

**CRIAÇÃO DE PORTFÓLIOS DE INVESTIMENTO EM AÇÕES A PARTIR DE  
MODELOS HIERÁRQUICOS NÃO LINEARES DE REGRESSÃO**

Tese apresentada ao Curso de Doutorado em Administração da Faculdade de Gestão e Negócios da Universidade Metodista de Piracicaba, como parte dos requisitos para obtenção do título de Doutor em Administração.

Campo de conhecimento: Finanças

Orientadora: Prof. Dra. Maria Imaculada de Lima Montebello

Data de aprovação: \_\_\_/\_\_\_/\_\_\_

Banca Examinadora:

---

Profa. Dra. Maria Imaculada de Lima Montebello  
(orientadora)  
UNIMEP/FGN

---

Prof. Dr. Thel Augusto Monteiro  
UNIMEP/FGN

---

Profa. Dra. Maria José de Camargo Machado  
Centro Universitário Salesiano de São Paulo  
(UNISAL)

---

Prof. Dr. Luiz Paulo Lopes Fávero  
Faculdade de Economia, Administração e  
Contabilidade da Universidade de São Paulo  
(FEA/USP)

---

Profa. Dra. Fabíola Cristina Ribeiro de Oliveira  
Pontifícia Universidade Católica de Campinas  
(PUC-Campinas)

*À minha querida mãe, Nair Borges da Silva, e à  
memória do meu amado pai, José Carlos Silva.*

## AGRADECIMENTOS

À **Prof. Dra. Maria José de Camargo Machado**, pelas aulas, orientações, cafés, conversas, amizade e, principalmente, por não permitir que meu trabalho fosse insignificante para a academia e para a sociedade.

Ao **Prof. Dr. Clóvis Luís Padoveze** (*in memoriam*), pelas aulas, orientações, confiança, livros e discussões.

À **Prof. Dra. Maria Imaculada de Lima Montebello**, por ter me apresentado ao mundo da estatística, inicialmente assustador e posteriormente encantador. Sou-lhe muito grato ainda pelas aulas, pelas orientações e por aceitar o desafio de me acompanhar no final dessa jornada.

Ao **Prof. Dr. José Francisco Calil**, pelas aulas, orientações, ideias e discussões.

Ao **Prof. Dr. Thel Augusto Monteiro**, pelas aulas, *slides* e apresentações impecáveis, que me serviram de incentivo e me trouxeram motivação. Agradeço-lhe também o encorajamento, as oportunidades e por aceitar participar da banca de defesa desta tese.

Ao **Prof. Dr. Luiz Paulo Lopes Fávero**, pela disponibilidade e oportunidade de conhecê-lo, um excelente professor, um profissional competente, uma pessoa incrível. Agradeço-lhe ainda as discussões, as aulas, os cursos, os livros e por aceitar participar da banca de defesa desta tese.

À **Prof. Dra. Fabíola Cristina Ribeiro de Oliveira**, por aceitar participar da banca de defesa desta tese e por contribuir para meu crescimento pessoal, profissional e acadêmico.

À **Prof. Dra. Lumila Souza Girioli Camargo**, por todos os ensinamentos, pela parceria, amizade e *feedbacks* ao longo desse grande ciclo. Sou-lhe grato ainda por aceitar participar da banca de defesa desta tese.

À **Prof. Dra. Nadia Kassouf Pizzinatto**, por aceitar participar da banca de defesa desta tese.

Ao **Prof. Dr. Levi Morgan**, pela disponibilidade e parcerias durante o curso de doutorado. Sou-lhe muito grato ainda por aceitar participar da banca de defesa desta tese.

Ao **Prof. Dr. Hygino Canhadas Belli**, pela atenção e prontidão em ler e contribuir para o desenvolvimento desse trabalho. Agradeço-lhe ainda por aceitar participar da banca de defesa desta tese.

À **Profa. Mirian Polla**, pelas correções, ideias e disponibilidade.

À minha mãe, **Nair Borges da Silva**, grande influenciadora da minha vida acadêmica, pelos incentivos e exemplo de fé, confiança e força de vontade.

À minha esposa, amiga e companheira de todas as horas, **Marina Franco**, pelo amor, incentivo e compreensão.

Às minhas amadas filhas, **Natalie**, **Bianca** e **Lara**, por, involuntariamente, todos os dias encherem minha vida de amor e de significado.

Ao meu irmão, parceiro de muitos desafios e aventuras, **Cadu Borges**.

Ao meu grande amigo e irmão, **Fábio Henrique de Paula**, pela parceria de uma vida inteira.

À **Evelin Yara Elias**, minha querida amiga e terapeuta, pelo acompanhamento, ideias, incentivos e boas conversas ao longo dessa jornada.

Ao companheiro de todas as horas no período do curso de doutorado, **Prof. Jorge Luiz dos Santos Silva**.

Aos **professores do Programa de Pós-Graduação em Administração da Unimep**, pelas discussões, contribuições e ensinamentos.

Aos **colaboradores da Unimep**, que sempre nos auxiliaram direta e indiretamente.

A **toda a minha família** e aos **meus amigos**, por contribuírem para meu crescimento pessoal e espiritual.

E acima de tudo, a **Deus** e a **Nossa Senhora**.

*“Goodness is the only investment that never fails.”*

(Henry David Thoreau)



## RESUMO

O Brasil vive atualmente o menor patamar de taxa de juro da história, fixada em 2,00% em agosto de 2020. Isso tem feito com que aplicações de renda fixa se tornem menos interessantes, e os investidores passem a buscar alternativas mais rentáveis. Essa busca por investimentos mais rentáveis refletiu diretamente no mercado de ações brasileiro, que viu seu número de investidores crescer bruscamente nos últimos quatro anos. O objetivo principal da pesquisa foi desenvolver um modelo de seleção de ações para compor portfólios *long and short* utilizando modelos hierárquicos não lineares de regressão para dados em painel. Por meio desse modelo, foi possível analisar o comportamento histórico de diversas variáveis que influenciam os retornos das ações e, com isso, prever a probabilidade de retornos futuros positivos ou negativos para formar as carteiras. Foram coletadas informações trimestrais das empresas listadas nas bolsas de valores NYSE, NASDAQ, NYSE EURONEXT e B3 referentes ao período de 1995 a 2T2019. A partir das informações e dos modelos estatísticos, formou-se diferentes carteiras *long and short* compreendendo diferentes cenários, o que totalizou nove carteiras. Após a apuração dos resultados, constatou-se que os melhores deles foram gerados pelas carteiras com a inserção de *stop loss*, porém sem o título de renda fixa. Esse comportamento pode ser visualizado nas carteiras mistas (Estados Unidos e Brasil), nas carteiras mistas com restrição de data e nas carteiras somente com ações listadas na B3. Notou-se que a inserção do *stop loss* promoveu grandes ganhos nos resultados das carteiras. Não se pode dizer o mesmo para a inserção dos títulos de renda fixa, que trouxeram discreta diminuição no risco. A principal contribuição do trabalho foi desenvolver um novo modelo de investimento diante de um mercado de ações em forte ascensão, o que viabiliza a seleção de ações a partir de modelos hierárquicos não lineares, complementado com técnicas de alocação, de diversificação e de minimização de risco. O estudo mostra originalidade ao utilizar modelos hierárquicos não lineares de regressão para seleção de ações.

**Palavras-chave:** Carteiras de investimento. *Long and short*. Modelos hierárquicos.

## ABSTRACT

Brazil is currently experiencing the lowest interest rate in history, set at 2.00% in August 2020. This has made fixed income investments less interesting, and investors are looking for more profitable alternatives. This search for more profitable investments was directly reflected in the Brazilian stock market, which has seen its number of investors grow sharply in the past four years. The main objective of the research was to develop a stock selection model to compose long and short portfolios using nonlinear hierarchical regression models for panel data. Through this model, it was possible to analyze the historical behavior of several variables that influence stock returns and, with this, predict the probability of future positive or negative returns to form the portfolios. Quarterly information was collected from companies listed on the NYSE, NASDAQ, NYSE EURONEXT and B3 stock exchanges for the period from 1995 to 2Q2019. From the information and statistical models, different long and short portfolios were formed, comprising different scenarios, which totaled nine portfolios. After determining the results, it was found that the best of them were generated by the portfolios with the insertion of stop loss, but without the fixed income security. This behavior can be seen in mixed portfolios (United States and Brazil), in mixed portfolios with restricted dates and in portfolios with only shares listed in B3. It was noted that the insertion of the stop loss promoted great gains in the results of the portfolios. The same cannot be said for the insertion of fixed income securities, which brought a slight decrease in risk. The main contribution of the work was to develop a new investment model in the face of a rapidly growing stock market, which makes it possible to select stocks based on non-linear hierarchical models, complemented by allocation, diversification and risk minimization techniques. The study shows originality when using non-linear hierarchical regression models for stock selection.

**Keywords:** Investment portfolios. Long and short. Hierarchical models.

## LISTA DE FIGURAS

**Figura 1 – Representação dos níveis dos aninhamentos utilizados no trabalho ... 45**

## LISTA DE GRÁFICOS

<b>Gráfico 1 – Curvas ROC modelos <i>logit</i> (azul) versus modelo hierárquico (vermelho) <i>in sample</i> .....</b>	<b>67</b>
<b>Gráfico 2 – Curvas ROC modelos <i>logit</i> (azul) versus modelo hierárquico (vermelho) <i>out of sample</i> .....</b>	<b>68</b>
<b>Gráfico 3 – Curvas ROC modelos <i>logit</i> (azul) versus modelo hierárquico (vermelho) completo .....</b>	<b>68</b>
<b>Gráfico 4 – Retorno acumulado trimestre a trimestre da carteira LS com <i>stop s/índice1</i> e demais indicadores .....</b>	<b>84</b>
<b>Gráfico 5 – Datas com maior concentração de publicações em 2018 .....</b>	<b>93</b>
<b>Gráfico 6 – Retorno acumulado trimestre a trimestre das carteiras LS <i>s/índice1</i>, LS <i>s/índice2</i> e demais indicadores .....</b>	<b>94</b>
<b>Gráfico 7 – Retorno acumulado trimestre a trimestre das carteiras LS <i>s/índice1</i>, LS <i>s/índice2</i>, LS <i>s/índice3</i> e demais indicadores .....</b>	<b>105</b>

## LISTA DE QUADROS

<b>Quadro 1 – Trabalhos de autores que fundamentaram o desenvolvimento da tese .....</b>	<b>39</b>
<b>Quadro 2 – Estudos sobre construção de portfólios ou previsão de retornos de ações .....</b>	<b>50</b>
<b>Quadro 3 – Grupos de carteiras .....</b>	<b>55</b>

## LISTA DE TABELAS

Tabela 1 – Correlação entre os valores das principais variáveis independentes do trabalho referente ao período de 2008-2T2019 .....	52
Tabela 2 – Estatística descritiva para as variáveis da amostra .....	58
Tabela 3 – Estimação de modelo hierárquico não linear – modelo nulo – período 1995 a 2007 .....	60
Tabela 4 – Correlação intraclasse referente ao período de 1995 a 2007 .....	61
Tabela 5 – Estimação de modelo hierárquico não linear – modelo nulo – período 1995 a 2019 .....	61
Tabela 6 – Correlação intraclasse referente ao período de 1995 a 2019.....	62
Tabela 7 – Estimação de modelo hierárquico não linear – modelo condicional – período 1995 a 2007 .....	63
Tabela 8 – Correlação intraclasse para o modelo condicional de 1995 e 2007 .....	65
Tabela 9 – Curvas ROC do modelo <i>logit versus</i> modelo hierárquico nos diferentes cenários da amostra .....	66
Tabela 10 – Tabela com os níveis de acerto <i>out of sample</i> para <i>cutoffs</i> 60%, 70%, 80% e 90%: seleção de ações para a carteira <i>long</i> ...	69
Tabela 11 – Tabela com os níveis de acerto para <i>cutoffs</i> 40%, 30%, 20% e 10%: seleção de ações para as carteiras <i>short</i> .....	70
Tabela 12 – Principais indicadores dos mercados brasileiro e americano ...	71
Tabela 13 – Exemplos da apuração dos retornos referentes aos índices utilizados .....	74
Tabela 14 – Carteira formada somente com ETFs com pesos iguais para cada ETF na carteira .....	74
Tabela 15 – Correlação entre as taxas de retorno trimestrais dos indicadores .....	76
Tabela 16 – Lado <i>long</i> da carteira .....	77
Tabela 17 – Lado <i>short</i> da carteira .....	78
Tabela 18 – Resultados consolidados da carteira <i>long and short</i> .....	80

<b>Tabela 19 – Resultados trimestrais acumulados para os indicadores e para a carteira LS .....</b>	<b>82</b>
<b>Tabela 20 – Two-sample Wilcoxon rank-sum (Mann-Whitney) test: carteira LS1 versus indicadores.....</b>	<b>85</b>
<b>Tabela 21 – Classificação setorial NAICS com os vinte setores mais representativos na carteira sem restrição de data .....</b>	<b>87</b>
<b>Tabela 22 – Quantidade de ações por bolsa .....</b>	<b>87</b>
<b>Tabela 23 – Quantidade de ações por bolsa com restrição de data de publicação .....</b>	<b>88</b>
<b>Tabela 24 – Lado <i>long</i> da carteira <i>long and short</i> com restrição de data de publicação .....</b>	<b>89</b>
<b>Tabela 25 – Lado <i>short</i> da carteira <i>long and short</i> com restrição de data de publicação .....</b>	<b>90</b>
<b>Tabela 26 – Resultados das carteiras <i>long and short</i> com restrição nas datas de publicação para alocação .....</b>	<b>91</b>
<b>Tabela 27 – Resultados trimestre a trimestre acumulados para as carteiras LS1 e LS2 .....</b>	<b>93</b>
<b>Tabela 28 – Two-sample Wilcoxon rank-sum (Mann-Whitney) test: carteiras LS1 e LS2 versus indicadores .....</b>	<b>95</b>
<b>Tabela 29 – Lado <i>short</i> da carteira <i>long and short</i> com ações listadas na B3 .....</b>	<b>97</b>
<b>Tabela 30 – Lado <i>short</i> da carteira <i>long and short</i> somente com ações listadas na B3 .....</b>	<b>98</b>
<b>Tabela 31 – Resultados da carteira <i>long and short</i> somente com ações listadas na B3 .....</b>	<b>99</b>
<b>Tabela 32 – Resultados da carteira <i>long and short</i> somente com ações listadas na B3 .....</b>	<b>102</b>
<b>Tabela 33 – Resultados trimestre a trimestre acumulados para as carteiras LS1, LS2 e LS3 .....</b>	<b>104</b>
<b>Tabela 34 – Two-sample Wilcoxon rank-sum (Mann-Whitney) test: carteiras LS1, LS2 e LS3 versus indicadores .....</b>	<b>105</b>
<b>Tabela 35 – Classificação setorial NAICS com os vinte setores mais representativos na carteira .....</b>	<b>108</b>

<b>Tabela 36 – Datas com maior quantidade de publicações em 2018 .....</b>	<b>110</b>
<b>Tabela 37 – Comparativo dos desvios padrão, taxas de retornos, médias e índice de Sharpe dos indicadores utilizados sem <i>stop loss</i> e com <i>stop loss</i> .....</b>	<b>111</b>
<b>Tabela 38 – Totalização das carteiras considerando custos de transações e inflação .....</b>	<b>112</b>
<b>Tabela 39 – Carteira LS3 em reais considerando custos de transações e inflação .....</b>	<b>113</b>
<b>Tabela 40 – Consolidado de todas as carteiras com as diferenças de médias .....</b>	<b>114</b>



## LISTA DE SIGLAS E ABREVIATURAS

AMEX	<i>American Exchange</i>
ARIMA	<i>Autoregressive Integrated Moving Average</i>
BP	<i>Back Propagation</i>
CDB	Certificado de Depósito Bancário
CDI	Certificado de Depósito Interbancário
CPI	<i>Consumer Price Index</i>
DY	<i>Dividend Yield</i>
ETF	<i>Exchange Traded Fund</i>
EUA	Estados Unidos da América
EGARCH	<i>Exponential Generalized Autoregressive Conditional Heteroscedasticity</i>
GARCH	<i>Generalized Autoregressive Conditional Heteroscedasticity</i>
HGLM	<i>Hierarchical Generalized Linear Models</i>
HLM	<i>Hierarchical Linear Models</i>
HME	Hipótese de Mercado Eficiente
HMM	<i>Hidden Markov Model</i>
HNM	Modelo Hierárquico não Linear
IA	<i>Artificial Intelligence</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
IDH	Índice de Desenvolvimento Humano
IPCA	Índice de Preços para o Consumidor Amplo
GJR-GARCH	<i>Glosten-Jagannathan-Runkle-Generalized Autoregressive Conditional Heteroscedasticity</i>
LFT	Letras Financeiras do Tesouro
LL	Lucro Líquido
LN	Logaritmo Natural ou Logaritmo Neperiano
LS	<i>Long and Short</i>
LSTM-DNN	<i>Long Short-Term Memory-Deep Neural Network</i>
MAE	<i>Mean Absolute Error</i>
MOP	<i>Momentum of Predictability</i>
MQO	Mínimos Quadrados Ordinários
NAICS	<i>North American Industry Classification System</i>
NASDAQ	<i>National Association of Securities Dealers Automated Quotations</i>

NYSE	<i>New York Stock Exchange</i>
OPEP	Organização dos Países Exportadores de Petróleo
P/L	Preço sobre Lucro
PIB	Produto Interno Bruto
RMSE	<i>Root Mean Square Error</i>
RNA	Redes Neurais Artificiais
ROA	<i>Return on Asset</i>
ROC	<i>Receiver Operating Characteristic</i>
ROE	<i>Return on Equity</i>
S&P	<i>Standard &amp; Poor's</i>
SVM	<i>Support Vector Machine</i>
T-BONDS	<i>Treasury Bonds</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	18
<b>1.1</b>	<b>Contexto</b> .....	18
<b>1.2</b>	<b>Problema de pesquisa</b> .....	23
<b>1.3</b>	<b>Objetivos</b> .....	24
<b>1.4</b>	<b>Tese proposta e hipóteses</b> .....	24
<b>1.5</b>	<b>Justificativas</b> .....	25
<b>1.6</b>	<b>Estrutura da tese</b> .....	27
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	28
<b>2.1</b>	<b>Estratégia <i>long and short</i></b> .....	30
<b>2.2</b>	<b>Estudos sobre carteiras de investimento</b> .....	33
<b>3</b>	<b>PROCEDIMENTOS METODOLÓGICOS</b> .....	41
<b>3.1</b>	<b>Amostra</b> .....	41
<b>3.2</b>	<b>Modelos hierárquicos</b> .....	42
<b>3.3</b>	<b>Procedimentos estatísticos</b> .....	44
<b>3.3.1</b>	<b>Nível 1 – Modelo de amostragem</b> .....	45
<b>3.3.2</b>	<b>Nível 1 – Função <i>link</i></b> .....	46
<b>3.3.3</b>	<b>Nível 1 – Modelo estrutural</b> .....	46
<b>3.3.4</b>	<b>Nível 1 – Modelo de amostragem</b> .....	47
<b>3.3.5</b>	<b>Nível 1 – Função <i>link</i></b> .....	47
<b>3.3.6</b>	<b>Nível 1 – Modelo estrutural</b> .....	48
<b>3.4</b>	<b>Variáveis</b> .....	49
<b>4</b>	<b>RESULTADOS</b> .....	58
<b>5</b>	<b>CONCLUSÕES, LIMITAÇÕES E SUGESTÕES</b> .....	117
	<b>REFERÊNCIAS</b> .....	120

# 1 INTRODUÇÃO

O capítulo introdutório apresenta o contexto do tema, o problema de pesquisa, os objetivos, a tese e as hipóteses, as justificativas e a estrutura da pesquisa.

## 1.1 Contexto

O Brasil vive atualmente o menor patamar de taxa de juro da história, fixada em 2,00% em agosto de 2020, segundo informações do Banco Central do Brasil (2020). Isso tem feito com que aplicações de renda fixa, conhecidas em virtude de seu alto retorno e baixo risco no mercado brasileiro, se tornem menos interessantes cada vez mais. Sendo assim, os investidores têm buscado alternativas mais arriscadas para poder manter o mesmo patamar de retorno que anteriormente era obtido sem muito esforço. Nesse contexto, o mercado de ações é uma das alternativas que possibilitam maior retorno, mas também maior risco.

O mercado de capitais brasileiro tem apresentado um número crescente de investidores, que vem se acentuando após as frequentes baixas na taxa de juro. Segundo a Brasil, Bolsa, Balcão (B3), considerando pessoas físicas, o número de investidores aumentou de 85.249, em 2002, para 813.291, em 2018, o que representa um aumento de 854% no período. No entanto, em uma atualização divulgada pela B3, em julho de 2020, ela registrava 2.648.975 investidores, ou seja, o número mais que triplicou em menos de dois anos (BRASIL, BOLSA, BALCÃO, 2020). Ainda é um número pequeno de investidores, quando comparado com a população brasileira, estimada em 212 milhões de habitantes, segundo projeção do Instituto Brasileiro de Geografia e Estatística (IBGE) (2020) para setembro de 2020. O número fica ainda mais destoado quando comparado com os mercados desenvolvidos, por exemplo, o mercado americano, com população estimada em 330 milhões de habitantes para setembro de 2020 (UNITED STATE CENSUS BUREAU, 2019), em que mais de 50% dessa população tem investimentos diretos ou indiretos em ações (UNITED NATIONS POPULATION FUND, 2019).

É notório que o investimento em ações apresenta alta volatilidade e riscos em curto prazo, mas há um consenso de que, em longo prazo, os retornos são mais elevados que a maioria das opções no mercado, principalmente quando o investimento é complementado com outros tipos de ativos como forma de proteção.

Com o crescimento do número de investidores, aumenta também a busca por modelos de investimento, tendo em vista que a concentração de uma grande quantidade de investidores em uma mesma estratégia tende a anular os ganhos. Para Damodaran (2012), uma vez que as estratégias de investimento se tornam públicas, quer diretamente, mediante livros e artigos, quer indiretamente, por meio de gerentes de carteiras que delas se utilizam, é de se esperar que se tornem menos eficazes. Nesse sentido, quando muitos investidores compartilham das mesmas estratégias de investimento, elas deixam de trazer resultados positivos, o que obriga investidores, pesquisadores e gerentes de carteiras e de fundos de maneira geral a estarem sempre buscando inovações nos modelos de investimento.

Em virtude do rápido crescimento da economia financeira e da tecnologia da informação, torna-se cada vez mais importante o uso de modernas técnicas de otimização para determinar portfólios confiáveis e robustos. Esse progresso tem causado impacto significativo no mercado financeiro global, e diferentes abordagens têm sido propostas para a criação de portfólios (CHEN; PENG; LISSER, 2020)

Demiguel, Garlappi e Uppal (2009) apontam que ao menos 14 estratégias de investimento foram estudadas de forma consistente após o artigo seminal de Markowitz (1952). No entanto, segundo Demiguel, Garlappi e Uppal (2009), esses modelos apresentam resultados discretos quando seus retornos são ajustados aos riscos. Os autores argumentam ainda que, apesar de os modelos serem sofisticados, são de difícil aplicação e ficam restritos a uma pequena parcela de investidores. Dessa forma, a grande maioria destes ainda utilizam modelos simplificados de investimento. No mesmo sentido, Black e Litterman (2006) alegam que o problema de muitos modelos quantitativos de alocação de ativos é que eles são difíceis de ser aplicados e tendem a apresentar resultados ruins.

Com o surgimento de computadores cada vez mais sofisticados, a possibilidade de expansão dos modelos de previsão é também incrementada, e isso possibilita inovações constantes nas estratégias de formação de carteiras de investimentos.

Hou *et al.* (2020) testaram diferentes modelos de seleção de ações utilizando séries temporais. Os autores utilizaram métodos clássicos de *machine learning*, abordagens com *deep learning* e modelos híbridos. Desse modo, por meio de um modelo híbrido, foi possível obter um retorno acumulado de 3.556% no período de 1989 a 2018. Ratto *et al.* (2018) testaram ferramentas para seleção de ações servindo-se de um modelo preditivo híbrido de análise técnica com *machine learning*. Mediante o sistema desenvolvido esses autores puderam prever a direção do preço das ações em 61,69% dos casos. Seguindo a mesma linha dos trabalhos anteriores e empregando uma abordagem baseada em *machine*

*learning*, Li e Song (2019) utilizaram o sistema de Redes Neurais Artificiais (RNA), uma poderosa ferramenta para previsão e modelagem não linear. Esses autores argumentam que por intermédio dela é possível prever eficientemente o retorno de ações desde que se utilize um modelo matemático apropriado. O modelo foi testado em três empresas, e o nível médio de assertividade apresentado foi de 79,6% para a Sony Corporation, 77,6% para a Panasonic Corporation e 70,6% para a Toyota Corporation. Ou seja, em mais de 70% dos casos, o modelo foi preciso ao prever se uma ação teria alta ou baixa em determinado período.

Wang *et al.* (2018) desenvolveram um modelo de previsão que se alterna entre dois submodelos no decorrer do tempo, de acordo com o comportamento preditivo de cada um deles. Os autores utilizaram, como amostra, informações mensais das empresas listadas no índice *Standard & Poor's* (S&P 500) de janeiro de 1927 a dezembro de 2015. As previsões são feitas baseadas em regressões univariadas valendo-se de informações econômicas e financeiras da empresa e do mercado, bem como a média de retorno de um *benchmark*. Entre as variáveis que trouxeram melhores retornos excedentes estão o *dividend yield*, *dividend payout* e lucro sobre preço, com retorno anualizado excedente de, respectivamente, 1,56%, 2,07% e 2,2%.

Lewin e Campani (2020) empregaram, na alocação de carteiras, um modelo introduzido por Hamilton (1989) chamado cadeias de Markov, o qual possibilita que as probabilidades de mudanças entre os estados da economia sejam capturadas. A alocação que não permitiu vendas a descoberto (*long only*) gerou um prêmio de risco anualizado de 4,66%, com volatilidade de 7,62% e índice de Sharpe de 0,61%. A alocação com vendas a descoberto (*long and short*) apresentou resultado anualizado significativamente maior, sendo o prêmio de 36,90%, volatilidade de 33,33% e índice de Sharpe de 1,11%. Apesar do alto prêmio de risco anual da carteira, nota-se também grande volatilidade.

Shchankina e Ping (2020) utilizaram dois métodos para construção de carteiras, com o objetivo de apurar qual deles geraria melhores resultados na Bolsa de Valores de Shanghai e na Bolsa de Valores de Hong Kong. O primeiro é o método clássico de otimização de carteiras criado por Markowitz (1952), o qual vem sendo estudado e empregado por pesquisadores e investidores há mais de 60 anos. No segundo método, foi utilizado um modelo de séries temporais por meio do modelo *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH). Os resultados do estudo confirmaram que as carteiras construídas a partir desse modelo tem menor desvio padrão e melhores retornos do que as construídas a partir do modelo Markowitz.

A despeito da complexidade de alguns modelos apresentados anteriormente, por exemplo, Markov e Redes Neurais Artificiais, modelos relativamente mais simples e transparentes continuam gerando bons resultados. A acurácia na previsão do retorno das ações e a elucidação clara dos fatores causais subjacentes são de interesse de várias partes, por exemplo, investidores e governo. Entretanto, essa é uma questão ainda notoriamente desafiadora devido à natureza complexa, dinâmica e caótica do mercado de ações (HOU *et al.*, 2020).

Um modelo estatístico que vem sendo testado e tem evoluído desde a década de 1960 é o de regressão logística, o qual é utilizado quando se deseja explicar determinado evento dicotômico – sim ou não, lucro ou prejuízo, morrer ou viver, aceitar ou rejeitar, entre outros – por meio de uma ou mais variáveis explicativas (CORRAR; PAULO; DIAS FILHO, 2017). Com o advento dos modelos longitudinais de regressão para dados em painel, tanto os modelos lineares, por exemplo, regressão linear simples ou múltipla, quanto os não lineares, por exemplo, regressão logística binária e multinomial, podem contribuir para a análise de informações em conjunto de diferentes indivíduos disponíveis ao longo do tempo. De acordo com Fávero e Belfiore (2017), a principal vantagem da utilização de modelos longitudinais de regressão consiste em possibilitar o estudo das diferenças entre indivíduos em determinado fenômeno em cada *cross-section*, além de permitir a análise da evolução temporal desse mesmo fenômeno para cada indivíduo.

Os modelos de regressão logística, apesar de bastante úteis e de fácil aplicação, ainda são pouco utilizados em muitas Áreas do Conhecimento. Embora o desenvolvimento de *softwares* e o incremento da capacidade de processamento dos computadores tenham propiciado sua aplicação de forma mais direta, muitos pesquisadores ainda desconhecem suas utilidades e, sobretudo, as condições para que seu uso seja correto (FÁVERO; BELFIORE, 2017). Os autores destacam ainda que, quando esses constructos consideram estruturas aninhadas de dados, eles recebem o nome de modelos hierárquicos de regressão para dados em painel.

Nas ciências sociais, as estruturas dos dados são frequentemente hierárquicas no seguinte sentido: variáveis descrevem indivíduos que estão agrupados em bairros, os quais estão agrupados em cidades, que estão agrupadas em estados. Pode-se ainda citar estudantes, que estão agrupados em classes, as quais estão agrupadas em escolas, que estão agrupadas em bairros ou cidades. Nesse sentido, pode-se mencionar ainda as empresas que estão agrupadas em segmentos, que estão agrupados em subsetores, que estão agrupados em setores (RAUDENBUSH; BRYK, 2002).

Diferentemente dos modelos tradicionais de regressão, que trabalham com todas as variáveis em um único nível, os modelos hierárquicos permitem que as características de cada nível sejam trabalhadas dentro do seu respectivo nível. Por exemplo: as características de alunos são tratadas dentro do nível de alunos – idade, gênero, raça, entre outras –, assim como as características de empresas – Lucro Líquido (LL), *Return On Equity* (ROE), *Return On Asset* (ROA), etc. – e de seus setores – índices setoriais, segmentos e outros – são tratadas, respectivamente, ao nível de empresas e de setores. Pode-se citar ainda que, na utilização de diversificação internacional, que tem sido explorada cada vez mais visando se beneficiar dos pontos fortes de diferentes países, as empresas podem ser agrupadas ao nível de países. Trabalha-se, assim, as características de empresas – ROE, ROA, etc. – ao nível de empresas e as características de países – Taxa de juro, Índice de Desenvolvimento Humano (IDH), Produto Interno Bruto (PIB) etc.) – ao nível de países.

Esse tratamento diferenciado faz com que os modelos hierárquicos sejam mais precisos para fins de previsão. Nesse sentido, enxerga-se nesses modelos uma oportunidade para servir de base na seleção de ações para compor portfólios de investimento. Tendo em vista a crescente evolução e acessibilidade de novos modelos estatísticos, bem como a crescente demanda por novos modelos de investimento, entende-se que antigas estratégias devem ser revistas e evoluídas, assim como outras devem ser criadas. Além de contribuir para maior estabilidade do mercado, a diversidade de estratégias contribui também positivamente para o mercado em momentos de crise e de recessão. Afinal, quando muitos investidores se agrupam em torno de uma mesma estratégia, os retornos até determinado ponto podem ser conjuntamente positivos, mas, em momentos de crise, esse comportamento pode ser devastador (AVRAMOV *et al.*, 2017).

Taleb (2005) destaca que as estratégias de investimento estão em constante mutação e adaptação. Exemplificando de forma simplista, se o preço das ações sempre aumenta às segundas-feiras, investidores racionais comprarão ações no pregão anterior. Com o passar do tempo, a dinâmica do mercado será alterada, e esse efeito eliminado.

Para obter retornos consistentes mesmo em tempos de crise ou de recessão, é importante que, no contexto da estratégia de investimentos, haja ainda uma estratégia de proteção. Uma das estratégias de investimento que une retorno e proteção em um só cenário é conhecida como *long and short*, com a qual o investidor minimiza o risco de mercado e obtém retornos em ambos os sentidos do mercado, seja de queda (*bear markets*), seja de alta (*bull markets*) (HARVEY *et al.*, 2019).



As estratégias de investimento *long and short* vêm crescendo e se popularizando, uma vez que permitem ao investidor trabalhar com um número grande de ativos de forma mais eficiente que os modelos *long only* – portfólios nos quais são alocadas somente ações compradas (GRINOLD; KAHN, 2000). Permitem ainda minimizar os custos de investimento pois o lado *short* da estratégia, que é a venda a descoberto, é utilizada como recurso para compra do lado *long* da estratégia (BEAVER; MCNICHOLS; PRICE, 2016).

O *long and short* combina a venda a descoberto de um lado, de uma ação que se encontra supervalorizada, com a compra do outro lado, de uma ação que está subvalorizada. Para que se possa vender a descoberto, é necessário que se dê garantias, as quais variam de um país para outro e podem ser desde saldo em conta bancária até investimento em Certificado de Depósito Bancário (CDB) de liquidez diária (ZAREMBA *et al.*, 2019). Destaca-se que as vendas a descoberto podem apresentar restrições para alguns tipos de investidores, entre eles, gestores de fundos mútuos ou de fundos de pensão, que são explicitamente proibidos de vender a descoberto (BELLALAH *et al.*, 2020).

## 1.2 Problema de pesquisa

Conforme exposto na seção anterior, nota-se um crescente interesse pelo investimento em ações. No caso do Brasil, percebe-se que esse interesse é devido ao número de investidores registrados na bolsa de valores nos últimos quatro anos. Ressalta-se que esse movimento acontece principalmente pela queda na taxa base de juros, a qual faz com que investimentos de renda fixa tragam menos retorno, e isso faz com que os investidores que querem manter seus investimentos rentáveis passem a procurar novas alternativas. Nesse contexto, novas estratégias são demandadas em virtude de que há uma tendência de velhas estratégias deixarem de ser interessantes à medida que aumenta o número de investidores.

Com vistas a atender à demanda de novas estratégias e destacar que os modelos hierárquicos ainda são pouco explorados no Brasil, em especial na área de investimentos, observou-se que há uma lacuna que pode ser investigada fornecendo aos investidores de maneira geral um novo modelo para seleção de ações para compor carteiras de investimento. Enfatiza-se que, a princípio, não foram encontrados trabalhos de criação de carteiras com as características utilizadas na presente tese. Desse modo, buscou-se responder à seguinte pergunta que norteou esta pesquisa: **quais elementos são necessários para desenvolver um modelo de seleção de ações que componha portfólios *long and short* utilizando modelos hierárquicos não lineares de regressão para dados em painel?**

### 1.3 Objetivos

O objetivo principal do presente trabalho foi desenvolver um modelo de seleção de ações para compor portfólios *long and short* utilizando modelos hierárquicos não lineares de regressão para dados em painel.

Especificamente, esta tese teve por objetivos:

- a) Analisar o comportamento histórico das variáveis que influenciam no retorno das ações.
- b) Prever a probabilidade de retornos futuros positivos ou negativos para formação das carteiras *long and short*.
- c) Comparar os resultados alcançados aos principais indicadores do mercado brasileiro e americano.
- d) Deflacionar os retornos para que seja possível observar os retornos reais.

### 1.4 Tese proposta e hipóteses

Dados o crescente interesse pelas estratégias *long and short* e os retornos consistentes encontrados nos trabalhos citados na contextualização desta tese, fatores em conjunto com o aumento do número de investidores interessados no mercado acionário, resolveu-se explorar esse cenário em forma de tese por meio da criação de um modelo dinâmico de seleção de ações para compor portfólios utilizando modelos hierárquicos de regressão para dados em painel.

Posto isso, formulou-se as seguintes hipóteses:

- H1: A construção de um portfólio utilizando o modelo proposto vai simular retornos maiores e com menor risco que o principal índice de ações do mercado brasileiro, o Ibovespa.
- H2: A construção de um portfólio utilizando o modelo proposto vai simular retornos maiores e com menor risco que o principal índice de ações do mercado americano, o S&P 500.
- H3: A construção de um portfólio utilizando o modelo proposto vai simular retornos maiores que os principais investimentos de baixo risco, como CDBs e títulos do tesouro.

## 1.5 Justificativas

Acredita-se que o modelo de seleção de ações proposto na presente tese trará retornos consistentes e seguros, pois combinará de forma inédita quatro fatores na composição de portfólios: (1) robustez e poder preditivo dos modelos hierárquicos de regressão para dados em painel; (2) efeito autoimune das carteiras *long and short*; (3) uma parcela de alocação em títulos de alta liquidez e baixa volatilidade (garantia compulsória); (4) diversificação internacional. Nesse sentido, esta tese contribuirá para o meio acadêmico ao fornecer um novo modelo de investimento. Ainda nesse contexto, contribuirá também para o trabalho de investidores, analistas e companhias de maneira geral.

Os modelos hierárquicos de regressão para dados em painel têm adquirido importância considerável em diversas áreas do conhecimento, principalmente decorrente da determinação de constructos de pesquisa que consideram a existência de estruturas aninhadas de dados. Tais estruturas podem ser, por exemplo, um grupo de companhias que apresentem dados que variam ao longo do tempo ao nível individual e outros dados, entre eles, os setoriais, que serão comuns para determinados grupos (FÁVERO; BELFIORE, 2017). Esses modelos permitem que determinadas características possam ser isoladas de forma que elas afetem somente as empresas que compartilham dessas mesmas características. Conforme destaca Furlanetti (2018), os modelos hierárquicos de regressão são pouco explorados no Brasil e se desconhece sua utilização no contexto da formação de portfólios.

Os modelos *long and short*, quando bem elaborados, promovem uma maneira segura, barata e rentável de investir no mercado de ações, sem a necessidade de utilizar ferramentas tradicionais de proteção que podem corroer uma parcela considerável dos retornos, como as vendas de opções de longo prazo (PUTS) (HARVEY *et al.*, 2019).

Guerard, Xu e Wang (2019) chamam o processo de buscar dados, formar carteiras e testar hipóteses de mineração de dados (*data mining*). Os autores destacam que esse é um processo inevitável no contexto da construção de portfólios, pois ninguém pode testar somente uma hipótese e investir dinheiro nela. Nesse sentido, ele deve ser constantemente aperfeiçoado de acordo com novas realidades do mercado, comportamento de novos investidores, novas empresas, novos governos e até mesmo novas leis. Destarte, mais uma vez se destaca a importância da busca por novos modelos de investimento.

Mesmo que discretamente, pode-se dizer ainda que esta tese dará uma parcela de contribuição para a estabilidade do mercado financeiro, pelo fato de fornecer-lhe um novo modelo para investimento em ações e, desse modo, fazer com que haja mais alternativas para

diversificar as formas de investir. Conforme apontado por Avramov *et al.* (2017), Damodaran (2012) e Taleb (2005), quando muitos investidores se concentram nos mesmos modelos de investimento, a tendência é que, com o tempo, os retornos se anulem e que momentos de crise ou recessão sejam potencializados. Portanto, estratégias antigas devem ser adaptadas ou até mesmo integradas às novas realidades do mercado, bem como novas estratégias devem ser desenvolvidas.

Este trabalho contribui ainda por ampliar a gama de modelos de seleção de ações que utilizam menos da intervenção humana, e por esse motivo estão menos sujeitas a fatores emocionais no processo decisório, seguindo uma tendência adotada pelos chamados fundos *quant*, os quais são ainda pouco utilizados no mercado brasileiro (FONTENELE, 2020). Com o desenvolvimento da inteligência artificial, do *Big Data* e da tecnologia de computação em nuvem, as pessoas conhecem gradualmente o investimento quantitativo, que emite instruções de negociação por meio de métodos quantitativos e programação de computadores (WANG *et al.*, 2020). Nas últimas duas décadas, o investimento quantitativo tornou-se um *hot spot* no desenvolvimento do mercado de capitais na Europa e nos Estados Unidos (EUA). A escala de mercado desse investimento chegou a 70% do mercado de investimento nos EUA e tornou-se um novo método para investidores, cujo desempenho é estável (WANG; XU; LIU, 2020).

Modelos de seleção de ações são um tópico especial em gerenciamento e finanças empresariais que é muito interessante ser discutido por investidores e pesquisadores. O gerente de um fundo de ações deve fazer boas escolhas, o que colabora para o crescimento tanto do fundo quanto das empresas nas quais investe. O bom desempenho de um fundo colabora ainda para os resultados obtidos pelos investidores que nele investem seu capital (MANURUNG; NATASHA; BUDIHARTO, 2020).

De forma indireta, pode-se afirmar que este trabalho contribui para a sociedade e para a economia. Entende-se que modelos consistentes de investimentos podem, além de manter os investidores no mercado, fazer com que mais investidores tenham interesse em investir e obter ganhos de capital ou de dividendos. Esses ganhos geram renda para os investidores e tributos para o Estado, seja de forma direta sobre a renda, seja de forma indireta sobre o consumo.

No contexto das companhias, o reduzido acesso a financiamentos, em especial em longo prazo, é um dos principais problemas do ambiente corporativo brasileiro, segundo Fonseca e Silveira (2016). Nesse sentido, o trabalho colabora para a manutenção de investidores em ações no mercado, e isso facilita o acesso a capital por parte das companhias,

as quais, se tiverem acesso facilitado ao capital e forem bem geridas, podem contribuir para a sociedade em forma de emprego e de renda.

Por fim, o sistema proposto pode ainda ser utilizado como ferramenta de apoio às decisões dos investidores, isto é, ao final da seleção das melhores ações para compor a carteira, estas podem passar por um segundo estágio de análise com base em outros critérios quantitativos e qualitativos, por exemplo, expectativa de crescimento e de geração de fluxo de caixa futuros na visão dos gestores. Consequentemente, o sistema vai minimizar o trabalho dos investidores analisando por meio do modelo estatístico um grande número de variáveis e de empresas e, a partir dessa análise, mostrar quais ações têm maior probabilidade de retornos – valorização do preço das ações – futuros positivos e quais devem seguir para o segundo estágio de análise. Nesse caso, as ações passariam por dois estágios de análise, sendo um quantitativo e outro qualitativo.

## **1.6 Estrutura da tese**

O trabalho está dividido em mais cinco capítulos além desta Introdução. No Capítulo 2, o qual concerne ao referencial teórico, aborda-se as principais teorias que fundamentam o tema seleção de portfólios, as estratégias de investimento e os estudos sobre o tema.

No Capítulo 3, dedica-se a descrever a metodologia empregada nesta pesquisa, em que serão expostas a amostra, com os procedimentos metodológicos utilizados na seleção e formação das carteiras. São ainda apresentados os raciais das variáveis utilizadas no trabalho.

No Capítulo 4, apresenta-se a análise e a discussão dos resultados.

Por fim, no Capítulo 5, trata-se das conclusões, das limitações e das recomendações do trabalho.

A seguir, no próximo capítulo, apresenta-se o referencial teórico, no qual estão compreendidas as principais teorias que embasaram o estudo.

## 2 REFERENCIAL TEÓRICO

Portfólios ou carteiras significam manter diferentes tipos de títulos para reduzir riscos associados aos investimentos. O patrimônio de um investidor não pode ser mantido em um único título. Em vez disso, ele deve investir em grupos de títulos. O gerenciamento de carteiras permite-lhe manter seus investimentos de maneira segura combinando entre diferentes carteiras. Um investidor com entendimento em gerenciamento de carteiras tem maiores chances de sucesso no mercado (SINGH, 2020).

Dependendo das metas e dos objetivos definidos pelo investidor na construção de uma carteira de investimento, um gestor de carteiras deve escolher uma certa porcentagem entre diferentes tipos de ativos que compõem o portfólio do investidor. Um portfólio construído com diferentes tipos de investimentos vai, em média, apresentar maiores retornos e menores riscos que qualquer investimento isolado dentro desse portfólio (SHCHANKINA; PING, 2020).

Segundo Markowitz (1952), devem-se considerar todos os investimentos como um só portfólio. Para ele, o investidor deve possuir investimentos que, combinados entre si, formam uma carteira que apresente o maior retorno possível com o menor risco possível. Ainda segundo o autor, existe uma relação direta entre o risco e o retorno, e quanto maior o risco que o investidor estiver disposto a assumir, maior poderá ser o retorno de sua carteira. O autor destaca que a importância da formação de carteiras ocorre principalmente pelo poder que elas têm de dissipar o risco, tendo em vista que o risco de um ativo isolado é maior do que o risco do mesmo ativo quando inserido em uma carteira.

Lyle e Yohn (2019) argumentam que a construção de portfólios consiste em dois estágios independentes: (1) levantamento das ações que se entende que apresentarão retornos futuros interessantes com base em seus fundamentos; (2) decidir quanto de recursos será alocado para cada uma dessas ações. Os autores chamam este último estágio de otimização do portfólio.

No que se refere ao primeiro estágio citado por Lyle e Yohn (2019), muitos modelos têm sido criados a partir do artigo de Markowitz (1952), com objetivo de obter a melhor relação risco *versus* retorno. Os modelos estatísticos, cada vez mais robustos e precisos, têm contribuído para o desenvolvimento de estratégias e para o desenvolvimento do mercado de maneira geral, entre eles, estão os modelos hierárquicos, ainda pouco utilizados no Brasil. A principal vantagem destes sobre os modelos tradicionais de regressão estimados,

por exemplo, por Mínimos Quadrados Ordinários (MQO), refere-se à possibilidade de que seja levado em consideração o aninhamento natural dos dados (FÁVERO; BELFIORE, 2017). Ainda segundo esses autores, em outras palavras, os modelos hierárquicos permitem que sejam identificadas e analisadas as heterogeneidades individuais dos componentes de grupos (investidores, estudantes, colaboradores, etc.) e entre os grupos aos quais esses indivíduos pertencem, tornando possível a especificação de componentes aleatórios em cada nível da análise.

Quanto ao segundo estágio citado por Lyle e Yohn (2019), no qual se define a proporção dos recursos que será alocada em cada ação, é uma questão que pode ser arbitrária e decidida com base na expectativa sobre a economia e sobre o mercado a partir da visão do gestor da carteira, bem como pode ser elaborada por meio de sistemas de otimização de portfólios.

Fama (1998) argumenta que o preço dos ativos reflete completamente todas as informações disponíveis, as quais são processadas por agentes maximizadores de utilidade e minimizadores de risco, ação conhecida como Hipótese de Mercado Eficiente (HME). De acordo com Ehrhardt e Brigham (2012), a HME afirma que as ações estão sempre em equilíbrio e que é impossível um investidor vencer o mercado e consistentemente obter uma taxa de retorno mais alta do que é justificado pelo risco das ações.

Damodaran (2012) defende que o mercado não é eficiente, e essa ineficiência faz com que as filosofias de investimento possam ser eficientes, dentro das quais existem várias estratégias que podem ser utilizadas. As estratégias de investimento visam explorar às ineficiências do mercado para obtenção de vantagens sobre elas.

Além das características de ineficiência encontradas no mercado, fazer previsões precisas utilizando modelos econométricos é um desafio, devido às características não lineares, caóticas e não paramétricas encontradas nos dados distribuídos em séries temporais, que muitas vezes são advindas de eventos inesperados (LUU; NGUYEN; PHAM, 2020; ZHANG; CHU; SHEN, 2020).

Partindo do princípio que, na maioria das vezes, o mercado não é eficiente, os investidores buscam identificar as companhias que estão supervalorizadas ou subvalorizadas (ARTUSO; CHAVES NETO, 2012). O preço das ações de uma empresa depende do histórico de lucros e da expectativa dos lucros futuros. Um investidor inteligente está preparado para ler, entender e implementar uma estratégia de investimentos que incorpora as informações divulgadas pelas companhias (GUERARD; XU; WANG, 2019).

Jacobs e Levy (1993) apontam que geralmente os investidores estão em busca de ações subvalorizadas que possam ser adquiridas para realização de ganhos futuros. Segundo esses autores, existe um lado pouco explorado, que é o de obter lucros com as ações supervalorizadas por meio da chamada venda a descoberto. No contexto desta, o investidor obtém lucro com a queda das ações e ainda pode utilizar a receita alcançada com essa prática financeira para comprar ações subvalorizadas ou outros ativos. Essa estratégia de investimento é conhecida como *long and short*, a qual é abordada na próxima seção.

## 2.1 Estratégia *long and short*

De maneira geral, a implementação dessa estratégia é simples. Após as estimações dos valores esperados de retornos para uma lista de ações, elas são classificadas em ordem decrescente. No topo, estarão as ações com estimativas mais promissoras ou, no caso do presente estudo, as ações com maior probabilidade de retornos futuros positivos. Na base da classificação, estarão as ações com estimativas ruins, ou seja, as ações para as quais não se há expectativa ou estimativa de retornos futuros positivos. A partir desse ponto, cabe ao investidor decidir qual o percentual das melhores ou das piores ele vai alocar na carteira.

Em estudo realizado por Hou *et al.* (2020), foi utilizada a seguinte estratégia: os autores classificaram e dividiram a lista de ações em decis. O primeiro decil de ações fez parte do lado *long* da carteira, e o décimo decil, do lado *short*. Os autores citam que um dos principais benefícios da estratégia *long and short* é que os movimentos do mercado impactam muito pouco a taxa de retorno da carteira. Por exemplo: suponha que um investidor aloque uma posição vendida no valor de 1 milhão de dólares com ações da empresa Apple e, ao mesmo tempo, aloque uma posição comprada também no valor de 1 milhão de dólares com ações da empresa Microsoft. Hipoteticamente, caso algum evento gere impacto no setor de tecnologia como um todo, o investidor vai perder com a queda das ações da Microsoft, mas vai lucrar ao mesmo tempo com as ações da Apple.

Jacobs, Levy e Starer (1999) citam três tipos de estratégia para implementar as estratégias *long and short*: *market neutral*, *equitized* e *hedge*. A primeira mantém o lado *long* e o lado *short* em equilíbrio o tempo todo. Dessa forma, elimina-se a exposição ao risco de mercado fazendo com que a carteira não seja afetada independentemente da direção do mercado. Nessa estratégia, os lucros são obtidos a partir da diferença de desempenho entre os lados *long* e *short*. A *equitized* complementa a anterior adicionando um índice de mercado de ações na carteira de igual valor ao capital investido. Os lucros ou os prejuízos são obtidos das



variações das pontas *long* e *short*, bem como das variações no índice de ações. Finalmente, a *hedge* é similar às estratégias anteriores, porém é adicionado à carteira um índice futuro de ações que varia de acordo com as expectativas do mercado. Assim como a *equitized*, os lucros ou os prejuízos são alcançados nos dois lados (*long* e *short*) e nos retornos do índice.

Desse modo, a estratégia de investimento *long and short* tem como vantagem aproveitar as duas direções de preços das ações, ou seja, ações que projetam alta futura devem ser compradas e ações que projetam baixa futura devem ser vendidas a descoberto. Venda a descoberto é quando um investidor vende as ações emprestadas (ou alugadas, visto que há o pagamento de uma taxa predefinida) por determinado preço com a expectativa de recomprá-las futuramente quando estiverem mais baratas (BHAGWAN *et al.*, 2021; SIMON, 2019).

Assumir uma posição vendida gera riscos, custos e atritos adicionais que não existem na posição comprada. O máximo que um investidor pode perder em uma posição comprada é 100%, enquanto o retorno máximo é ilimitado. Com uma posição vendida acontece o contrário, pois o máximo que se pode ganhar é 100%, mas o máximo que se pode perder é ilimitado, visto que, no lado vendido a descoberto, quanto mais a ação sobe, mais o investidor perde. Por isso, são importantes medidas para limitar perdas. Ademais, os locatários das ações vendidas podem eventualmente solicitar a devolução das ações sem aviso prévio, e isso força os vendedores a encerrarem a posição em momentos não favoráveis, caso não haja uma opção para substituição (BEAVER; MCNICHOLS; PRICE, 2016).

No contexto do portfólio, o objetivo do investidor é obter o máximo de retorno possível assumindo o mínimo de risco. Muitas medidas são tomadas para mensurar o risco, sendo as mais comuns o desvio padrão, a variância e o beta. Para a presente tese foi adotado o desvio padrão como medida de risco. De maneira geral, o desvio padrão mostra que, quanto mais distante um retorno está da sua média, mais arriscado ele é. Na construção de carteiras de ações, aversão ao risco é comumente entendida como aversão ao risco sistemático, ou seja, aversão ao risco para ativos mais arriscados e ativos menos arriscados (ALIAGA-DIAZ *et al.*, 2020).

Embora as posições vendidas sejam concentradas em empresas nas quais os investidores têm expectativas de quedas nos preços das ações, geralmente são utilizadas por fundos de *hedge* para cobrir as posições compradas (JIAO; MASSA; ZHANG, 2016). Essa mesma estratégia pode ser empregada por outros investidores e, assim, o modelo ficou conhecido como uma estratégia de zero investimento (*zero-investment strategy*), tendo em vista que o lado *short* vai servir de recursos para comprar o lado *long*. Vale lembrar que, de acordo com legislações específicas, nem todos podem trabalhar com vendas a descoberto

(*short*), por exemplo, alguns tipos de fundos de investimento. Além disso, para que se possa vender a descoberto, é necessário que se vinculem garantias, que podem ser desde saldo em conta até mesmo investimentos em CDBs de liquidez diária. Essas garantias variam de um país para outro (ZAREMBA *et al.*, 2019)

As estratégias *long and short* são vistas ainda como excelentes medidas de proteção para portfólios. Em estudo realizado por Harvey *et al.* (2019), foram testadas importantes estratégias no mercado americano, com objetivo de verificar qual é mais eficiente para mitigar perdas durante tempos de crise ou de recessão. É notório que as estratégias de *hedge* são difíceis e caras. Sendo assim, além de testar as principais estratégias utilizadas com esse objetivo, os autores testaram outras estratégias de investimento que podem exercer o mesmo papel de proteção. A primeira estratégia testada foi a utilização de opções de venda de longo prazo do índice S&P 500. Esse é o mais confiável e defensivo método utilizado, porém pode custar até 7% dos retornos da carteira, o que a torna muitas vezes inviável. Outra ferramenta de proteção, conhecida como “porto seguro” (*safe haven*), é a alocação de títulos do tesouro americano (*U.S. Treasury Bonds*). Em termos de confiabilidade e custo, existem ainda duas estratégias que podem ser inseridas entre as duas anteriores: investimento em ouro e proteção de crédito (*credit protection*). Com propósito de buscar estratégias de menor custo, os autores evoluíram os testes para mais duas estratégias diferentes. A primeira, estratégia de *momentum* de séries temporais, apresentou bom desempenho tanto em momentos de crise quanto de recessão. A segunda, estratégia *long and short* em ações, utilizou diferentes métricas de qualidade para classificar as empresas transversalmente. Os melhores resultados foram apresentados por esta última estratégia, que se mostraram satisfatórios ao longo do período estudado (1985-2018), em especial nos momentos de crise e recessão.

Beaver e McNichols e Price (2016) estudaram as estratégias *long and short* de maneira diferente em relação a que geralmente propõem os estudos acadêmicos. Os autores assumiram que o custo das carteiras não é zero e atribuíram um custo que foi descontado dos retornos das carteiras. Os resultados mostraram que o principal benefício das estratégias *long and short* é promover a diversificação em conjunto com outras estratégias. Isoladamente, a estratégia *long and short* não foi capaz de bater o retorno do mercado.

Mikutowski, Kambouris e Zaremba (2019) analisaram 124 companhias listadas nos Emirados Árabes Unidos. Foram elaboradas carteiras *long and short* utilizando quatro diferentes variáveis preditivas de retorno representando o valor da empresa: Preço sobre Lucro (P/L), valor contábil sobre valor de mercado (B/M), *Dividend Yield* (DY) e EBITDA sobre valor de mercado (E/EV). Cada um dos indicadores citados foi classificado e dividido

em tercis. As melhores classificações formaram o lado *long* da carteira, e as piores classificações, o lado *short* da carteira. Os indicadores foram revistos, e as carteiras, reformuladas mensalmente pelo período de janeiro de 2004 a março de 2019. O indicador B/M trouxe o melhor resultado, com média mensal de 1,66%.

A *long and short* pode ainda ser classificada como uma subcategoria de investimento, ou seja, o investidor tem uma estratégia principal que, de maneira geral, vai selecionar os ativos, os quais posteriormente serão classificados como melhores ou piores, de acordo com determinado indicador. Após essa classificação, são formadas as carteiras *long and short*.

Nesta tese, na estratégia principal para determinação dos melhores ou dos piores ativos, foram utilizados modelos hierárquicos de regressão para dados em painel. Seu funcionamento é detalhado nas seções 3.2 e 3.3 do próximo capítulo. A seguir, trata-se de estudos sobre carteiras de investimentos que serviram de base para este trabalho.

## 2.2 Estudos sobre carteiras de investimento

Esta seção apresenta alguns artigos recentes no contexto de formação de carteiras de investimento. Nota-se uma um número grande de metodologias, que vão desde modelos simples, como o proposto por Markowitz (1952), até modelos mais complexos e não muito transparentes envolvendo *machine learning*. Ressalta-se que, além de artigos que formaram as carteiras propriamente ditas, selecionou-se também artigos os quais, apesar de não formarem carteiras, contribuem para o estudo com ferramentas que são úteis nessas aplicações.

Conforme destacam Grinold e Kahn (2000), as instituições americanas vêm utilizando estratégias *long and short* desde os anos 1980, as quais, a despeito de gerarem controvérsias, ganharam aceitação ao longo do tempo e têm mostrado que valem a pena. Segundo esses autores, qualquer estratégia de investimentos pode ser aplicada no formato *long and short*, pois, sob o mesmo prisma em que se definem as melhores ações para investir, pode-se definir também, por exemplo, as piores ações e as ações neutras.

De acordo com Black *et al.* (2014), a literatura sobre previsão do retorno de ações foca especificamente a habilidade de variáveis técnicas e fundamentalistas concernentes às empresas para preverem os retornos das ações. Posto isso, os autores propuseram uma pesquisa com uma abordagem diferente das tradicionais analisando o poder das *commodities* para previsão dos retornos no mercado de ações. Para fazer essa análise, eles usaram o índice S&P 500, um índice representativo do mercado de ações americano, e o índice S&P GSCI, o

qual é composto por vinte *commodities*. Como *benchmark* para fins de comparação, esses autores utilizaram a estimativa na qual se incluiu apenas o termo constante que representa a média histórica. A evidência na amostra apresentou poder preditivo significativo, no entanto os resultados fora da amostra não foram capazes de superar a média histórica em várias simulações.

Gulay e Emec (2018) verificaram o poder de previsão dos modelos NoVas e *Exponential Generalized Autoregressive Conditional Heteroscedasticity* (GARCH), assim como seus derivados *Exponential Generalized Autoregressive Conditional Heteroscedasticity* (EGARCH) e *Glosten-Jagannathan-Runkle Generalized Autoregressive Conditional Heteroscedasticity* (GJR-GARCH), para os índices S&P 500 e BIST 100. Foram coletadas informações diárias de retornos para os períodos de 1983 a 1991 e 2006 a 2013. Não obstante os modelos apresentarem resultados parecidos *in sample*, o NoVas apresentou melhor desempenho *out of sample*, mesmo quando se alternavam em diferentes períodos *in sample* e *out of sample*. Como critério de avaliação para a superioridade, foram utilizados o *Root Mean Square Error* (RMSE) e *Mean Absolute Error* (MAE).

Baetje (2018) empregou regressão linear para prever o retorno de ações utilizando como amostra 124 variáveis macroeconômicas e indicadores financeiros. Ele usou como amostra o período de 1960 a 2012, dentro do qual alternou entre momentos de treino (*in sample*) e teste (*out of sample*) para fazer previsões de curtos e longos prazos. Evidências mostraram que, *in sample*, as variáveis macroeconômicas exercem crucial explicação nos movimentos dos retornos das ações e geram boas previsões. No que se refere ao comportamento fora da amostra, observou-se que o modelo apresenta boas previsões no curto prazo, contudo apresenta previsões ruins para o longo prazo. Acredita-se que a dinâmica do mercado mude constantemente, com novas empresas, novos investidores e novos comportamentos. Diante disso, quanto mais distante é o período de previsão, mais difícil é prever os retornos nos mercados.

Zaremba *et al.* (2019) inovaram ao aplicar a estratégia *momentum* ao mercado de *commodities*. A partir dessa estratégia e utilizando 25 diferentes *commodities*, os autores criaram carteiras *long and short* no período de 1986 a 2017. Apesar de as carteiras apresentarem melhor retorno que o *benchmark*, não apresentaram melhor índice de Sharpe. O conceito da estratégia *momentum* basicamente defende a hipótese de que, se uma ação gera retornos positivos por determinado período, a tendência é ela continuar gerando tais retornos. O contrário também é verdadeiro, ou seja, se uma ação gera retornos negativos por determinado período, a tendência é que ela continue se comportando desse modo.

Zhang *et al.* (2019) propuseram uma nova abordagem de investimentos indo ao sentido contrário do que as práticas de gestão de portfólios geralmente recomendam. De maneira geral, é difícil para um investidor ou gestor de portfólios confiar em previsões altamente discrepantes, ou seja, com valores extremamente positivos ou extremamente negativos, os chamados *outliers*. E foi justamente a partir desses destes que os autores formularam carteiras de investimentos. Os autores utilizaram como amostra informações das empresas listadas no S&P 500 referentes ao período de 1927 a 2017, sendo que os primeiros vinte anos foram empregados na construção do modelo econométrico. Para construção do modelo, os autores valeram-se de regressões univariadas e multivariadas utilizando variáveis comumente usadas na previsão de retornos de ações. Mesmo sendo um modelo simples, foi possível apresentar melhores resultados que os modelos tradicionais, por exemplo, o modelo de mínima variância.

Yu *et al.* (2019) utilizaram o *Autoregressive Integrated Moving Average* (ARIMA) para prever os retornos de séries temporais de 21 *Exchange Traded Funds* (ETFs) no período de 2001 a 2015. Os retornos previstos foram então integrados aos seguintes modelos de portfólios: *mean-variance*, *mean-absolute deviations*, *downside deviation*, *linearized value-at-risk*, *conditional value-at-risk* e *omega*. O objetivo principal do trabalho foi verificar se havia benefício nos retornos ao incluir nos modelos anteriores o retorno previsto de acordo com o ARIMA. Os resultados se mostraram consistentes e significantes para os modelos *omega* e *mean-variance*, e os demais modelos se apresentaram bastante voláteis ou com baixo retorno. Os autores observaram ainda que, durante os momentos de crise, a incorporação da previsão gerada pelo ARIMA foi menos efetiva, e isso sugere que o modelo proposto por eles é mais vantajoso em momentos de estabilidade no mercado.

Hou *et al.* (2020) testaram diferentes modelos de seleção de ações valendo-se de séries temporais. Os autores aplicaram métodos clássicos de *machine learning*, abordagens com *deep learning*, bem como modelos híbridos. O objetivo do trabalho foi prever mensalmente o retorno de ações com base no histórico dos dados. Como amostra, foram utilizados os preços das ações ordinárias das empresas listadas no índice S&P 500, ajustados, quando necessário, com os dividendos e desdobramentos. A amostra, que compreendeu dados de 1989 a 2018, foi dividida em diferentes períodos de treino e teste dependendo do modelo estatístico empregado. O modelo preditivo que gerou os melhores resultados foi um modelo híbrido em que se aplicou *Long Short-Term Memory* e *Deep Neural Network*, que os autores chamaram de (LSTM-DNN). A partir das previsões mensais geradas pelo modelo, as ações foram classificadas e divididas em decis. As ações enquadradas no primeiro decil fizeram

parte do lado *long* da carteira, enquanto as ações listadas no último decil fizeram parte do lado *short*. Mediante o modelo, foi possível obter um retorno acumulado no período de 3.556%.

Ratto *et al.* (2018) também testaram ferramentas para seleção de ações utilizando um modelo preditivo híbrido de análise técnica com *machine learning*. Segundo os autores, nos últimos vinte anos, pesquisadores e investidores têm investido no desenvolvimento de muitas maneiras de prever as tendências do mercado. No entanto, dada sua complexidade, esse ainda é um desafio sem uma solução satisfatória. O modelo proposto pelos autores foi aplicado nas vinte empresas mais capitalizadas listadas na *National Association of Securities Dealers Automated Quotations* (NASDAQ). Foram coletados e avaliados os preços das ações no intervalo de quinze em quinze minutos entre as datas de 17/07/2017 e 25/05/2018, compreendendo um total de 6 mil observações por empresa. O sistema desenvolvido pelos autores foi preciso em prever a direção do preço das ações em 61,69% dos casos.

De acordo com Dai e Zhu (2020), apesar de os modelos híbridos terem despertado o interesse de pesquisadores, alguns estudos mostram que a combinação de modelos pode levar ao *overfitting*. Nesse sentido, os modelos híbridos podem até atingir grandes níveis de acerto *in sample*, mas dificilmente alcançam bons resultados na previsão do retorno de ações *out of sample*. Previsões formadas a partir de vários modelos podem melhorar substancialmente a *performance* das previsões quando comparados aos modelos individuais, no entanto é importante saber como combiná-los (HUANG; CHAI; CHO, 2020; ROCCAZZELLA; GAMBETTIA; VRINS, 2020).

Wang *et al.* (2018) utilizaram uma amostra com informações mensais das empresas listadas no índice S&P 500 de janeiro de 1927 a dezembro de 2015. Alicerçados nessas informações, os autores desenvolveram um modelo de previsão que se alterna entre dois submodelos no decorrer do tempo. As previsões são feitas com base em regressões univariadas usando informações econômicas e financeiras da empresa e do mercado, bem como a média de retorno de um *benchmark*. Os autores mostraram que a *performance* de previsão de algumas variáveis, entre elas, *dividend yield*, *dividend payout*, taxa de lucro-preço, taxa da T-Bill e retornos de longo prazo, são persistentes. Eles argumentam que a boa *performance* recente na previsão é sempre seguida por uma boa previsibilidade futura e chamam esse fenômeno de *Momentum of Predictability* (MOP). Nesse sentido, quando determinado modelo supera o *benchmark* durante um período recente, ele é aplicado para prever o retorno de ações, senão é utilizada a média histórica do *benchmark*. Com objetivo de testar o modelo na alocação de ações, foi utilizado o período de janeiro de 1948 a dezembro

de 2015 *out of sample*. Entre as variáveis que trouxeram melhores retornos excedentes estão o *dividend yield*, o *dividend payout* e o lucro sobre preço, com retorno anualizado excedente de, respectivamente, 1,56%, 2,07% e 2,2%.

Li e Song (2019) empregaram o sistema de Redes Neurais Artificiais (RNA), uma poderosa ferramenta para previsão e modelagem não linear. Os autores argumentam que por meio dela é possível prever eficientemente o retorno de ações desde que se empregue um modelo matemático apropriado. De maneira geral, entre 80% e 90% dos modelos RNAs usam um algoritmo chamado *Back Propagation* (BP). Foram coletadas informações diárias das empresas listadas na Nikkei 225 de janeiro de 2008 a junho de 2018, totalizando 2.400 dias de negócios na amostra. Como predictoras, foram selecionadas cinco variáveis que tiveram seus valores normalizados antes de serem inseridos no modelo. No contexto dos modelos preditivos, é importante que sempre seja utilizado um período de treino, no qual o modelo “aprende” o comportamento de determinada amostra, e, em seguida, um período de teste, ou seja, um período de validação do modelo. Foram utilizados diferentes períodos de treino e de teste, sendo mantido sempre 230 dias como treino e 30 dias como teste. Por fim, o modelo foi testado em três empresas, e o nível médio de assertividade apresentado foi de 79,6% para a Sony Corporation, 77,6% para a Panasonic Corporation e 70,6% para a Toyota Corporation, isto é, em mais de 70% dos casos, o modelo foi preciso em prever se uma ação teria alta ou baixa em determinado período.

Wang *et al.* (2020) aplicaram regressão logística e *Support Vector Machine* (SVM) para prever os retornos de ações no mercado chinês. O modelo de teste foi criado a partir da amostra de 3.734 ações chinesas que apresentaram seus resultados entre 2008 e 2017. Após a formulação dos modelos, eles foram testados formando carteiras para o período de 2018 e 2019 e, desse modo, foi possível obter um retorno anualizado de 16,38% para a regressão logística e 17,13% para o SVM. Os resultados são superiores ao principal índice chinês de ações, o HS300.

Kim, Jeong e Lee (2019) utilizaram, na alocação de carteiras, um modelo introduzido por Hamilton (1989) chamado de cadeias de Markov, o qual possibilita que se capturem as probabilidades de mudanças de tendências em séries temporais. Isso pode ser aplicado, por exemplo, em séries temporais de índices e de preços de ativos. Os autores utilizaram 23 ETFs representando cada classe de ativos divididos em vários países, para formular portfólios de investimento no período de 2004 a 2019. Mediante o modelo foi possível identificar quando era hora de comprar e quando era hora de vender determinado ativo, em outras palavras, o peso de um ativo na carteira era incrementado quando o sistema

detectava probabilidade de alta e era diminuído – aumentando o peso de outros ativos – quando o sistema detectava probabilidade de baixa. Os resultados da estratégia foram comparados ao modelo de mínima variância e também à estratégia *momentum*, apresentando melhor resultado que ambos.

Lewin e Campani (2020), seguindo a mesma metodologia de Kim, Jeong e Lee (2019), aplicaram o modelo ao mercado brasileiro. Com objetivo de capturar as características desse mercado, foram utilizados o Certificado de Depósito Interbancário (CDI), a renda fixa – representada pelo índice IMA-G, que reflete um portfólio diversificado em títulos do tesouro –, a renda variável nacional (IBrX-100) e a renda variável internacional (S&P 500). Foram coletados dados semanais desses indicadores no período de 07 de dezembro de 2001 até 31 de agosto de 2018, totalizando 873 observações. A partir do modelo desenvolvido por Lewin e Campani (2020), foram criadas carteiras de investimento *long only* e *long and short* com diferentes tipos de alocações nos quatro indicadores citados anteriormente. Vale lembrar que o objetivo das cadeias de Markov foi detectar mudanças nas tendências de maneira geral, seja num indicador econômico, seja num indicador de mercado. Quando o modelo desses autores apontava para uma tendência otimista no mercado, uma parcela maior era alocada em ações, tendo em vista menor aversão ao risco e mais possibilidade de ganhos. Caso sinalizasse um cenário pessimista no mercado, havia maior alocação em renda fixa. A alocação que não permitiu vendas a descoberto (*long only*) gerou um prêmio de risco anualizado de 4,66% com volatilidade de 7,62% e índice de Sharpe de 0,61%. A alocação com vendas a descoberto (*long and short*) apresentou resultado anualizado significativamente maior, sendo o prêmio de 36,90%, volatilidade de 33,33% e índice de Sharpe de 1,11%. Apesar do alto prêmio de risco anual da carteira, nota-se também grande volatilidade. No entanto, mesmo com alta volatilidade, o modelo trouxe melhores resultados que os principais *benchmarks* dos mercados nacional e internacional, e isso faz com que seja uma alternativa interessante de investimento.

Por fim, Shchankina e Ping (2020) utilizaram dois métodos para construção de carteiras, com objetivo de apurar qual deles gera melhores resultados na Bolsa de Valores de Shanghai e na Bolsa de Valores de Hong Kong. O primeiro é o método clássico de otimização de carteiras criado por Markowitz (1952), que, por mais de 60 anos, vem sendo estudado e empregado por pesquisadores e investidores. No segundo método, foi utilizada série temporal por meio do modelo GARCH. Os princípios utilizados para seleção de ações usando o modelo Markowitz foram liquidez, *performance* da indústria, diversificação por indústria e os coeficientes do preço sobre lucro e preço sobre valor patrimonial. A estimação *in sample* foi



calculada por intermédio dos preços diários das ações entre o período de 04/01/2013 e 28/12/2018 tanto para o modelo de Markowitz quanto para o modelo GARCH. O teste para o período *out of sample* foi apurado com base nos dados referentes ao período de 02/01/2019 a 31/01/2019. Os resultados do estudo confirmaram que as carteiras construídas a partir do modelo GARCH tem menor desvio padrão e melhores retornos do que as carteiras construídas a partir do modelo Markowitz.

O Quadro 1 apresenta de forma resumida e cronológica os trabalhos apresentados nas seções 2.1 e 2.2 deste capítulo. As metodologias desses estudos serviram de principais referências para o desenvolvimento desta tese, apesar de utilizarem diferentes métodos. Nota-se uma grande diversidade de métodos, que variam desde a utilização de indicadores, estatística e inteligência artificial. A presente tese enquadra-se em um modelo estatístico.

**Quadro 1 – Trabalhos de autores que fundamentaram o desenvolvimento da tese**

<b>Autor</b>	<b>Amostra</b>	<b>Período</b>	<b>Método</b>
Black <i>et al.</i> (2014)	S&P 500 e S&P GSCI <i>Commodity</i>	1973 a 2012	Regressão linear
Beaver, McNichols e Price (2016)	Ações listadas na NYSE, na AMEX e na NASDAQ	1992 a 2013	Indicadores técnicos e <i>long and short</i>
Wang <i>et al.</i> (2018)	Ações listadas no S&P 500	1927 a 2015	Regressão linear
Gulay e Emec (2018)	S&P 500 e BIST 100	1983 a 1991 e 2006 a 2013	NoVas, GARCH, EGARCH e GJR-GARCH
Baetje (2018)	Variáveis macroeconômicas indicadores financeiros	1960 a 2012	Regressão linear
Harvey <i>et al.</i> (2019)	NYSE, AMEX e NASDAQ	1985 a 2018	Diversos métodos sem uso de estatística ou <i>Artificial Intelligence (IA)</i>
Zaremba <i>et al.</i> (2019)	25 <i>commodities</i> de diferentes setores	1986 a 2017	<i>Momentum e long and short</i>
Mikutowski, Kambouris e Zaremba (2019)	124 companhias listadas nos Emirados Árabes Unidos	2004 a 2019	Análise de indicadores
Ratto <i>et al.</i> (2018)	20 empresas mais capitalizadas listadas na NASDAQ	2017 a 2018	SVM
Li e Song (2019)	Ações listadas na Nikkei 225	2008 a 2018	RNA
Zhang <i>et al.</i> (2019)	Ações listadas na S&P 500	1927 a 2017	Regressão linear
Yu <i>et al.</i> (2019)	21 ETFs	2001 a 2015	ARIMA
Kim, Jeong e Lee (2019)	23 ETFs	2004 a 2019	Markov
Wang <i>et al.</i> (2020)	3.734 ações chinesas	2008 a 2019	Regressão Logística e SVM
Hou <i>et al.</i> (2020)	Ações listadas na NYSE, na AMEX e na NASDAQ	1977 a 2018	LSTM-DNN

(continua)

(conclusão)

<b>Autor</b>	<b>Amostra</b>	<b>Período</b>	<b>Método</b>
Lewin e Campani (2020)	CDI, IMA-G, IBrX-100 e S&P 500	2001 a 2018	Markov
Shchankina e Ping (2020)	Ações listadas nas bolsas de Shanghai e de Hong Kong	2013 a 2019	Markowitz e GARCH
Grinold e Kahn (2000)	Histórico da <i>performance</i> de 14 estratégias <i>long and short</i> americanas	1990 a 1999	Pesquisa exploratória ( <i>long and short</i> )
Dai e Zhu (2020)	Ações listadas na S&P 500	1927 a 2017	Regressão linear

**Fonte:** Elaborado pelo autor.

Nota-se a partir dos trabalhos apresentados que os métodos de construção de carteiras são muito variados, mas objetivam sempre maximizar os retornos e minimizar os riscos. Não foram localizados, a princípio, trabalhos que utilizassem a abordagem hierárquica na seleção de ações. O próximo capítulo trata da metodologia utilizada neste trabalho. Desse modo, apresenta-se a amostra, os procedimentos estatísticos, as variáveis e o racional de cada uma delas, bem como o método utilizado na construção das carteiras.

### 3 PROCEDIMENTOS METODOLÓGICOS

#### 3.1 Amostra

A amostra foi composta por todas as ações listadas nas três bolsas americanas, *New York Stock Exchange* (NYSE), NASDAQ e NYSE *Euronext* (antiga AMEX), e na Bolsa de Valores de São Paulo (B3). Foi utilizada apenas uma ação de cada empresa, dando preferência a de maior liquidez. Nesse caso, cabe ser evidenciado que a ação de maior liquidez nem sempre é uma ação líquida, quando considerado o mercado como um todo, ela simplesmente tem mais negócios que a outra classe de ação na mesma empresa. No total, foram utilizadas 193.445 informações de 5.441 empresas divididas em 258 setores.

O conceito de liquidez pode ser interpretado de diferentes maneiras, no entanto, no contexto do presente trabalho, a liquidez é representada pela quantidade de negócios diários das ações de uma empresa no mercado de capitais, pelo fato de ações mais líquidas poderem ser facilmente convertidas em dinheiro (ABENSUR *et al.*, 2020). A liquidez de uma ação pode ser vista ainda sob outra perspectiva, por exemplo, a NYSE é considerada a bolsa mais líquida do mundo, e a B3, uma bolsa de baixa liquidez. Para grandes fundos de investimento que negociam vultuosos valores, essa questão pode ser muito importante, mas no caso do pequeno investidor isso é irrelevante, tendo em vista que, para seus valores negociados, as ações serão sempre líquidas.

Foram coletadas informações trimestrais referentes ao período de 1995 até 2T2019 por meio do sistema Económica, tratando-se, dessa forma, de dados secundários. A amostra foi dividida em duas partes: utilizou-se o período de 1995 a 2007 *in sample*, para formulação do modelo, e o período de 2008 a 2T2019 *out of sample*, para teste do modelo estatístico. Em muitos trabalhos acadêmicos, sugere-se a prática de dividir a amostra em duas, de forma a se testar a robustez do modelo estatístico. O procedimento de dividir os dados em partes é chamado de método de validação (LANTZ, 2019).

O *out of sample* é um período no qual se simula não ter acesso às informações futuras, e isso permite que o modelo seja testado no mundo real em que também não se tem as informações do futuro. Para Gelman e Hill (2007), no caso de previsões, os dados podem futuramente se tornarem disponíveis, o que possibilita ao pesquisador ver como o modelo funciona para esse fim. Às vezes, previsões fora da amostra são feitas com o objetivo explícito da verificação do modelo.

As informações receberam um tratamento prévio em planilhas eletrônicas e posteriormente foram importadas para o sistema Stata (versão 14), em que ocorreram os procedimentos estatísticos. Na sequência, os resultados foram novamente transferidos para planilhas eletrônicas nas quais foram criadas as carteiras. Estas foram criadas trimestre a trimestre e depois tiveram os resultados consolidados.

A B3, em 31/01/2020, contava com 353 empresas ativas, valor capitalizado de mais de 4,6 bilhões e volume médio diário de 19 milhões. Apesar de ser uma bolsa/mercado menos amadurecida que as citadas anteriormente, tem apresentado crescimento exponencial em número de investidores nos últimos quatro anos. A NYSE, com mais de 2.400 empresas listadas e com volume médio diário de U\$ 123 bilhões, é a maior e mais líquida fonte de capital do mundo (NYSE, 2020). A NASDAQ, apesar de apresentar mais empresas listadas que a NYSE, não a supera em valor capitalizado. Nela estão listadas 1.920 empresas, apesar de nem todas terem sido utilizadas no trabalho. A NYSE *Euronext* (antiga *American Exchange*) é a bolsa de valores com o terceiro maior número de negócios nos EUA. A maior parte das negociações nela realizadas consiste em opções de índice – índice de tecnologia de computadores, índice institucional, índice de mercado – e predominam ações de pequenas e médias empresas.

### **3.2 Modelos hierárquicos**

Os modelos hierárquicos são utilizados quando os dados de determinado assunto de pesquisa estão aninhados de alguma forma como, por exemplo, estudantes podem ser aninhados em escolas, pessoas em bairros e empregados em firmas. Dados longitudinais são, por definição, agrupados, porque múltiplas observações ao longo do tempo são agrupadas dentro de unidades, geralmente pertencentes a determinado assunto (RABE-HESKETH; SKRONDAL, 2012). Os termos aninhamento e agrupamento utilizados nesta tese são equivalentes.

Ainda segundo esses autores, esses projetos em *cluster*, ou seja, com dados aninhados, geralmente fornecem informações valiosas sobre processos que operam em diferentes níveis, por exemplo, as características das pessoas interagindo com as características institucionais. Ressalta-se a probabilidade de a suposição padrão de observações independentes ser violada devido à dependência entre as observações dentro do mesmo *cluster*.

As pesquisas sociais geralmente envolvem estruturas de dados hierárquicas. Em estudos organizacionais, por exemplo, pesquisadores poderiam investigar de que forma as características do local de trabalho, como a centralização do processo decisório, influenciam a produtividade do trabalhador. Nesse caso, trabalhadores e empresas são unidades na análise, e variáveis podem ser mencionadas em ambos os níveis. Tais dados têm uma estrutura hierárquica com trabalhadores aninhados em empresas (RAUDENBUSH; BRYK, 2002).

Considere um estudo na área da educação com dados de estudantes coletados em várias escolas, com objetivo de prever as notas dos alunos em cada uma dessas instituições. Um modelo de regressão separado pode ser ajustado dentro de cada escola, e os parâmetros de cada escola podem ser modelados de acordo com as características individuais delas, tais como *status* socioeconômico das escolas do bairro, se são públicas ou privadas, etc., assim como as características de cada aluno, entre elas, sexo e idade. Desse modo, a regressão no nível do estudante e a regressão no nível da escola formam dois níveis de um modelo hierárquico ou multinível (GELMAN; HILL, 2007).

Quando utilizados em dados longitudinais, os modelos hierárquicos têm semelhanças com os modelos de regressão para dados em painel com efeito aleatório. A diferença é que esses últimos estão limitados a duas dimensões, sendo uma temporal e outra espacial, enquanto os primeiros podem ser compostos por outros níveis. Além disso, é possível incluir variáveis constantes ao longo do tempo de níveis superiores e analisar o comportamento de evolução dessas variáveis dentro do período pesquisado (SANTOS, 2018).

Os modelos hierárquicos permitem não só variáveis independentes em qualquer nível da estrutura, mas também pelo menos um efeito aleatório acima do nível um, segundo Guo e Zhao (2002). Os autores destacam ainda que tais modelos oferecem uma série de vantagens que os modelos tradicionais lineares ou não lineares não apresentam, por exemplo: (1) um modelo hierárquico, naturalmente, provê uma estrutura conveniente para estudar dados que estão disponíveis de forma hierárquica. Essa estrutura permite uma análise sistemática sobre como variáveis independentes mensuradas entre os níveis da estrutura afetam a variável dependente e, além disso, possibilita verificar como as interações entre as variáveis independentes nos diferentes níveis afetam a variável dependente; (2) a modelagem hierárquica corrige os vieses nas estimativas de parâmetros resultantes do armazenamento em *cluster*. Em contraste com a crença popular, ignorar a estrutura hierárquica pode resultar em vieses nas estimativas dos parâmetros e em seus erros padrão. Quanto maior a correlação entre as observações dentro dos *clusters*, maior é a probabilidade desses erros; (3) a modelagem hierárquica fornece erros padrão mais precisos e, portanto, intervalos de

confiança mais eficientes. Quando as observações são agrupadas em unidades de nível superior, as observações não são mais independentes. A independência é uma das suposições mais básicas subjacentes aos modelos tradicionais de regressão. Quando a estrutura de *cluster* nos dados é ignorada e a suposição de independência violada, os modelos lineares e binários tradicionais tendem a subestimar os erros padrão.

As bolsas de valores apresentam uma estrutura hierárquica e agregam as empresas em setores. Como as variáveis são analisadas ao longo do tempo, tem-se, dessa forma, uma estrutura aninhada em três níveis. O terceiro nível é o setor em que as empresas estão enquadradas, o segundo nível é a empresa e o primeiro nível são as características das empresas que variam ao longo do tempo.

### 3.3 Procedimentos estatísticos

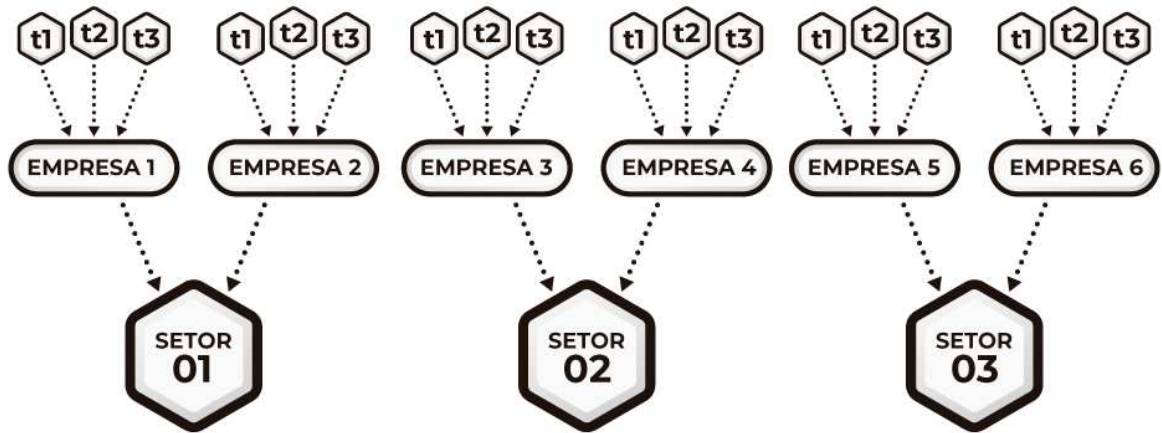
Unidades de observações geralmente estão enquadradas em grupos. Por exemplo, indivíduos podem ser aninhados em famílias, alunos em escolas, bairros em cidades, empregados em empresas e assim por diante. Os dados longitudinais também consistem em agrupamentos de observações feitas em diferentes ocasiões para o mesmo indivíduo ou agrupamento (RABE-HESKETH; SKRONDAL, 2012).

Os modelos hierárquicos de regressão para dados em painel, que também são conhecidos como modelos mistos ou modelos multiníveis, têm adquirido importância considerável em diversas áreas do conhecimento, muito em função da determinação de constructos de pesquisa que consideram a existência de estruturas aninhadas de dados. (FÁVERO; BELFIORE, 2017). Segundo esses autores,

A principal vantagem dos modelos hierárquicos sobre os modelos tradicionais de regressão estimados, por exemplo, por MQO, refere-se à possibilidade de que seja levado em consideração o aninhamento natural dos dados, ou seja, se as empresas estiverem aninhadas em setores, é possível que se defina um componente aleatório no nível de firma e outro no nível de setor, ao contrário do que permitiria um modelo tradicional de regressão, em que o efeito do setor sobre o desempenho das firmas seria considerado de maneira homogênea (p. 856).

Observando a estrutura das bolsas de valores, nota-se que elas apresentam uma estrutura hierárquica, agregando as empresas em setores. Como as variáveis são analisadas ao longo do tempo, tem-se, desta forma, uma estrutura alinhada em três níveis. O terceiro nível é o nível naics (setor), o segundo nível é a empresa (i) e o primeiro nível é o tempo (t). A Figura 1 ilustra de forma resumida os aninhamentos.

Figura 1 – Representação dos níveis dos aninhamentos utilizados no trabalho



Fonte: Adaptado de Fávero e Belfiore (2017, p. 859).

Tendo em vista o aninhamento dos dados e o fato de a variável dependente do modelo estudado ser categórica (*dummy*), o modelo estatístico utilizado na presente tese se enquadra no chamado Modelo Hierárquico não Linear (HNM), o qual é constituído de três níveis. Desse modo, a evolução das variáveis no tempo é representada no nível um, as variáveis referentes às empresas são alocadas no nível dois e, por fim, as variáveis que representam os setores são alocadas no nível três. As variáveis de níveis dois e três são características dos setores e das empresas, e tais características podem variar entre eles (níveis), mas são invariáveis no tempo (FÁVERO; BELFIORE, 2017).

De acordo com o modelo proposto por Raudenbush e Bryk (2002), o nível 1 em um modelo *Hierarchical Generalized Linear Models* (HGLM) consiste de três partes: um modelo de amostragem, uma função de *link* e um modelo estrutural. De fato, o *Hierarchical Linear Models* (HLM) pode ser visto como um caso especial de HGML, em que o modelo de amostragem é normal, a função *link* é a identidade e o modelo estrutural é linear.

### 3.3.1 Nível 1 – Modelo de amostragem

O modelo de amostragem para um HLM3 pode ser escrito como:

$$Y_{tij} | \mu_{tij} \sim NID(\mu_{tij}, \sigma^2) \quad (1)$$

Isso significa que o resultado de nível 1  $Y_{tij}$ , dado o valor previsto,  $\mu_{tij}$ , é normal e independentemente distribuído com um valor esperado de  $\mu_{ij}$  e variância constante  $\sigma^2$ . O valor esperado do nível 1 pode ainda ser representado de forma alternativa, conforme segue:

$$E(Y_{tij}|\mu_{tij}) = \mu_{tij}, \quad VAR(Y_{tij}|\mu_{tij}) = \sigma^2 \quad (2)$$

### 3.3.2 Nível 1 – Função *link*

Para Raudenbush e Bryk (2002), é possível transformar o valor previsto do nível 1,  $\mu_{ij}$ , para garantir que os valores sejam restritos a determinado intervalo. O valor transformado será chamado de  $\eta_{ij}$ , e essa transformação é chamada de função *link*. No entanto, pode-se tomar a decisão de não o transformar, a qual deve ser explicitamente anotada por escrito:

$$\eta_{tij} = \mu_{tij} \quad (3)$$

Nesse caso, a função *link* é vista como a “função *link* identidade”.

### 3.3.3 Nível 1 – Modelo estrutural

O valor transformado de  $\eta_{tij}$  é agora relacionado com os preditores por meio do modelo estrutural linear:

$$\eta_{tij} = \beta_{0tj} + \beta_{1tj} X_{1tij} + \beta_{2tj} X_{2tij} + \dots + \beta_{ptj} x_{ptij} \quad (4)$$

Observa-se que a combinação das equações 1, 3 e 4 leva ao modelo usual de HLM. Raudenbush e Bryk (2002) destacam que, a princípio, parece ser desnecessário se utilizar de três equações, quando, na realidade, simplesmente uma seria suficiente, ou seja, a 4. No entanto, os autores alertam que essas equações extras seriam mais úteis quando se aumenta a complexidade de casos não lineares.

Enquanto o modelo padrão de HLM usa um modelo de amostragem normal e uma função *link* de identidade, o modelo de resultado binário utilizado nesta tese utiliza um modelo de amostragem binomial e um *link logit*. As equações 5 a 12 apresentam os modelos para saídas binárias.



### 3.3.4 Nível 1 – Modelo de amostragem

Seguindo ainda a mesma linha de Raudenbush e Bryk (2002), digamos que  $Y_{ij}$  seja o número de acertos na seleção de ações com retornos positivos em  $m_{ij}$  tentativas e  $\varphi_{ij}$  seja a probabilidade de sucesso em cada tentativa. Dessa forma:

$$Y_{tij}|\varphi_{tij} \sim \beta(m_{tij}, \varphi_{tij}) \quad (5)$$

em que  $Y_{ij}$  tem uma distribuição binomial com  $m_{ij}$  tentativas e probabilidade de acerto por tentativa como  $\varphi_{ij}$ . Conforme a distribuição binomial, o valor esperado e a variância de  $Y_{ij}$  são:

$$E(Y_{tij}|\varphi_{tij}) = m_{tij}\varphi_{tij}, \quad VAR(Y_{tij}|\varphi_{tij}) = m_{tij}\varphi_{tij}(1 - \varphi_{ij}) \quad (6)$$

Quando  $m_{tij} = 1$ ,  $Y_{tij}$  é uma variável binária com valor zero ou um. Esse é um caso especial conhecido como distribuição Bernoulli. No entanto, verifica-se que, em algumas aplicações, os dados não seguem precisamente esse modelo. Nesses casos, é possível generalizar o modelo para estimar um componente de variância escalar,  $\sigma^2$ , para que a variância seja  $\sigma^2 m_{tij} \varphi_{tij} (1 - \varphi_{ij})$ .

### 3.3.5 Nível 1 – Função *link*

Muitas funções *link* são possíveis quando o modelo de amostragem nível 1 é binomial, no entanto o mais comum e conveniente é o *logit link*:

$$\eta_{tij} = \log\left(\frac{\varphi_{tij}}{1 - \varphi_{tij}}\right) \quad (7)$$

em que  $\eta_{tij}$  é o log das probabilidades de acerto. Se a probabilidade de acerto  $\varphi_{tij}$  é 0,5, as chances de sucesso  $\varphi_{tij}/(1 - \varphi_{tij}) = 0,5/0,5 = 1$  e o *log-odds* ou *logit* é  $\log(1) = 0$ . Quando a probabilidade de sucesso é menor que 0,5, as chances são menos que 1,0 e o *logit* é negativo. Quando a probabilidade é maior que 0,5, as chances são maiores que 1,0 e o *logit* é positivo. Observa-se que, enquanto  $\varphi_{tij}$  está restrito ao intervalo de 0 e 1,  $\eta_{tij}$  pode assumir qualquer valor real.

### 3.3.6 Nível 1 – Modelo estrutural

Nota-se a seguir que as estimativas de  $\beta_s$  na equação 4 permitem gerar um *logit* ( $\eta_{tij}$ ) para qualquer caso. Esses *logits* podem ser convertidos em chances utilizando-se  $\exp(\eta_{tij})$ . Uma previsão de *logit* pode também ser convertida em uma previsão de probabilidade mediante a seguinte equação:

$$\varphi_{tij} = \left( \frac{1}{1 + \exp\{-\eta_{tij}\}} \right) \quad (8)$$

Nesse sentido, independentemente do valor de  $\eta_{ij}$ , aplicando-se a equação 8, é possível obter um valor entre zero e um para  $\varphi_{ij}$ . Por fim, a equação 8 pode ser escrita da seguinte maneira, formando a equação final para o nível 1 do modelo *logit*:

$$\varphi_{tij} = \left( \frac{1}{1 + \exp\{-\{\beta_{0tj} + \beta_{1tj} X_{1tij} + \beta_{2tj} X_{2tij} + \dots + \beta_{ptj} X_{ptij}\}} \right) \quad (9)$$

em que  $\varphi_{tij}$  é a probabilidade do retorno positivo no tempo  $t$  para a empresa  $i$  pertencente ao setor  $j$ ,  $\beta_{ptj}$  se referem aos coeficientes do nível 1 e  $X_{ptij}$  é uma  $p$ -éssima variável de nível 1 para a observação  $t$  na unidade de nível 2  $i$  e na unidade de nível 3  $j$ .

O nível 2 pode ser representado de maneira geral por intermédio da seguinte equação:

$$\beta_{ptj} = \beta_{p0t} + \sum_{q=1}^{Q_p} \beta_{pqt} X_{qtj} + r_{ptj} \quad (10)$$

em que  $\beta_{pqt}$  se referem aos coeficientes de nível 2,  $X_{qtj}$  é uma  $q$ -éssima variável explicativa de nível 2 para a unidade  $i$  na unidade de nível 3  $k$ , e  $r_{ptj}$  são os efeitos aleatórios do nível 2, assumindo-se, para cada unidade  $i$ , que o vetor  $(r_{0tj}, r_{1tj}, \dots, r_{ptj})$  apresenta distribuição normal multivariada com cada elemento, portanto média zero e variância  $\tau_{rpp}$  (FÁVERO; BELFIORE, 2017).

O nível 3, por sua vez, pode ser representado de maneira geral mediante a seguinte equação:

$$\beta_{pqt} = \gamma_{pt0} + \sum_{s=1}^{S_{pq}} \gamma_{pqs} w_{st} + u_{pqt} \quad (11)$$

em que  $\gamma_{pqs}$  se referem aos coeficientes de nível 3,  $w_{st}$  é uma  $s$ -ésima variável explicativa de nível 3 para a unidade  $t$  e  $u_{pqt}$  são os efeitos aleatórios do nível 3, assumindo-se que, para cada unidade  $t$ , o vetor composto pelos termos  $u_{pqt}$  apresenta distribuição normal multivariada com cada elemento, portanto média zero e variância  $\tau_{\tau ppp}$ , o que resulta na matriz de variância-covariância  $T_b$  com dimensão máxima igual a:

$$Dim_{máx} = T_b + \sum_{p=0}^P (Q_p + 1) \sum_{p=0}^P (Q_p + 1) \quad (12)$$

que depende da quantidade de coeficientes de nível 3 especificados com termos aleatórios.

### 3.4 Variáveis

O desempenho do mercado acionário está significativamente relacionado às atividades econômicas. Nesse sentido, países que apresentam estabilidade econômica e crescimento consistente têm mercados acionários caminhando no mesmo sentido. O crescimento econômico, que se manifesta principalmente pelo sucesso das empresas em termos de desempenho financeiro favorável, será refletido no mercado de ações na forma de confiança dos investidores, os quais, por sua vez, atrairão mais capital ou fundos de investimento para a economia (ISA *et al.*, 2019).

Os estudos de Wang *et al.* (2018) e Welch e Goyal (2008) constataram que a utilização de variáveis fundamentalistas, por exemplo, *dividend yield* e *dividend-price ratio*, contribuem de forma eficiente para a previsão dos retornos das ações, o que gera retornos positivos e acima dos principais *benchmarks* enquanto o tratamento ocorre dentro da amostra. No entanto, ao realizar testes fora da amostra, essas variáveis deixam de ser interessantes, e as carteiras formadas a partir delas passam a gerar piores resultados. No Quadro 2, estão relacionados os estudos mais recentes sobre formação de carteiras e as variáveis utilizadas.

**Quadro 2 – Estudos sobre construção de portfólios ou previsão de retornos de ações**

<b>Autor / Ano</b>	<b>Método</b>	<b>Variáveis</b>
Dai e Zhu (2020)	Regressão linear	R, DP, DY, EP, DE, SVAR, BM, NTIS, TBL, LTY, LTR, TMS, DFY, DFR e INFL
Zhang <i>et al.</i> (2019)	Regressão linear	R, DP, DY, EP, DE, SVAR, BM, NTIS, TBL, LTY, LTR, TMS, DFY, DFR e INFL
Yu <i>et al.</i> (2019)	ARIMA	R e V
Kim, Jeong e Lee (2019)	HMM (Markov)	Retorno de 23 ETFs representando diferentes classes de ativos, por exemplo, ações, T-BONDS e <i>commodities</i>
Assagaf <i>et al.</i> (2019)	Regressão linear	R, INFL, IR (Indonésia), MS e US DOLLAR.
Gulay e Emec (2018)	NoVas, GARCH, EGARCH e GJR-GARCH	R e V
Baetje (2018)	Regressão linear	124 variáveis macroeconômicas americanas, entre elas, as utilizadas nesta tese
Wang <i>et al.</i> (2018)	Regressão linear	R, DP, DY, EP, SVAR, BM, NTIS, TBL, LTY, LTR, DFY, DFR e INFL
Black <i>et al.</i> (2014)	Regressão linear	S&P 500, S&P GSCI, Dow Jones – UBS, Índice Báltico Seco, Milho, Energia, Bolsa de Metais de Londres e Metais Preciosos

**Fonte:** Elaborado pelo autor.

Nota: Volatilidade (V), Taxa de Retorno (R), *Dividend-Price Ratio* (DP), *Dividend Yield* (DY), *Earnings-Price Ratio* (EP), *Dividend-Payout Ratio* (DE), *Stock Return Variance* (SVAR), *Book-to-Market Ratio* (BM), *Net Equity Expansion* (NTIS), *Treasury Bill Rate* (TBL), *Long-Term Yield* (LTY), *Long-Term Return* (LTR), *Term Spread* (TMS), *Default Yield Spread* (DFY), *Default Return Spread* (DFR), *Inflation* (INFL), *Interest Rate* (IR), *Money Supply* (MS).

Nesse sentido, têm ganhado destaque em estudos de alocação de portfólios não só o comportamento das variáveis endógenas, mas também o das exógenas, como é o caso das variáveis macroeconômicas, dos índices representativos de mercado e do preço de *commodities*, que refletem, de maneira geral, o comportamento da economia mundial (SETIADI; MASDUPI, 2019). Em última instância, pode-se afirmar que o risco das empresas

é, em grande parte, determinado pela exposição delas aos riscos macroeconômicos (BAETJE, 2018).

No presente estudo, também se notou o mesmo comportamento, ou seja, havia perda de desempenho da carteira à medida que se incluíam variáveis endógenas referentes às demonstrações das empresas, mesmo que essas variáveis se apresentassem estatisticamente significantes. Desse modo, optou-se por manter não mais que duas variáveis endógenas: (1) *dummy* da taxa de retorno das ações (variável dependente); (2) taxa de retorno das ações defasado (valor defasado de um a quatro trimestres – variável independente). Além delas, utilizou-se ainda como variáveis independentes a taxa de retorno do S&P 500, taxa de retorno do T-BOND 10 e taxa de retorno do ouro. Em todas as variáveis independentes, empregou-se o logaritmo natural.

A *dummy* da taxa de retorno das ações (*dummyret*) recebeu o valor 1, no caso de retorno positivo no trimestre subsequente à divulgação dos resultados, e o valor 0, no caso de retorno menor ou igual a zero. Na utilização de uma variável binária em um modelo estatístico, os valores previstos são dados em termos de probabilidade. Por exemplo, digamos que a empresa Vale S.A. divulgou as informações financeiras referentes ao 4T2017 em 10/02/2018. A partir do primeiro dia útil posterior à data da publicação, será apurada a taxa de retorno até a divulgação do próximo resultado, ou seja, até a divulgação do resultado referente ao 1T2018. Caso a taxa de retorno desse período seja positiva, a *dummy* receberá 1, caso contrário, receberá 0. A partir desse comportamento ao longo de todo o período, o sistema poderá mostrar qual a probabilidade de o retorno ser positivo, isto é, da *dummy* receber 1 a partir de determinado resultado.

Utilizar um modelo logístico, como no caso deste estudo, permite que sejam aliviadas as necessidades de linearidade, normalidade e homocedasticidade dos resíduos, o que torna o modelo mais flexível, pois os dados de mercado geralmente não atendem a esses pressupostos sem a necessidade de transformações ou adaptações.

A princípio, a variável dependente *dummy* também teve seu valor defasado utilizado como variável independente, porém sem apresentar resultados satisfatórios. Alternativamente, testou-se as próprias taxas de retornos defasadas, obtendo-se, dessa forma, melhores resultados.

*Standard & Poor's 500* (S&P 500) é um índice ponderado por capitalização composto por 500 ações, que tem o objetivo de ser uma amostra representativa das principais empresas dos setores mais importantes da economia americana. Nesse índice, as ações são selecionadas por tamanho de mercado, liquidez e representação de grupos industriais. Um

índice de mercado rastreia o desempenho de um grupo específico de ações que representam determinado setor, mercado ou economia. Há índices para quase todos os setores concebíveis da economia e do mercado de ações.

*Treasury Bonds* (T-BONDS) são títulos de renda fixa emitidos pelo governo americano e considerados os títulos mais seguros do mundo. Para o modelo estatístico adotado neste estudo, optou-se por utilizar os títulos com vencimento para dez anos. Ressalta-se que, além de ser parte integrante do modelo econométrico, o T-BOND fez parte ainda da composição de algumas carteiras, tendo como finalidade a diversificação e minimização do desvio padrão dos retornos delas.

Tendo em vista que o objetivo do trabalho foi não só formar carteiras rentáveis, mas também seguras, além de incluir a variação dos títulos do governo americano, considerados ativos livres de risco, foi também incluído o ouro, que é outro ativo frequentemente utilizado para proteção de carteiras de investimentos. Nesse caso, diferentemente do T-BOND, que faz parte de algumas carteiras, o ouro fez parte meramente do modelo estatístico, ou seja, apesar de o ouro não ter sido alocado na carteira, a variação do preço dele foi utilizada em tal modelo.

Títulos do governo americano e ouro são descritos como ativos “porto-seguro”. Uma posição de longo prazo em títulos públicos do governo americano é vista pelos investidores como proteção para crises, possivelmente com base na visão de segurança que uma economia avançada como a americana demonstra. O ouro, seja físico ou por negociação eletrônica, é um dos ativos mais utilizados para proteção de capital, especialmente em tempos de incerteza e de risco (ALSHAMMARI; ALTARTURI; SAITI, 2020; HARVEY *et al.*, 2019; JONES; SACKLEY, 2016; NAWAZ; AZAM; ASLAM, 2020; SAMUEL, 2020).

Na Tabela 1, apresenta-se a correlação entre as variáveis independentes.

**Tabela 1 – Correlação entre os valores das principais variáveis independentes do trabalho referente ao período de 2008-2T2019**

	S&P 500	OZ1D	T-BOND 10
S&P 500	1		
OZ1D	-0.5120	1	
T-BOND 10	0.4423	-0.5937	1

**Fonte:** Dados extraídos da pesquisa.

Nota: OZ1d = Contrato de compra de ouro.

Nota-se, na Tabela 1, que no período *out of sample* há correlação negativa entre as taxas de retorno do ouro e do S&P 500. Esse resultado está dentro do esperado tendo em vista

que, quando ocorrem quedas no mercado de ações, é natural que as pessoas migrem para aplicações consideradas mais seguras. Observa-se ainda uma correlação fraca, mas positiva entre o S&P 500 e o T-BOND 10, o que denota que as variáveis, de forma parcial, caminham na mesma direção. Não obstante a fraca correlação, destaca-se que há períodos em que os investidores preferem o ouro em vez de os T-BONDS. Isso varia em razão da taxa de juro atual e expectativas futuras para ela, bem como em virtude das expectativas do mercado em relação ao ouro. Dessa maneira, verifica-se que, para o período apurado, a correlação entre ouro e T-BOND 10 é negativa. A tabela completa de correlações (Tabela 15) é apresentada no próximo capítulo.

Além das variáveis apresentadas anteriormente, que estão todas dentro do nível 1 do modelo hierárquico proposto, foram utilizadas também as seguintes variáveis: *i*, uma variável de nível 2 que representa a empresa, e NAICS, uma variável de nível 3 que representa o setor no qual a empresa está enquadrada. Desse modo, resume-se a seguir a lista de variáveis utilizadas na presente tese:

- 1) *Dummy (dummyret)* da taxa de retorno do trimestre posterior à divulgação dos dados.
- 2) LN<sup>1</sup> do Retorno do S&P 500 (Lnsp500), que é o índice formado por ações listadas nas bolsas NYSE e NASDAQ, de acordo com critérios de tamanho, liquidez e participação industrial. É um dos índices mais utilizados no mercado americano e no mundo (primeiro nível).
- 3) LN do Retorno das Ações (Lnretournousd) nos quatro últimos trimestres (primeiro nível).
- 4) LN do Retorno do T-BOND 10 (Lntbond) do trimestre de coleta de dados e quatro anteriores (primeiro nível).
- 5) LN do Retorno OURO (Lnnoz) do trimestre de coleta de dados e quatro trimestres anteriores (primeiro nível).
- 6) NAICS representando o setor no qual a empresa está incluída (terceiro nível).
- 7) *i* representando a empresa (segundo nível).
- 8) *t* representando o trimestre de divulgação das informações (primeiro nível).

Utilizando-se das variáveis apresentadas, a equação do modelo hierárquico pode ser escrita da seguinte forma:

---

<sup>1</sup> Logaritmo Natural.

$$\begin{aligned}
 & \text{dummyret}_{tij} \\
 & = \frac{1}{1 + e^{-\left( \begin{array}{l} \pi_{0jk} + \sum_{a=1}^4 \pi_{ajk} \cdot \text{Lnretornousd}_{t-a} + \\ \sum_{b=0}^4 \tau_{bjk} \cdot \text{Lnsp500}_{t-b} + \\ \sum_{c=0}^4 \varphi_{cjk} \cdot \text{Lntbond}_{t-c} + \\ \sum_{d=0}^4 \mu_{dj k} \cdot \text{Lnoz}_{t-d} \end{array} \right)}} \\
 \text{Nível 1:} & \\
 \text{Nível 2:} & \quad \pi_{0jk} = b_{00k} + r_{0jk} \\
 \text{Nível 3:} & \quad b_{00k} = \gamma_{000} + \mu_{00k}
 \end{aligned}$$

em que  $i$  representa a empresa e  $j$  representa o setor.

Conforme mencionado anteriormente, as variáveis foram selecionadas com base nos estudos mencionados no Quadro 2. Em todas as variáveis de primeiro nível, foram utilizados o valor do trimestre de coleta de dados e o dos quatro trimestres anteriores. Pretendeu-se com isso verificar qual a influência das variáveis citadas, ou seja, quanto cada uma delas influenciava na probabilidade de retornos positivos futuros. A probabilidade para fins de formação das carteiras foi sempre analisada um passo à frente, isto é, divulgadas e analisadas as informações do 1T2018, em conjunto com as informações de 1T2017, 2T2017, 3T2017 e 4T2017, qual a probabilidade de a empresa “ $i$ ” obter retorno positivo no 2T2018. Foram selecionadas para a carteira todas as ações com probabilidade maior que 60% de retornos positivos, que fez parte do lado *long* (comprado), e todas as ações com probabilidade menor que 40% de retornos positivos, que fez parte do lado *short* (vendido). Ações com probabilidade dentro do intervalo de 40% a 60% foram descartadas. Assim, as ações de interesse para a formação da carteira foram as de probabilidade entre 0 e 39,9% para as ações *short*, e as probabilidades de 60,01% a 100% para as ações *long*.

A partir do processo de criação das carteiras, foram surgindo lacunas que precisaram ser exploradas em busca de resultados melhores. Com base no objetivo inicial, foram criadas carteiras *long and short* (LS) com e sem critérios de *stop loss*, bem como com e sem alocação de títulos de renda fixa.

Definidas as probabilidades, muitas vezes o modelo estatístico não atinge 100% de acerto na formação das carteiras. A definição de um critério de *stop loss* visa delimitar um nível máximo de perda por ação (VEZERIS; KARKANIS; KYRGOS, 2019; WANG *et al.*, 2020). No caso, foi definido que o *stop loss* seria acionado quando a ação perdesse 10% ou mais. Desse modo, a operação com a ação seria encerrada, e o dinheiro ficaria em caixa até a próxima alocação. Uma opção nesse caso para o “mundo real” seria alocar em renda fixa o



dinheiro disponível em caixa da operação encerrada até a alocação referente ao próximo trimestre.

Como a carteira é formada em ordem de divulgação dos resultados com um limite máximo de cem ações, houve um momento em sua criação em que não se pôde mais alocá-las, mesmo que surgissem empresas com alta probabilidade de retornos positivos. Por esse motivo, optou-se por criar dois sistemas de alocação: no primeiro não houve restrição de datas, e no segundo foram alocadas unicamente ações que divulgaram as informações após as datas de 31/03, 30/06, 30/09 e 31/12. Sendo assim, ações que publicaram fora desse “padrão”, que não é obrigatório, foram descartadas. Do primeiro sistema, surgiram três carteiras, do segundo sistema, mais três carteiras e, por fim, criou-se outras três carteiras utilizando exclusivamente ações listadas na B3, totalizando nove carteiras. Nesses três grupos de três carteiras, foram empregados os sistemas de alocação apresentados no Quadro 3.

**Quadro 3 – Grupos de carteiras**

<b>Carteiras sem restrição de data</b>
(1) carteira LS somente com ações (2) carteira LS com 2/3 de ações e 1/3 de T-BOND 10 (3) carteira LS com 2/3 de ações e 1/3 de LFT <sup>2</sup>
<b>Carteiras com restrição de data (31/03, 30/06, 30/09 e 31/12)</b>
(1) carteira LS somente com ações (2) carteira LS com 2/3 de ações e 1/3 de T-BOND 10 (3) carteira LS com 2/3 de ações e 1/3 de LFT
<b>Carteiras somente com ações listadas na B3</b>
(1) carteira LS somente com ações (2) carteira LS com 2/3 de ações e 1/3 de T-BOND 10 (3) carteira LS com 2/3 de ações e 1/3 de LFT

**Fonte:** Dados extraídos da pesquisa.

A alocação de títulos de renda fixa é utilizada como forma de diversificar e minimizar riscos na alocação de portfólios, principalmente nos portfólios *long only*.

Diferentemente de fundos conservadores, cujo principal foco é em produtos de renda fixa em vez de ações, o presente trabalho configura um investimento arrojado, tendo em vista que, no máximo, teve 33,3% de investimento em renda fixa, seguindo uma linha com grande foco no investimento em ações (LAU, 2019; YANG; QIAN; BELTON, 2019).

<sup>2</sup> Letras Financeiras do Tesouro.

O método adotado neste trabalho baseou-se ainda numa linha de diversificação internacional em que, ao mesmo tempo que se diversifica com diferentes empresas pertencentes a diversos setores da economia, se diversifica também com o investimento em empresas e títulos de renda fixa de diferentes países. Ao utilizarem como amostra os mercados de ações americano, alemão e indonésio, estudos mostraram que a diversificação internacional reduz significativamente o risco quando é comparada com um investimento isolado no mercado doméstico (PADMANTYO; PRASOJO, 2019). Como o mundo está se fechando numa grande comunidade global, tem se tornado comum a prática entre grandes investidores de distribuírem seus investimentos em diferentes países, como um esforço para aproveitar oportunidades onde quer que elas apareçam (AWA; NNAMETU; EMOH, 2019).

A carteira *long and short* é formada a partir de uma carteira *long* (comprada) e uma carteira *short* (vendida), a qual, desde que haja garantias, pode fornecer recursos para a compra da carteira *long*. Desse modo, a carteira LS é conhecida também como *zero-cost strategy* (BEAVER; MCNICHOLS; PRICE, 2016).

O método para alocação dos ativos nas carteiras segue o chamado modelo ingênuo, no qual não há diferenciação de pesos entre os ativos. De acordo com Demiguel, Garlappi e Uppal (2009) esse modelo é de fácil implementação, porque não depende da estimativa dos momentos de retornos das ações ou da otimização. Além disso, esses autores destacam que, apesar dos avanços nos modelos teóricos desenvolvidos nos últimos cinquenta anos e dos avanços nos métodos para estimar os parâmetros desses modelos, os investidores continuam a usar regras simples de alocação.

Tendo em vista que todas as ações têm o mesmo peso na carteira, o resultado de cada trimestre é uma média elaborada com base no retorno de todas as ações que compõem a carteira no trimestre. No fechamento do período total da carteira, será apresentado o desvio padrão baseando-se nos seguintes fatores: retornos trimestrais, retorno total (soma dos produtos de cada um dos trimestres), média trimestral e índice de Sharpe.

O índice de Sharpe é calculado dividindo-se a média do retorno excedente pelo desvio padrão da carteira. Entende-se por retorno excedente o retorno previsto (média) subtraído da taxa livre de risco. No entanto, não há um consenso no sentido de qual a melhor taxa para refletir a taxa brasileira livre de risco. Alguns autores defendem a utilização da taxa Selic, porém, no período utilizado neste trabalho, ela apresentou valores muito discrepantes em relação às taxas praticadas, por exemplo, no mercado americano, que também foi utilizado na amostra. Outros autores defendem que a taxa livre de risco brasileira deve ser baseada na americana, mas com alguns ajustes. Desse modo, conforme citado linhas acima, dada a falta

de consenso no sentido de qual a melhor referência de taxa livre de risco no mercado brasileiro, optou-se por aplicar o índice de Sharpe modificado considerando tão somente a média dos retornos dividida pelo desvio padrão, que também é uma das formas de avaliação utilizadas, ou seja, nesse sentido descarta-se a utilização da taxa livre de risco. Tal índice é um dos mais comumente utilizados na avaliação de retorno ajustado ao risco de portfólios (LEWIN; CAMPANI, 2020; PAI; ILANGO, 2020). No próximo capítulo, são apresentados os resultados da pesquisa.

## 4 RESULTADOS

O objetivo principal deste trabalho foi desenvolver um modelo de seleção de ações para compor portfólios *long and short* utilizando modelos hierárquicos não lineares de regressão para dados em painel. A partir da coleta dos dados e da implementação dos modelos econométricos, esta seção apresenta os resultados da pesquisa. Inicialmente, na Tabela 2, mostra-se a estatística descritiva das variáveis utilizadas.

**Tabela 2 – Estatística descritiva para as variáveis da amostra**

Var	dummyret	retorno	sp500	tbond10	oz1
N	150549	150549	150549	150549	150549
Mean	0,57	3,76	1,73	- 0,08	3,85
Sd	0,50	64,17	6,91	14,96	9,04
Min	0	- 99,97	- 32,46	- 33,54	- 14,02
Max	1,00	22.239,55	18,68	47,15	36,56

**Fonte:** *Output* do *software* Stata.

Nota: A linha N está descrita em unidades. As demais linhas estão em percentual.

Nota-se, na Tabela 2, que o tamanho total da amostra contempla 150.549 observações. A variável *dummyret* apresenta valor médio de 0,57, e isso significa que 57% das empresas da amostra apresentaram resultados positivos nos trimestres coletados. Quanto ao desvio padrão, o maior desvio é encontrado na variável *retorno*, com mínimo de -99% e máximo de mais de 22.000%. Esse desvio na variável já era esperado, pois ela engloba os retornos trimestrais de todas as empresas da amostra, ou seja, 25 anos mais 2 trimestres.

Optou-se por utilizar um modelo *in sample* com 13 anos (treino) abrangendo o período de 1T1995 a 4T2007 e um modelo *out of sample* de 11 anos e 2 trimestres (teste) referente ao período de 1T2008 a 2T2019. A seleção desse período de base foi por ele apresentar diferentes cenários de baixa e de alta. A opção de iniciar o modelo de teste, ou seja, fora da amostra, em um ano de crise foi justamente para testar o modelo num período de turbulência. Ressalta-se que a escolha dos períodos foi arbitrária e que todos os resultados publicados pelas empresas até poderiam estar dentro da amostra e apenas o próximo trimestre, que ainda é desconhecido, fora dela.

Uma das formas de desenvolver um modelo estatístico hierárquico é conhecida como *step-up strategy*. Essa metodologia inicia-se, basicamente, pela execução de um modelo nulo, no qual são inseridas apenas a variável dependente, que, no caso da presente tese, é a

dummyret, e as variáveis independentes que representam os níveis, no caso, a variável NAICS, que representa os setores, e a variável  $i$ , as empresas.

Nessa etapa, espera-se que o modelo seja estatisticamente significativo, inclusive os níveis (WEST; WELCH; GAŁECKI, 2014). Esse modelo é conhecido também como incondicional, ou seja, nenhuma variável preditora é especificada em nenhum dos níveis. Ele representa como a variação em uma medida de resultado é alocada por meio dos três diferentes níveis (tempo, empresa e setor) (RAUDENBUSH; BRYK, 2002).

Dessa forma, inicialmente é feita a estimação do modelo incondicional que permite verificar se existe variabilidade na taxa de retorno entre empresas provenientes de um mesmo setor e entre aquelas oriundas de setores distintos. Conforme mencionado anteriormente, nenhuma variável independente foi incluída na modelagem, sendo considerada apenas a existência de um intercepto e dos termos de erro de níveis 2 e 3  $u_{00k}$  e  $r_{0jk}$ , com variâncias respectivamente iguais a  $\tau_{u000}$  e  $\tau_{r000}$ , dado que os termos de erro independentes de nível 1 apresentam variância constante e igual a  $\pi^2/3$ . De acordo com Fávero e Belfiore (2017), o modelo apresenta a seguinte expressão:

$$\begin{aligned} \text{Modelo incondicional:} \quad p_{tij} &= \frac{1}{1 + e^{-(\pi_{0jk})}} & (13) \\ \pi_{0jk} &= b_{00k} + r_{0jk} \\ b_{00k} &= \gamma_{000} + u_{00k} \end{aligned}$$

resultando em:

$$p_{tij} = \frac{1}{1 + e^{-(\gamma_{000} + u_{00k} + r_{0jk})}}$$

Na equação  $\gamma_{000}$  representa o *log* de chances de retornos positivos entre as empresas em determinado trimestre enquadradas em determinado setor. O valor estimado para  $\gamma_{000}$  foi 0,3428 (s.e. = 0,01482), conforme mostrado na Tabela 3. Portanto, para uma ação com taxa de retorno “típica”, ou seja, para uma empresa com efeito aleatório  $r_{0jk} = 0$ , o valor esperado dos *logs* de chances é 0,3428, que corresponde a  $\exp\{-0,3428\} = 1,40$ . Utilizando as expressões apresentadas anteriormente, isso corresponde a uma probabilidade de  $1/(1+\exp\{-0,3428\}) = 0,5848$ . A probabilidade está muito próxima da média, que foi apresentada na Tabela 2. A Tabela 3 apresenta o modelo nulo referente ao período de 1T1995 a 4T2007.

**Tabela 3 – Estimação de modelo hierárquico não linear – modelo nulo – período 1995 a 2007**

Mixed-effects logistic regression		Number of obs = 61,187			
		Observations per Group			
Group Variable	No. Of groups	Minimum	Average	Maximum	
Naics	210	1	291,4	4.260	
I	2.022	1	30,3	52	
Integration method: mvaghermite		Integration pts. =		7	
Wald chi2(19) = .					
Log likelihood = - 44356.292		Prob > chi2 =		.	
Dummyret	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_cons	.3428344	.0148219	23.13	0.000	.313784 .3718848
naics var(_cons)	.0191299	.0037695			.0130012 .0281476
naics>i var(_cons)	9.85e-36	4.07e-20			
LR test vs. logistic model: chibar2(01) = 198.88 Prob >= chibar2 = 0.0000					

**Fonte:** Output do software Stata.

Pode-se observar, ainda na Tabela 3, que o modelo que compreende o período de 1995 a 2007 se mostrou estatisticamente significativo ao nível de 1%. Além da variável dependente *dummyret*, discutida anteriormente, o modelo inclui também as variáveis *i*, que representa as empresas no nível 2, e a variável *naics*, os setores no nível 3.

Apesar de o modelo não apresentar o nível de significância das variáveis representantes dos níveis, foi possível calcular com base nos valores dos coeficientes e do erro padrão. Sendo assim, nota-se que a variável *naics* é significativa, tendo em vista que  $0,0191299 / 0,0037695$  resulta em um *z* de 5,07, isto é, o valor está acima do nível crítico de 1,96. No que se refere à variável *i*, esta não apresentou significância estatística, visto que  $9.85e-36 / 4.07e-20$  resulta em  $2,420e-16$ , que está muito próximo de zero, ou seja, abaixo do valor crítico de 1,96. Dessa forma, faz sentido manter o *naics*, e não faria sentido manter o *i*. No entanto, como se trata de um modelo dinâmico, motivo pelo qual se pretende atualizá-lo em diferentes períodos, optou-se por manter a variável *i*, mesmo que ela, de acordo com a Tabela 3, não tenha apresentado significância estatística.

Após essa primeira etapa e havendo um resultado satisfatório quanto à significância estatística, foi executado o teste de correlação intraclasse, o qual mensura a proporção de variação total que é atribuída aos níveis (Tabela 4).

**Tabela 4 – Correlação intraclasse referente ao período de 1995 a 2007**

Intraclass correlation				
Level	ICC	Std. Err.	[95% Conf. Interval]	
Naics	.0057812	.0011326	.0039363	.0084833
ilnaics	.0057812	.0011326	.0039363	.0084833

Fonte: Output do software Stata.

Observa-se que, para o intervalo utilizado, não faz diferença incluir a variável  $i$  (empresa) como um dos níveis. Nota-se ainda que aproximadamente 0,5% da variância total dos termos de erro é devido à alteração do comportamento entre setores. À medida que novos dados são incluídos na amostra *in sample*, as estimações do modelo são alteradas. Na Tabela 5, é apresentada sua estimação utilizando toda a amostra. Apesar de essa estimação não ser usada no trabalho para fins de formação da carteira, optou-se por demonstrar por meio dela como a influência dos setores e das empresas se altera em decorrência da inserção de novas informações.

**Tabela 5 – Estimação de modelo hierárquico não linear – modelo nulo – período 1995 a 2019**

Mixed-effects logistic regression		Number of obs = 193,445		
		Observations per Group		
Group Variable	No. Of groups	Minimum	Average	Maximum
Naics	258	1	749,8	16.658
I	5.441	1	35,6	98

Integration method: mvaghermite      Integration pts. = 7

Wald chi2(19) = .

Log likelihood = -131797.82      Prob > chi2 = .

Dummyret	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
_cons	.2108266	.0127167	16.58	0.000	.1859023	.2357509
naics var(_cons)	.0194394	.0029926			.0130012	.0281476
naics>i var(_cons)	.0488316	.0034963			.0424381	.0561884

LR test vs. logistic model: chi2(2) = 1568.06      Prob >= chibar2 = 0.0000

Fonte: Output do software Stata.

Na Tabela 5, é possível observar que o valor estimado para  $\gamma_{000}$  foi 0,2108 (s.e. = 0,01271). Obviamente, com um período diferente esperava-se um resultado diferente. Para uma ação com taxa de retorno “típica”, isto é, para uma empresa com efeito aleatório  $r_{0jk} = 0$ , o valor esperado dos logs de chances é 0,2108, que corresponde a  $\exp\{-0,2108\} = 1,23$ .

Utilizando a expressão apresentada anteriormente (Fórmula 13), isso corresponde a uma probabilidade de  $1/(1+\exp\{-0,2108\}) = 0,5525$ . Nota-se que a probabilidade continua próxima da média apresentada na Tabela 5, no entanto o objetivo dessa nova estimação foi analisar o comportamento dos componentes de níveis 2 e 3 para a amostra total. Apresenta-se, na Tabela 6, a correlação intraclasse para a amostra completa.

**Tabela 6 – Correlação intraclasse referente ao período de 1995 a 2019**

<b>Intraclass correlation</b>				
<b>Level</b>	<b>ICC</b>	<b>Std. Err.</b>	<b>[95% Conf. Interval]</b>	
naics	.0057887	.0008861	.0042872	.0078119
i naics	.02033	.001328	.0178839	.0231028

**Fonte:** Output do software Stata.

Nesse caso, nota-se que a correlação intraclasse passa a ter maior relevância tanto ao nível de empresa quanto ao nível de setor e de empresa juntos. O percentual que era de aproximadamente 0,5% no modelo anterior, o qual considerou o período de 1995 a 2007, passa a ser de aproximadamente 2% para toda a amostra, ou seja, 2% da variância total dos termos de erro é devido à variação das empresas e dos setores.

Desse modo, fica clara a importância de ambos os níveis nos dias atuais, isto é, tanto o terceiro nível naics (setor) quanto o segundo nível i (empresa). Por questões acadêmicas, em razão das quais houve a necessidade de, a partir de um período fechado (*in sample* ou período de treino), se avaliar outro (*out of sample* ou período de teste) que teoricamente não se teria conhecimento das informações, utilizou-se o primeiro modelo de base com período de 1995 a 2007. No entanto, para fins práticos, o sistema pode ser reavaliado trimestre a trimestre com a inserção de novas informações, as quais, dessa maneira, gerarão novos coeficientes com provável melhoria no nível de predição.

Se o modelo for consideravelmente diferente de zero pela presença de, ao menos, um termo de erro significativo, são inseridas as demais variáveis independentes em todos os níveis, caso haja. O modelo pode ser utilizado normalmente, se algum dos níveis não apresentar variáveis independentes.

A seguir, mediante a Tabela 7, apresenta-se a estimação completa, chamada também de modelo condicional, para o período de 1995 a 2007. Assim, como os estudos apresentados no capítulo anterior, o qual trata dos procedimentos metodológicos, diversas variáveis foram testadas (Quadro 2). Após testar as mesmas variáveis por meio do modelo hierárquico, observou-se que as que trouxeram maior nível de assertividade para o período



estudado foram a variável de logaritmo natural do retorno da empresa defasada de um a quatro trimestres, bem como os logaritmos naturais dos retornos dentro do próprio trimestre defasados de um a quatro trimestres das variáveis S&P 500, T-BOND 10 e ouro. Mais detalhes sobre cada uma delas foi abordado no capítulo anterior.

**Tabela 7 – Estimação de modelo hierárquico não linear – modelo condicional - período 1995 a 2007**

<b>Mixed-effects logistic regression</b>		<b>Number of obs = 55.368</b>				
		Observations per Group				
Group Variable	No. Of groups	Minimum	Average	Maximum		
Naics	204	1	271,4	3.924		
I	1.917	1	28,9	48		
Integration method: mvaghermite		Integration pts. =		7		
Wald chi2(19) = 4498,30						
Log likelihood = -34997,958		Prob > chi2 =		0.0000		
<b>Dummyret</b>	<b>Coef.</b>	<b>Std. Err.</b>	<b>z</b>	<b>P&gt; z </b>	<b>[95% Conf. Interval]</b>	
<b>Lnreturnousd</b>						
L1.	-.184808	.0424034	-4.36	0.000	-.2679171	-.1016988
L2.	.2295716	.0424372	5.41	0.000	.1463963	.3127469
L3.	.2127382	.0420402	5.06	0.000	.130341	.2951354
L4.	.1217288	.0417141	2.92	0.004	.0399707	.2034869
<b>Insp500</b>						
-.	8.28161	.1791861	46.22	0.000	7.930412	8.632809
L1.	.1729139	.1722256	1.00	0.315	-.1646421	.51047
L2.	-2.449974	.1711438	-14.32	0.000	-2.78541	-2.114538
L3.	-1.433052	.175657	-8.16	0.000	-1.777334	-1.088771
L4.	-3.411227	.1818427	-18.76	0.000	-3.767632	-3.054822
<b>Lntbond</b>						
-.	.7204212	.1328889	5.42	0.000	.4599637	.9808787
L1.	1.457355	.1237192	11.78	0.000	1.21487	1.69984
L2.	-.0174138	.1371261	-0.13	0.899	-.286176	.2513485
L3.	.9153792	.1388195	6.59	0.000	.6432979	1.18746
L4.	.4768438	.1231562	3.87	0.000	.235462	.7182256
<b>Lnoz</b>						
-.	-.8847249	.118141	-7.49	0.000	-1.116277	-.6531728
L1.	-1.94452	.1164278	-16.70	0.000	-2.172714	-1.716326
L2.	.0509849	.1237788	0.41	0.680	-.191617	.2935868
L3.	-.6711394	.1216871	-5.52	0.000	-.9096418	-.4326371
L4.	1.752966	.1235832	14.18	0.000	1.510747	1.995185

(continua)

(conclusão)

Dummyret	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
_cons	.4703961	.0231489	20.32	0.000	.4250252 .5157671
naics var(_cons)	.0269288	.0051632	5.21		.0184932 .0392122
naics>i var(_cons)	7.87e-35	1.96e-19	4.015e-16	.	.
LR test vs. logistic model: chibar2(01) = 228.89 Prob >= chibar2 = 0.0000					

Fonte: Output do software Stata.

Tendo em vista o modelo apresentado na Tabela 7 e utilizando-se da equação proposta na metodologia, pode-se reescrevê-la aplicando os coeficientes da seguinte maneira:

$$dummyret_{tij} = \frac{1}{1 + e^{-r_{ojk} + \mu_{ook}}} \left( \begin{array}{l} 0,4703961 - 0,184808.(1/4)Lnretornousd + 0,2295716.(2/4)Lnretornousd \\ 0,2127382.(3/4)Lnretornousd + 0,1217288.(4/4)Lnretornousd \\ 8,28161.(0/4)lnsp500 + 0,1729139.(1/4)lnsp500 - 2,449974.(2/4)lnsp500 - \\ 1,433052.(3/4)lnsp500 + 3,411227.(4/4)lnsp500 + 0,7204212.(0/4)Lntbond + \\ 1,457355.(1/4)Lntbond + -0,0174138.(2/4)Lntbond + 0,9153792.(3/4)Lntbond + \\ 0,4768438.(4/4)Lntbond - 0,8847249.(0/4)Lnoz - 1,94452.(1/4)Lnoz + \\ 0,0509849.(2/4)Lnoz - 0,6711394.(3/4)Lnoz + 1,752966.(4/4)Lnoz + \end{array} \right)$$

Os resultados da estimação mostram que o modelo é estatisticamente significativo ao nível de 1% (Prob > chi2 = 0.0000). O modelo englobou no período 204 setores e 1.917 empresas, com 55.368 observações. No que diz respeito às variáveis e seus coeficientes, praticamente todas as variáveis, atuais e defasadas, foram estatisticamente significantes, com exceção do L1 do S&P 500, L2 do T-BOND 10 e L2 da variável lnoz, que representa a variação do valor do ouro. Apesar de o protocolo pedir para excluir essas variáveis insignificantes do modelo, optou-se por mantê-las no modelo de base, pois, à medida que novos dados fossem incluídos, todas elas se tornariam significantes.

O resultado apresentado na última linha da Tabela 7 (LR test vs. logistic model: chibar2(01) = 228.89 Prob > = chibar2 = 0.0000) mostra que o modelo é mais preciso e significativo quando comparado ao modelo logístico tradicional, no qual não há a diferenciação de níveis. Todavia, somente o nível setor foi estatisticamente significativo. Da mesma forma que as variáveis não significantes apresentadas na estimativa, optou-se por também manter o nível i (empresas) pelo mesmo motivo destacado anteriormente, ou seja, à medida que novos dados fossem incluídos, todas elas se tornariam significantes. Esse resultado para o “i” era esperado, visto que, conforme verificado por meio da classificação intraclasse, ele não mostrou poder de incremento no modelo.

Nesse ponto, há de se destacar que, no intervalo pré-crise, percebe-se que as diferenças entre setores se mostram significantes nos resultados positivos das empresas, contudo as diferenças entre empresas são estatisticamente insignificantes. Assim, considerando a amostra e o período utilizados, pode-se inferir que, no momento pré-crise, poder-se-ia apenas escolher um setor para serem obtidos resultados similares nas taxas de retornos, não sendo necessário a opção por essa ou por aquela empresa.

Após a inserção de todas as variáveis e a execução do modelo condicional, é importante verificar novamente a correlação intraclasse, com objetivo de visualizar o comportamento das variáveis dos níveis 2 e 3. Sendo assim, apresenta-se a Tabela 8, na qual é demonstrada a correlação intraclasse para os níveis naics e i.

**Tabela 8 – Correlação intraclasse para o modelo condicional de 1995 e 2007**

<b>Intraclass correlation</b>				
<b>Level</b>	<b>ICC</b>	<b>Std. Err.</b>	<b>[95% Conf. Interval]</b>	
naics	.0081189	.001544	.0055898	.0117787
i naics	.0081189	.001544	.0055898	.0117787

**Fonte:** *Output do software Stata.*

Visualiza-se por meio da Tabela 8 que a correlação intraclasse mostra um leve incremento com a inclusão de todas as variáveis, não obstante a variável i (empresa) continua sem fazer diferença.

A seguir, são apresentadas as curvas *Receiver Operating Characteristic (ROC)* para o período dentro da amostra (1995 a 2007), fora da amostra (1T2008 a 2T2019) e período total dos dados (1995 a 2T2019). É mostrada ainda a curva ROC do modelo *logit*, que não é exposto, mas foi estimado para fins de comparação entre as curvas (Tabela 9).

**Tabela 9 – Curvas ROC do modelo *logit* versus modelo hierárquico nos diferentes cenários da amostra**

<b>1995 - 2007</b>	<b>Obs</b>	<b>ROC Area</b>	<b>Std. Err.</b>	<b>Asymptotic Normal [95% Conf. Interval]</b>	
ylogit_is	55,368	0.6761	0.0023	0.67163	0.68062
ymelogit_is	55,368	0.6767	0.0023	0.67222	0.68117
Ho: area(ylogit_is) = area(ymelogit_is)					
chi2(1) = 0.68 Prob>chi2 = 0.4098					

<b>2008 - 2019t2</b>	<b>Obs</b>	<b>ROC Area</b>	<b>Std. Err.</b>	<b>Asymptotic Normal [95% Conf. Interval]</b>	
ylogit_oos	95,181	0.6328	0.0018	0.62927	0.63630
ymelogit_oos	95,181	0.6475	0.0018	0.64405	0.65101
Ho: area(ylogit_oos) = area(ymelogit_oos)					
chi2(1) = 502.01 Prob>chi2 = 0.0000					

<b>1995 - 2019t2</b>	<b>Obs</b>	<b>ROC Area</b>	<b>Std. Err.</b>	<b>Asymptotic Normal [95% Conf. Interval]</b>	
ylogit_full	150,549	0.6602	0.0014	0.65742	0.66292
ymelogit_full	150,549	0.6876	0.0014	0.68496	0.69030
Ho: area(ylogit_full) = area(ymelogit_full)					
chi2(1) = 2036.32 Prob>chi2 = 0.0000					

**Fonte:** Output do software Stata.

Na Tabela 9, tem-se o objetivo de não só mostrar a curva ROC do modelo em relação aos períodos *in sample*, *out of sample* e geral, como também demonstrar que o modelo hierárquico apresenta melhor desempenho que o modelo *logit*.

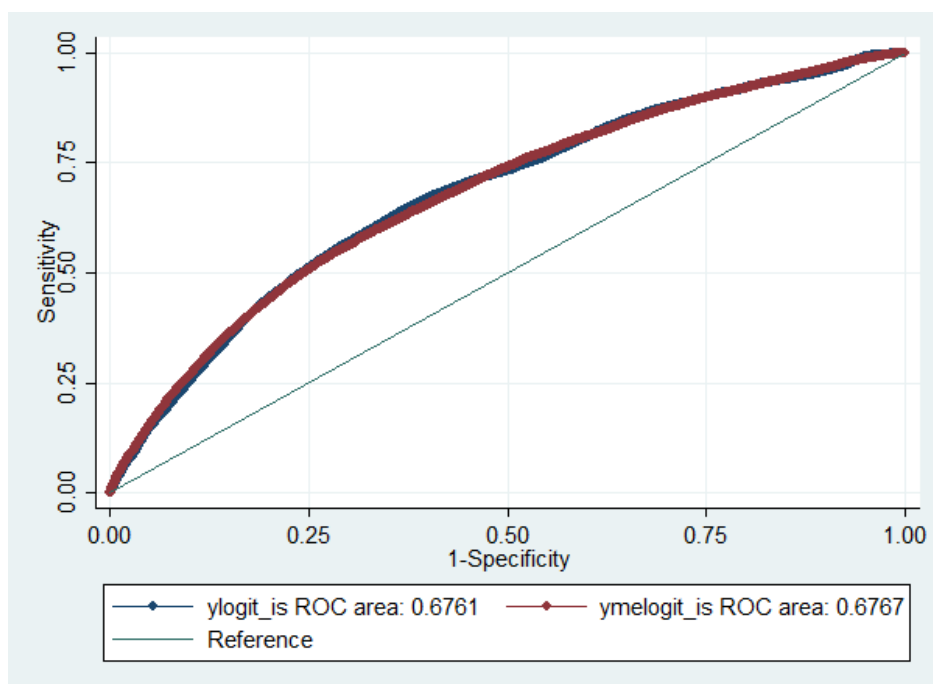
O grande desafio de modelos preditivos é alcançar o mesmo nível de acerto *out of sample* que se alcançou *in sample*. Constatou-se que os valores exibidos referentes ao intervalo de 1995 a 2007 são muito parecidos entre o modelo *logit* e o hierárquico, sendo 67,61 para o primeiro e 67,67 para o segundo. Apesar de mostrar uma pequena superioridade, a diferença não é estatisticamente significativa com  $\chi^2(1) = 0.68$ . No que se refere ao período de 2008 a 2T2019, observa-se novamente superioridade para o modelo hierárquico, mas dessa vez com diferença estatisticamente significativa. O modelo *logit* apresenta curva ROC de 63,28 contra 64,75 com  $\chi^2(1) = 502.01$ . Apesar disso, percebe-se que as curvas ROC têm valores menores que os apresentados pelas curvas *in sample*, o que já era esperado, pois o modelo *in sample*, em geral, mostra melhores curvas. Para o período inteiro, considerando toda a amostra *in sample*, novamente a curva ROC gerada pelo modelo hierárquico apresenta superioridade, com valor de 68,76, em comparação à curva formada por meio do modelo *logit*, com valor de 66,02. Destaca-se ainda o valor de  $\chi^2(1) = 2036.32$ . Além disso, verifica-se que, quando se compara o período de base com o período completo da

amostra, visualiza-se uma melhora na curva ROC, e isso sugere que, com o passar do tempo e com o acúmulo de informações, o modelo pode se tornar mais preciso (PAI; ILANGO, 2020).

Aparentemente, pode-se cair na tentação de dizer que a variação de 63,28 para 64,75 é pequena, embora, em outras palavras, seja possível dizer que, no período fora da amostra, o modelo *logit* foi preciso na previsão de 63,28% das empresas, enquanto o modelo hierárquico acertou 64,75%. Em números, tem-se que o modelo *logit* acertou na previsão de  $95.181 * 63,28\% = 60.230$ , enquanto o modelo hierárquico acertou na previsão de  $95.181 * 64,75\% = 61.630$ . Assim dizendo, o modelo hierárquico acertou na previsão de 1.400 observações a mais, o que é um número bastante relevante para a construção de uma carteira de investimentos.

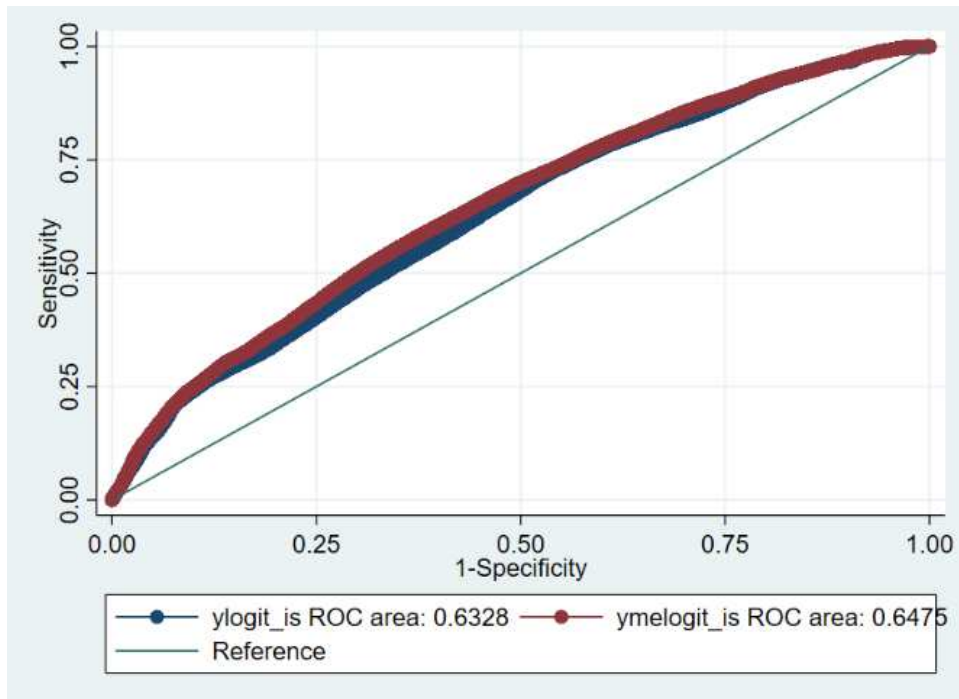
Nos Gráficos 1 a 3, ilustra-se as curvas ROC na mesma ordem que são apresentadas na Tabela 9.

**Gráfico 1 – Curvas ROC modelos *logit* (azul) versus modelo hierárquico (vermelho) *in sample***



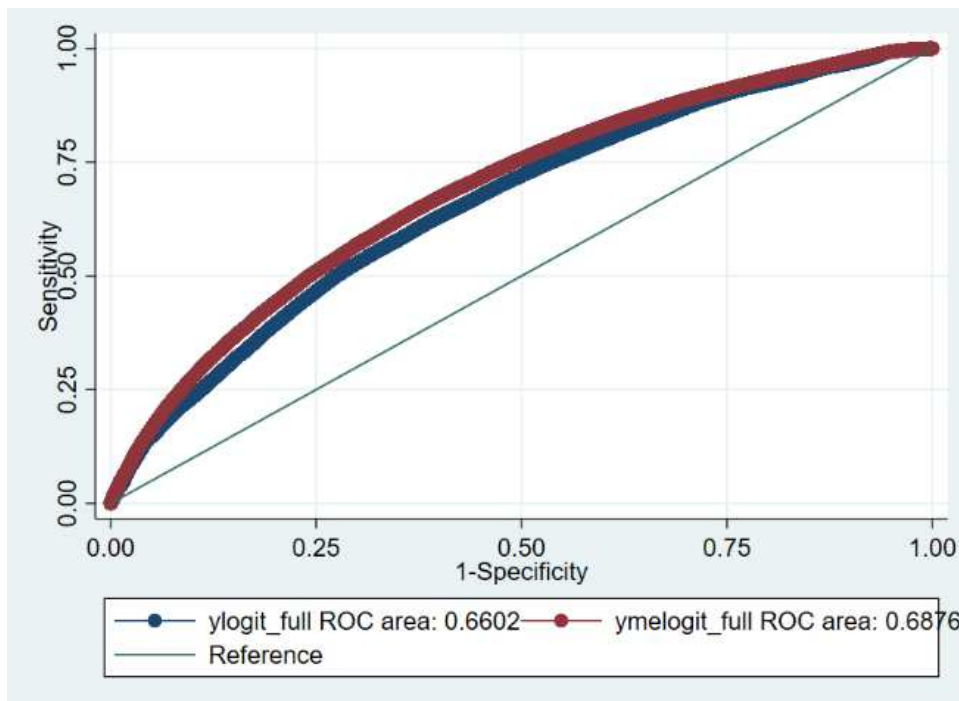
**Fonte:** Output do software Stata.

Gráfico 2 – Curvas ROC modelos *logit* (azul) versus modelo hierárquico (vermelho) *out of sample*



Fonte: Output do software Stata.

Gráfico 3 – Curvas ROC modelos *logit* (azul) versus modelo hierárquico (vermelho) completo



Fonte: Output do software Stata.

As curvas ROC apresentadas nos Gráficos 1 a 3 confirmam o que os números mostraram. A primeira delas, referente ao período de 1995 a 2007, fica quase que sobreposta,

visto que a diferença entre elas é insignificante. Nas outras duas, observa-se pequenas, mas estatisticamente significantes diferenças.

Partindo do princípio que o modelo hierárquico é superior para a construção das carteiras, as Tabelas 10 e 11 apresentam os níveis de acerto do modelo estimado. Repara-se, conforme esperado, que à medida que o *cutoff*<sup>3</sup> é incrementado, o nível de acerto também o é. Isso significa que, quanto maior a probabilidade exigida, maior é o nível de acerto. Desse modo, a partir de diferentes *cutoffs*, esperou-se aumentar também o nível de acerto e retornos das carteiras.

Por se tratar de uma carteira *long and short*, é desejável que o sistema seja preciso tanto na previsão do que será evento quanto na previsão do que não o será. Assim, considerando diferentes *cutoffs*, a Tabela 10 apresenta o poder de assertividade do modelo. Para o presente estudo, foi definido *cutoff* maior que 60% para a carteira *long* e menor que 40% para a carteira *short*.

**Tabela 10 – Tabela com os níveis de acerto *out of sample* para cutoffs 60%, 70%, 80% e 90%: seleção de ações para a carteira long**

Predicted cutoff 60%				Predicted cutoff 80%			
	dummyret2		Total		dummyret2		Total
1	0	1		1	0	1	
	14.302	27.851	42.153		1.698	6.431	8.129
	33,93	66,07	100		20,89	79,11	100

Predicted cutoff 70%				Predicted cutoff 90%			
	dummyret2		Total		dummyret2		Total
1	0	1		1	0	1	
	4.489	13.829	18.618		255	932	1.187
	25,72	74,28	100		21,48	78,52	100

**Fonte:** *Output* do software Stata.

Nota: Resultados em percentual.

Por intermédio da Tabela 10, testou-se diferentes níveis de *cutoffs* para verificar o comportamento dos níveis de acerto. Conforme era esperado, à medida que se aumenta o nível de corte da probabilidade prevista, tornam-se maiores os níveis de acerto. No *cutoff* de 60%, por exemplo, de 42.153 observações classificadas como evento, 14.302 (33,93%) estavam erradas e 27.851 estavam corretas (66,07%). No *cutoff* de 90%, por exemplo, identifica-se que o volume de observações diminuiu substancialmente, sendo que das 1.187 observações classificadas como evento, 255 (21,48%) estavam erradas e 932 (78,52% estavam certas). Constatou-se ainda que o nível de 80% de *cutoff*, além de promover um resultado discretamente melhor que o nível de 90%, aumentou significativamente o número

<sup>3</sup> Um ponto ou nível que é um limite designado de algo.

de observações. Cabe a cada investidor definir qual o melhor *cutoff* para atender à sua necessidade na formação de carteiras. *Cutoffs* mais baixos demonstram menores níveis de acerto, porém com uma disponibilidade de empresas maior. Por outro lado, *cutoffs* mais altos melhoram o nível de acerto, mas diminuem as quantidades de empresas. Analisando-se sob outra perspectiva, *cutoffs* mais baixos podem aumentar a gama de oportunidades de investimentos, porém podem aumentar o *turnover* das carteiras, o que incrementa seu custo. Em contrapartida, *cutoffs* mais altos aumentam o nível de acerto, mas diminuem o número de empresas disponíveis, e isso pode, eventualmente, comprometer a diversificação.

Apresenta-se, a seguir, a Tabela 11, a qual o faz a mesma classificação, porém diminuindo o *cutoff* para buscar os não eventos, que são as observações de interesse para o lado *short* da carteira. Nesse caso, foram testados os *cutoffs* de 40%, 30%, 20% e 10%.

**Tabela 11 – Tabela com os níveis de acerto para *cutoffs* 40%, 30%, 20% e 10%: seleção de ações para as carteiras *short***

Predicted cutoff 40%	dummyret2		Total	Predicted cutoff 20%	dummyret2		Total
	0	1			0	1	
0	6.045	2.865	8.910	0	1.474	189	1.663
	67,85	32,15	100		88,63	11,37	100

Predicted cutoff 30%	dummyret2		Total	Predicted cutoff 10%	dummyret2		Total
	0	1			0	1	
0	3.526	1.225	4.751	0	1.336	133	1.469
	<b>74,22</b>	<b>25,78</b>	<b>100</b>		<b>90,95</b>	<b>9,05</b>	<b>100</b>

**Fonte:** Output do software Stata.

Nota: Resultados em percentual.

Nota-se por meio da Tabela 11 que o comportamento está dentro do esperado e similar ao encontrado na Tabela 10. No *cutoff* de 40%, que é o nível utilizado no presente trabalho para seleção de ações, por exemplo, foram previstas 8.910 observações com não evento, ou seja, que teriam retornos iguais ou menores que zero. Destas, 6.045 (67,85%) estavam corretas e 2.865 (32,15%) estavam erradas. No *cutoff* de 10%, no qual se aumentou inversamente a exigência, conforme esperado, o nível de acerto também aumentou. Nesse nível de *cutoff*, foram selecionadas 1.469 observações, das quais 1.336 (90,95%) estavam corretas e 133 (9,05%) estavam erradas.

Na Tabela 12, apresenta-se os principais indicadores dos mercados brasileiro e americano, sendo que este último é inclusive utilizado como referência mundial. Na ilustração, são mostradas as taxas de retornos trimestrais apresentando ao final dela o desvio



padrão, soma dos produtos considerando aplicações trimestrais, média e índice Sharpe, no período de 2008 a 2019.

**Tabela 12 – Principais indicadores dos mercados brasileiro e americano**

<b>t</b>	<b>S&amp;P 500</b>	<b>WTICO</b>	<b>DJIA</b>	<b>NASDAQ</b>	<b>OZ1D</b>	<b>IBOV</b>	<b>LFT</b>	<b>T-BOND 10</b>
2008q1	-9%		-11%	-3%	-2%	-21%	6%	1%
2008q2	-32%	-50%	-27%	-38%	-14%	-54%	-27%	-4%
2008q3	-9%	-13%	-10%	1%	21%	13%	3%	-28%
2008q4	11%	14%	9%	10%	5%	34%	13%	16%
2009q1	14%	19%	13%	18%	1%	31%	16%	14%
2009q2	9%	13%	10%	9%	25%	23%	9%	-4%
2009q3	7%	9%	6%	11%	0%	5%	0%	12%
2009q4	-3%	-14%	-2%	-2%	6%	-11%	0%	-10%
2010q1	-5%	5%	-3%	-7%	2%	6%	4%	-22%
2010q2	11%	13%	9%	16%	20%	9%	6%	3%
2010q3	11%	25%	10%	10%	-3%	3%	10%	25%
2010q4	1%	-7%	2%	2%	2%	-8%	1%	-8%
2011q1	-12%	-14%	-11%	-11%	13%	-15%	4%	-30%
2011q2	6%	15%	7%	6%	1%	0%	-6%	-9%
2011q3	13%	5%	9%	16%	-7%	7%	1%	8%
2011q4	-6%	-9%	-4%	-6%	-4%	-20%	-8%	-20%
2012q1	6%	-1%	4%	4%	3%	2%	0%	-2%
2012q2	-3%	-8%	-5%	-6%	8%	-5%	0%	-8%
2012q3	16%	13%	16%	15%	-8%	3%	5%	18%
2012q4	6%	-3%	5%	6%	-12%	-3%	0%	4%
2013q1	2%	14%	0%	6%	-5%	-19%	-11%	40%
2013q2	6%	-12%	4%	8%	-5%	4%	2%	-1%
2013q3	5%	8%	4%	6%	1%	-3%	7%	1%
2013q4	0%	0%	0%	-3%	-1%	9%	3%	-8%
2014q1	5%	-6%	2%	9%	1%	1%	0%	-4%
2014q2	4%	-21%	6%	5%	-10%	-19%	-11%	-3%
2014q3	1%	-37%	1%	5%	1%	-20%	-15%	-17%
2014q4	3%	25%	3%	3%	2%	20%	9%	11%
2015q1	-1%	-29%	-4%	0%	-9%	-29%	-11%	3%
2015q2	-3%	-4%	-1%	-2%	-2%	-10%	-6%	4%
2015q3	2%	-9%	3%	-1%	16%	15%	12%	-21%
2015q4	-1%	25%	-1%	-3%	1%	5%	3%	-4%
2016q1	7%	-4%	6%	11%	5%	25%	15%	-12%
2016q2	-1%	-3%	2%	0%	-10%	-6%	-5%	47%
2016q3	9%	17%	10%	13%	1%	18%	14%	8%
2016q4	2%	-3%	2%	4%	0%	8%	3%	-2%
2017q1	3%	-3%	5%	3%	3%	-3%	0%	-6%
2017q2	5%	17%	6%	6%	1%	1%	-1%	9%
2017q3	2%	17%	3%	5%	4%	19%	1%	15%

(continua)

(conclusão)								
t	S&P 500	WTICO	DJIA	NASDAQ	OZ1D	IBOV	LFT	T-BOND 10
2017q4	3%	10%	3%	4%	-2%	-10%	-9%	12%
2018q1	5%	-6%	3%	7%	-9%	-13%	-4%	-6%
2018q2	-5%	-16%	-1%	-9%	1%	12%	4%	8%
2018q3	5%	7%	3%	8%	7%	8%	0%	-23%
2018q4	1%	3%	-1%	1%	1%	-7%	-2%	-1%
2019q1	0%	-11%	-1%	-1%	27%	9%	1%	-34%
2019q2	9%	3%	9%	9%	-7%	2%	-3%	15%
SD	8%	16%	7%	9%	9%	16%	8%	16%
RET	118%	-51%	114%	235%	73%	-41%	20%	-53%
MÉDIA	2%	0%	2%	3%	2%	0%	1%	0%
SHARPE	26%	-1%	27%	34%	18%	2%	9%	-2%

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio Padrão, RET = Retorno Acumulado, SHARPE = Média/SD. Retornos em dólar.

Conforme os dados apresentados na Tabela 12, observa-se que os menores desvios padrões são vistos no índice S&P 500 e na LFT. O resultado desta está dentro do esperado, visto que é a taxa brasileira básica de juros. No entanto, não se esperava que o T-BOND 10 apresentasse desvio padrão de 16%, que ficou no mesmo patamar da variação do petróleo e do Ibov. Notou-se ainda que o ouro (OZ1D), conhecido como um ativo muito utilizado para proteção de carteiras, não apresentou os melhores resultados, sendo 9% de desvio padrão e 73% de retorno acumulado no período. Vale ressaltar que, nesse caso, foi avaliado o retorno do ouro isoladamente, ou seja, sem considerá-lo inserido em uma carteira. Obviamente ao inseri-lo, o resultado não seria diferente, no entanto a combinação de diferentes ativos em diferentes momentos econômicos proporciona, conforme largamente estudado em finanças, resultados positivos advindos da diversificação (ALSHAMMARI; ALTARTURI; SAITI, 2020; HARVEY *et al.*, 2019; JONES; SACKLEY, 2016; KOUSAR; BATOOL, 2019; NAWAZ; AZAM; ASLAM, 2020; SAMUEL, 2020). Apesar de muitos estudos confirmarem o papel do ouro como um ativo de proteção do capital, há ainda pesquisadores que não encontraram os mesmos resultados, entre eles, Maghyreh e Abdoh (2020), segundo os quais o ouro não age como um diversificador ou ferramenta de proteção em médio e longo prazo para portfólios islâmicos. Como qualquer ativo, o preço do ouro está relacionado diretamente à demanda, assim o preço sobe quando há mais demanda e vice-versa. Nesse sentido, em momentos de crise, o ouro apresenta maior demanda e conseqüentemente melhores preços e retornos sobre o investimento. Portanto, sugere-se que, para apurar o comportamento válido do ouro em portfólios de investimento, sejam utilizados períodos que contemplem incertezas e instabilidades econômicas. No contexto do presente

estudo, o ouro não estará inserido na carteira como um ativo, mas sua variação foi considerada, no modelo estatístico, no papel de variável independente.

O petróleo, que é o bem de maior demanda no mercado internacional, fechou com valor negativo no período, além de se destacar com uma das maiores volatilidades entre os indicadores apurados. É importante frisar que o petróleo é uma *commodity* e que muitos fatores influenciam seu valor, por exemplo, oferta e demanda, descoberta de novos poços, guerras, acordos internacionais, cotas definidas pela Organização dos Países Exportadores de Petróleo (OPEP), entre outros (MONDAY; ABDULKADIR, 2020). O preço dele definitivamente influencia as economias de maneira geral e também os mercados de capitais, bem como desperta interesse de formuladores de políticas e investidores, como resultado de seu papel imprescindível no desenvolvimento econômico de todos os países que o importam ou o exportam (ALSHAMMARI; ALTARTURI; SAITI, 2020; MOREMA; BONGA-BONGA, 2020; WU *et al.*, 2020). Não obstante ser utilizado também com objetivo de diversificação em carteiras de investimento, conforme estudado por Gong, Bu e Chen (2020), demonstrando bons resultados em períodos de crise, o petróleo foi utilizado meramente como variável explicativa no presente estudo.

Nota-se que o maior índice de Sharpe, e, dessa forma, o maior retorno ajustado ao risco, foi apresentado pelo índice NASDAQ, com desvio padrão de 9%, retorno acumulado de 235%, média trimestral de 3% e Sharpe de 34%. O pior indicador foi o T-BOND 10 americano de 10 anos, seguido pelo índice Bovespa. Destaca-se que, quando considerado em reais, o Ibovespa trouxe resultado positivo no período; apesar disso, quando dolarizado, apresentou resultado negativo.

Salienta-se que todos os indicadores apresentados na Tabela 12 possuem ETFs ou fundos que os refletem, e, por conseguinte, é possível investir neles isoladamente e obter os resultados apresentados. Os ETFs, que existem, nos EUA, desde 1993 e, na Europa, desde 1999, tipicamente rastreiam um índice e são uma alternativa para investidores avessos a riscos. Destaca-se ainda que eles são altamente líquidos e permitem aos investidores uma rápida exposição ao índice subjacente (OZDURAK; ULUSOY, 2020). Além disso, uma das principais vantagens dos ETFs como meio de investimento são os baixos custos de transações (POTERBA; SHOVEN, 2002).

Como as datas utilizadas nas alocações de ações nas carteiras dependem das datas de publicação dos resultados, não se pôde utilizar essas mesmas datas na Tabela 12, que simula a alocação de recursos nos indicadores utilizados no trabalho. Não obstante isso, foi possível adotar datas aproximadas de acordo com as datas limites de publicação dos

resultados baseadas no mercado brasileiro. Dessa forma, definiu-se as datas e os períodos de apuração apresentados na Tabela 13.

**Tabela 13 – Exemplos da apuração dos retornos referentes aos índices utilizados**

<b>Data de demonstração</b>	<b>Data máxima de publicação</b>	<b>Período de apuração do retorno</b>
31/03/2019	15/05/2019	entre 16/05/2019 e 14/08/2019
30/06/2019	14/08/2019	entre 15/08/2019 e 14/11/2019
30/09/2019	14/11/2019	entre 15/11/2019 e 31/03/2019
31/12/2019	31/03/2020	entre 01/04/2020 e 15/05/2020

**Fonte:** Elaborada pelo autor.

Talvez, ao ver esses indicadores, um investidor possa pensar em formar carteiras compostas somente por ETFs deles. Alguns estudiosos defendem que investir em ETFs é uma maneira mais segura e menos trabalhosa de fazer investimento, além de reduzir custos operacionais e transacionais. Contudo, conforme apresentado na Tabela 14, uma carteira formada exclusivamente com esses indicadores não traria um resultado muito interessante para investidores mais arrojados.

**Tabela 14 – Carteira formada somente com ETFs com pesos iguais para cada ETF na carteira**

<b>t</b>	<b>Resultado Todos</b>	<b>Resultado Brasil</b>	<b>Resultado EUA</b>
2008q1	-5,6%	-4,4%	-3,9%
2008q2	-30,7%	-36,2%	-27,6%
2008q3	-2,9%	5,9%	-6,4%
2008q4	13,8%	16,3%	10,7%
2009q1	15,6%	16,6%	13,1%
2009q2	11,7%	17,2%	10,4%
2009q3	6,1%	3,5%	7,4%
2009q4	-4,5%	-4,8%	-4,3%
2010q1	-2,5%	4,3%	-5,0%
2010q2	10,7%	11,9%	11,8%
2010q3	11,4%	8,8%	13,1%
2010q4	-1,8%	-3,0%	-1,3%
2011q1	-9,4%	-3,0%	-10,8%
2011q2	2,6%	2,6%	4,5%
2011q3	6,6%	1,6%	7,5%
2011q4	-9,7%	-10,3%	-8,2%
2012q1	2,1%	1,2%	2,4%
2012q2	-3,3%	-1,1%	-3,6%
2012q3	9,6%	3,1%	11,5%
2012q4	0,3%	-4,5%	0,8%
2013q1	3,3%	-5,2%	9,4%

(continua)

(conclusão)

t	Resultado Todos	Resultado Brasil	Resultado EUA
2013q2	0,7%	-2,8%	0,0%
2013q3	3,6%	3,4%	4,1%
2013q4	0,1%	3,0%	-1,9%
2014q1	1,0%	-0,8%	1,1%
2014q2	-6,1%	-15,2%	-3,1%
2014q3	-10,0%	-17,6%	-7,6%
2014q4	9,5%	14,1%	7,8%
2015q1	-10,1%	-19,5%	-6,8%
2015q2	-3,2%	-5,7%	-1,7%
2015q3	2,0%	8,5%	-1,9%
2015q4	3,2%	8,7%	2,9%
2016q1	6,6%	10,3%	2,3%
2016q2	3,0%	-5,9%	5,9%
2016q3	11,3%	12,6%	9,7%
2016q4	1,6%	2,0%	0,3%
2017q1	0,1%	-0,8%	0,7%
2017q2	5,5%	4,4%	7,3%
2017q3	8,3%	10,2%	7,6%
2017q4	1,3%	-2,7%	4,9%
2018q1	-2,9%	-7,8%	-1,1%
2018q2	-0,8%	0,1%	-3,8%
2018q3	2,0%	5,5%	1,4%
2018q4	-0,7%	-1,2%	0,5%
2019q1	-1,1%	6,7%	-3,2%
2019q2	4,6%	-1,3%	6,3%
SD	7,9%	9,9%	7,5%
RET	45,4%	4,2%	63,7%
MÉDIA	1,1%	0,6%	1,4%
SHARPE	14,6%	6,4%	18,3%

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD. Retornos em dólar.

Para formação da carteira Todos, naturalmente se utilizou todos os indicadores da Tabela 14. Para a formação da carteira Brasil, utilizou-se ouro, petróleo, Ibov e LFT. Para a formação da carteira EUA, utilizou-se ouro, petróleo, S&P 500, NASDAQ, DJI e T-BOND 10. Conforme se pode observar nessa ilustração, o melhor resultado foi gerado com base nos indicadores americanos, seguido pela combinação de todos os indicadores e, por fim, pelos indicadores brasileiros (dolarizados). Na Tabela 15, apresenta-se a correlação entre as taxas de retorno acumuladas dos indicadores.

**Tabela 15 – Correlação entre as taxas de retorno trimestrais dos indicadores**

	sp500	wti	dji	nasdaq	oz1d	ibov	lft	tbond10
sp500	1							
wti	0.7188	1						
dji	0.9738	0.7688	1					
nasdaq	0.9578	0.6233	0.8900	1				
oz1d	-0.5552	-0.5392	-0.5923	-0.4446	1			
ibov	0.6909	0.7017	0.7059	0.6565	-0.3041	1		
lft	-0.0013	-0.0836	0.0119	0.0577	0.1316	0.0710	1	
tbond10	0.4506	0.5939	0.5393	0.3349	-0.5910	0.2863	0.0064	1

**Fonte:** Output do software Stata.

Todos os indicadores apresentados na Tabela 15 foram utilizados de alguma forma no trabalho, seja para fins de comparação as taxas de retorno com os retornos gerados pelas carteiras, seja como integrante do modelo econométrico como variável independente. Nesse sentido, entende-se ser importante demonstrar as correlações existentes entre eles.

Os principais índices de ações americanos (S&P 500, DJI e NASDAQ) apresentam forte correlação entre si. Isso decorre do fato de existirem muitas ações em comum entre eles, motivo pelo qual caminham na mesma direção. No que se refere ao principal índice de ações brasileiro (IBOV), este demonstrou correlação positiva com os índices americanos e com a variação do preço do petróleo (WTI), bem como correlação negativa, mas fraca, com o ouro (OZ1D). É possível observar que as letras financeiras do tesouro (LFT) não apresentaram correlação significativa com nenhum dos indicadores. O T-BOND 10 mostrou correlação positiva com o petróleo e com o DJI e negativa com o ouro.

O objetivo principal deste estudo, como mencionado anteriormente, foi criar carteiras de investimento *long and short* a partir de modelos hierárquicos não lineares para dados em painel. Seguindo uma tendência recente de diversificação internacional optou-se por utilizar quatro bolsas de valores sendo elas: a Bolsa de Valores de São Paulo (B3) e três bolsas do mercado americano (NYSE, NASDAQ e NYSE EURONEXT).

Conforme explorado no capítulo anterior, as carteiras *long and short* são formadas por um lado comprado e por um lado vendido, na maioria das vezes com valores iguais de investimento. A Tabela 16 apresenta o lado comprado (*long*) da carteira formada com ações americanas e brasileiras. A quantidade de ações de cada carteira foi influenciada por dois fatores: (1) a probabilidade de retornos positivos no trimestre posterior com base nos dados divulgados; (2) a data da divulgação dos resultados, limitando-se ao máximo de cem empresas no lado *long* e cem empresas no lado *short*, ou seja, foram incluídas nas carteiras as primeiras duzentas empresas que publicaram suas demonstrações e apresentaram probabilidade superior a 60% de retorno positivo no próximo trimestre.

**Tabela 16 – Lado long da carteira**

<b>t</b>	<b>nível de acerto</b>	<b>ret s/ stop</b>	<b>ret c/ stop</b>	<b>quant.</b>	<b>nível previsto</b>
2008q1	38%	-6%	0%	24	61%
2008q3	25%	1%	22%	4	62%
2008q4	50%	4%	10%	100	79%
2009q1	90%	48%	48%	100	91%
2009q2	83%	19%	19%	100	86%
2009q3	62%	4%	6%	100	73%
2010q2	77%	9%	9%	100	62%
2010q3	92%	26%	26%	100	81%
2010q4	91%	7%	7%	100	62%
2011q3	76%	9%	10%	100	79%
2012q1	65%	3%	3%	100	63%
2012q3	51%	1%	3%	100	71%
2012q4	66%	6%	7%	100	73%
2013q1	74%	8%	8%	100	64%
2013q2	70%	9%	10%	100	81%
2013q3	69%	7%	7%	100	63%
2013q4	67%	5%	6%	100	62%
2014q1	61%	1%	3%	100	62%
2014q2	53%	1%	2%	100	65%
2014q3	0%	-19%	-10%	1	60%
2015q3	35%	-5%	-2%	100	61%
2015q4	84%	14%	15%	100	62%
2016q1	71%	11%	12%	100	64%
2016q2	46%	1%	2%	100	63%
2016q3	56%	6%	7%	100	81%
2016q4	43%	-5%	1%	100	65%
2017q1	67%	9%	11%	100	66%
2017q2	49%	0%	3%	100	67%
2017q3	65%	7%	8%	100	62%
2017q4	53%	2%	4%	100	62%
2018q1	55%	4%	6%	100	63%
2018q2	0%	-67%	-10%	1	71%
2018q3	16%	-13%	-5%	100	68%
2018q4	60%	16%	17%	5	61%
2019q1	100%	10%	10%	1	74%
2019q2	74%	5%	5%	100	62%
SD		14%	9%		
RET		106%	1196%		
MÉDIA		3%	6%		
SHARPE		21%	65%		

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD. Retornos em dólar.

Ao analisar a Tabela 16, nota-se que nem todos os meses apresentam valores. Nesse caso, não foram localizadas empresas com probabilidade de retornos positivos igual ou maior que 60%, motivo pelo qual não se alocou nada nesses trimestres. A segunda coluna mostra o nível de acerto que significa o percentual de ações que foram incluídas na carteira e trouxeram resultados positivos. Um dos piores resultados, por exemplo, foi o 3T2008 no qual não mais que 25% das quatro empresas incluídas na carteira apresentaram resultado positivo, ou seja, apenas uma delas. Ainda no contexto do nível de acerto e ainda analisando o 3T2008, observa-se, na última coluna, que o nível médio de acerto previsto para essas empresas era de 62%. Com melhores resultados em termos de acerto, pode-se destacar os trimestres 1T2009, 3T2010 e 4T2010, com 90% ou mais. Na terceira e quarta colunas, são apresentados os retornos das carteiras no trimestre. A terceira apresenta os retornos sem critério de *stop loss*, e a quarta carteira mostra critério de *stop loss* limitando as perdas em 10%. A quinta coluna apresenta a quantidade de ações alocadas no lado *long* da carteira, sendo que o máximo foi definido como cem empresas. No rodapé da referida tabela, apresenta-se os resultados consolidados. Percebe-se que o desvio padrão das carteiras sem *stop loss* foi de 14%, enquanto as carteiras com *stop loss* foi de 9%. Isso significa que o objetivo do *stop loss* foi alcançado minimizando as perdas da carteira e consequentemente a variação e o risco dos retornos. Se fosse considerado unicamente esse lado como uma carteira (*long only*), ela teria apresentado retorno acumulado no período de 106% sem delimitação de *stop loss* e 1.196% com a delimitação. Por fim, é apresentado o índice Sharpe, que representa o retorno ajustado ao risco, melhor dizendo, qual o retorno aferido para cada unidade de risco assumida. No caso, a carteira apresentou 21% sem *stop* e 65% com *stop*. Ressalta-se que a Tabela 16 representa o lado *long* da carteira, e que a carteira consolidada depende ainda do lado *short*, o qual é apresentado na Tabela 17.

**Tabela 17 – Lado *short* da carteira**

<b>t</b>	<b>nível de acerto</b>	<b>ret s/ <i>stop</i></b>	<b>ret c/ <i>stop</i></b>	<b>quant.</b>	<b>nível previsto</b>
2008q1	35%	-4%	2%	100	62%
2008q2	59%	4%	7%	100	97%
2008q3	73%	14%	18%	90	73%
2009q4	28%	-10%	-3%	100	63%
2010q1	61%	3%	5%	100	76%
2010q4	50%	-24%	1%	2	64%
2011q1	46%	-1%	2%	100	75%
2011q2	57%	5%	8%	100	63%
2011q4	14%	-16%	-6%	100	70%

(continua)



(conclusão)

<b>t</b>	<b>nível de acerto</b>	<b>ret s/ stop</b>	<b>ret c/ stop</b>	<b>quant.</b>	<b>nível previsto</b>
2012q1	67%	11%	17%	3	61%
2012q2	38%	-2%	1%	100	63%
2013q4	33%	1%	6%	3	62%
2014q3	41%	-10%	3%	100	64%
2014q4	50%	1%	4%	100	63%
2015q1	66%	8%	12%	100	63%
2015q2	70%	10%	14%	100	63%
2015q3	77%	24%	30%	31	62%
2015q4	33%	-27%	0%	15	62%
2016q1	78%	-12%	28%	9	63%
2016q2	70%	15%	20%	22	63%
2016q4	100%	43%	43%	1	60%
2017q1	100%	44%	44%	2	62%
2017q3	43%	-18%	17%	16	62%
2017q4	75%	10%	15%	4	63%
2018q1	50%	15%	23%	2	61%
2018q2	64%	8%	11%	100	63%
2018q3	0%	-59%	-10%	2	71%
2018q4	59%	4%	7%	100	62%
2019q1	67%	13%	15%	100	63%
2019q2	76%	10%	19%	34	62%
SD		16%	12%		
RET		-6%	2291%		
MÉDIA		1%	8%		
SHARPE		8%	66%		

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD. Retornos em dólar.

De acordo com o exposto no capítulo anterior, para o lado *short* da carteira, teve-se a expectativa de que as ações das empresas selecionadas iriam desvalorizar no próximo trimestre. Em outros termos, a partir dos resultados publicados, que geralmente são ruins, o modelo econométrico vai mostrar qual a probabilidade de retorno positivo para o próximo trimestre. Nesse caso, é necessária uma adaptação para que os modelos mostrem a probabilidade de retorno negativo que é 1 menos a probabilidade de retorno positivo. Tanto com delimitação de *stop* quanto sem delimitação, os desvios padrão do lado *short* se mostraram mais altos. Nesse mesmo contexto, entretanto, os retornos acumulados se mostraram bastante diferentes, com retorno negativo de 6% para a carteira sem *stop* e 2.291% para a carteira com *stop*. Consequentemente tanto as médias trimestrais quanto o índice de Sharpe também foram mais altos para a carteira com *stop*.

Na Tabela 18, expõe-se a consolidação dos lados *long* e *short* da carteira, os quais, conforme vastamente documentado em trabalhos científicos, visam, entre outras coisas, reduzir riscos. Como apresentado anteriormente, além dos lados *long* e *short*, utilizou-se também, na carteira, a aplicação em um índice de renda fixa com objetivo de diversificar e minimizar o risco. Desse modo, simulou-se dois cenários: um em que o *long and short* alocou também o investimento em T-BONDS de dez anos e outro no qual o *long and short* alocou um investimento em LFT.

**Tabela 18 – Resultados consolidados da carteira *long and short***

t	<i>Long and short sem stop</i>			<i>Long and short com stop</i>		
	LS s/índice	c/ BOND	c/ LFT	LS s/índice	c/ BOND	c/ LFT
2008q1	-5%	-3%	-1%	1%	1%	2%
2008q2	2%	0%	-7%	3%	1%	-7%
2008q3	7%	-4%	6%	20%	4%	14%
2008q4	2%	7%	6%	5%	9%	8%
2009q1	24%	21%	21%	24%	21%	21%
2009q2	9%	5%	9%	10%	5%	9%
2009q3	2%	5%	1%	3%	6%	2%
2009q4	-5%	-7%	-3%	-2%	-4%	-1%
2010q1	2%	-6%	2%	2%	-6%	3%
2010q2	5%	4%	5%	5%	4%	5%
2010q3	13%	17%	12%	13%	17%	12%
2010q4	-8%	-8%	-5%	4%	0%	3%
2011q1	-1%	-10%	1%	1%	-9%	2%
2011q2	2%	-1%	0%	4%	0%	1%
2011q3	5%	6%	3%	5%	6%	4%
2011q4	-8%	-12%	-8%	-3%	-9%	-5%
2012q1	7%	4%	5%	10%	6%	7%
2012q2	-1%	-3%	-1%	1%	-2%	0%
2012q3	1%	6%	2%	2%	7%	3%
2012q4	3%	3%	2%	4%	4%	3%
2013q1	4%	16%	-1%	4%	16%	-1%
2013q2	5%	3%	4%	5%	3%	4%
2013q3	3%	3%	4%	4%	3%	5%
2013q4	3%	-1%	3%	6%	1%	5%
2014q1	1%	-1%	1%	1%	0%	1%
2014q2	0%	-1%	-3%	1%	0%	-3%
2014q3	-14%	-15%	-15%	-3%	-8%	-7%
2014q4	1%	4%	3%	2%	5%	4%
2015q1	4%	4%	-1%	6%	5%	0%
2015q2	5%	4%	1%	7%	6%	3%
2015q3	10%	-1%	10%	14%	2%	14%

(continua)

(conclusão)

t	<i>Long and short sem stop</i>			<i>Long and short com stop</i>		
	LS s/índice	c/ BOND	c/ LFT	LS s/índice	c/ BOND	c/ LFT
2015q4	-6%	-6%	-3%	7%	3%	6%
2016q1	-1%	-4%	4%	20%	9%	18%
2016q2	8%	21%	4%	11%	23%	6%
2016q3	3%	5%	7%	4%	5%	7%
2016q4	19%	12%	14%	22%	14%	16%
2017q1	27%	16%	18%	28%	16%	18%
2017q2	0%	3%	0%	2%	4%	1%
2017q3	-5%	1%	-3%	12%	13%	9%
2017q4	6%	8%	1%	10%	11%	3%
2018q1	10%	5%	5%	15%	8%	8%
2018q2	-29%	-17%	-18%	1%	3%	2%
2018q3	-36%	-32%	-24%	-7%	-12%	-5%
2018q4	10%	6%	6%	12%	8%	7%
2019q1	11%	-4%	8%	13%	-3%	9%
2019q2	7%	10%	4%	12%	13%	7%
SD	11%	10%	8%	7%	8%	6%
RET	104%	48%	87%	1855%	574%	719%
MÉDIA	2%	1%	2%	7%	5%	5%
SHARPE	21%	14%	21%	94%	60%	78%
PERC. POS.	74%	59%	65%	91%	76%	85%

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD., PERC. POS = percentual de trimestres com retorno positivo. Retornos em dólar.

Na Tabela 18, mostra-se o resultado das carteiras *long and short* em diferentes cenários: carteira *long and short sem stop loss*; carteira *long and short sem stop loss* com alocação de 1/3 em BONDS do tesouro americano de 10 anos; carteira *long and short sem stop loss* com alocação de 1/3 em LFT brasileiro. Essa mesma simulação é feita também com as carteiras com *stop loss*. Considerando o desvio padrão, a carteira que apresentou o menor risco foi a carteira com *stop loss* e diversificação por meio do título brasileiro – tal como destacado anteriormente, todos os valores são considerados em dólares. No entanto, quando se observa qual a melhor relação risco *versus* retorno, nota-se que o melhor índice de Sharpe (94%) foi encontrado na carteira *long and short com stop loss*, mas sem a diversificação de renda fixa. Essa carteira apresenta o segundo menor desvio padrão (7%) e o maior retorno acumulado (1.855%). Observa-se ainda que a carteira foi também a que apresentou maior frequência de resultados positivos (PERC. POS.), com 91% dos trimestres positivos ao longo do período estudado.

Na Tabela 19, criada a partir da Tabela 18, mostra-se a trajetória trimestre a trimestre dos principais indicadores e também a da carteira que apresentou a melhor rentabilidade na Tabela 18.

**Tabela 19 – Resultados trimestrais acumulados para os indicadores e para a carteira LS**

<b>t</b>	<b>LS</b>	<b>S&amp;P 500</b>	<b>WTICO</b>	<b>DJIA</b>	<b>NASDAQ</b>	<b>OZ1D</b>	<b>IBOV</b>	<b>LFT</b>	<b>T-BOND 10</b>
2008q2	4%	-39%	-50%	-35%	-40%	-16%	-64%	-23%	-2%
2008q3	25%	-44%	-57%	-41%	-40%	2%	-59%	-20%	-30%
2008q4	31%	-38%	-51%	-36%	-34%	7%	-45%	-10%	-19%
2009q1	63%	-29%	-41%	-28%	-22%	8%	-28%	4%	-7%
2009q2	78%	-23%	-34%	-21%	-14%	34%	-12%	13%	-11%
2009q3	84%	-18%	-27%	-16%	-5%	35%	-8%	13%	0%
2009q4	81%	-20%	-38%	-18%	-7%	42%	-18%	13%	-10%
2010q1	85%	-24%	-34%	-21%	-14%	45%	-13%	18%	-30%
2010q2	94%	-16%	-26%	-14%	-1%	74%	-5%	24%	-28%
2010q3	119%	-7%	-8%	-5%	10%	70%	-2%	37%	-10%
2010q4	129%	-6%	-14%	-3%	12%	72%	-10%	38%	-17%
2011q1	131%	-17%	-26%	-13%	-1%	94%	-23%	44%	-42%
2011q2	140%	-12%	-15%	-7%	5%	97%	-23%	35%	-47%
2011q3	152%	-1%	-10%	2%	22%	84%	-18%	36%	-42%
2011q4	144%	-7%	-18%	-3%	14%	76%	-35%	26%	-54%
2012q1	169%	-1%	-19%	1%	19%	82%	-33%	26%	-55%
2012q2	171%	-5%	-25%	-3%	12%	96%	-37%	26%	-59%
2012q3	175%	10%	-15%	12%	29%	81%	-35%	32%	-51%
2012q4	185%	17%	-18%	18%	37%	60%	-37%	33%	-50%
2013q1	197%	18%	-7%	18%	45%	52%	-49%	19%	-29%
2013q2	213%	26%	-19%	22%	57%	45%	-47%	21%	-30%
2013q3	224%	32%	-12%	27%	66%	47%	-48%	29%	-29%
2013q4	243%	31%	-12%	27%	61%	46%	-44%	33%	-35%
2014q1	247%	37%	-17%	29%	76%	48%	-43%	33%	-38%
2014q2	251%	43%	-34%	36%	85%	33%	-54%	19%	-40%
2014q3	239%	45%	-59%	37%	93%	35%	-63%	1%	-50%
2014q4	246%	49%	-48%	41%	99%	38%	-56%	10%	-44%
2015q1	267%	47%	-63%	35%	99%	26%	-68%	-2%	-43%
2015q2	293%	42%	-65%	33%	94%	22%	-72%	-8%	-41%
2015q3	348%	45%	-68%	36%	92%	42%	-67%	4%	-53%
2015q4	380%	44%	-60%	35%	86%	44%	-66%	7%	-56%
2016q1	476%	53%	-61%	43%	107%	52%	-57%	22%	-61%
2016q2	539%	52%	-62%	45%	106%	37%	-60%	16%	-42%
2016q3	562%	66%	-56%	59%	133%	39%	-52%	32%	-38%
2016q4	707%	69%	-58%	61%	143%	39%	-49%	37%	-39%
2017q1	932%	73%	-59%	69%	150%	43%	-50%	36%	-43%
2017q2	949%	81%	-52%	80%	166%	45%	-50%	35%	-38%
2017q3	1078%	86%	-44%	86%	179%	50%	-40%	37%	-29%

(continua)

(conclusão)

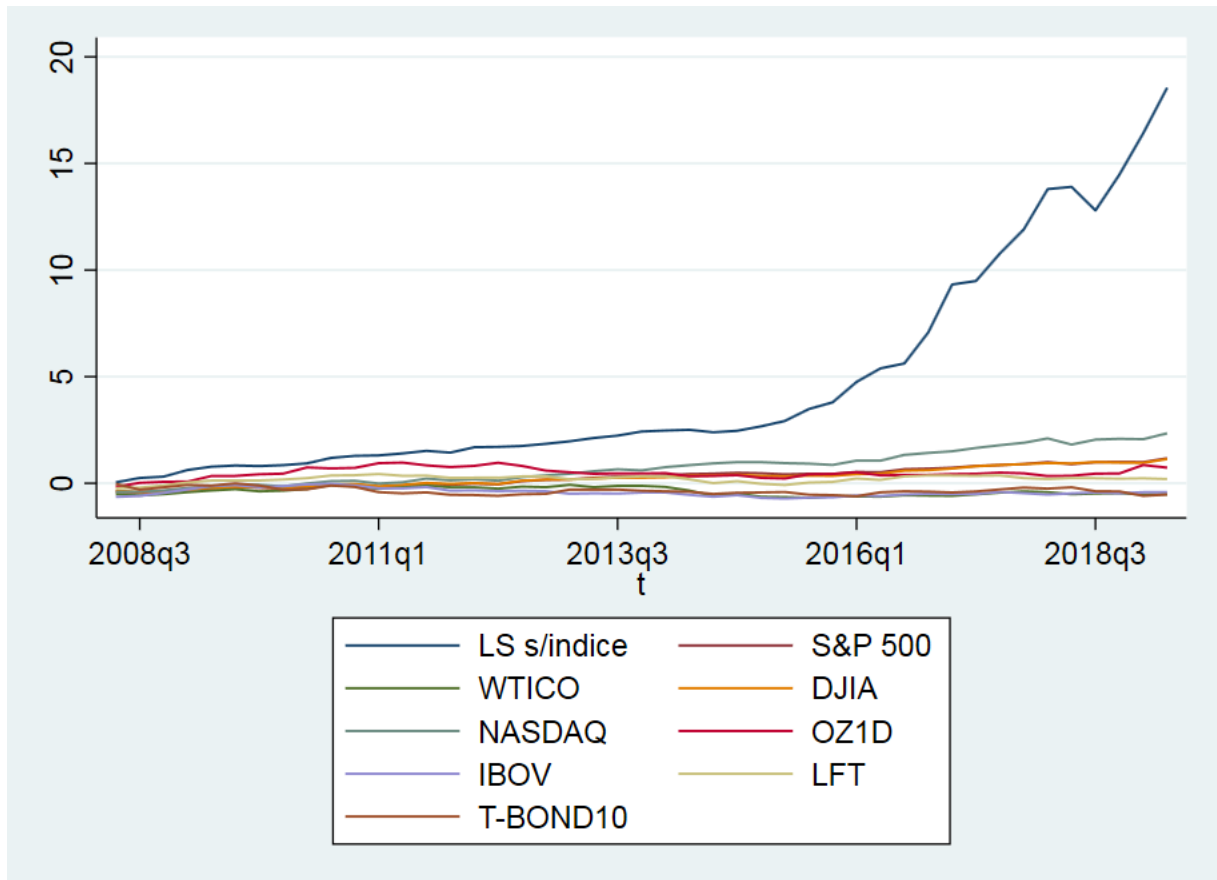
t	LS	S&P 500	WTICO	DJIA	NASDAQ	OZ1D	IBOV	LFT	T-BOND 10
2017q4	1191%	90%	-38%	90%	190%	47%	-46%	25%	-20%
2018q1	1380%	99%	-42%	95%	211%	34%	-53%	20%	-25%
2018q2	1390%	90%	-51%	93%	182%	35%	-47%	25%	-19%
2018q3	1280%	99%	-48%	100%	205%	45%	-43%	24%	-37%
2018q4	1447%	100%	-46%	97%	209%	46%	-47%	22%	-38%
2019q1	1642%	100%	-52%	96%	207%	86%	-42%	23%	-59%
2019q2	1855%	118%	-51%	114%	235%	73%	-41%	20%	-53%

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Observa-se, na Tabela 19, que, enquanto todos os indicadores apresentam resultados negativos no 2T2008 em decorrência da crise do *subprime*, a carteira LS apresenta resultado positivo em 4% e permanece subindo de maneira discreta, mas superior às demais, até que, a partir de 2016, se distancia substancialmente de todas as outras, o que pode ser visualizado no Gráfico 4. Cabe mencionar que a partir de 2016 os mercados entraram numa tendência de alta que se estendeu até o final da construção das carteiras, em 2T2019. Observa-se ainda que a carteira teórica do Ibovespa fechou o segundo trimestre de 2019 com resultado acumulado negativo. Isso aconteceu principalmente por ela ter sido computada em dólares, assim como todos os outros indicadores para fins de padronização. Quando considerado o valor em reais para o índice Bovespa, o valor acumulado no período é positivo.

Ilustra-se, no Gráfico 4, a trajetória acumulada da carteira mais rentável (*long and short* com *stop* sem índice – linha azul) comparada com os principais indicadores utilizados no mercado.

**Gráfico 4 – Retorno acumulado trimestre a trimestre da carteira LS com *stop s/índice1* e demais indicadores**



**Fonte:** Output do software Stata.

O Gráfico 4 permite visualizar, conforme apresentado na Tabela 19, que todos os indicadores seguiram, no período, uma linha muito parecida. É possível visualizar, ainda, a superioridade da carteira *long and short* com *stop* sem índice quando comparada com os principais indicadores de mercado.

A seguir, na Tabela 20, apresenta-se o teste de diferença de médias, com objetivo de comprovar se as diferenças médias são estatisticamente significantes ou se podem ser um comportamento do acaso.

**Tabela 20 – Two-sample Wilcoxon rank-sum (Mann-Whitney) test: carteira LS1 versus indicadores**

<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice	46	2451	2139
S&P 500	46	1827	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice}) = \text{ret}(i==\text{S\&P 500})$			
z=		2,436	
Prob >  z  =		0,0148	
Exact Prob =		0,0145	
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice	46	2438	2139
WTICO	46	1840	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice}) = \text{ret}(i==\text{WTICO})$			
z=		2,335	
Prob >  z  =		0,0196	
Exact Prob =		0,0193	
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice	46	2490	2139
DJIA	46	1788	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice}) = \text{ret}(i==\text{DJIA})$			
z=	-	2,741	
Prob >  z  =		0,0148	
Exact Prob =		0,0145	
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice	46	2328	2139
NASDAQ	46	1950	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice}) = \text{ret}(i==\text{NASDAQ})$			
z=		1,476	
Prob >  z  =		0,14	
Exact Prob =		0,1416	
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice	46	2608	2139
OZ1D	46	1670	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice}) = \text{ret}(i==\text{OZ1D})$			
z=		2,436	
Prob >  z  =		0,0148	
Exact Prob =		0,0145	

(continua)

(conclusão)			
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice	46	2420	2139
IBOV	46	1858	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice}) = \text{ret}(i==\text{IBOV})$			
z=	-	2,194	
Prob >  z  =		0,0282	
Exact Prob =		0,028	
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice	46	2600	2139
LFT	46	1678	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice}) = \text{ret}(i==\text{LFT})$			
z=	-	3,600	
Prob >  z  =		0,0003	
Exact Prob =		0,0003	
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice	46	2528	2139
T-BOND10	46	1750	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice}) = \text{ret}(i==\text{T-BOND10})$			
z=		3,038	
Prob >  z  =		0,0024	
Exact Prob =		0,0022	

**Fonte:** Output do software Stata.

De acordo com os dados expostos na Tabela 20, a carteira LS com *stop* e sem índice foi estatisticamente diferente e superior em relação aos indicadores apresentados, com exceção do índice NASDAQ.

No que se refere à participação setorial das empresas nas carteiras, apresenta-se, na Tabela 21, os vinte setores com maior participação. Como a classificação NAICS tem mais de 270 setores, optou-se, por questões de espaço, mostrar não mais que os vinte primeiros. Se uma empresa participou, por exemplo, em dois trimestres, ela é contada duas vezes, e assim sucessivamente.



**Tabela 21 – Classificação setorial NAICS com os vinte setores mais representativos na carteira sem restrição de data**

<b>Cod. Setor</b>	<b>Quant.</b>	<b>Perc.</b>
10 Indústria de remédios	637	14%
25 Outros fundos	331	7%
20 Loja de roupas	248	5%
6 Editoras de <i>software</i>	173	4%
13 Outras lojas de mercadorias variadas	135	3%
7 Indústria de equipamentos de comunicação	131	3%
23 Indústria de semicondutores e outros componentes eletrônicos	124	3%
11 Geração, transmissão e distribuição de energia elétrica	109	2%
134 Loja de departamentos	100	2%
123 Outros tipos de loja	84	2%
52 Loja de roupas e acessórios de vestir	83	2%
1 Comércio varejista	73	2%
2 Indústria de equipamentos de navegação, medição, controle e eletromédicos	68	1%
18 Indústria de equipamentos aeroespaciais	62	1%
35 Outras indústrias	60	1%
116 Distribuição de gás natural	60	1%
22 Extração de petróleo e gás	55	1%
31 Indústria de ferramentas de metal não motorizadas e cutelaria	54	1%
97 Indústria de roupas de tecido	54	1%
126 Transmissão por TV e rádio	52	1%

**Fonte:** Elaborada pelo autor.

Observa-se, na Tabela 21, que o setor mais representativo na carteira ao longo do período utilizado foi o de indústria de remédios, aparecendo 637 vezes e representando 14% do total de setores. Vale observar que, cada vez que determinada empresa entra na carteira, ela conta uma vez. Assim, por exemplo, se uma empresa aparecer durante dez trimestres, ela será contada dez vezes.

A quantidade de empresas de cada bolsa de valores alocadas nas carteiras é apresentada na Tabela 22.

**Tabela 22 – Quantidade de ações por bolsa**

<b>Bolsa</b>	<b>Quant.</b>	<b>Perc.</b>
NYSE	2.579	54,04%
NASDAQ	2.066	43,29%
B3	28	0,59%
NYSE Alternext U.S.	99	2,07%
Total	4.772	

**Fonte:** Elaborada pelo autor.

Conforme apresentado na Tabela 22, observa-se que a maior origem das ações alocadas nas carteiras está nas bolsas americanas, em especial NYSE e NASDAQ, as quais concentram 97,33% das ações. A B3, em virtude de sua pequena quantidade de empresas, é a que apresenta menor representatividade nas carteiras.

Objetivando testar a robustez do sistema com a alocação de diferentes empresas e, ao mesmo tempo, na tentativa de melhorar a representatividade das empresas brasileiras na carteira, foi proposta a seguinte alteração na alocação das ações: as ações que antes eram alocadas de acordo com a data de divulgação dos resultados continuaram sendo alocadas dessa forma, porém com a restrição de alocar exclusivamente empresas que divulgassem os resultados a partir das datas de fechamento de trimestres mais usuais no mercado brasileiro, ou seja, 31/03, 30/06, 30/09 e 31/12. No mercado americano, grande parte das empresas também praticam a divulgação após essas datas, no entanto as que não praticam eventualmente entram antes na carteira, se apresentarem a probabilidade definida de 60%.

Com essa alteração, houve uma discreta mudança na quantidade de ações alocadas de cada bolsa, todavia houve uma redução de 547 ações no quadro total. Apesar de serem pequenas as mudanças no que se refere à participação de cada bolsa no total, houve mudanças nos resultados, os quais são apresentados na Tabela 23.

**Tabela 23 – Quantidade de ações por bolsa com restrição de data de publicação**

<b>Bolsa</b>	<b>Quant.</b>	<b>Perc.</b>
NYSE	2.374	56,20%
NASDAQ	1.728	40,90%
B3	30	0,70%
NYSE Alternext U.S.	93	2,20%
<b>Total</b>	<b>4.225</b>	

**Fonte:** Elaborada pelo autor.

Na Tabela 24, mostra-se o lado *long* da carteira *long and short*. São apresentadas as colunas nível de acerto, retorno sem critério e com critério de *stop loss*, quantidade de ações alocadas para o trimestre e o nível previsto, ou seja, a probabilidade.

**Tabela 24 – Lado long da carteira long and short com restrição de data de publicação**

t	Nível de acerto	Ret. s/ stop	Ret. c/ stop	Quant.	Nível previsto
2008q1	35%	-6%	0%	23	61%
2008q3	25%	1%	22%	4	62%
2008q4	73%	25%	26%	100	77%
2009q1	77%	25%	26%	100	91%
2009q2	85%	23%	23%	100	89%
2009q3	76%	10%	11%	100	75%
2010q2	78%	9%	9%	100	63%
2010q3	85%	14%	14%	100	81%
2010q4	90%	7%	7%	100	62%
2011q3	83%	14%	15%	100	79%
2012q1	64%	3%	3%	100	63%
2012q3	81%	13%	13%	100	73%
2012q4	59%	7%	9%	100	73%
2013q1	76%	9%	9%	100	65%
2013q2	75%	8%	9%	100	81%
2013q3	81%	13%	13%	100	73%
2013q4	68%	3%	5%	100	63%
2014q1	60%	1%	3%	100	62%
2014q2	56%	3%	5%	100	66%
2014q3	0%	-19%	-10%	1	60%
2015q3	35%	-5%	-1%	100	61%
2015q4	86%	14%	15%	100	62%
2016q1	60%	5%	7%	100	64%
2016q2	36%	-2%	-1%	100	63%
2016q3	77%	12%	13%	100	82%
2016q4	50%	0%	2%	100	65%
2017q1	61%	2%	4%	100	66%
2017q2	67%	4%	5%	100	69%
2017q3	54%	2%	3%	100	63%
2017q4	48%	1%	2%	100	63%
2018q1	79%	10%	10%	100	64%
2018q2	0%	-67%	-10%	1	71%
2018q3	44%	-4%	1%	100	69%
2018q4	60%	16%	17%	5	61%
2019q1	100%	10%	10%	1	74%
SD		13%	8%		
RET		169%	1316%		
MÉDIA		4%	6%		
SHARPE		26%	76%		

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD. Retornos em dólar.

Assim como no caso da carteira anterior, a carteira apresentada na Tabela 24 também expõe trimestres sem nenhuma alocação, nos quais, conforme mencionado anteriormente, não houve empresas que apresentassem os critérios que possibilitassem a elas

fazer parte da carteira. Em tais critérios, pode acontecer de haver um número maior no lado *short* ou simplesmente não haver empresas para se investir em nenhum dos dois lados. O melhor retorno acumulado se deu na carteira com critério de *stop loss* (definido em 10% para todas as carteiras), com retorno acumulado de 1.316%, desvio padrão de 8% e índice de Sharpe de 76%. O lado *short* da carteira é apresentado por meio da Tabela 25.

**Tabela 25 – Lado *short* da carteira *long and short* com restrição de data de publicação**

t	Nível de acerto	Ret. s/ stop	Ret. c/ stop	Quant.	Nível previsto
2008q1	42%	-1%	9%	24	62%
2008q2	91%	26%	28%	100	96%
2008q3	73%	14%	18%	86	62%
2009q4	29%	-10%	-3%	100	63%
2010q1	86%	10%	11%	100	77%
2010q4	50%	-24%	1%	2	64%
2011q1	55%	2%	3%	100	75%
2011q2	55%	3%	7%	100	63%
2011q4	37%	-6%	-2%	100	71%
2012q1	67%	11%	17%	3	61%
2012q2	39%	-2%	1%	100	63%
2013q4	50%	14%	14%	2	62%
2014q3	41%	-9%	4%	100	64%
2014q4	50%	1%	4%	100	63%
2015q1	72%	11%	14%	100	63%
2015q2	67%	7%	12%	100	63%
2015q3	80%	26%	31%	30	62%
2015q4	33%	-27%	0%	15	62%
2016q1	78%	-12%	28%	9	63%
2016q2	74%	17%	22%	21	63%
2016q4	100%	43%	43%	1	60%
2017q1	100%	44%	44%	2	62%
2017q3	43%	-18%	17%	16	62%
2017q4	75%	10%	15%	4	63%
2018q1	50%	15%	23%	2	61%
2018q2	71%	9%	12%	100	63%
2018q3	0%	-59%	-10%	2	71%
2018q4	59%	4%	7%	100	62%
2019q1	70%	12%	15%	100	64%
SD		16%	12%		
RET		51%	3065%		
MÉDIA		2%	8%		
SHARPE		15%	70%		

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD. Retornos em dólar.

Por fim, consolida-se ambos os lados, o que pode ser observado na Tabela 26. Dentro dessa nova metodologia, com restrição de datas, formou-se mais seis carteiras, assim como as anteriores. Dessa forma, apresenta-se as carteiras *long and short* sem *stop* e com *stop* “pura”, com 1/3 de T-BOND 10 e com 1/3 de LFT.

Nota-se que, apesar de a quantidade de ações ter diminuído, os retornos em todas as carteiras foram incrementados e o risco em todas elas, com exceção da LS *c/stop c/LFT*, o qual permaneceu o mesmo. Destaca-se ainda que, em todos os casos, o índice de Sharpe aumentou indicando que a imposição de uma restrição para a data de publicação trouxe benefícios para a carteira.

**Tabela 26 – Resultados das carteiras *long and short* com restrição nas datas de publicação para alocação**

t	<i>Long and short sem stop</i>			<i>Long and short com stop</i>		
	LS s/índice	c/ BOND	c/ LFT	LS s/índice	c/ BOND	c/ LFT
2008q1	-4%	-2%	-1%	4%	3%	5%
2008q2	13%	7%	0%	14%	8%	0%
2008q3	7%	-4%	6%	20%	4%	14%
2008q4	12%	14%	13%	13%	14%	13%
2009q1	13%	13%	14%	13%	13%	14%
2009q2	11%	6%	10%	11%	6%	10%
2009q3	5%	7%	4%	5%	7%	4%
2009q4	-5%	-7%	-3%	-1%	-4%	-1%
2010q1	5%	-4%	4%	5%	-4%	5%
2010q2	4%	4%	5%	5%	4%	5%
2010q3	7%	13%	8%	7%	13%	8%
2010q4	-9%	-8%	-5%	4%	0%	3%
2011q1	1%	-9%	2%	2%	-9%	2%
2011q2	1%	-2%	-1%	3%	-1%	0%
2011q3	7%	8%	5%	8%	8%	5%
2011q4	-3%	-9%	-4%	-1%	-7%	-3%
2012q1	7%	4%	5%	10%	6%	7%
2012q2	-1%	-3%	-1%	1%	-2%	0%
2012q3	6%	10%	6%	7%	10%	6%
2012q4	4%	4%	3%	4%	4%	3%
2013q1	4%	16%	-1%	5%	16%	-1%
2013q2	4%	3%	3%	5%	3%	4%
2013q3	6%	5%	7%	7%	5%	7%
2013q4	9%	3%	7%	9%	4%	7%
2014q1	1%	-1%	0%	1%	0%	1%
2014q2	2%	0%	-3%	2%	0%	-2%
2014q3	-14%	-15%	-14%	-3%	-8%	-7%
2014q4	1%	4%	3%	2%	5%	4%
2015q1	5%	5%	0%	7%	6%	1%

(continua)

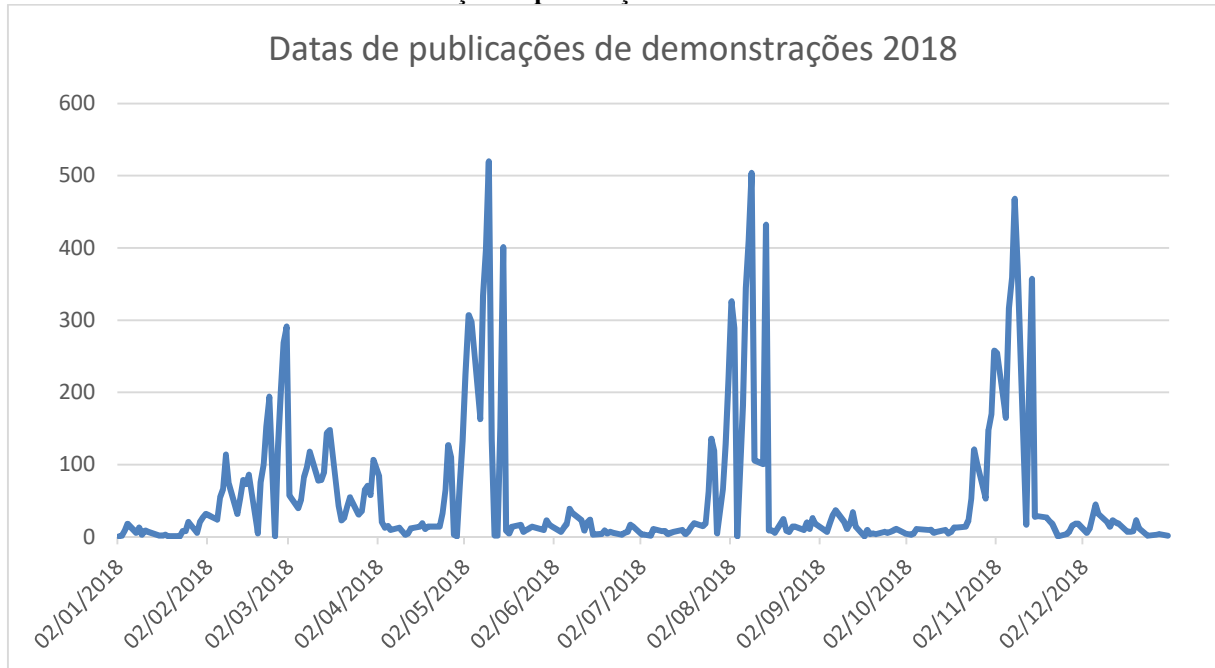
(conclusão)

t	<i>Long and short sem stop</i>			<i>Long and short com stop</i>		
	LS s/índice	c/ BOND	c/ LFT	LS s/índice	c/ BOND	c/ LFT
2015q2	4%	4%	1%	6%	5%	2%
2015q3	10%	0%	11%	15%	3%	14%
2015q4	-6%	-6%	-3%	7%	3%	6%
2016q1	-3%	-6%	3%	17%	8%	17%
2016q2	7%	21%	3%	10%	23%	5%
2016q3	6%	6%	9%	6%	7%	9%
2016q4	22%	14%	16%	23%	14%	16%
2017q1	23%	13%	15%	24%	14%	16%
2017q2	2%	4%	1%	2%	4%	1%
2017q3	-8%	0%	-5%	10%	11%	7%
2017q4	5%	8%	1%	9%	10%	3%
2018q1	13%	6%	7%	17%	9%	10%
2018q2	-29%	-16%	-18%	1%	3%	2%
2018q3	-31%	-29%	-21%	-5%	-11%	-3%
2018q4	10%	6%	6%	12%	8%	7%
2019q1	11%	-4%	8%	12%	-3%	9%
2019q2	0%	5%	-1%	0%	5%	-1%
SD	10%	9%	7%	6%	7%	6%
RET	201%	93%	142%	2284%	675%	837%
MÉDIA	3%	2%	2%	7%	5%	5%
SHARPE	30%	21%	30%	114%	70%	93%
PERC. POS.	74%	59%	65%	89%	78%	85%

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD., PERC. POS = percentual de trimestres com retorno positivo. Retornos em dólar.

Constata-se que, mesmo após a imposição das datas e das mudanças, que eventualmente aconteceram nas ações que formaram as carteiras, os resultados se mantiveram positivos, o que mostra a robustez do modelo econométrico. A imposição de datas ocorreu quando se percebeu que muitas empresas ainda não haviam divulgado as informações financeiras após as carteiras já estarem completas. Nesse sentido, apurou-se que oportunidades estavam sendo perdidas e optou-se por protelar a data de início da composição das carteiras. Com base no Gráfico 5, percebeu-se que há uma concentração de divulgação de resultados no entorno de algumas datas.

**Gráfico 5 – Datas com maior concentração de publicações em 2018**

**Fonte:** Elaborado pelo autor.

Assim como apresentado na Tabela 19, faz-se novamente o comparativo trimestre a trimestre acumulado da nova carteira mantendo na Tabela 27 a primeira carteira LS para fins de comparação. Como os indicadores acumulados já foram apresentados por meio da Tabela 19 e não houve alteração, eles não serão mostrados novamente.

**Tabela 27 – Resultados trimestre a trimestre acumulados para as carteiras LS1 e LS2**

<b>t</b>	<b>LS s/índice1</b>	<b>LS s/índice2</b>	<b>t</b>	<b>LS s/índice1</b>	<b>LS s/índice2</b>
2008q2	4%	19%	2014q1	247%	363%
2008q3	25%	43%	2014q2	251%	374%
2008q4	31%	61%	2014q3	239%	359%
2009q1	63%	82%	2014q4	246%	368%
2009q2	78%	103%	2015q1	267%	401%
2009q3	84%	113%	2015q2	293%	432%
2009q4	81%	110%	2015q3	348%	512%
2010q1	85%	122%	2015q4	380%	556%
2010q2	94%	132%	2016q1	476%	670%
2010q3	119%	148%	2016q2	539%	751%
2010q4	129%	159%	2016q3	562%	804%
2011q1	131%	163%	2016q4	707%	1010%
2011q2	140%	172%	2017q1	932%	1279%
2011q3	152%	192%	2017q2	949%	1312%
2011q4	144%	190%	2017q3	1078%	1448%
2012q1	169%	220%	2017q4	1191%	1581%
2012q2	171%	222%	2018q1	1380%	1866%

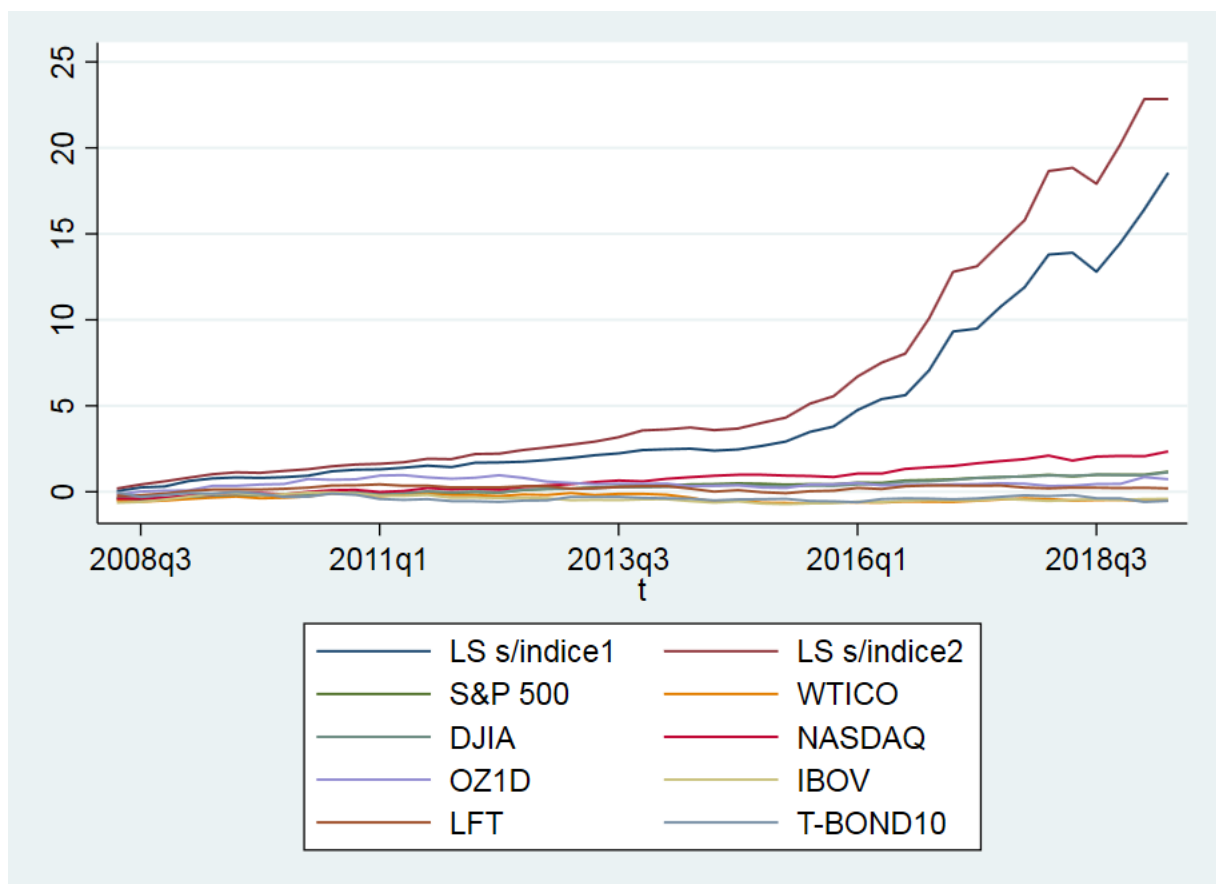
(continua)

(conclusão)					
t	LS s/índice1	LS s/índice2	t	LS s/índice1	LS s/índice2
2012q3	175%	243%	2018q2	1390%	1884%
2012q4	185%	258%	2018q3	1280%	1791%
2013q1	197%	274%	2018q4	1447%	2021%
2013q2	213%	292%	2019q1	1642%	2284%
2013q3	224%	318%	2019q2	1855%	2284%
2013q4	243%	357%			

**Fonte:** Elaborada pelo autor a partir de dados do *output* do software Stata.

Na sequência, os dados da Tabela 27 com as carteiras LS1 e LS2, inclusive os dos principais indicadores, são apresentados no Gráfico 6.

**Gráfico 6 – Retorno acumulado trimestre a trimestre das carteiras LS s/índice1, LS s/índice2 e demais indicadores**



**Fonte:** *Output* do software Stata.

Por meio do Gráfico 6 observa-se a superioridade da carteira LS2 em relação à carteira LS1. Nota-se ainda que ambas seguem uma tendência muito parecida, que



provavelmente é provocada por uma correlação entre elas advinda da alocação de ações em comum entre ambas.

Novamente é executado o teste de diferença de médias para verificar se a nova carteira tem as médias estatisticamente diferentes dos indicadores e da carteira anterior (Tabela 28).

**Tabela 28 – Two-sample Wilcoxon rank-sum (Mann-Whitney) test: carteiras LS1 e LS2 versus indicadores**

<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2533	2139
S&P 500	46	1745	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{S\&P 500})$			
z=		3,077	
Prob >  z  =		0,0021	
Exact Prob =		0,0019	
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2456,5	2139
WTICO	46	1821,5	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{WTICO})$			
z=		2,335	
Prob >  z  =		0,0132	
Exact Prob =		0,0128	
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2565	2139
DJIA	46	1713	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{DJIA})$			
z=		2,741	
Prob >  z  =		0,0009	
Exact Prob =		0,0007	
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2411	2139
NASDAQ	46	1867	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{NASDAQ})$			
z=		2,124	
Prob >  z  =		0,0337	
Exact Prob =		0,0334	

(continua)

(conclusão)			
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2662	2139
OZ1D	46	1616	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{OZ1D})$			
z=	4,084		
Prob >  z  =	0,0000		
Exact Prob =	0,0000		
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2461	2139
IBOV	46	1817	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{IBOV})$			
z=	2,514		
Prob >  z  =	0,0119		
Exact Prob =	0,015		
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2658	2139
LFT	46	1620	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{LFT})$			
z=	4,053		
Prob >  z  =	0,0001		
Exact Prob =	0,0000		
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2551	2139
T-BOND10	46	1727	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{T-BOND10})$			
z=	3,217		
Prob >  z  =	0,0013		
Exact Prob =	0,0000		
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2229,5	2139
LS s/índice	46	2048,5	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{T-BOND10})$			
z=	0,707		
Prob >  z  =	0,4797		
Exact Prob =	0,4828		

**Fonte:** Output do software Stata.

Conforme demonstrado na Tabela 28, apesar de a carteira LS2 ter apresentado resultado maior que a LS, esse resultado não é significativo estatisticamente. A despeito disso, todos os indicadores tiveram resultado significativo do ponto de vista estatístico, inclusive o NASDAQ, que não havia apresentado diferença estatística com a carteira LS.

Para a carteira LS2, não foi apresentada a classificação setorial novamente, tendo em vista que não houve mudanças significativas. Na sequência, optou-se por elaborar uma carteira utilizando só ações brasileiras, dado que, no critério anterior, a participação das brasileiras continuou pequena. Nesse caso, não houve limitação de data de publicação, no entanto se alterou o limite de 200 (100 *long* e 100 *short*) ações na carteira para 60 (30 *long* e 30 *short*) ações. A Tabela 29 apresenta o lado *long* da carteira formada apenas com ações listadas na B3.

**Tabela 29 – Lado *short* da carteira *long and short* com ações listadas na B3**

<b>t</b>	<b>Nível de acerto</b>	<b>Ret. s/ stop</b>	<b>Ret. c/ stop</b>	<b>Quant.</b>	<b>Nível previsto</b>
2008q1	13%	-20%	-6%	8	61%
2008q3	100%	118%	118%	1	62%
2008q4	97%	39%	39%	30	79%
2009q1	93%	28%	28%	30	93%
2009q2	93%	20%	20%	30	91%
2009q3	71%	4%	5%	30	81%
2010q2	72%	9%	9%	30	63%
2010q3	60%	6%	9%	30	84%
2010q4	64%	3%	3%	26	61%
2011q3	85%	14%	15%	30	83%
2012q1	57%	-2%	0%	30	63%
2012q3	77%	10%	11%	30	77%
2012q4	50%	-4%	0%	30	78%
2013q1	30%	-17%	-8%	30	70%
2013q2	67%	3%	5%	30	85%
2013q3	21%	-3%	0%	30	64%
2013q4	71%	8%	8%	30	64%
2014q1	70%	3%	4%	30	62%
2014q2	14%	-19%	-9%	30	71%
2014q3	0%	-19%	-10%	1	60%
2015q3	100%	95%	95%	2	61%
2015q4	60%	8%	9%	20	61%
2016q1	87%	37%	37%	30	68%
2016q2	32%	-15%	-8%	30	64%
2016q3	77%	26%	29%	30	87%
2016q4	69%	0%	1%	30	65%
2017q1	87%	13%	13%	30	67%
2017q2	83%	14%	14%	30	68%

(continua)

(conclusão)

t	Nível de acerto	Ret. s/ stop	Ret. c/ stop	Quant.	Nível previsto
2017q3	73%	9%	11%	30	63%
2017q4	40%	-8%	-5%	30	62%
2018q1	24%	-12%	-7%	30	65%
2018q3	69%	11%	11%	30	67%
2019q1	100%	10%	10%	1	74%
SD		25%	23%		
RET		1306%	3324%		
MÉDIA		8%	10%		
SHARPE		33%	42%		

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD. Retornos em dólar.

Os resultados acumulados da carteira apresentada na Tabela 29 mostraram-se interessantes, porém com alto risco julgado pelos desvios padrão de 25% para a carteira com *stop* e 23% para a carteira sem *stop*. Em razão desse alto desvio padrão, o índice de Sharpe da carteira ficou abaixo das carteiras anteriores. Todavia, trata-se apenas do lado *long* da carteira, ou seja, o lado comprado. Um dos objetivos quando se unem os lados *long* e *short* é justamente minimizar os riscos.

Assim como nos cenários anteriores, a inserção de um critério de *stop loss* incrementa substancialmente o retorno acumulado. Recapitulando: no caso sem *stop loss*, a ação só é retirada da carteira após a divulgação de novas informações financeiras e se não obedecer aos critérios para permanecer alocada na carteira. Na sequência, a Tabela 30 apresenta o lado *short* da carteira.

**Tabela 30 – Lado *short* da carteira *long and short* somente com ações listadas na B3**

t	Nível de acerto	Ret. s/ stop	Ret. c/ stop	Quant.	Nível previsto
2008q1	100%	72%	72%	1	61%
2008q2	93%	60%	60%	30	97%
2008q3	67%	1%	4%	3	67%
2009q4	89%	12%	12%	20	61%
2010q1	33%	-9%	-4%	30	78%
2011q1	79%	2%	11%	30	80%
2011q2	54%	9%	12%	30	63%
2011q4	86%	19%	19%	30	75%
2012q2	64%	4%	7%	30	62%
2013q4	100%	31%	31%	1	64%
2014q3	59%	26%	27%	19	61%
2014q4	38%	-24%	-6%	13	63%
2015q1	79%	32%	33%	30	69%
2015q2	68%	14%	19%	23	62%

(continua)

(conclusão)

t	Nível de acerto	Ret. s/ stop	Ret. c/ stop	Quant.	Nível previsto
2015q3	100%	33%	33%	2	63%
2015q4	0%	-12%	-10%	1	60%
2018q2	33%	-10%	-3%	30	62%
2018q4	100%	19%	19%	4	65%
2019q1	30%	-19%	-4%	22	62%
SD		17%	16%		
RET		635%	1479%		
MÉDIA		6%	7%		
SHARPE		33%	45%		

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD. Retornos em Dólar.

De maneira geral, tanto os lados *long* quanto os lados *short* das carteiras criadas a partir das ações da B3 apresentaram maior risco que as ações criadas com todas as ações da amostra. Acredita-se que a maior volatilidade dessa carteira se deu principalmente pela menor maturidade do mercado brasileiro de ações. Além disso, observou-se a capacidade que a diversificação internacional tem de minimizar o risco. Na Tabela 31, demonstra-se o consolidado dos lados *long* e *short* da carteira B3.

**Tabela 31 – Resultados da carteira *long and short* somente com ações listadas na B3**

t	<i>Long and short sem stop</i>			<i>Long and short com stop</i>		
	LS s/ índice	c/ BOND	c/ LFT	LS s/índice	c/ BOND	c/ LFT
2008q1	26%	18%	19%	33%	22%	24%
2008q2	30%	19%	11%	30%	19%	11%
2008q3	60%	30%	41%	61%	31%	42%
2008q4	20%	18%	17%	20%	18%	17%
2009q1	14%	14%	14%	14%	14%	14%
2009q2	10%	6%	10%	10%	6%	10%
2009q3	2%	5%	1%	3%	6%	2%
2009q4	6%	1%	4%	6%	1%	4%
2010q1	-4%	-10%	-2%	-2%	-9%	0%
2010q2	4%	4%	5%	5%	4%	5%
2010q3	3%	10%	5%	4%	11%	6%
2010q4	1%	-2%	1%	2%	-1%	1%
2011q1	1%	-9%	2%	6%	-6%	5%
2011q2	4%	0%	1%	6%	1%	2%
2011q3	7%	7%	5%	8%	8%	5%
2011q4	9%	0%	4%	9%	0%	4%
2012q1	-1%	-1%	0%	0%	-1%	0%
2012q2	2%	-1%	1%	3%	0%	2%
2012q3	5%	9%	5%	6%	10%	5%

(continua)

(conclusão)

t	<i>Long and short sem stop</i>			<i>Long and short com stop</i>		
	LS s/ índice	c/ BOND	c/ LFT	LS s/índice	c/ BOND	c/ LFT
2012q4	-2%	0%	-1%	0%	1%	0%
2013q1	-8%	8%	-9%	-4%	11%	-6%
2013q2	1%	1%	1%	2%	1%	2%
2013q3	-2%	-1%	1%	0%	0%	2%
2013q4	19%	10%	14%	19%	10%	14%
2014q1	2%	0%	1%	2%	0%	1%
2014q2	-10%	-8%	-10%	-5%	-4%	-7%
2014q3	4%	-3%	-2%	8%	0%	0%
2014q4	-12%	-4%	-5%	-3%	2%	1%
2015q1	16%	11%	7%	16%	12%	7%
2015q2	7%	6%	3%	9%	7%	4%
2015q3	64%	35%	47%	64%	35%	47%
2015q4	-2%	-3%	0%	-1%	-2%	1%
2016q1	19%	8%	17%	19%	8%	17%
2016q2	-8%	11%	-7%	-4%	13%	-4%
2016q3	13%	11%	13%	14%	12%	14%
2016q4	0%	-1%	1%	1%	0%	2%
2017q1	7%	2%	4%	7%	2%	4%
2017q2	7%	8%	4%	7%	8%	4%
2017q3	4%	8%	3%	6%	9%	4%
2017q4	-4%	1%	-6%	-3%	2%	-5%
2018q1	-6%	-6%	-5%	-4%	-5%	-4%
2018q2	-5%	-1%	-2%	-1%	2%	0%
2018q3	5%	-4%	3%	5%	-4%	3%
2018q4	9%	6%	6%	9%	6%	6%
2019q1	-4%	-14%	-2%	3%	-9%	2%
2019q2	0%	5%	-1%	0%	5%	-1%
SD	15%	10%	11%	14%	9%	10%
RET	1327%	512%	595%	2957%	920%	1057%
MÉDIA	7%	4%	5%	8%	6%	6%
SHARPE	46%	47%	45%	60%	61%	58%
PERC. POS.	67%	59%	70%	74%	72%	87%

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD., PERC. POS = percentual de trimestres com retorno positivo. Retornos em dólar.

O mercado brasileiro ainda é menos amadurecido em comparação ao mercado americano. Nesse sentido, esperava-se que os resultados fossem diferentes. A terceira e última carteira mostrou retornos maiores, mas riscos também maiores. O incremento no risco fez com que a carteira apresentasse menor índice de Sharpe, ou seja, a carteira formada tão somente com ações brasileiras trouxe retorno maior que as anteriores sem ajuste ao risco e menor quando os retornos são ajustados ao risco. À vista disso, percebeu-se que, no mercado

brasileiro, as possibilidades de retorno são maiores, mesmo sendo composto por um número menor de empresas, todavia assumem-se maiores riscos.

O resultado encontrado para as carteiras com ações está em linha com os achados de Araújo, Brito e Sanvicente (2020). Os autores levantaram os retornos do índice Bovespa de 1968 a 2019 e encontraram uma média aritmética anual de 21,3%, prêmio de risco de 20,1% por ano e um altíssimo desvio padrão de 67%, o que resultou em um Sharpe de 0,30%. Eles destacam que o prêmio de risco brasileiro é mais alto do que o americano, no entanto a alta volatilidade desencoraja investimentos mais robustos no mercado de ações brasileiro. Ainda segundo esses autores, para o mesmo período, foram constatados, para o mercado americano, prêmio de risco de 6,4% e desvio padrão de 17%, e isso resultou em um Sharpe de 0,38%.

Com base nos resultados apresentados por esses autores, duas perspectivas podem ser analisadas tanto em relação ao trabalho deles quanto no tocante à presente tese. Em comparação ao mercado brasileiro, o retorno do mercado americano é menor, porém com desvio padrão também menor. Isso resulta, portanto, em um índice de Sharpe maior, o que caracteriza um mercado ideal para pessoas com baixa tolerância ao risco. Apesar de apresentar Sharpe menor que o mercado americano, tendo em vista o desvio padrão bastante superior a ele, o mercado brasileiro tem uma característica interessante para os investidores interessados em maiores riscos, com alto desvio padrão e alto retorno.

Assim como nas demais carteiras, o melhor resultado foi apresentado pela carteira com delimitação de *stop loss* e sem a inclusão de nenhum índice de renda fixa, ou seja, uma carteira *long and short* composta unicamente por ações. O retorno total acumulado da melhor carteira foi de 2.957% no período estudado com desvio padrão de 14%, índice de Sharpe de 60% e com 74% dos trimestres apresentando resultado positivo.

Na Tabela 32, apresenta-se as ações brasileiras considerando o investimento em reais. Optou-se por expor a carteira dessa forma com objetivo de mostrar os retornos sob uma perspectiva abasileirada.

**Tabela 32 – Resultados da carteira *long and short* somente com ações listadas na B3 com valores em reais**

t	<i>Long and short sem stop</i>		<i>Long and short com stop</i>		IBOV_BRL	LFT_BRL
	LS s/índice	c/LFT	LS s/índice	c/LFT		
2008q1	25%	17%	32%	22%	14%	1%
2008q2	10%	8%	11%	8%	-24%	3%
2008q3	61%	42%	62%	43%	-34%	3%
2008q4	12%	10%	12%	10%	14%	5%
2009q1	6%	5%	6%	5%	17%	1%
2009q2	7%	5%	7%	5%	16%	2%
2009q3	4%	3%	5%	4%	15%	2%
2009q4	6%	5%	6%	5%	8%	3%
2010q1	-8%	-5%	-3%	-2%	-11%	1%
2010q2	3%	3%	3%	3%	4%	2%
2010q3	1%	1%	2%	2%	6%	3%
2010q4	1%	2%	2%	2%	-3%	4%
2011q1	5%	4%	5%	4%	-9%	1%
2011q2	2%	2%	3%	3%	-15%	3%
2011q3	9%	7%	9%	7%	7%	3%
2011q4	1%	2%	1%	2%	11%	4%
2012q1	0%	0%	1%	1%	-13%	1%
2012q2	0%	0%	2%	2%	4%	2%
2012q3	4%	3%	4%	4%	-3%	2%
2012q4	-2%	0%	0%	1%	0%	3%
2013q1	-3%	-1%	-1%	0%	-2%	1%
2013q2	2%	2%	3%	2%	-7%	2%
2013q3	-3%	-1%	-1%	0%	5%	2%
2013q4	19%	14%	19%	14%	-6%	4%
2014q1	4%	3%	4%	3%	7%	1%
2014q2	-4%	-2%	-2%	-1%	3%	3%
2014q3	5%	4%	6%	5%	-9%	3%
2014q4	-8%	-4%	-1%	0%	-1%	4%
2015q1	7%	5%	8%	6%	9%	1%
2015q2	3%	3%	7%	6%	-17%	3%
2015q3	60%	41%	60%	41%	-2%	3%
2015q4	4%	4%	5%	5%	8%	5%
2016q1	12%	8%	12%	9%	2%	2%
2016q2	-4%	-1%	-2%	0%	13%	3%
2016q3	7%	6%	9%	7%	1%	3%
2016q4	1%	2%	2%	3%	9%	5%
2017q1	6%	4%	6%	4%	5%	1%
2017q2	9%	7%	9%	7%	-1%	2%
2017q3	4%	3%	5%	4%	4%	2%
2017q4	0%	1%	0%	1%	21%	2%
2018q1	-3%	-2%	-2%	-1%	0%	1%
2018q2	-4%	-2%	-1%	0%	-9%	2%

(continua)



(conclusão)

t	<i>Long and short sem stop</i>		<i>Long and short com stop</i>		IBOV_BRL	LFT_BRL
	LS s/índice	c/LFT	LS s/índice	c/LFT		
2018q3	5%	4%	5%	4%	12%	2%
2018q4	8%	6%	8%	6%	11%	2%
2019q1	-4%	-3%	3%	2%	-5%	1%
2019q2	8%	6%	9%	7%	11%	2%
SD	13%	9%	13%	9%	11%	1%
RET	1009%	660%	1915%	1032%	44%	199%
MÉDIA	6%	5%	7%	6%	1%	2%
SHARPE	46%	55%	57%	66%	13%	218%
PERC. POS	76%	78%	83%	89%	59%	100%

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD., PERC. POS = percentual de trimestres com retorno positivo. Retornos em real.

No contexto elaborado em reais, observou-se que o indicador de renda fixa LFT cumpriu o que se esperava, o que diminuiu o risco das carteiras, e isso proporcionou incremento no índice de Sharpe. Apresentou-se ainda separadamente os retornos do IBOV e da LFT no período, que totalizaram, respectivamente, 44% e 199%. Destaca-se que o melhor retorno ajustado ao risco foi gerado justamente pela LFT, no entanto, tendo em vista o atual patamar da taxa de juro brasileira (2% em setembro de 2020), não é possível obter esse retorno por meio da LFT. No que se refere ao retorno acumulado, o melhor foi apresentado pela carteira LS com *stop* sem LFT.

A seguir, mostra-se, na Tabela 33, a evolução trimestral de cada uma das metodologias aplicadas, em outras palavras, a carteira com todas as ações, a carteira com ações delimitadas por data de publicação e, por fim, a carteira criada exclusivamente com ações listadas na B3. Nota-se que, nos três casos, os melhores desempenhos foram encontrados sem a adição de um índice de renda fixa. Nesse caso optou-se por não incluir a carteira com as ações em real, dado que em diferentes moedas a comparação pode se tornar equivocada.

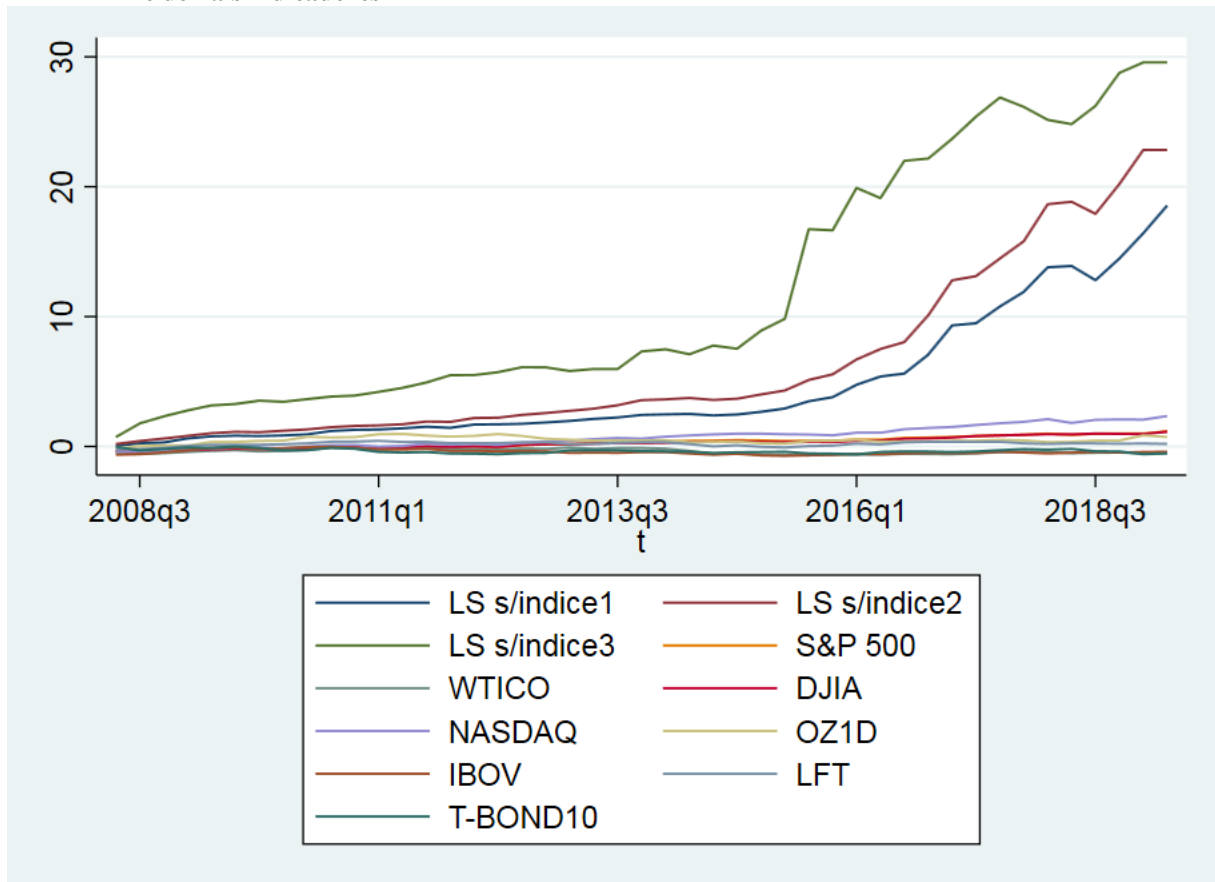
**Tabela 33 – Resultados trimestre a trimestre acumulados para as carteiras LS1, LS2 e LS3**

t	LS s/índice1	LS s/índice2	LS s/índice3	t	LS s/índice1	LS s/índice2	LS s/índice3
2008q2	4%	19%	72%	2014q1	247%	363%	749%
2008q3	25%	43%	178%	2014q2	251%	374%	710%
2008q4	31%	61%	232%	2014q3	239%	359%	777%
2009q1	63%	82%	278%	2014q4	246%	368%	753%
2009q2	78%	103%	316%	2015q1	267%	401%	891%
2009q3	84%	113%	327%	2015q2	293%	432%	983%
2009q4	81%	110%	353%	2015q3	348%	512%	1674%
2010q1	85%	122%	344%	2015q4	380%	556%	1665%
2010q2	94%	132%	364%	2016q1	476%	670%	1991%
2010q3	119%	148%	384%	2016q2	539%	751%	1912%
2010q4	129%	159%	392%	2016q3	562%	804%	2200%
2011q1	131%	163%	420%	2016q4	707%	1010%	2217%
2011q2	140%	172%	452%	2017q1	932%	1279%	2369%
2011q3	152%	192%	493%	2017q2	949%	1312%	2540%
2011q4	144%	190%	549%	2017q3	1078%	1448%	2687%
2012q1	169%	220%	550%	2017q4	1191%	1581%	2615%
2012q2	171%	222%	572%	2018q1	1380%	1866%	2515%
2012q3	175%	243%	610%	2018q2	1390%	1884%	2482%
2012q4	185%	258%	609%	2018q3	1280%	1791%	2622%
2013q1	197%	274%	581%	2018q4	1447%	2021%	2877%
2013q2	213%	292%	597%	2019q1	1642%	2284%	2957%
2013q3	224%	318%	596%	2019q2	1855%	2284%	2957%
2013q4	243%	357%	731%				

**Fonte:** Elaborada pelo autor a partir de dados do *output* do software Stata.

O Gráfico 7 ilustra o comportamento parecido das carteiras, mesmo tendo sido formadas em diferentes bolsas. Vale lembrar que o modelo de base para seleção das ações foi composto pelos retornos defasados das próprias ações, além da variação do ouro, S&P 500 e T-BOND 10. A questão que talvez suscite dúvidas é: como podem indicadores externos influenciarem o retorno das ações brasileiras? Segundo uma das linhas que explicam esse fato, isso ocorre porque há fatores externos que influenciam o comportamento do mercado como um todo, por exemplo, guerras, crises, epidemias e notícias de maneira geral, os quais podem influenciar empresas específicas ou o mercado como um todo (CHEN, 2020; HUANG; CHAI; CHO, 2020). Por fim, destaca-se que o ouro é utilizado por muitos investidores como forma de proteção para as carteiras de investimento e que, portanto, se justifica sua influência nas carteiras mostradas no Gráfico 7 (SHABBIR; KOUSAR; BATOOL, 2019).

**Gráfico 7 – Retorno acumulado trimestre a trimestre das carteiras LS s/índice1, LS s/índice2, LS s/índice3 e demais indicadores**



Fonte: Output do software Stata.

Assim como no caso das demais carteiras, para essa última também será executado o teste de diferença de médias (Tabela 34).

**Tabela 34 – Two-sample Wilcoxon rank-sum (Mann-Whitney) test: carteiras LS1, LS2 e LS3 versus indicadores**

<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice3	46	2378	2139
S&P 500	46	1900	2139
combined	92	4278	4278
Ho: ret(i==LS s/índice3) = ret(i==S&P 500)			
z=		1,866	
Prob >  z  =		0,0620	
Exact Prob =		0,0623	

(continua)

(continuação)

<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice3	46	2431,5	2139
WTICO	46	1846,5	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice3}) = \text{ret}(i==\text{WTICO})$			
z=	2,284		
Prob >  z  =	0,0224		
Exact Prob =	0,0220		

<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice3	46	2396	2139
DJIA	46	1882	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice3}) = \text{ret}(i==\text{DJIA})$			
z=	2,007		
Prob >  z  =	0,0448		
Exact Prob =	0,0448		

<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice3	46	2270	2139
NASDAQ	46	2008	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice3}) = \text{ret}(i==\text{NASDAQ})$			
z=	1,023		
Prob >  z  =	0,3063		
Exact Prob =	0,3097		

<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice3	46	2525	2139
OZ1D	46	1753	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice3}) = \text{ret}(i==\text{OZ1D})$			
z=	30,14		
Prob >  z  =	0,0026		
Exact Prob =	0,0023		

<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice3	46	2399	2139
IBOV	46	1879	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice3}) = \text{ret}(i==\text{IBOV})$			
z=	2,030		
Prob >  z  =	0,0423		
Exact Prob =	0,0423		

(continua)

(conclusão)			
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2498	2139
LFT	46	1780	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{LFT})$			
z=	2,803		
Prob >  z  =	0,0051		
Exact Prob =	0,0048		
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice2	46	2497	2139
T-BOND10	46	1781	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{T-BOND10})$			
z=	2,796		
Prob >  z  =	0,0052		
Exact Prob =	0,0049		
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice3	46	2073	2139
LS s/índice	46	2205	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{LS s/índice})$			
z=	0,515		
Prob >  z  =	0,6063		
Exact Prob =	0,6106		
<b>i</b>	<b>obs</b>	<b>rank sum</b>	<b>expected</b>
LS s/índice3	46	2005,5	2139
LS s/índice2	46	2272,5	2139
combined	92	4278	4278
Ho: $\text{ret}(i==\text{LS s/índice2}) = \text{ret}(i==\text{LS s/índice2})$			
z=	1,042		
Prob >  z  =	0,2972		
Exact Prob =	0,2996		

**Fonte:** Output do software Stata.

Conforme demonstrado na Tabela 34, as carteiras LS1, LS2 e LS3 não apresentam diferença de médias estatisticamente significantes. Para os demais indicadores, com exceção do índice S&P 500, que apresentou significância ao nível de 10%, e o índice NASDAQ, que não apresentou significância estatística, todos os outros são diferentes ao nível de, pelo menos, 5%.

Quanto à participação setorial das empresas nas carteiras, na Tabela 35, apresenta-se os vinte setores com maior participação. Como a classificação NAICS tem mais de 270 setores, optou-se, por questão de espaço, mostrar tão somente os vinte primeiros.

**Tabela 35 – Classificação setorial NAICS com os vinte setores mais representativos na carteira**

<b>Cod. NAICS</b>	<b>Quant.</b>	<b>Perc.</b>
11 Geração, transmissão e distribuição de energia elétrica	376	32%
4 Bancos	97	8%
54 Administração de empresas e empreendimentos	78	7%
35 Outras indústrias	52	4%
86 Água, esgoto e outros sistemas	49	4%
213 Telecomunicações	33	3%
70 Mineração de metais	27	2%
3 Seguradora	24	2%
48 Construção de edifícios residenciais	24	2%
116 Distribuição de gás natural	24	2%
231 Escola de ensino superior	20	2%
207 Forjarias e estamparias	16	1%
62 Tecelagens	15	1%
5 Indústria de fumo	14	1%
94 Outras indústrias de alimentos	14	1%
22 Extração de petróleo e gás	12	1%
30 Indústria de máquinas industriais	12	1%
97 Indústria de roupas de tecido	12	1%
255 Indústria de ferragens	12	1%
20 Loja de roupas	11	1%

**Fonte:** Elaborada pelo autor.

Observa-se destaque para três setores: Geração, transmissão e distribuição de energia elétrica, Bancos e Administração de empresas e empreendimentos. Em relação ao primeiro, que apresentou 376 observações, 32 empresas estiveram presentes nas carteiras, entre elas, CPFL energia, Elektro, Eletrobras e Rede.

No que diz respeito ao segundo setor, com 97 observações, 26 bancos passaram pelas carteiras, entre eles, os maiores bancos brasileiros, por exemplo, Itaú, Bradesco, Brasil e Santander.

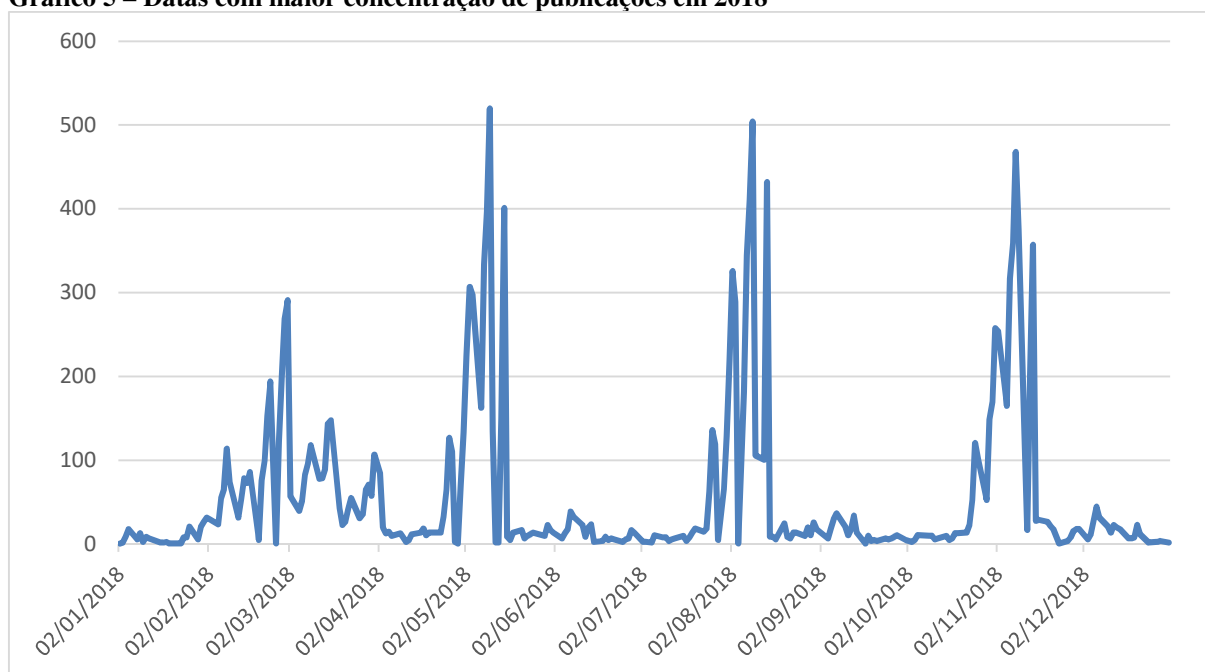
Quanto ao terceiro setor mais representativo, o de Administração de empresas e empreendimentos, com 78 observações, 13 empresas passaram pelas carteiras, entre elas, Itaúsa, GP Invest e Companhia de Participações Aliança Bahia. Enfatiza-se, novamente, que cada observação representa uma entrada na carteira, seja de empresas diferentes ou da mesma

empresa mais de uma vez. Destaca-se ainda que nem todas as empresas estão ativas hoje, mas estavam no momento do investimento.

Cabe mencionar que se pode ainda optar por investir nas carteiras *long* ou *short* isoladamente, pois dessa forma também apresentam bons resultados, apesar de apresentarem também maiores riscos.

Os modelos quantitativos têm ganhado evidência nos últimos anos em função de permitir a análise de um grande número de ações de forma rápida. Muitas vezes uma ação tem grandes altas ou grandes baixas após a divulgação dos resultados, visto que as informações nem sempre atendem às expectativas dos investidores. O Gráfico 5, aqui rerepresentado, mostra os períodos com grandes fluxos de divulgação de demonstrações financeiras.

**Gráfico 5 – Datas com maior concentração de publicações em 2018**



**Fonte:** Elaborado pelo autor.

Além da visualização gráfica, são expostos ainda por meio da Tabela 36 os dez dias com maiores números de publicações referentes ao ano de 2018. Dificilmente seria possível analisar esse volume de informações sem o apoio de um sistema.

**Tabela 36 – Datas com maior quantidade de publicações em 2018**

<b>Datas de pub.</b>	<b>Quant.</b>
10/05/2018	520
09/08/2018	504
08/11/2018	468
14/08/2018	432
08/08/2018	411
15/05/2018	401
09/05/2018	397
09/11/2018	376
07/11/2018	359
14/11/2018	357

**Fonte:** Elaborada pelo autor.

Destaca-se também a importância do *stop loss*, que foi utilizado no trabalho. Talvez se possa questionar a efetividade do modelo estatístico, tendo em vista que os resultados foram substancialmente menores quando utilizado somente ele sem o *stop loss*. No entanto, o modelo baseia-se no comportamento geral dos mercados e dos retornos das próprias ações. Nesse sentido, alguns momentos de maior volatilidade podem ser difíceis de serem previstos, por exemplo, o “Joesley Day”, que, em 17/05/2017, fez a B3 despencar mais de 10% – maior queda diária em nove anos – após vir à tona uma conversa entre o então presidente Michel Temer e o dono da JBS Joesley Batista. Está certo que a conversa não mudou os fundamentos da maioria das empresas, no entanto o conteúdo da gravação gerou insegurança para o mercado, que abandonou temporariamente as ações migrando para investimentos mais seguros. Mais recentemente se pode citar surgimento da Covid-19, a qual fez os mercados mundiais caírem por vários dias seguidos.

Destarte, não se pode deixar de destacar a importância de ambas as ferramentas, o modelo econométrico, com a função de selecionar as ações, e a técnica do *stop loss*, com o objetivo de limitar as perdas.

Em busca de verificar se o incremento encontrado nas carteiras de ações pode também ser obtido nos ETFs desses indicadores, simulou-se o investimento nos ETFs utilizando a técnica do *stop loss* para limitar perdas em eventuais quedas. Assim como nas carteiras de ações, o valor foi limitado em 10%. Para fins de comparação, foram inseridos tão somente o rodapé da Tabela 12 e o resumo dos resultados de como seria essa mesma tabela com o uso do *stop loss*. A seguir, na Tabela 37, apresenta-se os resultados.



**Tabela 37 – Comparativo dos desvios padrão, taxas de retornos, médias e índice de Sharpe dos indicadores utilizados sem *stop loss* e com *stop loss***

Sem stop	S&P 500	WTICO	DJIA	NASDAQ	OZ1D	IBOV	LFT	T-BOND 10
SD	8%	16%	7%	9%	9%	16%	8%	16%
RET	118%	-51%	114%	235%	73%	-41%	20%	-53%
MÉDIA	2%	0%	2%	3%	2%	0%	1%	0%
SHARPE	26%	-1%	27%	34%	18%	2%	9%	-2%

Com stop	S&P 500	WTICO	DJIA	NASDAQ	OZ1D	IBOV	LFT	T-BOND 10
SD	6%	11%	6%	7%	9%	12%	7%	13%
RET	196%	132%	168%	395%	86%	185%	60%	96%
MÉDIA	3%	2%	2%	4%	2%	3%	1%	2%
SHARPE	42%	22%	40%	54%	20%	25%	19%	17%

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: SD = Desvio padrão, RET = Retorno acumulado, SHARPE = Média/SD. Retornos em dólar.

Conforme exposto na Tabela 37, percebe-se que o benefício do *stop loss* também se reflete nos indicadores, no entanto com menor substancialidade. Observa-se melhora em todos os indicadores, desde o desvio padrão, que ficou menor, até o retorno ajustado ao risco, que ficou maior. Apesar disso, os resultados ainda ficam distantes dos resultados apurados nas carteiras formuladas a partir dos modelos hierárquicos com a utilização do *stop loss*.

Um grande desafio dos trabalhos acadêmicos envolvendo esse tema é o de transferir os custos de transações existentes no mundo real para as simulações acadêmicas. Assim, alguns autores têm sugerido utilizar o valor de 50 pontos base por transação. Desse modo, para cada trimestre, considerou-se o abatimento de 0,5% no retorno das carteiras (DAI; ZHU, 2020; NEELY *et al.*, 2014; ZHANG *et al.*, 2019).

É comum que as pesquisas de formação de portfólios desconsiderem além dos custos de transação a perda do poder de compra do dinheiro com o passar do tempo, ou seja, a inflação. Ações e títulos competem por espaço na carteira dos investidores. Se o rendimento dos títulos aumenta, o rendimento ajustado ao risco das ações também deve aumentar, para manter a competitividade das ações. Historicamente, a principal influência nos rendimentos nominais dos títulos tem sido a taxa de inflação. (CAMPBELL; VUOLTEENAHO, 2004). Portanto, optou-se por elaborar uma tabela considerando essas informações. Como todos os valores estão em dólares, o que naturalmente já traz uma proteção aos investimentos por ser uma moeda forte, utilizou-se o índice de referência para inflação americana, o *Consumer Price Index* (CPI). A Tabela 38 mostra os valores das melhores carteiras ajustadas pela inflação e pelos custos de transação. Para os índices, não obstante existirem custos, não faria

sentido incluir os mesmos custos de transações das ações, motivo pelo qual se considerou apenas a inflação.

**Tabela 38 – Totalização das carteiras considerando custos de transações e inflação**

<b>Retornos</b>	<b>LS1</b>	<b>LS2</b>	<b>LS3</b>	<b>S&amp;P</b>	<b>WTICO</b>	<b>DJIA</b>	<b>NASDAQ</b>	<b>OZ1D</b>	<b>IBOV</b>	<b>LFT</b>	<b>T-BOND</b>
Total	1855%	2284%	2957%	118%	-51%	114%	235%	73%	-41%	20%	-53%
Líqu. custos	1475%	1821%	2365%								
Líqu. inflação	1519%	1872%	2417%	75%	-61%	73%	170%	41%	-54%	-3%	-62%
Líquido	1203%	1488%	1928%	75%	-61%	73%	170%	41%	-54%	-3%	-62%

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: Total representa o valor total sem descontos. Líquido de custos é o valor total com o abatimento dos custos de transações. Líquido inflação é o valor total com o abatimento da inflação. Líquido é o valor total com o abatimento dos custos de transação e da inflação.

Os custos apresentados na Tabela 38 foram considerados tão somente para as carteiras LS, e, dessa forma, os indicadores não sofreram alterações na linha de custos. No entanto, assim como as carteiras de ações, a inflação também afeta os indicadores da mesma maneira. Apesar de a “perda” ser substancial, os resultados das carteiras LS continuam maiores que todos os indicadores. Em contrapartida, houve prejuízo para os investidores que alocaram seus recursos em títulos ou ETFs, que replicam os indicadores WTI, IBOV, LFT e T-BOND. Ressalta-se, mais uma vez, que todos os indicadores expostos na Tabela 38 estão em dólar, inclusive o IBOV e a LFT.

A partir do momento em que os modelos são colocados em operação no mercado real, os investidores passam a se interessar em reduzir o risco de inflação em suas carteiras. (PARIKH; MALLADI; FABOZZI, 2020). Investimentos de longo prazo são muito sensíveis ao risco de inflação e também a outros riscos emergentes, por exemplo, a Covid-19. (BELLALAH *et al.*, 2020). Conforme destacado anteriormente, o ouro é visto como um “porto seguro” e serve como proteção em momentos de incerteza. Além disso, é utilizado também como proteção contra a inflação (BAUR; LUCEY, 2010; JONES; SACKLEY, 2016). Nessa lógica, sugere-se que um percentual de alocação em ouro possa gerar benefícios para os investidores, principalmente em mercados mais voláteis como o brasileiro.

As carteiras foram inicialmente formuladas e apresentadas na mesma moeda, ou seja, em dólar. Desse modo, foi possível isolar as diferenças provenientes da variação das moedas e comparar as carteiras sob uma mesma perspectiva. Posto isso, foi possível observar, em consonância com o que outros estudos já revelaram, que o mercado brasileiro é mais volátil e apresenta melhor retorno sem ajuste ao risco, quando comparado ao mercado americano. A maior volatilidade se dá principalmente devido à baixa quantidade de ações listadas, grande número de investidores institucionais, baixa participação de pessoas físicas e

pouca maturidade do mercado brasileiro. Assim, com o objetivo de trazer os retornos mais próximo da realidade do mercado, converteu-se o investimento das ações brasileiras para o real. Posteriormente, apresentou-se ainda os valores deflacionados.

Por meio da Tabela 32, foi apresentada a carteira com ações brasileiras com retornos computados em reais. A seguir resume-se essa tabela novamente levando em conta, assim como na Tabela 38, os custos de transações e inflação, o que pode ser observado na Tabela 39. Foram considerados os mesmos custos de transações considerados no mercado americano, no entanto, no que se refere à inflação, foi utilizado o Índice de Preços para o Consumidor Amplo (IPCA). Os custos foram contemplados unicamente para as carteiras de ações. Apesar de existirem custos também para a alocação em LFT e ainda em ETF, que representa o IBOV, eles são quase que irrelevantes quando comparados aos custos das carteiras.

**Tabela 39 – Carteira LS3 em reais considerando custos de transações e inflação**

<b>Retornos</b>	<b>LS3 BR</b>	<b>IBOV</b>	<b>LFT</b>
Total	1915%	44%	199%
Líqu. custos	1522%		
Líqu. inflação	984%	-26%	57%
Líquido	770%	-26%	57%

**Fonte:** Elaborada pelo autor.

Nota: Total representa o valor total sem descontos. Líquido de custos é o valor total com o abatimento dos custos de transações. Líquido inflação é o valor total com o abatimento da inflação. Líquido é o valor total com o abatimento dos custos de transação e da inflação.

Conforme apresentado na Tabela 39, nota-se que o valor corrigido pelo IPCA causa um impacto maior que as carteiras corrigidas pelo CPI. Isso já era esperado, visto que o IPCA acumulado do primeiro trimestre de 2008 até o segundo trimestre de 2019 foi de 95%, enquanto o CPI para o mesmo período foi de 22%. Reforça-se, portanto, a importância de se manter posições diversificadas em moedas fortes, por exemplo, o dólar, bem como a alocação de ativos de proteção, por exemplo, o ouro. Além disso, a diversificação internacional tem apresentado ganhos em retornos e minimização de riscos.

Notou-se que, mesmo a LFT tendo apresentado um rendimento estável e elevado no período, a inflação foi suficiente para comprometer substancialmente os ganhos. A carteira brasileira alocada em real apresentou retorno acumulado líquido de 770% no período, considerando o abatimento de inflação e custos de transações. O IBOV, quando descontada a inflação, apresentou resultado negativo no período.

Na Tabela 40, consolidam-se as principais carteiras utilizadas no trabalho, inclusive com as *longs* separadas das *shorts*, mostrando também se há diferença nas médias

com os principais indicadores. A carteira LS3 em reais não é apresentada nessa análise em função de estar alocada em uma moeda diferente das demais carteiras. Algumas diferenças de médias já foram mostradas anteriormente, no entanto pretende-se consolidar todas elas novamente a fim de proporcionar uma ampla visão de todas as carteiras trabalhadas nesta tese.

**Tabela 40 – Consolidado de todas as carteiras com as diferenças de médias**

carteiras	sd	ret	média	Sharpe	diferença de médias (z)							
					500	WTI	DJIA	NAS	OZ	IB	LFT	BO
Sem restrição de data												
USA BR long s/stop	14,2%	106%	3,0%	21%	1,36	1,26	1,57	0,56	<b>2,37</b>	1,29	<b>2,38</b>	<b>1,97</b>
USA BR long c/stop	9,5%	1196%	6,1%	65%	<b>2,82</b>	<b>2,14</b>	<b>3,02</b>	<b>1,85</b>	<b>3,61</b>	<b>2,32</b>	<b>3,51</b>	<b>2,79</b>
USA BR short s/stop	15,8%	-6%	1,3%	8%	0,44	0,48	0,65	0,05	0,88	0,63	1,03	0,98
USA BR short c/stop	11,8%	2291%	7,7%	66%	<b>3,40</b>	<b>3,06</b>	<b>3,46</b>	<b>2,78</b>	<b>3,87</b>	<b>2,94</b>	<b>4,01</b>	<b>3,52</b>
USA BR LS s/stop	10,6%	104%	2,2%	74%	0,17	0,67	0,34	0,70	1,52	0,62	1,22	1,36
USA BR LS s/stop c/BOND	9,6%	48%	1,3%	59%	0,59	0,32	0,40	1,43	0,68	0,18	0,42	0,99
USA BR LS s/stop c/LFT	7,9%	87%	1,7%	65%	0,51	0,51	0,28	1,35	1,05	0,21	0,73	1,27
USA BR LS c/stop	7,4%	1855%	6,9%	94%	<b>2,44</b>	<b>2,22</b>	<b>2,74</b>	1,48	<b>3,66</b>	<b>2,19</b>	<b>3,60</b>	<b>3,04</b>
USA BR LS c/stop c/BOND	7,5%	574%	4,5%	76%	1,22	1,44	1,48	0,23	<b>2,44</b>	1,36	<b>2,23</b>	<b>2,13</b>
USA BR LS c/stop c/LFT	6,2%	719%	4,9%	85%	1,46	1,57	1,57	0,37	<b>2,91</b>	1,45	<b>2,61</b>	<b>2,32</b>
Com restrição de data												
USA BR long s/stop	13,3%	169%	3,5%	26%	<b>1,98</b>	1,70	<b>2,20</b>	1,27	<b>2,72</b>	<b>1,80</b>	<b>2,71</b>	<b>2,34</b>
USA BR long c/stop	8,2%	1316%	6,2%	76%	<b>3,06</b>	<b>2,42</b>	<b>3,34</b>	<b>2,23</b>	<b>3,69</b>	<b>2,57</b>	<b>3,72</b>	<b>3,16</b>
USA BR short s/stop	16,2%	51%	2,4%	15%	1,08	1,06	1,25	0,71	1,39	1,10	1,50	1,43
USA BR short c/stop	12,0%	3065%	8,4%	70%	<b>3,92</b>	<b>3,43</b>	<b>4,03</b>	<b>3,38</b>	<b>4,15</b>	<b>3,33</b>	<b>4,34</b>	<b>3,85</b>
USA BR LS s/stop	10,0%	201%	3,0%	30%	1,04	0,94	1,29	0,16	<b>2,09</b>	1,00	<b>2,03</b>	1,63
USA BR LS s/stop c/BOND	9,1%	93%	1,9%	21%	0,07	0,45	0,24	0,89	1,11	0,45	0,83	1,14
USA BR LS s/stop c/LFT	7,4%	142%	2,2%	30%	0,02	0,68	0,22	0,84	1,25	0,49	1,02	1,42
USA BR LS c/stop	6,4%	2284%	7,3%	114%	<b>3,19</b>	<b>2,41</b>	<b>3,45</b>	<b>2,22</b>	<b>4,20</b>	<b>2,58</b>	<b>4,17</b>	<b>3,26</b>
USA BR LS c/stop c/BOND	6,8%	675%	4,8%	70%	1,69	1,49	<b>1,98</b>	0,54	<b>2,84</b>	1,58	<b>2,59</b>	<b>2,26</b>
USA BR LS c/stop c/LFT	5,5%	837%	5,1%	93%	<b>1,77</b>	<b>1,73</b>	<b>2,00</b>	0,70	<b>3,07</b>	1,62	<b>2,83</b>	<b>2,55</b>
Somente brasileiras												
BR long s/stop	24,5%	1306%	8,0%	33%	1,48	1,43	<b>1,64</b>	0,91	<b>1,95</b>	1,60	<b>1,86</b>	<b>1,92</b>
BR long c/stop	23,5%	3324%	9,8%	42%	<b>2,12</b>	<b>2,14</b>	<b>2,29</b>	1,56	<b>2,68</b>	<b>2,26</b>	<b>2,68</b>	<b>2,59</b>
BR short s/stop	17,1%	635%	5,6%	33%	<b>1,82</b>	<b>1,98</b>	<b>1,93</b>	1,63	<b>2,05</b>	<b>1,93</b>	<b>2,25</b>	<b>2,18</b>
BR short c/stop	16,0%	1479%	7,2%	45%	<b>2,81</b>	<b>2,89</b>	<b>2,94</b>	<b>2,60</b>	<b>2,94</b>	<b>2,84</b>	<b>3,12</b>	<b>3,09</b>
BR LS s/stop	14,7%	1327%	6,8%	46%	0,95	1,59	1,10	0,28	<b>2,09</b>	1,47	<b>1,95</b>	<b>2,13</b>
BR LS s/stop c/BOND	9,5%	512%	4,4%	47%	0,59	1,28	0,71	0,05	1,43	1,05	1,34	1,84
BR LS s/stop c/LFT	10,6%	595%	4,8%	45%	0,19	1,35	0,38	0,48	<b>1,66</b>	0,98	1,46	<b>1,91</b>
BR LS c/stop	14,1%	2957%	8,5%	60%	<b>1,96</b>	<b>2,26</b>	<b>2,10</b>	1,10	<b>3,11</b>	<b>2,09</b>	<b>2,90</b>	<b>2,83</b>
BR LS c/stop c/BOND	9,1%	920%	5,6%	61%	1,16	<b>1,65</b>	1,33	0,48	<b>2,19</b>	1,46	<b>2,03</b>	<b>2,25</b>
BR LS c/stop c/LFT	10,2%	1057%	5,9%	58%	0,92	<b>1,65</b>	1,19	0,14	<b>2,64</b>	1,36	<b>2,41</b>	<b>2,34</b>

Nível de confiança de 90%  $Z > 1,64$

Nível de confiança de 95%  $Z > 1,96$

Nível de confiança de 99%  $Z > 2,58$

**Fonte:** Elaborada pelo autor a partir de dados do *output* do *software* Stata.

Nota: Resultados em negrito destacam valores estatisticamente significantes.

De maneira geral, observa-se que os melhores resultados foram trazidos pelas carteiras com critério de *stop loss*, o que mostra a importância de delimitar um limite de perdas. Percebe-se ainda que há carteiras que, mesmo sem o *stop loss*, apresentaram resultados estatisticamente acima dos indicadores utilizados para comparação. Como exemplo, no contexto das carteiras sem restrição de data, a carteira “USA BR *long s/stop*” apresentou retorno de 106% no período, sendo esse resultado estatisticamente maior que o retorno do ouro, da LFT e do T-BOND.

No contexto das ações com restrição de data de divulgação dos resultados, duas carteiras sem *stop loss* se destacaram: “USA BR *long s/stop*”, retornando 169% no período e estatisticamente acima da S&P 500, DJI, OURO, Ibov, LFT e T-BOND; “USA BR LS *s/stop*”, retornando 201% no período e estatisticamente acima do ouro e da LFT.

No que se refere às carteiras formadas somente com ações brasileiras, quatro carteiras sem *stop loss* se sobressaíram: “BR *long s/stop*”, com retorno de 1.306% e estatisticamente acima do DJI, do ouro, da LFT e do T-BOND; “BR *short s/stop*”, com retorno de 635% e estatisticamente acima de todos os indicadores, com exceção do índice NASDAQ; “BR LS *s/stop*”, com retorno de 1.327% e estatisticamente acima do ouro, da LFT e do T-BOND; “BR LS *s/stop c/LFT*”, com 595% de retorno e resultado estatisticamente acima do ouro e do T-BOND.

Considerando todas as carteiras, o melhor retorno foi apresentado pela carteira “BR *long c/stop*”, com retorno de 3.324% e com médias estatisticamente superiores a todos os indicadores, exceto ao índice NASDAQ. Quanto ao resultado ajustado ao risco, nota-se, conforme já mencionado, destaque para a carteira “USA BR LS *c/stop*”, com retorno ajustado ao risco de 114% e retorno acumulado de 2.284%. O resultado é estatisticamente superior a todos os indicadores.

Assim, o objetivo do trabalho foi alcançado mostrando ser possível construir carteiras de investimentos utilizando modelos hierárquicos não lineares de regressão. Validaram-se ainda todas as hipóteses desta pesquisa:

H1: A construção de um portfólio utilizando o modelo proposto vai gerar retornos maiores e com menor risco que o principal índice de ações do mercado brasileiro, o Ibovespa.

H2: A construção de um portfólio utilizando o modelo proposto vai gerar retornos maiores e com menor risco que o principal índice de ações do mercado americano, o S&P 500.

H3: A construção de um portfólio utilizando o modelo proposto vai gerar retornos maiores que os principais investimentos de baixo risco, como CDBs e títulos do tesouro.

Os trabalhos dos autores apresentados no decorrer deste estudo foram construídos a partir de diferentes metodologias e amostras, com poucas similaridades com a presente tese, o que torna arriscado comparar os resultados. No entanto, de maneira geral, notou-se maior retorno que o trabalho de Hou *et al.* (2020) e maior acurácia que o de Ratto *et al.* (2018).

No próximo capítulo, são apresentadas as conclusões, as limitações e as sugestões da pesquisa.

## 5 CONCLUSÕES, LIMITAÇÕES E SUGESTÕES

O objetivo principal deste estudo foi desenvolver um modelo de seleção de ações para compor portfólios *long and short* utilizando modelos hierárquicos não lineares de regressão para dados em painel. Trabalhou-se com as hipóteses de que as carteiras construídas a partir desses modelos podem trazer retornos maiores e com menores riscos que os principais índices de mercado. A utilização em conjunto de elementos como modelos hierárquicos não lineares, alocação *long and short*, diversificação internacional e técnicas de *stop loss* proporcionou os resultados alcançados e a validação das hipóteses H1, H2 e H3.

Foram levantados estudos sobre criação de carteiras de investimentos empregando diferentes modelos estatísticos. No entanto, não foram encontrados estudos que utilizassem o modelo proposto na criação de carteiras. Por meio de um período de treino e outro de teste, simulou-se diferentes modelos de carteira, e em todos eles foi possível obter retornos maiores que os índices utilizados para fins de comparação. O risco, por sua vez, mostrou-se menor na maioria dos casos.

Após não obter êxito na construção das carteiras com base em variáveis endógenas e exógenas, optou-se por manter exclusivamente as exógenas que trouxeram melhores resultados. A única variável endógena que fez parte do modelo foi o retorno das ações. Cabe destacar que não foram apresentados os resultados de diversas variáveis testadas, por exemplo, ROA, ROE, *dividend yield* e *dividend payout*, valor do ativo, valor de mercado, entre outras comumente encontradas em estudos do tipo, pois se optou por apresentar os resultados somente daquelas que funcionaram para o modelo final.

Conforme destacado nesta tese, observou-se ainda que, no intervalo pré-criese, as empresas não apresentaram diferenças estatisticamente significantes entre elas. Desse modo, pode-se inferir que, nesse período, haveria a possibilidade de escolher apenas um setor para obter resultados similares nas taxas de retornos, não sendo necessário optar por uma ou outra empresa.

No que se refere às carteiras, notou-se que os melhores resultados foram demonstrados nas carteiras criadas a partir do modelo hierárquico, com delimitação de perdas por meio do *stop loss* e sem a inserção de um título de renda fixa. Assim, tendo em vista que as carteiras com os melhores retornos ajustados ao risco não contavam com um título de renda fixa, o trabalho coloca em dúvida o poder da renda fixa para diversificação e minimização de

riscos. Esse comportamento se estende a diversos cenários dos âmbitos nacional, internacional e misto ao utilizar tanto T-BONDS quanto LFTs.

Conforme visualizado por intermédio dos Gráfico 4, 6 e 7, constatou-se o comportamento parecido das carteiras LS1, LS2 e LS3, mesmo sendo formadas em diferentes bolsas. Vale lembrar que o modelo de base para seleção das ações foi composto pelos retornos defasados das próprias ações, além da variação do ouro, do S&P 500 e do T-BOND 10. Isso demonstra que, de maneira geral, os mercados nacionais e internacionais caminham em direções parecidas.

Uma das limitações do trabalho é que não foi utilizada nenhuma estratégia visando otimizar a diversificação, com exceção da diversificação internacional e por meio dos títulos, ou seja, todas as carteiras tiveram os ativos distribuídos em pesos iguais, independentemente da probabilidade de retornos positivos. Nesse sentido, duas sugestões são dadas: primeira, utilizar algum dos tradicionais métodos de otimização e diversificação de carteiras, entre eles, média variância e mínima variância; segunda, atribuir peso maior às ações com maior probabilidade de retornos positivos. Dessa forma, uma ação com probabilidade de retorno positivo de 90% deverá ter peso maior que uma ação com probabilidade de retorno positivo de 60%, por exemplo. Outra limitação foi não ter havido nenhum controle para que não houvesse uma concentração grande de alocação dos recursos em determinados setores.

Sugere-se que, em estudos futuros, o modelo seja atualizado trimestralmente, e o resultado apurado na sequência, isto é, somente um passo à frente, como testado por Black *et al.* (2014). Nesse caso, o período *out of sample* ficará menor, sendo sempre utilizado somente um trimestre nesse formato. No entanto, o modelo de base compreenderá o máximo de atualizações possíveis. Sugere-se ainda a utilização de outros períodos *in sample* e *out of sample*, bem como diferentes níveis de *cutoff*. Observou-se, que cada nível e cada período geram diferentes resultados, e isso denota que naturalmente os resultados também são diferentes.

Por fim, sugere-se testar os modelos híbridos que recentemente têm ganhado destaque em estudos acadêmicos. Tais trabalhos associam dois ou mais modelos de previsão de retornos e têm historicamente apresentado melhores resultados que modelos isolados (HOU *et al.*, 2020; RATTO *et al.*, 2018; WANKHADE *et al.*, 2013; ZHANG, 2003). Complementarmente ao *stop loss* utilizado na presente tese e indo ao encontro da utilização de modelos híbridos, sugere-se o uso de modelos que avaliem diariamente ou semanalmente as tendências do mercado e que possam prever a mudança de direção antes do acionamento do



*stop loss*. O objetivo disso é minimizar as perdas delimitadas pelo *stop loss* que, neste estudo, foi de 10%. Uma das alternativas largamente estudadas e utilizadas para esse fim se baseia no artigo “*A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle*”, escrito por Hamilton (1989). Além de compor um modelo híbrido que auxilia nos momentos de *stop loss*, o modelo proposto por esse autor pode também ser utilizado exclusivamente para compor portfólios, como os desenvolvidos por Gómez e Jiménez (2020), Kim, Jeong e Lee (2019) e Lewin e Campani (2020).

A principal contribuição do trabalho foi, diante de um mercado de ações em forte ascensão, desenvolver um novo modelo de investimento que viabilize a seleção de ações a partir de modelos hierárquicos não lineares, complementado com técnicas de alocação, diversificação e minimização de risco. Ressalta-se a importância de atualizar o modelo econométrico sempre que novos dados forem disponibilizados pelas empresas, para garantir que ele contemple comportamentos recentes do mercado.

## REFERÊNCIAS

- ABENSUR, E. O. *et al.* Stochastic liquidity model and its applications to portfolio selection. *In: ANISIC, Z.; LALIC, B.; GRACANIN, D. (ed.). Proceedings on 25th International Joint Conference on Industrial Engineering and Operations Management – IJCIEOM: the next generation of production and service systems.* Switzerland: Springer, 2020. p. 42-51. Disponível em: <https://link.springer.com/book/10.1007%2F978-3-030-43616-2>. Acesso em: 18 abr. 2020.
- ALIAGA-DIAZ, R. *et al.* Portfolio Optimization with Active, Passive, and Factors: Removing the Ad Hoc Step. **The Journal of Portfolio Management**, v. 46, n. 4, p. 39-51, 2020.
- ALSHAMMARI, A. A.; ALTARTURI, B.; SAITI, B. The impact of exchange rate, oil price and gold price on the Kuwaiti stock market: a wavelet analysis. **The European Journal of Comparative Economics**, v. 17, n. 1, p. 31-54, 2020.
- ARAÚJO, E.; BRITO, R. D.; SANVICENTE, A. Long-term stock returns in Brazil: volatile equity returns for u.s.-like investors. **SSRN Electronic Journal**, p. 1-37, Mar. 2020.
- ARTUSO, A. R.; CHAVES NETO, A. Identificação de fatores relevantes e construção de portfólio diversificado no mercado acionário brasileiro. **Produção Online**, v. 12, n. 4, p. 1002-1030, 2012.
- ASSAGAF, A. *et al.* The effect of macro economic variables on stock return of companies that listed in stock exchange: empirical evidence from Indonesia. **International Journal of Business and Management**, v. 14, n. 8, p. 108-116, 2019.
- AVRAMOV, D. *et al.* Scaling up market anomalies. **The Journal of Investing**, v. 26, n. 3, p. 89-105, 2017.
- AWA, K. N.; NNAMETU, J.; EMOH, F. I. Global determinants of direct real estate investment returns in Nigeria. **PM World Journal Global Determinants of Direct Real Estate**, v. VIII, n. X, p. 1-13, Nov. 2019.
- BAETJE, F. Does a lot help a lot? Forecasting stock returns with pooling strategies in a data-rich environment. **Journal of Forecasting**, v. 37, n. 1, p. 37-63, 2018.
- BANCO CENTRAL DO BRASIL. **Taxas de juros básicas: histórico.** Disponível em <https://www.bcb.gov.br/controleinflacao/historicotaxasjuros>. Acesso em: 30 nov. 2019.
- BAUR, D. G.; LUCEY, B. M. Is gold a hedge or a safe haven? An analysis of Stocks, Bonds and Gold Dirk. **The Financial Review**, v. 45, p. 217-229, 2010.
- BEAVER, W.; MCNICHOLS, M.; PRICE, R. The costs and benefits of long-short investing: a perspective on the market efficiency literature. **Journal of Accounting Literature**, v. 37, p. 1-18, 2016.
- BELLALAH, M. *et al.* Long term optimal investment with regime switching: inflation, information and short sales. **Annals of Operations Research**, p. 1-14, 2020.

BHAGWAN, J. *et al.* A Q-learning agent for automated trading in equity stock markets. **Expert Systems With Applications**, v. 163, 2021.

BLACK, A. J. *et al.* Forecasting stock returns: do commodity prices help? **Journal of Forecasting**, v. 33, n. 8, p. 627-639, 2014.

BLACK, F.; LITTERMAN, R. Global portfolio optimization. **Financial Analysts Journal**, v. 48, n. 5, p. 28-43, 2006.

BRASIL, BOLSA, BALCÃO. **Histórico pessoas físicas**. Disponível em: [http://www.b3.com.br/pt\\_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/historico-pessoas-fisicas/](http://www.b3.com.br/pt_br/market-data-e-indices/servicos-de-dados/market-data/consultas/mercado-a-vista/historico-pessoas-fisicas/). Acesso em: 30 nov. 2019.

CAMPBELL, J. Y.; VUOLTEENAHU, T. Inflation illusion and stock prices. **American Economic Review**, v. 94, n. 2, p. 19-23, 2004.

CHEN, L. Using machine learning algorithms on prediction of stock price. **Journal of Modeling and Optimization**, v. 12, n. 2, p. 84-99, 2020.

CHEN, Z.; PENG, S.; LISSER, A. A sparse chance constrained portfolio selection model with multiple constraints. **Journal of Global Optimization**, v. 77, n. 4, p. 825-852, Aug. 2020.

CORRAR, L. J.; PAULO, E.; DIAS FILHO, J. M. (coord). **Análise multivariada: para os cursos de administração, ciências contábeis e economia**. São Paulo: Atlas, 2017.

DAI, Z.; ZHU, H. Stock return predictability from a mixed model perspective. **Pacific-Basin Finance Journal**, v. 60, p. 1-18, Jan. 2020.

DAMODARAN, A. **Investment philosophies: successful strategies and the investors who made them work**. 2th ed. Hoboken, NJ: John Wiley & Sons, 2012.

DEMIGUEL, V.; GARLAPPI, L.; UPPAL, R. Optimal versus naive diversification: how inefficient is the 1/N portfolio strategy? **Review of Financial Studies**, v. 22, n. 5, p. 1915-1953, 2009.

EHRHARDT, M. C; BRIGHAM, E. F. **Administração financeira: teoria e prática**. São Paulo: Cengage Learning, 2012.

FAMA, E. F. Market efficiency, long-term returns, and behavioral finance. **Journal of Financial Economics**, v. 49, n. 3, p. 283-306, 1998.

FÁVERO, L. P.; BELFIORE, P. **Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®**. Rio de Janeiro: LTC, 2017.

FONSECA, C. V. C.; SILVEIRA, R. L. F. DA. Governança corporativa e custo de capital de terceiros: evidências entre empresas brasileiras de capital aberto. **REAd: Revista Eletrônica de Administração**, Porto Alegre, v. 22, n. 1, p. 106-133, 2016.

FONTENELE, D. **O comportamento dos fundos quantitativos durante a crise**. 29 mar. 2020. Disponível em: <https://conteudos.xpi.com.br/fundos-de-investimento/relatorios/o-comportamento-dos-fundos-quantitativos-durante-a-crise/>. Acesso em: 25 abr. 2020.

FURLANETTI, C. E. **Análise das características das carteiras acionárias, das estratégias de gestão e dos retornos dos fundos de previdência privada complementar PGBL/VGBL que investem em renda variável**. Orientador: José Roberto Securato. 2008. 142 p. Tese (Doutorado em Ciências) – Universidade de São Paulo, São Paulo, 2018.

GELMAN, A.; HILL, J. **Data analysis using regression and multilevel/hierarchical models**. New York, NY: Cambridge University Press, 2007.

GÓMEZ, JUAN M.; JIMÉNEZ, J. A. Selección óptima de portafolios basada en cadenas de Markov de primer y segundo orden. **Lecturas de Economía**, v. 92, n. 1, p. 33-66, 2020.

GONG, Y.; BU, R.; CHEN, Q. What affects the relationship between oil prices and the U.S. stock market? A mixed-data sampling copula approach. **Journal of Financial Econometrics**, p. 1-25, 2020.

GRINOLD, R. C.; KAHN, R. N. The Efficiency Gains of Long-Short Investing. **Financial Analysts Journal**, v. 56, n. 6, p. 40-53, 2000.

GUERARD, J. B.; XU, G.; WANG, Z. **Portfolio and investment analysis with SAS: financial modeling techniques for optimization**. Cary, CN: SAS Institute, 2019.

GULAY, E.; EMEC, H. Comparison of forecasting performances: does normalization and variance stabilization method beat GARCH(1,1)-type models? Empirical evidence from the stock markets. **Journal of Forecasting**, v. 37, n. 2, p. 133-150, 2018.

GUO, G.; ZHAO, H. Multilevel Modeling for Binary Data. **Annual Review of Sociology**, v. 26, n. 1, p. 441-462, 2002.

HAMILTON, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle. **Econometrica**, v. 57, n. 2, p. 357-384, 1989.

HARVEY, C. R. *et al.* The Best of Strategies for the Worst of Times: Can Portfolios be Crisis Proofed? **The Journal of Portfolio Management**, v. 45, n. 5, p. 1-22, July 2019.

HOU, X. *et al.* An enriched time-series forecasting framework for long-short portfolio strategy. **IEEE Access**, v. 8, p. p. 31992-32002, 2020.

HUANG, J.; CHAI, J.; CHO, S. Deep learning in finance and banking: a literature review and classification. **Frontiers of Business Research in China**, v. 14, n. 1, p. 1-24, 2020.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **População: população do Brasil**. 2020. Disponível em: [https://www.ibge.gov.br/apps/populacao/projecao/index.html?utm\\_source=portal&utm\\_medium=popclock&utm\\_campaign=novo\\_popclock](https://www.ibge.gov.br/apps/populacao/projecao/index.html?utm_source=portal&utm_medium=popclock&utm_campaign=novo_popclock). Acesso em: 25 abr. 2020.

- ISA, M. A. *et al.* Co-movement and causal relationship between stock market and macroeconomic variables: latest evidence from Malaysia. **Advanced International Journal of Banking, Accounting and Finance**, v. 1, n. 1, p. 25-35, Dec. 2019.
- JACOBS, B. I.; LEVY, K. N. Long/short equity investing. **The Journal of Portfolio Management**, v. 20, n. 1, p. 52-63, 1993.
- JACOBS, B. I.; LEVY, K. N.; STARER, D. Long-short portfolio management: an integrated approach. **The Journal of Portfolio Management**, v. 25, n. 2, p. 23-32, 1999.
- JIAO, Y.; MASSA, M.; ZHANG, H. Short selling meets hedge fund 13F: an anatomy of informed demand. **Journal of Financial Economics**, v. 122, n. 3, p. 544-567, 2016.
- JONES, A. T.; SACKLEY, W. H. An uncertain suggestion for gold-pricing models: the effect of economic policy uncertainty on gold prices. **Journal of Economics and Finance**, v. 40, n. 2, p. 367-379, 2016.
- KIM, E.; JEONG, H.; LEE, N. Global Asset allocation strategy using a Hidden Markov Model. **Journal of Risk Financial Management**, v. 1, n. 4, p. 1-15, 2019.
- LANTZ, B. **Machine learning with R**. 3th ed. Birmingham, UK: Packt Publishing, 2019.
- LAU, W.-Y. Private retirement scheme funds: will the asset allocation strategy work for retirees? **Shanlax International Journal of Economics**, v. 8, n. 1, p. 1-14, 2019.
- LEWIN, M.; CAMPANI, C. H. Gestão de carteiras sob múltiplos regimes: estratégias que performam acima do mercado. **Revista de Administração Contemporânea**, v. 24, n. 4, p. 300-316, 2020.
- LI, C.; SONG, Y. Predicting direction of individual stock price movement using a hybrid model. **Journal of Economics, Business and Management**, v. 7, n. 2, p. 60-64, 2019.
- LUU, Q.; NGUYEN, S.; PHAM, U. Time series prediction : a combination of long short-term memory and structural time series models. **Science & Technology Development Journal – Economics - Law and Management**, v. 4, n. 1, p. 500-515, 2020.
- LYLE, M. R.; YOHN, T. L. Fundamental analysis and mean-variance optimal portfolios. **SSRN Electronic Journal**, 2019.
- MAGHYEREH, A.; ABDOH, H. Tail dependence between gold and islamic securities. **Finance Research Letters**, Apr. 2020.
- MANURUNG, A. H.; NATASHA, C.; BUDIARTO, W. Modelling shares choice to enter in a portfolio using artificial neural networks (ANN). **Journal of Big Data**, v. 7, n. 17, p. 1-9, 2020.
- MARKOWITZ, H. Portfolio selection. **The Journal of Finance**, v. 7, n. 1, p. 77-91, 1952.
- MIKUTOWSKI, M.; KAMBOURIS, G. D.; ZAREMBA, A. A note on value investing in the UAE stock market. **Journal of Research in Emerging Markets**, v. 1, n. 2, p. 33-38, 2019.

MONDAY, T. E.; ABDULKADIR, A. Modeling fluctuation of the price of crude oil in Nigeria using. **Asian Journal of Probability and Statistics**, v. 7, n. 1, p. 16-40, 2020.

MOREMA, K.; BONGA-BONGA, L. The impact of oil and gold price fluctuations on the South African equity market: Volatility spillovers and financial policy implications. **Resources Policy**, v. 68, p. 1-10, Jan. 2020.

NAWAZ, M. S.; AZAM, M.; ASLAM, M. Probable daily return on investments in gold. p. **Gold Bulletin**, v. 53, p. 47-54, 2020.

NEELY, C. J. et al. Forecasting the equity risk premium: the role of technical indicators. **Management Science**, v. 60, n. 7, p. 1772-1791, 2014.

NYSE. **NYSE listings**. 2020. Disponível em: <https://www.nyse.com/listings/international-listings>. Acesso em: 22 mar. 2020.

OZDURAK, C.; ULUSOY, V. Price discovery in crude oil markets: intraday volatility interactions between crude oil futures and energy exchange traded funds. **International Journal of Energy Economics and Policy**, v. 10, n. 3, p. 402-413, 2020.

PADMANTYO, S.; PRASOJO, P. Total risk stock investment: does international portfolio diversification give more impact to investors? *In: UNIVERSITY RESEARCH COLLOQUIUM*, 10., 2019, Gombong. **Annals [...]**. Gombong: Muhammadiyah STIKES Gombong, 2019.

PAI, N.; ILANGO, D. V. Neural network model for efficient portfolio management and time series forecasting. *In: INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING AND CONTROL SYSTEMS*, 4., 2020, Madurai. **Annals [...]**. Madurai: Vaigai College of Engineering, 2020.

PARIKH, H.; MALLADI, R. K.; FABOZZI, F. J. Preparing for higher inflation: portfolio solutions using U.S. equities. **Review of Financial Economics**, v. 38, n. 3, p. 542-554, 2020.

POTERBA, J. M.; SHOVEN, J. B. Exchange traded funds: a new investment option for taxable investors. **American Economic Review**, v. 92, n. 2, p. 422-427, May 2002.

RABE-HESKETH, S.; SKRONDAL, A. **Multilevel and longitudinal modeling using stata vol. II: categorical responses, counts, and survival**. 3. ed. College Station: Stata Press Publication, 2012.

RATTO, A. P. *et al.* Ensemble of technical analysis and machine learning for market trend prediction. *In: 2018 IEEE SYMPOSIUM SERIES ON COMPUTATIONAL INTELLIGENCE*, 2018, Bangalore. **Annals [...]**. Bangalore: IEEE Computational Intelligence Society, 2018. p. 2090-2096.

RAUDENBUSH, S. W.; BRYK, A. S. **Hierarchical linear models: applications and data analysis methods**. New York: Sage Publications, 2002. v. 1.

ROCCAZZELLA, F.; GAMBETTIA, P.; VRINS, F. **Optimal and robust combination of forecasts via constrained optimization and shrinkage**. Louvain-la-Neuve: Louvain Finance of Université Catholique de Louvain, 2020.

SAMUEL, S. S. A study on growth of gold ETFs : as an effective investment tool over physical gold. **UGC Care Journal**, v. 40, n. 40, p. 2328-2331, 2020.

SANTOS, M. A. **Hipótese de mercados adaptativos e fatores econômico-institucionais: uma abordagem multinível.** Orientador: Luiz Paulo Lopes Fávero. 2018. 176 p. Tese (Doutorado em Ciências) – Universidade de São Paulo, São Paulo, 2018.

SETIADI, J.; MASDUPI, E. The effect of macroeconomic variables on banking stock price index in Indonesia stock exchange. *In: PADANG INTERNATIONAL CONFERENCE ON EDUCATION, ECONOMICS, BUSINESS AND ACCOUNTING*, 4., 2019, West Sumatera. **Annals [...]**. West Sumatera: Faculty of Economics Universitas Negeri Padang, 2019.

SHABBIR, A.; KOUSAR, A. S. AND S.; BATOOL, S. A. Impact of gold and oil prices on the stock market in Pakistan. **Journal of Economics, Finance and Administrative Science**, p. 47-63, 2019.

SHCHANKINA, A.; PING, Z. An empirical study on the modeling of an optimal investment portfolio using multivariate model of conditional heteroscedasticity: evidence from the chinese stock exchanges. *In: INTERNATIONAL SCIENTIFIC CONFERENCE “FAR EAST CON” (ISCFEC 2020).* **Annals [...]**. 2020.

SIMON, S. **Developing a new stock scoring model for shariah: compliant investment.** Thesis. 2019. (Doctor of Philosophy in Islamic Finance) – University of Bolton, Bolton, 2019.

SINGH, J. K. Portfolio management: importance and challenges. **UGC Care Journal**, v. 40, n. 71, p. 113-117, 2020.

TALEB, N. N. **Fooled by randomness: the hidden role of chance in life and in the markets.** 2. ed. New York: Random House Publishing Group, 2005.

UNITED NATIONS POPULATION FUND. **World Population Dashboard: United States of America.** Disponível em: <https://www.unfpa.org/data/world-population/US>. Acesso em: 30 nov. 2019.

UNITED STATE CENSUS BUREAU. U.S. and World Population Clock. Disponível em: <https://www.census.gov/popclock/>. Acesso em: 11 abr. 2020.

VEZERIS, D.; KARKANIS, I.; KYRGOS, T. AdTurtle: an advanced turtle trading system. **Journal of Risk and Financial Management**, v. 12, n. 2, p. 1-52, 2019.

WANG, H. Stock price prediction based on machine learning approaches. *In: INTERNATIONAL CONFERENCE ON DATA SCIENCE AND INFORMATION TECHNOLOGY*, 3., 2020, New York. **Annals [...]**. New York: Association for Computing Machinery, 2020. p. 1-5.

WANG, M.; XU, P.; LIU, J. **Application of machine learning technology in financial market.** *In: INTERNATIONAL CONFERENCE ON BIG DATA APPLICATION & ECONOMIC MANAGEMENT*, 2020, Guiyang. **Annals [...]**. Guiyang: Guizhou University of Finance and Economics, 2020.

WANG, S. *et al.* Stock selection strategy of A-share market based on rotation effect and random forest. **AIMS Mathematics**, v. 5, n. 5 , p. 4563-4580, Apr. 2020.

WANG, Y. *et al.* Momentum of return predictability. **Journal of Empirical Finance**, v. 45, p. 141-156, Jan. 2018.

WANKHADE, S. B. *et al.* Hybrid model based on unification of technical analysis and sentiment analysis for stock price prediction. **International Journal of Computers & Technology**, v. 11, n. 9, p. 3025-3033, 2013.

WELCH, I.; GOYAL, A. A comprehensive look at the empirical performance of equity premium prediction. **Review of Financial Studies**, v. 21, n. 4, p. 1455-1508, 2008.

WEST, B.; WELCH, K.; GAŁECKI, A. **Linear mixed models: a practical guide using statistical software**. 2. ed. Boca Raton: CRC Press, 2014 .

WU, K. *et al.* Can crude oil drive the co-movement in the international stock market? Evidence from partial wavelet coherence analysis. **North American Journal of Economics & Finance**, v. 53, p. 1-17, Jul. 2020.

YANG, K.; QIAN, E.; BELTON, B. Protecting the downside of trend when it is not your friend. **The Journal of Portfolio Management**, p. 1-13, Jul. 2019.

YU, J.-R. *et al.* Portfolio models with return forecasting and transaction costs. **International Review of Economics and Finance**, v. 66, p. 118-130, Mar. 2019.

ZAREMBA, A. *et al.* Picking winners to pick your winners: the momentum effect in commodity risk factors. **North American Journal of Economics and Finance**, v. 50, p. 1-13, Nov. 2019.

ZHANG, P. G. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, v. 50, p. 159-175, 2003.

ZHANG, Y. *et al.* Economic constraints and stock return predictability: a new approach. **International Review of Financial Analysis**, v. 63, p. 1-9, May 2019.

ZHANG, Y.; CHU, G.; SHEN, D. The role of investor attention in predicting stock prices: the long short-term memory networks perspective. **Finance Research Letters**, p. 1-11, 2020.