

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE CIÊNCIAS ECONÔMICAS
DEPARTAMENTO DE CIÊNCIAS ADMINISTRATIVAS
CENTRO DE PÓS-GRADUAÇÃO E PESQUISAS EM ADMINISTRAÇÃO

**O EFEITO DO RISCO DE INFORMAÇÃO ASSIMÉTRICA SOBRE O RETORNO DE
AÇÕES NEGOCIADAS NA BM&FBOVESPA**

LEONARDO SOUZA SIQUEIRA

BELO HORIZONTE

2017

LEONARDO SOUZA SIQUEIRA

**O EFEITO DO RISCO DE INFORMAÇÃO ASSIMÉTRICA SOBRE O RETORNO DE
AÇÕES NEGOCIADAS NA BM&FBOVESPA**

Dissertação apresentada ao Centro de Pós-Graduação e Pesquisas em Administração da Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Administração.

Linha de pesquisa - Finanças

Orientador - Prof. Dr. Hudson Fernandes Amaral

Coorientadora - Prof^a. Dr^a. Laise Ferraz Correia

BELO HORIZONTE

2017

Ficha catalográfica

S617e
2017 Siqueira, Leonardo Souza.
O efeito do risco de informação assimétrica sobre o retorno de ações negociadas na BM&FBOVESPA [manuscrito] / Leonardo Souza Siqueira, 2017.
166 f.: il., gráfs. e tabs.

Orientador: Hudson Fernandes Amaral.

Coorientador: Laise Ferraz Correia.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Centro de Pós-Graduação e Pesquisas em Administração.

Inclui bibliografia (f. 151-161) e apêndices.

1. Mercado de ações – Previsão – Teses. 2. Bolsa de valores – Teses. 3. Informação assimétrica – Teses. I. Amaral, Hudson Fernandes. II. Correia, Laise Ferraz. IV. Universidade Federal de Minas Gerais. Centro de Pós-Graduação e Pesquisas em Administração. V. Título.

CDD: 332.642

Elaborada pela Biblioteca da FACE/UFMG. – NMM/015/2017



Universidade Federal de Minas Gerais
Faculdade de Ciências Econômicas
Departamento de Ciências Administrativas
Centro de Pós-Graduação e Pesquisas em Administração

ATA DA DEFESA DE DISSERTAÇÃO DE MESTRADO EM ADMINISTRAÇÃO do Senhor **LEONARDO SOUZA SIQUEIRA**, REGISTRO N° 600/2017. No dia 03 de fevereiro de 2017, às 14:00 horas, reuniu-se na Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais - UFMG, a Comissão Examinadora de Dissertação, indicada pelo Colegiado do Centro de Pós-Graduação e Pesquisas em Administração do CEPEAD, em 24 de janeiro de 2017, para julgar o trabalho final intitulado "**O EFEITO DO RISCO DE INFORMAÇÃO ASSIMÉTRICA SOBRE O RETORNO DE AÇÕES NEGOCIADAS NA BM&FBOVESPA**", requisito para a obtenção do **Grau de Mestre em Administração**, linha de pesquisa: **Finanças**. Abrindo a sessão, o Senhor Presidente da Comissão, Prof. Dr. Hudson Fernandes Amaral, após dar conhecimento aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Logo após, a Comissão se reuniu sem a presença do candidato e do público, para julgamento e expedição do seguinte resultado final:

APROVAÇÃO;

() APROVAÇÃO CONDICIONADA A SATISFAÇÃO DAS EXIGÊNCIAS CONSTANTES NO VERSO DESTA FOLHA, NO PRAZO FIXADO PELA BANCA EXAMINADORA (NÃO SUPERIOR A 90 NOVENTA DIAS);

() REPROVAÇÃO.

O resultado final foi comunicado publicamente ao candidato pelo Senhor Presidente da Comissão. Nada mais havendo a tratar, o Senhor Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora. Belo Horizonte, 03 de fevereiro de 2017.

NOMES

ASSINATURAS

Prof. Dr. Hudson Fernandes Amaral.....
ORIENTADOR (CEPEAD/UFMG)

Prof^a. Dr^a. Laise Ferraz Correia.....
COORIENTADORA (DCSA/CEFET)

Prof. Dr. Bruno Pérez Ferreira.....
(CEPEAD/UFMG)

Prof. Dr. José Roberto de Souza Francisco.....
(CIC/UFMG)

Prof. Dr. Robert Aldo Iquiapaza Coaguila.....
(CEPEAD/UFMG)

AGRADECIMENTOS

Ao meu pai Claudio, à minha mãe Andreia e ao meu irmão Filipe, minha amada família.

Ao Prof. Dr. Hudson Fernandes Amaral e à Profa. Dra. Laise Ferraz Correia, pela orientação e amizade nesses dois anos de crescimento pessoal e profissional.

Aos professores Prof. Dr. Aureliano Bressan, Prof. Dr. Robert Iquiapaza, Prof. Dr. Bruno Pérez e Prof. Dr. Marcos Antônio Camargos, pelas aulas no mestrado e incessante transmissão de conhecimento nesses dois anos. Ao Prof. Dr. Luiz Alberto Bertucci (*in memoriam*), por ter me dado a honra de ter sido seu aluno na graduação.

À UFMG, minha casa nos últimos sete anos e minha casa pelos próximos quatro.

À Júlia, minha companheira e suporte nos momentos difíceis.

Aos meus caros colegas de mestrado.

RESUMO

O risco informacional é tema central de discussões sobre precificação de ativos desde o trabalho seminal de Fama (1970). A forma de mensuração da probabilidade de negociações privilegiadas ganhou força a partir dos trabalhos de Easley e O'Hara (1987) e Easley *et al.* (1996). Visando verificar se existe uma parte dos retornos de portfólios no mercado brasileiro que pode ser explicada pelo seu conteúdo informacional, utilizou-se a métrica VPIN, de Easley, López de Prado e O'Hara (2011a), calculada através dos dados reais obtidos pelo *market data* da BM&FBovespa. Esta métrica foi aplicada juntamente aos modelos de três e cinco fatores de Fama e French (1993, 2015) e Carhart (1997). Foi criado um fator VPIN relacionado ao risco informacional das ações. Aplicando os testes GRS e *Average F-test* e através do método de simulação de dados *bootstrap* verificou-se que a combinação de fatores que melhor explica os retornos dos portfólios criados foi aquele que se valeu dos fatores de mercado, tamanho, lucratividade, investimento e informação. Assim como Fama e French (2015), observou-se que o fator HML se torna redundante no cenário nacional. Além disso, verificou-se que o fator relacionado ao risco informacional funciona como um complemento ao fator tamanho e que a inclusão deste leva a uma melhora no desempenho dos modelos, indicando um possível poder de explicação do conteúdo informacional sobre o retorno dos portfólios utilizados neste trabalho.

Palavras-chave: PIN, VPIN, risco informacional, *asset pricing*, fatores de precificação.

ABSTRACT

The Informational risk has been at the center of discussion regarding asset pricing since the seminal work of Fama (1970). Measuring the probability of informed trading gained strength with the works of Easley and O'Hara (1987) and Easley *et al.* (1996). In order to verify if there's a part of the portfolios' returns in the brazilian stock market that can be explained by its informational content, this study used the VPIN metric, developed by Easley, López de Prado and O'Hara (2011a), calculated with real data obtained by the BM&FBovespa's market data. This metric was applied alongside the three and five factors models of Fama and French (1993, 2015) and the Carhart (1997) four factor model. A VPIN factor related to the informational risk of the assets was created. By applying the GRS and Average F-test and the bootstrap simulation method it was found that the combination of factors that best explains the required returns of portfolios was the market, size, profitability, investment and informational factors. As in the Fama and French (2015) study, the HML factor became redundant in the analysis of the brazilian stock market. The results suggest the informational factor works as a complement to the size factor. It was also documented that the information risk factor increases the models' performance, which indicates that the information content has possibly a power of explanation on the portfolios' returns analyzed in this study.

Keywords: PIN, VPIN, informational risk, asset pricing, pricing factors.

LISTA DE ILUSTRAÇÕES

1. Lista de figuras

Figura 1 – Diagrama da árvore do processo de negociação	16
---	----

2. Lista de gráficos

Gráfico 1 – Distribuição do VPIN por intervalos.....	82
Gráfico 2 – Distribuição dos parâmetros simulados – média e desvio padrão.....	86
Gráfico 3 – Evolução do VPIN diário dos segmentos.....	90
Gráfico 4 – Evolução do VPIN diário para os grupos de tamanho das ações	94

LISTA DE TABELAS

Tabela 1 – Amostra utilizada na pesquisa	49
Tabela 2 – Modelos testados e suas respectivas bases teóricas	59
Tabela 3 – Informações acerca dos portfólios criados	61
Tabela 4 – Carteiras formadas a partir dos fatores tamanho e <i>book-to-market</i>	63
Tabela 5 – Carteiras formadas a partir dos fatores tamanho e momento.	65
Tabela 6 – Carteiras formadas a partir dos fatores tamanho, <i>book-to-market</i> , lucratividade e investimento	66
Tabela 7 – Carteiras formadas com base nas variáveis tamanho e VPIN	67
Tabela 8 – Desempenho dos métodos TR e BVC para a amostra irrestrita	74
Tabela 9 – Desempenho dos métodos TR e BVC para a amostra definida.....	75
Tabela 10 – Estatísticas descritivas dos VPINs de todas as ações	79
Tabela 11 – Estatísticas descritivas do VPIN para toda a amostra.....	78
Tabela 12 – Intervalo de confiança dos parâmetros do VPIN calculados a partir do método <i>bootstrap</i>	85
Tabela 13 – Estatísticas descritivas do VPIN diário por segmento de listagem da BM&FBovespa.....	88
Tabela 14 – Teste Jarque-Bera para a normalidade dos VPINs diários dos segmentos de listagem da BM&FBovespa.....	89
Tabela 15 – Teste t para a diferença entre as médias dos VPINs das ações por segmento de listagem da BM&FBovespa.....	89
Tabela 16 – Estatísticas do VPIN diário por grupo relacionado ao tamanho das empresas.....	92
Tabela 17 – Comparação entre os VPINs calculados por grupos de tamanho em diferentes mercados	93
Tabela 18 – Correlação entre o VPIN e o <i>bid-ask spread</i> das ações.....	99
Tabela 19 – Regressão MQO do <i>bid-ask spread</i> sobre o VPIN, testes para heterocedasticidade e autocorrelação e estimação dos coeficientes por Newey-West.....	100
Tabela 20 – Estimação do Intercepto e do Coeficiente Angular da regressão $BidAsk_{it} = \alpha + \beta VPIN_{it} + \varepsilon$, para cada ação, onde i se refere ao indexador do ativo i e t se refere aos dias, $t = 1, \dots, 512$	102
Tabela 21 – Estatísticas descritivas dos três, quatro e cinco fatores de Fama e French (1993, 2015) e Carhart (1997) e o fator IMU	107

Tabela 22 – Correlação entre os três, quatro e cinco fatores de Fama e French (1993, 2015) e Carhart (1997) e o fator IMU	109
Tabela 23 – Média dos fatores formadores dos portfólios, número de ações e média dos retornos das carteiras.....	112
Tabela 24 – Retornos das carteiras formadas por diferentes variáveis.....	114
Tabela 25 – Resultados do teste VIF de multicolinearidade das variáveis explicativas	116
Tabela 26 – Coeficientes para os modelos de 3 Fatores e 3 Fatores adicionado o IMU	120
Tabela 27 - Coeficientes para os modelos de 4 Fatores e 4 Fatores adicionado o IMU.....	124
Tabela 28 - Coeficientes para os modelos de 5 Fatores e 5 Fatores adicionado o IMU.....	128
Tabela 29 – Teste GRS para os modelos de três, quatro e cinco fatores acrescidos do fator IMU	132
Tabela 30 – Regressão dos três fatores de Fama e French (1993) e o IMU	135
Tabela 31 - Regressão dos quatro fatores de Carhart (1997) e o IMU	136
Tabela 32 - Regressão dos cinco fatores de Fama e French (2015) e o IMU.....	137
Tabela 33 – Teste GRS para as variações dos modelos de três, quatro e cinco fatores e o fator IMU	139
Tabela 34 – P.valores do <i>Average F-test</i> para diferentes combinações de fatores.....	144
Tabela 35 – Resultado da combinação dos p.valores através do método de Fisher	145
Tabela 36 – Amostra de ações passivas para composição dos fatores de Fama e French (1993, 2015) e Carhart (1997)	162
Tabela 37 – Estatísticas do teste de Durbin-Watson para os portfólios analisados.....	163
Tabela 38 – Estatísticas do teste de Breusch-Pagan para os portfólios analisados	165

LISTA DE SIGLAS E ABREVIACOES

BVC – *Bulk Volume Classification*

BM – *Book-to-Market*

CAPM – *Capital Asset Pricing Model*

CMA – *Conservative Minus Aggressive*

GRS – Gibbons, Ross e Shanken

HEM – Hiptese de Eficincia de Mercado

HML – *High Minus Low*

IMU – *Informed Minus Uninformed*

Inv. – Investimento

Luc. – Lucratividade

MCE – *Mandatory Call Event*

MKT – Fator de mercado

N1 – Nvel 1

N2 – Nvel 2

NM – Novo Mercado

NYSE – *The New York Stock Exchange*

P&G – Procter & Gamble

PIN – *Probability of Informed Trading*

RMW – *Robust Minus Weak*

SMB – *Small Minus Big*

Tam. – Tamanho

TR – *Tick-Rule*

Trad – Tradicional

UMD – *Up Minus Down*

VPIN – *Volume-Synchronized Probability of Informed Trading*

SUMÁRIO

1 INTRODUÇÃO.....	1
1.1 JUSTIFICATIVA	4
1.2 OBJETIVOS	6
1.2.1 <i>Objetivo geral</i>	6
1.2.2 <i>Objetivos específicos</i>	6
2 REFERENCIAL TEÓRICO	7
2.1 ASSIMETRIA INFORMACIONAL.....	7
2.2 MÉTRICAS INDIRETAS DE MENSURAÇÃO DA ASSIMETRIA INFORMACIONAL	12
2.3 <i>PROBABILITY OF INFORMED TRADING (PIN)</i>	14
2.4 <i>VOLUME-SYNCHRONIZED PROBABILITY OF INFORMED TRADING (VPIN)</i>	22
2.4.1 <i>Tempo por volume e tempo cronológico</i>	25
2.4.2 <i>Algoritmos de classificação de transações</i>	27
2.4.3 <i>Cálculo do VPIN</i>	32
2.5 MODELOS DE PRECIFICAÇÃO DE ATIVOS	37
2.6 PIN E OS MODELOS DE PRECIFICAÇÃO DE ATIVOS	41
2.7 ESTUDOS SOBRE ASSIMETRIA INFORMACIONAL NO MERCADO BRASILEIRO	45
3 METODOLOGIA.....	48
3.1 TIPO DE PESQUISA	48
3.2 POPULAÇÃO E AMOSTRA	49
3.3 COLETA E TRATAMENTO DOS DADOS	52
3.4 HIPÓTESES DE PESQUISA	52
3.5 MÉTODOS DE ANÁLISE DOS DADOS	55
3.5.1 <i>Aplicação dos métodos Tick-Rule e Bulk Volume Classification.</i>	55
3.5.2 <i>Modelos de precificação de ativos</i>	56
3.5.3 <i>Os modelos de fatores e o VPIN</i>	57
3.5.3.1 <i>Variáveis dependentes</i>	59
3.5.3.2 <i>Variáveis independentes</i>	61
3.5.3.3 <i>Três fatores de Fama e French (1993)</i>	63
3.5.3.4 <i>Quatro fatores de Carhart (1997)</i>	64
3.5.3.5 <i>Cinco fatores de Fama e French (2015)</i>	65

3.5.3.6 Fator VPIN (IMU)	67
3.5.4 Testes realizados nos modelos de precificação	69
4 ANÁLISE DOS RESULTADOS	73
4.1 DESEMPENHO DOS MÉTODOS TR E BVC	73
4.2 ESTATÍSTICAS DESCRITIVAS DO VPIN	76
4.2.1 Análise do VPIN de todas as ações da amostra	77
4.2.1.1 Análise da simulação do VPIN por bootstrap	84
4.2.2 Análise do VPIN por segmento de mercado da Bovespa	87
4.2.3 Análise do VPIN por valor de mercado da ação	91
4.3 ANÁLISE DA RELAÇÃO ENTRE O VPIN E O BID-ASK SPREAD DAS AÇÕES.....	95
4.4 ANÁLISE DOS MODELOS DE TRÊS, QUATRO E CINCO FATORES E O FATOR IMU	104
4.4.1 Estatísticas descritivas dos fatores	105
4.4.2 Características dos portfólios.....	109
4.4.3 Resultados das regressões dos modelos de fatores	116
4.4.4 Análise das relações entre os fatores de risco sistemático.....	134
4.4.5 Teste GRS aplicado em diferentes combinações de modelos	138
4.4.6 Análise dos modelos através do Bootstrap das carteiras.....	142
5 CONSIDERAÇÕES FINAIS.....	147
REFERÊNCIAS.....	151
APÊNDICE	162

1 INTRODUÇÃO

Devido ao crescente número de ações sendo transacionadas em alta frequência e a concomitante expansão dos bancos de dados *tick by tick* – que contêm dados acerca de todas as transações, como preço, horário, volume e lado agressor –, as pesquisas em microestrutura de mercado estão cada vez mais viáveis de serem realizadas. Em especial, permite que o campo de microestrutura deixe de ser visto como sendo somente um meio de se estudar um comportamento de curto prazo dos preços dos ativos e, possa, portanto, ser associado às demais áreas de estudos em finanças, como o *asset pricing*.

A área de microestrutura de mercado surgiu em meados dos anos 1980 e, segundo O'Hara (1995), estuda o processo e as consequências da compra e venda de ações. A principal diferença dessa área para a economia tradicional acontece pelo foco que a microestrutura dá em analisar como mecanismos de transações específicos afetam a formação do preço das ações. Uma das vertentes, portanto, da microestrutura é o estudo do conteúdo informacional carregado pelos preços das ações. A diferença entre as informações possuídas pelos agentes em um mercado é denominada assimetria informacional e é alvo de estudos desde, ao menos, a década de 1970.

Akerlof (1970) foi um dos primeiros a estudar tal fenômeno utilizando-se de um exemplo no mercado de carros usados e, posteriormente, expandindo sua pesquisa para o mercado financeiro. Outro trabalho fundamental acerca do papel do conjunto informacional detido pelos acionistas foi o de Fama (1970) que diferencia a eficiência do mercado em três formas distintas: fraca, semiforte e forte, a depender do modo como o preço do ativo reflete informações acerca do mesmo. Nesse mesmo contexto, surgiram os modelos de Kyle (1985) acerca da atuação de um *insider* e seu impacto no preço das ações, de Grossman e Stiglitz (1980) sobre a questão do impacto da informação no nível de eficiência do mercado e de Glosten e Milgrom (1985) que propõem um dos primeiros modelos de microestrutura de mercado ao considerar os efeitos de *inside trading* sobre os preços de *bid* e *ask* na perspectiva do formador de mercado. Concomitantemente, outros trabalhos incorporaram a assimetria informacional como um dos fatores principais na determinação da estrutura de capital como a teoria da *pecking order* de Myers e Majluf (1984) e como causador de conflitos entre os

stakeholders de uma companhia, abordado pela teoria da agência de Jensen e Meckling (1976).

Foi, no entanto, a partir do final da década de 1980 que a busca pela quantificação da assimetria informacional ganhou força. Isto se deve ao trabalho inicial de Easley e O'Hara (1987), em que os autores sustentam que a quantidade de ativos negociados em uma transação denuncia uma provável presença de um agente informado, na medida que este, em posse de informação privilegiada, busca concretizar a negociação independentemente do preço estipulado.

A partir de então, diversos trabalhos, encabeçados principalmente pelos autores supracitados (EASLEY *et al.*, 1996; EASLEY, HVIDKJAER e O'HARA, 2002; EASLEY *et al.*, 2004; EASLEY, O'HARA, 2004; EASLEY, HVIDKJAER e O'HARA, 2005; EASLEY *et al.*, 2008; EASLEY, LÓPEZ DE PRADO, O'HARA, 2011a), procuraram refinar e desenvolver uma forma de determinação da assimetria informacional presente em um mercado acionário, resultando primeiramente na *Probability of informed trading* (PIN) e, posteriormente, no *Volume-Synchronized Probability of informed trading* (VPIN). O primeiro modelo procura estimar, por meio de uma função de máxima verossimilhança, a probabilidade de ter havido transações baseadas em informações privilegiadas em um determinado período. Já o segundo procura medir diretamente qual o grau de toxicidade dos fluxos de ordens de uma ação, a qual foca a perda esperada de um *market maker* ao estar em um mesmo ambiente de um agente mais bem informado.

Considerando o mercado acionário brasileiro como mais arriscado do que os mercados de países desenvolvidos (MARTINS e PAULO, 2013) e levando em consideração que países emergentes são campos férteis para transações acionadas por agentes com informações privilegiadas (DUARTE e YOUNG, 2009), diversos pesquisadores se propuseram a estudar a assimetria informacional no mercado nacional, tanto via PIN (BARBEDO *et al.*, 2009; MARTINS e PAULO, 2013, 2014), quanto por modelos alternativos (IQUIAPAZA, LAMOUNIER e AMARAL, 2008; ALBANEZ e VALLE, 2009; ALBANEZ *et al.*, 2012). Os achados de tais pesquisas convergem para o mesmo resultado: existe uma alta probabilidade de a prática de *inside trading* estar presente no mercado acionário brasileiro.

O desequilíbrio informacional acerca de um ativo pode ser um dos fatores precificados pelos agentes em um mercado e, dessa forma, pode afetar o retorno requerido das ações

listadas no mercado financeiro. O cálculo desse custo enseja controvérsia na literatura, tendo gerado diversos modelos que se propõem a mensurá-lo, perpassando pelos trabalhos de Gordon e Shapiro (1956), Sharpe (1964), Ross (1976), Fama e French (1992, 1993 e 2015), Ohlson e Juettner-Nauroth (2005) dentre outros. Uma dificuldade em sua mensuração reside na determinação dos fatores explicativos que o compõem, tendo como modelo geral o *Arbitrage Pricing Theory* (ATP) de Ross (1976). Apesar de sua teórica capacidade explicativa do retorno requerido por um ativo, este não estabelece quais são, de fato, as variáveis independentes do modelo.

Proposto anteriormente ao APT, o *Capital Asset Pricing Model* (CAPM), formulado inicialmente por Sharpe (1964), mantém-se como uma das principais formas de precificação de ativos no mercado financeiro (FORTUNATO, MOTTA e RUSSO, 2010). Tal modelo, valendo-se de condições hipotéticas de equilíbrio e funcionamento eficiente dos mercados em que as informações estão amplamente disponíveis aos investidores, ausência de custos de transação e existência de um ativo livre de risco, preconiza que o retorno de um ativo é dado pelo prêmio de risco do mercado multiplicado pelo coeficiente de risco da ação (beta) adicionado o retorno de um ativo livre de risco.

Apesar de sua extensa utilização na prática e pesquisas em finanças, Easley, Hvidkjaer e O'Hara (2005) apontam a incongruência na utilização de modelos como o CAPM – que preconizam que, em equilíbrio, os agentes têm homogeneidade informacional – para estudar a precificação da assimetria de informações por parte dos investidores. Isto se deve ao fato de que os modelos PIN e VPIN derivam de um cenário em que os participantes dispõem de diferentes níveis de acesso à informação, portanto tal fator violaria o pressuposto supracitado, indicando a necessidade da utilização de outros modelos.

Dessa forma, neste trabalho, adotaram-se procedimentos semelhantes aos de Easley, Hvidkjaer e O'Hara (2005) e Mohanram e Rajgopal (2009), que estudam a presença de assimetria informacional no retorno requerido das ações por meio de um fator relacionado à PIN utilizando o modelo de Fama e French (1993). Este trabalho vai um passo além, adotando também o modelo de cinco fatores de Fama e French (2015) não utilizado nas pesquisas supracitadas.

Tendo em vista a necessidade de maiores estudos acerca da existência de investidores informados e o seu efeito sobre o mercado acionário brasileiro, definiu-se a seguinte pergunta de pesquisa:

Qual o efeito do nível de toxicidade dos fluxos de ordens de transações, medido pelo VPIN, no retorno requerido de ações negociadas no mercado acionário brasileiro no período de 01º de maio de 2014 a 31 de maio de 2016?

Esta pesquisa está dividida em cinco seções. A primeira delas versa sobre a contextualização do trabalho, além de sua justificativa e objetivos. A segunda uma revisão geral do alicerce teórico por trás deste trabalho. A terceira explica as metodologias utilizadas, enquanto a quarta analisa os resultados da aplicação dos modelos. A última seção retoma os objetivos e relata as considerações finais. Por fim, apresentam-se também as referências bibliográficas e um apêndice.

1.1 Justificativa

Segundo Duarte e Young (2009), países em desenvolvimento tendem a constatar uma atuação mais volumosa de *insiders* em seus mercados acionários. Apesar de tal ato ser caracterizado como crime (MARTINS e PAULO, 2013), as leis parecem não alcançar a aplicação ideal, deixando, portanto, aberturas para que o mercado seja informacionalmente desbalanceado, prejudicando os agentes nele envolvidos. Além disso, a existência de *insiders* pode levar à saída de investidores – especialmente os minoritários – ou a não entrada daqueles que desejam investir seu capital por se verem em desvantagem informacional (EASLEY, HVIDKJAER e O’HARA, 2002).

A literatura financeira, especialmente nos últimos anos, tem dado maior atenção aos trabalhos de microestrutura de mercado. Diversos autores se propuseram a explicar como a existência de *insiders* em um ambiente afeta o custo de capital das empresas. No entanto, os resultados divergem demasiadamente de estudo para estudo (MOHANRAM e RAJGOPAL, 2009). No cenário nacional, diversas pesquisas acerca de negociações com informação privilegiada vêm sendo realizadas, relacionando tal fenômeno a diversos fatores como: estrutura de capital (ALBANEZ e VALLE, 2012), governança corporativa (BARBEDO,

SILVA e LEAL, 2009; MARTINS e PAULO, 2013, 2014), custo de capital (ALBANEZ *et al.*, 2010) dentre outros.

Em relação ao papel da microestrutura no campo de finanças, Easley, Hvidkjaer e O'Hara (2005) afirmam que, apesar da aparente afinidade natural entre a área de *asset pricing* e a microestrutura, as pesquisas mantêm os campos em separado. Enquanto o primeiro foca no papel do mercado e outros riscos agregados, o segundo foca questões específicas dos ativos e o ajustamento dos preços. Segundo os autores, tal separação criou uma lacuna em ambas as áreas, já que os modelos de precificação de ativos não apresentam bom desempenho e o campo de microestrutura é considerado apenas na análise do comportamento de curto prazo dos preços.

A área de estudos de *asset pricing* não leva em consideração as mecânicas da evolução do preço dos ativos, deixando em voga a questão do *tatônnement* – atingimento do equilíbrio nos mercados. Enquanto isso, a microestrutura de mercados estuda a mecânica do processo de transações e seu efeito na evolução dos preços das ações, sendo que um dos maiores focos é justamente o papel da informação e como esta é incorporada nos preços (EASLEY, HVIDKJAER e O'HARA, 2002).

O'Hara (2003) diz que a área de *asset pricing* ignora justamente o que a microestrutura de mercado estuda, o fato de o preço dos ativos evoluir no mercado. Tendo em vista que o mercado de forma geral tem duas funções: a de liquidez e a de definição dos preços, e que ambas são importantes para a precificação de ativos, segundo a autora, os modelos de precificação devem incorporar o custo da liquidez e os riscos do chamado *price discovery*. Finalmente, a pesquisadora argumenta que os modelos de precificação de ativos não funcionam, pois presumem que os problemas relacionados à liquidez e à definição de preços já estão resolvidos e, como apontado por O'Hara (2003), ambos se relacionam com o grau de informação do mercado e sua assimetria entre os agentes.

Dessa forma, percebe-se a importância da pesquisa acerca da existência de assimetria informacional no mercado brasileiro e os efeitos que ele acarreta. Além disso, buscou-se contribuir para o melhor entendimento do processo gerador dos retornos das ações, na medida em que esse assunto ainda tem um vasto campo a ser explorado, não tendo, ainda, a área das finanças chegado a uma conclusão a esse respeito.

1.2 Objetivos

1.2.1 Objetivo geral

O objetivo desta pesquisa foi verificar se o nível de toxicidade das ações, quantificado pelo VPIN, é um fator de risco sistemático precificado no mercado de capitais brasileiro. Ou seja, buscou-se analisar se a existência de agentes informados representa um fator de risco sistemático precificado pelos investidores de ações negociadas na BM&FBovespa.

1.2.2 Objetivos específicos

- I. Comparar os resultados dos algoritmos de classificação de transações *Tick Rule* e *Bulk Volume Classification* com os dados reais.
- II. Estimar, via VPIN, o nível de toxicidade das ações transacionadas no mercado brasileiro.
- III. Verificar se há diferença entre os VPINs médios nos diferentes segmentos de mercado da BM&FBovespa.
- IV. Analisar a relação entre o VPIN e o valor de mercado das empresas.
- V. Verificar se um fator baseado no VPIN é precificado utilizando-se dos modelos de três, quatro e cinco fatores.
- VI. Identificar qual a melhor combinação de fatores para a explicação dos retornos dos portfólios formados por ações negociadas na BM&FBovespa.

2 REFERENCIAL TEÓRICO

2.1 Assimetria informacional

O conceito de assimetria informacional é definido por Lambert, Leuz e Verrecchia (2011) como resultante do fato de um conjunto de investidores não terem acesso a informações que estão disponíveis a outros participantes. A consequente utilização de tais informações para a compra e venda de ações no mercado financeiro é denominada *inside trading*, e o seu praticante, *insider*.

Um dos primeiros trabalhos voltados a analisar o impacto da existência de assimetria informacional foi o de Akerlof (1970) em estudo no qual o autor se valeu do mercado de carros usados dos Estados Unidos, conhecido como *market for lemons*, para demonstrar os efeitos negativos decorrentes desse fenômeno. Segundo o pesquisador, a assimetria surge na medida em que os vendedores de carros têm maior conhecimento sobre a sua qualidade do que os clientes. Os potenciais compradores, não dispendo de meios de averiguar a qualidade do produto, atribuirão aos carros bons e ruins o mesmo preço. Dessa forma, Akerlof (1970) afirma que a maioria dos carros vendidos será ruim, pois aqueles vendedores que colocarem bons carros à venda serão prejudicados por receberem o mesmo valor de um carro de qualidade inferior. Em um cenário mais extremo, o mercado pode desaparecer, na medida em que os vendedores param de negociar seus produtos por não conseguirem receber o que eles realmente valem.

A partir de então, diversos autores se voltaram às pesquisas acerca da assimetria informacional no contexto financeiro. Segundo Leland e Pyle (1977), inúmeros mercados são caracterizados por terem um desequilíbrio de informações entre os compradores e os vendedores. Tal fato é ainda mais acentuado nos mercados financeiros, em especial, nas relações entre tomadores e credores. Os primeiros conhecem de forma mais robusta seus riscos operacionais, financeiros e morais do que os segundos. Os credores, caso tenham em mãos informações precisas acerca dos tomadores, têm melhores condições de precificar um empréstimo ou avaliar os riscos de concretizar operações junto aos mutuários. A aquisição de tal conjunto de informações, no entanto, pode ser demasiado cara ou impossível, na medida

em que não se espera que os tomadores sejam totalmente abertos acerca das operações e características de suas empresas.

Os pesquisadores da teoria da agência, como Jensen e Meckling (1976), levam em consideração esse contexto em que, devido à divergência de interesses entre as partes envolvidas, pode ocorrer a sobreposição de interesses particulares (dos gestores, de acionistas antigos, de fornecedores e de outros) aos objetivos gerais de uma organização. Como consequência, surgem conflitos internos e externos nas mais diversas esferas, tais como: acionistas e executivos, controladores e minoritários, empresa e credores etc (BARBEDO, SILVA e LEAL, 2009). Esse contexto é permeado pela assimetria informacional entre os agentes. Tendo em vista que cada um age de forma a concretizar seus interesses à frente do interesse geral, omitem ou divulgam apenas aquelas informações que lhes servirão para atingir seus objetivos.

De forma a tratar da assimetria informacional nos mercados acionários, é necessário tentar entender quais as consequências de um desequilíbrio informacional para o mercado como um todo. Englobam-se a esse questionamento os efeitos para os acionistas minoritários, grupos de investidores, o preço da ação, seu risco e todos os fatores que podem ser afetados por tal desbalanceamento.

Leland (1992) aponta que, nos Estados Unidos, tal prática é considerada crime desde 1934, sendo regulada pela *Securities and Exchange Commission* (SEC). O autor levanta duas questões importantes: (i) é justo haver um mercado quando os indivíduos têm diferentes níveis de informações? (ii) É economicamente viável permitir o *inside trading*?

Levando em consideração a Hipótese de Eficiência de Mercado (HEM) de Fama (1970), Leland (1992) aponta os argumentos a favor e contra a prática do *inside trading*. Na medida em que os preços das ações refletem todas as informações disponíveis (públicas e privadas), a atuação do *insider* fará com que as novas informações sejam incorporadas aos preços dos ativos. Dessa forma, tomadores de decisões podem reduzir riscos e melhorar seu desempenho ao levarem em consideração tais informações. Em consequência, os riscos dos ativos serão menores e os preços, superiores, aumentando a liquidez e permitindo maiores investimentos. Por outro lado, potenciais investidores serão avessos a entrar nesse mercado ao considerá-lo injusto. Consequentemente, os investimentos, preços e liquidez dos ativos serão

menores, prejudicando aqueles investidores que operam no mercado sem informações privilegiadas.

Os resultados encontrados por Leland (1992) confirmam, tanto alguns dos argumentos a favor, quanto contra o *inside trading*. Resumindo seus achados, de fato, os preços dos ativos revelam as informações quando esta prática é permitida. O preço médio das ações aumenta, e o retorno esperado e o risco para os investidores serão menores. Em contrapartida, a liquidez do mercado diminuirá, prejudicando aqueles indivíduos que buscam liquidez ao investir no mercado acionário.

Grossman e Stiglitz (1980) estudam a questão do impacto da informação no nível de eficiência do mercado. Os autores propõem um modelo de equilíbrio em que há graus de desequilíbrio. Nesse caso, os preços refletem parcialmente as informações tidas pelos *inside traders*, sendo assim, aqueles que procuram adquirir informações privilegiadas obtêm uma compensação por tal ação.

Basicamente, o modelo dos autores propõe que um agente faz uma oferta, aumentando o preço de uma ação quando este consegue informações positivas sobre a mesma. O contrário ocorre caso a informação é de natureza negativa. Dessa forma, o sistema de preços traz à tona, de forma imperfeita, aquelas informações obtidas pelos agentes informados.

Os autores mostram que, de maneira geral, o sistema de preços não revela toda a informação sobre o valor real de um ativo de risco. A única maneira de os agentes informados terem vantagens sobre os investidores não informados é tomar posições mais vantajosas do que as tomadas pelos últimos. A Hipótese de Eficiência de Mercado (HEM) preconiza que, a qualquer momento, os preços refletem toda a informação disponível acerca de um ativo. No entanto, caso o mercado funcionasse dessa maneira, os *inside traders* não obteriam lucros ao adquirirem informações privilegiadas. Se a HEM se sustentasse, as informações seriam disponíveis a todos os participantes do mercado e sem custo. Nesse caso, não haveria o incentivo para o agente desembolsar recursos para obter informações: todos os agentes seriam não informados (em termos de privilégio). Dessa forma, os autores concluem que não é possível a existência de um mercado eficiente em seu componente informacional.

Glosten e Milgrom (1985) propõem um dos primeiros modelos de microestrutura de mercado ao considerarem os efeitos de *inside trading* sobre os preços de *bid* e *ask* na perspectiva do formador de mercado. Segundo os autores, o *spread* entre o *bid* e o *ask* pode ser puramente uma consequência do nível informacional do mercado. O modelo diz que o

formador de mercado se encontra em um problema de seleção adversa quando um potencial investidor aceita comprar ou vender uma ação nos preços por ele determinados, pois tal agente pode deter informações que o *market maker* não tem. Sendo assim, o formador de mercado deve compensar essas perdas transacionando com aqueles que compram ou vendem por motivos de liquidez.

Dessa forma, os autores mostram que o *spread* pode surgir inteiramente de seleção adversa. Ao mesmo tempo, no entanto, Glosten e Milgrom (1985) afirmam que, ao longo do tempo, os valores esperados, tanto para o investidor informado, quanto para o formador de mercado tendem a convergir, apontando que a informação privilegiada se inclina a ser completamente disseminada por meio dos preços das transações.

Wang (1993) propõe um modelo dinâmico de precificação de ativos sob um contexto de assimetria informacional. Neste, os investidores informados sabem ao certo a taxa de crescimento dos dividendos, enquanto os desinformados não têm acesso a tal informação. Estes últimos observam os pagamentos de dividendos e o preço das ações e deles extraem as informações acerca das expectativas futuras dos mesmos. Wang (1993) explora as implicações do modelo em relação ao comportamento dos preços dos ativos, volatilidade, prêmio de risco e também a existência de correlação serial entre o retorno das ações e as estratégias de investimento dos agentes.

O efeito da assimetria informacional sobre o retorno requerido estudado por Wang (1993) se mostra ambíguo. Primeiramente, um desequilíbrio informacional no mercado de capitais pode levar a um aumento da volatilidade dos preços das ações. Nesse contexto, agentes informados transacionam com informações melhores do que investidores não informados. Dessa forma, estes últimos irão requerer um prêmio adicional pelo risco de estarem em um mesmo ambiente juntamente a investidores mais bem informados.

Por outro lado, quando estão presentes mais investidores não informados do que sua contraparte, os preços dos ativos contêm menos informações sobre a ação. Assim, de forma geral, há mais incertezas sobre o fluxo de caixa futuro das empresas e, em consequência, os agentes demandarão maior prêmio ao investirem nessas ações. Dessa forma, na medida em que aumenta o número de investidores informados, também aumenta o prêmio de risco requerido (WANG, 1993). Segundo Easley e O'Hara (2004), como o modelo de Wang (1993) só contempla a existência de um ativo de risco, não fica claro como o conjunto informacional afeta os retornos ou a seleção de portfólios.

Diamond e Verrecchia (1991) investigam a relação da assimetria informacional e o custo de capital das empresas e chegam a resultados inconclusivos. Por um lado, a revelação de informações ao público reduz o custo de capital de uma empresa uma vez que atrai grandes investidores motivados pela maior liquidez do ativo (grandes firmas divulgam mais informações visto que o benefício angariado por elas serão maiores). No entanto, em situações menos comuns, a redução da assimetria informacional leva ao resultado contrário.

A lógica seguida por Diamond e Verrecchia (1991) é de que, quando investidores dispõem de informações relevantes, lançam grandes ofertas de compras ou vendas juntamente ao *market maker*. Considerando um cenário caracterizado pelo desequilíbrio de informações entre os agentes, a redução da assimetria informacional (via divulgações por parte das empresas) leva à diminuição da volatilidade do desequilíbrio das ordens dos investidores, o que diminui o *spread* dos *market makers*, levando-os a saírem do mercado.

Lambert, Leuz e Verrecchia (2011) reafirmam a divergência dos resultados dos estudos acerca do efeito da assimetria informacional sobre o custo de capital das empresas. Em seu trabalho, os autores formulam dois cenários, um em que há a presença de agentes mais bem informados e outro em que os indivíduos têm um nível informacional homogêneo. Ao mesmo tempo em que os cenários se diferem, o nível médio de informação entre os agentes nos dois ambientes é o mesmo. Isto se deve ao fato de que, no primeiro, os investidores com mais informações compensam aqueles menos informados, mantendo, portanto, uma média do nível de informação entre os agentes igual à do segundo cenário. Tendo em vista essas duas conjunturas, os autores procuram avaliar em qual delas o custo de capital das firmas será maior.

O resultado encontrado por Lambert, Leuz e Verrecchia (2011) é diretamente relacionado ao grau de competitividade do mercado em que as firmas estão presentes. Caso estas estejam imersas em um mercado perfeitamente competitivo, o custo de capital será igual nos dois contextos. Isto se deve ao fato de que tal fator é regido unicamente pelo grau médio de precisão informacional dos agentes, de forma que este, sendo igual nos dois cenários, fará com que o custo de capital também o seja. Tomando como base novamente a HEM, o grau de incerteza dos investidores diminui quando outros agentes adquirem mais informações (privadas ou públicas), pois tal conhecimento é passado para o restante do mercado por meio dos preços dos ativos. Em outras palavras, aumentar o grau de assimetria informacional quando os mercados são perfeitamente líquidos leva a um aumento na precisão média de

informação dos indivíduos e, conseqüentemente, a uma redução no custo de capital das firmas (LAMBERT, LEUZ e VERRECCHIA, 2011).

Trazendo esse raciocínio para um contexto em que o mercado não seja perfeitamente competitivo, tem-se que o grau de liquidez influencia diretamente o volume de informação que é transmitido via preços dos ativos. Como parte das informações não serão passadas para os investidores sem acesso a informações privadas, o nível médio informacional dos agentes será menor, aumentando, assim, o custo de capital das empresas. Dessa forma, Lambert, Leuz e Verrecchia (2011) afirmam que a influência da assimetria de informações sobre o custo de capital das firmas está diretamente relacionado ao grau de competitividade do mercado.

Hughes, Liu e Liu (2007) desenvolvem uma teoria em que o risco informacional, quando presente em uma grande economia, ou é capturado pelos fatores de risco já presentes no CAPM, ou é diversificável. Dessa forma, este não se caracteriza como um fator de risco adicional. Os autores mostram que, em um cenário com infinitos ativos, a PIN demonstrada no modelo de Easley e O'Hara (2004) não é precificada. No entanto, esta se torna um fator de precificação quando analisada em um ambiente em que o número de ativos é finito.

Tendo em vista as divergentes conclusões dos pesquisadores acerca do efeito da assimetria informacional, é necessário estabelecer uma forma de testá-la empírica e estatisticamente, o que requer um meio de quantificar esse fenômeno. Existem *proxies* indiretas para o *inside trading*, assim como formas diretas de mensurá-lo. A principal, PIN, cunhada por Easley *et al.* (1987, 1996, 2008) e depois modificada para o VPIN, procura quantificar a probabilidade da existência de atuação de agentes mais bem informados, também denominado nível de toxicidade de uma ação (*order flow toxicity*).

2.2 Métricas indiretas de mensuração da assimetria informacional

Quando os primeiros trabalhos acerca do papel da assimetria informacional começaram a aparecer nas décadas de 1970 e 1980, naturalmente sugeriram também formas de tentar mensurar tal fenômeno. A primeira medição proposta foi o *bid-ask spread*, utilizado por Bagehot (1971), Copeland e Galai (1983) e Glosten e Milgrom (1985). Bagehot (1971) utiliza essa medida para explicar o risco incorrido pelos formadores de mercado ao transacionarem com agentes informados. Copeland e Galai (1983) mostram que uma maior quantidade de

inside traders leva a um *spread* maior para que os *market makers* compensem as suas perdas por seleção adversa. Glosten e Milgrom (1985) formam um modelo de microestrutura de mercado para explicar a formação de preços a partir do *bid-ask spread*.

Segundo Jiang (2015), outras formas de quantificar a presença de *inside traders* foram propostas por Benston e Hagerman (1984), Hasbrouck (1991) e Sarin, Shastri e Shastri (2000). Benston e Hagerman (1984) testam a associação do custo de transação com o risco sistêmico e não sistêmico, mostrando que este último se relaciona com o *spread*. Stoll (1978) regride o *spread* com o volume de transação, preço da ação e variabilidade do retorno. O autor mostra que o *spread* é positivamente associado com o risco e negativamente associado com preço e volume e que a variabilidade do retorno poderia ser uma *proxy* para assimetria informacional.

Hasbrouck (1991) separa a variância da movimentação do preço em duas partes de acordo com a sua relevância no preço da transação baseada em um modelo autorregressivo e afirma que a variância da parte relevante do preço pode ser uma variável substituta do *inside trading*. Sarin, Shastri e Shastri (2000) constatam que maior *ownership* de indivíduos dentro da companhia é associada, tanto com um maior *spread*, quanto com a assimetria informacional enfrentada pelos investidores. Além desses trabalhos, Bharath, Pasquariello e Wu (2009) utilizam um índice de assimetria informacional baseada em medidas de seleção adversa para testar se tal fator é um determinante da estrutura de capital das empresas, como previsto pela teoria da *pecking order*.

No cenário nacional, Albanez e Valle (2009) seguem os passos de Iquiapaza, Lamounier e Amaral (2008) e utilizam como *proxy* para assimetria informacional uma variável *dummy* indicando se a empresa pertence ou não a segmentos diferenciados de governança corporativa da BM&FBovespa. Outras variáveis também foram usadas como *proxy* para assimetria informacional. Segundo Albanez e Valle (2009), a liquidez em bolsa e a volatilidade apresentam relações com o caráter informacional da empresa, além de serem correlacionados com problemas de seleção adversa.

Nenhuma das medidas apresentadas acima, porém, procura diretamente mensurar a assimetria informacional e pode, portanto, servir como *proxies* para outros componentes da ação. Dessa forma, considera-se na literatura que os métodos PIN e VPIN desenvolvidos por Easley *et al.* (1996) e Easley, López de Prado e O'Hara (2011a) respectivamente permitem

uma mensuração mais objetiva e assertiva da probabilidade de negociações privilegiadas e o grau de toxicidade dos fluxos das ordens das ações.

2.3 Probability of informed trading (PIN)

A quantificação de um fenômeno tal como a assimetria informacional em um contexto dinâmico e complexo como um mercado acionário é um dos maiores desafios em se tratando dos estudos de microestrutura em finanças (MARTINS e PAULO, 2013). Pesquisas anteriores utilizaram diferentes *proxies* para a assimetria informacional, como: erro dos analistas na previsão de ganhos por ação (ALBANEZ *et al.*, 2010); uso de variáveis *dummy* para adesão nos níveis diferenciados de governança corporativa da BOVESPA (IQUIAPAZA, LAMOUNIER e AMARAL, 2007); grau de liquidez ou volatilidade das ações (ALBANEZ e VALLE, 2009). No entanto, essas formas de mensuração de desequilíbrio informacional são carregadas de incertezas e não apresentam uma estimativa objetiva do fenômeno estudado, apontando, portanto, a necessidade de uma *proxy* mais robusta e direta.

A *Probability of Informed Trading* (PIN) foi primeiramente introduzida em Easley *et al.* (1996) baseada no trabalho de Easley e O'Hara (1987). Easley e O'Hara (1987) procuram investigar o efeito da quantidade de ordens de transações sobre o preço de um ativo. Os autores viram que o tamanho de uma ordem de compra ou venda pode ser indício de ação de um *insider*. A explicação reside no fato de que grandes ordens de compra ou venda (chamados de *blocks*) são feitos a preços inferiores a ofertas menores, indicando que os agentes - possivelmente detentores de informações privilegiadas - desejam concretizar a transação independentemente do preço.

Esta pesquisa abriu espaço para que Easley *et al.* (1996), buscando estudar atividades de *insiders* em ações com diferentes níveis de liquidez, propusessem um modelo para a estimativa da probabilidade de atuação de agentes com informações privilegiadas. Easley *et al.* (1996) procuraram estudar como transações baseadas em informações internas diferenciam-se entre ações ativas e inativas no mercado de ações de Nova York (NYSE). O resultado de maior impacto desse artigo é o de que a probabilidade de negociações privilegiadas é menor para ações com maior volume transacionado. Analisando as ações menos ativas, os autores apontam para o fato de que a chegada de ordens de compra ou

vendas podem vir de agentes informados. Dessa forma, o risco de transações informadas para esse tipo de ações maiores é maior.

A PIN é baseada na perspectiva do *market maker* e no *spread* entre os *bid* e *ask prices* de uma ação. Segundo Easley *et al.* (1996), o modelo se sustenta no problema de decisão do formador de mercado, no qual os preços são os *outputs* e as ordens de compra e venda são os *inputs* no seu processo decisório de estabelecimento de preço (*market maker's price-setting decision problem*).

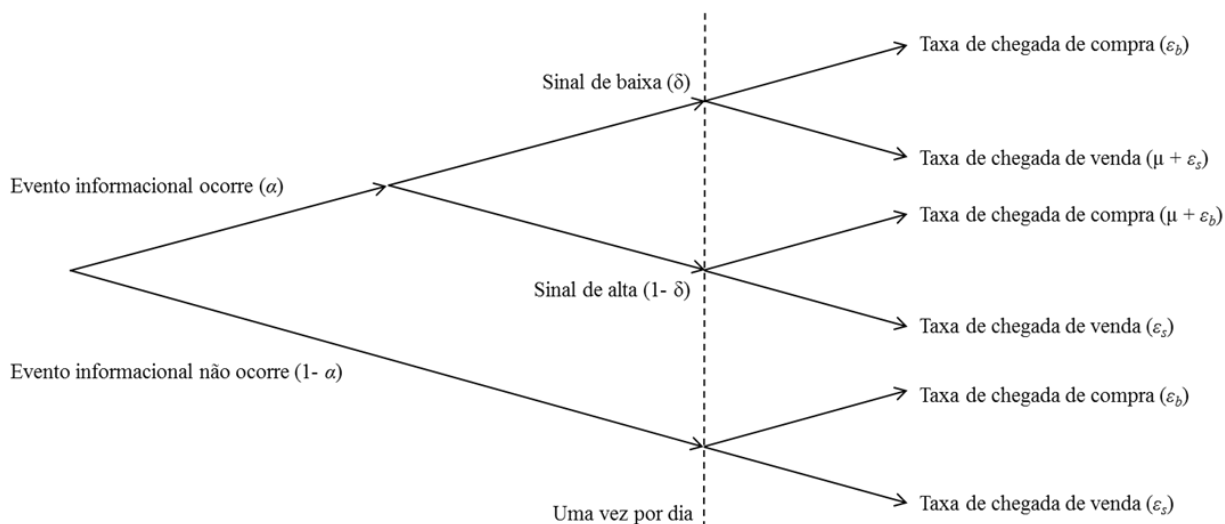
Primeiramente pressupõe-se que as operações de compra e venda de ativos aconteçam com base nas informações que os investidores detêm. Um negociador informado é aquele que tem informações privadas, ao passo que o negociador desinformado não as tem (MARTINS e PAULO, 2014). A construção do modelo se dá por meio de variáveis probabilísticas que relacionam o grau informativo dos participantes à ocorrência dos possíveis tipos de eventos num determinado período. Os indivíduos compram e vendem ações juntamente ao *market maker* ao longo de $i = 1, \dots, I$ dias. O tempo, dentro dos dias, é contínuo e indexado por $t \in [0, T]$. O *market maker* compra e vende ações no *bid* e *ask prices* a qualquer momento. Tomando-o como competitivo e neutro ao risco, esses preços são os valores esperados dos ativos com base nas informações de que o formador de mercado dispõe no momento da transação.

Easley *et al.* (1996) afirmam que o modelo se baseia no fato de que, ao longo do dia, a natureza determina o acontecimento de um evento informativo – entende-se nesse caso que a aparição de uma notícia sobre uma ação se dá de forma aleatória – e são independentemente distribuídos e ocorrem com uma probabilidade α ; δ representa a probabilidade de ser uma notícia ruim e $(1 - \delta)$ uma notícia boa. Dada uma boa notícia em t , a ação assume valor \bar{V}_i . Caso a notícia seja ruim, a ação valerá \underline{V}_i . Se não há nenhuma notícia no dia i a ação valerá V_i . Assume-se que $\underline{V}_i < V_i < \bar{V}_i$. Negociadores informados operam com uma taxa de negociação μ . Por sua vez, negociadores desinformados transacionam no mercado com taxa de chegada ε_b para compras e ε_s para vendas.

Dada à ocorrência de um evento informativo, o *market maker* avaliará a informação sob a qual tomará sua decisão de compra ou venda. Segundo Martins e Paulo (2013), os movimentos nos preços se darão conforme as crenças formadas por esse investidor em relação

ao ativo e às probabilidades supracitadas. Pode-se representar graficamente o processo de negociação conforme diagrama da figura 1.

Figura 1 – Diagrama da árvore do processo de negociação



Fonte: EASLEY, HVIDKJAER e O'HARA (2002), adaptado pelo autor da dissertação.

Assume-se que o *market maker* utiliza a chegada das ordens de transações para tomar decisões acerca da ocorrência ou não de eventos informativos. Como os dias são independentes, pode-se analisá-los separadamente. Tem-se que $P(t) = (P_b(t), P_g(t), P_n(t))$ são as probabilidades de ocorrência de notícias ruins, boas e a não ocorrência de notícias acerca da ação respectivamente. Dessa forma, as probabilidades no período 0 são $P(0) = (\alpha\delta, \alpha(1 - \delta), 1 - \alpha)$.

Para a determinação dos *bid* e *ask prices* o *market maker* atualiza suas crenças na medida que ordens de compra (*B*) e venda (*S*) chegam. Assim, pode-se calcular a probabilidade posterior à chegada de uma ordem de transação. As probabilidades de que tenha havido uma informação ruim, boa e nenhuma, quando há o lançamento de uma ordem de venda (*S*), por exemplo, são dadas respectivamente pelas equações (1), (2) e (3)

$$P_b(t|S_t) = \frac{P_b(t)(\varepsilon + \mu)}{\varepsilon + P_b(t)\mu} \quad (1)$$

$$P_g(t|S_t) = \frac{P_g(t)\varepsilon}{\varepsilon + P_b(t)\mu} \quad (2)$$

$$P_n(t|S_t) = \frac{P_n(t)}{\varepsilon + P_b(t)\mu} \quad (3)$$

No momento t , o *bid price* – $b(t)$ – é o valor que o *market maker* espera que o ativo tenha com base em suas expectativas. Dessa forma, $b(t)$ é dado por:

$$b(t) = \frac{P_n(t)\varepsilon V_i + P_b(t)(\varepsilon + \mu)\underline{V}_i + P_g(t)\varepsilon \bar{V}_i}{\varepsilon + P_b(t)\mu} \quad (4)$$

Simetricamente tem-se que o *ask price* será:

$$a(t) = \frac{P_n(t)\varepsilon V_i + P_b(t)\varepsilon \underline{V}_i + P_g(t)(\varepsilon + \mu)\bar{V}_i}{\varepsilon + P_g(t)\mu} \quad (5)$$

Anteriormente ao tempo t , tem-se que o valor esperado do ativo é:

$$E[V_i|t] = P_n(t)V_i + P_b(t)\underline{V}_i + P_g(t)\bar{V}_i \quad (6)$$

Substituindo (6) em (5) e (4), tem-se que os valores de *bid* e *ask* serão dados por:

$$b(t) = E[V_i|t] - \frac{\mu P_b(t)}{\varepsilon + \mu P_b(t)} (E[V_i|t] - \underline{V}_i) \quad (7)$$

$$a(t) = E[V_i|t] + \frac{\mu P_g(t)}{\varepsilon + \mu P_g(t)} (\bar{V}_i - E[V_i|t]) \quad (8)$$

As equações (7) e (8) demonstram, de forma explícita, o papel dos agentes informados e não informados na determinação dos preços. Caso não haja agentes informados ($\mu = 0$), o valor de $b(t)$ e $a(t)$ serão iguais ao valor esperado do ativo. Simetricamente, caso só haja a presença de agentes com informações privilegiadas ($\varepsilon = 0$) o valor de $b(t) = \underline{V}_i$ e $a(t) = \bar{V}_i$.

O *spread* entre o *bid* e o *ask prices* se dá de forma que o *market maker* possa se proteger de perdas para agentes informados. Tomando o *spread* por π , tem-se que este no tempo t é dado por:

$$\pi(t) = a(t) - b(t) \quad (9)$$

Portanto, o *spread* é a probabilidade de uma compra acionada por um agente informado vezes a perda esperada, somada pela probabilidade de uma venda acionada por um agente informado vezes a perda esperada decorrente dessa transação. Somando essas probabilidades, tem-se a probabilidade de que uma ordem seja baseada em informação privilegiada no tempo t , o que resulta em:

$$PI(t) = \frac{\mu(1 - P_n(t))}{\mu(1 - P_n(t)) + 2\varepsilon} \quad (10)$$

Caso o *market maker* não acredite na possibilidade de ter havido notícias acerca da ação, em outras palavras, $P_n(t) = 1$, então este não crê na atuação de agentes informados ($\mu = 0$), portanto, $PI(t) = 0$. Caso contrário, ou seja, só haja agentes informados no mercado ($\varepsilon = 0$), então $PI(t) = 1$. Neste caso, o *spread* ($\overline{V}_i - \underline{V}_i$) será tal que não permitirá que esses agentes lucrem com a transação. Tal fato fica evidente quando se considera o caso em que a probabilidade de ocorrerem notícias boas e ruins é a mesma ($\delta = 0.5$):

$$\pi(0) = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} [\overline{V}_i - \underline{V}_i] \quad (11)$$

Devido à impossibilidade de se observar diretamente a ocorrência de eventos informativos, Easley *et al.* (1996) apontam que a forma de se estimar os parâmetros $\theta = (\alpha, \mu, \varepsilon_b, \varepsilon_s, \delta)$, se dá por meio de uma função de máxima verossimilhança utilizando as únicas variáveis observáveis: o número de ordens de compra e venda.

Easley *et al.* (1996) argumentam que devido à natureza das ordens de transação, estas seguem um dos três processos de Poisson no tempo t . Apesar da incerteza sobre qual dos três opera naquele momento, tem-se que os dados refletem a estrutura de informação acerca da ação, com maior número de ordens de compra esperado em dias com notícias boas, e maior número de ordens de venda quando há notícias ruins. Em dias neutros (sem notícias de qualquer natureza), não há a presença de agentes informados, portanto, menos ordens de transações são lançadas.

Primeiramente considera-se a probabilidade da chegada de ordens em um dia com notícias ruins. As ordens de venda chegarão em uma taxa $(\mu + \varepsilon)$ enquanto as ordens de compra chegam com taxa (ε) . Isso significa que agentes informados e não informados vendem a ação e somente estes últimos compram tal ativo. Como a distribuição no modelo segue um processo independente de Poisson, a probabilidade de observar um número B de compras e S de vendas em um dia de notícias ruins é:

$$e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-(\mu + \varepsilon)T} \frac{[(\mu + \varepsilon)T]^S}{S!} \quad (12)$$

Seguindo a mesma lógica, em um dia sem notícias, exclui-se a taxa de chegada de agentes informados (μ), o que resulta na seguinte probabilidade:

$$e^{-\varepsilon T} \frac{(\varepsilon T)^B}{B!} e^{-\varepsilon T} \frac{(\varepsilon T)^S}{S!} \quad (13)$$

Por fim, em um dia com notícias boas, tem-se:

$$e^{-(\mu+\varepsilon)T} \frac{[(\mu+\varepsilon)T]^B}{B!} e^{\varepsilon T} \frac{(\varepsilon T)^S}{S!} \quad (14)$$

Easley, Hvidkjaer e O'Hara (2005) argumentam que os volumes físicos totais de negociações de compra e venda são suficientes para a estimação da PIN. Visto que todas as probabilidades de operação seguem distribuição de Poisson, estas podem ser estimadas a partir da maximização de uma função de máxima verossimilhança, que representa uma média das equações (12), (13) e (14), ponderada pelas probabilidades de ocorrer uma notícia ruim, boa e nenhuma notícia $(\alpha\delta, \alpha(1 - \delta), 1 - \alpha)$. Por fim, tal função é representada pela equação (15).

$$\begin{aligned} L(\theta|B, S) = & \alpha(1 - \delta)e^{-(\mu+\varepsilon_b+\varepsilon_s)} \frac{(\mu+\varepsilon_b)^B (\varepsilon_s)^S}{B!S!} \\ & + \alpha\delta e^{-(\mu+\varepsilon_b+\varepsilon_s)} \frac{(\mu+\varepsilon_s)^S (\varepsilon_b)^B}{B!S!} \\ & + (1 - \alpha)e^{-(\varepsilon_b+\varepsilon_s)} \frac{(\varepsilon_b)^B (\varepsilon_s)^S}{B!S!} \end{aligned} \quad (15)$$

Segundo Easley, Hvidkjaer e O'Hara (2002), as probabilidades diárias de compra e venda no período $t \in [1, T]$ são independentes. Desse modo, a probabilidade conjunta durante esse intervalo é dada pelo produto dos dados diários.

$$V = L(\theta|M) = \prod_{t=1}^T L(\theta|B_t, S_t) \quad (16)$$

Em que B_t e S_t são os volumes de compra e venda nos dias $t = 1, \dots, T$ e $M [(B_1, S_1), \dots, (B_T, S_T)]$ é o conjunto de dados do período de negociações. Os parâmetros do modelo, isto é, o vetor $\theta = (\alpha, \mu, \varepsilon_b, \varepsilon_s, \delta)$, são estimados a partir da maximização da equação (16).

O modelo preocupa-se com a identificação de diferenças nos níveis de compra e venda de uma ação (EASLEY *et al.*, 1996). Assim, dias em que o volume comprado se equilibra ao volume vendido são interpretados como dias de negociações originadas de transações sem a presença de informações privadas, servindo, portanto, para estimar os valores de ε_b e ε_s . Volumes anormais de compras e vendas são tidos como transações oriundas de negociadores informados, a partir das quais, o parâmetro μ é identificado. Por fim, os valores de α e δ são estimados a partir do número de dias em que houve a ocorrência de volumes anormais (EASLEY, HVIDKJAER e O'HARA, 2002).

Visto se tratar de modelo probabilístico que envolve a ocorrência de um conjunto de diferentes eventos relacionados, a probabilidade de negociação com informação privada (PIN), segue a formulação em (17)

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_b + \varepsilon_s} \quad (17)$$

Easley, Hvidkjaer e O'Hara (2005) argumentam que, na equação (17), o numerador ($\alpha\mu$) representa a taxa de probabilidade de que haja negociações informadas, ao passo que o denominador ($\alpha\mu + \varepsilon_b + \varepsilon_s$) é o valor percentual da provável ocorrência de qualquer tipo de negociação (informada ou não).

Retomando as equações (10) e (11) fica claro que o método desenvolvido por Easley *et al.* (1996) procura averiguar se o risco de haver transações informadas é um fator explicativo do *spread* na visão do *market maker*. De forma a testar essa hipótese e a robustez do modelo em si, os autores primeiramente testam se o *spread* entre as diferentes ações estudadas varia com base em sua probabilidade de haver negociações informadas. Posteriormente, os autores regridem o *spread* tendo como variáveis independentes a PIN de cada ação e o volume transacionado (ver Easley *et al.* 1996, p. 23). Ambos os testes apontam para resultados favoráveis ao método PIN.

Por ser um modelo que utiliza dados intradiários e avalia diretamente a probabilidade da atuação de agentes informados, a PIN tem sido largamente utilizada na literatura financeira, tendo sido testada empiricamente nos mercados norte-americano (EASLEY *et al.* 1996, 2002, 2005, 2010); espanhol (ABAD, RUBIA, 2005); brasileiro (BARBEDO, SILVA e LEAL, 2009; MARTINS e PAULO, 2013, 2014), francês (AKTAS *et al.*, 2007), sul-coreano (HWANG *et al.*, 2013) dentre outros.

Como qualquer modelo, a PIN é uma forma de representar um fenômeno e, portanto, não consegue capturar todos os aspectos relativos a ele. Dessa maneira, esta também está sujeita a testes e eventuais críticas que ajudam a lapidar e sanar os problemas e lacunas que se apresentam na medida em que este é aplicado à realidade.

Um dos principais estudos que visam atestar a aplicabilidade da PIN foi conduzido por Aktas *et al.* (2007). Os autores procuram estudar a PIN acerca de anúncios de fusões e aquisições por parte das empresas, eventos estes conhecidos por serem alvo da atuação de

insiders que, portando informações de grande impacto nos preços das ações, realizam ordens de transações antes que tal fato seja divulgado ao público.

Segundo Aktas *et al.* (2007), existem irrefutáveis evidências de *inside trading* ou vazamento de informações perto de anúncios de fusões e aquisições, sendo, portanto, um momento ideal para testar a robustez da PIN. Os autores, estudando casos específicos ocorridos na França, encontram, surpreendentemente, resultados anti-intuitivos para os parâmetros da PIN. Esperava-se que a probabilidade da presença de agentes informados fosse maior momentos antes do anúncio. No entanto, o resultado aponta o contrário. A PIN obtida decresce durante uma janela de tempo pré-evento, seguida por um aumento da mesma durante um período pós-evento.

Uma das maiores contribuições dos autores supracitados é o estudo dos motivos que levaram os resultados auferidos a estarem na contramão do que era esperado. De acordo com Aktas *et al.* (2007), isto pode ser devido a má classificação das ordens, a não incorporação do volume de transações, ou a possibilidade de o modelo refletir outros fatores que não sejam informações privadas, como tendências globais do mercado. Easley, Hvidkjaer e O'Hara (2002) descartam a possibilidade de a inclusão do volume ser um fator-chave no cálculo da PIN. Baseado no trabalho de Easley *et al.* (1996), os autores mostram que, ao adicionar o volume, a complexidade computacional aumenta bastante e o poder explicativo continua estatisticamente o mesmo do modelo tradicional da PIN.

Duarte e Young (2009) apontam também que a PIN condensa a infinidade de motivos que levam os investidores a lançarem ordens de transações a somente duas: informações privilegiadas ou busca por liquidez, ou seja, todas as transações que não forem iniciadas por agentes informados são consideradas transações por liquidez. Outro ponto a ser levado em conta resulta da forma como a PIN captura a assimetria informacional. O modelo se vale dos desequilíbrios anormais das ordens de compra e vendas para o cálculo dos parâmetros. Duarte e Young (2009), no entanto, atentam ao fato de que tais desequilíbrios podem não ser frutos de transações informadas, mas, sim, refletirem choques ou mudanças na demanda por liquidez.

Aslan *et al.* (2011) mostram, porém, que, tanto um fator de PIN, quanto o índice de Amihud (2002) afetam independentemente os retornos de ativos. Quando aplicados em conjunto nos modelos de precificação, o índice de Amihud (2002) é estatisticamente

insignificante, o que reforça a hipótese de que é a informação e não liquidez que afeta os retornos dos ativos.

Devido às diversas críticas e à necessidade de maior poder computacional para a estimação dos parâmetros não observáveis e eventuais problemas na fatoração de quantidades grandes de ordens de compra e venda, especialmente em mercados de transações em alta frequência, Easley, López de Prado e O'Hara (2011a) desenvolveram um novo modelo de estimação da toxicidade dos fluxos das ordens de transações, denominado *Volume-Synchronized Probability of informed trading* (VPIN).

2.4 *Volume-Synchronized Probability of informed trading* (VPIN)

Em 06 de maio de 2010, o índice Dow Jones Industrial Average teve sua maior queda no período de um dia jamais registrado (EASLEY, LÓPEZ DE PRADO e O'HARA, 2011a). O declínio de 998,5 pontos (9%) foi parcialmente recuperado em torno de 20 minutos depois, com um retorno de aproximadamente 600 pontos que haviam sido perdidos (SONG, WU e SIMON, 2014). Tal episódio ficou conhecido como '*Flash Crash*' (ABAD e YAGÜE, 2012).

A reverberação desse evento foi tal que diversos autores se propuseram a explicá-lo. Easley, López de Prado e O'Hara (2011a) listam algumas das justificativas propostas para a queda do índice, dentre elas: o disparo de uma cascata de ordens de *stop loss* causada por um *fat-finger trade* dentro da P&G; dificuldades técnicas na NYSE; mudanças bruscas na paridade dólar/iene; larga aquisição de opções de compra por um fundo de *hedge*; e a presença de algoritmos predatórios que impediram os competidores de processarem os dados corretamente. O relatório completo da SEC levou quatro meses para ser emitido, demonstrando a complexidade desse evento para o mercado (BETHEL *et al.*, 2012). A explicação mais aceita é a de que o declínio do índice foi devido a um novo paradigma de transações, resultado de mudanças nas legislações americana e europeia (ABAD e YAGÜE, 2012).

Tal paradigma foi o resultado de novos métodos e algoritmos utilizados por empresas de investimento para agilizar o lançamento de ordens de compra e venda, muitas vezes, ganhando uma margem mínima em cada transação obtendo, porém, lucros via operações em grande escala (ABAD e YAGÜE, 2012). Diversos autores se propõem a pesquisar as

consequências que essa nova forma de operar pode causar ao mercado, especialmente em vista de episódios como o *Flash Crash* (YILDIZ, VAN NESS e VAN NESS, 2016).

Em mercados de alta frequência, a toxicidade dos fluxos de ordens pode levar os *market makers* a saírem do mercado, causando eventos de iliquidez para os agentes envolvidos (EASLEY, LÓPEZ DE PRADO e O'HARA, 2011a). A toxicidade foca na perda esperada de um *market maker* ao estar em um mesmo ambiente de um agente mais bem informado, ou seja, na probabilidade de os provedores de liquidez serem alvos de seleção adversa (ABAD e YAGÜE, 2012). Um fluxo de ordens é considerado tóxico quando os *market makers* provêm liquidez para o mercado sem estarem cientes de que o estão fazendo do lado em desvantagem da operação (EASLEY, LÓPEZ DE PRADO e O'HARA, 2012a). Um fluxo de ordens tóxico é antagônico a um fluxo de ordens benigno que advém de investidores não informados (EASLEY, LÓPEZ DE PRADO e O'HARA, 2011b).

A provisão de liquidez por parte dos formadores de mercado se dá de modo complexo, de forma que tomadores de posições podem deter informações acerca de um ativo que os *market makers* não têm. Em mercados de alta frequência, estes últimos procuram ganhar pequenas margens que inflam em transações com grandes quantidades de ordens. Dessa forma, a habilidade de estar do lado certo da operação depende do controle do risco de ser alvo de seleção adversa. A possibilidade do ganho desses formadores de mercado ao transacionarem uma grande quantidade de ações é maior quando há um equilíbrio entre os fluxos de ordens. Quando não é o caso, a maior chance de serem alvos de seleção adversa e, conseqüentemente, a alta toxicidade estimada por eles faz com que liquidem suas posições e saiam do mercado diminuindo a sua liquidez.

Motivados pelo *Flash Crash* e buscando desenvolver uma forma de prever grandes mudanças nos volumes de compras e vendas ligados a volatilidade, liquidez e toxicidade dos fluxos das ordens, Easley, López de Prado e O'Hara (2011a) adaptaram o modelo PIN para a incorporação do desequilíbrio das ordens de compra e venda baseado no volume das transações, denominado *Volume-Synchronized Probability of informed trading* (VPIN).

O modelo original desenvolvido por Easley *et al.* (1996) se vale da estimação de parâmetros não observáveis via função de máxima verossimilhança a partir da quantidade diária de compras e vendas para cada ação. Conforme as transações em alta frequência se tornam mais comuns, essa abordagem tradicional da PIN encontra dificuldades na estimação

desses parâmetros devido à fatoração de números muito grandes de ordens de operações. Apesar da modificação feita por Easley *et al.* (2008) para facilitação do cálculo da PIN em mercados com volume acentuado, persiste ainda a dificuldade da classificação das transações que, aliada a um subsequente aumento e rapidez na emissão de ordens via algoritmos de investimento faz com que esse modelo encontre problemas, como o *numerical overflow*, impedindo a conversão da função de máxima verossimilhança (GAN, WEI e JOHNSTONE, 2015).

A abordagem baseada no VPIN, no entanto, não requer a estimação de parâmetros não observáveis, além de ser atualizada em tempo estocástico, calibrada para ter um volume igual de transações em cada intervalo de tempo. Dessa forma, o VPIN supera as dificuldades apresentadas pela PIN em mercados altamente ativos, apresentando uma forma mais concisa de se tratar da toxicidade dos fluxos das ordens de compra e venda (EASLEY, LÓPEZ DE PRADO e O'HARA, 2012a).

A base teórica do VPIN é a mesma do seu predecessor PIN. Baseando-se nos trabalhos de Easley e O'Hara (1987, 1992), este se vale da perspectiva do formador de mercado diante de um ambiente propício à existência de agentes mais informados. Nesse cenário, o *market maker* atualiza suas crenças na medida em que chegam as ordens de compra e venda, ajustando o preço do *bid* e *ask*, avaliando suas chances de ser alvo de seleção adversa. O VPIN procura calcular diretamente a probabilidade da existência de agentes informados, ou seja, a equação final (17)

$$PIN = \frac{\alpha\mu}{\alpha\mu + \varepsilon_b + \varepsilon_s} \quad (17)$$

O cálculo é baseado nos resultados apresentados por Easley *et al.* (2008), estudo no qual os autores mostram que o valor esperado para a soma das ordens de compra e venda em um dia, denotado por $E[TT]$ é igual a soma da taxa de chegada de ordens de agentes informados e não informados, dado por (18)

$$E[TT] = \alpha(1 - \delta)(2\varepsilon + \mu) + \alpha\delta(\mu + 2\varepsilon) + (1 - \alpha)(2\varepsilon) = \alpha\mu + 2\varepsilon \quad (18)$$

Esse valor é igual ao denominador da equação (17). Outro conceito importante é o de desequilíbrio nas ordens, dado por $K = S - B$ em que S representa as ordens de venda e B as ordens de compra. O valor esperado para K é dado por (19)

$$E[K] = \alpha\mu(2\delta - 1) \quad (19)$$

Segundo Easley *et al.* (2008), a esperança dos valores absolutos das variáveis em uma distribuição de Poisson se dá de forma complexa, porém, o termo de primeira ordem dessa esperança se relaciona diretamente à chegada de ordens advindas de agentes informados, dessa maneira $E[K]$ é dado por (20)

$$E[K] \approx \alpha\mu \quad (20)$$

que é uma aproximação do valor dado pelo numerador do cálculo da PIN. Dessa forma, o VPIN pode ser determinado ao se calcular a quantidade total do volume de compra e venda e o desequilíbrio entre elas em um período estipulado. A determinação da janela temporal a ser utilizada no cálculo de K é feita por meio de uma nova forma de separação das ordens em tempo por volume, estipulada por Easley, López de Prado e O'Hara (2012c) em detrimento do tradicional tempo cronológico. Além disso, Easley, López de Prado e O'Hara (2012b) desenvolvem uma nova maneira de classificação das transações entre compra e venda denominada *Bulk Volume Classification*. Essas são duas das maiores contribuições dos autores ao proporem a métrica VPIN.

2.4.1 Tempo por volume e tempo cronológico

A maneira mais comum de se medir uma sequência de observações é pelo tempo. A separação entre eventos é feita, na maioria das vezes, cronologicamente, ou seja, separam-se os acontecimentos por anos, meses, dias, minutos, segundos e afins. Quando se trata da forma como as máquinas operam, no entanto, essa lógica não é mais absoluta. Uma máquina pode operar com base na ocorrência de eventos, denominado ciclo (EASLEY, LÓPEZ DE PRADO e O'HARA, 2012c). Um ciclo pode se completar em diferentes medidas cronológicas. Esse é o raciocínio a ser empregado quando se pensa em tempo por volume, que se enquadra como uma forma de medir o tempo baseando-se em eventos.

Easley, López de Prado e O'Hara (2012a) dizem que eventos com cargas de informações podem ocorrer frequentemente durante um dia, tendo diferentes impactos nas perspectivas da movimentação dos preços dos ativos. Em se tratando de modelos de alta frequência, as transações não ocorrem igualmente espaçadas no tempo, pelo contrário, elas chegam de forma irregular, conseqüentemente, existem ordens que trazem uma carga informacional mais relevante que outras. Easley e O'Hara (1992) desenvolveram a ideia de

que o tempo entre as ordens de transações é correlacionado com a existência de novas informações, provendo a base para se olhar para um tempo baseado em volume de transação ao invés do convencional tempo cronológico.

De forma a trabalhar com o tempo baseado em volume, Easley, López de Prado e O'Hara (2012c) advogam a utilização dos chamados baldes de volume (*volume buckets*). A ideia se sustenta em torno da divisão do volume transacionado em todo o período em uma quantidade estabelecida de baldes iguais que contêm um determinado volume (V). A cada preenchimento de um balde, tem-se um evento completo. Easley, López de Prado e O'Hara (2012a) afirmam que essa forma de se dividir o volume transacionado faz com que se pondere de forma mais precisa a chegada das ordens ao longo dos dias. Por exemplo, se em um dia não há movimentação forte no mercado, o preenchimento de um balde (um evento) demorará mais para acontecer. Por outro lado, em um dia de forte movimentação, ou seja, provável chegada de informação acerca do ativo, vários eventos ocorrem, podendo haver o preenchimento de diversos baldes de volume. Dessa forma, esse dia terá maior peso para o cálculo do VPIN.

O que esse procedimento procura fazer é imitar a chegada de informações no mercado. Ou seja, se uma notícia gera uma quantidade maior de volume, considera-se uma quantidade maior de observações, aumentando o seu peso na amostra. Presume-se, portanto, que, quanto mais relevante uma notícia, maior volume de transação ela gera (EASLEY, LÓPEZ DE PRADO e O'HARA, 2012a).

Segundo Easley, López de Prado e O'Hara (2012c), trabalhar com tempo por volume também apresenta vantagens estatísticas, sendo elas: a remoção das sazonalidades intradiárias, e a possibilidade de recuperação parcial de normalidade e das premissas que os dados são identicamente e independentemente distribuídos o que permite a aplicação de técnicas padrão estatísticas.

Para o cálculo do VPIN, portanto, estabelece-se a utilização do tempo por volume, ou seja, os volumes totais e o desequilíbrio entre as ordens de transações serão calculadas dentro de cada balde, que procura simular o término de um ciclo. Procede-se, então, para a questão da determinação da quantidade de ordens de compra e venda dentro de cada balde. Para isso, Easley, López de Prado e O'Hara (2012b) desenvolveram uma forma de classificação denominada *Bulk Volume Classification* (BVC).

2.4.2 Algoritmos de classificação de transações

Cada transação envolve ao menos um comprador e um vendedor. A determinação do lado que iniciou a operação pode ser usada para identificar uma possível assimetria informacional, e o desequilíbrio entre os volumes pode indicar a presença de toxicidade no fluxo das ordens (EASLEY, LÓPEZ DE PRADO e O'HARA, 2012b). Um predomínio de ordens de compra pode sugerir que a toxicidade advém da existência de notícias boas acerca da ação, enquanto o contrário indicaria a existência de notícias ruins sobre o ativo (EASLEY, LÓPEZ DE PRADO e O'HARA, 2012a).

Assinalar qual lado iniciou a transação se encontra entre as tarefas mais árduas em se tratando de mercados com alto volume transacionado. A obtenção desses dados é usualmente problemática. Assim como reforçado por Easley, López de Prado e O'Hara (2012b), mesmo quando os dados estão disponíveis, a quantidade de informações, comprimidas em arquivos enormes torna a manipulação e orientação dos mesmos muito trabalhosa e, por vezes, requer computadores potentes não disponíveis na maior parte dos pesquisadores. Frente a esse problema, surgiram diversos meios de assinalar o lado agressor das operações a partir dos dados disponibilizados pelos bancos de dados convencionais, como o *Tick Rule*, o *Quote Rule*, o método de *Lee-Ready* e, finalmente, o *Bulk Volume Classification*.

(a) Tick Rule

O *Tick Rule* (TR) classifica a operação como compra (venda) se o preço da transação for maior (menor) que o preço da transação anterior. Caso o preço seja o mesmo, repete-se a classificação dada à operação antecedente. Esse método tem a vantagem de classificar todas as transações ocorridas utilizando somente os preços das mesmas. No entanto, como apontado por Easley, López de Prado e O'Hara (2012b), o TR está sujeito a erros, especialmente em mercados com maior volume transacional. Chakrabarty, Pascual e Shkilko (2015) apontam que a eficácia desse método vem diminuindo no mercado norte-americano, caindo de 79% utilizando dados dos anos 1990, para 75% ao utilizar dados de 2005. Easley, López de Prado e O'Hara (2012b), porém, apontam uma precisão de 86% para esse método entre novembro de 2010 e novembro de 2011.

Aikten e Frinos (1996), analisando o mercado australiano, viram que o TR teve uma precisão de 75%. Ellis, Michaely e O'Hara (2000) detectam que o TR classificou corretamente 81% das transações na NASDAQ. A pesquisa mais recente, de Omrane e Welch (2016), constata que o TR classificou somente 67% das transações de forma exata, reforçando os achados de Chakrabarty, Pascual e Shkilko (2015) de que sua eficácia vem diminuindo, possivelmente pela maior rapidez das transações no mercado, o que dificulta estimações baseadas em *trade-by-trade*.

Perlin, Brooks e Dufour (2014) criam uma variação do método TR em que a precisão do algoritmo depende de três variáveis: o *spread* entre os preços de *bid* e *ask*, a volatilidade das inovações e a probabilidade de transição das operações. Os autores mostram que seu modelo prevê, no mínimo, 50% das transações corretamente. Um desempenho sem erros, no entanto, só poderia ser obtido em um cenário não realista. Ao analisar o mercado brasileiro, Perlin, Brooks e Dufour (2014) encontram uma precisão de 72% para esse novo método baseado no TR.

(b) Quote Rule

O *Quote Rule* (QR) necessita dos preços das transações, além do *best-bid-or-offer* (BBO) de cada ação para a classificação das ordens. O QR identifica a transação como compra (venda) se ela ocorre acima (abaixo) do ponto médio do BBO e deixa sem classificação aquelas que ocorrem exatamente no ponto médio. Devido à quantidade de transações que acontecem nessa cotação, o QR deixa, por vezes, um número alto de transações não identificadas (EASLEY, LÓPEZ DE PRADO e O'HARA, 2012b). Odders-White (2000), analisando dados da NYSE, relata que esse método deixou de classificar 15% das transações, além de ter identificado erroneamente 9% das demais. Ellis, Michaely e O'Hara (2000) apontam uma eficácia de 78% para esse método analisando a NASDAQ. Omrane e Welch (2016) encontram uma precisão de 77% para o QR.

(c) Método de Lee-Ready

O método de *Lee-Ready* (L-R) é o mais utilizado na identificação de transações em pesquisas em finanças (CHAKRABARTY, PASCUAL e SHKILKO, 2015) e, conseqüentemente, nos trabalhos que necessitam dessas informações como, por exemplo, o cálculo da PIN (OMRANE e WELCH, 2016). Esse algoritmo foi cunhado por Lee e Ready (1991) como uma combinação dos dois métodos supracitados. O L-R aplica a QR para todas

as transações fora do ponto médio e a TR, para aquelas que se encontram nessa cotação. Dessa forma, o L-R não deixa nenhuma transação sem ser identificada e utiliza dados mais refinados (*bid e ask prices*) para a classificação das ordens.

A precisão do L-R varia consideravelmente, sendo consenso entre os pesquisadores que, quanto mais desenvolvido e rápido o mercado, menor a eficácia desse método na identificação das transações (ASLAN *et al.*, 2011). Odders-White (2000) apresenta uma precisão de 85% aplicando esse método nos dados da NYSE, em 1991. Ellis, Michaely e O'Hara (2000) encontram uma eficácia de 83% investigando os dados da NASDAQ de 1996. Chakrabarty, Moulton e Shkilko (2012) apresentam uma diminuição considerável na precisão do L-R para o ano de 2005, quando os autores verificaram que esse método classificou corretamente cerca de 69% das transações.

Asquith, Oman e Safaya (2009) aplicaram o L-R para verificar a sua precisão juntamente com as vendas a descoberto e relatam que esse algoritmo identificou erroneamente cerca de 33% das transações (valor maior do que os métodos TR e QR, testados também pelos autores). Theissen (2000), analisando o mercado alemão, constata que o L-R classificou corretamente 73% das operações. O resultado de maior peso contrário à aplicação do método *Lee-Ready* vem de Hwang *et al.* (2013). Em estudo no mercado sul-coreano, os autores mostram que o L-R classifica erroneamente cerca de 58% das ordens iniciadas por compradores e 62% das ordens iniciadas por vendedores, o que condiz com o apontado por Aslan *et al.* (2011) de que esse algoritmo não tem bom desempenho em mercados de alta frequência. No mercado brasileiro, Silva (2009) aponta que sua precisão é de cerca de 72%.

(d) Bulk Volume Classification

Frente à necessidade de um modelo de classificação de transações com bom desempenho em mercados de alta frequência, Easley, López de Prado e O'Hara (2012b) desenvolveram um método que mitiga os efeitos da divisão de ordens (*order splitting*) e usa a mudança de preço padronizada para classificar o volume de transações em termos probabilísticos. Esse algoritmo, denominado *Bulk Volume Classification* (BVC), agrega as transações em intervalos de tempo ou volume (denominados barras de tempo ou barras de volume) - definidos de forma arbitrária ou limitada pela formatação do banco de dados - assinalando a cada período o último preço incluso no mesmo e utilizando a mudança de preço padronizada para determinar a porcentagem do volume de compras e vendas (EASLEY,

LÓPEZ DE PRADO e O'HARA, 2012b). A classificação se dá por meio da assinalação de cada barra τ à mudança de preços $P_\tau - P_{\tau-1}$ em que essas variáveis são os últimos preços inclusos nas barras τ e $\tau - 1$ respectivamente. Após tal procedimento, as frações do volume iniciadas pelos compradores e vendedores são calculadas como mostrado em (21) e (22).

$$V_\tau^B = V_\tau \cdot Z\left(\frac{P_\tau - P_{\tau-1}}{\sigma_{\Delta P}}\right) \quad (21)$$

$$V_\tau^S = V_\tau \cdot \left[1 - Z\left(\frac{P_\tau - P_{\tau-1}}{\sigma_{\Delta P}}\right)\right] = V_\tau - V_\tau^B \quad (22)$$

V_τ representa o volume total da barra τ , Z é a função de distribuição acumulada da distribuição normal padrão, e $\sigma_{\Delta P}$ é o desvio-padrão estimado da mudança de preços entre as barras. Segundo Easley, López de Prado e O'Hara (2012a), esse procedimento divide o volume da barra igualmente entre compra e venda caso não haja mudança de preço do início ao fim desta. Alternativamente, caso o preço aumente (diminua), maior fração do volume é classificada como tendo sido iniciada por compradores (vendedores). Quanto maior a variação do preço, maior fração do volume é classificada como compras ou vendas.

O BVC trata a mudança de preços como sendo i.i.d com média zero e variância $\sigma_{\Delta P}$. Easley, López de Prado e O'Hara (2012b) afirmam que se pode calcular a verdadeira função de distribuição acumulada da mudança de preços. No entanto, esse processo levaria a diversos problemas, incluindo uma diminuição na eficácia da classificação das transações. Dessa forma, a suposição de normalidade é o procedimento mais adequado na estimação do volume de compra e venda de uma ação.

Segundo Chakrabarty, Pascual e Shkilko (2015), o BVC é proposto de forma a substituir classificações discretas baseadas em *tick-by-tick*, como o TR, QR e L-R por uma classificação contínua de natureza probabilística. Uma diferença-chave entre o BVC e as demais é que este identifica uma fração do volume como compra e o resto como venda, enquanto as outras identificam a transação ou como compra ou como venda.

Easley, López de Prado e O'Hara (2012b) afirmam que, por utilizar a aproximação da normal, o BVC tem bom desempenho em ações com alto volume transacionado. Os autores atentam ao fato de que, para aqueles ativos pouco líquidos e, portanto, com poucas transações, os métodos *tick-by-tick* são mais adequados. Para estabelecer a precisão do BVC, Easley, López de Prado e O'Hara (2012b) comparam os resultados dos métodos TR e BVC com os dados reais obtidos e exaustivamente tratados pelos autores. Os pesquisadores dizem que o

BVC classifica corretamente 86,61% das transações dos contratos futuros do e-mini S&P500 usando barras de tempo de um minuto e 87,35% para barras de tempo de cinco minutos. Ao utilizar barras de volume de 10.000 transações, a precisão aumenta para 90,72%. Enquanto isso, a precisão do TR se mantém em 86,43%. O mesmo resultado é alcançado pelos autores trabalhando com outros contratos futuros como o *WTI Crude Futures* e *Gold Futures*.

Outra grande vantagem em aplicar o BVC em detrimento dos métodos baseados em *tick-by-tick* está no tamanho do banco de dados utilizado. Os dados em barras de volume representam 0,60% do total de dados em *tick*, enquanto as barras de tempo de um minuto representam 0,35%. Ou seja, atinge-se maior precisão com uma quantidade extraordinariamente menor de dados a se processar.

Chakrabarty, Pascual e Shkilko (2015) atentam para o fato de que Easley, López de Prado e O'Hara (2012b) testam a precisão do BVC no mercado de futuros. Dessa forma, os autores procuram comparar os desempenhos dos métodos TR, L-R e BVC nos mercados à vista. De porte dos dados reais, os pesquisadores analisam 1.471 ações e afirmam que o TR e o L-R apresentam precisão maior que o BVC. Especificamente, o TR e o L-R classificam corretamente 90,8% e 92,6% das transações. A melhor precisão apresentada para a abordagem BVC é de aproximadamente 80% quando se utilizam barras de tempo de um minuto. Omrane e Welch (2016) reforçam esses achados ao analisarem o mercado de moedas estrangeiras. Os autores verificam que, tanto o TR, quanto o QR apresentam resultados melhores do que o BVC.

Easley, López de Prado e O'Hara (2016), em vista das críticas ao BVC, destacam que, para pesquisas acerca de informações privilegiadas, o que deve ser buscado é a informação subjacente contida nas transações, a qual é um dado não observável, dependendo, portanto, de *proxies* para essa informação. Os autores comparam a precisão do TR contra o BVC utilizando três *proxies* para esse dado subjacente das transações. Easley, López de Prado e O'Hara (2016) afirmam que o BVC tem um melhor desempenho para duas das três *proxies* e concluem que o BVC é útil para a distinção das intenções relativas às transações observadas, além das vantagens computacionais supracitadas.

2.4.3 Cálculo do VPIN

Retomando-se os resultados apresentados por Easley *et al.* (2008), tem-se que, para cada período, o número total de transações e o desequilíbrio entre elas são dados por (23) e (24).

$$E[TT] = E[V_{\tau}^B + V_{\tau}^S] = \alpha\mu + 2\varepsilon \quad (23)$$

$$E[K] = E[V_{\tau}^S - V_{\tau}^B] \approx \alpha\mu \quad (24)$$

Divide-se um dia em baldes de volumes iguais, tratando cada um equivalente a um período de chegada de informação. Dessa forma, $V_{\tau}^B + V_{\tau}^S$ é constante e igual a V para todo τ . Aproxima-se, então, o desequilíbrio de transações pela média calculada sobre n baldes de volume. Dessa forma, o VPIN pode ser calculado por (25).

$$VPIN = \frac{\alpha\mu}{\alpha\mu + 2\varepsilon} = \frac{\alpha\mu}{V} \approx \frac{\sum_{\tau=1}^n |V_{\tau}^S - V_{\tau}^B|}{nV} \quad (25)$$

A estimação do VPIN necessita da determinação das variáveis V e n . Para se calcular o VPIN para um dia, por exemplo, estabelece-se V como sendo um cinquenta avos da média diária do volume e n como sendo 50. Dessa forma, calcula-se o VPIN para cinquenta baldes, correspondente ao VPIN diário. Se há o intuito de calcular o VPIN semanal, substitui-se o valor de n por 250, por exemplo.

O VPIN é atualizado a cada balde de volume. Assim, quando se preenche o balde número 51, descarta-se o primeiro e calcula-se o novo VPIN utilizando os baldes 2 a 51. Segundo Easley, López de Prado e O'Hara (2012a), deseja-se que a velocidade de atualização do VPIN simule a velocidade de chegada de informações ao mercado. Além disso, é interessante que cada atualização seja feita com base em uma quantidade comparável de informação. Como os volumes podem ser muito desbalanceados durante dias com poucos participantes, ou seja, baixa probabilidade de chegada de informação, atualizar o VPIN com base em tempo cronológico levaria a resultados baseados em quantidades desiguais de informação.

A métrica VPIN ganhou atenção da literatura acadêmica por se basear na estrutura teórica do já consolidado PIN procurando sanar suas críticas, além de ter tido sucesso na previsão do *Flash Crash*. Abad e Yagüe (2012) apontam as principais diferenças entre os dois

modelos e apresentam as vantagens do VPIN em relação ao seu predecessor. O modelo PIN foca o valor real de uma ação e leva em consideração a probabilidade das informações de impactar o mercado na medida em que *insiders* atuam desequilibrando a quantidade de ofertas de compra e vendas. Já o VPIN mensura a toxicidade do fluxo das ordens, que se relaciona diretamente com a probabilidade dos provedores de liquidez de serem alvos de seleção adversa. Dessa forma, o conceito de informação no VPIN tem a ver com os eventos subjacentes que provocam desequilíbrio nas ordens em um curto período de tempo.

Uma segunda diferença entre os dois modelos é o sistema de tempo empregado por eles. O PIN utiliza o tempo cronológico para sua estimação, baseando-se na quantidade de ordens de compra e venda diárias sob o pressuposto de independência entre os dias. O VPIN, por outro lado, computa o desequilíbrio entre as ordens a cada vez que o mercado transaciona certa quantidade de ações, completando um ciclo, também chamado de tempo por volume. Como dito anteriormente, essa metodologia simula a chegada ao mercado de informações de relevância comparável.

Outra vantagem do VPIN é que ele pode indicar a probabilidade da atuação de *insiders* ao longo do dia por ser atualizado conforme terminam os ciclos (baldes de volume). Além disso, essa métrica não requer a estimação de parâmetros não observáveis, tarefa particularmente árdua nos mercados de alta frequência e com grande quantidade de volume transacionado.

De forma a validar o VPIN empiricamente, Easley, López de Prado e O'Hara (2011a) avaliaram essa métrica de janeiro de 2008 até outubro de 2010, focando os dias em torno do *Flash Crash*. Os autores mostram que o VPIN atentou para problemas de liquidez horas antes do colapso do índice Dow Jones Industrial Average. Segundo Easley, López de Prado e O'Hara (2011a), o VPIN ultrapassou o limite de 90% da função de distribuição acumulada na manhã do dia 06 de maio de 2010, indo para 95% no início da tarde e atingindo o seu maior valor em todo o período analisado minutos antes da queda brusca do índice. Observando especificamente o dia do *Flash Crash*, o VPIN atinge seu valor máximo concomitante à queda do índice e se mantém incomumente alto até o fim do dia, diminuindo o seu valor conforme o índice recuperava os pontos perdidos.

Easley, López de Prado e O'Hara (2011a) sugerem que o *Flash Crash* foi causado por uma crise de liquidez devido a características estruturais do novo paradigma de mercado de

alta frequência (responsável por mais de 70% do volume transacionado no mercado americano desde 2009). De acordo com os autores, em mercados de alta frequência, as transações são comandadas por algoritmos programados para lançar ordens de compra e venda, sendo, portanto, as firmas de alta frequência as grandes provedoras de liquidez nos novos mercados. Conforme a toxicidade do fluxo das ordens aumenta, esses formadores de mercado se encontram diante de perdas significantes e diminuem seus riscos ao reduzirem ou liquidarem suas posições. A consequente queda de liquidez no mercado pode ter sérias repercussões para os agentes nele envolvidos, como evidenciado pelo *Flash Crash*.

A fama angariada pelo VPIN por ter previsto o *Flash Crash* chamou a atenção de diversos pesquisadores, entre eles, Andersen e Bondarenko (2014a) que, ao replicarem o estudo de Easley, López de Prado e O'Hara (2011a), afirmam que o VPIN atinge seu valor máximo somente depois da queda do índice, contrariando os resultados a favor do poder preditivo dessa nova métrica. Andersen e Bondarenko (2014a), no entanto, ao calcularem o VPIN, classificam as transações agrupadas em barras de volume como sendo todas vendas ou compras, método este rapidamente criticado por Easley, López de Prado e O'Hara (2014). Estes últimos apontam que, longe de replicar os resultados, Andersen e Bondarenko (2014a) criticam uma metodologia não utilizada pelos autores, uma análise não realizada e conclusões não tiradas por eles.

Andersen e Bondarenko (2014b) reiteram a disputa iniciada em Andersen e Bondarenko (2014a) afirmando que os resultados alcançados quando utilizado o BVC foram os mesmos, ou seja, o VPIN não atingiu um valor máximo momentos antes do *Flash Crash*. Os autores terminam dizendo que, ao utilizarem as quantidades reais de ordens de compra e venda, eles encontraram valores para o VPIN opostos àqueles achados nos estudos de Easley, López de Prado e O'Hara (2011a, 2012a).

Easley, López de Prado e O'Hara (2014) explicam que o VPIN é relacionado à toxicidade das ordens, a qual prejudica os *market makers* que saem do mercado quando tal métrica apresenta valores acentuados, afetando, assim, a liquidez e podendo resultar em uma volatilidade de curto prazo do preço da ação. Dessa forma, os autores apontam que há um componente de curto prazo na volatilidade relacionada à toxicidade. No entanto, reiteram que a volatilidade não é somente devida à toxicidade das ordens.

A partir da divulgação dessa nova métrica, outros autores se propuseram a testá-la. Bethel *et al.* (2012) utilizam computadores super potentes para encontrar indicadores que possam prever episódios de iliquidez do mercado e implementar um sistema de aviso para que os reguladores possam tomar medidas para evitar quedas de mercado como o *Flash Crash*. Os autores replicam os estudos de Easley, López de Prado e O'Hara (2011a) analisando ações individuais ao invés do índice e-mini S&P500 e reforçam as conclusões sobre o potencial preditivo do VPIN. Segundo os autores, indicadores como o proposto por Easley, López de Prado e O'Hara (2011a, 2012a) podem substituir medidas bruscas como a suspensão das transações quando o preço ou a volatilidade atingirem um limite por uma diminuição gradual das transações que permitiriam ao mercado se estabilizar e se corrigir.

Wei, Gerace e Frino (2013) analisam o mercado acionário australiano e encontram fortes evidências de que o VPIN, assim como a PIN, analisado por Easley, Hvidkjaer e O'Hara (2002), é correlacionado com o tamanho da empresa e liquidez das ações. Os autores mostram que um maior valor de VPIN é associado a empresas menores e ações menos líquidas, enquanto menores VPINs são encontrados para empresas maiores e ações mais líquidas. Os pesquisadores afirmam também que o VPIN é relacionado com características intradiárias das ações, como volatilidade do preço, tempo necessário para o enchimento de um balde de volume e desequilíbrio do *quote* (diferença entre os *bid* e *ask prices* dentro de um balde).

Cheung, Chou e Lei (2015) aplicam o tradicional PIN juntamente ao VPIN no mercado acionário de Hong Kong para investigar a validação deste último como medida de toxicidade dos fluxos de ordens. Os autores analisam contratos lançados nesse mercado que consistem de um limite para o preço, que, se alcançado, aciona o chamado evento de compra mandatório (MCE) no qual o contrato é imediatamente terminado. Diferentemente dos estudos acerca do *Flash Crash*, nos EUA, os pesquisadores argumentam que os MCEs provêm de um conjunto maior de eventos para testar a validade do VPIN como uma medida de risco de liquidez intradiário.

Cheung, Chou e Lei (2015) asseveram que a toxicidade dos fluxos de ordens, representado por VPIN, aumenta significativamente antes de um MCE e continua a crescer por um curto período após o evento. A métrica retorna a níveis normais após o restabelecimento do contrato. Os pesquisadores também mostram que, nos dias em que

acontecem os MCEs, existe uma quantidade considerável de investidores não informados atuando, o que sugere uma porção significativa de transações especulativas dos ativos subjacentes aos contratos analisados.

Por fim, Cheung, Chou e Lei (2015) se direcionam às críticas de Andersen e Bondarenko (2014a) de que, para determinar a validade do VPIN, deve ser evidenciado o poder preditivo incremental para volatilidade e volume. Os autores regridem a ocorrência ou não dos MCEs com os valores encontrados para o VPIN, volatilidade defasada e volume transacionado. Cheung, Chou e Lei (2015) mostram que o VPIN tem um poder preditivo incremental significativo sobre os MCEs quando controlado para as duas últimas variáveis.

Yildiz, Van Ness e Van Ness (2016) estudam a relação do VPIN com os mercados de alta frequência procurando responder três questões: como os provedores de liquidez afetam a toxicidade das ordens nos mercados? Períodos em que os *market makers* saírem do mercado, ou diminuirão suas posições, podem ser previstos? Como os agentes que transacionam em alta frequência afetam a variação do preço durante períodos de alta toxicidade de ordens?

Os autores dizem que existe uma relação negativa entre transações em alta frequência e toxicidade das ordens. Além disso, eles mostram que a métrica VPIN pode detectar mudanças na demanda e oferta de liquidez pelos agentes que transacionam em alta frequência para até dez baldes de volume à frente. Tal assertiva leva a crer que os reguladores de mercado podem usar o VPIN, calculado em tempo real, para traçar e implementar estratégias protecionistas a fim de evitar crises de liquidez. Yildiz, Van Ness e Van Ness (2016) constatam também que a atuação dos agentes de alta frequência pode ser benéfico para ações com médio e alto volumes. No entanto, para ativos de baixo volume, esses agentes podem ajudar a aumentar a volatilidade dos retornos. Assim, os autores concluem que os impactos dessas transações não são uniformes e variam dependendo do nível de toxicidade das ordens e do volume transacionado.

Para poder fazer a conexão entre essa área de microestrutura de mercado, que trata basicamente do conteúdo informacional e a mecânica das transações, aos estudos sobre *asset pricing*, é necessário, primeiramente, fazer uma revisão sobre os modelos e em quais deles pode ser incorporado um fator relativo ao risco informacional de uma ação.

2.5 Modelos de precificação de ativos

Diversos modelos de precificação de ativos foram propostos na literatura ao longo dos anos, sendo um dos primeiros e mais impactantes o modelo de precificação de ativos de capital (CAPM) de Sharpe (1964). Este trabalho foi fortemente influenciado por aquele que representa uma extensão da moderna teoria de portfólios de Markowitz (1952) considerado como o início das finanças modernas. O pressuposto central do modelo é que o portfólio de mercado é eficiente no sentido média-variância. O CAPM pressupõe que o retorno esperado dos ativos seja uma função linear do seu beta, multiplicado pelo prêmio de risco de mercado, somado a um retorno de um ativo livre de risco, como mostra (26).

$$R_i = R_f + \beta_i(R_M - R_f) \quad (26)$$

Em que R_f e R_M representam respectivamente o ativo livre de risco e o retorno de um portfólio *proxy* de mercado. A inclinação da regressão, β_i , representa o risco sistemático do ativo i .

Seguindo o impacto que esse modelo teve na literatura financeira, diversos autores contribuíram para a sua construção (LINTNER, 1965; MOSSIN, 1966; BLACK, 1972) e, naturalmente, outros procuraram testar a validade do modelo em diferentes mercados (BANZ, 1981; BHANDARI, 1998; CHAN, HAMAIO e LAKONISHOK, 1991; FAMA e MACBETH, 1973). Fama e MacBeth (1973) testaram quatro hipóteses, nas quais, o CAPM se baseia, para o período de 1926 a 1968: (i) na linearidade do risco capturado pelo beta; (ii) na existência de outra fonte de risco além do beta da ação; (iii) no *trade-off* positivo entre o risco e retorno; e (iv) no intercepto da regressão que representa o ativo livre de risco. Tal teste proposto pelos autores pode ser entendido por meio da regressão

$$R_{it} = \gamma_{0t} + \gamma_{1t}\beta_i + \gamma_{2t}\beta_i^2 + \gamma_{3t}S_i + \varepsilon_{it} \quad (27)$$

em que Fama e MacBeth (1973) testaram, respectivamente, se $\gamma_{2t} = \gamma_{3t} = 0$, hipóteses (i) e (ii); $\gamma_{1t} > 0$, hipótese (iii); e, finalmente, se $\gamma_{0t} = R_f$, hipótese (iv). Em resumo, nenhuma das quatro hipóteses pôde ser rejeitada, ou seja, a precificação dos ativos está de acordo com os pressupostos do modelo de dois parâmetros para os retornos esperados.

Os resultados apresentados por Fama e MacBeth (1973) indicam que o modelo de dois parâmetros (risco e retorno) poderia ajudar a explicar os retornos dos ativos. No entanto, isso não significa que o único fator explicativo seja o beta da ação. Banz (1981) assevera que o tamanho da empresa – *market equity* (ME), ou seja, o preço de uma ação multiplicado pelo número de ações em circulação – também é um fator explicativo, na medida em que retornos de ações com um baixo ME tendem a ser altos e vice-versa.

Bhandari (1988) mostra que existe uma relação positiva entre alavancagem e retorno de uma ação. Isso se apresenta como uma contradição ao modelo CAPM, tendo em vista que o risco de alavancagem deveria ser capturado pelo beta da ação. O autor fala que, incluindo um índice de alavancagem em uma regressão com o beta e o ME, este ajuda a explicar os retornos dos ativos. Rosenberg, Reid e Lanstein (1985) e Chan, Hamao e Lakonishok (1991) afirmam que a razão entre o valor contábil e o *market equity*, chamado de *book-to-market* (BM ou BE/ME), também é um fator explicativo dos retornos médios nos mercados americano e japonês. Além disso, Basu (1983) mostra que um índice formado pela lucro sobre preço (E/P), ao lado do beta, explica os retornos médios dos ativos no mercado norte-americano.

Fama e French (1992) afirmam que os índices de alavancagem, ME, E/P, e BE/ME são versões escalonadas do preço do ativo. Com isso existe a possibilidade de esses fatores serem redundantes, ou seja, explicarem o mesmo fenômeno. De forma a verificar tal hipótese, seguem a estimação de regressões uni e multivariadas com esses fatores ao lado do beta dos ativos.

Ao analisar o período 1963-1990, os autores não encontram uma relação significativa entre o beta e o retorno médio, o que contradiz o achado de Fama e MacBeth (1973) – que trabalham com o período 1926-1968 – e também a premissa básica do modelo CAPM da existência de um *trade-off* positivo entre risco e retorno – terceira hipótese testada por Fama e MacBeth (1973). Em relação às demais variáveis, os autores dizem que a utilização conjunta do tamanho e *book-to-market* absorvem os efeitos da alavancagem e do índice E/P. Dessa forma, Fama e French (1992) concluem que os riscos das ações são multidimensionais, em que uma das dimensões é capturada pelo tamanho e a outra é capturada pelo *book-to-market*, ao menos para o período 1963-1990.

Em um trabalho posterior, Fama e French (1993) incorporam os retornos de títulos públicos e debêntures, além dos retornos de ações, reforçando os achados de Fama e French (1992) para as variáveis tamanho e *book-to-market*. Além disso, os autores dizem que existem,

pelo menos, três fatores de mercado e dois fatores estruturais que afetam os retornos dos ativos pesquisados. Estes dois últimos se relacionam aos títulos públicos e debêntures e são, respectivamente: a diferença entre o retorno de um título público de longo prazo e o retorno mensal da letra do tesouro norte-americano; e a diferença entre o retorno de um portfólio de mercado de debêntures de longo prazo e o retorno dos títulos públicos de longo prazo.

Em relação aos três fatores de mercado, eles são os conhecidos: SMB (*small minus big*), que é a diferença de retornos entre um portfólio composto por ações de pequena capitalização de mercado e outro portfólio composto por ações grandes; HML (*high minus low*), que representa a diferença de retornos entre um portfólio de ações com alto *book-to-market* e um outro portfólio de ações com baixo *book-to-market*; e, finalmente, $R_M - R_f$, que constitui o fator de mercado, também utilizado pelo CAPM, em que o retorno de mercado é subtraído pelo retorno de um ativo livre de risco. Esse trabalho abriu caminho para a construção do consagrado modelo de três fatores de Fama e French (1993,1996), representado por (28)

$$R_i - R_f = \alpha_i + b_i(R_M - R_f) + s_iSMB + h_iHML + \varepsilon_i \quad (28)$$

Fama e French (1996) argumentam que esse modelo consegue explicar os padrões nos retornos não interpretados pelo modelo CAPM, comumente chamados de anomalias. As evidências na aplicação de (28) sugerem um intercepto (α_i) não estatisticamente diferente de zero e um R^2 médio de 0,93 nas 25 regressões rodadas pelos autores. Além disso, Fama e French (1996) falam que os resultados empíricos mostram que ele se apresenta como um modelo de precificação em equilíbrio, ou seja, uma versão de três fatores do modelo geral do APT de Ross (1976).

Apesar do aparente sucesso do modelo de três fatores comparado ao CAPM, Fama e French (1996) atentam que, por ser um modelo, este não é capaz de explicar os retornos de todos os ativos e portfólios. Dessa forma, Carhart (1997) se debruça sobre o problema e sugere a adição de um quarto fator, denominado fator de momento (*UMD – up minus down*), baseado no trabalho de Jagadeesh e Titman (1993). Esse fator é calculado tomando-se a média igualmente ponderada das empresas com os 30% maiores retornos nos últimos onze meses defasados em um mês, menos a média igualmente ponderada das empresas com os 30% menores retornos nos últimos onze meses defasados em um mês. O modelo proposto por Carhart (1997) é representada por (29)

$$R_i - R_f = \alpha_i + b_i(R_M - R_f) + s_iSMB + h_iHML + m_iUMD + \varepsilon_i \quad (29)$$

A adição do fator de momento (*UMD*) reduz os erros de precificação apresentados, tanto pelo CAPM, quanto pelo modelo de três fatores (CARHART, 1997). O denominado modelo de quatro fatores eliminou quase por completo os padrões nos erros de precificação, indicando uma robustez na explicação dos retornos médios das ações no período analisado pelo autor.

Devido a evidências surgidas na literatura ao longo dos anos (NOVY-MARX, 2013; TITMAN, WEI e XIE, 2004) de que o modelo de três fatores não é capaz de explicar a variação nos retornos médios relativos à lucratividade e investimento, Fama e French (2015) revisitam o modelo, adicionando a ele dois fatores. O primeiro deles, denominado *RMW* (*robust minus weak*), é obtido pela diferença entre os retornos de portfólios de ações com alta e baixa lucratividade. Já o fator referente ao investimento, denominado *CMA* (*conservative minus aggressive*), é a diferença entre o retorno de portfólios de ações com baixo e alto investimentos. O modelo em sua forma completa é dado por (30).

$$R_i - R_f = \alpha_i + b_i(R_M - R_f) + s_iSMB + h_iHML + r_iRMW + c_iCMA + \varepsilon_i \quad (30)$$

Fama e French (2015) fazem várias contribuições em relação ao modelo anterior. Primeiramente, há a possibilidade da criação dos fatores por meio de combinações diferentes da usada em Fama e French (1993, 1996). Os cinco fatores podem ser obtidos classificando as ações de 2×2 e $2 \times 2 \times 2 \times 2$, diferentemente do 2×3 utilizado nos trabalhos anteriores. A divisão das ações, nesses novos casos, é dada pela mediana das variáveis analisadas. Além disso, os autores afirmam que, para a amostra utilizada, a variável HML se torna redundante, podendo ser substituída pela chamada HMLO (*orthogonal high minus low*) produzindo um modelo de cinco fatores alternativos. De forma a corroborar a decisão pela utilização do modelo (30), Fama e French (2015) mostram que o valor para a estatística de Gibbons, Ross e Shanken (1989) – GRS – é menor para o modelo de cinco fatores do que para o de três. Ainda sim, os autores admitem a incompletude do modelo. Afirmam, porém, que ele tem um desempenho melhor do que seus predecessores.

Baseando-se nos modelos de precificação com solidez na literatura, diversos autores procuraram encontrar uma relação entre a probabilidade de haver negociações privilegiadas em um mercado e o retorno requerido das ações. Por vezes, esses trabalhos se valeram da

mensuração direta pela *proxy* PIN e, em outros casos, formou-se um fator PIN, relacionado aos fatores propostos por Fama e French (1996) e Carhart (1997). Os resultados encontrados pelos autores apresentam divergências, tendo, em alguns trabalhos, sido constatada uma relação positiva entre a PIN e o retorno requerido e, em outros, inexistência de relação entre eles.

2.6 PIN e os modelos de precificação de ativos

Entre os trabalhos que procuraram incorporar um fator de risco informacional aos modelos de precificação de ativos destacam-se o de Easley, Hvidkjaer e O'Hara (2002, 2005) e Aslan *et al.* (2011), estudos nos quais os pesquisadores se propuseram a analisar empiricamente a influência da PIN sobre o retorno requerido de ações negociadas no mercado acionário norte-americano; além do de Easley e O'Hara (2004), no qual os autores estudaram a diferença de composição de informações pública e privada de uma ação e seus efeitos sobre o custo de capital. Diversos estudiosos seguiram essa trilha devido às questões ainda em aberto sobre o assunto, com destaque para Mohanram e Rajgopal (2009) que encontram resultados contraditórios aos de Easley, Hvidkjaer e O'Hara (2002) e Hwang *et al.* (2013) que reforçam os resultados encontrados por Easley, Hvidkjaer e O'Hara (2002) afirmando que, se o número de ordens de compra e vendas pode ser precisamente apurado, a PIN é um fator de precificação ao utilizar retornos esperados.

Easley, Hvidkjaer e O'Hara (2002) abriram caminho para a incorporação da PIN como fator explicativo em modelos de precificação de ativos ao procurarem responder a seguinte pergunta: uma maior probabilidade de transações informadas afeta o retorno requerido de um ativo? Baseando-se no trabalho de Fama e French (1992), os autores chegaram à conclusão de que um aumento de 10% na PIN de uma ação é associado a um aumento do retorno anual realizado de 2,5%.

Os modelos de precificação clássicos não incluem fatores que influenciam a estrutura informacional de uma ação, como a microestrutura do mercado e análises financeiras por partes de profissionais especializados. Modelos como o de Fama e French (1992), porém, admitem a possibilidade de outros fatores além do risco de mercado de influenciarem o retorno requerido de uma ação (EASLEY e O'HARA, 2004). Easley e O'Hara (2004)

constroem um modelo de equilíbrio de expectativas racionais em que eles dividem a informação em parte pública e parte privada. Os autores dizem que a distribuição de informação privada afeta o retorno requerido pelos investidores quando em equilíbrio. Basicamente uma ação que tem mais informações privadas tem um maior custo de capital.

Os autores reforçam também o papel dos analistas e da precisão de dados contábeis em relação aos retornos requeridos. Segundo Easley e O'Hara (2004), as empresas se beneficiam quando diversos analistas apresentam previsões acerca das mesmas, na medida em que coletivamente haverá maior precisão informacional no mercado o que levará a uma diminuição do custo de capital dessas companhias. Além disso, relatórios contábeis confiáveis diminuem o risco para o investidor não informado. O modelo dos autores demonstra que, dado que a informação afeta o preço dos ativos, a quantidade e a qualidade de tal informação são bastante relevantes na precificação das ações.

Easley, Hvidkjaer e O'Hara (2005) revisitam a questão da precificação de ativos ao reforçarem o papel da PIN como fator explicativo do retorno dos ativos. Segundo os autores, o risco derivado da assimetria informacional se dá uma vez que agentes informados podem reestruturar seus portfólios mais rapidamente do que aqueles indivíduos sem acesso a tais informações. Por consequência, estes últimos não poderão diversificar eficientemente, abrindo espaço para que agentes informados recebam maiores retornos baseados em suas informações.

A principal contribuição de Easley, Hvidkjaer e O'Hara (2005) em relação a Easley, Hvidkjaer e O'Hara (2002) é a criação de um fator PIN. A partir deste, os autores criaram os chamados *PIN-size portfolios* e regrediram seus retornos com os três fatores de Fama e French (1993, 1996) e o fator de momento de Carhart (1997), adicionado o fator PIN. Os resultados encontrados apontam para um intercepto estatisticamente insignificante para oito dos dez portfólios quando acrescido o fator PIN na regressão. O modelo utilizado pelos autores é apresentado em (31).

$$R_i = \alpha_i + \beta_i R_M + s_i SMB + h_i HML + m_i UMD + p_i PINF_{-i} + \epsilon_i \quad (31)$$

em que se encontram os três fatores de Fama e French (1996), o fator de momento de Carhart (1997) e o fator PIN de Easley, Hvidkjaer e O'Hara (2005).

Aslan *et al.* (2011), seguindo a mesma linha de Easley, Hvidkjaer e O'Hara (2002 e 2005), criam um fator PIN (PPIN) que serve como *proxy* para o risco informacional de uma ação. Os autores afirmam que, investindo \$1 (uma unidade de dólar americano) em um

portfólio com alto valor para esse fator, em 1969, equivaleria a \$590, em 2009, em contraste com os \$40,62 resultantes se esse mesmo dólar fosse investido em um portfólio com baixo valor para o PPIN.

De forma a verificar a relação da PIN com os diferentes setores do mercado, os autores regridem a PIN de cada ação com as características setoriais como variáveis explicativas, por exemplo: tamanho, *turnover*, Q de Tobin, número de analistas que fizeram previsões no ano anterior dentre outras – ver Aslan *et al.* (2011). Os resultados encontrados são consistentes com o esperado: empresas maiores têm menos probabilidade de serem alvos de transações privilegiadas; o número de acionistas ordinários e a cobertura de analistas são negativamente relacionados à PIN, o que indica que essa classe de acionistas tende a não ter informações privilegiadas e que um maior número de previsões feitas por analistas leva a uma menor probabilidade de negociações informadas – corroborando a hipótese de Easley e O'Hara (2004).

O resultado mais expressivo de Aslan *et al.* (2011) se relaciona à separação entre liquidez e risco informacional atentado primeiramente por Duarte e Young (2009). Aslan *et al.* (2011) realizam três regressões. Primeiramente regridem o retorno dos ativos com o beta, tamanho, *book-to-market*, retornos passados e o fator PIN (PPIN). Em seguida, substituem o PPIN pelo índice de liquidez de Amihud (2002). Ambos os testes apontam para a precificação do fator e índice separadamente. Por fim realizam a regressão com ambos atuando como variáveis explicativas. Tanto o índice de Amihud (2002), quanto o PPIN são negativamente correlacionados com o tamanho (-0,92 e -0,94 respectivamente) e positivamente correlacionados entre si (0,96). Isso indica que ambos podem capturar características similares (ASLAN *et al.*, 2011). O resultado da regressão com os dois índices em questão indica que o PPIN continua a ser precificado enquanto o coeficiente do índice de Amihud (2002) declina de 0,175, na primeira regressão, para 0,039 quando o PPIN está presente. Aslan *et al.* (2011) concluem que o risco informacional capturado por PPIN é um fator de precificação de ativos juntamente com o tamanho e o momento. O resultado da última regressão também indica que o PPIN compromete o efeito do índice de Amihud (2002) quando os dois estão presentes.

Apesar do aparente sucesso dos resultados alcançados pelas pesquisas supracitadas, alguns autores contestam: (i) o fato de a PIN capturar a assimetria informacional; (ii) o papel da PIN nos retornos requeridos dos investidores. A primeira dessas questões foi mais bem

endereçada por Hwang *et al.* (2013), enquanto Mohanram e Rajgopal (2009) tratam do segundo questionamento ao replicarem os estudos de Easley, Hvidkjaer e O'Hara (2002,2005) estendendo-o para períodos posteriores utilizando o modelo de Carhart (1997) e não encontrando evidências de que a PIN seja um risco informacional precificado.

Hwang *et al.* (2013) realizam um estudo único a respeito da PIN. Os autores são os primeiros a utilizarem dados reais do número de ordens de compra e venda de ações em um mercado acionário. Até o momento, devido a não possibilidade de obtenção de tais dados, os pesquisadores, de forma geral, se valiam do método de Lee e Ready (1991) de classificação de ordens de transações. Hwang *et al.* (2013) mostram que esse método classifica erroneamente cerca de 58% das ordens iniciadas por compradores e 62% das ordens iniciadas por vendedores, o que condiz com o apontado por Aslan *et al.* (2011) de que o método de Lee e Ready (1991) não tem bom desempenho em mercados de alta frequência.

Em posse dos dados reais, Hwang *et al.* (2013) calculam a PIN e o retorno esperado representado por quatro estimativas diferentes do *implied cost of equity capital* (ICOE), ou custo de capital próprio implícito (ver Hwang *et al.* (2013), p. 12). Seguindo os passos de Duarte e Young (2009), os autores separaram a PIN em seus efeitos de liquidez e de risco informacional. Hwang *et al.* (2013) chegam a resultados empíricos que sustentam a hipótese de que existe uma relação entre risco informacional e os retornos esperados, assim como reportado por Easley, Hvidkjaer e O'Hara (2002, 2005). Além disso, os autores colocam que a PIN pode ser confiável e precisamente estimada quando de posse dos dados reais do número de ordens de compra e venda (o que não está presente na grande maioria das pesquisas utilizando a PIN). Além disso, a opção de se utilizar o ICOE como *proxy* para retornos *ex-ante* ao invés dos realizados se deve ao fato de que o primeiro captura a relação temporal do *trade-off* risco e retorno mais habilmente do que os retornos realizados tradicionalmente utilizados nos modelos de precificação.

Por fim, o trabalho de maior peso crítico em relação aos resultados da PIN advém de Mohanram e Rajgopal (2009) em que os autores avaliam a PIN em quatro testes. Primeiramente, eles replicam a pesquisa de Easley, Hvidkjaer e O'Hara (2002) estendendo o período em quatro anos – 1984-2002, em contraste com o período de 1984-1998 de Easley, Hvidkjaer e O'Hara (2002). Em seguida, os retornos são regredidos com o fator de PIN e os três fatores de Fama e French (1993, 1996), adicionado ao fator de momento de Carhart

(1997). De posse do fator PIN, os autores seguem para estimar regressões dos retornos com o fator PIN e o crescimento do PIB, tendo em vista que, segundo Mohanram e Rajgopal (2009), essa é uma *proxy* para informações sobre o conjunto de investimentos futuros. Por fim, é calculada a correlação da PIN com o custo de capital *ex-ante* derivado de previsões de analistas. Esperava-se que houvesse uma correlação positiva entre os dois, o que evidenciaria que a PIN é de fato precificada. Os resultados gerais contradizem os trabalhos anteriormente citados. Para Mohanram e Rajgopal (2009), os retornos para o fator PIN são negativamente correlacionados com os retornos de ações com alto PIN. Além disso, o fator PIN não apresentou um coeficiente significativo no teste com os três fatores de Fama e French (1993) e o fator de momento, além de ter apresentado uma associação negativa com o custo de capital *ex-ante*. Por fim, não foi encontrada nenhuma evidência de associação da PIN com o crescimento futuro do PIB. Os autores concluem que a diferença dos retornos atribuída anteriormente ao fator PIN não pode ser confiavelmente vista como compensação pelo risco informacional.

Os resultados apresentados na literatura apontam para uma necessidade ainda mais profunda de se estudar a relação do risco informacional e o retorno requerido. No Brasil, Martins e Paulo (2014) encontraram uma relação positiva da PIN com o custo de capital e retorno das ações. Os autores, no entanto, não se valeram de dados reais da quantidade de ordens de transação.

2.7 Estudos sobre assimetria informacional no mercado brasileiro

Segundo Duarte e Young (2009), mercados emergentes são campos férteis para a prática do *inside trading*. Muitas vezes, nos países latino-americanos, os acionistas minoritários são prejudicados quando atuam em um mesmo ambiente juntamente a grandes grupos controladores que, por meio do poder e do acesso a informações privadas, expropriam os ganhos e afastam pequenos investidores dos mercados acionários (BARBEDO, SILVA e LEAL, 2009).

Dessa forma, o estudo do desequilíbrio informacional no mercado nacional se mostra pertinente. No Brasil, alguns autores se valeram da PIN para a avaliação desse fenômeno (BARBEDO, SILVA e LEAL, 2009; SILVA, 2009; MARTINS e PAULO, 2013, 2014),

enquanto outros optaram por outras *proxies* (IQUIAPAZA, LAMOUNIER, AMARAL, 2007; ALBANEZ e VALLE, 2009; ALBANEZ *et al.*, 2010).

Iquiapaza, Lamounier e Amaral (2007) utilizam como *proxy* para assimetria informacional uma variável *dummy* representando a adesão das empresas aos níveis de governança corporativa da BM&FBOVESPA e procuram estudar a essa relação com o pagamento de dividendos. Albanez e Valle (2009) utilizam uma *proxy* semelhante e procuram analisar a sua relação com a estrutura de capital das empresas. Albanez *et al.* (2010) se valem do erro dos analistas na previsão de ganhos por ação como forma de mensuração da assimetria informacional para estudar as decisões de financiamento e o custo de capital de empresas brasileiras.

Tais *proxies*, como a utilizada em Albanez *et al.* (2010), dependem profundamente de fatores como a qualidade das previsões dos acionistas e não refletem diretamente a presença de assimetria informacional no contexto pesquisado. Dessa forma, a opção pela PIN e, mais recentemente, pelo VPIN são mais indicados pela literatura para o estudo desse fenômeno.

No mercado brasileiro, Barbedo, Silva e Leal (2009) estudaram a relação da PIN com os níveis de governança corporativa da BM&FBOVESPA. O valor médio da probabilidade de negociação privilegiada encontrada foi menor para aquelas empresas alocadas no mercado tradicional, seguida das empresas no nível 1, Novo Mercado e, por fim, nível 2.

Já Martins e Paulo (2013, 2014) aplicam a PIN no mercado brasileiro nos períodos de 2010 e 2011, procurando relacionar o resultado encontrado com os níveis de governança corporativa e as características econômico-financeiras das empresas como: risco, retorno, liquidez, custo de capital e tamanho das firmas dentre outras. Os autores encontram uma probabilidade média de transações privilegiadas de 25% para as empresas no período analisado, valor maior do que o encontrado por Barbedo, Silva e Leal (2009) de 12,5%.

Um problema inerente aos artigos supracitados se refere à classificação das transações entre compra e venda. Os artigos (tanto nacionais, quanto internacionais) se valem do algoritmo de Lee e Ready (1991). No entanto, Silva (2009) afirma que esse método tem uma porcentagem de acerto de 72% no cenário brasileiro, indicando que 28% das ordens são erroneamente classificadas, podendo comprometer o resultado da PIN.

Grammig e Theissen (2002) investigam as consequências da má classificação das ordens de compra e venda na estimação da PIN. Segundo os autores, ao se calcular a

probabilidade de negociação informada utilizando o algoritmo de Lee e Ready (1991), os resultados encontrados são ainda menos confiáveis para ações menos líquidas. Dessa forma, a diferença entre a PIN das ações com grandes disparidades de liquidez é ainda maior, o que está em contradição com os achados de Easley *et al.* (1996). De acordo com Grammig e Theissen (2002), no entanto, quando a classificação entre as ordens pode ser feita de maneira correta, a PIN produz resultados robustos e pode ser usada sem maiores problemas. Tal linha de pensamento pode ser estendida à métrica VPIN.

Na seção seguinte serão expostos os procedimentos metodológicos da pesquisa, desde a coleta de dados, passando pelo seu tratamento, até os modelos usados para responder a pergunta de pesquisa definida anteriormente.

3 METODOLOGIA

3.1 Tipo de pesquisa

O objetivo desta pesquisa foi verificar se o nível de toxicidade das ações, quantificado pelo VPIN, é um fator de risco sistemático precificado no mercado de capitais brasileiro. A fim de se chegar à resposta da questão levantada pelo pesquisador, é necessário estabelecer uma série de procedimentos metodológicos válidos, que vão desde a abordagem do objeto de pesquisa até a coleta e análise de dados. Nas seções seguintes esses aspectos serão descritos e mais bem detalhados.

Cada estudo desenvolvido por um pesquisador tem um objetivo específico. Gil (2008) agrupa os diversos tipos de pesquisa em um conjunto contendo três grandes grupos: as pesquisas exploratórias, as descritivas e as explicativas. O presente trabalho se enquadra no grupo das pesquisas descritivas, pois descreve um fenômeno – o retorno requerido das empresas e sua relação com transações com informações privilegiadas – e aproxima-se das pesquisas explicativas na medida em que tenta apresentar uma explicação sobre o relacionamento entre os retornos requeridos e a assimetria informacional.

Os métodos utilizados neste trabalho são quantitativos. Algumas características desse tipo de pesquisa como: testar e validar teorias sobre como os fenômenos acontecem; testar as hipóteses que são colocadas antes da coleta dos dados e analisar os resultados obtidos possibilitou ao pesquisador responder a pergunta de pesquisa.

Creswell (2007) afirma que a escolha, ao selecionar entre as diferentes técnicas de pesquisa - quantitativa, qualitativa e métodos mistos -, envolve três considerações: o problema de pesquisa, as experiências pessoais do pesquisador e o público-alvo. Analisando tais questões, entende-se que a abordagem quantitativa é mais indicada para o desenvolvimento deste trabalho, pois permite avaliar o efeito de fatores de risco (dentre eles, o proposto neste estudo – assimetria informacional) sobre os retornos requeridos das ações. Além disso, trata-se de um estudo no campo de finanças, em que se vale de métodos econométricos para analisar os relacionamentos entre variáveis de estudos (retorno e fatores de risco).

Para testar o relacionamento entre variáveis de risco e retorno estabelecido pela teoria de precificação de ativos, formularam-se hipóteses que foram submetidas ao teste mediante métodos econométricos. Embora tais métodos não permitam verificar uma relação de causa e efeito, propiciam testar as associações entre fatores de risco e o retorno, apresentadas na teoria e nos estudos empíricos em finanças. Segundo Creswell (2007), a técnica quantitativa emprega alegações pós-positivistas e segue assim ao testar uma teoria especificando hipóteses, coletando dados e utilizando procedimentos estatísticos que corroborarão ou refutarão tais hipóteses.

Quanto aos procedimentos, adotou-se a pesquisa *ex-post-facto* que se caracteriza, segundo Gil (2008), como uma investigação empírica e sistemática na qual o pesquisador não tem controle sobre as variáveis independentes, pois estas já ocorreram, chegando, portanto, às mãos do observador já tendo exercido seus efeitos. Situações em que as variáveis independentes não são manipuláveis requerem pesquisas *ex-post-facto* e se faz necessário considerar outras variáveis, controlando-as estatisticamente. Dessa forma, a sua influência pode ser analisada e neutralizada na análise dos resultados do trabalho (GIL, 2008).

3.2 População e amostra

A população objeto de estudo desta pesquisa foi composta por todas as ações transacionadas na BM&FBovespa. A amostra foi dividida em duas: a primeira, menor, se restringe àquelas ações que apresentaram, no mínimo, uma transação por dia para o período entre 01º de maio de 2014 e 31 de maio de 2016. Tal restrição foi necessária para o cálculo da variável VPIN utilizada para captar a assimetria informacional da ação. Dessa forma, o número de ativos disponíveis para o cálculo dessa variável foi de 142 ações como mostrado na tabela 1.

Tabela 1

Amostra utilizada na pesquisa

Empresa	Código	Tipo	Empresa	Código	Tipo
Banco ABC Brasil	ABCB4	PN N2	Kroton	KROT3	ON NM
Ambev	ABEV3	ON EJ	Lojas Americanas	LAME3	ON
Alpargatas	ALPA4	PN N1	Lojas Americanas	LAME4	PN
Aliansce	ALSC3	ON NM	Mahle Metal Leve	LEVE3	ON NM

Empresa	Código	Tipo	Empresa	Código	Tipo
Anima	ANIM3	ON NM	Light	LIGT3	ON NM
Arezzo	ARZZ3	ON NM	Linx	LINX3	ON NM
Banco do Brasil	BBAS3	ON EJ NM	Le Lis Blanc	LLIS3	ON NM
Bradesco	BBDC3	ON N1	Log-in Logística	LOGN3	ON NM
Bradesco	BBDC4	PN N1	LPS Brasil	LPSB3	ON NM
Brasil Brokers	BBRK3	ON NM	Lojas Renner	LREN3	ON NM
BB Seguridade	BBSE3	ON NM	Magnesita	MAGG3	ON NM
BTG Pactual	BBTG11	UNT	M. Dias Branco	MDIA3	ON NM
Minerva	BEEF3	ON NM	Magazine Luiza	MGLU3	ON NM
Brasil Pharma	BPHA3	ON NM	Mills Estrutura	MILS3	ON NM
Bradespar	BRAP4	PN N1	MMX Mineração	MMXM3	ON NM
BRF	BRFS3	ON NM	Multiplus	MPLU3	ON NM
BR Insurance	BRIN3	ON NM	Marfrig	MRFG3	ON NM
Braskem	BRKM5	PNA N1	MRV Engenharia	MRVE3	ON NM
BR Malls	BRML3	ON NM	Multiplan	MULT3	ON N2
BR Properties	BRPR3	ON NM	Iochpe-Maxion	MYPK3	ON NM
Banrisul	BRSR6	PNB N1	Natura	NATU3	ON NM
B2W Digital	BTOW3	ON NM	Odontoprev	ODPV3	ON NM
BM&F Bovespa	BVMF3	ON EJ NM	Oleo e Gás Part.	OGXP3	ON NM
Card System	CARD3	ON NM	Oi	OIBR3	ON N1
CCR	CCRO3	ON NM	Oi	OIBR4	PN N1
CESP	CESP6	PNB N1	Pão de Açúcar	PCAR4	PN N1
Comgás	CGAS5	PNA	PDG Realty	PDGR3	ON NM
Cielo	CIEL3	ON NM	Petrobras	PETR3	ON
CESC	CLSC4	PN N2	Petrobras	PETR4	PN
CEMIG	CMIG4	PN N1	Profarma	PFRM3	ON NM
CPFL	CPFE3	ON NM	Paranapanema	PMAM3	ON NM
COPEL	CPLE6	PNB N1	Marcopolo	POMO4	PN N2
Cosan	CSAN3	ON NM	Positivo Informática	POSI3	ON NM
Copasa	CSMG3	ON EJ NM	Prumo Logística	PRML3	ON NM
CSN	CSNA3	ON	Porto Seguro	PSSA3	ON NM
Cetip	CTIP3	ON NM	Qgep Participações	QGEP3	ON NM
CVC	CVCB3	ON NM	Qualicorpo	QUAL3	ON NM
Cyrela	CYRE3	ON NM	Raia Drogasil	RADL3	ON NM
Direcional Engenharia	DIRR3	ON NM	Localiza Rent a Car	RENT3	ON NM
Duratex	DTEX3	ON NM	Industrias Romi	ROMI3	ON NM
Ecorodovias	ECOR3	ON NM	Rossi Residencial	RSID3	ON NM
Eletróbrás	ELET3	ON N1	Sabesp	SBSP3	ON NM
Eletróbrás	ELET6	PNB N1	Ser Educacional	SEER3	ON NM
Eletropaulo	ELPL4	PN N2	Springs Global	SGPS3	ON NM
Embraer	EMBR3	ON NM	SLC Agrícola	SLCE3	ON NM
EDP	ENBR3	ON NM	Saraiva	SLED4	PN N2
Equatorial Energia	EQTL3	ON NM	Smiles	SMLE3	ON NM
Estácio Participações	ESTC3	ON NM	São Martino	SMTO3	ON NM
Eternit	ETER3	ON NM	Santos Brasil	STBP11	UNT
Even	EVEN3	ON NM	Sul America	SULA11	UNT N2
Eztec	EZTC3	ON NM	Suzano	SUZB5	PN N1

Empresa	Código	Tipo	Empresa	Código	Tipo
Comp. F Liga Bahia	FESA4	PN N1	Trans. Aliança Ener	TAE11	UNT N2
Fibria Celulose	FIBR3	ON NM	Engie Brasil Energia	TBLE3	ON NM
Fleury	FLRY3	ON NM	Tecnisa	TCSA3	ON NM
Gafisa	GFSA3	ON NM	Technos	TECN3	ON NM
Gerdau	GGBR4	PN ED N1	Tereos Internacional	TERI3	ON
Gerdau Met	GOAU4	PN N1	Tegma G. Logística	TGMA3	ON NM
Gol	GOLL4	PN N2	Tim	TIMP3	ON NM
Grendene	GRND3	ON NM	TOTVS	TOTS3	ON NM
Helbor	HBOR3	ON NM	TPI Participações	TPIS3	ON NM
Cia Hering	HGTX3	ON NM	Cteep	TRPL4	PN N1
Hypermarcas	HYPE3	ON NM	Tupy	TUPY3	ON NM
Iguatemi	IGTA3	ON NM	Unicasa	UCAS3	ON NM
Itaúsa	ITSA4	PN ED N1	Ultrapar	UGPA3	ON NM
Itaú Unibanco	ITUB4	PN ED N1	Usiminas	USIM5	PNA N1
JBS	JBSS3	ON NM	Vale	VALE3	ON N1
JHSF	JHSF3	ON NM	Vale	VALE5	PNA N1
Grupo JSL	JSLG3	ON NM	Telefônica Vivo	VIVT4	PN
Kepler Weber	KEPL3	ON	Valid	VLID3	ON NM
Klabin	KLBN11	UNT N2	Via Varejo	VVAR11	UNT N2
Klabin	KLBN4	PN N2	Weg	WEGE3	ON NM

Fonte - Elaborada pelo autor da dissertação.

Nota - Dados coletados em: <http://www.bmfbovespa.com.br/pt_br/produtos/listados-a-vista-e-derivativos/renda-variavel/empresas-listadas.htm> Acesso em: 29 ago. 2016.

Já a segunda parte da amostra se refere às ações utilizadas para as formações dos fatores de Fama e French (1993, 2015) e Carhart (1997). De forma a ter a maior representatividade do mercado nacional, buscou-se utilizar o maior número possível de ações no cálculo desses fatores. A amostra disponível para tal foi de 349 ações, quantidade esta relativa a todas as ações passíveis de serem pesquisadas por meio do banco de dados da Bloomberg ® cujo acesso foi disponibilizado pelo Grupo de Pesquisas em Finanças Corporativas e de Mercado (GFIN) do CEFET-MG . A tabela das ações utilizadas se encontra no apêndice.

O período analisado neste trabalho foi definido em função dos dados necessários ao cálculo da variável VPIN disponibilizados pela BM&FBovespa. Esta entidade disponibiliza, em seu sítio, na internet, somente dois anos de informações. Sendo assim, o período analisado teve que se restringe para aquele previamente citado.

3.3 Coleta e tratamento dos dados

A principal limitação dos trabalhos que se propuseram a aplicar a PIN ou, neste caso, o VPIN, se deve à classificação errônea das ordens de compra e venda. De forma a contornar tal problema, esta pesquisa se vale dos dados reais do volume transacionado no mercado brasileiro. A base de dados utilizada foi o diretório *market data* disponibilizado pela BM&FBovespa, a qual contém as ordens emitidas de todas as ações transacionadas no mercado acionário nacional nos últimos dois anos. Não há, até agora, registrada na literatura do País, pesquisa que utiliza essa mesma base de dados para o cálculo do VPIN, o que evidencia a originalidade deste trabalho na área de microestrutura de mercado em finanças.

O principal motivo de esse diretório não ter sido utilizado até o momento pode residir na dificuldade do tratamento dos dados. Estes são disponibilizados via arquivos de texto contendo inúmeras outras informações além do horário em que a ordem de transação foi emitida. Assim, lidar com esses dados requer um árduo e longo trabalho, além da necessidade de maior poder computacional para a separação e filtragem das informações relevantes para a aplicação do modelo.

Os dados coletados por meio do *market data* da BM&FBovespa foram tratados exclusivamente em rotinas desenvolvidas no *software* estatístico R com auxílio dos principais pacotes de manipulação de dados disponibilizados para esse programa. As diversas outras informações, como excesso de retorno da ação, tamanho, *book-to-market*, lucratividade, ativo, passivo não circulante e demais dados necessários para a aplicação dos modelos foram coletados das bases de dados Bloomberg® e Quantum Axis®.

3.4 Hipóteses de pesquisa

Para atingir os objetivos desta pesquisa, algumas hipóteses foram testadas em relação às variáveis estudadas, são elas:

Hipótese 1 - Quanto menor o tamanho da empresa, maior a probabilidade de negociação com informação privilegiada.

Essa hipótese procurou averiguar se empresas maiores apresentam menores níveis de probabilidade de negociações privilegiadas como reportado por Easley *et al.* (1996) e diversos outros trabalhos empíricos que utilizaram da PIN ou VPIN (Mohanram & Rajgopal, 2009; Abad & Yagüe, 2012; Wei, Gerace & Frino, 2013).. Segundo Easley *et al.* (1996), espera-se que ações de grandes empresas tenham maior cobertura de analistas, ou seja, um número maior de profissionais produzindo pareceres que se referem a esse tipo de empresa e também maior atenção dos investidores. Dessa forma, a chance de haver transações privilegiadas é em tese menor para essas ações, resultando em um VPIN mais baixo do que o encontrado para ações de companhias menores.

Hipótese 2 - Os segmentos da Bovespa apresentam VPINs diferenciados

Esperava-se que as empresas que compõem os segmentos diferenciados da Bovespa apresentassem menor VPIN, assim como verificado por Barbedo *et al.* (2009) e Martins e Paulo (2013). Dessa forma, estipula-se a hipótese de que o VPIN para segmento NM é o mais baixo, seguido do N2, N1 e, por fim, o tradicional.

Hipótese 3 - Quanto maior a probabilidade de negociação com informações privilegiadas, menor a liquidez das ações (maior o *bid-ask spread*).

Apesar de teoricamente não ser uma *proxy* para liquidez, o VPIN está intrinsecamente ligado a esse conceito. Quando há sinais de investidores informados atuando no mercado, Easley *et al.* (1996) mostram que o *bid-ask spread* é diretamente afetado. Quanto maior o indício de negociações com informações privilegiadas, maior é o *spread*. Dessa forma, espera-se que essa *proxy* para a toxicidade dos fluxos tenha um poder de explicação sobre a liquidez das ações.

Foi utilizado como *proxy* para liquidez o *spread* diário entre os preços de *bid* e *ask* das ações, e a seguinte regressão foi testada (32).

$$BidAskSpread_{it} = \alpha_i + \beta_i VPIN_{it} \quad (32)$$

Quanto maior o *spread* entre o *bid* e o *ask* de uma ação, menor a sua liquidez. Assim, espera-se que o coeficiente dessa regressão seja positivo, indicando que o VPIN tem um poder estatisticamente significativo de prever a liquidez de uma ação para o período posterior àquele ao qual o VPIN foi calculado.

Tal hipótese também foi testada a partir do cálculo da correlação entre o VPIN e o *bid-ask spread* das ações. Espera-se que a correlação entre ambas as variáveis seja positiva e estatisticamente significativa.

Hipótese 4 - Um fator relacionado ao VPIN ajuda a explicar os retornos de portfólios

De acordo com a teoria de eficiência de mercado de Fama (1970, 1991), o risco informacional seria diversificável e, por consequência, não precificado pelos investidores. Hughes, Liu e Liu (2007) e Lambert, Leuz e Verrecchia (2011) afirmam que o risco associado ao PIN encontrado por Easley, Hvidkjaer e O'Hara (2002) é diversificável ou incorporado aos fatores de risco já existentes. Contrariando essa argumentação, Easley, Hvidkjaer e O'Hara (2005) se valeram dos modelos de fatores de Fama e French (1993) e Carhart (1997) para verificar se a probabilidade de negociação com informação privilegiada constitui um fator adicional de risco de mercado, tendo os seus resultados corroborado essa hipótese.

Easley, Hvidkjaer e O'Hara (2005) dizem que um fator baseado nos valores do PIN das ações ajudam a explicar os retornos de portfólios criados com base nas variáveis tamanho e PIN dos ativos, tanto quando aplicado ao modelo de três fatores de Fama e French (1993), quando aplicado ao modelo de quatro fatores de Carhart (1997). Mohanram e Rajgopal (2009) replicam esse estudo estendendo em quatro anos os dados investigados. Esses autores constatam que o fator baseado no PIN é precificado somente em uma parte do período analisado, não sendo possível, portanto, afirmar que o risco informacional seja precificado.

De forma a apresentar uma contribuição aos estudos de microestrutura de mercado e precificação de ativos, conjectura-se a hipótese de que um fator criado com base no VPIN seja precificado por meio de testes em diferentes conjuntos de carteiras. Esperava-se que a adição de um fator VPIN nos modelos de três e cinco fatores de Fama e French (1993, 2015) e de quatro fatores de Carhart (1997) apresente uma redução no intercepto geral das carteiras analisadas pelo teste GRS – Gibbons, Ross e Shanken (1989), aplicado por Fama e French (1996, 2015) e pelo *Average F-test* de Hwang e Satchell (2014) para os dados simulados.

3.5 Métodos de análise dos dados

Nesta parte, descrevem-se as etapas percorridas para obter produzir uma resposta à pergunta de pesquisa. Buscou-se preencher algumas das lacunas apresentadas na literatura de microestrutura em finanças, como a verificação do desempenho dos métodos de classificação de transações e da existência ou não da precificação do desequilíbrio informacional por parte dos modelos de fatores. Os próximos capítulos detalharão quais modelos e etapas foram adotados para atingir os objetivos desta pesquisa.

3.5.1 Aplicação dos métodos *Tick-Rule* e *Bulk Volume Classification*.

De posse dos dados reais coletados no *market data* da BM&FBovespa, o primeiro passo do trabalho foi na direção de contribuir para a literatura que trata dos algoritmos de classificação de transações. Buscou-se comparar a precisão de dois métodos distintos: o *Tick-Rule* (TR) e o mais recente *Bulk Volume Classification* (BVC). O TR, utilizado mais frequentemente nos estudos empíricos, propõe uma classificação *tick-by-tick*, comparando a transação atual com a sua precedente, classificando-a como compra, caso o preço tenha sido superior, e venda em caso contrário. Em situações em que o preço se mantém, repete-se a classificação da transação anterior. O TR foi testado empiricamente diversas vezes, tendo se mostrado eficaz na maioria dos mercados, com uma precisão variando de 67% (OMRANE e WELCH, 2016) até 86% (EASLEY, LÓPEZ DE PRADO e O'HARA, 2012b) no mercado norte-americano.

Já o BVC, cunhado por Easley, López de Prado e O'Hara (2011a) – e mais detalhado em Easley, López de Prado e O'Hara (2012b, 2016) – propõe uma nova forma de classificação de transações em que se agrega o volume transacionado e se classifica-o como parte compra e parte venda, dependendo da variação do preço no período em que tais ações foram negociadas. Easley, López de Prado e O'Hara (2012b) argumentam que esse método tem uma precisão de cerca de 92% a depender dos parâmetros utilizados, sendo este superior ao TR. Chakrabarty, Pascual e Shkilko (2015) e Omrane e Welch (2016), analisando outros mercados, atestam a superioridade dos métodos *tick-by-tick* em relação ao BVC.

Foram, portanto, aplicados os métodos TR e BVC para a classificação das transações no cenário nacional. A apuração da eficácia do TR é direta. Como esse método classifica cada transação, basta apurar quantas delas foram identificadas corretamente, como mostra (33).

$$Ar_{TR} = \frac{Q_C}{Q_T} \quad (33)$$

em que Ar_{TR} é a taxa de precisão – *Accuracy rate* – do método TR na previsão dos sinais das transações; Q_C se refere à quantidade de transações identificadas de forma correta; e Q_T representa a quantidade total de transações no dia.

Para a apuração da precisão do BVC, por se tratar de um método que agrega o volume transacionado em um tempo específico, utilizou-se o índice de precisão cunhado por Easley, López de Prado e O'Hara (2012b), definido por (34).

$$Ar_{BVC} = \frac{\sum_{\tau} [\min(\hat{V}_{\tau}^B, V_{\tau}^B) + \min(\hat{V}_{\tau}^S, V_{\tau}^S)]}{\sum_{\tau} V_{\tau}} \quad (34)$$

em que (V_{τ}^B, V_{τ}^S) são os volumes reais de compra e venda; $(\hat{V}_{\tau}^B, \hat{V}_{\tau}^S)$ são os volumes estimados para a barra de volume ou tempo τ ; e V_{τ} representa o volume total transacionado em tal barra.

Buscou-se, primeiramente, trabalhar com o máximo de ações possíveis no mercado nacional, de forma a testar a eficácia dos métodos TR e BVC de forma mais abrangente. Posteriormente, os modelos foram testados na amostra específica para o cálculo do VPIN. Os resultados apresentados para ambos os casos são apresentados na seção de resultados.

3.5.2 Modelos de precificação de ativos

Este trabalho se inspirou nos passos e nas técnicas utilizados principalmente nos trabalhos de Easley, Hvidkjaer e O'Hara (2005) e Mohanram e Rajgopal (2009) de forma a trabalhar com uma gama de testes e análises estatísticas que se provaram robustas no mercado norte-americano, especialmente no que se refere ao cálculo dos retornos requeridos, testes de correlações e análise dos modelos. Além disso, esses artigos utilizam modelos de microestrutura, que se diferenciam das técnicas tradicionais na medida em que eles podem ser vistos como sistemas de aprendizado em que o *market maker* observa os dados do mercado e tira conclusões acerca do valor real de um ativo (EASLEY, HVIDKJAER e O'HARA, 2005).

3.5.3 Os modelos de fatores e o VPIN

Easley, Hvidkjaer e O'Hara (2005) e Mohanram e Rajgopal (2009) procuraram testar a hipótese de que a PIN é precificada. Enquanto o primeiro encontrou resultados que sustentam a hipótese de uma possível precificação do risco informacional, o segundo trabalho, ampliando o período analisado, achou resultados contrários àqueles de Easley, Hvidkjaer e O'Hara (2005). Dessa forma, procurou-se se basear nos testes e nas regressões aplicados por esses trabalhos, de forma a verificar como eles se comportam no mercado brasileiro. A principal ideia se sustenta na criação de um fator baseado no VPIN de cada ação. Easley, Hvidkjaer e O'Hara (2005) afirmam que, para que a probabilidade de negociação privilegiada possa ser utilizada como um fator, deve existir uma variação comum entre os retornos de ações com alta e baixa probabilidades de *inside trading*.

O primeiro sinal dado por Easley, Hvidkjaer e O'Hara (2005) para a necessidade de se criar um fator PIN reside no retorno anormal de um portfólio composto, em sua maioria, por ações com alta probabilidade de negociação privilegiada não ser explicada pelos modelos de três fatores de Fama e French (1993) e o fator de momento de Carhart (1997). A partir dessa constatação, os autores encontraram uma correlação significativa entre portfólios construídos com base nos valores da PIN e tamanho das empresas, o que sugere que a *proxy* para a probabilidade de negociações privadas poderia ser um fator na precificação de ativos.

Criando um fator PIN, os autores conseguem reduzir os interceptos apresentados pelos modelos previamente utilizados (três e quatro fatores) para explicar os retornos de oito dos dez portfólios construídos com base nas variáveis tamanho e PIN. Interpretou-se esse resultado como justificativa suficiente para proceder-se à criação de um fator neste estudo baseado no VPIN, método mais recente e robusto para estimação da probabilidade de negociações privilegiadas.

Em sintonia com os trabalhos de Easley, Hvidkjaer e O'Hara (2005), Mohanram e Rajgopal (2009) e Fama e French (1993, 1996, 2015), os fatores foram regredidos com os retornos de portfólios e, não, ações individuais. Seguiram-se os passos de Fama e French (2015) para a criação desses portfólios, baseada na variável principal tamanho e permutando o segundo componente entre as demais variáveis: *book-to-market*, lucratividade, investimento e

VPIN. A justificativa para a formação das carteiras baseadas no tamanho e VPIN reside na evidência empírica de que a probabilidade de negociação privilegiada de uma ação apresenta uma forte correlação negativa com o seu tamanho. Dessa forma, esse procedimento busca isolar os efeitos do VPIN dentro das carteiras.

Por fim, apresentam-se as regressões utilizadas nesta pesquisa. Primeiramente, regrediu-se o retorno dos portfólios com base no modelo de três fatores de Fama e French (1993), como mostrado em (35).

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + \varepsilon_i \quad (35)$$

Em seguida, acrescentou-se o fator IMU ao modelo exposto acima, resultando em (36).

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + p_iIMU + \varepsilon_i \quad (36)$$

Posteriormente, foi utilizado o modelo de quatro fatores de Carhart (1997), apresentado em (37).

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + u_iUMD + \varepsilon_i \quad (37)$$

Acrescentou-se o fator IMU a esse modelo, resultando em (38)

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + u_iUMD + p_iIMU + \varepsilon_i \quad (38)$$

Por fim, calculou-se o modelo de cinco fatores de Fama e French (2015), mostrado pela equação (39).

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + r_iRMW + c_iCMA + \varepsilon_i \quad (39)$$

O último passo foi a adição do fator IMU de forma a verificar o seu efeito no modelo exposto em (40).

$$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + r_iRMW + c_iCMA + p_iIMU + \varepsilon_i \quad (40)$$

Tabela 2
Modelos testados e suas respectivas bases teóricas

Modelo	Especificação do modelo	Base empírica
Três Fatores	$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + \varepsilon_i$	Fama e French (1993)
Três Fatores + IMU	$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + p_iIMU_i + \varepsilon_i$	Fama e French (1993)
Quatro Fatores	$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + u_iUMD + \varepsilon_i$	Carhart (1997)
Quatro Fatores + IMU	$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + u_iUMD + p_iIMU_i + \varepsilon_i$	Carhart (1997)
Cinco Fatores	$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + r_iRMW + c_iCMA + \varepsilon_i$	Fama e French (2015)
Cinco Fatores + IMU	$R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + r_iRMW + c_iCMA + p_iIMU_i + \varepsilon_i$	Fama e French (2015)

Fonte - Elaborada pelo autor da dissertação.

3.5.3.1 Variáveis dependentes

A criação dos portfólios utilizados nos testes dos modelos de precificação seguiu duas etapas. Procurou-se, primeiramente, seguir os passos de Fama e French (2015), que estendem o trabalho de Fama e French (1993) em relação à construção de portfólios. Os autores, em seu trabalho mais recente, testaram os modelos propostos em quatro conjuntos de portfólios. O primeiro deles é composto por 25 carteiras formadas a partir das variáveis tamanho e *book-to-market*. Para cada ano, dividem-se as ações em cinco quintis baseadas em seus tamanhos. O procedimento é repetido para a variável *book-to-market*. A interseção entre os portfólios (quintis) constituídos a partir da variável tamanho - ações menores - com cada um dos portfólios (quintis) formados de acordo com uma segunda variável produz cinco portfólios. Repetindo-se o procedimento para cada divisão das ações resulta-se em 25 carteiras diferentes.

Fama e French (2015) repetem o procedimento para as combinações das variáveis tamanho e lucratividade; e tamanho e investimento. De forma a expandir a forma de construção dessas carteiras, os autores procuram formar portfólios com a interseção entre três variáveis ao mesmo tempo, montando 32 carteiras baseadas nas variáveis tamanho, lucratividade e investimento.

De forma a definir quantos portfólios seriam criados para cada combinação de variáveis, foram seguidos os resultados encontrados por Ceretta e Costa Jr (2000) e Sanvicente e Bellato (2004). Tais trabalhos demonstram que uma carteira devidamente diversificada no mercado brasileiro tem entre seis e dezoito ações. Dessa forma, devido à impossibilidade de se formar carteiras com o mesmo número de ativos dos trabalhos de Fama e French (2015), prosseguiu-se à montagem de portfólios que tivessem, em média, um mínimo de seis ações. A partir disso, foi possível estabelecer qual seria a relação entre as variáveis analisadas.

Para a variável tamanho e *book-to-market*, foi possível dividir as ações pela primeira variável em três grupos e, pela segunda, em quatro. Tal procedimento resultou em doze carteiras, com o mínimo de seis ações e máximo de vinte e cinco. A tentativa de aumentar a quantidade de carteiras resultou em portfólios com menos de três ações, o que violaria a restrição de mínimo de seis ações por carteira.

Para o tamanho e a lucratividade, foi possível dividir as ações em quatro grupos para ambas as variáveis. Assim, obtiveram-se dezesseis carteiras, sendo a menor delas com seis ações e a maior delas com quinze.

Para o tamanho e investimento, dividiram-se as ações pela primeira variável em grupos de três e, pelo investimento, em grupos de quatro. Tal procedimento resultou em doze carteiras, com um mínimo de sete ações e um máximo de vinte e três.

Finalmente, para o tamanho, investimento e lucratividade, foi possível a divisão em dois grupos para o tamanho, três para o investimento e novamente dois para a lucratividade. Sendo assim, foram criadas doze carteiras com mínimo de sete ações e máximo de dezoito.

Como neste trabalho utiliza-se também a variável VPIN para a formação dos fatores, incluiu-se um conjunto de carteiras formadas pelo tamanho e pelo VPIN das ações. Foi possível dividir as ações pela primeira variável em três grupos e, pela segunda, em quatro. Das doze carteiras formadas apenas a primeira e a última tinham menos de seis ações. Assim, excluíram-se tais portfólios, resultando em dez carteiras no total, em que a menor delas contém seis ações e a maior, vinte e nove.

Tabela 3
Informações acerca dos portfólios criados

Variáveis	Quantidade de portfólios	Mínimo de ações	Máximo de ações	Média de ações
Tamanho (3) / <i>Book-To-Market</i> (4)	12	6	25	14
Tamanho (3) / Investimento (4)	12	7	23	14
Tamanho (2) / Inv. (3) / Lucratividade (2)	12	7	18	13
Tamanho (3) / VPIN (4)	10	6	29	11
Tamanho (4) / Lucratividade (4)	16	6	15	10
Total	62	6	29	12

Fonte - Elaborada pelo autor da dissertação.

3.5.3.2 Variáveis independentes

Para a criação dos fatores utilizados neste trabalho, seguiram-se os procedimentos realizados por Fama e French (1993, 2015) e Carhart (1997) utilizando, no total, 349 ações negociadas na BM&FBovespa. Os dados para a criação dos fatores foram coletados na base de dados Quantum Axis ®. A base para a criação dos fatores é a mesma para os modelos de três e quatro fatores. Esta consiste na classificação das ações baseadas em: tamanho, *book-to-market* e, para o fator de Carhart (1997), desempenho no ano anterior. Para a criação do fator IMU, a amostra foi reduzida àquelas ações com negociações em todos os dias durante o período analisado, resultando em 142 ações passíveis de comporem o fator VPIN.

Primeiramente, em junho de cada ano, os ativos foram classificados em ordem de tamanho (capitalização de mercado, que é igual ao preço vezes a quantidade de ações em circulação). Após tal classificação, os ativos foram divididos pela mediana em dois grupos, formando o grupo grande (*big*) e o grupo pequeno (*small*).

O segundo passo consistiu em classificar as ações baseado em seus valores *book-to-market* (valor patrimonial da ação em relação a seu valor de mercado, ou *BE/ME*). Após tal classificação, dividiram-se as ações em três grupos, os 30% menores consistem das ações com baixo *BE/ME* (*low*), os próximos 40% são as ações com médio *BE/ME* (*medium*), e, por fim, as últimas 30% são aquelas com alto *BE/ME* (*high*). Segundo Fama e French (1993), a divisão em três grupos para a variável *BE/ME* foi realizada porque Fama e French (1992)

atestaram que tal variável tem um papel mais importante do que a variável tamanho (*size*) ao explicar os retornos médios analisados.

O terceiro passo se deu pela divisão das ações em três grupos, os 30% menores consistem das ações com menor retorno nos últimos onze meses. Os próximos 40% são as ações com retornos medianos, finalmente, as últimas 30% são aquelas ações que apresentaram melhores retornos nos últimos onze meses. Para a criação do fator utilizam-se somente o primeiro e o terceiro grupos criados, ou seja, ignoram-se as ações com retornos medianos. Tal procedimento foi feito baseado no trabalho de Carhart (1997) a fim de se criar o fator de momento.

Para a aplicação do modelo de cinco fatores, seguiu-se o procedimento sugerido em Fama e French (2015) de classificar as ações em dois grupos por meio da mediana utilizando as variáveis: tamanho, *book-to-market*, lucratividade e investimento. Fama e French (2015) asseveram que a classificação pode ser feita de diversas maneiras. No entanto, os autores apontam que essa forma resulta em portfólios mais bem diversificados e, portanto, apresenta melhores resultados do que as demais opções. Os fatores são criados para esse modelo dando atenção especial à variável tamanho, ou seja, eles são obtidos via carteiras de ações em que o tamanho está presente, e a segunda variável muda ao longo dos portfólios.

Para a obtenção dos cinco fatores de Fama e French (2015), são criados dois grupos para cada variável. Para a variável *size*, as ações são divididas pela sua mediana entre o grupo *small* e o *big* – este procedimento é análogo ao apresentado anteriormente. Em relação ao índice *book-to-market*, as ações são divididas entre *high* e *low book-to-market*, resultando novamente em dois grupos.

Para a alocação das ações em portfólios com base em sua lucratividade, foi necessário, primeiramente, calcular tal variável. Esta foi obtida dividindo-se a receita líquida do ano $t - 1$ pelo valor contábil do patrimônio líquido no ano $t - 1$. A receita líquida foi obtida subtraindo-se da receita bruta o custo das mercadorias vendidas, as despesas gerais e administrativas e, por fim, os juros. Após esse procedimento, dividiram-se as ações em dois grupos, chamados de robustos (*robust*) e fracos (*weak*).

Por fim, calculou-se a variável investimento para as ações analisadas. Tal procedimento foi feito por meio do cálculo do crescimento do ativo total de um ano para o outro, ou seja, dividiu-se o ativo total do ano $t - 1$ pelo total de ativos da empresa no ano

$t - 2$. A amostra foi então ordenada em função dessa variável e separada em dois grupos: conservadores (*conservatives*) e agressivos (*aggressives*).

Após tais classificações, procedeu-se à criação dos fatores, o que se deu por meio de três diferentes metodologias, tendo em vista que foram três os modelos analisados – três e cinco fatores de Fama e French (1993, 2015) e quatro fatores de Carhart (1997). As formas de cálculo dos fatores são explicadas a seguir.

3.5.3.3 Três fatores de Fama e French (1993)

A primeira regressão testada foi o modelo de três fatores de Fama e French (1993). Dessa forma, o procedimento é análogo àquele empregado pelos autores, ou seja, são utilizadas as classificações a partir das variáveis tamanho (*size*) e *book-to-market*. Para a construção dos fatores são construídas seis carteiras, como mostrado pela tabela 4.

Tabela 4

Carteiras formadas a partir dos fatores tamanho e *book-to-market*

Carteira	Sigla	Descrição
<i>Small e Low</i>	SL	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e as ações com baixo valor de <i>book-to-market</i>
<i>Small e Medium</i>	SM	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e as ações com valor intermediário de <i>book-to-market</i>
<i>Small e High</i>	SH	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e as ações com alto valor de <i>book-to-market</i>
<i>Big e Low</i>	BL	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e as ações com baixo valor de <i>book-to-market</i>
<i>Big e Medium</i>	BM	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e as ações com valor intermediário de <i>book-to-market</i>
<i>Big e High</i>	BH	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e as ações com alto valor de <i>book-to-market</i>

Fonte - Elaborada pelo autor da dissertação.

O primeiro fator, *SMB* (*small minus big*) é obtido a partir da diferença entre os retornos das carteiras *small* ($R_{S,t}$) e *big* ($R_{B,t}$), ou seja,

$$SMB_t = R_{S,t} - R_{B,t} \quad (41)$$

Tais retornos foram obtidos a partir da média aritmética dos retornos das carteiras que os representam, como mostrado em (42) e (43).

$$R_{S,t} = \frac{SL_t + SM_t + SH_t}{3} \quad (42)$$

$$R_{B,t} = \frac{BL_t + BM_t + BH_t}{3} \quad (43)$$

O segundo fator, *HML* é obtido a partir da diferença entre os retornos das carteiras *high* ($R_{H,t}$) e *low* ($R_{L,t}$), ou seja,

$$HML_t = R_{H,t} - R_{L,t} \quad (44)$$

Em que $R_{H,t}$ e $R_{L,t}$ são obtidos pela média aritmética dos retornos das carteiras que têm ações com alto índice *book-to-market* e baixo índice *book-to-market*, respectivamente, como mostrado em (45) e (46).

$$R_{H,t} = \frac{SH_t + BH_t}{2} \quad (45)$$

$$R_{L,t} = \frac{SL_t + BL_t}{2} \quad (46)$$

O fator de mercado, assim como para a grande parte dos estudos envolvendo modelos de precificação de mercados, foi definido como o retorno do índice de mercado – neste trabalho utilizou-se o Ibovespa – subtraído pelo retorno do ativo livre de risco, definido com o CDI diário, como mostrado pela equação (47).

$$MKT = R_{Ibovespa,t} - R_{CDI,t} \quad (47)$$

3.5.3.4 Quatro fatores de Carhart (1997)

Para a aplicação das equações (37) e (38), empregou-se a metodologia adotada por Carhart (1997) em que um fator de momento é adicionado ao modelo de três fatores de Fama e French (1993). Dessa forma, utilizaram-se os fatores calculados para o modelo de três fatores com a adição do quarto fator de Carhart (1997), calculado pela subtração de carteiras compostas por ações vencedoras pelas carteiras compostas por ações perdedoras. O procedimento, portanto, envolveu a divisão das ações entre o grupo *small* e *big* e, posteriormente, procedeu-se à obtenção da interseção entre esses grupos com os 30% de ações

com melhores retornos no último ano e os 30% de ações com piores retornos neste período. As carteiras são formadas como exposto pela tabela 5.

Tabela 5
Carteiras formadas a partir dos fatores tamanho e momento

Carteira	Sigla	Descrição
<i>Small e Winner</i>	SW	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e maior retorno passado
<i>Big e Winner</i>	BW	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e maior retorno passado
<i>Small e Loser</i>	SL	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e menor retorno passado
<i>Big e Loser</i>	BL	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e menor retorno passado

Fonte - Elaborada pelo autor da dissertação.

O processo para o cálculo dos fatores no modelo de Carhart (1997) é similar ao que foi apresentado na seção anterior. Para a obtenção do fator UMD, subtrai-se o retorno diário das ações nos portfólios *winner* dos retornos diários das ações dos portfólios *loser*, como mostram as equações (48), (49) e (50).

$$UMD_t = R_{w,t} - R_{l,t} \quad (48)$$

$$R_{w,t} = \frac{SW+BW}{2} \quad (49)$$

$$R_{l,t} = \frac{SL+BL}{2} \quad (50)$$

O fator de mercado é o mesmo do modelo de três fatores.

3.5.3.5 Cinco fatores de Fama e French (2015)

O modelo de cinco fatores de Fama e French (2015) apresenta procedimentos metodológicos similares aos precedentes. A principal diferença é a forma como as carteiras são formadas. Como apresentado anteriormente, os portfólios para esse modelo são criados a partir da variável tamanho como sendo constante entre as carteiras, mudando o segundo componente, como apresentado na tabela 6.

Tabela 6

Carteiras formadas a partir dos fatores tamanho, *book-to-market*, lucratividade e investimento

Carteira	Sigla	Descrição
<i>Small e Low</i>	SL	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e baixo valor de <i>book-to-market</i>
<i>Small e High</i>	SH	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e alto valor de <i>book-to-market</i>
<i>Small e Robust</i>	SR	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e alta lucratividade
<i>Small e Weak</i>	SW	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e baixa lucratividade
<i>Small e Conservative</i>	SC	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e alto investimento
<i>Small e Aggressive</i>	SA	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e baixo investimento
<i>Big e Low</i>	BL	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e baixo valor de <i>book-to-market</i>
<i>Big e High</i>	BH	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e alto valor de <i>book-to-market</i>
<i>Big e Robust</i>	BR	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e alta lucratividade
<i>Big e Weak</i>	BW	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e baixa lucratividade
<i>Big e Conservative</i>	BC	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e alto investimento
<i>Big e Aggressive</i>	BA	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e baixo investimento

Fonte - Elaborada pelo autor da dissertação.

A criação do fator *SMB* se dá diretamente por meio da diferença entre os retornos diários das carteiras com ações *small* e ações *big*, ou seja,

$$SMB = \frac{(SH+SL+SR+SW+SC+SA)}{6} - \frac{(BH+BL+BR+BW+BC+CA)}{6} \quad (51)$$

Já os fatores *HML*, *RMW* e *CMA* são obtidos via equações (52), (53) e (54)

$$HML = \frac{(SH+BH)}{2} - \frac{(SL+BL)}{2} \quad (52)$$

$$RMW = \frac{(SR+BR)}{2} - \frac{(SW+BW)}{2} \quad (53)$$

$$CMA = \frac{(SC+BC)}{2} - \frac{(SA-BA)}{2} \quad (54)$$

O fator de mercado permanece o mesmo dos modelos anteriores.

Como apontado por Fama e French (2015), existem outras formas de se calcular esses fatores. No entanto, o método apresentado acima resulta em portfólios bem diversificados e que apresentaram melhores resultados nos testes conduzidos pelos autores. Dessa forma, optou-se por seguir os passos de Fama e French (2015) para a continuação deste trabalho.

3.5.3.6 Fator VPIN (IMU)

Para a criação de um fator baseado no VPIN das ações, seguiram-se os passos de Mohanram e Rajgopal (2009), mudando, basicamente, o número de portfólios criados – tal mudança foi necessária devido à diferença na quantidade de ativos nas amostras dos trabalhos. Primeiramente, as ações foram divididas em três grupos com base em seus tamanhos. Posteriormente, dividiram-se as ações em dois grupos: baixo e alto VPIN. Calculou-se o retorno ponderado de cada grupo para, então, determinar a diferença entre os retornos das ações com valores maiores e valores menores para a variável VPIN. A tabela 7 resume a forma de criação dos portfólios.

Tabela 7

Carteiras formadas com base nas variáveis tamanho e VPIN

Carteira	Sigla	Descrição
<i>Small e Low</i>	SL	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e as ações com baixo valor de VPIN
<i>Small e High</i>	SH	Interseção entre as ações do grupo <i>small</i> para a variável tamanho e as ações com alto valor de VPIN
<i>Medium e Low</i>	ML	Interseção entre as ações do grupo <i>medium</i> para a variável tamanho e as ações com baixo valor de VPIN
<i>Medium e High</i>	MH	Interseção entre as ações do grupo <i>medium</i> para a variável tamanho e as ações com alto valor de VPIN
<i>Big e Low</i>	BL	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e as ações com baixo valor de VPIN
<i>Big e High</i>	BH	Interseção entre as ações do grupo <i>big</i> para a variável tamanho e as ações com alto valor de VPIN

Fonte - Elaborada pelo autor da dissertação.

Dessa forma, o fator IMU foi obtido como mostrado em (55)

$$IMU = \frac{SH+MH+BH}{3} - \frac{SL+ML+BL}{3} \quad (55)$$

Os motivos que deram suporte à criação do IMU na literatura repousam na relação entre a probabilidade de negociações privilegiadas e o retorno das ações. Easley, Hvidkjaer e O'Hara (2002) encontram uma correlação positiva entre essas duas variáveis. Segundo os autores, ações com maior PIN apresentam um retorno requerido maior e, conseqüentemente, maior custo de capital. Além disso, os autores constatam que existe uma correlação negativa entre o tamanho e o PIN das empresas, indicando que companhias maiores têm um menor nível de probabilidade de negociações privilegiadas.

Neste trabalho, foi observado que a relação entre o VPIN e o tamanho das empresas apresenta uma correlação negativa de -0,3080 com p-valor de 0,000. Este resultado condiz com os trabalhos empíricos que mostram que ações de empresas maiores apresentam menor probabilidade de negociação privilegiada. Já, para a correlação entre o VPIN e o retorno, foi encontrado um valor de 0,01409 com p-valor de 0,0003. Apesar de baixo, condiz com a literatura de que ações com maior VPIN apresentam, na média, maiores retornos. Em relação à correlação entre o tamanho das empresas e seus retornos, o valor encontrado foi de -0,0105 com p-valor 0,0068. Novamente, apesar de ser um valor próximo de zero, este indica um maior retorno para empresas menores. Esperava-se que os valores para as correlações entre o VPIN e os retornos e os retornos e o tamanho fossem mais expressivos, ou seja, que ações com maior VPIN e ações de empresas menores apresentassem consideravelmente maiores retornos.

Apesar disso, supõe-se possível a existência de um prêmio para o investimento em ações com maior toxicidade dos fluxos das ordens. Dessa forma, buscou-se testar a pertinência de se introduzir um fator baseado no VPIN, o qual representaria o prêmio pelo risco de se negociar com informações privilegiadas na explicação dos retornos dos portfólios examinados no contexto dos modelos de fatores.

3.5.4 Testes realizados nos modelos de precificação

(a) Teste de Gibbons, Ross e Shanken (1989)

O teste usado para verificar conjuntamente se as carteiras apresentam intercepto estatisticamente igual a zero foi o teste de Gibbons, Ross e Shanken (1989), ou teste GRS, aplicado por Fama e French (1996, 2015) para averiguar a eficiência dos modelos propostos.

O teste GRS é de grande importância por dois motivos: este é sensível, tanto à escolha do portfólio, quanto ao número de fatores testados. Dessa forma, assim como em Fama e French (2015), o GRS foi aplicado a todos os conjuntos de carteiras montados neste trabalho.

O GRS se apresenta como um teste F com N e $T - N - K$ graus de liberdade, em que

$N =$ número de portfólios analisados

$T =$ número de observações (dias)

$K =$ quantidade de fatores na regressão

A partir desses valores, calcula-se o GRS, como mostrado em (56)

$$\left(\frac{T}{N}\right) \left(\frac{T-N-K}{T-N-1}\right) \left[\frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + \bar{\mu}' \hat{\Omega}^{-1} \bar{\mu}}\right] \sim F(N, T - N - K) \quad (56)$$

Em que $\hat{\alpha}$, $\hat{\Sigma}^{-1}$, $\bar{\mu}$ e $\hat{\Omega}^{-1}$ são respectivamente: um vetor com os interceptos estimados das regressões, o inverso da matriz de covariância dos resíduos, a média dos fatores e o inverso da matriz de covariância dos fatores.

O teste, portanto, testa as seguintes hipóteses:

H_0 : os interceptos dos portfólios não são estatisticamente diferentes de zero

H_1 : os interceptos dos portfólios são estatisticamente diferentes de zero

Espera-se que os testes não rejeitem a hipótese nula nos portfólios testados, isto é, que os interceptos não sejam, estatisticamente, diferentes de zero.

(b) Average F -test

Em situações em que a matriz de covariância dos resíduos, $\hat{\Sigma}$, não pode ser invertida, o teste GRS fica impossibilitado de ser aplicado. Tendo isso em vista, Hwang e Satchell (2014)

desenvolveram uma variação desse teste, chamado de *Average F-test*, que se caracteriza como um teste *F* com 1 e $T - K - 1$ graus de liberdade.

O *Average F-test* é definido como:

$$S_2 = \frac{Tc}{N} \sum_{n=1}^N \frac{\hat{\alpha}_n^2}{\hat{\sigma}_n^2} \quad (57)$$

em que

$$\hat{\sigma}_n^2 = \sum_{t=1}^T \frac{\varepsilon_t^2}{(T-K-1)} \quad (58)$$

$$c = 1 + \bar{\mu}' \hat{\Omega}^{-1} \bar{\mu} \quad (59)$$

Os termos são os mesmos para o teste GRS. Além disso, ε_t^2 representa o quadrado dos resíduos das regressões. Neste estudo, o Average F-test foi aplicado nos retornos estimados dos portfólios, tendo em vista que eles produzem resíduos linearmente dependentes devido à repetição dos dados pela reamostragem com reposição. Além disso, esse teste se mostra robusto em situações em que o número de portfólios testados (N) e o período analisado (T) são significativamente maiores do que o número de fatores (K).

(c) *Variance Inflation Factor – VIF*

A estimação do *Variance Inflation Factor* (VIF) permite identificar se as variáveis independentes são correlacionadas, ou seja, se há multicolinearidade. A correlação entre as variáveis afeta a estimação dos parâmetros, tendo em vista que tais variáveis seriam *proxies* para o mesmo fenômeno. Segundo Greene (2002), VIFs acima de 20 indicariam multicolinearidade, enquanto Gujarati (2006) indica que valores acima de 10 já indicariam que esse fenômeno está presente entre as variáveis.

(d) *Teste de Durbin-Watson*

O teste de Durbin-Watson procura testar se os resíduos são correlacionados ao longo do tempo. Quando existe tal relação, a estimação dos erros-padrão dos estimadores é afetada. O teste de Durbin-Watson utiliza os resíduos dos mínimos quadrados por meio da seguinte estatística de teste:

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} = 2(1 - r) - \frac{e_1^2 + e_T^2}{\sum_{t=1}^T e_t^2} \quad (60)$$

Segundo Heij *et al.* (2004), a equação satisfaz $0 \leq d \leq 4$ e

$$d \approx 2(1 - r_1) \quad (61)$$

Na ausência de correlação de primeira ordem, o valor de $d \approx 2$. Ao mesmo tempo, valores de d próximos a zero indicam correlação positiva, enquanto valores de d próximos a 4 indicam correlação negativa. Tomando ρ como a correlação entre ε_t e ε_{t-1} , o teste de Durbin-Watson testa as seguintes hipóteses

$$H_0: \rho = 0$$

$$H_1: \rho \neq 0$$

(e) Teste de Breusch-Pagan

Um problema que ocorre ao utilizar regressões do tipo MQO é a presença de heterocedasticidade dos resíduos, ou seja, variância não constante entre eles, o que pode diminuir a eficiência dos estimadores. Para verificar se os resíduos são homocedásticos, utilizou-se o teste de Breusch-Pagan.

O primeiro passo deste teste é a aplicação de uma regressão MQO nos dados. Após o cálculo dos resíduos, se faz necessário aplicar regressões auxiliares em que os resíduos são as variáveis dependentes. Por fim, faz-se o cálculo $LM = nR^2$ que testa as seguintes hipóteses:

$$H_0: \text{a variância dos resíduos é homocedástica}$$

$$H_1: \text{a variância dos resíduos não é homocedástica}$$

(f) Teste de Jarque-Bera

Para testar a normalidade dos dados, necessário em diversos momentos ao longo do trabalho, foi utilizado o teste de Jarque-Bera que averigua qual o desvio dos valores da distribuição normal com assimetria e curtose iguais a zero e três respectivamente. Caso os valores encontrados destoem da distribuição gaussiana, a hipótese nula de normalidade é rejeitada.

(g) Bootstrap

O método de *bootstrap* para simulação dos dados foi utilizado em dois momentos nesta pesquisa. Primeiramente para o cálculo dos parâmetros do VPIN da amostra e posteriormente para a aplicação dos modelos de fatores testados.

Este método consiste em selecionar os dados através de reamostragens com reposição, e sob cada reamostragem, calcula-se os parâmetros requeridos. Ruppert e Matteson (2015)

colocam que os parâmetros estimados se aproximam de uma distribuição normal na medida que o número de simulações cresce. Finalmente, utiliza-se de um intervalo de confiança para se estabelecer um limite inferior e superior onde os parâmetros reais se encontram.

4 ANÁLISE DOS RESULTADOS

Nesta seção serão apresentados os resultados para a análise dos desempenhos dos métodos *Tick Rule* e *Bulk Volume Classification* na definição do lado agressor das transações. O resultado do VPIN para a amostra geral e dividido em relação aos segmentos da listagem da BM&FBovespa das ações e seus valores de mercado. Posteriormente, serão evidenciados os resultados envolvendo o objetivo central deste trabalho: a precificação do risco informacional no mercado nacional. Por meio dos testes GRS e *Average F-test* e com o suporte de simulações, foi destacada a combinação de fatores que otimiza a explicação dos retornos dos portfólios criados neste trabalho.

4.1 Desempenho dos métodos TR e BVC

A primeira contribuição deste trabalho se refere à aplicação dos métodos de classificação de sinais de transações. O primeiro algoritmo, *Tick Rule*, utiliza os dados *tick-by-tick* e apresenta solidez na literatura financeira, tendo sido comprovada sua eficiência em diversos mercados como citado anteriormente. Já o BVC se apresenta como uma solução para aqueles trabalhos que necessitam das classificações das transações, porém não têm acesso aos dados *tick-by-tick*. A vantagem desse modelo, segundo Easley, López de Prado e O'Hara (2012b), é a utilização de uma quantidade consideravelmente menor de dados em relação ao TR, o que implica: (i) menor tempo para coleta de dados e (ii) menor necessidade computacional para a aplicação do algoritmo.

O primeiro passo em direção à análise dos métodos no cenário nacional foi a aplicação de ambos na maior quantidade de ações possíveis. A janela de tempo analisada consiste do dia 02/05/2014 ao dia 31/05/2016, totalizando 512 dias de transações. Como o TR e o BVC podem ser aplicados ao longo dos dias independentemente – ou seja, os dados de outros dias não interferem no resultado do dia sendo estudado – aplicou-se somente um único filtro: um mínimo de duas transações no dia analisado. Este filtro abrange ambos os métodos, pois a metodologia do TR implica comparação do preço da transação que está sendo classificada com o preço da operação anterior e, portanto, necessita de, no mínimo, duas transações no dia analisado. Já o BVC leva em consideração a variação do preço em relação às ordens para a

classificação parcial do volume em compra e venda, evidenciando a necessidade, também, de, no mínimo, duas transações por dia para aplicação do método. Dessa forma, excluíram-se da amostra aqueles ativos com somente uma transação e, naturalmente, ambos os modelos, por não conseguirem classificar a primeira transação de cada ação, perdem, já de início, um poder preditivo de, no mínimo, uma ordem, dada a impossibilidade de classificá-la.

A tabela 8 evidencia, na primeira linha, a quantidade de ações analisadas por dia. A segunda linha mostra o resultado da taxa de precisão aplicada ao TR e a última linha indica a taxa de sucesso do método BVC. Como se pode perceber, o número mínimo de ações analisadas ao longo dos dias foi de 207, e o máximo, 240, com uma média de 221,56 ações por dia. O desvio-padrão da quantidade de ações foi relativamente baixo: de 5,67 ações. A quantidade de ações por dia varia, pois não se determinou uma amostra com tamanho definido, somente o filtro supracitado.

Tabela 8

Desempenho dos métodos TR e BVC para a amostra irrestrita

	Mínimo	Máximo	Média	Desv. Pad
Número de Ações	207	240	221,56	5,6703
TR	0,7954	0,8707	0,8323	0,0119
BVC	0,2967	0,8990	0,7162	0,0636

Fonte - Elaborada pelo autor da dissertação.

Verifica-se que a média de acertos do *Tick Rule* foi de 83,23%, com um desvio-padrão de somente 1,19%. Ao mesmo tempo, a amplitude entre a precisão mínima e a máxima não foi alta, tendo em vista que o valor mínimo foi de 79,54% e o máximo de 87,07%. Esse resultado está de acordo com a literatura, que afirma que o TR apresenta um bom desempenho na classificação *tick-by-tick*, como predito pela teoria de que um aumento no preço da transação geralmente indica uma ordem de compra e um decréscimo uma ordem de venda.

Em relação ao BVC, este apresentou um *spread* entre os valores mínimo e máximo de precisão muito maior do que o TR, tendo em vista que seu mínimo foi de 29,67% e seu máximo de 89,90%. Com uma média de 71,62%, esse método teve um desempenho geral inferior ao TR, o que está em concordância com o estudo de Easley, López de Prado e O'Hara (2012b) de que o BVC não apresenta uma boa performance em ativos menos líquidos, como os negociados no mercado brasileiro. O desvio-padrão do BVC também foi relativamente alto, de 6,36%, indicando que houve dias em que o BVC teve um desempenho perto de 80% –

considerado satisfatório para estudos que utilizam dados dessa natureza – e dias em que seu desempenho foi perto de 65%, apontando que cerca de 35% das transações foram classificadas erroneamente, o que causaria problemas no cálculo do VPIN ou da PIN, como apontado por Gremmig e Theissen (2002).

De forma geral, analisando uma amostra que contém concomitantemente ações transacionadas frequentemente e ações pouco líquidas com, por exemplo, duas transações em um dia, os resultados apontam que a aplicação do TR é mais indicada para a classificação dos sinais das ordens de transações, como verificado por Chakrabarty, Pascual e Shkilko (2015) e Omrane e Welch (2016).

Após a aplicação dos métodos TR e BVC na maior quantidade possível de ações, conforme tabela 8, prosseguiu-se à sua aplicação na amostra analisada nesta dissertação apresentada na tabela 1 que se constituiu de 142 ações transacionadas em todos os 512 dias analisados. Dessa forma, garante-se que os ativos em questão são os mais líquidos da BM&FBovespa, possibilitando a apresentação do desempenho do TR e BVC nessa categoria de ações, conforme tabela 9.

Tabela 9

Desempenho dos métodos TR e BVC para a amostra definida

	Mínimo	Máximo	Média	Desv. Pad
Número de Ações	142	142	142	0
TR	0,7957	0,8711	0,8320	0,0118
BVC	0,6494	0,8243	0,7477	0,0195

Fonte - Elaborada pelo autor da dissertação.

Novamente a primeira linha da tabela 9 indica o número de ações. Como a amostra foi fixada previamente, em todos os dias, os mesmos 142 ativos foram analisados, não tendo havido, portanto, nenhuma variação.

Em relação ao desempenho médio do TR apresentado na tabela 9, este foi praticamente igual àquele da tabela 8, tendo, somente, variações nos números da quarta casa decimal. Tal fato corrobora a evidência de que esse método tem boa performance, tanto em amostras com ativos menos líquidos, quanto em amostras com ativos frequentemente transacionados. Para essa amostra, a média do TR foi de 83,20% com um desvio-padrão de 1,18% indicando pouca variabilidade na precisão de seu algoritmo de classificação. O mínimo

e o máximo foram semelhantes àqueles apresentados na tabela 8 de 79,54% e 87,11% respectivamente.

Em relação ao BVC, como esperado, este apresentou um melhor desempenho quando se restringiu à aplicação do método a somente ações mais líquidas. O valor mínimo que, para a amostra irrestrita era de 29,67%, passou para 64,94%, ao mesmo tempo em que o desvio-padrão passou de 6,36% para 1,95%. Dessa forma, a variabilidade quando o BVC é aplicado a ações líquidas é consideravelmente menor, indicando que a utilização desse método é adequada quando se tem o cuidado de selecionar ações frequentemente transacionadas.

Em conclusão, apesar da clara melhora do desempenho do BVC em relação às amostras, o TR ainda assim apresentou uma melhor performance, tendo classificado corretamente, na média, 83,20% das transações contra 74,77% do BVC (para a amostra mais líquida). Dessa forma, por meio do que foi examinado, conclui-se que o TR se mostrou superior ao BVC, conforme já constatado em estudos anteriores (CHALRABARTY, PASCUAL e SHKILKO, 2015; OMRANE e WELCH, 2016). As vantagens do BVC apresentadas por Easley, López de Prado e O'Hara (2012b) ainda se sustentam, como a compactação dos dados em relação ao *tick-by-tick* – os autores mostram que o BVC pode ser aplicado com somente 0,35% da quantidade de dados utilizadas pelo TR. Isto posto, ele se apresenta como uma alternativa viável ao se utilizar base de dados intradiários como a Bloomberg®, especialmente quando se opta por restringir a amostra às ações mais líquidas do mercado.

4.2 Estatísticas descritivas do VPIN

Tendo averiguado os desempenhos dos métodos TR e BVC no cenário nacional, passa-se agora para a análise do VPIN calculado com os dados reais para as 142 ações expostas na tabela 1.

Ressalta-se que primeiramente serão mostrados os valores do VPIN e suas estatísticas descritivas de forma indiscriminada para todas as ações analisadas em ordem alfabética. Posteriormente, tal amostra será dividida em relação a seus segmentos da BM&FBovespa para avaliar se há diferença estatisticamente significativa entre o VPIN de empresas de diferentes segmentos de mercado. Por fim, o VPIN das ações será analisado a partir da

separação da amostra em relação ao tamanho das empresas. Um dos principais resultados, comuns a todas as pesquisas que aplicaram, tanto a PIN, quanto o VPIN, era a relação entre o tamanho da empresa e sua probabilidade de ser alvo de transações privilegiadas. De forma geral Easley *et al.* (1996), Easley, Hvidkjaer e O'Hara (2002, 2005), Mohanram e Rajgopal (2009) e Hwang *et al.* (2013) demonstraram que companhias maiores apresentavam PIN (e VPIN) mais baixos, indicando que tamanho e probabilidade de negociação privilegiada são negativamente correlacionados. Para testar essa afirmação, a correlação entre as variáveis também foi calculada. Este é o primeiro passo para analisar a relação do VPIN com o retorno requerido, o que será feito adiante, juntamente aos modelos de precificação de ativos apresentados na metodologia.

4.2.1 Análise do VPIN de todas as ações da amostra

A tabela 10 evidencia as estatísticas descritivas do VPIN para todas as 142 ações analisadas, classificadas por ordem alfabética de seus *tickers*. Percebe-se que a grande maioria delas apresentou mínimos iguais a 0 e máximos iguais a 1, isto significa que houve momentos em que a proporção entre o volume de compras e o de vendas foi igual para todos os cinquenta baldes usados para o cálculo do VPIN e momentos em que todo o volume estava concentrado ou em ordens de compra ou em ordens de venda, acarretando, respectivamente, em VPINs iguais a 0 e 1. De forma geral, percebe-se que a média do VPIN para a maioria das ações se distingue bastante, sendo que algumas ações apresentam média no intervalo de 0,20 a 0,30 e outras têm VPIN acima de 0,60. Em relação ao desvio-padrão, percebe-se que, quanto maior a média do VPIN para a ação, maior é seu desvio-padrão. Tal relação é confirmada pela correlação positiva e forte de 0,7640 entre a média do VPIN e seu desvio-padrão, indicando que, quanto menor o VPIN de uma ação, menor é a tendência de esse valor se desviar da média.

Em relação aos seus momentos superiores, assimetria e curtose, os valores para o terceiro momento geralmente são positivos, indicando uma tendência de o VPIN se concentrar em valores abaixo de sua média. Em relação ao quarto momento, todas as ações apresentaram valores positivos e predominantemente acima de três para a curtose, indicando que as distribuições do VPIN são, em geral, leptocúrticas, com valores concentrados ao redor da

média. De forma geral os VPINs não apresentam uma distribuição uniforme entre as ações. O teste de Jarque-Bera rejeitou a hipótese nula de normalidade (p.valor igual a zero) para o VPIN de todas as ações, o que já era esperado tendo em vista o comportamento dos momentos estatísticos das ações apresentado na tabela 10.

Para comparar os resultados obtidos nesta pesquisa com os de outros trabalhos que: (i) procuraram mensurar a probabilidade de negociações privilegiadas no mercado brasileiro e (ii) aplicaram a metodologia do VPIN em diferentes mercados acionários, se faz necessário o cálculo do VPIN para a amostra como um todo. Tais resultados são apresentados na tabela 11.

Tabela 10

Estatísticas descritivas do VPIN para toda a amostra

Nº de ações	Mínimo	Máximo	Média	Desvio Padrão	Assimetria	Curtose
142	0	1	0,4548	0,2218	0,8251	2,9426

Fonte - Elaborada pelo autor da dissertação.

Os valores mínimos e máximos haviam sido previamente mostrados na tabela 10, e a média do VPIN para o mercado acionário brasileiro foi de 0,4548, com desvio-padrão igual a 0,2218. A variação do VPIN já era esperada, tendo em vista que diversas ações apresentaram VPINs em torno de 0,30 e outras superiores a 0,60. Os terceiro e quarto momentos, quando analisado o VPIN em conjunto, apresentam valores menos acentuados do que para as ações individualmente. A assimetria positiva mostra novamente que os valores se concentraram abaixo da média, e a curtose próxima de três mostra uma distribuição platicúrtica, diferentemente dos valores acima de três apresentados para as ações individualmente.

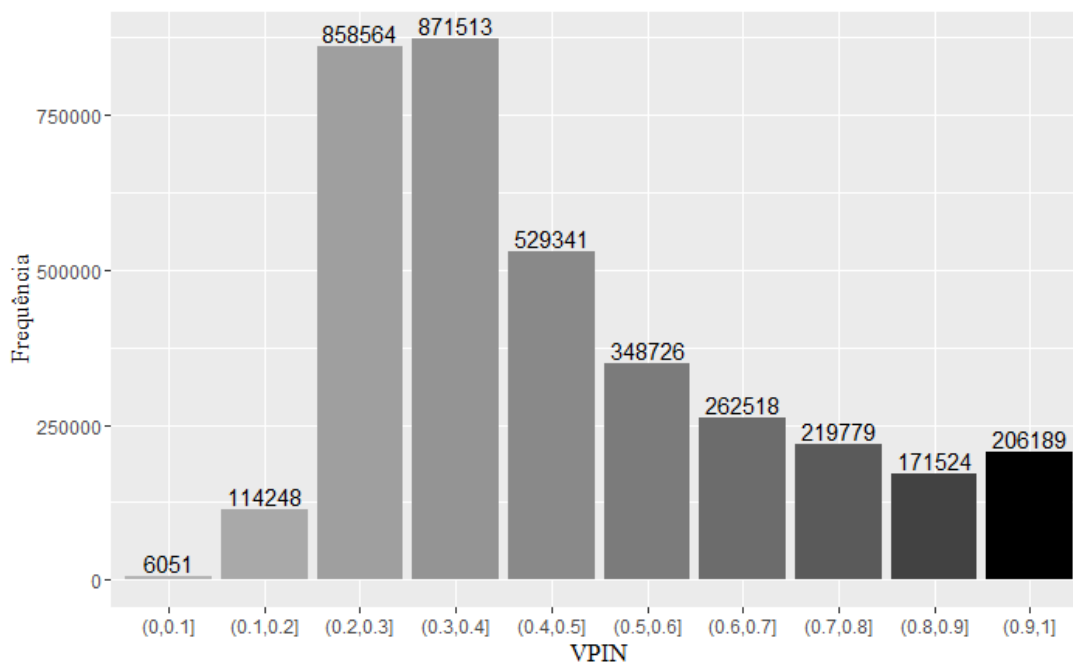
O gráfico 1 mostra a distribuição dos VPINs em relação a seus intervalos. Como o valor médio para o VPIN no mercado nacional foi de 0,4548, e a maior ocorrência de valores foi no intervalo de (0,2,0,4] a assimetria da distribuição do VPIN apresenta, naturalmente, valor positivo. Apesar desta assimetria positiva, fica claro, a partir da gráfico 1, que a ocorrência de valores no intervalo entre (0,9,1] foi responsável pelo aumento da média geral do VPIN no mercado nacional, o que indica fortes momentos de polaridade no volume de compra ou de vendas de ações.

Assimetria	1,3126	0,7494	0,8075	2,7105	1,8036	1,2782	-1,1974	1,1734	2,1858	-1,7820	2,8618	0,4428	1,0043	0,5020
Curtose	6,8731	6,0487	3,9690	19,3124	7,7328	6,2099	6,6812	6,1672	11,1828	9,2809	18,0958	4,1894	6,0242	3,7316
Ações	GOAU4	GOLL4	GRND3	HBOR3	HGTX3	HYPE3	IGTA3	ITSA4	ITUB4	JBSS3	JHSF3	JSLG3	KEPL3	KLBN11
Média	0,5029	0,4489	0,4788	0,5930	0,3830	0,2932	0,3694	0,2906	0,2552	0,2956	0,7329	0,6833	0,7770	0,3677
Desv. Padrão	0,1551	0,1145	0,2074	0,1728	0,1163	0,0905	0,0984	0,0663	0,0526	0,0934	0,1918	0,2322	0,1326	0,1319
Mínimo	0,1171	0,1821	0,0000	0,0000	0,0000	0,0000	0,0895	0,1300	0,0920	0,0474	0,0000	0,0000	0,0000	0,0000
Máximo	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,6720	0,6931	1,0000	1,0000	1,0000	1,0000	1,0000
Assimetria	-0,6342	0,5623	1,1037	0,2144	1,5437	1,9004	1,3492	0,6078	1,0190	2,5201	-1,6792	-1,0004	-0,9662	1,8348
Curtose	2,7006	3,1366	4,1757	3,3415	7,7893	11,7813	7,2053	3,5083	6,8971	15,7319	7,4771	4,2647	5,4757	8,6860
Ações	KLBN4	KROT3	LAME3	LAME4	LEVE3	LIGT3	LINX3	LLIS3	LOGN3	LPSB3	LREN3	MAGG3	MDIA3	MGLU3
Média	0,6436	0,3314	0,4418	0,3140	0,4765	0,3467	0,5293	0,5918	0,7457	0,6125	0,3330	0,6833	0,4665	0,5099
Desv. Padrão	0,2321	0,0944	0,1465	0,0992	0,1533	0,0943	0,1895	0,3149	0,1874	0,2257	0,1287	0,3109	0,1198	0,1099
Mínimo	0,0000	0,0086	0,0000	0,0000	0,0000	0,0946	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,1173
Máximo	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Assimetria	-1,1180	1,7439	1,1256	3,1753	0,5068	1,7917	-0,1675	-0,7672	-1,3621	-0,5421	2,1664	-1,0649	1,4122	0,4054
Curtose	4,3835	10,0245	6,7783	20,0204	5,1308	9,7240	4,4691	2,5408	6,5576	3,7123	10,8962	3,1058	8,4267	3,1709
Ações	MILS3	MMXM3	MPLU3	MRFG3	MRVE3	MULT3	MYPK3	NATU3	ODPV3	OGXP3	OIBR3	OIBR4	PCAR4	PDGR3
Média	0,5102	0,7982	0,3567	0,3827	0,3177	0,2945	0,4149	0,3240	0,4061	0,7162	0,5789	0,5340	0,2998	0,7111
Desv. Padrão	0,1709	0,1340	0,1463	0,1063	0,1005	0,0965	0,1350	0,0848	0,1251	0,3466	0,1606	0,1488	0,0814	0,1547
Mínimo	0,0000	0,0481	0,0000	0,0000	0,0381	0,0000	0,0000	0,1453	0,0080	0,0000	0,0537	0,0000	0,1250	0,0000
Máximo	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9141	1,0000	1,0000	1,0000	1,0000	0,7840	1,0000
Assimetria	0,5629	-0,9228	1,7933	1,7144	1,3371	2,2095	1,1187	1,3331	1,7033	-1,1951	-0,0726	-0,5567	1,0468	-0,6438
Curtose	4,1425	4,6412	8,1348	10,3012	8,0009	12,2780	5,5180	6,2039	7,7772	2,9593	2,5442	3,4429	4,7209	5,0163
Ações	PETR3	PETR4	PFRM3	PMAM3	POMO4	POSI3	PRML3	PSSA3	QGEP3	QUAL3	RADL3	RENT3	ROMI3	RSID3
Média	0,2261	0,2538	0,6901	0,7001	0,4609	0,6728	0,7722	0,3054	0,4521	0,3012	0,2909	0,2992	0,7945	0,6488
Desv. Padrão	0,0447	0,0458	0,2310	0,1678	0,1009	0,1851	0,1465	0,0807	0,0913	0,1117	0,0974	0,1129	0,2074	0,1319
Mínimo	0,1183	0,1290	0,0000	0,0000	0,1195	0,0000	0,0000	0,0854	0,0000	0,0000	0,0000	0,0130	0,0000	0,2163
Máximo	0,4985	0,5318	1,0000	1,0000	1,0000	1,0000	1,0000	0,9266	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000
Assimetria	0,7532	0,5072	-1,2960	-1,5710	0,4895	-0,6153	-1,0961	1,7311	1,2820	2,4136	2,4683	2,7008	-1,6500	-0,2022

Curtose	4,3327	3,7945	4,9751	8,2548	3,5254	4,2519	5,7815	9,6069	8,9605	12,5192	15,6680	14,1992	6,2437	3,0299
Ações	SBSP3	SEER3	SGPS3	SLCE3	SLED4	SMLE3	SMT03	STBP11	SULA11	SUZB5	TAE11	TBLE3	TCSA3	TECN3
Média	0,3224	0,4359	0,8644	0,4854	0,7181	0,3117	0,3992	0,5920	0,3288	0,3126	0,3705	0,2851	0,5892	0,7755
Desv. Padrão	0,1008	0,1222	0,1631	0,1525	0,1750	0,1049	0,1269	0,2188	0,0846	0,0917	0,1253	0,1014	0,1579	0,2052
Mínimo	0,0000	0,0000	0,0000	0,0000	0,0000	0,0434	0,0000	0,0000	0,1185	0,0000	0,0000	0,0440	0,0000	0,0000
Máximo	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,8178	1,0000	1,0000	1,0000	1,0000	1,0000
Assimetria	2,2329	0,8094	-2,2022	0,7996	-0,8745	2,3153	1,5703	-0,4166	1,3642	3,1329	2,2594	2,9041	0,2875	-1,9663
Curtose	12,8248	6,0776	9,7038	6,0035	5,2793	12,9936	6,8841	3,9672	6,4649	22,8332	11,1082	18,1724	4,0697	7,7544
Ações	TERI3	TGMA3	TIMP3	TOTS3	TPIS3	TRPL4	TUPY3	UCAS3	UGPA3	USIM5	VALE3	VALE5	VIVT4	VLID3
Média	0,7902	0,5688	0,2882	0,3026	0,7182	0,3812	0,5746	0,7546	0,2899	0,4453	0,2566	0,2501	0,2999	0,3504
Desv. Padrão	0,2475	0,1710	0,0954	0,1160	0,2483	0,1102	0,2018	0,2861	0,0902	0,1190	0,0557	0,0470	0,0881	0,0787
Mínimo	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,1912	0,1311	0,1163	0,0564	0,1554
Máximo	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	0,9058	0,5566	0,5793	1,0000	0,7504
Assimetria	-1,8339	0,0861	3,1009	2,3367	-1,6470	1,8426	-0,0411	-1,4517	2,1449	0,4574	1,0334	1,0791	1,9056	0,7798
Curtose	6,0676	3,9434	21,7099	13,8451	5,4490	9,1768	3,6341	4,3294	13,5110	2,6114	4,8656	6,9444	11,3626	4,2489
Ações	VVAR11	WEGE3												
Média	0,4655	0,3238												
Desv. Padrão	0,1301	0,0827												
Mínimo	0,0308	0,1312												
Máximo	1,0000	0,8532												
Assimetria	0,9321	1,0920												
Curtose	4,9824	5,3219												

Fonte - Elaborada pelo autor da dissertação.

Gráfico 1 – Distribuição do VPIN por intervalos



Fonte - Elaborado pelo autor da dissertação.

A partir da média encontrada para o VPIN, é possível traçar um paralelo com os valores encontrados por Barbedo, Silva e Leal (2009) e Martins e Paulo (2013, 2014). Barbedo, Silva e Leal (2009) foram os primeiros a aplicar o PIN no Brasil. Os autores, utilizando uma amostra de 48 ações líquidas entre o período de 2001 e 2006, encontraram um PIN médio de 0,1250, valor menor do que o valor do VPIN encontrado nesta pesquisa. Martins e Paulo (2013, 2014) analisaram 229 ações de 194 empresas entre os anos de 2010 e 2011 e encontraram um PIN médio de 0,2490, valor este duas vezes maior do que o encontrado por Barbedo, Silva e Leal (2009), porém quase metade do valor encontrado para o VPIN das ações aqui analisadas.

A diferença entre os resultados das pesquisas supracitadas e esta pode residir, tanto na metodologia utilizada, já que os autores aplicaram o PIN, quanto na questão da classificação dos dados. Barbedo, Silva e Leal (2009) e Martins e Paulo (2013, 2014) aplicaram algoritmos de classificação do lado iniciador da transação – Barbedo, Silva e Leal (2009) não discutem o algoritmo utilizado, enquanto Martins e Paulo (2013) utilizaram o método de Lee e Ready (1991). Grammig e Theissen (2002) discutem os problemas relacionados a má classificação das ordens e dizem que este pode gerar valores enviesados para o PIN estimado. Além disso, ambos os trabalhos aplicaram o

PIN em momentos e em amostras diferentes. Levando em consideração essas condições, a comparação dos resultados desses autores com os deste estudo encontra dificuldades inerentes aos procedimentos utilizados em cada pesquisa. Tendo isto em vista, passa-se à comparação do VPIN encontrado no mercado nacional com o VPIN calculado em outros países.

Dos trabalhos que se valeram do VPIN para o cálculo do nível de toxicidade dos fluxos de transações, Borochin e Rush (2016) analisaram uma amostra no mercado norte-americano que consiste em todas as ações com preço acima de cinco dólares para o período de 1993 a 2013. Os autores encontraram um valor mínimo para o VPIN de 0 e média de 0,1774. O valor máximo foi de aproximadamente 0,50, resultado este próximo do VPIN médio encontrado para o cenário nacional.

Cheung, Chon e Lei (2015), analisando trinta e seis ações do mercado chinês, encontram um VPIN médio de 0,1943, com desvio-padrão de 0,0904, valor mínimo de 0 e máximo de 1. Isto indica que, de maneira geral, o desequilíbrio do volume de compra e de vendas no mercado chinês foi menor do que no Brasil. No entanto, momentos de completo desequilíbrio entre o volume de compra e de venda ($VPIN = 1$) e puro equilíbrio entre as ordens de transações ($VPIN = 0$) também foram encontrados para aquele mercado. De forma geral, esse é o único trabalho que também apresentou os dois valores extremos para o VPIN.

Abad e Yagüe (2012) foram os primeiros a aplicar o VPIN em ações. Estudando o mercado acionário espanhol, utilizam uma amostra de 15 ações transacionadas no ano de 2009. Os autores dividem as ações em três grupos baseados no valor de mercado: pequenas, médias e grandes ações. Abad e Yagüe (2012) encontram para as ações do grupo *pequeno* um VPIN de 0,532, para o grupo *médio* este valor foi de 0,402, e, finalmente, para o grupo das maiores empresas, o VPIN foi de 0,254. Os valores encontrados pelos autores ao classificarem as ações em diferentes grupos foi similar ao encontrado por Yildez, Van Ness e Van Ness (2016) e Wei, Gerace e Frino (2013).

Yildez, Van Ness e Van Ness (2016), analisando 120 ações transacionadas no mercado norte-americano para o ano de 2009, encontraram, utilizando a mesma separação em grupos de Abad e Yagüe, os valores de 0,6154, 0,40 e 0,2378 para o VPIN dos grupos pequeno, médio e grande, respectivamente. Wei, Gerace e Frino (2013) dividem uma amostra de 30 ações em três grupos em relação ao peso de cada

ativo no índice de mercado S&P/ASX 200, referente ao mercado acionário australiano, entre os anos de 2008 e 2010. Os autores constataram que o VPIN é maior para o grupo de ações com menor peso no índice do que para aquelas ações com maior peso. O VPIN médio calculado pelos autores é de 0,6903. Os resultados encontrados por esses trabalhos não diferem significativamente do VPIN encontrado para o mercado acionário brasileiro, indicando, portanto, que o nível de toxicidade das ações na BM&FBovespa se encontra em um valor similar àquele apresentado para outros mercados proeminentes como o espanhol, o australiano e o americano – este último referente à amostra de Yildez, Van Ness e Van Ness (2016).

Os trabalhos de Abad e Yagüe (2012), Gerace, Wei e Frino (2013) e Yildez, Van Ness e Van Ness (2016) reforçam os achados de Easley *et al.* (1996) de que a probabilidade de negociações privilegiadas é negativamente correlacionada com o tamanho da empresa. Os resultados encontrados no presente trabalho também seguem na mesma direção e serão apresentados mais à frente.

Não serão feitas comparações dos resultados deste trabalho com os de Easley, López de Prado e O'Hara (2011a,b, 2012a). Os autores focam o estudo do *Flash Crash* – evento que envolveu uma queda brusca do índice Dow Jones Industrial Average em 2010 – e calculam o VPIN para o mercado de futuros, em especial, o índice E-mini S&P500. Por essa razão, optou-se por comparar o valor encontrado neste trabalho somente com aquelas pesquisas que aplicaram o VPIN no mercado à vista em diferentes países.

4.2.1.1 Análise da simulação do VPIN por bootstrap

De forma a expandir os resultados apresentados na seção 4.2.1, optou-se por utilizar o método de reamostragem, ou *bootstrap*, em relação aos valores do VPIN encontrados no cenário nacional. Como o VPIN foi aplicado em 142 ações, em um período de cerca de dois anos (512 dias de transações), a aplicação do método do *bootstrap* permitiu fazer inferências acerca de dois dos principais parâmetros do VPIN – média e desvio-padrão.

O principal objetivo do *bootstrap*, segundo Ruppert e Matteson (2015), é o cálculo dos intervalos de confiança para os parâmetros. Apesar de não ser necessário

que a amostra apresente uma distribuição normal, o intervalo de confiança calculado presume a distribuição normal dos parâmetros calculados, obedecendo o teorema central do limite. De forma a seguir este procedimento, foram feitas 100.000 simulações a partir da reamostragem com reposição dos VPINs calculados. O alfa utilizado foi de 5%, o que indica, portanto, um intervalo de confiança de 95%. Os valores encontrados para as simulações são apresentados na tabela 12.

Tabela 12

Intervalo de confiança dos parâmetros do VPIN calculados a partir do método *bootstrap*

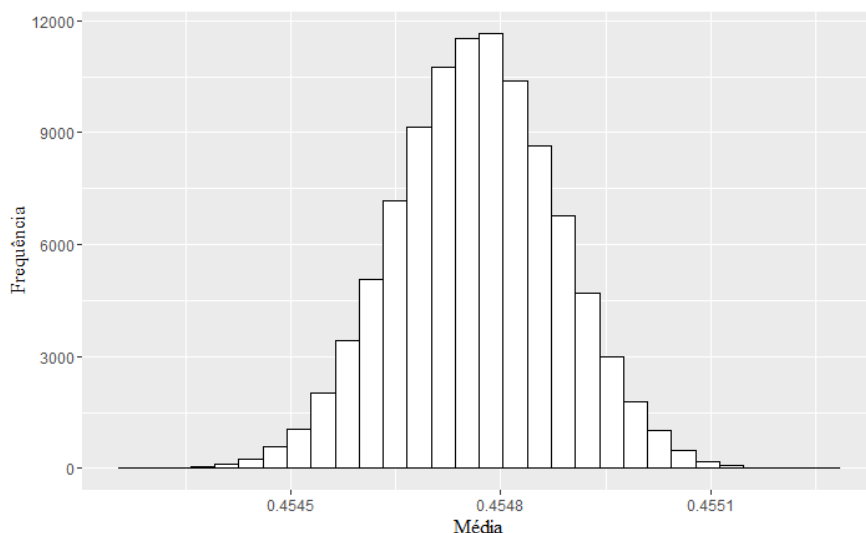
Intervalo de Confiança	Lim. Inferior	Média	Lim. Superior
Média	0,4545	0,4548	0,4551
Desvio Padrão	0,2217	0,2219	0,2221

Fonte - Elaborada pelo autor da dissertação.

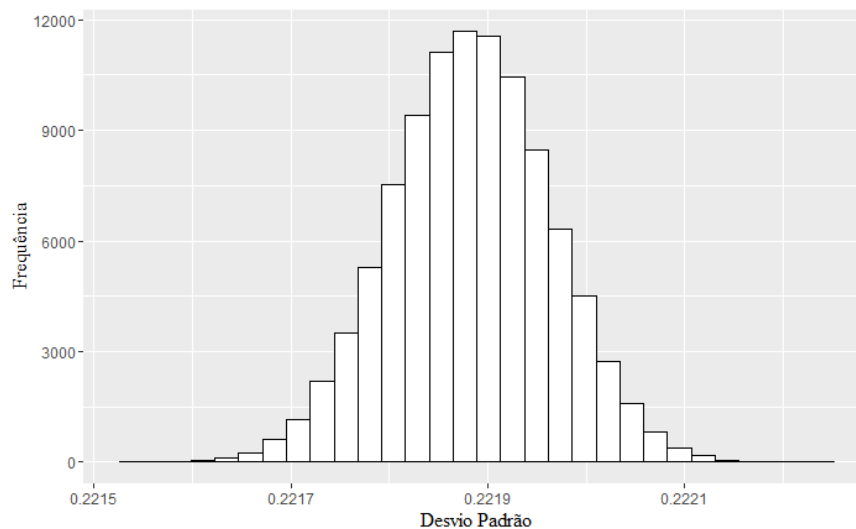
Nota: Foram feitas 100.000 simulações. Além disso, foi utilizado um alfa de 5% para o cálculo dos limites inferior e superior dos parâmetros.

A partir dos resultados da tabela 12, pode-se concluir com 95% de confiança que o VPIN das ações no mercado brasileiro se encontra em um intervalo entre [0.4545, 0.4551], enquanto seu desvio-padrão se encontra em um intervalo entre [0.2217, 0.2221]. Foram calculados também os vieses da média e do desvio-padrão subtraindo a média dos parâmetros simulados do valor dos parâmetros calculados da amostra original. Esses valores foram respectivamente $2,97 \times 10^{-7}$ e $1,96 \times 10^{-7}$. O valor próximo de zero para ambos os vieses dos parâmetros mostram uma solidez nos resultados apresentados pelo *bootstrap*, tendo em vista que, para esse método, a amostra original é, para a amostra simulada, a população, portanto, o viés entre o parâmetro amostral e populacional é irrisório. De forma a corroborar esses resultados, as distribuições da média e do desvio-padrão calculados para cada simulação são apresentadas abaixo.

Gráfico 2 – Distribuição dos parâmetros simulados – média e desvio padrão

a) Distribuição do *bootstrap* da média

Nota: a média foi calculada para cada reamostragem, portanto foram obtidas 100.000 médias diferentes para os valores do VPIN encontrados para as 142 ações.

b) Distribuição do *bootstrap* do desvio padrão

Fonte - Elaborada pelo autor da dissertação.

Nota: o desvio-padrão foi calculado para cada reamostragem, portanto foram obtidos 100.000 desvios-padrão diferentes para os valores do VPIN encontrados para as 142 ações.

Ruppert e Matteson (2015) apontam que o cálculo dos intervalos presume que os parâmetros seguem uma distribuição normal. Os gráficos acima apresentam a

distribuição da média e dos desvios-padrão calculados via reamostragem. Ambos aparentam seguir a distribuição gaussiana. Para verificar, foi realizado o teste de Jarque-Bera nos parâmetros. Para a média, o p.valor do teste foi de 0,194, indicando não rejeição da hipótese nula de normalidade. Para o desvio-padrão, o p.valor do teste foi de 0,032, indicando que, com um alfa igual a 5%, a hipótese nula de normalidade seria rejeitada. Para valores de alfa menores que 5%, no entanto, a hipótese nula é aceita. Dessa forma, entende-se que os parâmetros calculados satisfazem o objetivo da reamostragem que é o cálculo do intervalo de confiança dos parâmetros.

Por fim, conclui-se, com 95% de confiança, que a média e o desvio-padrão do VPIN das ações do mercado brasileiro para o período analisado neste trabalho está dentro do intervalo apresentado na tabela 12. As demais estatísticas descritivas não foram submetidas ao processo de simulação por reamostragem.

4.2.2 Análise do VPIN por segmento de mercado da Bovespa

Os segmentos de listagem diferenciados da BM&FBovespa – Nível 1 (N1), Nível 2 (N2) e Novo Mercado (NM) – foram criados com o objetivo de passar ao investidor maior segurança ao transacionar ações no mercado financeiro, ou seja, menor risco de, por exemplo, ser exposto a transações privilegiadas (BARBEDO, SILVA e LEAL, 2009). Cada nível demanda que a empresa siga uma série de procedimentos que visam gerar maior segurança para os investidores do mercado. Quanto maior o segmento de listagem em termos da governança diferenciada, maior o comprometimento com a transparência por parte da empresa. Dessa forma, empresas que se encontram no segmento NM têm, em tese, maior transparência do que aquelas dos níveis N1 e N2 e do segmento Tradicional. Tendo isto em vista, Iquiapaza, Lamounier e Amaral (2007), Albanez e Valle (2009), Barbedo, Silva e Leal (2009) e Martins e Paulo (2013) procuraram estudar o impacto da assimetria informacional no contexto de governança corporativa, sendo que os últimos dois trabalhos utilizaram o PIN para tal fim.

No presente trabalho, as ações foram primeiramente separadas em relação a seus segmentos de listagem. O segmento de cada ação pode ser visto na tabela 1. O processo do cálculo do VPIN para cada segmento foi feito de forma diferente do que foi apresentado nas seções anteriores. O primeiro passo foi o cálculo do VPIN diário para

cada ação. Posteriormente, foi calculada a média diária do VPIN para cada segmento. Esse processo gerou resultados diferentes daqueles apresentados na seção 4.2, pois: (i) o VPIN usado foi o diário e (ii) a média diária por segmento faz com que o comportamento dos VPINs seja diferente daquele apresentado para os dados gerais. A tabela 13 apresenta os resultados obtidos.

Tabela 13

Estatísticas descritivas do VPIN diário por segmento de listagem da BM&FBovespa

Segmentos	N1	N2	NM	Trad.
Nº de ações	24	12	94	12
Mínimo	0,3168	0,3669	0,4270	0,3466
Máximo	0,4310	0,6001	0,5039	0,5695
Média	0,3761	0,4870	0,4700	0,4597
Desvio Padrão	0,0191	0,0355	0,0124	0,0379
Assimetria	-0,2811	-0,1433	-0,1102	-0,1851
Curtose	3,0784	3,4434	3,0461	2,9702

Fonte - Elaborada pelo autor da dissertação.

Nota: 'N1' se refere ao Nível 1 de Governança Corporativa. 'N2' ao Nível 2. 'NM' ao Novo Mercado e 'Trad' ao Tradicional.

Verifica-se que o VPIN do Nível 2 foi o maior entre os segmentos de listagem da Bovespa, seguido pelo Novo Mercado, Tradicional e, finalmente, o Nível 1. Em relação ao desvio-padrão, o segmento Tradicional apresentou o maior desvio em relação à média do VPIN e o NM, o menor. Os terceiro e quarto momentos apresentaram resultados divergentes daqueles do VPIN para toda a amostra. Analisando o VPIN diário para cada segmento, tem-se assimetria negativa, ou seja, valores mais concentrados após a média e curtose próxima daquela apresentada pela distribuição normal. É importante notar que a quantidade de ações em cada segmento se difere bastante, sendo o Novo Mercado o segmento com maior número de empresas. Tal fato impacta fortemente os resultados por segmento, como será explicado na seção seguinte.

De forma a verificar estatisticamente a diferença entre as médias dos VPINs dos segmentos, utilizou-se o t-test de Student. Para a aplicação desse teste, foi aplicado primeiramente o teste de Jarque-Bera para a normalidade dos VPINs dos segmentos de listagem. Os resultados para esse teste são apresentados na tabela 14.

Tabela 14

Teste Jarque-Bera para a normalidade dos VPINs diários dos segmentos de listagem da BM&FBovespa

Segmentos	N1	N2	NM	Trad.
χ^2	6,8632	5,9362	1,0798	2,9380
<i>p-value</i>	0,0323	0,0514	0,5828	0,2302

Fonte - Elaborada pelo autor da dissertação.

Nota: 'N1' se refere ao Nível 1 de Governança Corporativa. 'N2' ao Nível 2. 'NM' ao Novo Mercado e 'Trad' ao Tradicional.

Analisando os resultados da tabela 14, verifica-se que a hipótese de normalidade dos VPINs não é rejeitada para praticamente todos os segmentos. Para o N1, a hipótese nula é aceita para níveis de significância de 1%, mas rejeitada para α igual a 5%. De qualquer forma, procede-se à utilização do teste t para a verificação da hipótese de diferença entre as médias.

Tabela 15

Teste t para a diferença entre as médias dos VPINs das ações por segmento de listagem da BM&FBovespa

Segmentos	N1	N2	NM	Trad.
N1	0 (1,000)			
N2	-62,022 (0000)	0 (1,000)		
NM	-92,915 (0000)	10,158 (0000)	0 (1,000)	
Trad	-44,489 (0000)	11,852 (0000)	5,8504 (0000)	0 (1,000)

Fonte - Elaborada pelo autor da dissertação.

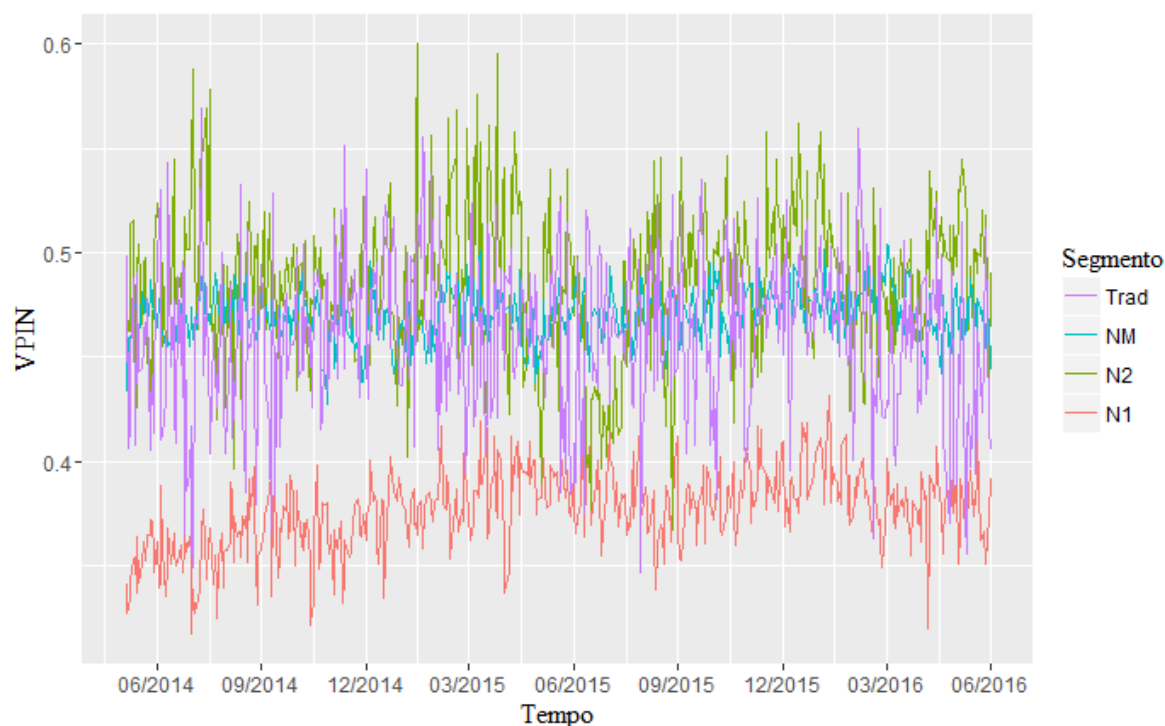
Nota: os valores em parênteses de referem aos p-valores do teste t de Student para a diferença entre as médias dos setores. Todos os testes rejeitam a hipótese nula de que a diferença é igual a zero, portanto aceita-se a hipótese alternativa de que há diferença entre as médias dos setores. 'N1' se refere ao Nível 1 de Governança Corporativa. 'N2' ao Nível 2. 'NM' ao Novo Mercado e 'Trad' ao Tradicional.

A tabela 15 mostra a rejeição da hipótese nula de igualdade de valores para a média do VPIN para todos os segmentos, ou seja, mesmo os segmentos N2, NM e Tradicional apresentando valores similares, a diferença entre eles é estatisticamente diferente de zero. Tal fato permite a conclusão de que, para a amostra analisada, o segmento N1 apresenta o menor nível de toxicidade dos fluxos de transações seguido do Tradicional, NM e, por fim, o N2.

Para melhor apresentar o comportamento dos VPINs diários dos segmentos, os valores foram colocados em um gráfico de série temporal, mostrando desde o início do

período até o final. O gráfico 3 mostra que o VPIN para as ações pertencentes ao segmento N1 se manteve mais baixo do que para os outros segmentos para todo o período analisado. O VPIN diário das ações do Novo Mercado se manteve mais estável, como demonstrado pelo menor desvio-padrão entre os segmentos (0,0124). A média similar entre os segmentos N1, NM e Tradicional, apresentada na tabela 13, não traduz de forma adequada o comportamento dos VPINs dos segmentos. O gráfico 3 demonstra que o VPIN das ações do Nível 2 atingiu valores acima da média (chamados de picos) mais frequentemente do que os outros segmentos. De forma oposta, o segmento Tradicional apresentou valores mais baixos para o VPIN mais frequentemente que os outros segmentos. É interessante ressaltar também que os VPINs dos segmentos não apresentaram correlação significativa – a maior correlação foi entre o NM e o N1, de 0,10 e p.valor de 0,016.

Gráfico 3 – Evolução do VPIN diário dos segmentos



Fonte - Elaborado pelo autor da dissertação.

Nota: 'N1' se refere ao Nível 1 de Governança Corporativa. 'N2' ao Nível 2. 'NM' ao Novo Mercado e 'Trad' ao Tradicional.

Esperava-se que os valores apresentados pelo NM que, em tese, é o segmento mais transparente da Bovespa, fossem significativamente mais baixos do que os dos outros segmentos, em especial, do Tradicional. Os resultados encontrados, foram, no entanto, contrários à hipótese definida nesta pesquisa. Constatou-se que o segmento N1

apresentou o menor VPIN para a amostra analisada e que o segmento N2 exibiu maior VPIN médio no período em questão. Os motivos de tais disparidades serão discutidos na seção seguinte, quando o VPIN é analisado juntamente ao tamanho da empresa.

Por fim, a partir dos resultados obtidos com as análises dos VPINs dos segmentos de governança diferenciada, é possível traçar um paralelo com os trabalhos de Barbedo, Silva e Leal (2009) e Martins e Paulo (2013). Barbedo, Silva e Leal (2009) analisaram os PINs de cada segmento e verificaram que o Tradicional apresentou o menor PIN (0,083), seguido do N1 (0,089), N2 (0,176) e NM (0,153). Dessa forma, os resultados para o VPIN se assemelham aos de Barbedo, Silva e Leal (2009), visto que o N1 apresentou resultado substancialmente baixo e próximo ao Tradicional; e o N2 exibiu o maior nível de probabilidade de negociações privilegiadas, seguido do NM. Em comparação com este trabalho, a única diferença entre os resultados auferidos por Barbedo, Silva e Leal (2009) foi em relação ao segmento Tradicional.

Martins e Paulo (2013), no entanto, encontraram resultados opostos. Para os autores, o segmento N1 apresentou o maior PIN no período analisado (0,256), seguido pelo Tradicional (0,25), NM (0,245) e N2 (0,244). O teste de Mann-Whitney aplicado pelos autores rejeita a igualdade da média para todos os segmentos. Dessa forma, para Martins e Paulo (2013), o segmento N2 apresentou o menor PIN enquanto o N1, o maior. Novamente, devem-se deixar evidentes as ressaltas feitas anteriormente em relação às diferenças metodológicas entre as pesquisas, já que Barbedo, Silva e Leal (2009) e Martins e Paulo (2013) se valem do PIN e de algoritmos de classificação de transações.

4.2.3 Análise do VPIN por valor de mercado da ação

Um dos principais resultados presentes em praticamente todos os estudos que aplicaram o PIN e, mais recentemente, o VPIN é de que a probabilidade de negociação privilegiada ou nível de toxicidade dos fluxos de transações é menor para ações maiores. Easley *et al.* (1996), ao aplicarem a PIN, verificam tal relação negativa entre esses dois fatores. As explicações dadas pelos autores residiam na maior presença de analistas produzindo pareceres a respeito de empresas maiores, além da dispersão dos agentes privilegiados em relação ao montante total transacionado diariamente para uma empresa de grande porte.

Para o VPIN, por exemplo, a disparidade entre os volumes de compra e de venda não são tão acentuados para empresas maiores, reduzindo, portanto, o seu nível de toxicidade dos fluxos. Abad e Yagüe (2012) foram os primeiros a constatar tal relação em ações transacionadas no mercado espanhol. Wei, Gerace e Frino (2013) e Yildiz, Van Ness e Van Ness (2016) encontraram evidências que dão suporte a tal afirmação em estudos nos mercados australianos e norte-americanos, respectivamente. Tendo isto em vista, procura-se agora verificar a hipótese levantada de que, no mercado brasileiro, o VPIN é negativamente correlacionado com o tamanho da empresa.

De forma a investigar essa hipótese, procedeu-se a etapas similares às aquelas empregadas para a análise dos segmentos de listagens da Bovespa. Primeiramente calculou-se o VPIN diário das ações, posteriormente, dividiram-nas em três grupos, relacionados ao seu tamanho. O tamanho da empresa foi calculado tomando-se a média do valor de mercado diário das empresas para o período analisado. Ao fim, obtiveram-se três grupos com cerca de 47 ações cada, denominados *pequeno*, *médio* e *grande*.

O primeiro indício relacionando o VPIN com o tamanho das empresas veio com o cálculo da correlação entre esses dois fatores. O resultado para a correlação foi de -0,3080 e p.valor de 0. Tal valor era esperado, tendo em vista as constantes evidências empíricas da relação negativa entre o VPIN e o tamanho. Objetivando o aprofundamento na análise dessa relação, procedeu-se ao cálculo das estatísticas descritivas para cada grupo, como mostrado na tabela 16.

Tabela 16

Estatísticas do VPIN diário por grupo relacionado ao tamanho das empresas

Grupo	Pequeno	Médio	Grande
Mínimo	0,5582	0,3544	0,2793
Máximo	0,7134	0,4565	0,3528
Média	0,6364	0,3989	0,3164
Desvio Padrão	0,0284	0,0164	0,0123
Assimetria	-0,1216	0,1622	0,0492
Curtose	2,5049	3,1365	2,8315

Fonte - Elaborada pelo autor da dissertação.

Assim como esperado, a tabela 16 evidencia a relação verificada nos artigos que aplicaram o VPIN na comparação entre os tamanhos das ações, ou seja, quanto maior o valor de mercado da empresa, menor o seu nível de toxicidade de fluxos das ordens. Verifica-se uma grande diferença entre os VPINs das menores empresas da amostra em

relação às empresas médias e grandes. Para efeito de comparação, a tabela 17 evidencia os resultados obtidos em outros mercados em que foi aplicado o VPIN.

Tabela 17

Comparação entre os VPINs calculados por grupos de tamanho em diferentes mercados

Mercados	Pequeno	Médio	Grande	Autores
Brasil	0,6364	0,3989	0,3164	Presente Trabalho
Espanha	0,5320	0,4020	0,2540	Abad e Yagüe (2012)
Australia	0,8459	0,7249	0,5002	Wei, Gerace e Frino (2013)
Estados Unidos	0,6154	0,4001	0,2378	Yildez, Van Ness e Van Ness (2016)

Fonte - Elaborada pelo autor da dissertação.

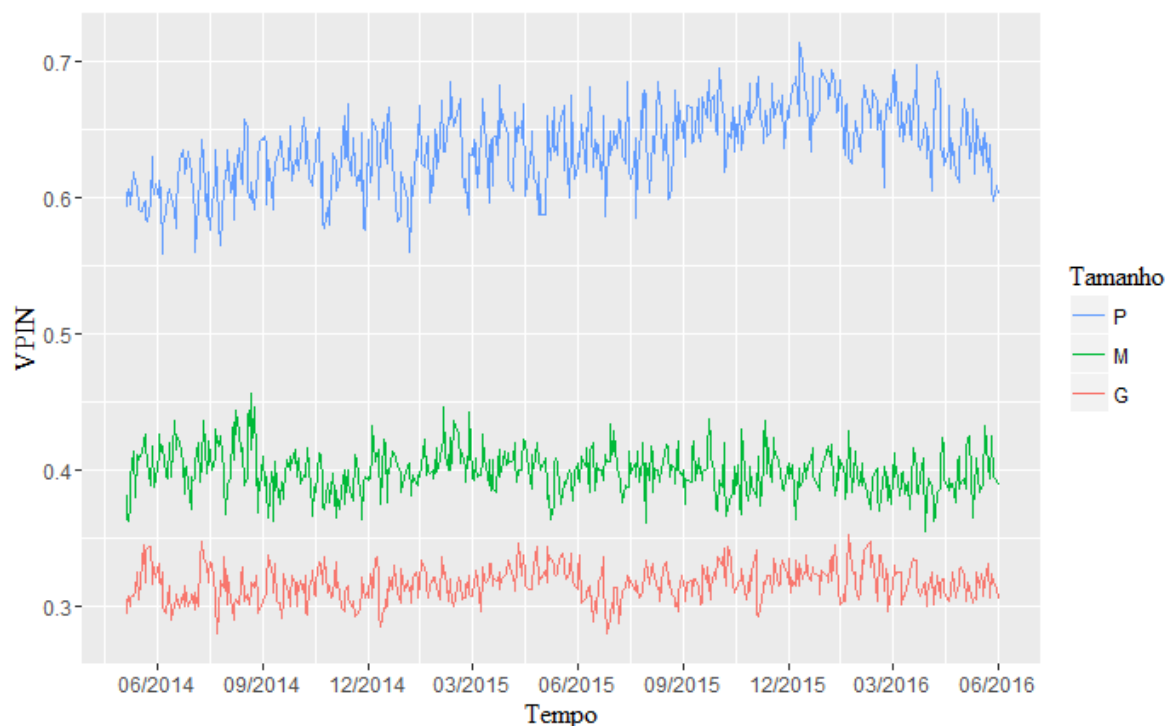
De forma geral, os resultados obtidos para as empresas médias e grandes nos mercados espanhol e americano foram próximos dos do mercado brasileiro. O VPIN das empresas menores no Brasil apresentou valor relativamente similar ao das empresas de menor tamanho dos Estados Unidos. O maior VPIN encontrado entre os mercados analisados é para as empresas pequenas do mercado australiano. Tal fato pode ser fruto da escolha da amostra por parte de Wei, Gerace e Frino (2013), tendo em vista que os autores utilizam para esse grupo as 10 ações com menor peso no índice S&P200/ASX.

As ações do grupo médio do mercado brasileiro apresentaram valores de VPIN próximos aos das ações dos grupos médios do mercado espanhol (0,402) e norte-americano (0,4001), indicando uma similaridade do nível de toxicidade para ações desse grupo. Novamente, as ações medianas australianas apresentaram valores mais elevados, assim como para seus outros grupos de tamanho. As ações de grande porte brasileiras apresentaram VPIN maior do que aquelas do mercado espanhol e norte-americano. Pode-se concluir que o VPIN das empresas brasileiras apresenta comportamento similar ao das empresas de outros mercados, com correlação negativa em relação ao tamanho das companhias. De forma geral, o VPIN calculado para o mercado nacional de 0,4548 não se diferencia tanto daqueles do mercado espanhol (0,396) e norte-americano (0,4178), sendo, no entanto, relativamente maior, como era esperado, levando em consideração o caráter emergente do mercado acionário brasileiro.

Voltando à análise das características do VPIN em relação ao tamanho das empresas, o gráfico 4 mostra o comportamento do VPIN de cada grupo. Pode-se observar, a partir do gráfico 4, a diferença substancial do nível de toxicidade das ações em relação a seus valores de mercado. O grupo *pequeno* apresentou um VPIN diário

sempre acima de 0,55, enquanto o grupo *médio* ficou em torno de 0,40, com leves picos chegando a 0,45. O grupo *grande* se manteve mais estável, com o menor desvio-padrão entre os três, apresentando um máximo de 0,35.

Gráfico 4 – Evolução do VPIN diário para os grupos de tamanho das ações



Fonte - Elaborado pelo autor da dissertação.

Nota: 'P' se refere as ações do grupo pequeno. 'M' se refere ao grupo médio e 'G' ao grupo grande.

Tendo sido constatada uma diferença substancial do VPIN das ações em relação a seus tamanhos, a explicação para os comportamentos dos VPINs para os segmentos da Bovespa pode estar contida nos valores de mercado das ações que compõem cada segmento. Das 47 empresas do grupo *pequeno*, 36 – ou seja, 75% das ações – se encontram no segmento Novo Mercado que tem 94 ações no total, fazendo com que o VPIN médio do NM aumente substancialmente. Excluindo as ações do grupo *pequeno*, o VPIN do NM cairia para 0,3418, valor significativamente menor do que o VPIN atual do segmento. Para o segmento N1, dentre as 24 ações que o compõem, uma está dentro do grupo *pequeno*, onze dentro do *médio* e doze dentro do grupo das grandes empresas, fazendo com que o VPIN seja levado para baixo, o que explicaria o fato de o N1 apresentar o menor VPIN entre os segmentos. Caso as ações do grupo *pequeno* e *médio* fossem excluídas do cálculo do VPIN do N1, seu VPIN cairia de 0,3761 para 0,2972.

Em relação ao segmento N2, das 12 ações que o compõem, seis são do grupo *pequeno*, três do grupo *médio* e três do grupo *grande*. Como o VPIN não considera a ponderação pelo tamanho da empresa e dá pesos iguais para as empresas, o fato de metade do segmento ser composto de ações do grupo *pequeno* explicaria por que o N2 apresenta o maior VPIN entre os segmentos. Já, para o segmento Tradicional, das 12 ações que o compõem, três são do grupo *pequeno*, três do grupo *médio* e seis do grupo *grande*. Dentre as ações do grupo de maior valor de mercado se encontram os seguintes ativos: ABEV3, LAME3, LAME4, PETR3, PETR4 e VIVT4. Juntas, essas ações apresentariam um VPIN médio de 0,3030, significativamente inferior aos 0,4597 apresentados pelo segmento como um todo. A razão para tal diferença reside no VPIN das ações menores, que apresentam um VPIN médio de 0,7835, fazendo com que o valor do VPIN do segmento Tradicional suba para 0,4597.

A hipótese, portanto, de que o tamanho da empresa e seu VPIN são negativamente correlacionados foi verificada a partir da amostra analisada no mercado acionário nacional. Tal evidência está em concordância com o esperado e com o apresentado na literatura internacional.

4.3 Análise da relação entre o VPIN e o *bid-ask spread* das ações

A hipótese levantada de que, quanto maior o VPIN de uma ação, menor a sua liquidez – representado pelo *bid-ask spread* – está calcada nos resultados encontrados pelos diversos estudos que relacionaram os dois conceitos. O próprio método VPIN surgiu do estudo de Easley, López de Prado e O'Hara (2011a) a respeito do *Flash Crash* que foi, em última análise uma crise de liquidez. Yildez, Van Ness e Van Ness (2016) verificam que o VPIN pode ser uma boa métrica preditiva das mudanças de liquidez entre investidores de alta frequência. Segundo Easley, López de Prado e O'Hara (2011a), mesmo em um momento de alto volume de transações, pode haver baixa liquidez. Dessa forma, a relação entre a quantidade de ações transacionadas e liquidez não é direta. Tal relação pode, no entanto, ser vista por meio da métrica do VPIN, que é calculado pelo desequilíbrio entre as ordens de transações. Easley, López de Prado e O'Hara (2011a) mostram que o VPIN alertava para sinais de crise de liquidez horas antes do *Flash Crash*, indicando, assim, a possível relação entre os dois conceitos.

Aslan *et al.* (2011) encontram uma correlação de 0,96 entre o PIN e o índice de iliquidez de Amihud (2002), indicando que, possivelmente, ambas as *proxies* representam o mesmo fenômeno. Duarte e Young (2009) separam a PIN em dois fatores: um relacionado à assimetria informacional e outro relacionado à parte de liquidez e concluem que esta última é precificada em modelos *cross-section*. Aslan *et al.* (2011) encontram resultados contrários aos de Duarte e Young (2009) e afirmam que, quando o PIN e o índice de Amihud (2002) estão presentes em uma regressão, este último se torna irrelevante para a explicação dos retornos dos ativos. Hwang *et al.* (2013), analisando a relação entre o PIN e o índice de iliquidez de Amihud (2002), constatam que existe uma associação positiva entre os dois fatores, tanto utilizando o PIN de Easley *et al.* (1996), quanto o PIN decomposto de Duarte e Young (2009).

Martins e Paulo (2013) investigam a relação da PIN com a liquidez por meio de uma regressão tobit. Os resultados encontrados pelos autores foram contrários àqueles de Duarte e Young (2009). Para a amostra analisada por Martins e Paulo (2013), o coeficiente angular foi positivo e significativo, porém baixo, indicando um grau de explicação do PIN das ações por meio de sua liquidez. Tay *et al.* (2009) afirmam que os efeitos da assimetria informacional estão ligados à liquidez do mercado. Espera-se que mercados com alto grau de assimetria informacional sejam menos líquidos e, portanto, apresentem maior *bid-ask spread*. Segundo os autores, espera-se que o PIN seja positivamente correlacionado com o *spread* entre o *bid* e o *ask* das ações.

Tendo em vista as evidências empíricas relacionando a probabilidade de negociações privilegiadas e a liquidez das ações, procedeu-se à análise desses dois fatores no mercado brasileiro em duas etapas: a primeira, mais simples, consistiu no cálculo da correlação entre o VPIN e o *bid-ask spread* diário das ações. Foi feita também a correlação do VPIN com o *bid-ask spread* considerando todo o mercado. Como o *bid-ask spread* representa a diferença entre o maior preço de oferta de compra e o menor preço de oferta de venda, tem-se que, quanto maior esse *spread*, menor a liquidez da ação para aquele momento. Posteriormente, aplicou-se a regressão apresentada na seção de Hipóteses, representada pela equação (32).

$$BidAskSpread_{it} = \alpha_i + \beta_i VPIN_{it} + \varepsilon \quad (32)$$

A abordagem apresentada por (32) não procura analisar a capacidade preditiva do VPIN. Pelo fator intradiário dessa métrica, acredita-se que tal capacidade seria mais

bem representada analisando-se o VPIN intradiário juntamente ao *bid-ask spread* na mesma frequência. Dessa forma, o que se pretende com a equação (32) é analisar se o VPIN diário das ações tem poder explicativo sobre o *bid-ask spread* das ações. Esperava-se que essa relação fosse positiva e estatisticamente significativa para as ações, indicando que parte do *bid-ask spread* é explicado pelo desequilíbrio entre as ofertas de compra e venda.

Para a correlação entre o VPIN e o *bid-ask spread*, procedeu-se ao cálculo utilizando todos os dados da amostra, ou seja, procurando uma correlação geral entre esses dois fatores. O resultado obtido foi de 0,2672 com p.valor igual a 0, o que indica uma correlação positiva e significativa como esperado. Isto significa, portanto, que, quanto maior o VPIN, maior o *bid-ask spread*, o que representa menor liquidez das ações. Parte da iliquidez das ações, segundo O'Hara (2015), advém do desequilíbrio entre o volume de ofertas de compra e venda o que, portanto, leva à consideração de que não necessariamente um alto volume transacionado indica alta liquidez da ação, haja vista que esse volume pode estar altamente desbalanceado, fator este indicado pelo VPIN. A título de comparação, o valor da correlação encontrado no presente trabalho é substancialmente menor do que aquele encontrado por Aslan *et al.* (2011), de 0,96, trabalho em que os autores utilizam o índice de Amihud (2002) como *proxy* para iliquidez e a PIN como *proxy* para probabilidade de negociações privilegiadas.

A tabela 18 mostra a correlação entre o VPIN e o *bid-ask spread* para cada ação da amostra, além do p.valor. A tabela 18 está classificada em ordem decrescente da correlação desses fatores para os ativos. Percebe-se que a maior correlação auferida foi da ação USIM5, de 0,4145. O menor valor para a correlação foi da ação OIBR4, de -0,3621.

De modo geral, 88 ações apresentaram correlação positiva. Além disso, 75 ações mostraram correlação no intervalo entre $[-0.05, 0.05]$, indicando, portanto, que, para mais da metade das ações analisadas, a correlação entre seu VPIN e seu *bid-ask spread* é estatisticamente insignificante.

Foram analisadas as correlações entre: (i) os segmentos da Bovespa e (ii) entre os diferentes grupos de tamanho das ações. Não foi constatada nenhuma relação significativa das correlações intragrupos. Isso pode ser verificado pela tabela 18 que mostra, por exemplo, que a ação PETR4 apresentou a segunda maior correlação

(0,3326) enquanto a PETR3 evidenciou uma correlação de -0,0921. Finalmente, entende-se que a relação entre o VPIN e o *bid-ask spread* dependa propriamente da ação e suas características intrínsecas, sem se ater a propriedades comuns entre elas.

O próximo passo da análise se refere à regressão representada pela equação (32). Os resultados referentes a esse procedimento estão apresentados na tabela 19. Como esperado, o coeficiente angular apresentou resultado positivo e estatisticamente significativo (0,0256 com p.valor igual a 0), indicando que existe, de forma geral, uma relação positiva entre o VPIN e o *bid-ask spread*, ou seja, quanto maior o desequilíbrio entre os volumes de ordens de transação, menor a liquidez diária das ações. O VPIN, no entanto, não se apresenta como a única variável explicativa para o *bid-ask spread* das ações, tendo em vista que o intercepto da regressão é diferente de zero ao mesmo tempo em que se rejeita a hipótese nula de que esse valor é estatisticamente igual a zero.

Para verificar a robustez do modelo, utilizaram-se os testes de Breusch-Pagan e Durbin-Watson para testar, respectivamente, se existe: (i) heteroscedasticidade na amostra e (ii) autocorrelação dos resíduos. Os resultados de ambos os testes são apresentados no painel (b) da tabela 19. Verifica-se que o teste de Breusch-Pagan rejeita a hipótese nula de homocedasticidade, enquanto o teste de Durbin-Watson rejeita a hipótese nula de ausência de autocorrelação. Este último teste apresenta estatística de teste próxima de 0, o que sugere, portanto, uma correlação serial positiva. Além do teste de Durbin-Watson, foi feita também a correlação entre os resíduos. Ela apresentou um valor de 0,7504 com p.valor igual a 0, indicando, portanto, forte correlação entre eles, o que pode afetar substancialmente a estimação dos coeficientes da regressão MQO.

Tendo sido comprovada a presença, tanto de heterocedasticidade, quanto de correlação serial no modelo testado, deixa-se clara a preocupação com a subestimação dos erros-padrão pela MQO. Assim, Newey e West (1987) propõem uma correção aplicada tanto, para a heterocedasticidade, quanto para a autocorrelação. O primeiro passo para a estimação dos erros-padrão dos coeficientes é a seleção de um *lag* para o qual é feita essa correção. Foi utilizado o procedimento sugerido por Newey e West (1994) para escolha automática do *lag* apropriado. Os resultados dos coeficientes e seus erros-padrão por Newey-West são apresentados no painel (c) da tabela 19. Verifica-se que, tanto o intercepto, quanto o coeficiente angular são os mesmos para a regressão MQO. Os erros-padrão, no entanto, são substancialmente maiores, passando de 0,00017 para 0,00122 para o intercepto e de 0,00034 para 0,00423 para o coeficiente angular.

Tabela 18

Correlação entre o VPIN e o *bid-ask spread* das ações

Ação	Correlação	P.valor	Ação	Correlação	P.valor	Ação	Correlação	P.valor	Ação	Correlação	P.valor	Ação	Correlação	P.valor
USIM5	0,4145	0,0000	ROMI3	0,0749	0,0906	BBDC3	0,0280	0,5276	CTIP3	0,0015	0,9729	OGXP3	-0,0475	0,2841
PETR4	0,3326	0,0000	BBRK3	0,0684	0,1223	LINX3	0,0256	0,5638	QGEP3	-0,0004	0,9922	BBAS3	-0,0511	0,2491
GOLL4	0,3191	0,0000	UGPA3	0,0604	0,1727	BRFS3	0,0251	0,5714	CSMG3	-0,0016	0,9710	EQTL3	-0,0515	0,2450
PRML3	0,2832	0,0000	PFRM3	0,0596	0,1785	LAME3	0,0248	0,5757	LEVE3	-0,0019	0,9654	ENBR3	-0,0533	0,2288
MMXM3	0,2276	0,0000	BRKM5	0,0596	0,1786	FIBR3	0,0242	0,5858	HGTX3	-0,0025	0,9551	CGAS5	-0,0541	0,2219
MILS3	0,1987	0,0000	BRAP4	0,0579	0,1914	STBP11	0,0209	0,6373	BRPR3	-0,0095	0,8302	GFSA3	-0,0580	0,1903
GGBR4	0,1929	0,0000	SGPS3	0,0567	0,2005	UCAS3	0,0206	0,6414	ODPV3	-0,0114	0,7969	LAME4	-0,0593	0,1805
PDGR3	0,1721	0,0000	PMAM3	0,0560	0,2065	ETER3	0,0205	0,6435	QUAL3	-0,0143	0,7477	KLBN4	-0,0600	0,1759
SLED4	0,1710	0,0001	KEPL3	0,0557	0,2085	KLBN11	0,0185	0,6770	SMT03	-0,0160	0,7187	KROT3	-0,0645	0,1457
EVEN3	0,1670	0,0001	POMO4	0,0549	0,2152	TIMP3	0,0180	0,6842	EZTC3	-0,0167	0,7072	FESA4	-0,0659	0,1371
HBOR3	0,1620	0,0002	ABEV3	0,0549	0,2153	LLIS3	0,0177	0,6901	IGTA3	-0,0187	0,6726	ANIM3	-0,0678	0,1257
BRIN3	0,1486	0,0008	ESTC3	0,0529	0,2330	BEEF3	0,0176	0,6909	SEER3	-0,0206	0,6419	MRFG3	-0,0754	0,0886
LPSB3	0,1401	0,0015	SUZB5	0,0521	0,2394	EMBR3	0,0161	0,7167	NATU3	-0,0223	0,6153	TUPY3	-0,0768	0,0828
VVAR11	0,1359	0,0021	RENT3	0,0488	0,2707	ALSC3	0,0154	0,7284	RADL3	-0,0223	0,6150	MPLU3	-0,0862	0,0514
CSNA3	0,1251	0,0046	BPFA3	0,0485	0,2742	VLID3	0,0137	0,7580	BBTG11	-0,0227	0,6090	PCAR4	-0,0886	0,0453
ECOR3	0,1227	0,0055	SMLE3	0,0477	0,2822	CESP6	0,0133	0,7641	MDIA3	-0,0228	0,6067	PSSA3	-0,0911	0,0396
VALE5	0,1212	0,0061	TPIS3	0,0475	0,2841	BBDC4	0,0114	0,7978	DTEX3	-0,0268	0,5452	PETR3	-0,0921	0,0374
JHSF3	0,1200	0,0066	WEGE3	0,0439	0,3220	BRML3	0,0111	0,8026	HYPE3	-0,0272	0,5391	CPLE6	-0,0930	0,0355
DIRR3	0,1187	0,0072	TAE11	0,0434	0,3275	VIVT4	0,0108	0,8081	SLCE3	-0,0280	0,5270	CYRE3	-0,0963	0,0295
GOAU4	0,1124	0,0110	CIEL3	0,0421	0,3427	TERI3	0,0108	0,8081	BVMF3	-0,0295	0,5062	GRND3	-0,0994	0,0246
TCSA3	0,1069	0,0156	TOTS3	0,0400	0,3675	SULA11	0,0090	0,8383	TBLE3	-0,0330	0,4573	MULT3	-0,1136	0,0102
JSLG3	0,1058	0,0167	BBSE3	0,0389	0,3803	BTOW3	0,0082	0,8534	CVCB3	-0,0351	0,4291	TECN3	-0,1364	0,0020
ELPL4	0,1012	0,0221	POS3	0,0387	0,3826	CMIG4	0,0081	0,8543	LREN3	-0,0367	0,4079	SBSP3	-0,1393	0,0016
ITSA4	0,0998	0,0240	CLSC4	0,0379	0,3927	OIBR3	0,0081	0,8544	ARZZ3	-0,0381	0,3899	TRPL4	-0,1983	0,0000
CARD3	0,0960	0,0300	MAGG3	0,0354	0,4243	VALE3	0,0063	0,8874	ABCBA	-0,0391	0,3776	RSID3	-0,2426	0,0000
TGMA3	0,0958	0,0304	JBSS3	0,0345	0,4359	CCRO3	0,0062	0,8889	CSAN3	-0,0403	0,3627	OIBR4	-0,3621	0,0000
LOGN3	0,0907	0,0404	CPFE3	0,0338	0,4452	ALPA4	0,0050	0,9109	BRSR6	-0,0406	0,3598			
MRVE3	0,0867	0,0502	ELET6	0,0330	0,4563	ELET3	0,0034	0,9393	LIGT3	-0,0420	0,3436			
MGLU3	0,0781	0,0779	ITUB4	0,0293	0,5083	FLRY3	0,0028	0,9490	MYPK3	-0,0441	0,3197			

Fonte - Elaborada pelo autor da dissertação.

Como supracitado, na presença de heterocedasticidade e autocorrelação, é esperada uma subestimação dos erros-padrão por parte do MQO. Dessa forma, aplicando-se a correção de Newey-West, esses valores se apresentam maiores. O aumento dos erros-padrão fez com que os p.valores para o teste de hipótese acerca dos valores dos coeficientes se distanciassem de zero, porém não a ponto de não rejeitar a hipótese nula de que os coeficientes são estatisticamente iguais a zero. Assim, mantém-se a conclusão apresentada pela regressão MQO de que, tanto o intercepto, quanto o coeficiente angular são diferentes de zero e que, portanto, o VPIN tem um certo poder de explicação sobre o *bid-ask spread* diário das ações.

Tabela 19

Regressão MQO do *bid-ask spread* sobre o VPIN, testes para heterocedasticidade e autocorrelação e estimação dos coeficientes por Newey-West

Painel (a): Regressão MQO do <i>bid-ask spread</i> sobre o VPIN				
	Estimação	Erros Padrão	t.value	P.valor
Intercepto	-0,0033	0,00017	-19,65	0,0000
Coefficiente Angular	0,0256	0,00034	74,69	0,0000
Painel (b): Testes de Breusch-Pagan e de Durbin-Watson				
	Estatística de Teste	P.valor		
Breusch-Pagan	167,77	0,0000		
Durbin-Watson	0,4991	0,0000		
Painel (c): Estimação dos coeficientes por Newey-West				
	Estimação	Erros Padrão	t.value	P.valor
Intercepto	-0,0033	0,00122	-2,7554	0,0058
Coefficiente Angular	0,0256	0,00423	6,0598	0,0000

Fonte - Elaborada pelo autor da dissertação.

Nota: O painel (a) se refere aos resultados obtidos através da regressão MQO com os dados empilhados; o painel (b) se refere aos resultados para os testes de heterocedasticidade (Breusch-Pagan) e autocorrelação (Durbin-Watson); o painel (c) se refere aos resultados para os coeficientes obtidos pela correção proposta por Newey-West com *lag* automático calculado através do procedimento estabelecido por Newey e West (1994).

O próximo passo foi a aplicação da regressão (32) para cada ação, de forma a verificar o grau de explicação do VPIN para o *bid-ask spread* individualmente. A tabela 20 apresenta os resultados. Das 142 ações, apenas 13 delas apresentaram um p.valor para seu intercepto maior que um alfa de 5%, indicando que somente para essas ações, a hipótese nula de que o valor do intercepto é zero não é rejeitada. Dessa forma, 129 ações apresentaram intercepto positivo e estatisticamente diferente de zero, o que indica que o VPIN não foi suficiente para explicar o *bid-ask spread* dessas ações. Tal valor era esperado, tendo em vista que o desequilíbrio das ordens é um dentre outros fatores que

formam a liquidez da ação. Dentre os interceptos que apresentaram valores negativos (sete ações), seis delas apresentam p.valor maior do que 0,05, indicando a não rejeição da hipótese nula, ou seja, aceita-se que tais valores, apesar de negativos, não são estatisticamente diferentes de zero.

Em relação ao coeficiente do VPIN, o maior valor estimado foi para a ação PDGR3 (0,0455 e p.valor de 0,001) e o menor para a ação TUPY3 (-0,0224 e p.valor de 0,0828). Os resultados para os p.valores dos coeficientes das ações indicam a aceitação da hipótese nula de que os coeficientes são iguais a zero para 103 das 142 ações analisadas. Naturalmente, tais ações apresentaram interceptos estatisticamente significantes. Portanto, para 103 ações, o VPIN não tem poder de explicação sobre seu *bid-ask spread*. As demais ações, que não tiveram a hipótese nula acerca de seus coeficientes aceita, são aquelas que apresentaram maior correlação entre o VPIN e o *bid-ask spread*, o que já era esperado.

Acerca do coeficiente de determinação, R^2 , o maior valor apresentado foi para a ação USIM5 (0,1718), seguido da ação OIBR4 (0,1311), PETR4 (0,1106), GOLL4 (0,1018) e PRML3 (0,0802), que foram aquelas ações que mostraram maior correlação entre o VPIN e o *bid-ask spread*. Apesar de não ser um coeficiente de determinação alto, este confirma que, para as ações que apresentaram maior valor de correlação entre os dois fatores analisados, o VPIN tem algum grau de explicação sobre o *bid-ask spread*.

De forma geral, não existe um padrão a ser seguido para a relação entre o VPIN e o *bid-ask spread*. Esperava-se que existisse uma relação positiva e estatisticamente significativa entre os dois para todas as ações, o que só foi constatado para aquelas ações que apresentaram uma correlação significativa para os dois fatores. Conclui-se que a relação entre o desequilíbrio dos volumes e o *spread* entre os preços de *bid* e *ask* sigam as características intrínsecas de cada ação, sem obedecer uma regra geral para o mercado. Optou-se por não testar outras *proxies* para a liquidez (ou iliquidez) das ações, tendo em vista que o *bid-ask spread* é, para a literatura, uma *proxy* robusta para esse fator por indicar diretamente qual o grau de liquidez de uma ação – quanto mais próximo de zero, mais líquida é a ação, pois indica que existem compradores e vendedores ofertando o ativo a um preço similar, mostrando, portanto, a facilidade de executar as ordens de compra e venda para aquela ação naquele momento (O'HARA, 2003).

Tabela 20

Estimação do Intercepto e do Coeficiente Angular da regressão $BidAsk_{it} = \alpha + \beta VPIN_{it} + \varepsilon$, para cada ação, onde i se refere ao indexador do ativo i e t se refere aos dias, $t = 1, \dots, 512$.

Ação	PETR4	USIM5	GOLL4	GGBR4	GOAU4	EVEN3	VALE5	ITSA4	PRML3	MILS3	CSNA3	ECOR3	SLED4	MRVE3	HBOR3
Intercepto	0,2306 (0)	0,3846 (0)	0,4105 (0)	0,2953 (0)	0,4769 (0)	0,4006 (0)	0,2440 (0)	0,2816 (0)	0,7223 (0)	0,4827 (0)	0,3250 (0)	0,3394 (0)	0,6901 (0)	0,3069 (0)	0,5606 (0)
Coef. Ang.	17,3434 (0)	16,9407 (0)	9,4301 (0)	6,0817 (0)	5,1047 (0,011)	4,7242 (0,0001)	4,6853 (0,0061)	3,9076 (0,024)	3,3744 (0)	3,2266 (0)	2,7083 (0,0046)	2,6125 (0,0055)	2,4625 (0,0001)	2,4598 (0,0502)	2,3829 (0,0002)
Ação	ABEV3	ELPL4	LPSB3	VVAR11	TCSA3	UGPA3	CIEL3	SUZB5	JHSF3	DIRR3	BRKM5	POMO4	TGMA3	BBSE3	JSLG3
Intercepto	0,2792 (0)	0,4328 (0)	0,5817 (0)	0,4472 (0)	0,5732 (0)	0,2846 (0)	0,3142 (0)	0,3080 (0)	0,7096 (0)	0,4905 (0)	0,2958 (0)	0,4518 (0)	0,5509 (0)	0,3273 (0)	0,6588 (0)
Coef. Ang.	2,2035 (0,2153)	2,1890 (0,0221)	2,1820 (0,0015)	2,1748 (0,0021)	2,1322 (0,0156)	2,0499 (0,1727)	2,0237 (0,3427)	1,9015 (0,2394)	1,8952 (0,0066)	1,8007 (0,0072)	1,6547 (0,1786)	1,6484 (0,2152)	1,6073 (0,0304)	1,5568 (0,3803)	1,5202 (0,0167)
Ação	RENT3	BRAP4	BRIN3	MMXM3	MGLU3	ESTC3	SMLE3	ITUB4	PFRM3	TAE11	WEGE3	JBSS3	LOGN3	MAGG3	BRFS3
Intercepto	0,2939 (0)	0,3382 (0)	0,6519 (0)	0,7644 (0)	0,5010 (0)	0,3053 (0)	0,3064 (0)	0,2536 (0)	0,6803 (0)	0,3630 (0)	0,3198 (0)	0,2923 (0)	0,7290 (0)	0,6716 (0)	0,2895 (0)
Coef. Ang.	1,3814 (0,2707)	1,3105 (0,1914)	1,2620 (0,0008)	1,2509 (0)	1,2035 (0,0779)	1,1843 (0,233)	1,1371 (0,2822)	1,1333 (0,5083)	1,1317 (0,1785)	1,1263 (0,3275)	1,1188 (0,322)	1,1010 (0,4359)	0,9749 (0,0404)	0,9335 (0,4243)	0,9138 (0,5714)
Ação	FIBR3	BBRK3	ELET6	PMAM3	TOTS3	BPHA3	CPFE3	KLBN11	BBDC3	POSI3	CARD3	ROMI3	PDGR3	LINX3	TPIS3
Intercepto	0,2797 (0)	0,5845 (0)	0,3773 (0)	0,6902 (0)	0,2978 (0)	0,7082 (0)	0,2747 (0)	0,3653 (0)	0,3307 (0)	0,6631 (0)	0,8620 (0)	0,7803 (0)	0,6990 (0)	0,5243 (0)	0,7064 (0)
Coef. Ang.	0,8840 (0,5858)	0,8738 (0,1223)	0,8658 (0,4563)	0,8580 (0,2065)	0,8388 (0,3675)	0,8111 (0,2742)	0,7454 (0,4452)	0,7266 (0,677)	0,7147 (0,5276)	0,7023 (0,3826)	0,6879 (0,03)	0,6670 (0,0906)	0,6508 (0,0001)	0,6355 (0,5638)	0,6210 (0,2841)
Ação	CLSC4	LAME3	TIMP3	SGPS3	BEEF3	EMBR3	LLIS3	ALSC3	BBDC4	STBP11	VIVT4	BRML3	CESP6	ETER3	CMIG4
Intercepto	0,7573 (0)	0,4383 (0)	0,2859 (0)	0,8543 (0)	0,4140 (0)	0,2906 (0)	0,5863 (0)	0,4719 (0)	0,2494 (0)	0,5878 (0)	0,2988 (0)	0,3223 (0)	0,3115 (0)	0,7898 (0)	0,3251 (0)
Coef. Ang.	0,5968 (0,3927)	0,5770 (0,5757)	0,5185 (0,6842)	0,4839 (0,2005)	0,4297 (0,6909)	0,4220 (0,7167)	0,4123 (0,6901)	0,3998 (0,7284)	0,3832 (0,7978)	0,3730 (0,6373)	0,3590 (0,8081)	0,3082 (0,8026)	0,3026 (0,7641)	0,2974 (0,6435)	0,2911 (0,8543)

Ação	UCAS3	VLID3	BTOW3	SULA11	ALPA4	TERI3	OIBR3	CCRO3	VALE3	KEPL3	ELET3	CTIP3	FLRY3	QGEP3	CSMG3
Intercepto	0,7479 (0)	0,3494 (0)	0,4179 (0)	0,3280 (0)	0,5260 (0)	0,7874 (0)	0,5775 (0)	0,2837 (0)	0,2562 (0)	0,7749 (0)	0,4165 (0)	0,3387 (0)	0,4889 (0)	0,4521 (0)	0,4016 (0)
Coef. Ang.	0,2486 (0,6414)	0,2183 (0,758)	0,1664 (0,8534)	0,1575 (0,8383)	0,1503 (0,9109)	0,1477 (0,8081)	0,1464 (0,8544)	0,1323 (0,8889)	0,1147 (0,8874)	0,1142 (0,2085)	0,1130 (0,9393)	0,0654 (0,9729)	0,0589 (0,949)	-0,0052 (0,9922)	-0,0244 (0,971)
Ação	LEVE3	HGTX3	CVCB3	OGXP3	TUPY3	BRPR3	SEER3	SMT03	ODPV3	IGTA3	QUAL3	EZTC3	RADL3	MDIA3	NATU3
Intercepto	0,4768 (0)	0,3833 (0)	0,4940 (0)	0,7375 (0)	0,5784 (0)	0,4186 (0)	0,4383 (0)	0,4009 (0)	0,4078 (0)	0,3716 (0)	0,3030 (0)	0,3897 (0)	0,2931 (0)	0,4692 (0)	0,3259 (0)
Coef. Ang.	-0,0470 (0,9654)	-0,0630 (0,9551)	-0,1103 (0,4291)	-0,1254 (0,2841)	-0,2631 (0,0828)	-0,2647 (0,8302)	-0,2896 (0,6419)	-0,3100 (0,7187)	-0,3173 (0,7969)	-0,3424 (0,6726)	-0,4413 (0,7477)	-0,4442 (0,7072)	-0,5309 (0,615)	-0,5442 (0,6067)	-0,5521 (0,6153)
Ação	SLCE3	LIGT3	DTEX3	ABCB4	ANIM3	BBTG11	ARZZ3	BRSR6	TBLE3	FESA4	HYPE3	ENBR3	MYPK3	CSAN3	BVMF3
Intercepto	0,4894 (0)	0,3502 (0)	0,3776 (0)	0,5422 (0)	0,4370 (0)	0,5540 (0)	0,5421 (0)	0,3745 (0)	0,2888 (0)	0,7553 (0)	0,2957 (0)	0,2874 (0)	0,4210 (0)	0,3309 (0)	0,3283 (0)
Coef. Ang.	-0,5608 (0,527)	-0,5936 (0,3436)	-0,7009 (0,5452)	-0,7806 (0,3776)	-0,8103 (0,1257)	-0,8163 (0,609)	-0,8878 (0,3899)	-0,8950 (0,3598)	-0,9080 (0,4573)	-0,9109 (0,1371)	-0,9328 (0,5391)	-0,9824 (0,2288)	-1,0351 (0,3197)	-1,2162 (0,3627)	-1,3237 (0,5062)
Ação	CGAS5	LREN3	TECN3	PSSA3	LAME4	BBAS3	PETR3	GFSA3	KLBN4	MRFG3	CPLE6	EQTL3	MPLU3	KROT3	PCAR4
Intercepto	0,6712 (0)	0,3371 (0)	0,7993 (0)	0,3136 (0)	0,3191 (0)	0,2672 (0)	0,2306 (0)	0,5048 (0)	0,6619 (0)	0,3917 (0)	0,3107 (0)	0,3699 (0)	0,3699 (0)	0,3378 (0)	0,3069 (0)
Coef. Ang.	-1,3995 (0,2219)	-1,4167 (0,4079)	-1,4455 (0,002)	-1,6254 (0,0396)	-1,6768 (0,1805)	-1,7379 (0,2491)	-1,7380 (0,0374)	-1,7633 (0,1903)	-1,9275 (0,1759)	-1,9470 (0,0886)	-1,9852 (0,0355)	-2,1399 (0,245)	-2,3261 (0,0514)	-2,5321 (0,1457)	-2,8372 (0,0453)
Ação	MULT3	CYRE3	SBSP3	GRND3	RSID3	TRPL4	OIBR4								
Intercepto	0,3054 (0)	0,3568 (0)	0,3361 (0)	0,5038 (0)	0,6925 (0)	0,4038 (0)	0,6285 (0)								
Coef. Ang.	-3,0141 (0,0102)	-3,0186 (0,0295)	-3,5519 (0,0016)	-4,0014 (0,0246)	-4,4678 (0)	-4,9992 (0)	-15,164 (0)								

Fonte - Elaborada pelo autor da dissertação.

Nota: os valores entre parênteses se referem aos p.valores do Intercepto e do Coeficiente Angular, respectivamente.

4.4 Análise dos modelos de três, quatro e cinco fatores e o fator IMU

O objetivo principal deste trabalho foi analisar se um fator relacionado ao risco informacional das ações ajuda a explicar os retornos de portfólios formados com base em diferentes variáveis (tamanho, *book-to-market*, momento, lucratividade, investimento e VPIN). Easley, Hvidkjaer e O'Hara (2005), seguindo Easley, Hvidkjaer e O'Hara (2002), verificaram que um fator relacionado ao PIN reduziu os interceptos de portfólios montados com base no tamanho e no PIN das empresas, tanto utilizando o modelo de três fatores de Fama e French (1993), quanto o modelo de quatro fatores de Carhart (1997).

Hwang *et al.* (2013), usando dados de transações reais que não necessitam de algoritmos de classificação encontraram evidências empíricas da relação entre o risco informacional e os retornos esperados. Os autores mostram que erros relacionados ao método de Lee-Ready (1991) podem ter produzido inferências incorretas em outros estudos sobre a PIN. É importante ressaltar que os autores, utilizando o procedimento de Duarte e Young (2009), separam a parte relacionada à liquidez da parte informacional da PIN, e viram que esse segundo componente funciona como uma melhor *proxy* para o risco informacional.

Contrariamente aos achados de Easley, Hvidkjaer e O'Hara (2005), Mohanram e Rajgopal (2009) replicam os estudos de Easley, Hvidkjaer e O'Hara (2002) e Easley, Hvidkjaer e O'Hara (2005) e concluíram que, para o primeiro, o resultado só é robusto para um subperíodo analisado pelos autores, não sendo, portanto, o fator PIN precificável para toda a janela temporal estudada (1984 – 2002). Utilizando os modelos de três e quatro fatores, Mohanram e Rajgopal (2009) não encontram um coeficiente estatisticamente significativo para o fator baseado no PIN. Finalmente, no Brasil, Martins e Paulo (2014) encontraram uma relação positiva da PIN com o custo de capital e retorno das ações. No entanto, esses autores não se valeram dos dados reais da quantidade de ordens de transação.

Mais recentemente, Fama e French (2015) estenderam o modelo de três fatores, adicionando um fator relacionado à lucratividade das empresas e outro ao investimento. Os resultados apresentados pelos autores mostram uma melhora no desempenho desse modelo em relação ao anterior representado pelo teste GRS – que testa conjuntamente se os interceptos de diferentes portfólios não são estatisticamente diferentes de zero.

Tendo em vista as divergências empíricas de outros estudos, apresentam-se agora os resultados obtidos para os testes de três, quatro e cinco fatores, utilizando ainda um fator adicional referente ao risco informacional criado a partir do VPIN das ações utilizando os dados reais das transações.

4.4.1 Estatísticas descritivas dos fatores

O primeiro passo para analisar o desempenho dos modelos se refere às estatísticas descritivas dos fatores, o que inclui também estudar a correlação entre eles. O painel (a) da tabela 21 apresenta o sumário dos três fatores de Fama e French (1993), o fator de momento de Carhart (1997) e o fator IMU. Como o modelo de quatro fatores se vale dos três fatores de Fama e French (1993), não se fez necessária a separação em dois painéis para a análise dos resultados.

Verifica-se que, assim como em Easley, Hvidkjaer e O'Hara (2005) e Mohanram e Rajgopal (2009), o fator UMD foi aquele que apresentou a maior média dos retornos. A diferença principal é o valor auferido por esse fator. Nos trabalhos supracitados, a média encontrada para o UMD foi de 0,99%, enquanto, neste estudo, encontrou-se uma média de 0,05%, valor este significativamente menor. A segunda maior média entre os fatores encontrados neste trabalho foi o fator relacionado ao risco informacional, IMU. Foi verificada uma correlação positiva, porém baixa entre o VPIN e os retornos da ação (0,0140 com p.valor igual a 0,0003). Isto indica que ações com maior VPIN tendem a apresentar retornos maiores, o que explica o desempenho do IMU superior em relação aos fatores de mercado (MKT), tamanho (SMB) e *book-to-market* das empresas (HML). Easley, Hvidkjaer e O'Hara (2005) e Mohanram e Rajgopal (2009) também constatarem que o fator relacionado ao PIN apresenta retorno superior ao de tamanho e *book-to-market*, porém não o de mercado.

Em relação ao fator SMB, este apresentou a menor média para os retornos entre os fatores. Tal fato poderia ser explicado pela correlação entre o tamanho e os retornos das empresas (-0,0105 com p-valor 0,0068). Tal correlação mostra que, de forma geral, empresas menores podem gerar maiores retornos. No entanto, pela magnitude de essa correlação não ser substancialmente alta, muitas vezes, ações maiores podem apresentar retornos mais altos – o que de fato acontece na amostra. Ações presentes no grupo *grande*, muitas vezes, apresentaram retornos superiores àqueles das empresas

pertencentes ao grupo *pequeno*. Vale notar que Easley, Hvidkjaer e O'Hara (2005) e Mohanram e Rajgopal (2009) viram que o fator SMB também apresentou a menor média de retornos na amostra por eles analisada – -0,0003 e -0,0072, respectivamente – o que contraria os resultados de Fama e French (1992) de que os retornos das ações do grupo *small* são superiores aqueles do grupo *big*, chamado de *size effect*.

Em relação às demais estatísticas descritivas, o valor mínimo apresentado pelos fatores não variou significativamente, ficando entre -0,0592 (HML) e -0,0413 (IMU). Analisando os máximos auferidos pelos fatores, o mais baixo foi 0,0365 (UMD) e o mais alto o fator de mercado (0,0654). Os demais momentos estatísticos não apresentaram grande diferença entre os fatores e o desvio-padrão se manteve em cerca de 0,01. Assim como para Easley, Hvidkjaer e O'Hara (2005) e Mohanram e Rajgopal (2009), o fator relacionado ao risco informacional apresentou o menor valor para o segundo momento estatístico (0,0092), enquanto o MKT obteve o maior (0,0167). Em relação aos terceiros e quartos momentos, todos os fatores apresentaram assimetria positiva (exceto pelo UMD) e distribuições leptocúrticas. Em resultados não reportados, o teste de Jarque-Bera rejeitou a hipótese de normalidade para os retornos de todos os fatores.

Os resultados relacionados aos cinco fatores de Fama e French (2015) apresentados no painel (b) da tabela 21 mostram que os fatores SMB e HML se comportam de forma semelhante àqueles formados pela metodologia de Fama e French (1993). Os dois fatores adicionais de Fama e French (2015) apresentaram resultados para as médias de retorno superiores aos demais. O fator relacionado à lucratividade (RMW) auferiu o maior retorno médio, seguido pelo IMU e, posteriormente, pelo CMA e MKT com os mesmos valores. Esperava-se que a lucratividade se relacionasse com os retornos das ações das empresas. Dessa forma, não surpreende o desempenho dos retornos desse fator. Contrariamente a esses resultados, Fama e French (2015) verificam que o fator RMW apresenta o menor retorno médio (para duas das três metodologias de formação dos fatores). O maior retorno reportado pelos autores se refere ao fator de mercado, seguido pelos seguintes fatores: HML, CMA, SMB e RMW. Percebe-se, portanto, a diferença entre os resultados encontrados.

Tabela 21

Estatísticas descritivas dos três, quatro e cinco fatores de Fama e French (1993, 2015) e Carhart (1997) e o fator IMU

Painel (a): Estatísticas descritivas dos três fatores de Fama e French (1993), o fator de momento de Carhart (1997) e o fator IMU						
	MKT	SMB	HML	UMD	IMU	
Média	-0,0005	-0,0013	-0,0007	0,0005	-0,0002	
Desvio Padrão	0,0167	0,0121	0,0144	0,0101	0,0092	
Mínimo	-0,0492	-0,0480	-0,0592	-0,0416	-0,0413	
Máximo	0,0654	0,0559	0,0635	0,0365	0,0518	
Assimetria	0,3963	0,0535	0,2299	-0,2342	0,1223	
Curtose	3,6963	5,1985	5,3045	4,8185	6,0164	

Painel (b): Estatísticas descritivas dos cinco fatores de Fama e French (2015) e o fator IMU						
	MKT	SMB	HML	RMW	CMA	IMU
Média	-0,0005	-0,0014	-0,0007	-0,0001	-0,0005	-0,0002
Desvio Padrão	0,0167	0,0116	0,0144	0,0098	0,0104	0,0092
Mínimo	-0,0492	-0,0482	-0,0592	-0,0549	-0,0366	-0,0413
Máximo	0,0654	0,0574	0,0635	0,0477	0,0570	0,0518
Assimetria	0,3963	0,1194	0,2299	-0,1659	0,6032	0,1223
Curtose	3,6963	5,7775	5,3045	6,6668	6,6209	6,0164

Fonte - Elaborada pelo autor da dissertação.

Nota: o painel (a) se refere às estatísticas descritivas dos três fatores de Fama e French (1993), o fator de momento de Carhart (1997) e o fator IMU. Como o modelo de quatro fatores de Carhart (1997) utiliza os três fatores de Fama e French (1993) não se fez necessário a análise em separado dos fatores. O painel (b) se refere às estatísticas descritivas dos cinco fatores de Fama e French (2015) e o fator IMU. Dado que os fatores foram construídos com base metodologia de Fama e French (2015) os resultados são diferentes dos fatores construídos com base em Fama e French (1993), apontando, portanto a necessidade de se fazer uma análise separadamente.

No que se refere à correlação entre os fatores apresentados na tabela 22, de forma geral, os resultados – exceto para o fator de mercado – apresentam diferenças substanciais daqueles auferidos por Easley, Hvidkjaer e O’Hara (2005) e Mohanram e Rajgopal (2009). Para esses autores, o fator relacionado ao risco informacional apresentou correlação fraca, porém negativa com o SMB, fraca e positiva com o HML – para Mohanram e Rajgopal (2009) – e moderada e negativa para Easley, Hvidkjaer e O’Hara (2005) e forte e positiva com o UMD. Percebe-se, pelo painel (a) da tabela 22, que o IMU apresentou correlação moderada e negativa com o fator de mercado, sendo esse resultado próximo daquele reportado pelos autores supracitados. Em relação ao fatores HML e UMD, o IMU não apresentou correlação significativa, com p.valores de 0,1643 e 0,9638, o que sinaliza que, estatisticamente, esses fatores não são correlacionados com o IMU.

Verifica-se que a correlação de maior peso para o fator construído pelo VPIN é com o SMB. Este último é construído tomando-se os retornos das ações menores e subtraindo com os retornos das maiores ações. Foi constatado que o VPIN tem uma correlação negativa com o tamanho das ações (-0,3080), o que indica que, quanto maior a ação, menor seu VPIN. A construção do fator IMU se deu tomando-se o retorno das ações com maior VPIN e subtraindo com os retornos das ações com menor VPIN. É possível, portanto, que os retornos das ações com maior VPIN sejam daquelas empresas menores, enquanto as ações com menor VPIN representam as ações maiores, sendo esta, portanto, a explicação para a forte correlação entre os dois fatores. Será demonstrado mais a frente como essas duas variáveis atuam na explicação dos retornos dos portfólios construídos.

É válido notar que esse resultado era esperado, pois, como constatado em diversos estudos e reafirmado neste trabalho, a probabilidade de negociações privilegiadas é negativamente correlacionada com o tamanho das empresas. No entanto, Hvidkjaer e O'Hara (2005) e Mohanram e Rajgopal (2009) apresentam resultados contraditórios nesse ponto, tendo em vista que o fator construído para o PIN das ações apresenta correlação negativa com o SMB em ambas as pesquisas de -0,10 e -0,09 respectivamente. Mohanram e Rajgopal (2009) atentam para essa incongruência e afirmam que esse resultado é devido à metodologia da formação do fator PIN.

Quando analisadas as correlações entre os cinco fatores de Fama e French (2015) e o IMU, verifica-se que, mesmo em diferentes metodologias, o IMU apresenta uma correlação ligeiramente maior com o fator tamanho. Outra correlação moderada e positiva surge entre o IMU e o CMA. As demais correlações se mantêm semelhantes, exceto entre o SMB e HML, que passa de negativo para os três fatores de Fama e French (1993) para positivo e fraco. De forma geral, os resultados não se assemelham àqueles apresentados por Fama e French (2015), exceto para a correlação positiva entre os fatores CMA e HML. Novamente, os fatores MKT e o SMB apresentam correlação moderada e negativa, enquanto, para os autores citados, esta apresenta uma correlação com a mesma proporção, porém positiva.

Tabela 22

Correlação entre os três, quatro e cinco fatores de Fama e French (1993, 2015) e Carhart (1997) e o fator IMU

Painel (a): Correlação entre os três fatores de Fama e French (1993), o fator de momento de Carhart (1997) e o fator IMU

	MKT	SMB	HML	UMD	IMU
MKT	1				
SMB	-0,4463	1			
HML	0,3806	-0,1140	1		
UMD	-0,2801	-0,0110	-0,5750	1	
IMU	-0,2892	0,5633	0,0614	-0,0020	1

Painel (b): Correlação entre os cinco fatores de Fama e French (2015) e o fator IMU

	MKT	SMB	HML	RMW	CMA	IMU
MKT	1					
SMB	-0,3334	1				
HML	0,3806	0,0398	1			
RMW	0,0236	-0,0104	-0,4864	1		
CMA	0,0885	0,3766	0,3489	0,0220	1	
IMU	-0,2892	0,5650	0,0614	-0,1106	0,2861	1

Fonte - Elaborada pelo autor da dissertação.

Nota: o painel (a) se refere às correlações entre os três fatores de Fama e French (1993), o fator de momento de Carhart (1997) e o fator IMU. Como o modelo de quatro fatores de Carhart (1997) utiliza os três fatores de Fama e French (1993) não se fez necessário a análise em separado dos fatores. O painel (b) se refere às correlações entre os cinco fatores de Fama e French (2015) e o fator IMU. Dado que os fatores foram construídos com base metodologia de Fama e French (2015) os resultados são diferentes dos fatores construídos com base em Fama e French (1993), apontando, portanto a necessidade de se fazer uma análise separadamente.

4.4.2 Características dos portfólios

Nesta seção analisam-se as características dos portfólios criados para serem testados por meio dos modelos apresentados na seção de metodologia. Essas carteiras foram criadas com base nos fatores tamanho, *book-to-market*, lucratividade, investimento e VPIN, tendo como fator constante o tamanho. A tabela 3 apresenta as variáveis, segundo as quais, foram formados os portfólios, a quantidade de carteiras por combinação de variáveis, as quantidades mínima e máxima de ações e a média de ações por combinação. Foram criadas no total 62 carteiras para serem utilizadas como variáveis dependentes nas regressões testadas. Duas carteiras criadas a partir do tamanho e do VPIN das ações foram excluídas por não apresentarem o mínimo de ações necessárias para uma diversificação eficiente.

A tabela 23 mostra as médias para os fatores formadores das carteiras, o número de ações para cada uma delas, além de seu retorno médio. A nomenclatura de cada carteira se refere às variáveis que a compõem. A primeira letra se refere sempre ao tamanho, por ser a variável constante, a segunda letra se refere a outra variável, podendo ser o *book-to-market*, lucratividade, investimento ou VPIN. No painel (c) tem-se também uma terceira variável, representando as carteiras formadas com base no tamanho, no investimento e na lucratividade das variáveis. Percebe-se que o número de ações que compõem as carteiras variam, assim como reportado na tabela 3. O painel (d) mostra que não há ações na interseção entre ativos do grupo *pequeno* em relação ao tamanho e *baixo* em relação ao VPIN. Tal fato já era esperado, já que ações do grupo *pequeno* têm, em sua maioria, VPINs maiores. Além disso, a interseção entre as ações *grande* em tamanho e *alta* em VPIN só resultou em uma só ação. Este é outro fato relatado na relação entre VPIN e tamanho, já que ações grandes apresentam, geralmente, VPINs baixos. Dessa forma, explica-se a exclusão das carteiras feitas com a interseção entre S-L e B-H.

A maior combinação possível foi entre o tamanho e a lucratividade, que resultou em 16 carteiras com um mínimo de seis ações e máximo de 15. Como um todo, a carteira com maior quantidade de ações foi aquela com resultado entre a interseção das ações *grande* em tamanho e *pequeno* em VPIN, possivelmente pela evidência empírica de que ações grandes apresentam menores probabilidades de negociações privilegiadas.

Em comparação com os demais estudos que aplicaram o modelo de Fama e French (1993) no Brasil, o número de ações que compõem os portfólios presentes neste trabalho são similares àqueles utilizados por Argolo, Leal e Almeida (2012) estudo no qual a quantidade de ações dentre os portfólios varia entre 4 e 37. Rogers e Securato (2009), aplicando o modelo de Fama e French (1993), construíram carteiras que continham entre oito e 11 ações. De forma a constituir 25 portfólios para as variáveis tamanho e *book-to-market*, os autores tiveram que reduzir a amostra para esse número de ações o que diminuiu a possibilidade de diversificação dentro das carteiras.

Securato e Málaga (2004) aplicam o modelo de três fatores de forma semelhante à qual foi usada neste trabalho. Os autores dividem em três grupos para o tamanho e *book-to-market*, formando assim nove portfólios. Dentre estes portfólios, o número de ações varia entre 3 e 30. Aplicando o modelo de quatro fatores, Mussa, Famá e dos Santos (2012) formam 12 portfólios diferentes, com o número médio de ações variando de cinco a 18. Lucena e

Pinto (2008), utilizando uma metodologia diferenciada para a criação dos portfólios, constituem 25 carteiras diferentes com o número médio de ações variando de 1 a 35, sendo este o único trabalho a analisar o resultado com carteiras formadas apenas por um ativo.

De forma geral percebe-se que a quantidade de ativos a comporem os fatores utilizados pelos modelos de Fama e French (1993, 2015) é um problema inerente a qualquer estudo que se valha desse método de cálculo do retorno requerido no mercado brasileiro. Analisando comparativamente os demais estudos feitos no Brasil, acredita-se que a quantidade de ações esteja dentro da média para a avaliação do modelo de três e cinco fatores de Fama e French (1993, 2015).

Em relação ao desempenho dessas carteiras, foram formadas duas tabelas, sendo a tabela 23 com o objetivo de se analisar a evolução dos fatores que formam os portfólios e o número de ações que os compõem e a tabela 24 apresentando somente os retornos das carteiras seguindo uma formatação similar à de Fama e French (1993, 2015).

De forma geral, analisando os painéis da tabela 23, o retorno das carteiras tipicamente aumenta na medida em que o tamanho da empresa também aumenta, ou seja, o chamado *size effect* não é observado para as carteiras criadas. Esse resultado contraria o reportado por Fama e French (1993, 2015), já que se esperava que ações menores apresentassem maior retorno médio.

Argolo, Leal e Almeida (2012) afirmam que isso se apresenta como uma dificuldade na utilização do modelo de três fatores no Brasil. Rogers e Securato (2009) encontram resultados semelhantes nos 25 portfólios construídos com base no tamanho (5) e *book-to-market* (5). Para esses autores, os retornos das ações são maiores à medida que aumentam os tamanhos das empresas, fato este também reportado em todos os painéis da tabela 23.

Securato e Málaga (2004) chegam a resultado semelhante a respeitados nove portfólios analisados por eles. Para as ações com médio *book-to-market*, há um claro crescimento do retorno à medida que o tamanho das empresas aumenta. Já, para as ações com baixo *book-to-market*, essa relação existe, porém não tão forte, enquanto para aquelas com alto *book-to-market* não existe nenhuma relação aparente.

Tabela 23

Média dos fatores formadores dos portfólios, número de ações e média dos retornos das cateiras

Painel (a): 12 Carteiras formadas por Tamanho (3) - <i>Book-to-Market</i> (4)												
	S-L	S-LM	S-MH	S-H	M-L	M-LM	M-MH	M-H	B-L	B-LM	B-MH	B-H
Tamanho	760.201	694.574	762.036	565.983	4.274.102	3.421.649	3.113.558	3.317.681	35.219.299	26.933.520	34.530.434	80.672.083
BM	0,1327	1,7202	2,3811	11,9889	0,4035	1,0436	2,2021	9,6173	0,3811	1,0376	2,1455	7,5438
Nº de ações	6	7	17	25	16	19	12	7	19	15	11	8
Méd. do Ret.	-0,0026	-0,0011	-0,0028	-0,0007	-0,0005	0,0000	-0,0004	-0,0025	0,0001	0,0001	-0,0005	0,0000
Painel (b): 12 Carteiras formadas por Tamanho (3) - Investimento (4)												
	S-C	S-CM	S-MA	S-A	M-C	M-CM	M-MA	M-A	B-C	B-CM	B-MA	B-A
Tamanho	522.177	671.661	807.686	913.906	3.678.301	3.919.278	3.232.149	3.609.518	31.515.516	35.542.300	54.095.945	36.448.056
Investimento	-0,1632	0,0421	0,1022	0,9481	-0,1089	0,0443	0,1121	0,8545	-0,0672	0,0486	0,1153	1,8103
Nº de ações	23	14	10	8	11	13	15	16	7	14	17	15
Méd. do Ret.	-0,0032	-0,0002	-0,0012	-0,0012	-0,0001	-0,0007	-0,0007	-0,0004	0,0007	0,0002	-0,0002	-0,0002
Painel (c): 12 Carteiras formadas por Tamanho (2) - Investimento (3) - Lucratividade (2)												
	S-C-W	S-C-R	S-M-W	S-M-R	S-A-W	S-A-R	B-C-W	B-C-R	B-M-W	B-M-R	B-A-W	B-A-R
Tamanho	849.555	908.826	1.109.431	1.810.962	1.483.115	1.752.577	9.095.173	10.571.695	34.021.693	29.713.771	17.255.268	31.323.601
Investimento	-0,0726	-0,1472	0,0676	0,0721	0,2373	1,1844	-0,0375	-0,0637	0,0838	0,0724	0,2119	2,6569
Lucratividade	-0,6502	0,0114	-0,6815	-0,0322	-0,8692	0,0134	-0,5562	0,0032	-0,8943	0,0154	-0,8138	0,0993
Nº de ações	18	16	15	9	12	10	6	12	13	16	14	13
Méd. do Ret.	-0,0030	-0,0016	0,0008	-0,0017	-0,0010	-0,0007	0,0000	-0,0003	-0,0002	-0,0001	-0,0002	-0,0004
Painel (d): 12 Carteiras formadas por Tamanho (3) - VPIN (4)												
	S-L	S-LM	S-MH	S-H	M-L	M-LM	M-MH	M-H	B-L	B-LM	B-MH	B-H
Tamanho	0	1.073.377	923.803	735.634	4.566.823	4.231.166	3.423.443	2.879.412	53.144.379	25.213.573	18.970.675	10.884.802
VPIN	0	0,3518	0,4945	0,7524	0,2562	0,3461	0,4582	0,7092	0,2414	0,3401	0,4524	0,7494
Nº de ações	0	6	8	25	6	13	19	8	29	15	7	1
Méd. do Ret.	0	-0,0012	-0,0003	-0,0030	-0,0009	-0,0004	-0,0003	-0,0005	-0,0003	0,0002	0,0004	0,0012

Painel (e): 16 Carteiras formadas por Tamanho (4) - Lucratividade (4)

	S-W	S-WM	S-MR	S-R	SM-W	SM-WM	SM-MR	SM-R	MB-W	MB-WM	MB-MR	MB-R
Tamanho	452.193	437.089	412.135	487.576	2.029.404	1.834.094	2.124.392	2.120.793	5.489.320	6.090.563	5.410.900	6.520.701
Lucratividade	-1,1412	-0,3437	-0,1517	0,2251	-1,0668	-0,3129	-0,1438	0,1400	-0,8675	-0,3825	-0,0911	0,1899
Nº de ações	11	15	9	6	12	7	9	11	8	8	12	12
Méd. do Ret.	-0,0019	-0,0015	-0,0008	-0,0019	-0,0006	-0,0007	-0,0006	-0,0013	-0,0011	0,0002	-0,0006	-0,0003
	B-W	B-WM	B-MR	B-R								
Tamanho	21.855.742	53.878.319	42.785.750	54.607.414								
Lucratividade	-1,5975	-0,3747	-0,1063	0,1822								
Nº de ações	9	9	10	8								
Méd. do Ret.	-0,0004	-0,0001	-0,0002	0,0002								

Fonte - Elaborada pelo autor da dissertação.

Nota: o painel (a) mostra as características das 12 carteiras formadas a partir do tamanho e do *book-to-market* das ações. Cada carteira possui uma nomenclatura. A primeira letra remete à primeira variável (tamanho) onde 'S' é para as ações do grupo 'pequeno' (*small*), 'M' ações do grupo 'médio' (*medium*) e 'B' ações do grupo 'grande' (*big*). A segunda letra remete à segunda variável (*book-to-market*) onde 'L' é para as ações do grupo 'baixo' (*low*), 'LM' é para as ações do grupo 'baixo para médio' (*low to medium*), 'MH' para as ações do grupo 'médio para alto' (*medium to high*) e 'H' para as ações do grupo 'alto' (*high*). O painel (b) mostra as características das 12 carteiras formadas a partir do tamanho e investimento das ações. Novamente cada carteira possui uma nomenclatura. A primeira letra remete à primeira variável (tamanho) onde 'S' é para as ações do grupo 'pequeno' (*small*), 'M' ações do grupo 'médio' (*medium*), 'B' ações do grupo 'grande' (*big*). A segunda letra remete à segunda variável (investimento) onde 'C' é para as ações do grupo 'conservador' (*conservative*), 'CM' é para as ações do grupo 'conservador para médio' (*conservative to medium*), 'MA' para as ações do grupo 'médio para agressivo' (*medium to aggressive*) e 'A' para as ações do grupo 'agressivo' (*aggressive*). O painel (c) mostra as características das 12 carteiras formadas a partir do tamanho, investimento e lucratividade das ações. Novamente cada carteira possui uma nomenclatura. A primeira letra remete à primeira variável (tamanho) onde 'S' é para as ações do grupo 'pequeno' (*small*) e 'B' ações do grupo 'grande' (*big*). A segunda letra remete à segunda variável (investimento) onde 'C' é para as ações do grupo 'conservador' (*conservative*), 'M' para as ações do grupo 'médio' (*medium*) e 'A' para as ações do grupo 'agressivo' (*aggressive*). Em relação à terceira variável (lucratividade), 'W' se refere as ações do grupo 'fraco' (*weak*) e 'R' às ações do grupo 'robusto' (*robust*). O painel (d) mostra as características das 12 carteiras formadas a partir do tamanho e VPIN das ações. A primeira letra remete à primeira variável (tamanho) onde 'S' é para as ações do grupo 'pequeno' (*small*), 'M' ações do grupo 'médio' (*medium*), 'B' ações do grupo 'grande' (*big*). A segunda letra remete à variável VPIN onde 'L' é para as ações do grupo 'baixo' VPIN (*low*), 'LM' é para as ações do grupo 'baixo para médio' (*low to medium*), 'MH' para as ações do grupo 'médio para alto' (*medium to high*) e 'H' para as ações do grupo 'alto' (*high*). Finalmente, o painel (e) mostra as características das 16 ações formadas por base no tamanho e na lucratividade das ações. A primeira letra novamente se refere ao tamanho da empresa, onde onde 'S' é para as ações do grupo 'pequeno' (*small*), 'SM' é para as ações do grupo 'pequeno para médio' (*small to medium*), 'MB' ações do grupo 'médio para grande' (*medium to big*) e 'B' ações do grupo 'grande' (*big*). Para a variável lucratividade, a primeira letra, 'W' se refere as ações do grupo 'fraco' (*weak*), 'WM' se refere as ações do grupo 'fraco para médio' (*weak to medium*), 'MR' às ações do grupo 'médio para robusto' (*medium to robust*) e 'R' às ações do grupo 'robusto' (*robust*).

Tabela 24

Retornos das carteiras formadas por diferentes variáveis

Painel (a): 12 Carteiras formadas por Tamanho (3) - <i>Book-to-market</i> (4)					Painel (b): 12 Carteiras formadas por Tamanho (3) - Investimento (4)				
	L	LM	MH	H		C	CM	MA	A
S	-0,0026	-0,0011	-0,0028	-0,0007	S	-0,0032	-0,0002	-0,0012	-0,0012
M	-0,0005	0,0000	-0,0004	-0,0025	M	-0,0001	-0,0007	-0,0007	-0,0004
B	0,0001	0,0001	-0,0005	0,0000	B	0,0007	0,0002	-0,0002	-0,0002
Painel (c): 12 Carteiras formadas por Tamanho (2) - Investimento (3) - Lucratividade (2)									
	S		B			W		R	
	W	R	W	R		W	WM	MR	R
C	-0,0030	-0,0016	0,0000	-0,0003	S	-0,0019	-0,0015	-0,0008	-0,0019
M	0,0008	-0,0017	-0,0002	-0,0001	SM	-0,0006	-0,0007	-0,0006	-0,0013
A	-0,0010	-0,0007	-0,0002	-0,0004	MB	-0,0011	0,0002	-0,0006	-0,0003
					B	-0,0004	-0,0001	-0,0002	0,0002
Painel (d): 12 Carteiras formadas por Tamanho (3) - VPIN (4)					Painel (e): 16 Carteiras formadas por Tamanho (4) - Lucratividade (4)				
	L	LM	MH	H		W	WM	MR	R
S	0	-0,0012	-0,0003	-0,0030	S	-0,0019	-0,0015	-0,0008	-0,0019
M	-0,0009	-0,0004	-0,0003	-0,0005	SM	-0,0006	-0,0007	-0,0006	-0,0013
B	-0,0003	0,0002	0,0004	0,0012	MB	-0,0011	0,0002	-0,0006	-0,0003
					B	-0,0004	-0,0001	-0,0002	0,0002

Fonte - Elaborada pelo autor da dissertação.

Nota: o painel (a) mostra os retornos das 12 carteiras formadas a partir do tamanho e do *book-to-market* das ações. O nome das linhas remetem a primeira variável, em que a primeira letra 'S' se refere as ações do grupo 'pequeno' (*small*), 'M' ações do grupo 'médio' (*medium*) e 'B' ações do grupo 'grande' (*big*). As colunas remetem à segunda variável (*book-to-market*) onde 'L' é para as ações do grupo 'baixo' (*low*), 'LM' é para as ações do grupo 'baixo para médio' (*low to medium*), 'MH' para as ações do grupo 'médio para alto' (*medium to high*) e 'H' para as ações do grupo 'alto' (*high*). O painel (b) mostra os retornos das 12 carteiras formadas a partir do tamanho e investimento das ações. O nome das linhas remetem a primeira variável, em que a primeira letra 'S' se refere as ações do grupo 'pequeno' (*small*), 'M' ações do grupo 'médio' (*medium*) e 'B' ações do grupo 'grande' (*big*). As colunas remetem à segunda variável (investimento) onde 'C' é para as ações do grupo 'conservador' (*conservative*), 'CM' é para as ações do grupo 'conservador para médio' (*conservative to medium*), 'MA' para as ações do grupo 'médio para agressivo' (*medium to aggressive*) e 'A' para as ações do grupo 'agressivo' (*aggressive*). O painel (c) mostra os retornos das 12 carteiras formadas a partir do tamanho, investimento e lucratividade das ações. O painel foi dividido em duas partes referentes ao tamanho das empresas. A primeira parte 'S' se refere às empresas do grupo 'pequeno' (*small*) e a segunda parte 'B' ações do grupo 'grande' (*big*). Os nomes das linhas remetem à segunda variável (investimento) onde 'C' é para as ações do grupo 'conservador' (*conservative*), 'M' para as ações do grupo 'médio' (*medium*) e 'A' para as ações do grupo 'agressivo' (*aggressive*). Os nomes das colunas se referem a lucratividade, 'W' se refere as ações do grupo 'fraco' (*weak*) e 'R' às ações do grupo 'robusto' (*robust*). O painel (d) mostra os retornos das 12 carteiras formadas a partir do tamanho e VPIN das ações. Os nomes das linhas remetem à primeira variável (tamanho) onde 'S' é para as ações do grupo 'pequeno' (*small*), 'M' ações do grupo 'médio' (*medium*), 'B' ações do grupo 'grande' (*big*). Os nomes das colunas remetem à variável VPIN onde 'L' é para as ações do grupo 'baixo' VPIN (*low*), 'LM' é para as ações do grupo 'baixo para médio' (*low to medium*), 'MH' para as ações do grupo 'médio para alto' (*medium to high*) e 'H' para as ações do grupo 'alto' (*high*). Finalmente, o painel (e) mostra os retornos das 16 ações formadas por base no tamanho e na lucratividade das ações. Os nomes das linhas se referem ao tamanho da empresa, onde 'S' é para as ações do grupo 'pequeno' (*small*), 'SM' é para as ações

do grupo ‘pequeno para médio’ (*small to medium*), ‘MB’ ações do grupo ‘médio para grande’ (*medium to big*) e ‘B’ ações do grupo ‘grande’ (*big*). Em relação às colunas, a primeira letra, ‘W’ se refere as ações do grupo ‘fraco’ (*weak*), ‘WM’ se refere as ações do grupo ‘fraco para médio’ (*weak to medium*), ‘MR’ às ações do grupo ‘médio para robusto’ (*medium to robust*) e ‘R’ às ações do grupo ‘robusto’ (*robust*).

O efeito relacionado à variável investimento, adicionada por Fama e French (2015), apresenta comportamento similar àquele relatado pelos autores. Como reportado no painel (b) da tabela 23, controlando para o tamanho, a medida que se passa das ações consideradas conservadoras, ou seja, baixo investimento, para aquelas agressivas – alto grau de investimento – o retorno médio diminui. Em resultados não tabulados, foram formados 25 portfólios baseados em tamanho e investimento. O número médio das ações foi de sete, sendo o mínimo duas ações e máximo 14 ações. O comportamento se confirma para essas carteiras, ou seja, à medida que as ações apresentam maior investimento, seus retornos diminuem.

O painel (c) apresenta os retornos dos portfólios formados pelas variáveis tamanho, investimento e lucratividade. Dentro das ações menores em tamanho, estas apresentam um aumento no retorno à medida que se aumenta a lucratividade, exceto para aquelas de médio investimento. Para as ações maiores, o grupo conservador e agressivo apresentou diminuição do retorno à medida que elas se tornam mais robustas em relação à lucratividade. Novamente, as ações consideradas médias para o investimento apresentaram resultado inverso aos outros grupos.

Em relação às variáveis tamanho e VPIN, o painel (d) reporta os retornos dos portfólios criados. Para todos os grupos de tamanho verifica-se que à medida que se aumenta o VPIN, o retorno também cresce. Esse resultado é especialmente acentuado para as ações do grupo *grande*. Como já foi verificado, estas, contrariamente ao esperado pelo modelo de Fama e French (1993), tendem a apresentar maiores retornos. Na maior parte dos casos, em sintonia com os trabalhos de Easley, Hvidkjaer e O’Hara (2005), à medida que aumenta a probabilidade de negociações privilegiadas e, assim, o risco informacional, o retorno apresentado também cresce, o que sugere a precificação de um fator informacional.

Finalmente, o painel (e) reporta os resultados para os portfólios formados por tamanho e lucratividade. De forma geral, esperava-se que, quanto maior a lucratividade, maior os retornos, assim como em Fama e French (2015). Os resultados não apresentam relação aparente, não sendo, portanto, possível confirmar essa hipótese.

A diferença entre os resultados apresentados aqui e aqueles reportados pelos trabalhos que se valem dos modelos de fatores no mercado americano são relevantes, especialmente no

tocante à justificativa para a criação dos fatores, haja vista a não confirmação do efeito relacionado ao tamanho, *book-to-market* e lucratividade. Para as variáveis investimento e VPIN, foi verificado o efeito esperado. Essa dificuldade já era esperada, como atentado por Argolo, Leal e Almeida (2012). Procede-se, na seção seguinte, à apresentação dos resultados das regressões dos modelos de fatores.

4.4.3 Resultados das regressões dos modelos de fatores

Em vista de se apresentar os resultados referentes às regressões aplicadas, faz-se necessária a análise dos testes econométricos que visam testar a robustez estatística dos modelos analisados. Foram realizados três testes para verificar respectivamente se existe multicolinearidade, se os resíduos das regressões são autocorrelacionados e se estes são heterocedásticos.

O primeiro teste, *Variance Inflation Factor*, ou VIF, testa se as variáveis explicativas são correlacionadas, o que afetaria a estimação de seus coeficientes. A tabela 25 mostra os resultados para os diferentes modelos testados. Seguindo Gujarati (2006), valores de VIF acima de 10 indicam multicolinearidade. Verifica-se que em nenhum dos seis modelos o resultado para o VIF das variáveis foi grande, sendo o maior valor 1,9937 para a HML nos modelos de cinco fatores e no de 5 fatores como o fator IMU adicionado. Os resultados, portanto, indicam que não há multicolinearidade entre as variáveis, sendo possível a sua inclusão em um modelo de regressão sem aparente prejuízo nas estimativas dos coeficientes.

Tabela 25

Resultados do teste VIF de multicolinearidade das variáveis explicativas

VIF - <i>Variance Inflation Factor</i>						
Variável	3 Fatores	3 Fatores + IMU	4 Fatores	4 Fatores + IMU	5 Fatores	5 Fatores + IMU
MKT	1,4479	1,4682	1,4754	1,4914	1,4485	1,5163
SMB	1,2545	1,6828	1,2813	1,7366	1,3880	1,7299
HML	1,1747	1,2165	1,6211	1,6963	1,9937	1,9937
IMU		1,5210		1,5370		1,5468
UMD			1,5359	1,5521		
RMW					1,5030	1,5194
CMA					1,4314	1,4546

Fonte - Elaborada pelo autor da dissertação.

Os resultados para os testes de Durbin-Watson e Breusch-Pagan são apresentados nas tabelas 37 e 38 do apêndice. Dos 62 portfólios apenas um rejeitou a hipótese nula de não autocorrelação dos resíduos, tendo apresentado estatística de teste de 1,77, ou seja, a correlação foi positiva, porém não se mostrou forte. Em relação ao teste de Breusch-Pagan, apenas seis portfólios rejeitaram a hipótese nula de homocedasticidade. Tendo em vista esses resultados conclui-se que os estimadores MQO padrões são suficientes para a estimação concisa dos coeficientes dos modelos, não sendo, portanto, necessária a aplicação de métodos de correção como o de Newey-West, por exemplo.

Dividiu-se a apresentação dos resultados em três tabelas segmentadas em painéis para cada um dos cinco conjuntos de portfólios testados. Cada tabela contém um dos modelos aplicados com a respectiva adição do fator IMU em cada um deles. São apresentados, assim, os coeficientes dos fatores e seus p.valores. Além disso, são apresentados os valores para o R^2 de cada regressão.

A tabela 26 apresenta os resultados da aplicação do modelo de três fatores e o modelo de três fatores adicionado o IMU nos cinco conjunto de carteiras criadas para serem usadas como variáveis dependentes. Como regra geral, o fator de mercado derivado do CAPM tem coeficiente positivo e estatisticamente significativo para todas as 62 carteiras testadas. Esse resultado está de acordo com os demais estudos que aplicaram os modelos de fatores. Via de regra, a maior parte da variação dos retornos é efetivamente capturada pelo fator de mercado sendo o objetivo dos demais fatores capturar a parte da variação não explicada pelo fator proposto pelo CAPM.

Outro efeito já verificado por Fama e French (1993) é que a medida que o tamanho das ações dentro do portfólios aumenta, o coeficiente do fator SMB diminui e se torna estatisticamente insignificante. Isto foi constatado para os cinco conjuntos de portfólios, tendo em vista que todos eles têm como variável constante o tamanho das ações. Segundo Fama e French (1993), isso se deve à relação dos coeficientes do SMB com o tamanho das ações e já era esperado pela composição desse fator.

Em relação ao fator HML, somente para quatro dos 62 portfólios ele se mostrou estatisticamente significativo – considerando um alfa de 5% – sendo três deles dentro da interseção entre tamanho e *book-to-market* e somente um fora desse conjunto. Isto indica que esse fator não captura praticamente nenhuma variação dos retornos dentro dos portfólios,

exceto para aqueles em que o próprio fator se mostra um componente na criação das carteiras. Os resultados de Fama e French (1993, 1996) mostraram que o coeficiente para o fator HML era em geral menor do que para os outros dois fatores, porém estatisticamente significativo. Easley, Hvidkjaer e O'Hara (2005) e Mohranram e Rajgopal (2009) encontram os mesmos resultados, ou seja, coeficientes baixos, porém estatisticamente significantes. Resultados semelhantes aos deste trabalho, no qual o fator HML não é capaz de explicar os retornos das carteiras, foram encontrados por Fama e French (2015), o que será comentado mais adiante.

Em relação ao R^2 das regressões, a média ficou em torno de 0,06, indicando pouca capacidade das variáveis independentes de explicar a variância dos retornos das carteiras. O maior valor foi de 0,17 para um portfólio dentro da combinação tamanho e investimento; e o menor, 0,02.

Analisando a adição de um fator relacionado ao risco informacional, o IMU, verifica-se que 45 dos 62 portfólios apresentam coeficientes negativos. Como esperado, os coeficientes negativos são mais frequentes à medida que o tamanho das ações aumenta dentro dos conjuntos de carteiras, o que indica que maiores empresas têm uma sensibilidade negativa para o fator de risco informacional o que está de acordo com o encontrado por Mohanram e Rajgopal (2009). Como há uma correlação negativa entre a probabilidade de negociações privadas e o tamanho das empresas e há maior número de analistas seguindo tais empresas, o prêmio pelo risco de informações privilegiadas deve ser menor. Apesar desse resultado favorável ao fator de risco informacional, verifica-se também que apenas em cinco portfólios este apresentou p.valores menores do que 0,05, o que indica que, em 57 das 62 carteiras, não rejeita-se a hipótese nula de que o coeficiente relacionado ao risco informacional não é estatisticamente diferente de zero e que, portanto, não explica os retornos dos portfólios. Com a adição do fator IMU, passa-se a rejeitar a hipótese nula para o fator SMB em 15 portfólios, em contraste com a rejeição em 11 para o modelo de três fatores de Fama e French (1993), ou seja, a inclusão do fator IMU pode ter auxiliado no aumento do poder explicado do fator SMB. Em relação aos alfas das regressões, não rejeita-se a hipótese nula de que eles não são estatisticamente diferentes de zero para 56 regressões para ambos os modelos. Os resultados em conjunto serão apresentados a partir do teste GRS mais adiante.

Mohanram e Rajgopal (2009), em seus estudos, afirmam que os coeficientes negativos para o fator relacionado ao PIN estão mais presentes nas carteiras formadas por ativos

menores, padrão contrário ao esperado em se tratando do risco informacional. Apesar disso, os autores concluem que o coeficiente para o fator PIN é estatisticamente significativo na explicação dos retornos dos portfólios propostos. Easley, Hvidkjaer e O'Hara (2005), analisando dez carteiras formadas com base no tamanho e PIN das ações, constatam que o fator PIN é positivo e decrescente à medida que aumenta o tamanho das empresas. Para esses autores, tal fator apresentou maior poder de explicação, inclusive, do que o fator de mercado. O R^2 dessas regressões, no entanto, se manteve baixo e próximo de 0,20 para todos os portfólios.

A tabela 27 apresenta os resultados para as regressões utilizando o modelo de quatro fatores de Carhart (1997) e, posteriormente, a adição do IMU. Para esse modelo, as análises a respeito do MKT, SMB e HML se mantiveram. Ou seja, o MKT é o fator predominante na explicação dos retornos dos portfólios, enquanto o SMB se torna significativamente menor à medida que o tamanho das ações cresce e o HML apresenta p.valor maior que 0,05 para a grande maioria das regressões – nesse caso, rejeita-se a hipótese nula de que o coeficiente é igual a zero em somente oito das 62 regressões. Já, para o fator relacionado ao momento das ações, UMD, o resultado apresentado foi similar àquele de Mohanram e Rajgopal (2009), tanto em magnitude (valor do coeficiente), quanto em sinal. Em 57 das 62 regressões, o coeficiente UMD foi negativo e, em 37 delas, rejeita-se a hipótese nula de que este é estatisticamente igual a zero. Mohanram e Rajgopal (2009) encontram coeficientes negativos para o fator UMD em 29 dos 30 portfólios analisados, enquanto, para Easley, Hvidkjaer e O'Hara (2005), esse fator apresentou coeficiente positivo em oito das 10 carteiras utilizadas. Verifica-se, portanto, que o fator UMD apresentou melhor poder de explicação do que os demais fatores – exceto o MKT –, representado também pelo aumento do R^2 para 0,075 ao aplicar o modelo de quatro fatores de Carhart (1997).

Tabela 26

Coeficientes para os modelos de 3 Fatores e 3 Fatores adicionado o IMU

	3 Fatores						3 Fatores + IMU					
	Nº	α	MKT	SMB	HML	R ²	α	MKT	SMB	HML	IMU	R ²
12 Portfólios: Tamanho e <i>Book-to-market</i>	1	-0.002 (0.06)	0.547 (0.00)	0.372 (0.00)	-0.217 (0.01)	0.08	-0.002 (0.06)	0.542 (0.00)	0.404 (0.00)	-0.209 (0.02)	-0.08 (0.61)	0.08
	2	-0.001 (0.11)	0.267 (0.00)	0.117 (0.07)	0.108 (0.04)	0.09	-0.001 (0.12)	0.258 (0.00)	0.175 (0.02)	0.123 (0.02)	-0.144 (0.12)	0.09
	3	-0.003 (0.00)	0.197 (0.00)	0.044 (0.61)	-0.01 (0.89)	0.02	-0.003 (0.00)	0.191 (0.01)	0.086 (0.40)	0.001 (0.99)	-0.104 (0.41)	0.02
	4	-0.001 (0.64)	0.364 (0.00)	0.179 (0.10)	0.046 (0.61)	0.04	-0.001 (0.64)	0.363 (0.00)	0.184 (0.15)	0.047 (0.60)	-0.013 (0.93)	0.04
	5	-0.001 (0.21)	0.228 (0.00)	0.027 (0.61)	-0.047 (0.27)	0.07	-0.001 (0.22)	0.224 (0.00)	0.055 (0.37)	-0.04 (0.36)	-0.069 (0.37)	0.07
	6	0 (0.66)	0.204 (0.00)	0.021 (0.61)	0.023 (0.51)	0.1	0 (0.68)	0.201 (0.00)	0.042 (0.39)	0.028 (0.42)	-0.051 (0.40)	0.1
	7	0 (0.51)	0.28 (0.00)	0.077 (0.24)	0.034 (0.53)	0.07	0 (0.49)	0.284 (0.00)	0.048 (0.53)	0.026 (0.63)	0.072 (0.45)	0.08
	8	-0.002 (0.03)	0.465 (0.00)	0.215 (0.02)	0.145 (0.05)	0.11	-0.002 (0.03)	0.453 (0.00)	0.29 (0.01)	0.165 (0.03)	-0.187 (0.16)	0.12
	9	0 (0.66)	0.165 (0.00)	-0.041 (0.41)	-0.015 (0.70)	0.06	0 (0.66)	0.164 (0.00)	-0.036 (0.52)	-0.014 (0.73)	-0.011 (0.87)	0.06
	10	0 (0.68)	0.145 (0.00)	-0.001 (0.99)	-0.001 (0.99)	0.04	0 (0.72)	0.14 (0.00)	0.03 (0.60)	0.008 (0.85)	-0.076 (0.28)	0.04
	11	-0.001 (0.48)	0.357 (0.00)	0.037 (0.59)	0.075 (0.18)	0.12	-0.001 (0.49)	0.355 (0.00)	0.049 (0.54)	0.078 (0.17)	-0.029 (0.77)	0.12
	12	0 (0.92)	0.361 (0.00)	-0.061 (0.49)	0.188 (0.01)	0.11	0 (0.96)	0.353 (0.00)	-0.01 (0.92)	0.201 (0.01)	-0.126 (0.33)	0.12
12 Portfólios: Tamanho e Investimento	Nº	α	MKT	SMB	HML	R ²	α	MKT	SMB	HML	IMU	R ²
	1	-0.003 (0.00)	0.315 (0.00)	0.2 (0.03)	0.01 (0.89)	0.04	-0.003 (0.00)	0.322 (0.00)	0.157 (0.14)	-0.001 (0.99)	0.108 (0.42)	0.04
	2	0 (1.00)	0.517 (0.00)	0.181 (0.23)	-0.112 (0.36)	0.04	0 (0.99)	0.516 (0.00)	0.187 (0.28)	-0.11 (0.37)	-0.013 (0.95)	0.04
	3	-0.001 (0.37)	0.296 (0.00)	0.208 (0.06)	0.068 (0.45)	0.03	-0.001 (0.42)	0.273 (0.00)	0.356 (0.01)	0.107 (0.24)	-0.369 (0.02)	0.04
	4	-0.001 (0.09)	0.21 (0.00)	0.059 (0.43)	0.077 (0.21)	0.04	-0.001 (0.10)	0.203 (0.00)	0.1 (0.25)	0.087 (0.16)	-0.103 (0.34)	0.04
	5	0 (0.62)	0.242 (0.00)	0.041 (0.45)	0.009 (0.83)	0.08	0 (0.61)	0.244 (0.00)	0.033 (0.60)	0.007 (0.87)	0.021 (0.80)	0.08
6	-0.001	0.314	0.083	0.007	0.08	-0.001	0.312	0.099	0.012	-0.04	0.08	

	Nº	α	MKT	SMB	HML	R ²	α	MKT	SMB	HML	IMU	R ²
		(0.32)	(0.00)	(0.23)	(0.90)		(0.33)	(0.00)	(0.21)	(0.84)	(0.69)	
	7	-0.001	0.193	0.069	0.014	0.05	-0.001	0.19	0.09	0.019	-0.052	0.05
		(0.14)	(0.00)	(0.18)	(0.74)		(0.14)	(0.00)	(0.14)	(0.66)	(0.50)	
	8	-0.001	0.25	0.058	-0.022	0.08	-0.001	0.245	0.094	-0.013	-0.089	0.09
		(0.26)	(0.00)	(0.26)	(0.60)		(0.29)	(0.00)	(0.12)	(0.77)	(0.24)	
	9	0.001	0.359	-0.011	0.214	0.17	0.001	0.35	0.042	0.228	-0.131	0.17
		(0.34)	(0.00)	(0.88)	(0.00)		(0.31)	(0.00)	(0.61)	(0.00)	(0.20)	
	10	0	0.135	-0.044	0.011	0.03	0	0.138	-0.067	0.005	0.058	0.03
		(0.81)	(0.00)	(0.45)	(0.81)		(0.79)	(0.00)	(0.32)	(0.91)	(0.50)	
	11	0	0.235	-0.025	0.01	0.09	0	0.234	-0.017	0.012	-0.021	0.09
		(0.46)	(0.00)	(0.64)	(0.83)		(0.47)	(0.00)	(0.78)	(0.79)	(0.79)	
	12	-0.001	0.185	-0.012	-0.008	0.06	0	0.177	0.041	0.006	-0.133	0.07
		(0.36)	(0.00)	(0.81)	(0.84)		(0.40)	(0.00)	(0.49)	(0.89)	(0.07)	
	Nº	α	MKT	SMB	HML	R ²	α	MKT	SMB	HML	IMU	R ²
12 Portfólios: Tamanho, Investimento e Lucratividade	1	-0.003	0.38	0.24	-0.061	0.06	-0.003	0.381	0.235	-0.062	0.011	0.06
		(0.00)	(0.00)	(0.01)	(0.41)		(0.00)	(0.00)	(0.03)	(0.41)	(0.93)	
	2	-0.001	0.305	0.22	0.148	0.05	-0.001	0.304	0.227	0.15	-0.016	0.05
		(0.22)	(0.00)	(0.04)	(0.09)		(0.22)	(0.00)	(0.06)	(0.09)	(0.92)	
	3	0.001	0.388	0.064	0.057	0.04	0.001	0.385	0.084	0.062	-0.05	0.04
		(0.57)	(0.00)	(0.62)	(0.60)		(0.57)	(0.00)	(0.58)	(0.57)	(0.79)	
	4	-0.002	0.403	0.25	-0.123	0.06	-0.002	0.404	0.243	-0.125	0.018	0.06
		(0.11)	(0.00)	(0.01)	(0.09)		(0.11)	(0.00)	(0.02)	(0.09)	(0.89)	
	5	-0.001	0.248	0.113	0.016	0.05	-0.001	0.234	0.199	0.038	-0.215	0.06
		(0.15)	(0.00)	(0.11)	(0.78)		(0.18)	(0.00)	(0.01)	(0.51)	(0.03)	
	6	-0.001	0.244	0.099	0.044	0.08	-0.001	0.228	0.197	0.07	-0.246	0.09
		(0.22)	(0.00)	(0.08)	(0.34)		(0.27)	(0.00)	(0.00)	(0.14)	(0.00)	
7	0	0.227	0.089	0.01	0.04	0	0.227	0.087	0.009	0.004	0.04	
	(0.86)	(0.00)	(0.19)	(0.86)		(0.86)	(0.00)	(0.27)	(0.87)	(0.97)		
8	-0.001	0.17	-0.042	-0.024	0.04	-0.001	0.169	-0.037	-0.022	-0.013	0.04	
	(0.31)	(0.00)	(0.46)	(0.61)		(0.31)	(0.00)	(0.58)	(0.64)	(0.88)		
9	-0.001	0.19	-0.043	-0.012	0.06	-0.001	0.194	-0.066	-0.018	0.058	0.06	
	(0.31)	(0.00)	(0.43)	(0.79)		(0.30)	(0.00)	(0.30)	(0.70)	(0.47)		
10	0	0.223	0.004	0.041	0.08	0	0.224	-0.003	0.039	0.018	0.08	
	(0.56)	(0.00)	(0.94)	(0.35)		(0.56)	(0.00)	(0.96)	(0.38)	(0.81)		
11	-0.001	0.203	0.011	-0.039	0.07	0	0.198	0.041	-0.031	-0.075	0.07	
	(0.30)	(0.00)	(0.81)	(0.31)		(0.33)	(0.00)	(0.44)	(0.42)	(0.27)		
12	-0.001	0.208	0.001	0	0.06	-0.001	0.202	0.044	0.011	-0.107	0.07	
	(0.24)	(0.00)	(0.99)	(1.00)		(0.27)	(0.00)	(0.49)	(0.81)	(0.18)		

	Nº	α	MKT	SMB	HML	R^2	α	MKT	SMB	HML	IMU	R^2
10 Portfólios: Tamanho e VPIN	1	-0.001 (0.16)	0.305 (0.00)	0.064 (0.43)	0.066 (0.31)	0.07	-0.001 (0.22)	0.272 (0.00)	0.271 (0.00)	0.12 (0.07)	-0.516 (0.00)	0.1
	2	0 (1.00)	0.431 (0.00)	0.219 (0.04)	0.091 (0.30)	0.07	0 (0.99)	0.428 (0.00)	0.244 (0.05)	0.098 (0.27)	-0.062 (0.69)	0.07
	3	-0.003 (0.00)	0.337 (0.00)	0.154 (0.08)	-0.082 (0.25)	0.05	-0.003 (0.00)	0.34 (0.00)	0.137 (0.18)	-0.086 (0.23)	0.043 (0.74)	0.05
	4	-0.001 (0.15)	0.25 (0.00)	-0.019 (0.79)	0.03 (0.61)	0.06	-0.001 (0.16)	0.246 (0.00)	0.007 (0.94)	0.036 (0.54)	-0.064 (0.54)	0.06
	5	0 (0.42)	0.31 (0.00)	0.053 (0.34)	-0.019 (0.67)	0.11	0 (0.44)	0.304 (0.00)	0.089 (0.17)	-0.01 (0.83)	-0.091 (0.27)	0.11
	6	-0.001 (0.33)	0.249 (0.00)	0.016 (0.76)	-0.027 (0.53)	0.09	-0.001 (0.36)	0.243 (0.00)	0.057 (0.35)	-0.017 (0.71)	-0.102 (0.18)	0.09
	7	-0.001 (0.23)	0.124 (0.01)	0.011 (0.86)	0.029 (0.56)	0.02	-0.001 (0.22)	0.127 (0.01)	-0.006 (0.93)	0.025 (0.63)	0.043 (0.63)	0.02
	8	-0.001 (0.30)	0.199 (0.00)	-0.038 (0.51)	0 (1.00)	0.06	-0.001 (0.32)	0.194 (0.00)	-0.009 (0.89)	0.008 (0.87)	-0.073 (0.38)	0.06
	9	0 (0.91)	0.181 (0.00)	-0.004 (0.95)	0.039 (0.37)	0.06	0 (0.90)	0.181 (0.00)	-0.007 (0.91)	0.038 (0.39)	0.008 (0.92)	0.06
	10	0 (0.56)	0.294 (0.00)	0.068 (0.23)	0.049 (0.29)	0.12	0 (0.57)	0.296 (0.00)	0.056 (0.38)	0.046 (0.33)	0.028 (0.73)	0.12
	Nº	α	MKT	SMB	HML	R^2	α	MKT	SMB	HML	IMU	R^2
16 Portfólios: Tamanho e Lucratividade	1	-0.002 (0.10)	0.335 (0.00)	0.041 (0.71)	-0.021 (0.82)	0.04	-0.002 (0.10)	0.337 (0.00)	0.03 (0.82)	-0.024 (0.80)	0.027 (0.87)	0.04
	2	-0.001 (0.26)	0.399 (0.00)	0.27 (0.01)	-0.01 (0.91)	0.05	-0.001 (0.26)	0.398 (0.00)	0.276 (0.02)	-0.008 (0.92)	-0.015 (0.92)	0.05
	3	-0.001 (0.40)	0.268 (0.00)	0.076 (0.41)	0.144 (0.06)	0.05	-0.001 (0.41)	0.267 (0.00)	0.082 (0.44)	0.145 (0.06)	-0.016 (0.90)	0.05
	4	-0.002 (0.35)	0.319 (0.01)	0.332 (0.03)	0.115 (0.36)	0.02	-0.002 (0.34)	0.324 (0.01)	0.298 (0.10)	0.106 (0.41)	0.087 (0.70)	0.02
	5	-0.001 (0.36)	0.236 (0.00)	0.104 (0.10)	0.054 (0.30)	0.06	-0.001 (0.39)	0.225 (0.00)	0.171 (0.02)	0.071 (0.17)	-0.167 (0.07)	0.07
	6	-0.001 (0.32)	0.257 (0.00)	0.155 (0.02)	0.008 (0.88)	0.05	-0.001 (0.33)	0.253 (0.00)	0.179 (0.03)	0.015 (0.80)	-0.061 (0.55)	0.05
	7	-0.001 (0.30)	0.245 (0.00)	0.079 (0.24)	0.025 (0.65)	0.05	-0.001 (0.35)	0.232 (0.00)	0.163 (0.04)	0.047 (0.40)	-0.21 (0.03)	0.06
	8	-0.001 (0.16)	0.4 (0.00)	0.216 (0.00)	-0.065 (0.30)	0.09	-0.001 (0.17)	0.398 (0.00)	0.225 (0.01)	-0.063 (0.32)	-0.022 (0.85)	0.09
	9	-0.001 (0.03)	0.177 (0.00)	-0.026 (0.67)	-0.02 (0.70)	0.04	-0.001 (0.03)	0.178 (0.00)	-0.028 (0.70)	-0.02 (0.70)	0.005 (0.96)	0.04
	10	0	0.214	0.001	-0.001	0.07	0	0.215	-0.004	-0.003	0.013	0.07

Nº	α	MKT	SMB	HML	R ²	α	MKT	SMB	HML	IMU	R ²
	(0.95)	(0.00)	(0.99)	(0.98)		(0.94)	(0.00)	(0.94)	(0.95)	(0.87)	
11	-0.001	0.303	0.046	0.019	0.1	-0.001	0.299	0.074	0.026	-0.069	0.1
	(0.26)	(0.00)	(0.45)	(0.71)		(0.27)	(0.00)	(0.29)	(0.61)	(0.44)	
12	-0.001	0.172	-0.024	0.053	0.06	-0.001	0.171	-0.02	0.054	-0.01	0.06
	(0.31)	(0.00)	(0.67)	(0.26)		(0.32)	(0.00)	(0.76)	(0.26)	(0.90)	
13	-0.001	0.201	0.025	-0.014	0.05	-0.001	0.196	0.055	-0.006	-0.076	0.05
	(0.32)	(0.00)	(0.68)	(0.77)		(0.34)	(0.00)	(0.42)	(0.91)	(0.38)	
14	0	0.269	0.011	-0.045	0.09	0	0.269	0.014	-0.044	-0.009	0.09
	(0.56)	(0.00)	(0.85)	(0.32)		(0.56)	(0.00)	(0.83)	(0.34)	(0.92)	
15	-0.001	0.12	-0.047	-0.039	0.02	-0.001	0.119	-0.043	-0.038	-0.01	0.02
	(0.30)	(0.01)	(0.43)	(0.42)		(0.30)	(0.01)	(0.53)	(0.44)	(0.91)	
16	0	0.242	0.006	-0.008	0.07	0	0.235	0.053	0.004	-0.118	0.07
	(0.96)	(0.00)	(0.92)	(0.86)		(0.99)	(0.00)	(0.44)	(0.93)	(0.17)	

Fonte: Elaborada pelo autor da dissertação.

Nota: os valores entre parênteses representam os p.valores para os coeficientes.

Ao adicionar o fator IMU no modelo de quatro fatores, os resultados auferidos são similares àqueles reportados para o modelo de três fatores. O valor para o IMU é em geral negativo, porém com a adição do UMD passa-se para 37 coeficientes negativos ao invés dos 45 para o modelo de três fatores. Além disso, tem-se que em somente quatro regressões os p.valores foram menores do que 0,05, ou seja, o poder de explicação do IMU diminuiu com a inclusão do fator UMD, o que pode indicar que este último captura parte das variações que se esperava serem explicadas pelo IMU. É de interesse notar que, com a inclusão do IMU, a hipótese nula para o SMB é rejeitada em 12 portfólios, em contraste com a rejeição em apenas sete para o modelo de quatro fatores de Carhart (1997), assim como verificado para o modelo de três fatores. Novamente para os alfas das regressões não rejeita-se a hipótese nula em 56 delas, assim como para os modelos de três fatores. Em relação ao R² das regressões, este não apresentou aumento substancial ao se adicionar o IMU, o que já era esperado haja vista a não rejeição da hipótese de que este não é diferente de zero para a grande maioria dos portfólios analisados.

Tabela 27

Coeficientes para os modelos de 4 Fatores e 4 Fatores adicionado o IMU

	4 Fatores							4 Fatores + IMU						
	Nº	α	MKT	SMB	HML	UMD	R ²	α	MKT	SMB	HML	UMD	IMU	R ²
12 Portfólios: Tamanho e Book-to-market	1	-0.002 (0.06)	0.53 (0.00)	0.348 (0.00)	-0.299 (0.00)	-0.217 (0.13)	0.08	-0.002 (0.06)	0.527 (0.00)	0.372 (0.00)	-0.291 (0.01)	-0.212 (0.14)	-0.057 (0.72)	0.08
	2	-0.001 (0.11)	0.254 (0.00)	0.099 (0.12)	0.048 (0.43)	-0.158 (0.06)	0.09	-0.001 (0.12)	0.247 (0.00)	0.152 (0.04)	0.066 (0.29)	-0.147 (0.09)	-0.127 (0.17)	0.1
	3	-0.003 (0.00)	0.17 (0.01)	0.007 (0.94)	-0.137 (0.10)	-0.335 (0.00)	0.04	-0.003 (0.00)	0.166 (0.02)	0.035 (0.73)	-0.127 (0.13)	-0.329 (0.00)	-0.067 (0.60)	0.04
	4	-0.001 (0.64)	0.364 (0.00)	0.18 (0.11)	0.048 (0.65)	0.005 (0.97)	0.04	-0.001 (0.64)	0.363 (0.00)	0.185 (0.15)	0.05 (0.64)	0.007 (0.96)	-0.014 (0.93)	0.04
	5	-0.001 (0.21)	0.208 (0.00)	0 (1.00)	-0.141 (0.00)	-0.248 (0.00)	0.09	-0.001 (0.21)	0.206 (0.00)	0.017 (0.78)	-0.135 (0.01)	-0.244 (0.00)	-0.042 (0.58)	0.09
	6	0 (0.66)	0.194 (0.00)	0.008 (0.86)	-0.025 (0.53)	-0.126 (0.02)	0.11	0 (0.68)	0.192 (0.00)	0.023 (0.64)	-0.02 (0.63)	-0.123 (0.03)	-0.038 (0.54)	0.11
	7	0 (0.51)	0.27 (0.00)	0.064 (0.34)	-0.011 (0.86)	-0.118 (0.18)	0.08	0 (0.49)	0.275 (0.00)	0.028 (0.71)	-0.023 (0.72)	-0.126 (0.15)	0.086 (0.37)	0.08
	8	-0.002 (0.03)	0.445 (0.00)	0.188 (0.04)	0.053 (0.54)	-0.243 (0.05)	0.12	-0.002 (0.03)	0.436 (0.00)	0.255 (0.02)	0.076 (0.40)	-0.228 (0.06)	-0.162 (0.23)	0.12
	9	0 (0.66)	0.153 (0.00)	-0.057 (0.25)	-0.072 (0.13)	-0.149 (0.02)	0.07	0 (0.66)	0.153 (0.00)	-0.06 (0.30)	-0.073 (0.13)	-0.15 (0.02)	0.005 (0.94)	0.07
	10	0 (0.69)	0.133 (0.00)	-0.017 (0.73)	-0.057 (0.22)	-0.149 (0.02)	0.05	0 (0.71)	0.13 (0.00)	0.008 (0.89)	-0.048 (0.31)	-0.143 (0.03)	-0.06 (0.40)	0.05
	11	-0.001 (0.48)	0.324 (0.00)	-0.006 (0.93)	-0.076 (0.24)	-0.398 (0.00)	0.15	-0.001 (0.48)	0.325 (0.00)	-0.013 (0.87)	-0.078 (0.24)	-0.4 (0.00)	0.016 (0.88)	0.15
	12	0 (0.93)	0.328 (0.00)	-0.105 (0.24)	0.035 (0.68)	-0.403 (0.00)	0.13	0 (0.95)	0.324 (0.00)	-0.071 (0.49)	0.047 (0.59)	-0.396 (0.00)	-0.082 (0.53)	0.14
12 Portfólios: Tamanho e Investimento	1	-0.003 (0.00)	0.296 (0.00)	0.174 (0.06)	-0.08 (0.37)	-0.239 (0.05)	0.05	-0.003 (0.00)	0.304 (0.00)	0.118 (0.28)	-0.099 (0.27)	-0.252 (0.04)	0.136 (0.31)	0.05
	2	0 (1.00)	0.52 (0.00)	0.185 (0.22)	-0.098 (0.49)	0.036 (0.86)	0.04	0 (0.99)	0.519 (0.00)	0.192 (0.28)	-0.096 (0.51)	0.038 (0.85)	-0.017 (0.94)	0.04
	3	-0.001 (0.37)	0.277 (0.00)	0.182 (0.10)	-0.02 (0.85)	-0.234 (0.11)	0.04	-0.001 (0.42)	0.258 (0.00)	0.324 (0.01)	0.028 (0.79)	-0.202 (0.17)	-0.346 (0.03)	0.04
	4	-0.001 (0.09)	0.213 (0.00)	0.062 (0.41)	0.09 (0.21)	0.035 (0.72)	0.04	-0.001 (0.10)	0.207 (0.00)	0.107 (0.22)	0.105 (0.15)	0.045 (0.65)	-0.108 (0.32)	0.04
	5	0 (0.62)	0.227 (0.00)	0.02 (0.71)	-0.063 (0.23)	-0.191 (0.01)	0.09	0 (0.61)	0.229 (0.00)	0.003 (0.97)	-0.069 (0.20)	-0.195 (0.01)	0.042 (0.60)	0.09
	6	-0.001 (0.32)	0.305 (0.00)	0.071 (0.31)	-0.036 (0.58)	-0.115 (0.21)	0.08	-0.001 (0.33)	0.303 (0.00)	0.082 (0.31)	-0.032 (0.63)	-0.113 (0.22)	-0.028 (0.78)	0.08

	Nº	α	MKT	SMB	HML	UMD	R ²	α	MKT	SMB	HML	UMD	IMU	R ²
	7	-0.001 (0.14)	0.183 (0.00)	0.055 (0.29)	-0.034 (0.49)	-0.127 (0.07)	0.06	-0.001 (0.14)	0.181 (0.00)	0.071 (0.25)	-0.029 (0.57)	-0.124 (0.08)	-0.038 (0.62)	0.06
	8	-0.001 (0.26)	0.235 (0.00)	0.037 (0.48)	-0.096 (0.05)	-0.196 (0.00)	0.1	-0.001 (0.28)	0.231 (0.00)	0.064 (0.28)	-0.087 (0.08)	-0.19 (0.01)	-0.067 (0.37)	0.1
	9	0.001 (0.33)	0.329 (0.00)	-0.051 (0.47)	0.075 (0.26)	-0.368 (0.00)	0.19	0.001 (0.31)	0.324 (0.00)	-0.014 (0.87)	0.088 (0.20)	-0.359 (0.00)	-0.091 (0.37)	0.19
	10	0 (0.81)	0.119 (0.01)	-0.066 (0.26)	-0.064 (0.25)	-0.2 (0.01)	0.04	0 (0.78)	0.123 (0.01)	-0.099 (0.15)	-0.076 (0.18)	-0.208 (0.01)	0.081 (0.34)	0.05
	11	0 (0.46)	0.221 (0.00)	-0.046 (0.40)	-0.059 (0.25)	-0.182 (0.01)	0.1	0 (0.46)	0.221 (0.00)	-0.045 (0.47)	-0.059 (0.26)	-0.182 (0.01)	0 (1.00)	0.1
	12	-0.001 (0.36)	0.169 (0.00)	-0.035 (0.50)	-0.086 (0.08)	-0.205 (0.00)	0.08	0 (0.39)	0.162 (0.00)	0.011 (0.85)	-0.07 (0.16)	-0.195 (0.00)	-0.111 (0.13)	0.08
	Nº	α	MKT	SMB	HML	UMD	R ²	α	MKT	SMB	HML	UMD	IMU	R ²
12 Portfólios: Tamanho, Investimento e Lucratividade	1	-0.003 (0.00)	0.374 (0.00)	0.231 (0.01)	-0.091 (0.30)	-0.079 (0.51)	0.06	-0.003 (0.00)	0.375 (0.00)	0.222 (0.04)	-0.094 (0.29)	-0.081 (0.51)	0.021 (0.88)	0.06
	2	-0.001 (0.22)	0.288 (0.00)	0.197 (0.06)	0.069 (0.50)	-0.21 (0.14)	0.05	-0.001 (0.22)	0.289 (0.00)	0.194 (0.12)	0.068 (0.51)	-0.211 (0.14)	0.008 (0.96)	0.05
	3	0.001 (0.57)	0.375 (0.00)	0.046 (0.73)	-0.005 (0.97)	-0.162 (0.35)	0.04	0.001 (0.57)	0.373 (0.00)	0.059 (0.70)	0 (1.00)	-0.159 (0.36)	-0.032 (0.87)	0.04
	4	-0.002 (0.11)	0.388 (0.00)	0.229 (0.01)	-0.196 (0.02)	-0.191 (0.11)	0.07	-0.002 (0.11)	0.39 (0.00)	0.212 (0.04)	-0.201 (0.02)	-0.195 (0.10)	0.039 (0.76)	0.07
	5	-0.001 (0.15)	0.253 (0.00)	0.12 (0.09)	0.037 (0.58)	0.058 (0.53)	0.05	-0.001 (0.18)	0.24 (0.00)	0.211 (0.01)	0.069 (0.31)	0.079 (0.40)	-0.224 (0.03)	0.06
	6	-0.001 (0.22)	0.226 (0.00)	0.074 (0.20)	-0.042 (0.44)	-0.227 (0.00)	0.09	-0.001 (0.26)	0.213 (0.00)	0.165 (0.01)	-0.011 (0.85)	-0.207 (0.01)	-0.222 (0.01)	0.11
	7	0 (0.86)	0.221 (0.00)	0.081 (0.24)	-0.017 (0.79)	-0.071 (0.43)	0.04	0 (0.86)	0.222 (0.00)	0.076 (0.34)	-0.019 (0.78)	-0.072 (0.43)	0.012 (0.91)	0.04
	8	-0.001 (0.31)	0.155 (0.00)	-0.063 (0.28)	-0.094 (0.09)	-0.185 (0.02)	0.05	-0.001 (0.31)	0.156 (0.00)	-0.066 (0.33)	-0.095 (0.09)	-0.185 (0.02)	0.008 (0.93)	0.05
	9	-0.001 (0.31)	0.179 (0.00)	-0.059 (0.28)	-0.066 (0.20)	-0.144 (0.05)	0.07	-0.001 (0.29)	0.183 (0.00)	-0.089 (0.16)	-0.077 (0.15)	-0.151 (0.04)	0.074 (0.35)	0.07
	10	0 (0.57)	0.207 (0.00)	-0.017 (0.74)	-0.033 (0.51)	-0.196 (0.01)	0.1	0 (0.55)	0.209 (0.00)	-0.034 (0.58)	-0.039 (0.45)	-0.2 (0.00)	0.041 (0.60)	0.1
	11	-0.001 (0.31)	0.191 (0.00)	-0.005 (0.91)	-0.094 (0.03)	-0.147 (0.02)	0.08	-0.001 (0.32)	0.188 (0.00)	0.019 (0.72)	-0.086 (0.06)	-0.142 (0.02)	-0.059 (0.38)	0.08
	12	-0.001 (0.24)	0.188 (0.00)	-0.027 (0.63)	-0.095 (0.07)	-0.25 (0.00)	0.08	-0.001 (0.26)	0.184 (0.00)	0.006 (0.92)	-0.083 (0.12)	-0.242 (0.00)	-0.08 (0.32)	0.09
Portfólios: Tamanho e	Nº	α	MKT	SMB	HML	UMD	R ²	α	MKT	SMB	HML	UMD	IMU	R ²
	1	-0.001 (0.16)	0.276 (0.00)	0.025 (0.76)	-0.069 (0.37)	-0.355 (0.00)	0.09	-0.001 (0.21)	0.249 (0.00)	0.223 (0.02)	-0.001 (0.99)	-0.31 (0.00)	-0.481 (0.00)	0.12
	2	0	0.401	0.177	-0.053	-0.381	0.08	0	0.4	0.185	-0.05	-0.379	-0.019	0.08

	Nº	α	MKT	SMB	HML	UMD	R ²	α	MKT	SMB	HML	UMD	IMU	R ²
	3	(1.00)	(0.00)	(0.10)	(0.60)	(0.01)		(0.99)	(0.00)	(0.14)	(0.63)	(0.01)	(0.90)	
		-0.003	0.329	0.142	-0.123	-0.11	0.05	-0.003	0.332	0.119	-0.131	-0.115	0.056	0.05
		(0.00)	(0.00)	(0.11)	(0.14)	(0.35)		(0.00)	(0.00)	(0.25)	(0.13)	(0.33)	(0.66)	
	4	-0.001	0.221	-0.059	-0.107	-0.361	0.09	-0.001	0.22	-0.049	-0.104	-0.359	-0.023	0.09
		(0.15)	(0.00)	(0.41)	(0.11)	(0.00)		(0.15)	(0.00)	(0.55)	(0.13)	(0.00)	(0.82)	
	5	0	0.291	0.027	-0.109	-0.238	0.13	0	0.287	0.053	-0.1	-0.232	-0.065	0.13
		(0.42)	(0.00)	(0.63)	(0.04)	(0.00)		(0.44)	(0.00)	(0.41)	(0.07)	(0.00)	(0.42)	
	6	-0.001	0.237	-0.001	-0.084	-0.149	0.09	-0.001	0.232	0.035	-0.071	-0.141	-0.086	0.1
		(0.34)	(0.00)	(0.99)	(0.10)	(0.03)		(0.36)	(0.00)	(0.57)	(0.17)	(0.05)	(0.26)	
	7	-0.001	0.113	-0.005	-0.025	-0.142	0.03	-0.001	0.116	-0.029	-0.033	-0.148	0.059	0.03
		(0.23)	(0.02)	(0.94)	(0.67)	(0.08)		(0.22)	(0.02)	(0.68)	(0.58)	(0.07)	(0.51)	
	8	-0.001	0.181	-0.063	-0.084	-0.222	0.08	-0.001	0.178	-0.043	-0.077	-0.217	-0.049	0.08
		(0.30)	(0.00)	(0.28)	(0.13)	(0.00)		(0.31)	(0.00)	(0.53)	(0.17)	(0.00)	(0.56)	
	9	0	0.169	-0.019	-0.013	-0.138	0.07	0	0.171	-0.029	-0.017	-0.141	0.023	0.07
		(0.91)	(0.00)	(0.72)	(0.79)	(0.05)		(0.90)	(0.00)	(0.65)	(0.75)	(0.05)	(0.76)	
	10	0	0.268	0.032	-0.074	-0.324	0.15	0	0.272	0.005	-0.083	-0.33	0.065	0.15
		(0.55)	(0.00)	(0.56)	(0.16)	(0.00)		(0.57)	(0.00)	(0.94)	(0.12)	(0.00)	(0.42)	
	Nº	α	MKT	SMB	HML	UMD	R ²	α	MKT	SMB	HML	UMD	IMU	R ²
16 Portfólios: Tamanho e Lucratividade	1	-0.002	0.318	0.017	-0.103	-0.217	0.04	-0.002	0.321	-0.004	-0.11	-0.222	0.052	0.04
		(0.10)	(0.00)	(0.88)	(0.34)	(0.15)		(0.09)	(0.00)	(0.98)	(0.32)	(0.14)	(0.75)	
	2	-0.001	0.376	0.239	-0.117	-0.283	0.06	-0.001	0.377	0.232	-0.119	-0.285	0.017	0.06
		(0.26)	(0.00)	(0.02)	(0.23)	(0.03)		(0.26)	(0.00)	(0.05)	(0.23)	(0.03)	(0.91)	
	3	-0.001	0.244	0.044	0.035	-0.288	0.06	-0.001	0.245	0.037	0.032	-0.289	0.016	0.06
		(0.40)	(0.00)	(0.64)	(0.69)	(0.02)		(0.40)	(0.00)	(0.73)	(0.72)	(0.02)	(0.90)	
	4	-0.002	0.297	0.302	0.011	-0.277	0.03	-0.002	0.303	0.253	-0.006	-0.288	0.119	0.03
		(0.35)	(0.02)	(0.06)	(0.94)	(0.18)		(0.34)	(0.01)	(0.17)	(0.97)	(0.17)	(0.60)	
	5	-0.001	0.236	0.104	0.052	-0.006	0.06	-0.001	0.226	0.173	0.075	0.01	-0.168	0.07
		(0.36)	(0.00)	(0.11)	(0.40)	(0.94)		(0.39)	(0.00)	(0.02)	(0.22)	(0.91)	(0.07)	
	6	-0.001	0.262	0.161	0.029	0.056	0.05	-0.001	0.258	0.189	0.039	0.062	-0.068	0.05
		(0.32)	(0.00)	(0.02)	(0.66)	(0.54)		(0.33)	(0.00)	(0.02)	(0.57)	(0.50)	(0.50)	
7	-0.001	0.22	0.046	-0.09	-0.305	0.08	-0.001	0.21	0.119	-0.065	-0.288	-0.178	0.08	
	(0.30)	(0.00)	(0.50)	(0.16)	(0.00)		(0.34)	(0.00)	(0.13)	(0.31)	(0.00)	(0.07)		
8	-0.001	0.39	0.202	-0.111	-0.123	0.09	-0.001	0.389	0.206	-0.11	-0.123	-0.008	0.09	
	(0.16)	(0.00)	(0.01)	(0.13)	(0.22)		(0.17)	(0.00)	(0.02)	(0.14)	(0.23)	(0.94)		
9	-0.001	0.175	-0.03	-0.033	-0.036	0.04	-0.001	0.175	-0.034	-0.034	-0.037	0.009	0.04	
	(0.03)	(0.00)	(0.63)	(0.58)	(0.66)		(0.03)	(0.00)	(0.64)	(0.57)	(0.66)	(0.92)		
10	0	0.2	-0.018	-0.065	-0.168	0.08	0	0.202	-0.031	-0.069	-0.171	0.032	0.08	
	(0.95)	(0.00)	(0.74)	(0.20)	(0.02)		(0.94)	(0.00)	(0.62)	(0.18)	(0.02)	(0.68)		
11	-0.001	0.272	0.004	-0.127	-0.384	0.14	-0.001	0.271	0.014	-0.123	-0.382	-0.026	0.14	
	(0.26)	(0.00)	(0.95)	(0.03)	(0.00)		(0.26)	(0.00)	(0.84)	(0.03)	(0.00)	(0.76)		
12	-0.001	0.156	-0.046	-0.022	-0.197	0.07	-0.001	0.156	-0.051	-0.023	-0.198	0.012	0.07	
	(0.31)	(0.00)	(0.42)	(0.69)	(0.01)		(0.31)	(0.00)	(0.45)	(0.68)	(0.01)	(0.88)		

Nº	α	MKT	SMB	HML	UMD	R ²	α	MKT	SMB	HML	UMD	IMU	R ²
13	-0.001 (0.32)	0.183 (0.00)	0 (1.00)	-0.099 (0.08)	-0.225 (0.00)	0.06	-0.001 (0.34)	0.18 (0.00)	0.021 (0.76)	-0.092 (0.11)	-0.22 (0.00)	-0.051 (0.55)	0.06
14	0 (0.56)	0.255 (0.00)	-0.009 (0.87)	-0.114 (0.03)	-0.182 (0.01)	0.1	0 (0.56)	0.255 (0.00)	-0.014 (0.83)	-0.116 (0.03)	-0.183 (0.01)	0.012 (0.88)	0.1
15	-0.001 (0.30)	0.108 (0.02)	-0.063 (0.29)	-0.094 (0.10)	-0.145 (0.07)	0.03	-0.001 (0.30)	0.109 (0.02)	-0.066 (0.35)	-0.095 (0.11)	-0.145 (0.07)	0.006 (0.95)	0.03
16	0 (0.97)	0.226 (0.00)	-0.017 (0.77)	-0.087 (0.12)	-0.207 (0.01)	0.08	0 (1.00)	0.22 (0.00)	0.022 (0.75)	-0.073 (0.20)	-0.199 (0.01)	-0.096 (0.27)	0.09

Fonte - Elaborada pelo autor da dissertação.

Nota: os valores entre parênteses representam os p.valores para os coeficientes.

Para o modelo mais recente de cinco fatores de Fama e French (2015), são adicionados os fatores relacionados à lucratividade (RMW) e investimento (CMA). A tabela 28 mostra os resultados para as 62 regressões nos cinco conjuntos de carteiras. Em relação ao RMW, verifica-se que, quanto maiores as empresas dentro da carteira, maior a tendência de esse fator ser negativo. Isso indica que, para essa categoria de companhias, maior lucratividade operacional não necessariamente leva a um maior retorno de suas ações, tendo em vista que elas apresentam uma sensibilidade negativa a esse fator. Esse resultado está de acordo com Fama e French (2015), à medida que se aumenta o tamanho das empresas, o retorno relacionado à lucratividade decresce.

Analisando o fator CMA, verifica-se que à medida que o tamanho das empresas aumenta, o coeficiente se torna positivo e maior, observação reforçada também pelo seu p.valor diminuindo na mesma proporção. Rejeita-se a hipótese nula em 13 dos 62 portfólios para esse fator, em contraste com a rejeição dessa hipótese em apenas quatro portfólios para o fator RMW. Para o fator SMB, rejeita-se a hipótese nula em 10 portfólios, similar ao que foi verificado para os modelos de três e quatro fatores que se valem de um fator SMB construído de forma ligeiramente diferente.

Ao se incluir o fator IMU no modelo de cinco fatores, passa-se a rejeitar a hipótese nula em 11 portfólios para o fator SMB. Para os demais fatores, não há mudança em relação a seus p.valores. O coeficiente para o IMU permanece com um sinal negativo na maioria das carteiras, assim como para os demais modelos. Em relação ao R², há um aumento de 0,071 para o modelo de cinco fatores para 0,074 para esse modelo acrescido do IMU. Novamente não se observou melhora aparente para os modelos quando há a inclusão de um fator relacionado ao risco informacional.

Tabela 28

Coeficientes para os modelos de cinco fatores e cinco fatores adicionado o IMU

	5 Fatores								5 Fatores + IMU							
	Nº	α	MKT	SMB	HML	RMW	CMA	R ²	α	MKT	SMB	HML	RMW	CMA	IMU	R ²
12 Portfólios: Tamanho e Book-to-market	1	-0.002 (0.05)	0.539 (0.00)	0.348 (0.00)	-0.325 (0.00)	-0.123 (0.39)	0.092 (0.49)	0.08	-0.002 (0.05)	0.532 (0.00)	0.384 (0.00)	-0.324 (0.00)	-0.132 (0.36)	0.102 (0.45)	-0.097 (0.54)	0.08
	2	-0.001 (0.11)	0.259 (0.00)	0.102 (0.15)	0.077 (0.26)	-0.008 (0.92)	0.067 (0.40)	0.09	-0.001 (0.12)	0.247 (0.00)	0.162 (0.04)	0.078 (0.25)	-0.024 (0.78)	0.084 (0.29)	-0.161 (0.09)	0.1
	3	-0.003 (0.00)	0.18 (0.01)	0.101 (0.29)	0.09 (0.33)	0.184 (0.12)	-0.153 (0.15)	0.03	-0.003 (0.00)	0.173 (0.01)	0.133 (0.21)	0.09 (0.33)	0.175 (0.14)	-0.144 (0.18)	-0.085 (0.50)	0.03
	4	-0.001 (0.67)	0.356 (0.00)	0.222 (0.07)	0.059 (0.61)	0.053 (0.72)	-0.063 (0.64)	0.05	-0.001 (0.67)	0.354 (0.00)	0.229 (0.09)	0.059 (0.61)	0.051 (0.73)	-0.061 (0.66)	-0.019 (0.90)	0.05
	5	-0.001 (0.19)	0.228 (0.00)	-0.016 (0.78)	-0.103 (0.06)	-0.063 (0.38)	0.13 (0.05)	0.08	-0.001 (0.20)	0.221 (0.00)	0.018 (0.78)	-0.103 (0.07)	-0.071 (0.32)	0.14 (0.03)	-0.091 (0.24)	0.08
	6	0 (0.66)	0.206 (0.00)	0.012 (0.79)	-0.003 (0.95)	-0.024 (0.68)	0.058 (0.27)	0.1	0 (0.69)	0.2 (0.00)	0.038 (0.46)	-0.003 (0.95)	-0.03 (0.59)	0.065 (0.21)	-0.07 (0.26)	0.11
	7	0 (0.53)	0.268 (0.00)	0.09 (0.21)	0.058 (0.40)	0.074 (0.40)	-0.021 (0.80)	0.08	0 (0.52)	0.273 (0.00)	0.064 (0.43)	0.058 (0.40)	0.081 (0.36)	-0.028 (0.73)	0.072 (0.46)	0.08
	8	-0.002 (0.03)	0.46 (0.00)	0.232 (0.02)	0.106 (0.28)	-0.014 (0.91)	0.034 (0.77)	0.12	-0.002 (0.04)	0.443 (0.00)	0.312 (0.01)	0.106 (0.28)	-0.035 (0.78)	0.057 (0.62)	-0.216 (0.11)	0.12
	9	0 (0.60)	0.167 (0.00)	-0.094 (0.08)	-0.071 (0.17)	-0.069 (0.30)	0.153 (0.01)	0.07	0 (0.61)	0.164 (0.00)	-0.081 (0.18)	-0.071 (0.17)	-0.072 (0.28)	0.157 (0.01)	-0.035 (0.63)	0.07
	10	0 (0.64)	0.149 (0.00)	-0.038 (0.48)	-0.055 (0.29)	-0.07 (0.29)	0.122 (0.04)	0.05	0 (0.68)	0.141 (0.00)	-0.001 (0.99)	-0.055 (0.29)	-0.08 (0.23)	0.133 (0.03)	-0.1 (0.16)	0.05
	11	0 (0.52)	0.359 (0.00)	0.059 (0.43)	0.067 (0.36)	0.019 (0.84)	0.03 (0.73)	0.12	0 (0.54)	0.354 (0.00)	0.082 (0.33)	0.067 (0.36)	0.013 (0.89)	0.036 (0.67)	-0.061 (0.55)	0.12
	12	0 (0.97)	0.349 (0.00)	-0.082 (0.40)	0.193 (0.04)	0.108 (0.37)	0.175 (0.11)	0.12	0 (0.99)	0.334 (0.00)	-0.013 (0.90)	0.194 (0.04)	0.09 (0.46)	0.195 (0.08)	-0.187 (0.15)	0.12

	Nº	α	MKT	SMB	HML	RMW	CMA	R ²	α	MKT	SMB	HML	RMW	CMA	IMU	R ²
12 Portfólios: Tamanho e Investimento	1	-0.003 (0.00)	0.257 (0.00)	0.16 (0.12)	0.075 (0.45)	0.268 (0.03)	0.104 (0.36)	0.06	-0.003 (0.00)	0.265 (0.00)	0.121 (0.28)	0.074 (0.45)	0.277 (0.03)	0.093 (0.42)	0.104 (0.44)	0.06
	2	0 (0.99)	0.533 (0.00)	0.173 (0.29)	-0.228 (0.15)	-0.175 (0.39)	0.11 (0.56)	0.04	0 (1.00)	0.529 (0.00)	0.193 (0.29)	-0.228 (0.15)	-0.18 (0.37)	0.115 (0.54)	-0.053 (0.81)	0.04
	3	-0.001 (0.39)	0.35 (0.00)	0.368 (0.00)	0.026 (0.83)	-0.226 (0.13)	-0.349 (0.01)	0.05	-0.001 (0.44)	0.322 (0.00)	0.501 (0.00)	0.026 (0.82)	-0.26 (0.08)	-0.31 (0.02)	-0.357 (0.03)	0.06
	4	-0.001 (0.11)	0.213 (0.00)	0.138 (0.09)	0.134 (0.09)	0.066 (0.51)	-0.187 (0.04)	0.05	-0.001 (0.12)	0.206 (0.00)	0.171 (0.06)	0.134 (0.09)	0.057 (0.57)	-0.177 (0.06)	-0.089 (0.41)	0.05
	5	0 (0.62)	0.225 (0.00)	0.005 (0.93)	0.002 (0.98)	0.057 (0.44)	0.118 (0.08)	0.09	0 (0.62)	0.225 (0.00)	0.005 (0.94)	0.002 (0.98)	0.057 (0.44)	0.118 (0.09)	0.001 (0.99)	0.09
	6	-0.001 (0.29)	0.313 (0.00)	0.017 (0.83)	-0.091 (0.21)	-0.108 (0.24)	0.213 (0.01)	0.09	-0.001 (0.30)	0.307 (0.00)	0.045 (0.59)	-0.09 (0.22)	-0.115 (0.22)	0.222 (0.01)	-0.077 (0.44)	0.09
	7	-0.001 (0.13)	0.21 (0.00)	0.073 (0.21)	-0.054 (0.34)	-0.126 (0.07)	0.037 (0.57)	0.06	-0.001 (0.14)	0.204 (0.00)	0.099 (0.12)	-0.053 (0.34)	-0.133 (0.06)	0.044 (0.50)	-0.071 (0.35)	0.06
	8	-0.001 (0.26)	0.248 (0.00)	0.048 (0.40)	-0.045 (0.41)	-0.017 (0.80)	0.044 (0.49)	0.09	-0.001 (0.28)	0.24 (0.00)	0.085 (0.18)	-0.045 (0.41)	-0.027 (0.70)	0.055 (0.39)	-0.1 (0.19)	0.09
	9	0.001 (0.28)	0.349 (0.00)	0.013 (0.87)	0.242 (0.00)	0.135 (0.16)	0.085 (0.33)	0.17	0.001 (0.25)	0.334 (0.00)	0.083 (0.34)	0.242 (0.00)	0.116 (0.22)	0.106 (0.23)	-0.19 (0.07)	0.18
	10	0 (0.78)	0.135 (0.00)	-0.09 (0.16)	-0.033 (0.59)	-0.037 (0.64)	0.156 (0.03)	0.04	0 (0.77)	0.137 (0.00)	-0.101 (0.16)	-0.034 (0.59)	-0.034 (0.66)	0.153 (0.04)	0.029 (0.74)	0.04
	11	0 (0.43)	0.241 (0.00)	-0.063 (0.29)	-0.047 (0.42)	-0.073 (0.32)	0.134 (0.05)	0.1	0 (0.45)	0.237 (0.00)	-0.045 (0.50)	-0.046 (0.42)	-0.077 (0.29)	0.139 (0.04)	-0.048 (0.54)	0.1
	12	-0.001 (0.33)	0.186 (0.00)	-0.048 (0.39)	-0.049 (0.36)	-0.049 (0.48)	0.107 (0.09)	0.06	-0.001 (0.37)	0.174 (0.00)	0.008 (0.89)	-0.049 (0.37)	-0.063 (0.36)	0.123 (0.05)	-0.152 (0.04)	0.07
12 Portfólios: Tamanho, Investimento e Lucratividade	1	-0.003 (0.00)	0.34 (0.00)	0.181 (0.07)	-0.081 (0.40)	0.104 (0.40)	0.17 (0.13)	0.06	-0.003 (0.00)	0.339 (0.00)	0.184 (0.10)	-0.081 (0.40)	0.103 (0.40)	0.171 (0.13)	-0.009 (0.95)	0.06
	2	-0.001 (0.22)	0.243 (0.00)	0.172 (0.14)	0.211 (0.06)	0.278 (0.05)	0.122 (0.35)	0.06	-0.001 (0.23)	0.241 (0.00)	0.18 (0.16)	0.211 (0.06)	0.276 (0.05)	0.124 (0.35)	-0.022 (0.88)	0.06
	3	0.001 (0.55)	0.44 (0.00)	0.166 (0.25)	-0.012 (0.93)	-0.226 (0.20)	-0.163 (0.31)	0.05	0.001 (0.54)	0.435 (0.00)	0.19 (0.24)	-0.012 (0.93)	-0.232 (0.19)	-0.156 (0.34)	-0.066 (0.73)	0.05
	4	-0.002 (0.10)	0.389 (0.00)	0.193 (0.05)	-0.223 (0.02)	-0.076 (0.52)	0.198 (0.07)	0.07	-0.002 (0.10)	0.388 (0.00)	0.199 (0.07)	-0.223 (0.02)	-0.078 (0.52)	0.2 (0.07)	-0.017 (0.90)	0.07
	5	-0.001 (0.17)	0.263 (0.00)	0.176 (0.02)	0.004 (0.95)	-0.061 (0.52)	-0.128 (0.14)	0.05	-0.001 (0.19)	0.247 (0.00)	0.255 (0.00)	0.005 (0.95)	-0.082 (0.39)	-0.105 (0.23)	-0.214 (0.04)	0.06
	6	-0.001 (0.24)	0.219 (0.00)	0.109 (0.08)	0.102 (0.09)	0.144 (0.06)	-0.046 (0.51)	0.08	-0.001 (0.28)	0.201 (0.00)	0.196 (0.00)	0.102 (0.09)	0.121 (0.11)	-0.021 (0.77)	-0.236 (0.00)	0.1
	7	0 (0.83)	0.231 (0.00)	0.061 (0.42)	-0.062 (0.39)	-0.1 (0.28)	0.101 (0.23)	0.05	0 (0.83)	0.23 (0.00)	0.067 (0.42)	-0.062 (0.39)	-0.101 (0.27)	0.103 (0.23)	-0.016 (0.87)	0.05
	8	-0.001 (0.28)	0.169 (0.00)	-0.097 (0.12)	-0.073 (0.23)	-0.043 (0.58)	0.169 (0.02)	0.05	-0.001 (0.29)	0.166 (0.00)	-0.082 (0.25)	-0.073 (0.23)	-0.047 (0.55)	0.173 (0.02)	-0.041 (0.63)	0.05

	Nº	α	MKT	SMB	HML	RMW	CMA	R ²	α	MKT	SMB	HML	RMW	CMA	IMU	R ²
	9	-0.001 (0.30)	0.21 (0.00)	-0.06 (0.31)	-0.085 (0.14)	-0.13 (0.08)	0.107 (0.11)	0.07	-0.001 (0.29)	0.212 (0.00)	-0.07 (0.30)	-0.085 (0.14)	-0.128 (0.08)	0.104 (0.13)	0.025 (0.75)	0.07
	10	0 (0.55)	0.215 (0.00)	-0.031 (0.59)	0.015 (0.79)	0.005 (0.94)	0.122 (0.06)	0.09	0 (0.55)	0.215 (0.00)	-0.03 (0.65)	0.015 (0.79)	0.005 (0.95)	0.123 (0.07)	-0.004 (0.95)	0.09
	11	-0.001 (0.29)	0.205 (0.00)	-0.002 (0.96)	-0.066 (0.18)	-0.039 (0.54)	0.049 (0.40)	0.07	-0.001 (0.32)	0.199 (0.00)	0.03 (0.60)	-0.066 (0.18)	-0.047 (0.46)	0.058 (0.32)	-0.087 (0.20)	0.08
	12	-0.001 (0.21)	0.207 (0.00)	-0.049 (0.42)	-0.052 (0.37)	-0.061 (0.41)	0.133 (0.05)	0.07	-0.001 (0.24)	0.198 (0.00)	-0.002 (0.97)	-0.052 (0.37)	-0.074 (0.32)	0.146 (0.03)	-0.125 (0.12)	0.07
	Nº	α	MKT	SMB	HML	RMW	CMA	R ²	α	MKT	SMB	HML	RMW	CMA	IMU	R ²
10 Portfólios: Tamanho e VPIN	1	-0.001 (0.18)	0.289 (0.00)	0.075 (0.39)	0.102 (0.23)	0.109 (0.31)	-0.005 (0.96)	0.07	-0.001 (0.23)	0.247 (0.00)	0.273 (0.00)	0.104 (0.22)	0.058 (0.58)	0.052 (0.60)	-0.531 (0.00)	0.11
	2	0 (1.00)	0.388 (0.00)	0.17 (0.15)	0.095 (0.40)	0.152 (0.29)	0.157 (0.24)	0.08	0 (0.98)	0.381 (0.00)	0.202 (0.12)	0.095 (0.40)	0.144 (0.32)	0.166 (0.21)	-0.086 (0.59)	0.08
	3	-0.003 (0.00)	0.335 (0.00)	0.195 (0.04)	-0.078 (0.40)	0.018 (0.88)	-0.072 (0.51)	0.05	-0.003 (0.00)	0.338 (0.00)	0.18 (0.10)	-0.078 (0.40)	0.022 (0.85)	-0.076 (0.49)	0.042 (0.74)	0.05
	4	-0.001 (0.16)	0.253 (0.00)	-0.016 (0.83)	0.027 (0.72)	-0.008 (0.94)	0.006 (0.94)	0.06	-0.001 (0.16)	0.248 (0.00)	0.009 (0.91)	0.027 (0.72)	-0.014 (0.88)	0.014 (0.88)	-0.07 (0.50)	0.06
	5	-0.001 (0.40)	0.306 (0.00)	0.012 (0.84)	-0.072 (0.22)	-0.039 (0.61)	0.143 (0.04)	0.12	0 (0.42)	0.296 (0.00)	0.057 (0.40)	-0.072 (0.23)	-0.05 (0.50)	0.155 (0.03)	-0.12 (0.14)	0.12
	6	-0.001 (0.32)	0.25 (0.00)	-0.013 (0.83)	-0.069 (0.22)	-0.048 (0.50)	0.096 (0.14)	0.09	-0.001 (0.34)	0.241 (0.00)	0.033 (0.61)	-0.069 (0.22)	-0.059 (0.41)	0.109 (0.10)	-0.122 (0.12)	0.09
	7	-0.001 (0.22)	0.154 (0.00)	0.019 (0.77)	-0.057 (0.38)	-0.185 (0.02)	0.039 (0.61)	0.03	-0.001 (0.22)	0.155 (0.00)	0.013 (0.86)	-0.057 (0.38)	-0.183 (0.03)	0.037 (0.63)	0.018 (0.84)	0.03
	8	-0.001 (0.29)	0.202 (0.00)	-0.073 (0.25)	-0.043 (0.49)	-0.044 (0.57)	0.129 (0.07)	0.07	-0.001 (0.31)	0.194 (0.00)	-0.035 (0.62)	-0.042 (0.49)	-0.054 (0.49)	0.14 (0.05)	-0.102 (0.23)	0.07
	9	0 (0.86)	0.183 (0.00)	-0.058 (0.32)	-0.032 (0.56)	-0.079 (0.27)	0.182 (0.01)	0.08	0 (0.87)	0.181 (0.00)	-0.048 (0.46)	-0.032 (0.57)	-0.082 (0.25)	0.185 (0.01)	-0.027 (0.73)	0.08
	10	0 (0.56)	0.29 (0.00)	0.053 (0.39)	0.019 (0.74)	-0.004 (0.96)	0.085 (0.22)	0.12	0 (0.56)	0.29 (0.00)	0.051 (0.46)	0.019 (0.74)	-0.003 (0.96)	0.085 (0.23)	0.005 (0.95)	0.12
16 Portfólios: Tamanho e Lucratividade	1	-0.002 (0.10)	0.339 (0.00)	0.05 (0.69)	-0.042 (0.72)	-0.025 (0.87)	0.022 (0.88)	0.04	-0.002 (0.10)	0.34 (0.00)	0.046 (0.74)	-0.042 (0.72)	-0.024 (0.88)	0.021 (0.88)	0.011 (0.95)	0.04
	2	-0.001 (0.29)	0.377 (0.00)	0.342 (0.00)	0.05 (0.64)	0.153 (0.26)	-0.141 (0.26)	0.06	-0.001 (0.29)	0.376 (0.00)	0.346 (0.01)	0.05 (0.64)	0.152 (0.27)	-0.14 (0.27)	-0.009 (0.95)	0.06
	3	-0.001 (0.45)	0.219 (0.00)	0.096 (0.34)	0.276 (0.00)	0.34 (0.01)	-0.031 (0.79)	0.07	-0.001 (0.46)	0.218 (0.00)	0.102 (0.36)	0.276 (0.00)	0.339 (0.01)	-0.029 (0.80)	-0.017 (0.90)	0.07
	4	-0.002 (0.35)	0.231 (0.06)	0.26 (0.13)	0.201 (0.22)	0.363 (0.08)	0.132 (0.49)	0.03	-0.002 (0.34)	0.239 (0.05)	0.224 (0.24)	0.201 (0.22)	0.373 (0.08)	0.122 (0.53)	0.099 (0.67)	0.03
	5	-0.001 (0.39)	0.259 (0.00)	0.186 (0.01)	0.042 (0.53)	-0.078 (0.36)	-0.159 (0.04)	0.07	-0.001 (0.43)	0.246 (0.00)	0.248 (0.00)	0.042 (0.53)	-0.094 (0.27)	-0.141 (0.07)	-0.169 (0.07)	0.08
	6	-0.001	0.264	0.153	-0.064	-0.105	0.057	0.05	-0.001	0.258	0.184	-0.064	-0.113	0.066	-0.083	0.06

Nº	α	MKT	SMB	HML	RMW	CMA	R ²	α	MKT	SMB	HML	RMW	CMA	IMU	R ²
	(0.31)	(0.00)	(0.04)	(0.38)	(0.26)	(0.51)		(0.32)	(0.00)	(0.03)	(0.38)	(0.23)	(0.44)	(0.41)	
7	-0.001	0.225	0.082	0.063	0.11	-0.013	0.06	-0.001	0.208	0.16	0.064	0.09	0.01	-0.209	0.07
	(0.32)	(0.00)	(0.26)	(0.37)	(0.22)	(0.88)		(0.35)	(0.00)	(0.05)	(0.37)	(0.32)	(0.91)	(0.03)	
8	-0.001	0.383	0.165	-0.14	-0.032	0.182	0.09	-0.001	0.379	0.186	-0.14	-0.038	0.188	-0.055	0.09
	(0.15)	(0.00)	(0.05)	(0.08)	(0.75)	(0.05)		(0.16)	(0.00)	(0.05)	(0.08)	(0.72)	(0.05)	(0.62)	
9	-0.001	0.191	0.005	-0.022	-0.037	-0.056	0.04	-0.001	0.192	0.004	-0.022	-0.037	-0.056	0.002	0.04
	(0.03)	(0.00)	(0.94)	(0.74)	(0.66)	(0.47)		(0.03)	(0.00)	(0.96)	(0.74)	(0.66)	(0.47)	(0.98)	
10	0	0.214	-0.024	-0.035	-0.031	0.094	0.07	0	0.214	-0.021	-0.035	-0.032	0.095	-0.008	0.07
	(0.93)	(0.00)	(0.68)	(0.54)	(0.66)	(0.15)		(0.94)	(0.00)	(0.75)	(0.54)	(0.66)	(0.15)	(0.92)	
11	-0.001	0.29	0.02	0.012	0.044	0.087	0.1	-0.001	0.283	0.052	0.012	0.036	0.096	-0.085	0.11
	(0.26)	(0.00)	(0.76)	(0.85)	(0.59)	(0.25)		(0.27)	(0.00)	(0.49)	(0.85)	(0.66)	(0.20)	(0.34)	
12	-0.001	0.168	-0.082	0.001	-0.035	0.183	0.07	-0.001	0.164	-0.066	0.002	-0.039	0.188	-0.042	0.07
	(0.29)	(0.00)	(0.19)	(0.98)	(0.65)	(0.01)		(0.30)	(0.00)	(0.35)	(0.98)	(0.61)	(0.01)	(0.61)	
13	-0.001	0.197	-0.001	-0.043	-0.012	0.098	0.05	-0.001	0.189	0.035	-0.043	-0.021	0.108	-0.097	0.05
	(0.32)	(0.00)	(0.98)	(0.49)	(0.88)	(0.18)		(0.34)	(0.00)	(0.63)	(0.49)	(0.79)	(0.14)	(0.26)	
14	0	0.297	-0.006	-0.152	-0.211	0.093	0.1	0	0.294	0.008	-0.152	-0.214	0.097	-0.037	0.1
	(0.51)	(0.00)	(0.93)	(0.01)	(0.01)	(0.18)		(0.52)	(0.00)	(0.90)	(0.01)	(0.00)	(0.16)	(0.65)	
15	-0.001	0.124	-0.099	-0.093	-0.079	0.132	0.03	-0.001	0.122	-0.09	-0.093	-0.081	0.135	-0.026	0.03
	(0.27)	(0.01)	(0.13)	(0.14)	(0.33)	(0.07)		(0.27)	(0.01)	(0.22)	(0.14)	(0.31)	(0.07)	(0.77)	
16	0	0.239	-0.031	-0.045	-0.023	0.122	0.08	0	0.228	0.021	-0.044	-0.036	0.137	-0.142	0.08
	(0.94)	(0.00)	(0.63)	(0.48)	(0.77)	(0.10)		(0.99)	(0.00)	(0.77)	(0.48)	(0.65)	(0.06)	(0.10)	

Fonte - Elaborada pelo autor da dissertação.

Nota: os valores entre parênteses representam os p.valores para os coeficientes.

Após analisar cada modelo separadamente, faz-se necessária a comparação do desempenho de cada um deles. Tal procedimento foi feito com base na metodologia adotada por Fama e French (1996, 2015). Utilizou-se o teste de Gibbons, Ross e Shanken (1989) que visa testar se um conjunto de portfólios apresenta, no geral, interceptos estatisticamente não diferentes de zero. Aplicou-se esse teste para os seis modelos analisados. Os resultados são apresentados na tabela 29.

Tabela 29

Teste GRS para os modelos de três, quatro e cinco fatores acrescidos do fator IMU

Portfólios testados	12 Portfólios: Tam. e BM		12 Portfólios: Tam. e Inv.		12 Portfólios: Tam., Inv. e Luc.		10 Portfólios: Tam. e VPIN		16 Portfólios: Tam. e Luc.	
	GRS	p.valor	GRS	p.valor	GRS	p.valor	GRS	p.valor	GRS	p.valor
3 Fatores	1,7726	0,0498	1,6089	0,0853	1,0109	0,4372	1,8052	0,0570	0,8763	0,5970
3 Fatores + IMU	1,7336	0,0568	1,6116	0,0846	0,9900	0,4571	1,7529	0,0666	0,8844	0,5876
4 Fatores	1,7895	0,0470	1,6112	0,0847	1,0108	0,4372	1,8050	0,0571	0,8762	0,5973
4 Fatores + IMU	1,7554	0,0528	1,6121	0,0845	0,9913	0,4558	1,7538	0,0664	0,8854	0,5864
5 Fatores	1,7754	0,0493	1,6559	0,0734	1,0258	0,4232	1,7715	0,0630	0,8589	0,6176
5 Fatores + IMU	1,7459	0,0545	1,6699	0,0701	1,0076	0,4403	1,7318	0,0708	0,8677	0,6072

Fonte - Elaborada pelo autor da dissertação.

Nota: 'Tam.' significa Tamanho. 'BM', *book-to-market*, 'Inv.', investimento, 'Luc.', lucratividade, 'GRS', Gibbons, Ross e Shanken.

O objetivo do teste GRS, segundo Fama e French (2015), é verificar qual combinação de fatores explica melhor os retornos dos portfólios. Dessa forma, busca-se aquele modelo que minimiza a estatística de teste e apresenta o maior p.valor. O teste apresenta hipótese nula de que os interceptos do conjunto de carteiras não são estatisticamente diferentes de zero.

Analisando a tabela 29, verifica-se que, para os portfólios construídos com base no tamanho e *book-to-market*, a adição do fator IMU aos modelos de três, quatro e cinco fatores leva a uma melhora em seus desempenhos. Para as carteiras formadas pelo tamanho e investimento das empresas, a adição do IMU não acarreta diferenças significativas em relação aos modelos tradicionais. Um detalhe a ser destacado é que o modelo de cinco fatores apresentou pior desempenho do que os de três e de quatro fatores para esse conjunto de portfólios.

Para o conjunto construído com base em tamanho, lucratividade e investimento, verifica-se uma significativa melhora do desempenho dos modelos em relação aos conjuntos analisados anteriormente. Nesse caso, novamente, a adição do fator IMU incorre em uma pequena melhora na explicação dos retornos. Para esses portfólios, o modelo com melhor performance foi o de quatro fatores adicionado o IMU.

Quando analisadas as carteiras construídas a partir do tamanho e VPIN das ações, a adição do fator IMU implica melhora dos modelos. Neste caso, o modelo que apresentou melhor estatística de teste foi o de cinco fatores com a adição do IMU. É de interesse notar

que, apesar disso, o desempenho dos modelos foi inferior em comparação com os demais conjuntos de carteiras – exceto para investimento e *book-to-market*.

Finalmente, para os portfólios formados com base no tamanho e na lucratividade, os modelos apresentaram os melhores desempenhos entre os cinco conjuntos. A adição do fator IMU em cada um dos modelos nesse caso prejudicou as suas performances, elevando a estatística de teste e diminuindo o p.valor. O modelo de cinco fatores de Fama e French (2015) foi o que apresentou melhor desempenho dentre aqueles analisados.

O que se pode verificar é que não existe um modelo que obteve melhor desempenho independentemente do conjunto analisado. Fama e French (2015) mostram que o modelo de cinco fatores apresenta melhora substancial frente ao modelo de três fatores para os sete conjuntos de carteiras por eles analisados. No presente trabalho, o modelo de cinco fatores foi superior naqueles conjuntos criados a partir das variáveis tamanho e VPIN e tamanho e lucratividade. A formação dos portfólios com base na lucratividade apresentou o melhor grau de explicação por parte dos modelos em geral. Quando formados pelo *book-to-market*, investimento ou VPIN, os modelos não apresentaram desempenhos satisfatórios.

As estimativas em relação à performance dos modelos em portfólios criados com base no tamanho e na lucratividade das ações estão em concordância com o apresentado por Fama e French (2015). Para os autores, as 25 carteiras formadas com base nessas duas variáveis foram as mais bem explicadas pelos modelos, tendo se destacado o modelo de cinco fatores. Quando analisadas as 32 carteiras formadas pelo tamanho, investimento e lucratividade, os modelos também apresentaram bons desempenhos, com destaque, novamente, para o modelo de cinco fatores, assim como apresentado no presente trabalho.

Por meio dos resultados apresentados pelo teste GRS há indícios de que um fator relacionado ao risco informacional, quando adicionado aos tradicionais modelos de fatores, seja apropriado na explicação dos retornos. De forma a analisar mais profundamente essa hipótese, procede-se à regressão dos fatores com base nos trabalhos de Mohanram e Rajgopal (2009) e Fama e French (2015).

4.4.4 Análise das relações entre os fatores de risco sistemático

Seguindo os procedimentos de Mohanram e Rajgopal (2009) e Fama e French (2015) e com base nas evidências apresentadas nas tabelas acima, os fatores foram regredidos entre eles para verificar se algum deles já se apresenta incorporado aos demais. Esse procedimento, realizado pelos autores supracitados, visa testar se os interceptos das regressões são estatisticamente diferentes de zero. Um intercepto igual a zero significaria que o fator não é precificado e que seu poder preditivo já estaria incorporado aos fatores existentes. Portanto, segundo Mohanram e Rajgopal (2009), um intercepto significativamente diferente de zero sugeriria que há um prêmio pela exposição ao risco referente ao fator testado além do que é captado pela correlação deste com os demais fatores.

O principal intuito desta seção é verificar se o fator IMU já está incorporado aos demais fatores assim como dito por Hughes *et al.* (2007). No entanto, devido às evidências sugeridas pelas regressões nos conjuntos de carteiras e por Fama e French (2015), procedeu-se às regressões de todos os fatores.

A tabela 30 mostra as estimativas para as regressões dos três fatores de Fama e French (1993) e o fator IMU. O fator MKT apresenta intercepto zero quando regredido com os fatores SMB, HML e, posteriormente, o IMU. Isto é possivelmente devido à correlação moderada do MKT com esses fatores evidenciados na tabela 22. Os p.valores para as rejeições das hipóteses nulas referentes ao intercepto não são significativamente altos. Além disso, as evidências das regressões dos retornos dos portfólios mostram que o MKT é aquele que apresentou maior poder de explicação. Dessa forma, acredita-se que os outros fatores não o capturam de forma completa. Isso será evidenciado mais adiante.

O fator SMB apresenta intercepto estatisticamente diferente de zero, ou seja, os demais fatores não o incorporam em seus retornos. Quando se adiciona o IMU, verifica-se substancial aumento do R^2 da regressão. Isto provavelmente se deve à forte correlação positiva de 0,56 entre os dois fatores. No entanto, apesar desse aumento no poder de explicação da regressão, o fator SMB continua não sendo capturado pelos demais.

Analisando o fator HML, este é o que apresenta maior p.valor para seus interceptos. Os fatores MKT e IMU são os que mostram maior poder de explicação sobre o HML. Fama e French (2015) mostram que esse fator se torna redundante para o modelo de cinco fatores. Os

resultados da tabela 30 sugerem que, mesmo para o modelo de três fatores, o HML não se apresenta como fator relevante, haja vista a sua incorporação especialmente pelo fator de mercado. Evidências utilizando o teste de GRS mostram que a exclusão do HML do modelo de três fatores, deixando somente o fator de mercado e o SMB, aumenta o poder de explicação do modelo. Esse resultado será evidenciado na seção seguinte.

Para o fator IMU não rejeita-se a hipótese nula a respeito de seu intercepto. Este aparenta ser incorporado pelos demais fatores, em especial, o SMB. Essa evidência vai contra o que foi encontrado por Mohanram e Rajgopal (2009), cujos resultados apontam para interceptos diferentes de zero no modelo de três fatores.

Tabela 30

Regressão dos três fatores de Fama e French (1993) e o IMU

	α	MKT	SMB	HML	IMU	R^2
MKT	-0,0008		-0,5627	0,3885		0,3067
	(0,148)		(0)	(0)		
	-0,0008		-0,4674	0,4062	-0,2183	0,3149
	(0,188)		(0)	(0)	(0)	
SMB	-0,001	-0,3418		0,0515		0,1998
	(0,003)	(0)		(0,126)		
	-0,001	-0,2146		-0,026	0,6349	0,4023
	(0,002)	(0)		(0,412)	(0)	
HML	-0,004	0,3539	0,0827			0,1454
	(0,427)	(0)	(0,127)			
	-0,005	0,3623	-0,0505		0,3253	0,1732
	(0,334)	(0)	(0,413)		(0)	
IMU	0	-0,0633	0,4004	0,105		0,3387
	(0,314)	(0)	(0)	(0)		

Fonte - Elaborada pelo autor da dissertação.

Nota: os valores entre parênteses representam os p.valores para os coeficientes.

Na análise das regressões dos quatro fatores de Carhart (1997) evidenciados na tabela 31, verifica-se que os p.valores dos interceptos do fator HML aumentaram com a inclusão do UMD. Além disso, parte do que era capturado pelo fator SMB foi incorporado pelo fator de momento, que apresenta maior coeficiente absoluto na explicação do HML.

A respeito do próprio fator UMD, este apresentou os maiores p.valores para seus interceptos, de 0,869 e 0,95, sendo este último relativo à adição do fator IMU. Isso significa que há fortes evidências de que os outros fatores capturam por completo o fator UMD. Fama e French (2015) comentam a respeito da inclusão de um fator de momento e afirmam que seus

coeficientes são próximos de zero e não ajudam a explicar os retornos das carteiras. Esse não foi o caso observado neste trabalho. Na seção seguinte, será evidenciado que a inclusão do fator UMD produz melhores resultados para o teste GRS. Em relação ao fator IMU, o UMD pouco tem a acrescentar na explicação do risco informacional em comparação com os resultados para a regressão desse fator com os três fatores de Fama e French (1993).

Tabela 31

Regressão dos quatro fatores de Carhart (1997) e o IMU

	α	MKT	SMB	HML	UMD	IMU	R ²
MKT	0		-0,5776	0,2938	-0,2133		0,3183
	(0,159)		(0)	(0)	(0)		
	0		-0,4928	0,3177	-0,2108	-0,1913	0,3242
	(0,195)		(0)	(0)	(0,004)	(0,019)	
SMB	-0,001	-0,35		-0,0181	-0,1906		0,215
	(0,004)	(0)		(0,666)	(0)		
	-0,001	-0,2227		-0,1033	-0,2002	0,6376	0,4197
	(0,003)	(0)		(0,005)	(0)	(0)	
HML	-0,0002	0,1976	-0,0201		-0,7281		0,3795
	(0,556)	(0)	(0,667)		(0)		
	-0,0003	0,2064	-0,1485		-0,7251	-0,3144	0,4059
	(0,433)	(0)	(0)		(0)	(0)	
UMD	0	-0,0807	-0,1099	-0,3781			0,3451
	(0,869)	(0)	(0)	(0)			
	0	-0,0736	-0,1547	-0,39		0,1119	0,3507
	(0,950)	(0,004)	(0)	(0)		(0,02)	
IMU	0	-0,0557	0,4111	0,141	0,0933		0,3443
	(0,320)	(0)	(0)	(0)	(0,020)		

Fonte - Elaborada pelo autor da dissertação.

Nota: os valores entre parênteses representam os p.valores para os coeficientes.

Finalmente, em relação às regressões dos dois fatores adicionados por Fama e French (2015), apresentados na tabela 32, tem-se que o HML é, novamente, capturado pelos demais fatores. A mudança agora se dá no fato de que, com a inclusão do RMW e CMA, o IMU deixa de ser relevante na explicação do HML evidenciado pelo p.valor de 0,936 para seu coeficiente.

Os fatores RMW e CMA apresentaram interceptos próximos de zero e com alto p.valor. Essa evidência, surpreendentemente, indica que os demais fatores capturam os fatores relacionados à lucratividade e ao investimento propostos por Fama e French (2015). Os resultados da tabela 29 indicavam que, como regra geral, a inclusão destes não produz

melhora na explicação dos retornos, fato este evidenciado pelos desempenhos similares dos modelos de três, quatro e cinco fatores. Fama e French (2015) encontram substancial melhora na performance do modelo de cinco fatores na explicação dos retornos dos portfólios em comparação com o modelo de três fatores, mas esse resultado não foi corroborado para a amostra analisada neste trabalho.

Portanto, a partir das estimativas das regressões dos fatores apresentadas nas tabelas 30, 31 e 32, verifica-se que os dois principais fatores responsáveis pela explicação dos retornos das carteiras analisadas são o MKT e o SMB, sendo os demais incorporados à medida que outros fatores são adicionados às regressões. De forma a verificar qual combinação de fatores resulta no melhor modelo, realiza-se o procedimento estabelecido por Fama e French (2015) de análise dos resultados do teste GRS para diferentes arranjos de modelos.

Tabela 32

Regressão dos cinco fatores de Fama e French (2015) e o IMU

	α	MKT	SMB	HML	RMW	CMA	IMU	R ²
MKT	0,002 (0,261)		-5294 (0)	0,5933 (0)	0,4237 (0)	0,0687 (0,324)		0,3213
	0,001 (0,326)		-0,4199 (0)	0,5817 (0)	0,4188 (0)	0,0958 (0,168)	-0,2652 (0)	0,334
SMB	-0,001 (0,004)	-0,2722 (0)		0,0517 (0,227)	0,0253 (0,641)	0,4315 (0)		0,2739
	-0,001 (0,003)	-0,1769 (0)		0,04 (0,296)	0,0413 (0,145)	0,289 (0)	0,5318 (0)	0,4163
HML	-0,0003 (0,463)	0,3266 (0)	0,0553 (0,227)		-0,7332 (0)	0,4262 (0)		0,4945
	-0,0003 (0,462)	0,3269 (0)	0,0535 (0,296)		-0,7327 (0)	0,4257 (0)	0,0049 (0,936)	0,4935
RMW	-0,0001 (0,593)	0,1551 (0)	0,0168 (0,641)	-0,4554 (0)		0,2105 (0)		0,3295
	-0,0001 (0,662)	0,1446 (0)	0,0585 (0,145)	-0,4502 (0)		0,2203 (0)	-0,1126 (0,018)	0,3354
CMA	0,0002 (0,593)	0,0277 (0,324)	0,3392 (0)	0,313 (0)	0,2489 (0)			0,2959
	0,0001 (0,678)	0,0389 (0,168)	0,2786 (0)	0,3076 (0)	0,2591 (0)		0,1483 (0,004)	0,3058
IMU	0,0002 (0,364)	-0,078 (0)	0,3715 (0)	0,002 (0,935)	-0,096 (0,018)	0,1075 (0,004)		0,3471

Fonte - Elaborada pelo autor da dissertação.

Nota: os valores entre parênteses representam os p.valores para os coeficientes.

4.4.5 Teste GRS aplicado em diferentes combinações de modelos

Fama e French (2015) mostram que o fator HML se torna redundante no contexto do modelo de cinco fatores. Para encontrar o modelo que melhor explica os retornos dos portfólios, os autores testam diferentes combinações de fatores. O resultado de Fama e French (2015) indica que o modelo de quatro fatores – MKT, SMB, RMW e CMA – apresenta um desempenho igual ou levemente superior àquele que inclui o HML, indicando, portanto, que não é necessária a inclusão deste para a explicação dos retornos.

Realizou-se o mesmo procedimento de Fama e French (2015) com o objetivo de verificar qual combinação de fatores produz o melhor modelo para a amostra analisada. Foram feitas alterações entre as variáveis dos modelos de três, quatro e cinco fatores. Os resultados são apresentados na tabela 33.

Antes de analisar qual modelo obteve melhor desempenho em cada conjunto de carteiras, atenta-se ao fato de que, para todos os modelos, a retirada do fator HML incorreu em uma melhora na performance na explicação dos retornos. Essa melhora é verificada, tanto para os modelos de três, quanto de quatro e cinco fatores. A única exceção se dá para o conjunto de portfólios formado por tamanho e lucratividade, no qual a exclusão desse fator não acarretou em mudanças significativas para o teste GRS.

Este resultado está em parte de acordo com o reportado por Fama e French (2015). Os autores afirmam que o modelo de cinco fatores apresenta o mesmo desempenho quando excluído o HML, sendo este, portanto, um fator redundante. A partir dos resultados da tabela 33, percebe-se que essa análise pode ser estendida para os modelos de três e quatro fatores, levando, portanto, à conclusão de que o fator HML não só é redundante como também é prejudicial na performance dos modelos para a amostra analisada.

Tabela 33

Teste GRS para as variações dos modelos de três, quatro e cinco fatores e o fator IMU

Portfólios testados		12 Portfólios: Tam. e BM		12 Portfólios: Tam. e Inv.		12 Portfólios: Tam., Inv. e Luc.		10 Portfólios: Tam. e VPIN		16 Portfólios: Tam. e Luc.	
Modelos		GRS	p.valor	GRS	p.valor	GRS	p.valor	GRS	p.valor	GRS	p.valor
3 Fatores + IMU	MKT, SMB, HML	1,7726	0,0498	1,6089	0,0853	1,0109	0,4372	1,8052	0,0570	0,8763	0,5970
	MKT, SMB	1,7465	0,0543	1,5668	0,0975	0,9892	0,4578	1,7639	0,0645	0,8770	0,5962
	MKT, SMB, HML, IMU	1,7336	0,0568	1,6116	0,0846	0,9900	0,4571	1,7529	0,0666	0,8844	0,5876
	MKT, SMB, IMU	1,7096	0,0615	1,5576	0,1004	0,9664	0,4800	1,7148	0,0744	0,8841	0,5879
4 Fatores + IMU	MKT, SMB, HML, UMD	1,7895	0,0470	1,6112	0,0847	1,0108	0,4372	1,8050	0,0571	0,8762	0,5973
	MKT, SMB, HML, UMD, IMU	1,7554	0,0528	1,6121	0,0845	0,9913	0,4558	1,7538	0,0664	0,8854	0,5864
	MKT, SMB, UMD	1,7379	0,0559	1,5881	0,0912	0,9897	0,4573	1,7688	0,0636	0,8742	0,5996
	MKT, SMB, UMD, IMU	1,7005	0,0634	1,5806	0,0934	0,9670	0,4794	1,7198	0,0734	0,8814	0,5911
5 Fatores + IMU	MKT, SMB, HML, CMA, RMW	1,7754	0,0493	1,6559	0,0734	1,0258	0,4232	1,7715	0,0630	0,8589	0,6176
	MKT, SMB, CMA, RMW	1,7533	0,0531	1,6169	0,0832	1,0118	0,4364	1,7599	0,0652	0,8597	0,6166
	MKT, SMB, HML, CMA, RMW, IMU	1,7459	0,0545	1,6699	0,0701	1,0076	0,4403	1,7318	0,0708	0,8677	0,6072
	MKT, SMB, CMA, RMW, IMU	1,7237	0,0587	1,6284	0,0802	0,9924	0,4547	1,7198	0,0734	0,8684	0,6064
	MKT, SMB, RMW	1,7699	0,0502	1,6304	0,0797	1,0010	0,4465	1,7812	0,0613	0,8677	0,6072
	MKT, SMB, RMW, IMU	1,7337	0,0567	1,6373	0,0779	0,9813	0,4655	1,7312	0,0710	0,8739	0,5999
	MKT, SMB, CMA	1,7547	0,0529	1,6181	0,0829	1,0090	0,4390	1,7615	0,0649	0,8604	0,6158
	MKT, SMB, CMA, IMU	1,7230	0,0588	1,6259	0,0808	0,9851	0,4618	1,7191	0,0735	0,8684	0,6064

Fonte - Elaborada pelo autor da dissertação.

Analisando o conjunto de carteiras formadas pelo tamanho e *book-to-market* das ações, verifica-se que, dentre os 16 arranjos de fatores analisados, o que apresentou melhor performance foi aquele com os fatores MKT, SMB, UMD e IMU. Uma questão surge quando a inclusão do UMD melhora o desempenho dos modelos, tendo em vista que ele foi o que apresentou menor intercepto e com maiores p.valores na tabela 31. A explicação repousa na relação entre o UMD e o HML. Analisando a regressão do UMD com os demais fatores, constata-se que o HML apresenta o maior coeficiente absoluto dentre as variáveis, indicando, portanto, que este seria o fator que capturaria melhor a variação dos retornos do UMD. Em resultados não tabulados, as regressões desse fator com o MKT, SMB e, posteriormente, com a adição do IMU, verifica-se substancial diminuição do p.valor do intercepto, o que sugere que esses fatores não conseguem capturar completamente o fator UMD, deixando margem para que ele tenha um papel na explicação dos retornos dos portfólios.

Constata-se também que a adição do IMU melhora o desempenho de todos os modelos analisados. Novamente deve-se buscar a explicação para essa evidência, já que, via regressões dos fatores, verificou-se que o IMU era capturado pelo SMB e pelo HML. O SMB é o fator que tem maior poder de explicação sobre o IMU. Isto se deve à alta correlação entre ambos, haja vista a forma como eles foram formados. As empresas do grupo *informed* são aquelas de menor tamanho, enquanto as do grupo *uninformed* são as maiores empresas – como foi mostrado na análise entre tamanho e VPIN –, enquanto o SMB é construído pela subtração dos retornos das empresas menores pelos retornos das empresas maiores, ou seja, ambos os fatores são construídos de forma similar. Como o IMU visa captar a parte informacional das ações e a relação com o tamanho das empresas é uma consequência de como o mercado de capitais lida com o conteúdo informacional das empresas, é possível que o IMU seja capaz de explicar parte das variações dos retornos das carteiras não captadas pelo fator SMB – a parte relacionada ao risco informacional –, o que explicaria as melhoras nos desempenhos dos modelos que incluem o fator IMU.

Para o conjunto de carteiras formadas pelo tamanho e investimento das ações, o melhor modelo foi o de três fatores, porém com a substituição do HML pelo IMU, que culmina em um modelo que se vale das variáveis MKT, SMB e IMU. Constata-se, pela tabela 30, que o IMU apresenta alto coeficiente para a explicação do fator HML. Parte, portanto, da variação do HML pode ser capturada pelo IMU, que se propõe também a

refletir o conteúdo informacional dos ativos. Os modelos de fatores que incluem o fator investimento (CMA) não apresentaram bom desempenho, tendo se mostrado inferior aos modelos de quatro fatores e suas variações.

Em relação às carteiras formadas pelo tamanho, investimento e lucratividade, o melhor modelo foi novamente o de três fatores com o IMU – MKT, SMB e IMU – seguido do modelo que inclui os fatores MKT, SMB, UMD e IMU. As explicações para as inclusões dessas variáveis são as mesmas dos conjuntos de carteiras previamente analisados. Vale destacar que os modelos que incluem justamente as variáveis CMA e RMW, relacionadas a investimento e lucratividade, tiveram desempenho inferior aos demais modelos.

Para os portfólios formados pelo tamanho e VPIN, tem-se novamente como melhor modelo a combinação das variáveis MKT, SMB e IMU, seguido por três modelos com desempenhos extremamente próximos: MKT, SMB, CMA e IMU; MKT, SMB, CMA, RMW e IMU; MKT, SMB, UMD e IMU. Este último novamente se mostra entre aqueles com performance superior. A evidência contrária ao esperado fica por conta do modelo de cinco fatores com IMU e um modelo restrito que inclui o MKT, SMB, CMA e IMU. A explicação pode residir na relação do IMU com o fatores SMB e CMA. Como exposto acima, a composição do IMU é feita de forma similar ao SMB. Foi apresentado na tabela 32 que o CMA apresenta o maior coeficiente entre os fatores na explicação do SMB. Concomitantemente, o SMB apresentou o maior coeficiente na explicação do CMA. Como portfólios criados a partir do tamanho e VPIN podem assemelhar-se àqueles relacionados ao fator SMB, a inclusão do CMA pode ocasionar em uma melhora na explicação dos retornos dos portfólios, o que foi, efetivamente, verificado por meio dos resultados da tabela 33.

Finalmente, para os portfólios formados pelo tamanho e lucratividade das ações, obteve-se resultado semelhante àquele de Fama e French (2015). Os modelos de cinco (MKT, SMB, HML, RMW e CMA) e quatro fatores (MKT, SMB, RMW e CMA) apresentaram desempenhos próximos, assim como reportado pelos autores supracitados. Além disso, por meio de todas as combinações de fatores, pode-se verificar que a exclusão do HML não gerou forte impacto no desempenho dos modelos, assim como relatado em Fama e French (2015). Pode-se destacar também dois pontos a partir da análise da aplicação dos modelos nessas carteiras. A primeira é que a inclusão das variáveis HML e RMW não produziu forte impacto no desempenho dos modelos, haja

vista a proximidade da performance do modelo de cinco fatores com modelo que utilizou os fatores MKT, SMB, RMW e CMA e o modelo que se valeu do MKT, SMB e CMA. Isto difere dos resultados de Fama e French (2015) de que o fator RMW reduz substancialmente a estatística de teste GRS para os portfólios analisados, porém reforça a conclusão acerca do HML. A outra evidência se refere à inclusão da variável IMU. Para todas as combinações de fatores, a presença do IMU prejudica o desempenho dos modelos, fato este não verificado para os outros conjuntos de carteiras previamente analisados em que o IMU auxiliou na explicação dos retornos das carteiras.

Em conclusão, de forma geral, verificou-se que as melhores combinações de fatores foram aquelas que excluíram o HML e incluíram o IMU. O modelo MKT, SMB e IMU foi o que apresentou melhor desempenho para três dos cinco conjuntos de carteiras. O modelo MKT, SMB, UMD e IMU apresentou performance superior para os portfólios formados pelo tamanho e *book-to-market*, porém demonstrou também bom desempenho nos testes com os demais conjuntos, com exceção da combinação entre tamanho e lucratividade. Para esta última, a melhor combinação foi o modelo de cinco fatores, seguido do modelo de quatro fatores (MKT, SMB, RMW e CMA). A próxima seção procura trabalhar com simulações dos retornos das carteiras para verificar qual dessas combinações de fatores melhor explica os retornos na amostra analisada.

4.4.6 Análise dos modelos através do *Bootstrap* das carteiras

Objetiva-se nesta seção destacar o melhor modelo entre as quatro seguintes combinações de fatores: MKT, SMB, IMU; MKT, SMB, UMD, IMU; MKT, SMB, HML, RMW, CMA; e MKT, SMB, RMW, CMA. Optou-se por incluir os modelos de cinco fatores de Fama e French (2015) e a sua versão restrita sem o fator HML com o objetivo de verificar, por meio de simulações, o efeito da exclusão do HML no desempenho do modelo. Em vista disso, utilizou-se o método *bootstrap* para a simulação dos retornos dos portfólios. O procedimento realizado se valeu da reamostragem com reposição das carteiras dentro de cada conjunto previamente construído. Como exemplo, tomam-se os 12 portfólios criados a partir do tamanho e *book-to-market*. Para cada um desses portfólios, formaram-se 1.000 carteiras via reamostragem com reposição de seus retornos, e sobre essas 1.000 carteiras foi aplicado o teste estatístico cuja hipótese nula era de que os interceptos desse conjunto não eram

diferentes de zero. Portanto, ao invés de se obter somente um resultado para os 12 portfólios criados pelo tamanho e *book-to-market*, obtêm-se doze resultados, um para cada conjunto de 1.000 carteiras.

Ao se realizar a reamostragem, encontra-se um problema na utilização do teste GRS. Como explicado na metodologia, esse teste requer a inversão da matriz de covariância dos resíduos das regressões, $\hat{\Sigma}^{-1}$, no entanto, quando 1000 carteiras são construídas a partir da reamostragem com reposição de um único portfólio, essa matriz se torna não inversível devido à dependência linear acarretada pela repetição dos dados originais – para reamostragens menores como, por exemplo, 50, essa matriz pode ser, na maioria das vezes, invertida, porém perde-se a eficiência proposta pelo *bootstrap*. Visando contornar esse problema, adotou-se o *Average F-test* proposto por Hwang e Satchell (2014). Este teste é uma adaptação do GRS para os casos em que não é possível realizar a inversão da matriz de covariância dos resíduos e se mostra robusto em situações em que o número de portfólios testados (N) e o período analisado (T) são significativamente maiores do que o número de fatores (K) – o que é o caso neste trabalho, já que $N = 1000$, $T = 512$ e $K = \{3,4,5\}$.

A tabela 34 mostra os resultados dos p.valores do *Average F-test* para as simulações realizadas. A aplicação do teste a partir do *bootstrap* permite a observação em separado de cada carteira dentro de cada conjunto. Assim, analisam-se quais características das ações pioram o desempenho dos modelos.

A primeira análise a ser feita via resultados se refere ao comportamento semelhante do desempenho das combinações dos modelos na medida em que as empresas dentro dos portfólios crescem. Quanto maiores as ações nas carteiras, melhor tende a ser a performance do modelo, representado pelo crescimento do p.valor referente ao *Average F-test*. Para todos os conjuntos de carteiras, os maiores p.valores se referem aos últimos portfólios, formados pelas ações com maiores valores de mercado. Assim como apontado por Fama e French (2015), as ações menores, também chamadas de *microcaps*, prejudicam as performances dos modelos, que não conseguem capturar de forma precisa os retornos das ações desta natureza.

Tabela 34

P.valores do *Average F-test* para diferentes combinações de fatores

MKT, SMB, IMU	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Tam. (3) e BM (4)	0,006	0,017	0,000	0,132	0,059	0,160	0,113	0,003	0,256	0,232	0,108	0,282				
Tam. (3) e Inv. (4)	0,000	0,217	0,084	0,024	0,178	0,069	0,027	0,063	0,299	0,276	0,136	0,124				
Tam. (2), Inv. (3) e Luc. (2)	0,000	0,037	0,317	0,017	0,025	0,037	0,240	0,126	0,120	0,151	0,097	0,080				
Tam. (3) e VPIN (4)		0,036	0,219	0,000	0,049	0,098	0,104	0,080	0,109	0,285	0,307					
Tam. (4) e Luc. (4)	0,021	0,036	0,091	0,070	0,076	0,067	0,070	0,024	0,013	0,300	0,063	0,099	0,103	0,168	0,156	0,306
MKT, SMB, UMD, IMU	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Tam. (3) e BM (4)	0,007	0,015	0,000	0,139	0,060	0,147	0,131	0,003	0,231	0,235	0,121	0,268				
Tam. (3) e Inv. (4)	0,000	0,227	0,076	0,025	0,152	0,072	0,030	0,064	0,304	0,276	0,162	0,121				
Tam. (2), Inv. (3) e Luc. (2)	0,000	0,032	0,304	0,016	0,028	0,041	0,248	0,126	0,126	0,156	0,088	0,077				
Tam. (3) e VPIN (4)		0,034	0,234	0,000	0,053	0,104	0,108	0,086	0,114	0,268	0,307					
Tam. (4) e Luc. (4)	0,025	0,042	0,098	0,075	0,076	0,058	0,076	0,022	0,013	0,297	0,060	0,114	0,105	0,169	0,133	0,277
MKT, SMB, HML, RMW, CMA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Tam. (3) e BM (4)	0,006	0,015	0,000	0,139	0,066	0,170	0,127	0,003	0,225	0,229	0,112	0,255				
Tam. (3) e Inv. (4)	0,000	0,219	0,083	0,022	0,169	0,075	0,028	0,065	0,315	0,305	0,136	0,125				
Tam. (2), Inv. (3) e Luc. (2)	0,000	0,028	0,306	0,014	0,028	0,047	0,257	0,132	0,118	0,155	0,104	0,071				
Tam. (3) e VPIN (4)		0,038	0,224	0,000	0,049	0,095	0,085	0,079	0,114	0,274	0,311					
Tam. (4) e Luc. (4)	0,021	0,049	0,092	0,069	0,084	0,059	0,072	0,021	0,014	0,284	0,065	0,109	0,113	0,179	0,156	0,264
MKT, SMB, RMW, CMA	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Tam. (3) e BM (4)	0,006	0,015	0,000	0,128	0,053	0,179	0,135	0,002	0,234	0,233	0,118	0,252				
Tam. (3) e Inv. (4)	0,000	0,220	0,079	0,022	0,170	0,079	0,026	0,061	0,297	0,301	0,148	0,134				
Tam. (2), Inv. (3) e Luc. (2)	0,000	0,035	0,302	0,014	0,027	0,040	0,223	0,135	0,126	0,163	0,092	0,074				
Tam. (3) e VPIN (4)		0,039	0,215	0,000	0,050	0,098	0,104	0,077	0,120	0,295	0,322					
Tam. (4) e Luc. (4)	0,022	0,045	0,107	0,083	0,070	0,064	0,072	0,021	0,012	0,289	0,066	0,115	0,107	0,165	0,130	0,285

Fonte - Elaborada pelo autor da dissertação.

Nota: a tabela apresenta os p.valores do *Average F-test* para a simulação das carteiras dentro de cada conjunto. A primeira e a décima segunda carteiras formadas pelo tamanho e VPIN não foram simuladas por não terem em seus portfólios o número mínimo de ações necessário estabelecido no presente trabalho.

Analisando as características individuais de cada portfólio evidenciadas na tabela 23, não foi encontrada nenhuma relação entre o desempenho dos modelos e os demais fatores formadores dos portfólios, como o número de ações em cada carteira. Dessa forma, conclui-se que a performance do *Average F-test* está diretamente relacionada somente ao tamanho das empresas que compõem a carteira analisada para a amostra em questão, tendo em vista que foi constatada uma correlação de 0,5710 entre os p.valores das simulações das carteiras e o valores de mercado das empresas, confirmando as evidências apresentadas por Fama e French (2015).

Por fim, procura-se determinar qual desses modelos apresentou melhor desempenho de forma geral. Para isso, utilizou-se o método de Fisher para combinação dos p.valores apresentados pelos modelos. A distribuição dos p.valores por meio desse método segue uma distribuição χ^2 pelo fato de eles sob a hipótese nula seguirem uma distribuição uniforme limitada entre 0 e 1. Dessa forma, a combinação dos p.valores se dá como mostrado em (62)

$$-2 \sum_{i=1}^n \ln X_i \sim \chi^2(2n) \quad (62)$$

em que $\ln X_i$ é o logaritmo natural de cada p.valor. Os resultados para as combinações dos p.valores são mostrados na tabela 35. Para cada conjunto de portfólios, um modelo superou os demais. Dessa forma, fez-se o cálculo da combinação dos p.valores das 62 simulações – relacionados às 62 carteiras dos cinco conjuntos formados – o que culminou na coluna *geral* da tabela 35. Verifica-se que o modelo MKT, SMB, RMW e CMA apresentou o maior p.valor combinado seguido do MKT, SMB, UMD e IMU. Vale notar que os valores apresentam grandezas de baixo valor devido à natureza da multiplicação de probabilidades (LI *et al.*, 2014).

Tabela 35

Resultado da combinação dos p.valores através do método de Fisher

Combinações de Fatores	Tam. (3) e BM(4)	Tam. (3) e Inv. (4)	Tam. (2), Inv. (3) e Luc. (2)	Tam. (3) e VPIN (4)	Tam. (4) e Luc. (4)	Geral
MKT, SMB, IMU	$2,06 \times 10^{-7}$	$4,22 \times 10^{-6}$	$1,38 \times 10^{-6}$	$2,91 \times 10^{-5}$	$1,92 \times 10^{-6}$	$3,55 \times 10^{-24}$
MKT, SMB, UMD, IMU	$2,11 \times 10^{-7}$	$5,51 \times 10^{-6}$	$1,16 \times 10^{-6}$	$3,08 \times 10^{-5}$	$2,09 \times 10^{-6}$	$4,55 \times 10^{-24}$
MKT, SMB, HML, RMW, CMA	$1,87 \times 10^{-7}$	$5,54 \times 10^{-6}$	$1,13 \times 10^{-6}$	$2,81 \times 10^{-5}$	$2,82 \times 10^{-6}$	$4,06 \times 10^{-24}$
MKT, SMB, RMW, CMA	$2,29 \times 10^{-7}$	$5,38 \times 10^{-6}$	$1,11 \times 10^{-6}$	$3,21 \times 10^{-5}$	$2,27 \times 10^{-6}$	$5,22 \times 10^{-24}$

Fonte - Elaborada pelo autor da dissertação.

Nota: devido à grandeza dos valores optou-se por mostrar os resultados através de notações científicas.

Verifica-se via resultados obtidos pelas simulações que, de maneira geral, o HML é um fator redundante e que não apresenta melhora significativa em comparação com o modelo restrito. O modelo MKT, SMB e IMU, apesar de ter apresentado melhores resultados para os conjuntos, como apresentado pela tabela 33, não se sustentou em relação aos demais quando aplicado aos dados simulados. A inclusão do fator de momento, resultando no modelo MKT, SMB, UMD e IMU, entretanto, conseguiu manter um bom desempenho nas simulações das carteiras, o que sugere que o fator UMD pode capturar variações dos retornos de forma geral na amostra analisada.

Finalmente, devido à evidência apresentada na tabela 33 de que a inclusão do fator IMU no lugar do fator HML leva, de forma geral, a uma melhora no desempenho dos modelos, foi analisada a performance de um quinto modelo: MKT, SMB, RMW, CMA e IMU. O resultado da combinação dos p.valores das simulações foi de $7,72 \times 10^{-24}$, ou seja, superior ao desempenho do modelo MKT, SMB, RMW e CMA de $5,22 \times 10^{-24}$ o que indica que a adição do fator relacionado ao risco informacional levou a uma melhora do modelo restrito de Fama e French (2015). Foram realizadas também variações desses modelos com o UMD, devido ao papel desse fator na melhora do modelo MKT, SMB e IMU como verificado anteriormente. No entanto, os resultados foram inferiores aos dos modelos apresentados, tendo sido constatado, inclusive, que o modelo MKT, SMB, RMW, CMA e UMD apresentou um p.valor combinado menor do que o modelo MKT, SMB e IMU. Isto aponta então que a melhora referente ao UMD não se sustenta perante o modelo com os fatores de Fama e French (2015), assim como observado pelos próprios autores.

5 CONSIDERAÇÕES FINAIS

O objetivo desta pesquisa foi: (i) analisar o VPIN ou o nível de toxicidade das ações no mercado brasileiro e (ii) verificar, via modelos de fatores propostos por Fama e French (1993, 2015) e Carhart (1997), se um fator de risco sistemático relacionado ao conteúdo informacional das ações é precificado pelos investidores da BM&FBovespa. Através de uma base de dados ainda não utilizada para pesquisas a respeito do risco informacional no cenário nacional, a qual contém todos os dados a respeito das transações ocorridas no mercado acionário brasileiro nos últimos dois anos, foi definida uma amostra de 142 ações e 512 dias de transações para o cálculo do VPIN e uma amostra irrestrita – apresentada no apêndice – para a formação dos fatores relacionados ao tamanho, *book-to-market*, momento, lucratividade e investimento.

Foi encontrado um VPIN médio de 0,4548 com um desvio-padrão igual a 0,2219 para o mercado brasileiro. Por meio do método de *bootstrap*, foi calculado um intervalo de confiança que varia em três pontos para cima e para baixo na quarta casa decimal para esses dois parâmetros. Na análise relacionada ao segmento de listagem das ações, foi verificado que o segmento N1 apresentou menor VPIN, seguido do Tradicional, NM e N2. Esperava-se que ações do NM apresentassem menor VPIN, visto que o objetivo da segmentação da Bovespa é levar maior transparência para o investidor, o que implicaria menor probabilidade de negociações privilegiadas. Os resultados sugerem que a probabilidade de negociações privilegiadas dos segmentos se relaciona com a quantidade de empresas e com as características das ações que os compõem, em especial, o tamanho das empresas. Dessa forma, a hipótese estipulada nesta pesquisa de que os segmentos teoricamente mais transparentes da BM&FBovespa apresentariam menor probabilidade de negociações privilegiadas não pôde ser confirmada.

A hipótese de que existe uma correlação negativa entre o tamanho e o VPIN (ou PIN) das ações, largamente discutida em trabalhos que se valem dessas metodologias foi confirmada. Os resultados deste trabalho indicam que há uma correlação de -0,3080 entre o tamanho e o VPIN das empresas. A amostra analisada foi dividida em três grupos relacionados ao valor de mercado das companhias: pequeno, médio e grande. Os VPINs médios para esses grupos foram de 0,6364, 0,3989 e 0,3164, respectivamente, indicando clara diminuição do VPIN na medida em que o tamanho das empresas cresce. Esses resultados são similares àqueles apresentados em estudos nos mercados espanhol,

australiano e norte-americano, indicando convergência das evidências empíricas da relação entre o VPIN e o valor de mercado das empresas.

Em relação à hipótese acerca da relação entre o VPIN e o *bid-ask spread* das ações, foi encontrada uma correlação geral positiva de 0,2672 indicando que, quanto maior o desequilíbrio entre as ordens de compra e venda, menor a liquidez das ações no mercado, fato este reportado previamente na literatura. Ao analisar ações individualmente, verificou-se que mais da metade das ações da amostra apresentaram correlação no intervalo entre $[-0.05, 0.05]$ e com p.valores iguais a 0, ou seja, estatisticamente insignificantes. Ao se analisar a regressão entre o *bid-ask spread* e o VPIN das ações, encontrou-se coeficiente positivo e estatisticamente significativo, indicando que existe, de forma geral, um poder de explicação do VPIN referente ao *bid-ask spread* das ações. Um intercepto diferente de zero encontrado pela regressão aponta que existem outros fatores, como previsto, que explicam a iliquidez das ações. Regressões individuais entre essas duas variáveis indicaram que a hipótese de que o VPIN apresenta poder de explicação sobre o *bid-ask spread* das ações somente é confirmado para aquelas que apresentaram correlação positiva entre esses dois fatores.

A última hipótese, referente ao papel de um fator relacionado ao risco informacional das ações, foi analisada por meio da construção do fator IMU, ou seja, a subtração do retorno das ações com alto VPIN e baixo VPIN. Acrescentou-se esse fator aos modelos de três e cinco fatores de Fama e French (1993, 2015) e quatro fatores de Carhart (1997) e utilizaram-se, como variáveis dependentes nas regressões, 62 portfólios construídos com base no tamanho, *book-to-market*, lucratividade, investimento e VPIN das ações. O fator MKT apresentou coeficiente positivo e estatisticamente significativo em todas as regressões. Para o modelo de três fatores, o fator SMB se mostrou significativo em parte dos portfólios, enquanto o HML somente em quatro deles. Para o coeficiente referente ao IMU, foi rejeitada a hipótese nula em cinco dos portfólios. A adição do fator IMU, no entanto, aumentou o poder preditivo do SMB na amostra analisada.

Para o modelo de quatro fatores, os resultados para o MKT, SMB, HML e IMU são similares. O fator de momento apresentou coeficiente negativo e significativo em 37 portfólios. Analisando o modelo de cinco fatores, o fator RMW tem coeficiente estatisticamente significativo em quatro portfólios enquanto o fator CMW apresenta

coeficiente significativo em 13. Foi constatado, em todos os modelos, um aumento do poder de explicação do SMB quando adicionado o fator IMU.

De forma geral, a melhora no desempenho dos modelos com a inclusão do IMU foi verificada por meio do teste GRS. Para investigar mais afundo essa afirmação, foram realizadas regressões entre os fatores. Os resultados indicam que todos os fatores, exceto o SMB e MKT – em relação ao fator de mercado as evidências empíricas indicam que este deve estar presente nos modelos – são, em algum momento, capturados pelos demais. Fama e French (2015) atentam que, no contexto do modelo de cinco fatores, o fator HML se torna redundante. De forma a verificar quais desses fatores auxiliam na explicação dos retornos dos portfólios, o teste GRS foi aplicado em diferentes combinações de fatores.

Os resultados indicam que os modelos seguintes apresentaram melhor desempenho: MKT, SMB e IMU; MKT, SMB, UMD e IMU; MKT, SMB, HML, CMA, RMW; e MKT, SMB, CMA e RMW. Além de corroborar o resultado de Fama e French (2015) de que o fator HML é redundante para o modelo de cinco fatores, estendeu-se essa afirmação para os modelos de três e quatro fatores, tendo sido constatado, inclusive, que, quando presente, o HML prejudica o desempenho dos modelos.

Para estender essas conclusões, realizou-se o procedimento de *bootstrap* dos retornos dos portfólios sendo eles regredidos com os modelos citados no parágrafo anterior. Posteriormente aplicou-se o *Average F-test*, versão modificada do GRS, que corrige o problema da inversão da matriz de covariância dos resíduos que surge com os dados simulados. Os resultados para esse teste indicam que o modelo que melhor explica os retornos simulados é o MKT, SMB, RMW e CMA. A partir das evidências apresentadas anteriormente de que o fator IMU auxilia no desempenho dos modelos, fez-se o procedimento de simulação dos retornos com o modelo MKT, SMB, RMW, CMA e IMU. O resultado encontrado foi que este último apresenta performance superior ao dos demais modelos, o que fornece suporte para a hipótese central deste estudo de precificação do risco informacional no mercado acionário brasileiro.

A partir dos resultados das estimativas, entende-se que o fator IMU funciona como um complemento ao fator SMB – este último vital para o desempenho dos modelos – relacionado ao risco informacional das ações. A explicação se dá pela forma com que ambos são construídos, já que as empresas menores (*small*) estão fortemente

presentes no grupo *informed* e as grandes empresas (*big*) compõem o grupo *uninformed*. Caso esses dois fatores fossem *proxies* um para o outro, o teste VIF apresentaria valor alto. Além disso, verificou-se que a correlação entre os dois, apesar de positiva, não é suficiente para que um fator incorpore por completo o outro, fato este evidenciado pela regressão do IMU com os demais fatores.

Concluindo, o fator relacionado ao risco informacional aparenta ter um papel significativo na explicação dos retornos dos portfólios criados. O fator de mercado e o SMB são os mais importantes no desempenho dos modelos, enquanto o HML não só é redundante como diminui o p.valor da estatística de teste do GRS e Average F-test aplicados nos modelos. Os fatores adicionados por Fama e French (2015) ajudam a compor o modelo que melhor explica os retornos das 62 carteiras analisadas neste trabalho.

A principal limitação desta pesquisa reside no período analisado. Enquanto trabalhos de precificação tendem a cobrir o maior período possível, devido à natureza dos dados utilizados para a criação do VPIN, não foi possível analisar um período maior do que dois anos (512 dias de transações). Essa limitação é compensada, em parte, pela utilização dos dados reais no cálculo do VPIN, procedimento este raro nos trabalhos de microestrutura de mercado pela não possibilidade de obtenção dos dados pela maior parte dos pesquisadores.

Como guia futuro das pesquisas nesta área, sugerem-se maiores análises da relação entre o fator SMB e IMU, objetivando entender a profundidade do relacionamento entre elas. Outra lacuna se refere à extração do conteúdo informacional das transações, ramo este liderado por Easley, López de Prado e O'Hara (2016) e que ainda necessita de muito avanço, por se tratar de uma variável de extrema complexidade e volatilidade.

REFERÊNCIAS

ABAD, David.; RUBIA, Antonio. Modelos de estimación de la probabilidad de negociación informada: una comparación metodológica en el mercado español. **Revista de Economia Financeira**, v. 7, p. 1-37, 2005.

ABAD, David.; YAGÜE, José. From PIN to VPIN: An introduction to order flow toxicity. **The Spanish Review of Financial Economics**, v. 2, n. 10, p. 74-83, 2012.

AITKEN, Michael.; FRINO, Alex. The accuracy of the tick test: evidence from the Australian stock exchange. **Journal of Banking & Finance**, v. 20, p. 1715-29, 1996.

AKERLOF, George A. The market for “lemons”: quality uncertainty and the Market mechanism. **The Quarterly Journal of Economics**, v. 84, n. 3, p. 488-500, Aug. 1970.

AKTAS, Nihat.; BODT, Eric de.; DECLERCK, Fany.; OPPENS, Hervé Vans. The PIN anomaly around M&A announcements. **Journal of Financial Markets**, v. 10, p. 169-91, 2007.

ALBANEZ, Tatiana.; LIMA, Gerlando.; LOPES, Alexsandro.; VALLE, Maurício. The relationship of asymmetric information in brazilian public companies. **Review of Business**, v. 31, p. 1-21, 2010.

ALBANEZ, Tatiana.; VALLE, Maurício. Impactos da assimetria de informação na estrutura de capital de empresas brasileiras abertas. **Revista de Contabilidade e Finanças**, São Paulo, v. 20, n. 51, p. 6-27, jan./abr. 2009.

AMIHUD, Yakov. Illiquidity and stock returns: cross-section and time-series effects. **Journal of Financial Markets**, v. 5, p. 31-56, 2002.

ANDERSEN, Torben.; BONDARENKO, Oleg. VPIN and the flash crash. **Journal of Financial Markets**, v. 17, p. 1-46, 2014a.

ANDERSEN, Torben.; BONDARENKO, Oleg. Reflecting on the VPIN dispute. **Journal of Financial Markets**, v. 17, p. 53-64, 2014b.

ARGOLO, Érico.; LEAL, Ricardo.; ALMEIDA, Vinício. **O modelo de Fama e French é aplicável no Brasil?** Rio de Janeiro, Relatórios COPPEAD, 2012.

ASLAN, Hadiye.; EASLEY, David.; HVIDKJAER, Soeren.; O'HARA, Maureen. The characteristics of informed trading: implications for asset pricing. **Journal of Empirical Finance**, v. 18, p. 782-801, 2011.

ASQUITH, Paul.; OMAN, Rebecca.; SAFAYA, Christopher. Short sales and trade classification algorithms. **Journal of Financial Markets**, v. 13, p. 157-73, 2010.

BAGEHOT, Walter. The only game in town. **Financial Analysis Journal**, v. 27, n. 2, p. 12-4, 1971.

BANZ, Rolf. The relationship between return and market value of common stocks. **Journal of Financial Economics**, v. 9, p. 3-18, 1981.

BASU, Sanjoy. The relationship between earnings yield, market value, and return for NYSE common stocks: further evidence. **Journal of Financial Economics**, v. 12, p. 129-56, 1983.

BENSTON, George.; HAGERMAN, Robert. Determinants of bid-ask spreads in the over-the-counter market. **Journal of Financial Economics**, v. 1, n. 4, p. 353-64, 1984.

BETHEL, Wes.; LEINWEBER, David.; RÜBEL, Oliver.; WU, Kesheng. Federal market informational technology in the post Flash Crash era: roles for supercomputing. **Journal of Trading**, v. 7, n. 2, p. 9-25, 2012.

BHANDARI, Laxmi. Debt/equity ratio and expected common stock returns: empirical evidence. **Journal of Finance**, v. 43, p. 507-28, 1998.

BHARATH, Sreedhar.; PASQUARIELLO, Paolo.; WU, Guojun. Does asymmetric information drive capital structure decisions? **The Review of Financial Studies**, v. 22, n. 8, p. 3.212-43, 2009.

BARBEDO, Claudio Henrique.; SILVA, Eduardo Camilo.; LEAL, Ricardo Pereira Câmara. Probabilidade de informação privilegiada no mercado de ações, liquidez intradiária e níveis de governança corporativa. **Revista Brasileira de Economia**, Rio de Janeiro, v. 63, n. 1, p. 49-60, 2009.

BLACK, Fischer. Capital market equilibrium with restricted borrowing. **Journal of Business**, v. 45, p. 444-55, 1972.

BOROCHIN, Paul.; RUSH, Stephen. **Identifying and pricing adverse selection risk with VPIN**. Storrs, University of Connecticut, 2016. *Working Paper*.

BREUSCH, Trevor.; PAGAN, Adrian. A simple test for heteroscedasticity and random coefficient variation. **Econometrica**, v. 47, p. 1.287-94, 1979.

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

CERETTA, Paulo Sergio ., COSTA JR., Newton. Quantas ações tornam um portfólio diversificado no mercado de capitais brasileiro ? Mercado de Capitais: Análise Empírica no Brasil. São Paulo: Atlas, 2000.

CHAKRABARTY, Bidisha.; MOULTON, Pamela.; SHKILKO, Andriy. Short sales, long sales, and the Lee-Ready trade classification algorithm revisited. **Journal of Financial Markets**, v. 15, n. 4, p. 467-91, 2012.

CHAKRABARTY, Bidisha.; PASCUAL, Roberto.; SHKILKO, Andriy. Evaluating trade classification algorithms: bulk volume classification versus the tick rule and the Lee-Ready algorithm. **Journal of Financial Markets**, v. 25, p. 52-79, 2015.

CHAN, Louis.; HAMAOKA, Yasushi.; LAKONISHOK, Josef. Fundamentals and stock returns in Japan. **Journal of Finance**, v. 46, p. 1.739-89, 1991.

CHEUNG, William.; CHOU, Robin.; LEI, Adrian. Exchange-traded barrier option and VPIN: evidence from Hong Kong. **The Journal of Future Markets**, v. 35, n. 6, p. 561-81, 2015.

CRESWELL, J. W. **Projeto de pesquisa: Métodos qualitativo, quantitativo e misto**. 2 ed. Porto Alegre: Artmed, 2007.

COPELAND, Thomas.; GALAI, Dan. Information effects on the bid-ask spread. **The Journal of Finance**, v. 38, n. 5, p. 1.457-69, 1983.

DIAMOND, Douglas.; VERRECCHIA, Robert. Disclosure, liquidity and the cost of capital. **The Journal of Finance**, v. 46, n. 4, p. 1.325-59, 1991.

DUARTE, Jefferson.; YOUNG, Lance. Why is PIN priced? **Journal of Financial Economics**, v. 91, p. 119-38, 2009.

EASLEY, David.; O'HARA, Maureen. Price, trade size, and information in securities markets. **Journal of Financial Economics**, v. 19, p. 69-90, 1987.

EASLEY, David.; O'HARA, Maureen. Time and the process of security price adjustment. **Journal of Finance**, v. 47, p. 576-605, 1992.

EASLEY, David.; KIEFER, Nicholas M. O'HARA, Maureen; PAPERMAN, Joseph B. Liquidity, information, and infrequently traded stocks. **The Journal of Finance**, v. 51, n. 4, p. 1.405-36, Sep. 1996.

EASLEY, David.; HVIDKJAER, Soeren; O'HARA, Maureen. Is information risk a determinant of asset returns? **The Journal of Finance**, v. 52, n. 5, p. 2.185-221, 2002.

EASLEY, David.; O'HARA, Maureen. Information and the cost of capital. **The Journal of Finance**, v. 59, n. 4, p. 1.553-83, Aug. 2004.

EASLEY, David.; HVIDKJAER, Soeren; O'HARA, Maureen. Factoring information into returns. **Journal of Financial and Quantitative Analysis**, v. 45, n. 2, p. 293-309, Apr. 2005.

EASLEY, David.; ENGLE, Robert.; O'HARA, Maureen.; WU, Liuren. Time-varying arrival rates of informed and uninformed trades. **Journal of Financial Econometrics**, v. 6, n. 2, p. 171-207, 2008.

EASLEY, David.; LÓPEZ DE PRADO, Marcos; O'HARA, Maureen. The microstructure of the "Flash Crash": flow toxicity, liquidity crashes, and the probability of informed trading. **The Journal of Portfolio Management**, n. 37, v. 2, p.118-28, 2011a.

EASLEY, David.; LÓPEZ DE PRADO, Marcos; O'HARA, Maureen. The exchange of flow toxicity. **Johnson School Research Paper Series**, v. 10, p.1-11, 2011b.

EASLEY, David.; LÓPEZ DE PRADO, Marcos.; O'HARA, Maureen. Flow toxicity and liquidity in a high frequency world. **Review of Financial Studies**, v. 25, p. 1.457-93, 2012a.

EASLEY, David.; LÓPEZ DE PRADO, Marcos.; O'HARA, Maureen. Bulk classification of trading activity. **Johnson School Research Paper Series**, v. 8, p. 1-40, 2012b.

EASLEY, David.; LÓPEZ DE PRADO, Marcos.; O'HARA, Maureen. The volume clock: insights into the high frequency paradigm. **Journal of Portfolio Management**, v. 39, n. 1, p. 19-29, 2012c.

EASLEY, David.; LÓPEZ DE PRADO, Marcos.; O'HARA, Maureen. VPIN and the flash crash: a rejoinder. **Journal of Financial Markets**, v. 17, p. 47-52, 2014.

EASLEY, David.; LÓPEZ DE PRADO, Marcos.; O'HARA, Maureen. Discerning information from trade data. **Journal of Financial Economics**, v. 120, n. 2, p. 269-86, 2016.

ELLIS, Katrina.; MICHAELY, Roni.; O'HARA, Maureen. The accuracy of trade classification rules: evidence from NASDAQ. **Journal of Finance and Quantitative Analysis**, v. 35, n. 4, p. 529-51, 2000.

FAMA, Eugene F. Efficient capital markets: a review of theory and empirical work. **The Journal of Finance**, Pittsburgh, PA, v. 25, n. 2, p. 383-417, May 1970.

FAMA, Eugene F. Efficient capital markets: II. **The Journal of Finance**, v. 46, n. 5, p. 1.575-617, Dec. 1991.

FAMA, Eugene. F.; FRENCH, Kenneth. The cross-section of expected stock returns. **The Journal of Finance**, v. 47, n. 2, p. 427-65, 1992.

FAMA, Eugene. F.; FRENCH, Kenneth. Common risk factors in the returns on stocks and bonds. **Journal of Financial Economics**, v. 33, p. 3-56, 1993.

FAMA, Eugene. F.; FRENCH, Kenneth. Multifactor explanations of asset pricing anomalies. **The Journal of Finance**, v. 51, n. 1, p. 55-84, 1996.

FAMA, Eugene. F.; FRENCH, Kenneth. A five-factor asset pricing model. **Journal of Financial Economic**, v. 116, p. 1-22, 2015.

FAMA, Eugene. F.; MACBETH, James. Risk, return and equilibrium: empirical tests. **The Journal of Political Economy**, v. 81, n. 3, p. 607-36, 1973.

FORTUNATO, Gabriela.; MOTTA, Luiz Felipe.; RUSSO, Giuseppe. Custo de capital próprio em mercados emergentes: uma abordagem empírica no Brasil com *downside risk*. **Revista de Administração Mackenzie**, v. 11, n. 1, p. 92-116, 2010.

GAN, Quan.; WEI, Wang.; JOHNSTONE, David. A faster estimation method for the probability of informed trading using hierarchical agglomerative clustering. **Quantitative Finance**, v. 15, n. 11, p. 1.805-21, 2015.

GIBBONS, Michael.; ROSS, Stephen.; SHANKEN, Jay. A test of the efficiency of a given portfolio. **Econometrica**, v. 57, p. 1.121-52, 1989.

GIL, A.C. **Métodos e técnicas de pesquisa social**. 5. ed. São Paulo: Atlas, 2008.

GLOSTEN, Lawrence.; MILGROM, Paul. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. **Journal of Financial Economics**, v. 14, p. 71-100, 1985.

GRAMMIG, Joachim.; THEISSEN, Eric. Estimating the probability of informed trading: does trade misclassification matter? **Bonn Econ Discussion Papers**, v. 37, p. 1-21, 2002.

GREENE, William. **Econometric theory**. New York: McGraw-Hill, 2002.

GROSSMAN, Sanford.; STIGLITZ, Joseph. On the impossibility of informationally efficient markets. **The American Economic Review**, v. 70, n. 3, p. 393-408, 1980.

GORDON, Myron.; SHAPIRO, Eli. Capital investment analysis: the required rate of profit. **Management Science**, v. 3, p. 102-10, 1956.

GUJARATI, Damodar. **Econometria básica**. São Paulo: Campus, 2006.

HASBROUCK, Joel. Measuring the information content of stock trades. **The Journal of Finance**, v. 46, n. 1, p. 179-207, 1991.

HUGHES, John.; LIU, Jing.; LIU, Jun. Information asymmetry, diversification, and asset pricing. **Accounting Review**, v. 82, p. 705–29, 2007.

HWANG, Lee-Seok.; LEE, Woo-Jong.; LIM, Seung-Yeon.; PARK, Kyung-Ho. Does information risk affect the implied cost of equity capital? An analysis of PIN and adjusted PIN. **Journal of Accounting and Economics**, v. 55, p. 148-67, 2013.

HWANG, Soosung.; SATCHELL, Stephen. Testing linear factor models on individual stocks using the average F-test. **The European Journal of Finance**, v. 20, n. 5, p. 463-98, 2014.

IQUIAPAZA, Robert; LAMOUNIER, Wagner; AMARAL, Hudson. Assimetria de informações e pagamento de dividendos na Bovespa. **Advances in Scientific and Applied Accounting**, v. 1, n. 1, p. 1-15, 2007.

JAGADEESH, Narasimham.; TITMAN, Sheridan. Returns to buying winners and selling losers: implications for stock market efficiency. **Journal of Finance**, v. 48, p. 65-91, 1993.

JENSEN, Michael.; MECKLING, William. Theory of the firm: managerial behavior, agency costs and ownership structure. **Journal of Financial Economics**, v. 3, n. 4, p. 305-60, 1976.

JIANG, Jinzhi. **Volume-synchronized probability of informed trading (VPIN), market volatility, and high-frequency liquidity**, 2015, 140 f.. Dissertação (Mestrado em Administração) – Goodman School of Business, Brock University St. Catherines, Ontario, 2015.

KYLE, Albert. Continuous auctions and insider trading. **Econometrica**, v. 53, n. 6, p. 1315-1336, 1985.

LAMBERT, Richard; LEUZ, Christian; VERRECCHIA, Robert. Informational asymmetry, information precision, and the cost of capital. **Review of Finance**, v. 16, p. 1-29, 2011.

LEE, Charles M.C.; READY, Mark J. Inferring trade direction from intraday data. **The Journal of Finance**, v. 46, n. 2, p. 733-46, June 1991.

LELAND, Hayne. Insider trading: should it be prohibited? **The Journal of Political Economy**, v. 100, n. 4, p. 859-87, 1992.

LELAND, Hayne.; PYLE, David. Information asymmetries, financial structure, and financial intermediation. **The Journal of Finance**, v. 32, n. 2, p. 371-87, 1977.

LI, Qizhai.; HU, Jiyuan.; DING, Juan.; ZHENG, Gang. Fisher's method of combining dependent statistics using generalizations of the gamma distribution with application to genetic pleiotropic associations. **Biostatistics**, v. 15, n. 2, p. 284-295, 2014.

LINTNER, John. The valuation of risk assets and selection of risky investments in stocks portfolios and capital budgets. **Review of Economics and Statistics**, v. 47, p. 13-37, 1965.

LUCENA, Pierre.; PINTO, Antônio Carlos. Anomalias no mercado de ações brasileiro: uma modificação no modelo de Fama e French. **RAC-Eletrônica**, v. 2, n. 3, p. 509-30, 2008.

MARKOWITZ, Harry. Portfolio selection. **The Journal of Finance**, v. 7, n. 1, p. 77-91, 1952.

MARTINS, Orleans S.M.; PAULO, Edilson. A probabilidade de negociação com informação privilegiada no Mercado acionário brasileiro. **Revista Brasileira de Finanças** (online), Rio de Janeiro, v. 11, n. 2, p. 249-80, June. 2013.

MARTINS, Orleans S.M.; PAULO, Edilson. Assimetria de informação na negociação de ações, características econômico-financeiras e governança corporativa no mercado acionário brasileiro. **Revista Contabilidade e Finanças**, São Paulo, v. 25, n. 64, p. 33-45, jan./abr. 2014.

MOHANRAM, Partha.; RAJGOPAL, Shiva. Is PIN priced risk? **Journal of Accounting and Economics**, v. 47, p. 226-43, 2009.

MOSSIN, Jan. Equilibrium in capital asset market. **Econometrica**, v. 34, n. 4, p. 768-83, 1966.

MUSSA, Adriano.; FAMÁ, Rubens.; DOS SANTOS, José. A adição do fator de risco momento ao modelo de precificação de ativos dos três fatores de Fama e French aplicado ao mercado acionário brasileiro. **REGE**, v. 19, n. 3, p. 431-47, 2012.

MYERS, Stewart; MAJLUF, Nicholas. Corporate financing and investment decisions when firms have information that investors do not have. **Journal of Financial Economics**, v. 13, n. 2, p. 187-221, 1984.

NEWKEY, Whitney.; WEST, Kenneth. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. **Econometrica**, v. 55, n. 3, p. 703-8, 1987.

NEWKEY, Whitney.; WEST, Kenneth. Automatic lag selection in covariance matrix estimation. **Review of Economic Studies**, v. 61, p. 631-53, 1994.

NOVY-MARX, Robert. The other side of value: the gross profitability premium. **Journal of Financial Economics**, v. 108, p. 1-28, 2013.

ODDERS-WHITE, Elizabeth. On the occurrence and consequences of inaccurate trade classification. **Journal of Financial Markets**, v. 3, p. 259-86, 2000.

OHLSON, James.; JUETTNER-NAUROTH, Beate. Expected EPS and EPS growth as determinants of value. **Review of Accounting Studies**, v. 10, p. 349-65, 2005.

OMRANE, Walid.; WELCH, Robert. Tick test accuracy in foreign exchange ECN markets. **Research in International Business and Finance**, v. 37, p. 135-52, 2016.

O'HARA, Maureen. Presidential address: liquidity and price discovery. **The Journal of Finance**, v. 63, n. 4, p. 1335-1354, 2003.

O'HARA, Maureen. **Market Microstructure Theory**. Cambridge, MA. Ed. Blackwell, 1995.

O'HARA, Maureen. High frequency market microstructure. **Journal of Financial Economics**, v. 116, p. 257-70, 2015.

PERLIN, Marcelo.; BROOKS, Chris.; DUFOUR, Alfonso. On the performance of the tick test. **The Quartely Review of Economics and Finance**, v. 54, p. 42-50, 2014.

ROGERS, Pablo.; SECURATO, José Roberto. Estudo comparative no mercado brasileiro do CAPM, modelo de 3 fatores de Fama e French e *Reward Beta Approach*. **RAC-Eletrônica**, v. 3, n. 1, p. 159-79, 2009.

ROSENBERG, Barr.; REID, Kenneth.; LANSTEIN, Ronald. Persuasive evidence of market inefficiency. **Journal of Portfolio Management**, v. 11, p. 9-17, 1985.

ROSS, Stephen. The arbitrage theory of capital asset pricing. **Journal of Economic Theory**, v. 13, p. 341-60, 1976.

RUPPERT, D.; MATTESON, D. **Statistics and data analysis for financial engineering**, New York: Springer Science, 2015. Springer Texts in Statistics.

SANVICENTE, Antônio.; BELLATO, Letícia. Determinação do grau necessário de diversificação de uma carteira de ações no mercado de capitais brasileiro. In: SEMINÁRIOS EM ADMINISTRAÇÃO - SEMEAD, 7., 2004, São Paulo. **Anais...** São Paulo: FEA-USP, 2004.

SARIN, A.; SHASTRI, K. A.; SHASTRI, K. Ownership structure and stock market liquidity. **International Journal of Accounting**, v. 38, p. 465-91, 2000.

SECURATO, José Roberto.; MÁLAGA, Flávio. Aplicação do modelo de três fatores de Fama e French no mercado acionário brasileiro – um estudo empírico do period 1995-2003. In: XXVIII – ENANPAD 2004, Curitiba. **Anais....** Curitiba: ANPAD, 2004.

SILVA, Eduardo. **Dois ensaios sobre microestrutura de mercado e probabilidade de informação privilegiada no Mercado de ações brasileiro**. 2009, 71 ff. Tese

(Doutorado em Administração) – Instituto COPPEAD de Administração, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.

SHARPE, William. Capital asset prices: a theory of market equilibrium under conditions at risk. **The Journal of Finance**, v. 19, p. 425-42, 1964.

SONG, Jung.; WU, Kesheng.; SIMON, Horst. Parameter analysis of the VPIN metric. **Lawrence Berkeley National Lab**, p. 1-26, 2014.

STOLL, Hans. The pricing of security dealer services: an empirical study of NASDAQ stocks. **The Journal of Finance**, v. 33, n. 4, p. 1.153-72, 1978.

TAY, Anthony.; TING, Christopher.; TSE, Yiu.; WARACHKA, Mitch. Using high-frequency transaction data to estimate the probability of informed trading. **Journal of Financial Econometrics**, v. 7, n. 3, p. 288-311, 2009.

THEISSEN, Erik. A test of the accuracy of the Lee-Ready trade classification algorithm. **Journal of International Financial Markets, Institutions and Money**, v. 11, p. 147-65, 2000.

TITMAN, Sheridan.; WEI, John.; XIE, Feixue. Capital investments and stock returns. **Journal of Financial and Quantitative Analysis**, v. 39, p. 677-700, 2004.

WANG, Jiang. A model of intertemporal asset prices under asymmetric information. **Review of Economic Studies**, v. 60, p. 249-82, 1993.

WEI, Wang.; GERACE, Dionigi.; FRINO, Alex. Informed trading, flow toxicity and the impact on intraday trading factors. **Australasian Accounting, Business and Finance Journal**, v. 7, n. 2, p. 3-24, 2013.

YILDIZ, Serhat.; VAN NESS, Robert.; VAN NESS, Bonnie. **The role of HFTs in order flow toxicity and stock price variance, and predicting changes in HFTs liquidity provisions**. Oxford, University of Mississippi Press, 2016. *Working paper*.

APÊNDICE

Tabela 36

Amostra das ações passíves para composição dos fatores de Fama e French (1993, 2015) e Carhart (1997)

Ações								
ABCB4	BPAR3	CIEL3	ELET6	HBTS5	LREN3	OIBR4	RLOG3	TCSA3
ABEV3	BPHA3	CLSC4	ELPL4	HCBR3	LUPA3	OSXB3	RNEW11	TECN3
AELP3	BRAP4	CMIG4	EMAE4	HETA4	LUXM4	PARC3	ROMI3	TEKA4
AFLT3	BRFS3	CMSA4	EMBR3	HGTX3	MAGG3	PATI3	RPAD6	TELB3
AFLU3	BRGE5	COCE5	ENBR3	HOOT4	MAPT4	PCAR4	RPMG3	TENE5
AGRO3	BRIN3	CORR4	ENEV3	HYPE3	MDIA3	PDGR3	RSID3	TERI3
AHEB5	BRIV4	CPFE3	ENGI3	IBOV1	MEAL3	PEAB3	RSUL3	TGMA3
ALPA4	BRKM5	CPLE6	ENMT4	IDNT3	MEND6	PETR3	RUMO3	TIET11
ALSC3	BRML3	CPRE3	EQTL3	IDVL4	MERC4	PETR4	SANB1	TIET3
ALTS3	BRPR3	CRDE3	ESTC3	IGBR3	MGEL4	PFRM3	SAPR4	TIMP3
ALUP1	BRQB3	CREM3	ESTR4	IGTA3	MGLU3	PINE4	SBSP3	TKNO4
AMAR3	BRSR6	CRIV4	ETER3	IMBI4	MILS3	PLAS3	SCAR3	TOTS3
ANIM3	BSEV3	CRPG5	EUCA4	INEP4	MLFT4	PMAM3	SCLO4	TOYB3
APTI4	BSLI3	CSAB4	EVEN3	ITEC3	MMAQ4	PNVL4	SEDU3	TPIS3
ARZZ3	BTOW3	CSAN3	EZTC3	ITSA4	MMXM3	POMO4	SEER3	TRIS3
ATOM3	BTTL4	CSMG3	FBMC4	ITUB4	MNPR3	POSI3	SFSA4	TRPL4
AZEV4	BVMF3	CSNA3	FESA4	JBDU4	MOAR3	PPAR3	SGAS4	TRPN3
BAHI3	CABB3	CSRN3	FHER3	JBSS3	MPLU3	PQTM4	SGPS3	TUPY3
BALM4	CALI4	CTIP3	FIBR3	JFEN3	MRFG3	PRBC4	SHOW3	TXRX4
BAUH3	CAMB4	CTKA4	FIGE3	JHSF3	MRVE3	PRCA3	SHUL4	UCAS3
BAZA3	CARD3	CTNM4	FJTA4	JOPA4	MSPA4	PRIO3	SJOS4	UGPA3
BBAS3	CASN4	CTSA4	FLRY3	JSLG3	MSRO3	PRML3	SLCE3	UNIP6
BBDC3	CATA3	CVCB3	FNCN3	KEPL3	MTIG4	PSEG4	SLED4	USIM5
BBDC4	CBEE3	CYRE3	FOMS3	KLBN11	MTSA4	PSSA3	SMLE3	VAGR3
BBRK3	CBMA4	DASA3	FRIO3	KLBN4	MULT3	PSVM11	SMTO3	VALE3
BBSE3	CCHI3	DAYC4	FRTA3	KROT3	MWET4	PTBL3	SNSL3	VALE5
BBTG11	CCPR3	DHBI4	GEP4	LAME3	MYPK3	PTNT4	SNSY5	VIGR3
BDLL4	CCRO3	DIRR3	GFS3	LAME4	NAFG4	PTPA4	SOND6	VIVR3
BEEF3	CCXC3	DOHL4	GGBR4	LCAM3	NATU3	QGEP3	SPRI5	VIVT4
BEES3	CEBR5	DTCY3	GOAU4	LEVE3	NCNE5	QUAL3	SQRM3	VLID3
BGIP4	CEDO4	DTEX3	GOLL4	LFPE4	NORD3	QUSW3	SSBR3	VSPT4
BMEB4	CEEB3	EALT4	GPAP3	LHER4	NRTQ3	RADL3	STBP11	VTLM3
BMIN4	CEED3	ECOR3	GPCP3	LIGT3	NUTR3	RANI4	STKF3	VULC3
BMKS3	CEGR3	ECPR4	GRND3	LINX3	ODER4	RAPT4	SULA11	VVAR11
BMTO4	CELP5	EEEL3	GRUC3	LIPR3	ODPV3	RCSL4	SULT4	VVAR3
BNPA3	CEPE5	EGIE3	GSHP3	LIXC4	OFSA3	RDNI3	SUZB5	WEGE3
BOBR4	CESP6	EKTR4	GUAR4	LLIS3	OGSA3	REDE4	TAE11	WHRL4
BPAC3	CGAS5	ELEK4	HAGA4	LOGN3	OGXP3	RENT3	TBLE3	
BPAN4	CGRA4	ELET3	HBOR3	LPSB3	OIBR3	RJCP3	TCNO4	

Fonte - Elaborada pelo autor da dissertação.

Tabela 37

Estatísticas do teste de Durbin-Watson para os portfólios analisados

12 Portfólios: Tamanho e <i>Book-to-market</i>												
Nº	3 Fatores		3 Fatores + IMU		4 Fatores		4 Fatores + IMU		5 Fatores		5 Fatores + IMU	
	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.
1	1,885	0,095	1,886	0,096	1,881	0,087	1,882	0,088	1,896	0,116	1,898	0,121
2	2,024	0,605	2,033	0,643	2,041	0,676	2,047	0,701	2,020	0,585	2,028	0,621
3	1,946	0,267	1,937	0,237	1,975	0,383	1,969	0,357	1,981	0,411	1,972	0,373
4	2,209	0,991	2,209	0,991	2,209	0,991	2,209	0,991	2,201	0,988	2,201	0,989
5	2,189	0,984	2,183	0,981	2,180	0,979	2,176	0,977	2,157	0,961	2,146	0,951
6	2,188	0,983	2,184	0,981	2,184	0,981	2,181	0,980	2,182	0,980	2,175	0,976
7	2,148	0,953	2,139	0,943	2,142	0,945	2,131	0,931	2,151	0,956	2,145	0,949
8	2,003	0,509	2,003	0,511	1,996	0,477	1,996	0,479	2,006	0,521	2,006	0,522
9	2,306	1,000	2,306	1,000	2,298	1,000	2,298	1,000	2,292	1,000	2,291	1,000
10	2,254	0,998	2,249	0,998	2,239	0,997	2,235	0,996	2,253	0,998	2,244	0,997
11	2,240	0,997	2,241	0,997	2,256	0,998	2,256	0,998	2,237	0,996	2,240	0,997
12	2,346	1,000	2,343	1,000	2,336	1,000	2,334	1,000	2,331	1,000	2,327	1,000
12 Portfólios: Tamanho e Investimento												
Nº	3 Fatores		3 Fatores + IMU		4 Fatores		4 Fatores + IMU		5 Fatores		5 Fatores + IMU	
	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.
1	2,064	0,766	2,065	0,769	2,078	0,810	2,081	0,818	2,121	0,914	2,123	0,917
2	2,101	0,872	2,100	0,871	2,101	0,872	2,100	0,871	2,127	0,924	2,127	0,923
3	1,909	0,149	1,912	0,156	1,868	0,066	1,876	0,077	2,061	0,752	2,057	0,736
4	1,711	0,000	1,706	0,000	1,711	0,000	1,707	0,000	1,742	0,002	1,738	0,001
5	2,225	0,995	2,225	0,995	2,215	0,992	2,214	0,992	2,217	0,993	2,217	0,993
6	1,914	0,162	1,915	0,165	1,909	0,147	1,909	0,149	1,938	0,237	1,940	0,245
7	2,203	0,989	2,198	0,987	2,199	0,988	2,195	0,986	2,196	0,987	2,186	0,982
8	2,060	0,749	2,053	0,724	2,052	0,720	2,047	0,700	2,051	0,713	2,041	0,675
9	2,281	0,999	2,275	0,999	2,304	1,000	2,298	1,000	2,291	1,000	2,280	0,999
10	2,287	0,999	2,288	0,999	2,278	0,999	2,279	0,999	2,283	0,999	2,284	0,999
11	2,344	1,000	2,343	1,000	2,330	1,000	2,330	1,000	2,335	1,000	2,332	1,000
12	2,200	0,988	2,211	0,991	2,197	0,987	2,206	0,990	2,182	0,980	2,192	0,985
12 Portfólios: Tamanho, Investimento e Lucratividade												
Nº	3 Fatores		3 Fatores + IMU		4 Fatores		4 Fatores + IMU		5 Fatores		5 Fatores + IMU	
	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.
1	1,875	0,076	1,874	0,075	1,875	0,076	1,874	0,076	1,893	0,110	1,893	0,111
2	1,912	0,157	1,912	0,158	1,922	0,186	1,922	0,186	2,012	0,550	2,012	0,551
3	2,162	0,966	2,159	0,964	2,167	0,970	2,165	0,969	2,194	0,986	2,191	0,985
4	1,955	0,301	1,956	0,306	1,935	0,228	1,937	0,235	1,965	0,339	1,964	0,335
5	1,862	0,057	1,863	0,059	1,881	0,086	1,889	0,103	1,928	0,204	1,924	0,191
6	2,205	0,990	2,223	0,994	2,235	0,996	2,249	0,998	2,221	0,994	2,233	0,996
7	2,219	0,994	2,220	0,994	2,217	0,993	2,218	0,993	2,227	0,995	2,226	0,995
8	2,196	0,987	2,196	0,987	2,190	0,984	2,190	0,984	2,180	0,979	2,178	0,978

9	2,391	1,000	2,397	1,000	2,382	1,000	2,389	1,000	2,397	1,000	2,400	1,000
10	2,288	0,999	2,288	0,999	2,280	0,999	2,281	0,999	2,285	0,999	2,285	0,999
11	2,142	0,946	2,144	0,948	2,151	0,956	2,152	0,957	2,137	0,938	2,138	0,940
12	2,156	0,961	2,149	0,954	2,144	0,948	2,138	0,941	2,128	0,925	2,117	0,906

10 Portfólios: Tamanho e VPIN

Nº	3 Fatores		3 Fatores + IMU		4 Fatores		4 Fatores + IMU		5 Fatores		5 Fatores + IMU	
	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.
1	2,002	0,506	2,229	0,995	2,042	0,678	2,251	0,998	1,999	0,491	2,229	0,995
2	2,021	0,589	2,018	0,579	2,048	0,704	2,047	0,700	2,043	0,683	2,039	0,665
3	2,090	0,844	2,094	0,856	2,088	0,839	2,094	0,855	2,093	0,851	2,097	0,861
4	2,291	1,000	2,290	0,999	2,307	1,000	2,306	1,000	2,290	0,999	2,289	0,999
5	2,311	1,000	2,315	1,000	2,299	1,000	2,302	1,000	2,283	0,999	2,285	0,999
6	2,198	0,987	2,180	0,979	2,201	0,989	2,186	0,982	2,173	0,975	2,149	0,954
7	2,082	0,824	2,089	0,843	2,071	0,786	2,080	0,815	2,091	0,846	2,094	0,854
8	2,323	1,000	2,327	1,000	2,312	1,000	2,314	1,000	2,314	1,000	2,317	1,000
9	2,282	0,999	2,282	0,999	2,269	0,999	2,269	0,999	2,265	0,999	2,264	0,999
10	2,231	0,996	2,236	0,996	2,256	0,998	2,269	0,999	2,221	0,994	2,222	0,994

16 Portfólios: Tamanho e Lucratividade

Nº	3 Fatores		3 Fatores + IMU		4 Fatores		4 Fatores + IMU		5 Fatores		5 Fatores + IMU	
	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.	DW	P.val.
1	2,200	0,988	2,202	0,989	2,211	0,991	2,215	0,992	2,211	0,991	2,212	0,992
2	2,074	0,796	2,074	0,798	2,080	0,815	2,080	0,814	2,061	0,751	2,061	0,753
3	2,131	0,931	2,131	0,931	2,149	0,953	2,149	0,954	2,200	0,988	2,200	0,988
4	1,893	0,110	1,890	0,105	1,904	0,136	1,902	0,130	1,969	0,359	1,968	0,356
5	1,976	0,389	1,969	0,361	1,974	0,381	1,972	0,373	2,075	0,801	2,065	0,766
6	1,964	0,340	1,966	0,347	1,961	0,326	1,963	0,332	1,971	0,364	1,972	0,370
7	2,157	0,962	2,164	0,968	2,162	0,966	2,166	0,970	2,170	0,973	2,171	0,973
8	1,895	0,115	1,894	0,113	1,887	0,098	1,887	0,098	1,936	0,229	1,933	0,221
9	1,995	0,475	1,995	0,476	1,993	0,463	1,993	0,464	2,004	0,514	2,004	0,515
10	2,286	0,999	2,287	0,999	2,299	1,000	2,300	1,000	2,277	0,999	2,276	0,999
11	2,220	0,994	2,211	0,992	2,238	0,996	2,234	0,996	2,219	0,993	2,207	0,990
12	2,243	0,997	2,243	0,997	2,232	0,996	2,232	0,996	2,213	0,992	2,213	0,992
13	2,180	0,979	2,186	0,982	2,185	0,982	2,189	0,984	2,173	0,974	2,179	0,978
14	2,375	1,000	2,374	1,000	2,357	1,000	2,359	1,000	2,395	1,000	2,390	1,000
15	2,127	0,924	2,127	0,924	2,112	0,897	2,112	0,897	2,127	0,924	2,127	0,923
16	2,262	0,999	2,259	0,998	2,260	0,998	2,257	0,998	2,252	0,998	2,246	0,997

Fonte - Elaborada pelo autor da dissertação.

Tabela 38

Estatísticas do teste de Breusch-Pagan para os portfólios analisados

12 Portfólios: Tamanho e <i>Book-to-market</i>												
Nº	3 Fatores		3 Fatores + IMU		4 Fatores		4 Fatores + IMU		5 Fatores		5 Fatores + IMU	
	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.
1	3,088	0,378	3,754	0,440	8,301	0,081	9,284	0,098	4,069	0,540	4,758	0,575
2	0,897	0,826	2,842	0,585	0,988	0,912	2,762	0,737	2,101	0,835	3,919	0,688
3	4,100	0,251	6,806	0,147	3,083	0,544	4,917	0,426	12,400	0,030	14,740	0,022
4	3,191	0,363	3,250	0,517	4,386	0,356	4,506	0,479	3,493	0,624	3,550	0,737
5	1,081	0,782	1,636	0,802	1,983	0,739	2,429	0,787	6,385	0,271	7,127	0,309
6	4,608	0,203	5,399	0,249	6,423	0,170	7,075	0,215	5,524	0,355	6,470	0,373
7	1,580	0,664	1,685	0,793	1,933	0,748	2,056	0,841	2,953	0,707	3,164	0,788
8	5,204	0,157	7,003	0,136	7,363	0,118	8,587	0,127	10,242	0,069	11,699	0,069
9	2,626	0,453	2,866	0,581	2,565	0,633	2,658	0,752	3,746	0,587	4,188	0,651
10	2,299	0,513	2,256	0,689	3,153	0,533	3,091	0,686	4,101	0,535	4,106	0,662
11	10,513	0,015	10,650	0,031	6,923	0,140	7,252	0,203	10,931	0,053	11,157	0,084
12	13,787	0,003	14,801	0,005	15,214	0,004	15,773	0,008	14,767	0,011	16,065	0,013
12 Portfólios: Tamanho e Investimento												
Nº	3 Fatores		3 Fatores + IMU		4 Fatores		4 Fatores + IMU		5 Fatores		5 Fatores + IMU	
	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.
1	1,856	0,603	2,006	0,735	2,262	0,688	2,503	0,776	2,992	0,701	3,042	0,804
2	4,433	0,218	4,491	0,344	6,852	0,144	7,005	0,220	4,674	0,457	4,751	0,576
3	4,917	0,178	7,866	0,097	5,365	0,252	7,935	0,160	8,014	0,155	10,110	0,120
4	1,229	0,746	2,492	0,646	1,415	0,842	2,876	0,719	2,398	0,792	3,922	0,687
5	0,395	0,941	1,030	0,905	0,534	0,970	1,465	0,917	3,781	0,581	4,457	0,615
6	12,528	0,006	13,122	0,011	12,673	0,013	13,297	0,021	15,349	0,009	15,518	0,017
7	5,457	0,141	7,531	0,110	5,584	0,232	7,819	0,167	8,275	0,142	9,974	0,126
8	3,521	0,318	3,867	0,424	4,061	0,398	4,273	0,511	5,452	0,363	5,949	0,429
9	25,275	0,000	27,556	0,000	26,809	0,000	28,337	0,000	29,149	0,000	32,722	0,000
10	9,155	0,027	8,887	0,064	8,192	0,085	7,908	0,161	9,404	0,094	9,312	0,157
11	4,930	0,177	5,135	0,274	4,642	0,326	4,670	0,457	5,266	0,384	5,584	0,471
12	0,562	0,905	0,757	0,944	0,597	0,963	0,815	0,976	3,291	0,655	3,734	0,713
12 Portfólios: Tamanho, Investimento e Lucratividade												
Nº	3 Fatores		3 Fatores + IMU		4 Fatores		4 Fatores + IMU		5 Fatores		5 Fatores + IMU	
	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.
1	6,557	0,087	7,183	0,127	6,334	0,176	6,893	0,229	8,894	0,113	9,389	0,153
2	4,648	0,199	4,651	0,325	4,765	0,312	4,802	0,441	6,142	0,293	6,201	0,401
3	4,778	0,189	4,846	0,303	7,123	0,130	7,299	0,199	5,159	0,397	5,216	0,516
4	2,195	0,533	2,968	0,563	2,778	0,596	3,427	0,634	2,644	0,755	3,718	0,715
5	0,954	0,812	1,295	0,862	1,303	0,861	1,676	0,892	1,152	0,949	1,332	0,970
6	2,410	0,492	3,179	0,528	4,794	0,309	5,787	0,327	5,836	0,323	6,526	0,367
7	0,899	0,826	1,356	0,852	1,068	0,899	1,582	0,903	3,077	0,688	3,410	0,756
8	5,003	0,172	5,073	0,280	4,841	0,304	4,814	0,439	6,907	0,228	7,038	0,317
9	9,685	0,021	9,505	0,050	9,122	0,058	8,964	0,111	9,125	0,104	9,280	0,158
10	8,507	0,037	8,520	0,074	9,142	0,058	9,050	0,107	10,776	0,056	10,831	0,094
11	0,405	0,939	0,531	0,970	0,333	0,988	0,446	0,994	1,824	0,873	1,985	0,921
12	2,042	0,564	2,276	0,685	3,066	0,547	3,142	0,678	5,812	0,325	6,385	0,381

10 Portfólios: Tamanho e VPIN

Nº	3 Fatores		3 Fatores + IMU		4 Fatores		4 Fatores + IMU		5 Fatores		5 Fatores + IMU	
	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.
1	0,388	0,943	0,427	0,980	1,060	0,901	1,286	0,936	3,107	0,683	1,337	0,970
2	5,141	0,162	5,256	0,262	5,317	0,256	5,417	0,367	6,087	0,298	6,300	0,390
3	1,691	0,639	1,647	0,800	1,825	0,768	1,823	0,873	2,213	0,819	2,146	0,906
4	12,281	0,006	12,572	0,014	10,876	0,028	11,036	0,051	13,389	0,020	13,695	0,033
5	1,365	0,714	2,510	0,643	1,766	0,779	2,480	0,779	2,355	0,798	3,544	0,738
6	0,293	0,961	1,695	0,792	1,537	0,820	2,978	0,703	2,518	0,774	4,372	0,626
7	3,670	0,299	3,734	0,443	5,219	0,266	5,201	0,392	7,656	0,176	7,769	0,255
8	5,526	0,137	6,097	0,192	5,847	0,211	6,014	0,305	7,231	0,204	7,991	0,239
9	6,740	0,081	6,930	0,140	5,867	0,209	6,129	0,294	8,975	0,110	9,573	0,144
10	5,600	0,133	5,576	0,233	4,914	0,296	4,679	0,456	6,248	0,283	6,416	0,378

16 Portfólios: Tamanho e Lucratividade

Nº	3 Fatores		3 Fatores + IMU		4 Fatores		4 Fatores + IMU		5 Fatores		5 Fatores + IMU	
	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.	BP	P.val.
1	2,531	0,470	2,563	0,633	4,040	0,401	4,176	0,524	2,989	0,702	3,010	0,808
2	4,995	0,172	8,638	0,071	6,673	0,154	9,813	0,081	8,113	0,150	11,148	0,084
3	5,525	0,137	5,609	0,230	11,013	0,026	11,078	0,050	7,410	0,192	7,663	0,264
4	3,223	0,358	3,352	0,501	3,379	0,497	3,516	0,621	5,320	0,378	5,595	0,470
5	0,623	0,891	0,980	0,913	0,898	0,925	1,343	0,930	0,602	0,988	0,977	0,986
6	3,250	0,355	3,356	0,500	4,861	0,302	5,088	0,405	10,921	0,053	10,852	0,093
7	1,756	0,625	1,787	0,775	3,869	0,424	3,815	0,576	2,606	0,760	2,761	0,838
8	3,356	0,340	4,041	0,400	3,559	0,469	4,209	0,520	3,752	0,586	4,768	0,574
9	1,099	0,777	1,120	0,891	1,275	0,866	1,290	0,936	3,580	0,611	3,580	0,733
10	1,264	0,738	1,571	0,814	1,280	0,865	1,549	0,907	1,711	0,888	2,285	0,892
11	6,618	0,085	8,413	0,078	6,410	0,171	7,740	0,171	10,711	0,057	12,334	0,055
12	3,806	0,283	4,054	0,399	4,753	0,314	4,849	0,435	4,723	0,451	4,953	0,550
13	4,733	0,192	5,037	0,284	4,001	0,406	4,374	0,497	7,878	0,163	8,404	0,210
14	5,202	0,158	5,453	0,244	5,994	0,200	6,023	0,304	7,214	0,205	7,520	0,275
15	3,221	0,359	3,419	0,490	4,230	0,376	4,273	0,511	4,293	0,508	4,481	0,612
16	2,687	0,442	2,924	0,571	3,781	0,437	4,031	0,545	4,977	0,419	5,260	0,511

Fonte - Elaborada pelo autor da dissertação.