

Avaliação do grau de atenção dos investidores individuais como indutor de volatilidade adicional no mercado brasileiro de ações

**Marcelo Guzella
Henrique Castro**

Resumo

Neste trabalho verificamos a influência do nível de atenção dos investidores individuais sobre a dinâmica dos preços dos ativos. Nos mercados financeiros, essa influência é explicada pela hipótese de pressão nos preços induzida por excesso de atenção. Adotamos como medida de atenção o volume de buscas na Internet, que tem sido utilizado em recentes estudos internacionais devido às suas vantagens em termos de frequência, versatilidade e facilidade de obtenção. Foram aplicados vetores autorregressivos em séries diárias de tempo do Ibovespa, índice representativo do mercado de ações brasileiro. Sua volatilidade condicional foi medida através de modelos da família GARCH. Como confirmado em mercados mais desenvolvidos, a quantidade de buscas contribuiu significativamente para determinar a volatilidade futura do Ibovespa, mesmo considerando como informação adicional o volume de transações, que também apresentou poder preditivo. Os resultados, robustos a testes adicionais, indicaram a influência da atuação dos *noise traders* sobre movimentos de preço.

Palavras-chave: finanças comportamentais, atenção do investidor, volume de buscas, volatilidade

Abstract

In this paper we verified whether individual investors' attention levels influence asset prices behavior. In financial markets, this relationship is explained by the attention-induced price pressure hypothesis. As measure of attention level, we adopted Google search queries, used in previous studies due to better frequency, versatility and availability. We applied vector autoregression of Ibovespa timeseries, index that represents the Brazilian stock market. Its conditional volatility was measured through GARCH models. Indeed, as confirmed in more developed markets, search queries helped in determining future price volatility of Ibovespa, even considering, as additional information, the trading volume, which also presented significant predictive power for price volatility. The results, which were robust to different volatility models and time lags, indicated the effect of noise traders activity on price movements.

Keywords: behavioral finance, investor attention, search queries, volatility

1. Introdução

Atentar-se a algo é concentrar-se mentalmente em uma tarefa ou objeto. Atenção implica em seleção, pois sempre há atividades ou coisas alternativas para se engajar. Além de seleção, é também uma questão de intensidade. Como descreve Kahneman (1973), um aluno que lê uma revista em quadrinhos, enquanto seu professor está dando aula, é culpado por selecionar inadequadamente a que se atentar. Já o aluno em sala de aula imerso em um agradável estado de sonolência não só falha em direcionar a atenção, como tem menos volume de atenção para direcionar.

Atenção é, portanto, condição necessária para absorção plena da informação disponibilizada. Por outro lado, é crescente e abundante o volume de informação mundialmente gerada e armazenada, especialmente em arquivo digital. Gillings, Hilbert e Kemp (2016) estimam que, durante as últimas três décadas, a quantidade de informação digital armazenada dobrou a cada 2,5 anos, aproximadamente. Os autores calculam que esse armazenamento pode ter chegado a 5 *zettabytes* (5×10^{21} *bytes*) em 2014, o que equivale a 500 vezes o volume de informação presente no genoma de todos os 7,2 bilhões de habitantes do planeta (considerando o espaço ocupado em *bytes* após codificação).

Dada essa atual perspectiva, não há dúvidas de que vivemos na era da informação, marcada por uma crescente integração dos mercados e necessidade de agilidade nas decisões. Excesso de informação consome muita atenção, um escasso recurso (Kahneman, 1973). A incorporação imediata das informações nos preços requer alocação eficiente da atenção, que pode não ocorrer. A atenção ajuda a restringir o leque de opções (e minimizar custos de pesquisa) mas nem sempre tem relação direta com utilidade. Essa realidade justifica uma análise do comportamento dos agentes sob esse enfoque e suas implicações sobre os preços dos ativos.

Neste estudo, chamamos de agentes aquelas entidades com alguma participação ou influência nos mercados. Na bolsa de valores, investidores institucionais e individuais (ou não-profissionais) estão entre os agentes mais relevantes, assim como governo, agências de regulação e empresas que operacionalizam as negociações ou que têm ações negociadas na bolsa.

Indivíduos usam uma variedade de heurísticas para estimar as chances de ocorrência de eventos. Esses atalhos mentais podem levar a vieses que resultam em avaliações consistentemente ruins de probabilidades, reveladas em diversos

experimentos (Tversky & Kahneman, 1974). Como esses vieses têm o potencial de fazê-los enxergar tendências e padrões (que de fato não existem) em séries de preços de ações (Andreassen & Kraus, 1990), isso os coloca como fortes candidatos a *noise traders*. Esse grupo de investidor não atua com base em valores fundamentais e tende a reagir de forma exagerada a notícias ruins ou boas (De Long, Shleifer, Summers, & Waldmann, 1991).

Medidas que representam a atenção dos participantes do mercado têm sido utilizadas com frequência para uma melhor compreensão de atributos como racionalidade limitada e heterogeneidade de crenças, e de seus efeitos. Pesquisas indicaram que essas medidas permitem a identificação ou antecipação do grau de agitação do mercado de capitais (Barber & Odean, 2008; Huddart, Lang, & Yetman, 2012; Dimpfl & Jank, 2012), do desempenho geral da economia, ou até de variáveis macroeconômicas específicas, como taxa de desemprego, inflação e produção industrial (Gomes & Taamouti, 2016).

Variações abruptas de volatilidade no mercado de capitais capturam a atenção do investidor e podem levar a um comportamento de manada. Isso sugere uma associação entre medidas de atenção e o estado do mercado. Lux e Marchesi (1999) se basearam nessa dinâmica ao propor um modelo de volatilidade da bolsa baseado no comportamento de fundamentalistas e de *noise traders*, atuando como indutores de volatilidade adicional. Além disso, ao contribuírem para explicar a dinâmica econômica, medidas de atenção podem ter relação relevante com alterações nos comovimentos de diferentes classes de ativos, ou de mercados de diferentes países (Gomes & Taamouti, 2016).

Graus diferenciados de atenção do investidor individual devem também apresentar implicações no desequilíbrio das transações do mercado de ações, já que, para essa classe de investidores, o universo de possibilidades de compra é maior do que o de venda, dadas as restrições em relação a vendas a descoberto (Barber & Odean, 2008). Como forças compradoras levam à valorização dos ativos, suspeita-se de uma relação entre medidas de atenção e retornos anormais.

Como atenção é uma variável não-observada, são adotadas *proxies* que capturam parte das suas variações, como volumes, retornos anormais passados e presença no noticiário. Diversos estudos anteriores que utilizaram algumas dessas variáveis apresentaram resultados que corroboram com a hipótese de atenção do investidor (Barber & Odean, 2008; Da, Engelberg, & Gao, 2011; Huddart et al., 2012).

No entanto, variáveis como volume de transações e retornos, por serem resultantes de equilíbrios, podem não cumprir satisfatoriamente um papel de medida de grau de atenção. Brooks (1998) e Dimpfl e Jank (2012), por exemplo, não encontraram evidências significativas de poder preditivo do volume de transações sobre a volatilidade dos preços. A presença no noticiário, por si só, também não garante a captura da atenção pelos agentes (Da et al., 2011).

Volumes de buscas de determinadas palavras-chave em *sites* populares de pesquisa podem ser uma *proxy* direta e eficaz para o grau de atenção do investidor não-profissional. Esse tipo de agente tem menor acesso a ferramentas sofisticadas e tipicamente direciona atenção apenas quando vêem a necessidade de obter informações sobre o comportamento do mercado. Da et al. (2011) apontaram que essa é uma medida indubitável de atenção revelada, pois, quando pesquisamos sobre um assunto, automaticamente estamos nos atentando a ele. O potencial proporcionado por essa medida já foi aproveitado para prever a atividade de mercados imobiliários, automotivos e de turismo (Choi & Varian, 2009), ou até de epidemias de influenza (Ginsberg et al., 2009).

Pesquisas recentes começaram a utilizar o volume de buscas no Google como indicador de atenção do investidor não-profissional, com resultados satisfatórios. Relacionando o volume de buscas com o risco da ação, Kristoufek (2013) testou uma estratégia de diversificação que penaliza ações com grande volume de buscas. Essa estratégia apresentou performance superior a índices de mercado. Preis, Moat, e Stanley (2013) também encontraram evidências de que portfólios construídos segundo o volume de buscas, para determinadas palavras-chave, podem apresentar desempenho superior ao do mercado.

Por outro lado, essa medida teve seu poder preditivo testado sem sucesso sobre retornos anormais de mercado, no curto prazo, nos estudos de Preis, Reith, e Stanley (2010) e Challet e Ahmed (2013). De forma geral, nesses trabalhos não foi encontrada correlação significativa entre volume de buscas e retornos, ou desempenho superior de estratégias com palavras-chave determinadas em relação ao daquelas com palavras-chave aleatórias.

Dentro desse contexto, abordamos neste trabalho o problema das limitações da atenção como recurso cognitivo e suas implicações no mercado de ações brasileiro, analisando a relação entre volatilidade e o grau de atenção dos agentes. Até onde se tem conhecimento, ainda não existe no Brasil uma análise da influência do grau de atenção

dos investidores individuais, medido pelas buscas na Internet, sobre a dinâmica dos preços, apesar de o país possuir o 12º maior mercado de capitais do mundo (Ministério do Planejamento, Desenvolvimento e Gestão, 2015).

Os resultados contribuem para que os participantes do mercado tenham melhores condições de prever e se preparar para os momentos de turbulência. Além disso, visam servir para que investidores consigam constituir portfólios mais adequados ao seu grau de aversão ou propensão a risco.

Este estudo propicia um maior entendimento da dinâmica dos mercados financeiros, sob a ótica da atenção, e o aprimoramento dos modelos de previsão de importantes atributos, como média, volatilidade e covariância. Esses modelos são fundamentais para áreas como gestão de riscos e de portfólio, e para a precificação de ativos e opções.

Este documento está estruturado da seguinte forma: na próxima seção, são apresentadas pesquisas que dão suporte teórico e empírico à nossa abordagem e à hipótese de que ela parte. Em seguida, é apresentada a metodologia utilizada para a análise da relação entre indicadores de atenção do investidor e a volatilidade dos preços no mercado brasileiro de ações. Logo após, descrevemos e analisamos os resultados.

2. Referencial teórico

2.1 Indução de volatilidade adicional por *noise traders*

O modelo de mercado financeiro com diferentes grupos de agentes, descrito e simulado por Lux e Marchesi (1999), revelou os efeitos das interações entre esses grupos. Os autores segregaram os participantes em fundamentalistas, que compram ações com preço abaixo de seu valor fundamental (tipicamente o fluxo de caixa descontado da soma dos ganhos futuros esperados), e *noise traders*, que não acreditam na tendência imediata de reversão do preço a seu valor fundamental.

O grupo dos *noise traders* se subdivide em otimistas e pessimistas. Eles buscam analisar graficamente os preços e o comportamento dos demais *traders* para definir tendências, adotando típico comportamento de manada. Enquanto fundamentalistas possuem expectativas puramente racionais sobre os retornos dos títulos, as opiniões e decisões de *noise traders* podem estar sujeitas a vieses sistemáticos. Sabe-se que esse grupo de investidores pode mudar de opinião baseado em expectativas ou sentimentos que não são completamente justificados ou garantidos pelas informações existentes,

como conselhos de gurus ou estratégias “populares” (Shleifer & Summers, 1990). Na prática essa divisão entre grupos pode não ser tão clara, mas modelos teóricos requerem uma distinção precisa.

O modelo Lux e Marchesi (1999) se baseia na origem exógena dos valores fundamentais e endógena dos preços de mercado (de acordo com oferta e demanda). Variações nos preços fundamentais foram tratadas como variáveis estocásticas distribuídas normalmente, com média nula e variância constante. Além disso, considerou-se que agentes migram oportunamente de um grupo para outro.

A mudança entre os estados de pessimismo e otimismo foi tratada no modelo como um resultado da combinação da tendência corrente dos preços com a opinião prevalecente (representada pela diferença entre a quantidade de pessimistas e otimistas no mercado). Já a conversão entre o comportamento de fundamentalista e de *noise trader* é consequência dos lucros momentâneos de cada grupo: *noise traders* otimistas lucram instantaneamente com mudanças positivas de preço; o resultado para os fundamentalistas acontece em um futuro incerto, com a reversão do preço ao valor fundamental; pessimistas lucram com a diferença entre o retorno de aplicações alternativas (assumido constante) e a mudança de preço do ativo vendido. Lucros favoráveis para um grupo mudam o comportamento dos participantes do outro grupo.

No longo prazo, a simulação resulta em um mercado eficiente com preços aproximando dos valores fundamentais. No entanto, esses preços apresentam propriedades estatísticas diferentes, com maior frequência de eventos extremos e *clusters* de volatilidade. Diferente de seus valores originais (brutos), os retornos em valor absoluto apresentaram memória longa e previsibilidade. Como os valores fundamentais não apresentam essas características, os autores concluem que elas advêm das interações entre os agentes, com diferentes estratégias e crenças, e particularmente de migrações entre os três grupos. Volatilidade anormal surge quando o número de *noise traders* está crítico, mas acaba quando os preços desviam substancialmente de seus valores fundamentais, ou seja, de forma endógena. Diferentes parâmetros levam a resultados similares e são confirmados por dados empíricos.

A proposição de Lux e Marchesi (1999) suporta a hipótese de que a participação de *noise traders* induz volatilidade adicional e contribui substancialmente para os fatos estilizados dos mercados financeiros. Como a participação é precedida por atenção, entender a dinâmica do grau de atenção dessa classe de investidores permite um melhor conhecimento dos mercados.

2.2 Efeitos do nível atenção nos mercados financeiros

Diversas medidas já foram utilizadas para representar o despertar de graus diferenciados da atenção do investidor e, em geral, dos diversos participantes dos mercados. Em alguns casos, essas métricas serviram para compreender melhor os resultados potencialmente sub-ótimos da atenção limitada dentre as diferentes classes de investidores.

Dentro dessa abordagem, Barber e Odean (2008) analisaram dados de transações em quatro diferentes fontes: uma corretora de ações grande destinada ao público em geral (por volta de 66 mil investidores, após filtros, de 1991 a 1996), uma corretora pequena (com em torno de 14 mil contas de investidores mais ativos e qualificados, de 1996 a 1999), uma corretora grande com pacote completo de serviços (por volta de 665 mil investidores individuais, cobrindo 30 meses finalizando em junho de 1999) e uma base de transações executadas por 43 gestores profissionais para clientes institucionais (de 1993 a 1996, classificados por sua estratégia de momento, de valor ou diversificada). Como a compra de ações demanda maior atenção de investidores individuais (considerando que as vendas a descoberto por esse público são menos comuns), o objetivo foi verificar se existe nesse nicho uma tendência de comprar ações (em relação a vender) nos momentos de atenção exacerbada.

A atenção do investidor foi medida por três diferentes variáveis: volume anormal de transações, retornos no dia anterior e presença do *ticker* em notícias divulgadas. A análise envolveu ações da NYSE, Amex e NASDAQ.

Os autores calcularam o volume financeiro anormal de transações, em um dia, pela sua razão em relação à média dos últimos 252 dias úteis. Tanto no caso do volume quanto dos retornos, as ações foram agrupadas em decis. O primeiro e último decis foram subdivididos em dois vintis.

Para construir as séries diárias de desequilíbrio entre compras e vendas, foi calculada, para cada uma das quatro fontes de informação de investidores, a diferença entre o total de compras e de vendas executadas. Esse valor foi então dividido pela soma de ambos. Os desequilíbrios foram apurados tanto em quantidade quanto em valor monetário das transações.

Os resultados mostraram desequilíbrios significativos nas transações feitas por investidores não-profissionais (ou seja, nas bases de corretoras voltadas para esse público, maiores e menores). Esse grupo apresentou comportamento predominante de compra, intensificando monotonicamente nos decis de maior volume. No caso de

retornos, esse padrão ocorreu nas duas pontas das bases de corretoras grandes, indicando que retornos extremos, tanto positivos quanto negativos, de fato capturam a atenção do investidor. Os autores também encontraram evidências de que esses investidores não se beneficiam desse comportamento.

Já no caso dos gestores de patrimônio de investidores institucionais, o comportamento das transações foi o oposto: maior tendência a comprar nos dias de menor volume, ou quando há retornos não-extremos no dia anterior. O comportamento dessa classe de investidores mostrou tendência a seguir sua estratégia pré-definida. Em relação à divulgação de notícias, não foi identificado nesse grupo o desequilíbrio expressivo observado nos demais.

Os resultados indicaram que, dentre as medidas analisadas, volume anormal é a mais informativa de atenção do investidor, seguida de retornos e presença nas notícias. Os resultados foram robustos à inclusão de variáveis para identificação de eventos como a divulgação de resultados e de distribuição de dividendos. O padrão (maior proporção de compras nos dias de maior atenção) persistiu quando foram consideradas somente as transações envolvendo ações presentes no portfólio do investidor.

Seguindo a hipótese de atenção do investidor proposta por Barber e Odean (2008), Huddart et al. (2012) verificaram a influência de preços de fechamento extremos no volume e desequilíbrio de transações, que refletem o padrão de decisão dos investidores. Preços extremos foram considerados aqueles que saem da amplitude registrada nas últimas 52 semanas. Esse horizonte foi utilizado por sua distinção e seu desempenho de explicação superior verificado em estudos anteriores.

Foram utilizados dados semanais de uma amostra aleatória de 2.000 empresas listadas na NYSE, Amex e NASDAQ, de 1982 a 2006. Volume anormais foram obtidos a partir dos resíduos de regressões dos volumes médios diários de cada empresa, percentualmente em relação ao total de ações, contra o volume médio do mercado.

Após obtidos, esses volumes foram regredidos contra *dummies* que indicam o atingimento de preços extremos, além de diversas variáveis de controle: desvio-padrão dos preços históricos, tamanho da empresa, *dummies* que indicam eventos de distribuição de dividendos e divulgação de resultados, quantidade e relevância de notícias correlatadas veiculadas.

Mesmo controlando por essas variáveis, constatou-se que preços extremos determinam um volume anormal de transações. Nos resultados, esses preços levaram o volume da mediana para o 75º percentil. O aumento no volume foi mais relevante

quanto mais tempo a amplitude é definida, quanto menor ou mais jovem é a empresa, e quanto mais volátil é o preço da ação. Após o aumento, o volume reduz gradualmente.

Também foi encontrada relação entre preços extremos e retornos subsequentes. Os autores se basearam na hipótese de que um aumento nos volumes indicaria maior interesse do investidor, levando a um desequilíbrio direcionado para compras, que por sua vez resultaria em retornos superiores. Retornos anormais foram apurados através de regressões utilizando os fatores de risco de Fama e French (1993) e Carhart (1997). Portfólios contendo ações de empresas menores e cujos preços saem da amplitude histórica apresentaram desempenho superior, tanto ponderando os pesos igualmente quanto pelo valor. O resultado mostrou-se mais expressivo quando a ação está abaixo da amplitude do que quando está acima.

Os autores segregaram as transações em dois grupos segundo o valor (um grupo até 10 mil dólares e o outro acima de 50 mil dólares) para avaliar diferenças entre investidores individuais e institucionais. Os resultados confirmaram maior volume anormal e desequilíbrio (voltado para compras) quando os preços atingem valores extremos, especialmente dentre investidores que negociam valores menores.

Séries como retornos e volumes de transação atendem apenas parcialmente ao propósito de indicar a intensidade da atenção, pois são resultantes de equilíbrios de mercado. Se excesso de atenção determina a dinâmica do mercado, faz sentido identificar medidas que traduzem o grau de atenção com maior efetividade.

2.3 Buscas na Internet como indicador de atenção dos investidores

O desenvolvimento de novas ferramentas para análise da atividade na Internet contribuiu para a identificação de medidas potencialmente mais eficazes do grau de atenção dos participantes dos mercados. Da et al. (2011) examinaram o papel do volume de buscas no Google como medida de atenção do investidor. Os autores obtiveram dados semanais do índice do volume de buscas dos *tickers* para as 3.606 ações individuais que já fizeram parte do índice Russell 3000 no período de janeiro de 2004 a junho de 2008. Esse índice contém as três mil maiores empresas dos Estados Unidos (mais de 90% do mercado). Volumes de busca de palavras-chave relacionados ao nome da empresa ou ao seu principal produto também foram extraídos.

Para ter acesso a dados a respeito de notícias, os autores buscaram no arquivo do Dow Jones se a empresa é noticiada em artigos, e quantas vezes, durante cada semana, diferenciando notícias mais de menos importantes (pela presença de chamadas, com

trechos isolados) e o percentual de palavras com sentimento negativo em relação ao tamanho da notícia.

Foram capturadas informações a respeito da quantidade de ações negociadas e de ordens executadas em cada mercado. Além disso, a análise contemplou preços e volumes de transação, informações contábeis, como o percentual de gastos com propaganda sobre a receita, e a quantidade de analistas de mercado na cobertura.

Primeiramente, foram apuradas as correlações entre o volume de buscas no Google e as diversas *proxies* de atenção. A correlação entre buscas pelo *ticker* e buscas pelo nome ou produto da empresa foi de apenas 9%, o que se explica pelas diferenças no público e nos propósitos de cada tipo de busca. A correlação entre retornos extremos e volume de transações foi de mais de 30%, mas sua correlação com o volume de buscas foi pequena, o que se justifica pelos inúmeros fatores que também influenciam essas duas variáveis. Entre volume de buscas e cobertura de notícias, a correlação foi também baixa. Segundo os autores, notícias são esporádicas e não garantem a atenção de investidores. A correlação entre volume de buscas e medidas de sentimento negativo também foi baixa.

Foi então aplicado um modelo VAR com defasagem de uma unidade nas séries de volume de buscas, volume de transações, retornos anormais absolutos e presença em notícias. Os autores encontraram coeficientes positivos e significativos para o volume de buscas defasado, em todos os cenários, indicando seu poder preditivo sobre as demais variáveis. Os resultados também mostraram um comportamento de reversão à média do volume de buscas após maiores volumes e presença em notícias, e de persistência na atenção após semanas de retornos extremos.

Regressões utilizando o volume anormal de buscas como variável dependente confirmaram sua relação positiva com o tamanho da empresa, retornos extremos e volumes anormais. Notícias mais importantes apresentaram melhor poder explicativo do volume de buscas do que notícias em geral. A regressão apresentou coeficiente de determinação de apenas 3,3%, indicando que essas *proxies* capturam apenas uma pequena parte da variação do volume anormal de buscas.

Foram então segregadas somente as ordens cobertas (que não requerem nenhum tratamento especial para execução), envolvendo menos que 10 mil ações, e com limites de preço (*marketable limit orders*). Esse tipo de ordem é feito tipicamente por investidores individuais. O volume mensal dessas ordens foi regredido contra o volume (médio) mensal de buscas e outras variáveis que tipicamente afetam volume de

transações, como *book-to-market* e percentual das ações em posse de investidores institucionais. Em cada um dos quatro grupos de tamanhos de ordens (até 10 mil ações), foi registrada relação positiva e significativa entre volume de buscas e volume de transações. Segregando as ordens por mercado, a relação mostrou-se mais intensa em centros de negociação tipicamente usados por investidores menos informados.

A partir desses resultados, foi abordada a relação entre o nível de atenção e os retornos das ações. Como verificado por Barber e Odean (2008) a força compradora advinda de investidores não-profissionais, sinalizada por choques de atenção, levaria a retornos positivos no curto prazo. O efeito encontrado foi ainda mais significativo em empresas menores, para as quais forças compradoras teriam maior impacto. A análise também evidenciou relação mais expressiva nas ações mais negociadas por investidores individuais. Retornos positivos também apresentaram relação positiva com volume anormal de transações e presença importante no noticiário.

Os autores verificaram então a hipótese de choques de preço induzidos por atenção nos casos de oferta pública inicial de ações. Nesses eventos, excessos de entusiasmo de investidores não-profissionais podem levar a retornos exagerados no primeiro dia, e performance reduzida no longo prazo.

Foram obtidas informações de empresas cujo processo de IPO ocorreu no período de janeiro de 2004 a dezembro de 2007, totalizando 185 ofertas com dados disponíveis. Nesse caso, como não há *tickers*, foi utilizado para obtenção do volume de buscas o nome da empresa ou seu principal produto. Informações referentes a notícias foram obtidas através da busca manual pelos atributos de cada empresa.

Foram confirmadas mudanças significativas no volume de buscas na época do IPO, começando de 2 a 3 semanas antes e revertendo de 2 a 3 semanas depois. Em relação aos retornos, as regressões confirmaram que IPOs com maiores volumes anormais de busca apresentam retornos no primeiro dia significativamente maiores que os demais. A análise dos retornos no longo prazo confirmou a performance inferior de ações com alto retorno no primeiro dia pós-IPO e com alto volume anormal de buscas pré-IPO. Essa reversão no longo prazo não foi encontrada para empresas com alto retorno no primeiro dia pós-IPO que não apresentaram alto volume anormal de buscas pré-IPO.

Pesquisas recentes, com base em dados mais atuais, apresentaram resultados diferentes ao investigar a relação entre volume de buscas e retornos anormais. Bijl, Kringhaug, Molnár, e Sandvik (2016) investigaram o poder preditivo de retornos de

ações a partir do volume de buscas semanal obtido no Google Trends. Os autores evidenciaram que grande volume de buscas leva a retornos negativos no curto prazo. Foram obtidos preços de abertura, ajustados por dividendos, de 431 ações que fazem parte do S&P 500, de 2008 a 2013. Os excessos de retorno foram apurados subtraindo do retorno da ação o seu beta multiplicado pelo retorno de mercado. Foi utilizado o volume padronizado de buscas dos nomes das empresas realizadas nos Estados Unidos. As volatilidades semanais e mensais foram apuradas pela raiz quadrada da soma dos quadrados dos retornos diários. Foram também utilizados excessos de volume de transações (apurados através da subtração de médias móveis históricas).

Os autores implementaram regressões em painel dos excessos de retorno, utilizando como variáveis explicativas 5 defasagens do excesso de volume, de retorno e do volume de buscas padronizado, assim como as volatilidades semanal e mensal. Um modelo estendido contemplou *dummies* para o mês de janeiro e para o período da crise financeira, além de variáveis de interação entre volume de transações, volume de buscas e retornos (defasados de um período).

O coeficiente da primeira defasagem do volume de buscas mostrou-se negativo e significativo nos dois modelos, resultado ainda mais expressivo após a crise. O da segunda defasagem mostrou-se positivo, mas menos significativo. Em ambos os modelos, o poder preditivo do volume de buscas mostrou-se superior ao do volume de transações, mas inferior ao da volatilidade e dos retornos defasados. Os resultados apresentaram robustez a variações de medidas. A robustez foi reforçada pelo resultado não-significativo do uso de palavras-chave aleatórias para o volume de buscas, ou de volumes deslocados no tempo.

Foi então implementada uma estratégia de compra predominante de ações com baixo volume de buscas, com rebalanceamento semanal. Ela apresentou desempenho aproximadamente 16% superior a uma estratégia simples, com pesos igualmente ponderados, ao longo do período de 5 anos. No entanto, em um cenário com custos de transação que incluíram corretagem de 0,02% e *spread* entre compra e venda de 0,08%, a estratégia ativa apresentou performance inferior em cerca de 5%.

A capacidade preditiva do volume de buscas sobre a volatilidade também tem sido objeto de estudos recentes. Dimpfl e Jank (2012) analisaram o efeito que o volume de buscas no Google transmite à volatilidade dos preços no mercado dos Estados Unidos. A análise contemplou séries de julho de 2006 a dezembro de 2011 do índice Dow Jones, bolsa cujas informações são mais buscadas por investidores não-

profissionais, segundo os autores. A volatilidade diária do índice foi obtida pela raiz quadrada do somatório dos quadrados de seus retornos intradiários, utilizando 10 minutos de intervalo.

Os autores consideraram o índice representativo do volume de buscas da palavra-chave “dow” feitas nos Estados Unidos em dias de negociação. Foi constatado que essa palavra-chave é mais usada por investidores interessados no índice.

As séries de volatilidade realizada e volume de buscas apresentaram uma correlação de 82% no período. Já o volume de transações na bolsa apresentou correlação de 51% com o volume de buscas e 57% com a volatilidade.

O modelo VAR desenvolvido contemplou séries logarítmicas diárias da volatilidade realizada, quantidade de buscas e volume de transações, com defasagem máxima de 3 dias. Como esperado, para todas as variáveis, os resultados indicaram significância estatística nos termos de autocorrelação em quase todas as defasagens. Além disso, os resultados mostraram que o volume de buscas não é determinado pela volatilidade passada, mas apenas pela volatilidade presente.

Como esperado, os resultados indicaram que o volume de buscas do dia anterior determina de forma positiva e significativa a volatilidade. Já o volume de transações apresentou poder preditivo sobre o volume de buscas apenas marginal, na terceira defasagem. O volume de buscas, por outro lado, determinou significativamente o volume de transações. A relação entre as variáveis foi também confirmada nos testes de causalidade de Granger.

Para testar a capacidade de previsão, os autores fizeram testes dentro e fora-da-amostra, para diferentes horizontes e frequências de volatilidade e volumes de buscas. Além de modelos autorregressivos tradicionais (com defasagem autoregressiva de 1 e 4 períodos, e defasagem na variável externa de 1 período), foi testado também o modelo HAR (autorregressivo heterogêneo), que melhor captura aspectos de memória longa. Nessa configuração, os componentes autoregressivos da volatilidade são: um para o período anterior, outro com a média dos últimos 5 dias e outro com a média dos últimos 22 dias. Os modelos VAR e VHAR (bivariados) foram comparados com modelos AR e HAR (univariados). O desempenho de previsão da volatilidade foi feito através do erro quadrático médio, a função de quase-verossimilhança e o coeficiente de determinação da regressão dos valores reais contra os previstos.

Em relação às configurações univariadas, os resultados dentro-da-amostra, prevendo 1 dia à frente, indicaram melhor desempenho do modelo AR(4) em relação

ao AR(1) e ao HAR (exceto pela função de quase-verossimilhança). Os modelos bivariados, no entanto, apresentaram performance superior aos univariados sob todos os indicadores. Apresentaram melhor desempenho geral os modelos VAR(4) e VHAR.

As previsões fora-da-amostra contemplaram o horizonte de 1 dia, 1 semana e 2 semanas. Horizontes de 1 e 2 semanas foram feitos pela consolidação dos dados diários. O modelo foi estimado inicialmente utilizando os primeiros 2 anos da amostra, e reestimado a cada instante utilizando toda a informação passada disponível. Dentre as configurações univariadas, os resultados apontaram superioridade no desempenho do HAR, especialmente para os horizontes maiores. Por modelar a dinâmica do grau de atenção dos investidores individuais, as configurações bivariadas desempenham melhor que as univariadas, especialmente para horizontes de 1 e 2 semanas. A diferença de performance de predição do modelo VHAR em relação ao HAR, tanto dentro quanto fora-da-amostra, mostrou-se mais expressiva nos momentos de maior turbulência no mercado. Mesmo considerando o atraso de 1 dia para liberação dos dados de volume de busca, os modelos bivariados apresentaram performance fora-da-amostra superior (apesar de reduzida), e o VHAR continuou sendo o modelo de melhor performance geral.

Os resultados dos estudos feitos em mercados desenvolvidos acerca das particularidades e efeitos do grau de atenção dos agentes justificam uma abordagem no contexto do mercado brasileiro. Optamos nesse estudo por buscar evidências de indução de volatilidade adicional por excesso de atenção dos investidores individuais no mercado de ações do Brasil. Partimos então da hipótese de que medidas de atenção do investidor individual contribuem para explicar volatilidades condicionais no mercado de ações.

3. Metodologia

Nesta parte do documento, é feita uma caracterização dos dados amostrais utilizados, fontes das informações, modelos empíricos, coeficientes de interesse e sinais esperados.

3.1.1 Obtenção dos dados

Optamos por extrair dados referentes ao índice Ibovespa, representativo do mercado brasileiro de ações, já que a busca por índices de mercado tende a ser menos ambígua do que a de ações, o que favorece sua utilização na análise do impacto do volume de buscas (Da et al., 2011). Foram então obtidas séries de log-retornos nominais diários e do volume de negócios relativos ao índice Ibovespa, no software Economática, correspondentes ao período de 2004 ao primeiro trimestre de 2016. Os retornos são ajustados pela distribuição de proventos.

Como medida de atenção, foi utilizado o volume de buscas feitas no Google. O Google é com larga vantagem a ferramenta de busca mais utilizada no Brasil (Statista, 2014). Estudos como o de Dimpfl e Jank (2012) investigaram esse efeito no mercado estadunidense. No Google Trends, foi então extraída a série diária do índice que representa o volume relativo de buscas da palavra-chave “Ibovespa” no Google. Consideramos apenas as buscas feitas no Brasil, para minimizar efeitos relacionados a fusos horários. O período escolhido para o estudo é suficientemente longo por conter períodos de calma e turbulência, e limita-se pela data mais antiga em que o Google Trends disponibiliza informações.

Foram considerados apenas os dias em que houve pregão na BM&FBovespa. Como o Google Trends só disponibiliza dados diários em intervalos trimestrais, extraímos cada trimestre separadamente e concatenamos as séries conforme proposto por Johansson (2014). A série foi então normalizada para ter média unitária.

3.1.2 Modelagem

A modelagem da volatilidade dos retornos das ações é feita por estudiosos desde quando ela foi adotada como importante medida de risco. Os modelos da família ARCH (autoregressivo de heteroscedasticidade condicional) foram inicialmente propostos por Engle (1982) e estendidos por Bollerslev (1986), que propôs o ARCH generalizado (GARCH), mais abrangente. Esses modelos consideram que a volatilidade varia com o tempo e pode ser prevista através dos valores realizados. Dada a sua relativa popularidade e eficácia, a volatilidade condicional do Ibovespa (σ_t^2) foi medida por um modelo GARCH, representado pelas equações a seguir:

$$r_t = \mu + a_t \qquad a_t = \sigma_t \varepsilon_t \qquad (1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^s \alpha_i a_{t-i}^2 + \sum_{j=1}^m \beta_j \sigma_{t-j}^2 \qquad (2)$$

O modelo assume que a série de log-retornos r_t do índice Ibovespa não apresenta correlações seriais, o que é averiguado pelo teste de Ljung-Box e por critérios de informação. A série $a_t = \mu - r_t$ corresponde ao choque no tempo t , com variância incondicional finita. O componente ε_t é uma variável aleatória independente e identicamente distribuída, com média nula e variância unitária.

As ordens s e m indicam defasagens dos componentes do modelo de volatilidade. As ordens e distribuição mais adequadas dos parâmetros são obtidas através da análise da significância dos coeficientes, dos resultados dos testes LM e de Ljung-Box dos resíduos e dos critérios de informação bayesiano (BIC) e de Akaike (AIC). Espera-se que os coeficientes não-negativos α_i e β_j apresentem valores estatisticamente significativos.

Examinamos a relação entre volatilidade e volume de buscas através da modelagem de vetores autorregressivos (VAR). Definimos a defasagem máxima (p) mais adequada avaliando os critérios de informação. Considerando \mathbf{x}_t um vetor contendo as séries de volatilidade (σ_t^2), volume de buscas (BU_t) e volume de negociações (NE_t), todas na forma logarítmica, o modelo VAR(p) é apresentado a seguir:

$$\mathbf{x}_t = \mathbf{c} + \sum_{j=1}^p \beta_j \mathbf{x}_{t-j} + \boldsymbol{\varepsilon}_t \qquad (3)$$

O componente \mathbf{c} é o vetor de constantes e $\boldsymbol{\varepsilon}_t$ é o vetor de choques, i.i.d. Foram construídos três diferentes modelos, combinando as três séries objetos de análise: um modelo contempla as séries de volatilidade e volume de buscas (especificação 3.1); outro considera as séries de volatilidade e volume de negócios (3.2); o terceiro leva em conta as três séries (3.3). Espera-se que os coeficientes dos volumes de busca apresentem valores positivos em significativos na modelagem da volatilidade, mesmo

na presença do volume de transações como regressor. Confirmamos as relações esperadas através do teste de causalidade de Granger.

4. Resultados

São apresentados nesta parte do trabalho os achados do experimento que envolve medidas de atenção e volatilidade do mercado brasileiro de ações.

4.1.1 Análise descritiva

Os dados referentes ao volume de buscas foram obtidos entre 15 e 18 horas de 18/06/16. A palavra-chave “Ibovespa” foi escolhida por ter mais quantidade de buscas do que outras similares, como “Ibov”, e por se referir especificamente ao índice, diferente de “Bovespa”. A correlação entre a palavra-chave “Ibovespa” e correlatas é mostrada na tabela 1.

Tabela 1
Correlação entre palavras-chave relacionadas com o Ibovespa

Correlações e representatividades (razão) de palavras-chave únicas e correlatas em relação ao termo de busca “Ibovespa”, feitas no Brasil desde 2004, obtidas no Google Correlate e Google Trends. Extração feita em às 22 horas em 23/07/16.

Palavra-chave	Correlação	Representatividade
Bovespa	0,9570	6,25
Ibov	0,7939	0,04
Ações	0,7831	8,17
Cotação	0,7829	14,67
BMF	0,5289	0,09

Os valores correspondentes ao período de 11 a 24 de abril de 2004 foram ignorados por indicarem volume de buscas nulo. A tabela 2 mostra a estatística descritiva das séries utilizadas no trabalho, em sua forma normal e logarítmica. Cada série contempla 3.018 observações. Na tabela é possível evidenciar a alta variabilidade das séries de retornos e volume de buscas, assim como a redução substancial na assimetria e curtose do volume de buscas após conversão em escala logarítmica.

Tabela 2
Estatística descritiva das séries temporais

Principais medidas estatísticas das séries de log-retornos (nominais ajustados), volume de buscas no Google feitas no Brasil (índice do Google Trends), quantidade de negócios (em milhões de negociações) e valor dos negócios (em bilhões de reais) relativos ao índice Ibovespa, sendo as 3 últimas nas formas normal e logarítmica. Os dados são diários de 2004 ao primeiro trimestre de 2016.

	Log- retornos r_t	Volume de buscas BU_t	$\log BU_t$	Volume de negócios NE_t	$\log NE_t$	Valor dos negócios VA_t	$\log VA_t$
	%			Milhões		R\$ bilhões	
Média	0,03	1,23	-0,10	0,38	0,67	4,60	0,58
Desvio- padrão	1,79	2,15	0,37	0,33	0,54	2,44	0,30
Mediana	0,06	0,74	-0,13	0,30	-0,53	4,88	0,69
Mínimo	-12,10	0,07	-1,16	0,01	-2,10	0,20	-0,70
Máximo	13,68	54,43	1,74	2,41	0,38	25,9	1,41
Assimetria	-0,01	12,02	0,51	0,76	-0,48	0,75	-0,96
Curtose	4,95	219,54	0,92	0,11	-1,13	-0,96	0,17

4.1.2 Modelagem GARCH

Primeiramente, utilizamos as séries diárias de log-retornos do Ibovespa para apurar sua volatilidade condicional. O teste t de média não rejeitou a hipótese de média nula (p-valor 38,63%). Apesar de o teste de Ljung-Box rejeitar a hipótese de inexistência de autocorrelação (p-valor 0,19%), o critério de informação bayesiano indicou ordens nulas como mais adequadas para a modelagem da média (equação 1).

Como esperado, o teste de Ljung-Box da série ao quadrado e o teste do multiplicador de Lagrange (LM) indicaram a presença de efeitos ARCH (p-valor abaixo de 0,00%). Os modelos GARCH (equação 2) de ordens (2,1) e (2,2), com distribuições t e normal, mostraram-se adequados pelos testes LM e de Ljung-Box dos resíduos, como indica a tabela 3. Modelos de ordens superiores apresentaram coeficientes não-significativos. Selecionamos o modelo de ordem (2,1), com distribuição t, por ser mais parcimonioso de acordo com os critérios de informação.

Na figura 1, mostramos graficamente, na forma normal e logarítmica, as quatro séries utilizadas na modelagem de vetores autorregressivos: a volatilidade resultante do modelo GARCH(2,1) com distribuição t, a quantidade de buscas na Internet, a quantidade e o volume financeiro de transações associadas ao índice. Nos gráficos com

a evolução da volatilidade e volume de buscas fica nítido o pico referente à crise financeira de 2008.

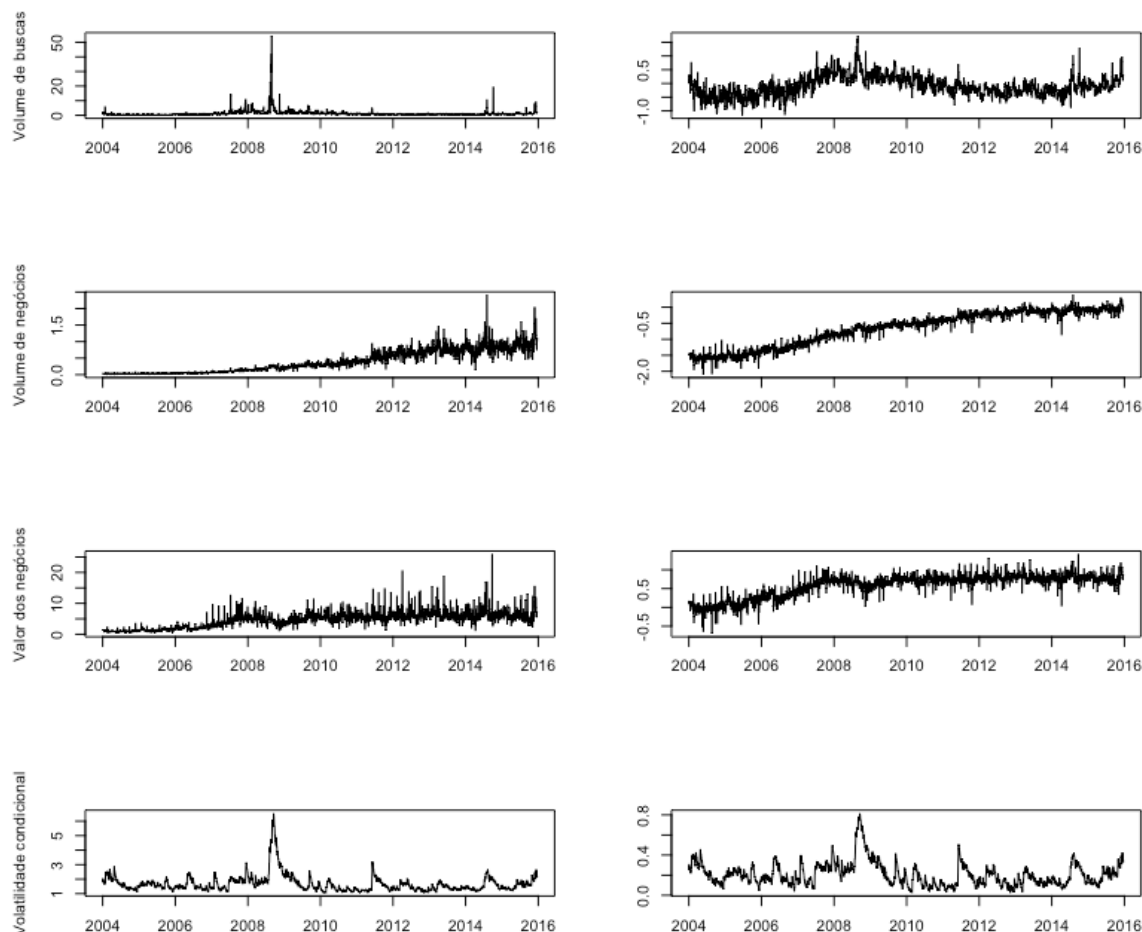
Tabela 3
Modelagem GARCH dos log-retornos do Ibovespa

Resultado da modelagem GARCH (equação 2) dos log-retornos diários nominais ajustados do índice Ibovespa, de 2004 ao primeiro trimestre de 2016. Foram selecionadas as ordens (2,1) e (2,2), com distribuições normal e t, por serem adequados segundo os testes de Ljung-Box e LM. ***: significativo ao nível de 1%; **: significativo ao nível de 5%; *: significativo ao nível de 10%.

Modelagem GARCH				
	GARCH(2,1)		GARCH(2,2)	
	Normal	T	Normal	t
μ	0,05*	0,06**	0,05**	0,06**
ω	0,07***	0,07***	0,10***	0,08***
α_1	0,03*	0,02	0,03*	0,02
α_2	0,05***	0,06***	0,09***	0,08***
β_1	0,90***	0,90***	0,36	0,57*
β_2			0,49*	0,30
Distrib.		10,00***		10,00***
<i>Log likelihood</i>	-5.695,86	-5.678,64	-5.694,18	-5.678,13
Testes de Ljung-Box				
R				
Q(10)	9,90	9,85	10,03	9,89
Q(15)	14,20	14,11	14,17	14,06
Q(20)	24,50	24,44	24,06	24,15
R²				
Q(10)	12,07	12,79	9,97	10,44
Q(15)	18,85	19,82	16,51	17,24
Q(20)	25,95	26,86	23,53	24,18
Teste LM	14,01	14,78	11,51	12,13
Critério de informação				
AIC	3,778	3,767	3,777	3,767
BIC	3,788	3,779	3,789	3,781
SIC	3,778	3,767	3,777	3,767
HQIC	3,781	3,771	3,782	3,772

Figura 1
Visualização gráfica das séries

Evolução do volume de buscas (índice do Google Trends), da quantidade de negócios (em milhões), do valor negociado (em R\$ bilhões) e da volatilidade condicional, apurada segundo um modelo GARCH(2,1) com distribuição t, relativos ao índice Ibovespa, desde 2004 ao primeiro trimestre de 2016. Os gráficos à esquerda correspondem à forma normal e, à direita, à forma logarítmica das séries.



A tabela 4 mostra as correlações entre as séries utilizadas, nas formas normal e logarítmica. Como esperado, a correlação entre quantidade e volume financeiro de negócios é relativamente alta e positiva (variando de 0,73 a 0,90).

Os valores mostram uma correlação relativamente baixa (variando de -0,03 a 0,39) entre volume de buscas e quantidade (ou volume financeiro) de negócios. Como afirma Da et al. (2011), esse resultado ocorre porque o volume de negócios é resultante do equilíbrio entre diversos fatores econômicos, além do grau de atenção.

Já a volatilidade do Ibovespa apresenta correlação razoável (variando de 0,41 a 0,46) com o volume de buscas. Isso indica que a atenção desperta imediatamente na ocorrência de um choque de volatilidade, como também verificado em estudos como o

de Dimpfl e Jank (2012). Por outro lado, a volatilidade apresenta correlação relativamente baixa (e negativa, variando de -0,15 a -0,07) com a quantidade ou volume financeiro de negócios.

Tabela 4
Matriz de correlações entre as séries

Correlações contemporâneas amostrais entre as séries diárias, de 2004 ao primeiro trimestre de 2016, de volume de buscas no Google, de quantidade e de volume financeiro dos negócios e da volatilidade condicional do Ibovespa, medida através de um modelo GARCH(2,1) com distribuição t.

Matriz de correlações								
	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
a. Volume de buscas (BU_t)	1							
b. $\log BU_t$	0,67	1						
c. Volume de negócios (NE_t)	-0,03	0,04	1					
d. $\log NE_t$	0,06	0,22	0,90	1				
e. Valor dos negócios (VA_t)	0,14	0,31	0,77	0,83	1			
f. $\log VA_t$	0,16	0,39	0,73	0,90	0,92	1		
g. Volatilidade condicional (σ_t^2)	0,46	0,43	-0,13	-0,09	-0,08	-0,07	1	
h. $\log \sigma_t^2$	0,41	0,42	-0,15	-0,13	-0,11	-0,10	0,96	1

4.1.2 Modelagem VAR

A tabela 5 apresenta o resultado da modelagem de vetores autoregressivos, conforme equação 3, a partir da forma logarítmica das séries de volatilidade, volume de buscas e a quantidade de negócios. As especificações 3.1 e 3.2 combinam duas das três séries, enquanto a 3.3 contempla as três séries.

Usando o BIC e o AIC como critérios de seleção, os modelos VAR(6) e VAR(10) apresentaram as ordens mais adequadas. Optamos pelo modelo com defasagem máxima de 6 dias por ser mais parcimonioso, visto que os resultados do VAR(10) não são qualitativamente diferentes. O painel A da tabela 5 mostra as estimativas e valores p de cada coeficiente.

Como esperado, os resultados mostram que todas as séries apresentam forte autocorrelação. Em todas as configurações, a variável dependente apresenta efeitos significativos em boa parte de seus componentes defasados. Esses efeitos são sempre positivos na série de quantidade de negócios. Já na volatilidade e no volume de buscas, parte das defasagens apresenta sinal negativo e parte mostra efeito positivo.

Na modelagem da volatilidade ($\log \sigma_t^2$), mesmo após considerar o efeito da autocorrelação, o coeficiente relativo ao volume de buscas na primeira defasagem ($\log BU_{t-1}$) é positivo e estatisticamente significativo a 1%, tanto na configuração 3.1 quanto na 3.3. Isso vai ao encontro da hipótese de que o grau de atenção contribui para antecipar maiores flutuações nos preços. Na terceira defasagem, o coeficiente apresenta valor negativo e significativo, indicando que o grau de agitação se assenta após um período.

Ademais, na modelagem do volume de buscas ($\log BU_t$) não se verifica contribuição dos componentes da volatilidade. Não foram encontrados coeficientes estatisticamente significativos para $\log \sigma^2$ em nenhuma defasagem. Isso também ocorre tanto na configuração 3.1 quanto na 3.3. Esse resultado indica que a indução de volatilidade adicional por atenção exacerbada é mais acentuada do que a relação inversa, o que corrobora com os achados de Dimpfl e Jank (2012) na bolsa dos Estados Unidos.

A inclusão da quantidade de negócios ($\log NE_t$) no modelo 3.3 não altera a relação encontrada entre buscas e volatilidade. Além disso, a primeira defasagem de $\log NE_t$ apresentou coeficiente positivo e estatisticamente significativo a 1% na modelagem do volume de buscas, tanto no modelo 3.2 quanto no 3.3. Na terceira defasagem, o coeficiente é negativo e estatisticamente significativo a 1%.

O coeficiente que indica relevância do volume de buscas no modelo da quantidade de transações também foi estatisticamente significativo a 1% na primeira defasagem. Esses resultados indicam que maior atenção dos *noise traders* à bolsa precede um aumento no volume de transações que, por sua vez, chama maior atenção (por algum tempo). Essa evidência também foi verificada em trabalhos como o de Da et al. (2011) no cenário estadunidense.

O modelo 3.3 indicou relevante poder de determinação de $\log NE_t$ no modelo de $\log \sigma_t^2$, com sinais positivos nas defasagens mais recentes e negativos nas menos recentes. A reciprocidade (determinação no sentido inverso) também ocorre, mas de forma bem menos expressiva, com poucos componentes significativos.

Tabela 5
Modelagem por vetores autorregressivos

Resultados dos três modelos autorregressivos (equação 3) com defasagem de 6 dias: (3.1) utilizando as séries de volatilidade e buscas, (3.2) de buscas e volume de negócios e (3.3) de volatilidade, buscas e volume de negócios. Todas as séries apresentam-se em sua forma logarítmica. A volatilidade condicional foi feita por um modelo GARCH(2,1) com distribuição t, a partir dos log-retornos do Ibovespa. Os dados diários compreendem o período de 2004 ao primeiro trimestre de 2016. No painel A, os coeficientes dos modelos são apresentados com seus respectivos valores p entre parênteses. No painel B, são exibidas as estatísticas do teste $\chi^2(4)$ de causalidade de Granger, com os valores p entre parênteses. 1, 2 e 3 asteriscos denotam significância nos níveis de 10%, 5% e 1%, respectivamente.

Painel A – Estimativa por vetores autorregressivos							
	Modelo 3.1		Modelo 3.2		Modelo 3.3		
	$\log \sigma_t^2$	$\log BU_t$	$\log BU_t$	$\log NE_t$	$\log \sigma_t^2$	$\log BU_t$	$\log NE_t$
$\log \sigma_{t-1}^2$	1,38*** (0,00)	-0,28 (0,18)			1,36*** (0,00)	-0,11 (0,61)	-0,04 (0,73)
$\log \sigma_{t-2}^2$	-0,55*** (0,00)	0,29 (0,41)			-0,51*** (0,00)	0,18 (0,63)	-0,14 (0,44)
$\log \sigma_{t-3}^2$	0,23*** (0,00)	-0,03 (0,95)			0,22*** (0,00)	-0,05 (0,90)	-0,33* (0,10)
$\log \sigma_{t-4}^2$	-0,13*** (0,00)	0,43 (0,25)			-0,10*** (0,00)	0,44 (0,24)	0,32* (0,10)
$\log \sigma_{t-5}^2$	0,06** (0,04)	-0,54 (0,12)			0,04 (0,15)	-0,59* (0,09)	-0,01 (0,97)
$\log \sigma_{t-6}^2$	-0,02 (0,36)	0,19 (0,36)			-0,03 (0,11)	0,22 (0,28)	0,14 (0,18)
$\log BU_{t-1}$	0,01*** (0,00)	0,64*** (0,00)	0,64*** (0,00)	0,03*** (0,00)	0,01*** (0,00)	0,63*** (0,00)	0,03*** (0,00)
$\log BU_{t-2}$	0,00 (0,16)	0,18*** (0,00)	0,18*** (0,00)	0,00 (0,79)	0,00 (0,39)	0,18*** (0,00)	0,00 (0,72)
$\log BU_{t-3}$	-0,01*** (0,00)	0,04* (0,06)	0,05** (0,03)	0,00 (0,84)	-0,01*** (0,01)	0,05** (0,04)	0,00 (0,99)
$\log BU_{t-4}$	0,00 (0,68)	-0,03 (0,25)	-0,02 (0,29)	-0,02** (0,04)	0,00 (0,49)	-0,03 (0,23)	-0,02* (0,10)
$\log BU_{t-5}$	0,00 (0,23)	-0,09*** (0,00)	-0,09*** (0,00)	0,00 (0,83)	0,00* (0,08)	-0,09*** (0,00)	0,00 (0,96)
$\log BU_{t-6}$	0,00 (0,28)	0,16*** (0,00)	0,16*** (0,00)	-0,01 (0,54)	0,00 (0,49)	0,16*** (0,00)	-0,01 (0,61)

(cont.) Painel A – Estimativa por vetores autorregressivos							
	Modelo 3.1		Modelo 3.2		Modelo 3.3		
	$\log \sigma_t^2$	$\log BU_t$	$\log BU_t$	$\log NE_t$	$\log \sigma_t^2$	$\log BU_t$	$\log NE_t$
$\log NE_{t-1}$			0,10***	0,46***	0,03***	0,12***	0,44***
			(0,00)	(0,00)	(0,00)	(0,00)	(0,00)
$\log NE_{t-2}$			-0,02	0,12***	0,03***	-0,01	0,11***
			(0,59)	(0,00)	(0,00)	(0,75)	(0,00)
$\log NE_{t-3}$			-0,15***	0,04**	-0,04***	-0,14***	0,04**
			(0,00)	(0,04)	(0,00)	(0,00)	(0,05)
$\log NE_{t-4}$			0,06	0,14***	0,00	0,05	0,16***
			(0,14)	(0,00)	(0,40)	(0,18)	(0,00)
$\log NE_{t-5}$			-0,01	0,16***	-0,02***	-0,02	0,17***
			(0,88)	(0,00)	(0,00)	(0,68)	(0,00)
$\log NE_{t-6}$			0,02	0,08***	-0,01***	0,02	0,07***
			(0,50)	(0,00)	(0,01)	(0,69)	(0,00)
Const. c	0,01***	-0,02***	0,00	0,00	0,00***	-0,02**	0,01**
	(0,00)	(0,01)	(0,79)	(0,32)	(0,00)	(0,02)	(0,04)

Painel B – Testes de causalidade de Granger							
	Modelo 3.1		Modelo 3.2		Modelo 3.3		
	$\log \sigma_t^2$	$\log BU_t$	$\log BU_t$	$\log NE_t$	$\log \sigma_t^2$	$\log BU_t$	$\log NE_t$
$\log \sigma_{t-j}^2$		1,65				1,93*	8,73***
		(0,13)				(0,07)	(0,00)
$\log BU_{t-j}$	10,95***			3,43***	7,22***		3,96***
	(0,00)			(0,00)	(0,00)		(0,00)
$\log NE_{t-j}$			4,09***		59,53***	4,36***	
			(0,00)		(0,00)	(0,00)	

As estatísticas dos testes de causalidade de Granger são apresentadas no painel B da tabela 5. Os resultados evidenciam que volume de buscas hoje Granger-cause volatilidade amanhã, o que vai ao encontro dos coeficientes encontrados no painel A e da hipótese de pesquisa. Os testes também indicam que a quantidade de negócios em dias anteriores fornece informação relevante sobre o volume de buscas (nível de atenção) e volatilidade presentes. Por outro lado, choques no nível de atenção têm relação direta com o volume de negócios futuro. Essas relações ocorreram tanto nos modelos bivariados (3.1 e 3.2) quanto no trivariado (3.3). Assim como na análise dos

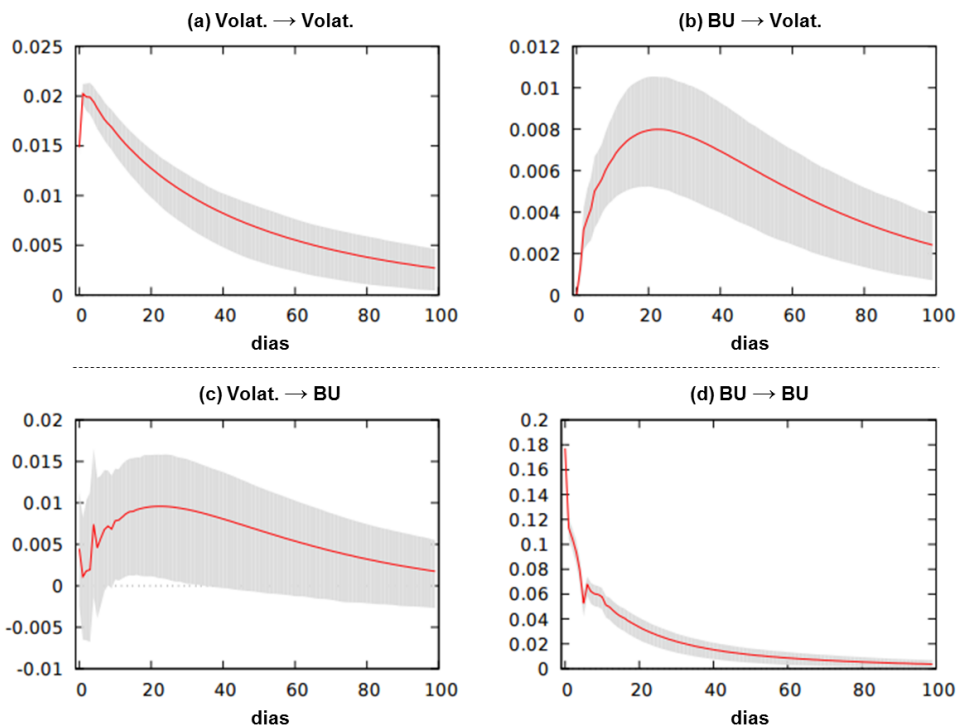
coeficientes, não houve indícios conclusivos de que volatilidade passada Granger-cause atenção dos *noise traders*, apesar de determinar volume total de negócios.

Repetimos a modelagem utilizando o volume financeiro ou invés da quantidade de negócios. Todas as relações encontradas entre volume de buscas, de transações e volatilidade do índice Ibovespa persistiram, exceto para casos alguns isolados em defasagens superiores.

Na figura 2, são apresentados os resultados das funções de resposta ao impulso sobre o modelo bivariado envolvendo volatilidade e volume de buscas (3.1). Os gráficos (a) e (d) confirmam as respostas persistentes de volatilidade e buscas a seus próprios choques, como esperado. No gráfico (c), a resposta com decaimento lento evidencia o impacto também persistente de um choque no volume de buscas sobre a volatilidade, o que corrobora com a hipótese de pressão dos preços induzida por excesso de atenção. O gráfico (b) indica que também há elevação da atenção em resposta a choques de volatilidade.

Figura 2
Funções de resposta ao impulso das séries

(a) Resposta da volatilidade condicional do Ibovespa (modelo GARCH(2,1) com distribuição t, a partir dos log-retornos) a um choque de volatilidade. (b) Resposta da volatilidade a choque de volume de buscas (índice do Google Trends). (c) Resposta do volume de buscas a choque de volatilidade. (d) Resposta do volume de buscas a choque de volume de buscas. Os resultados foram obtidos pelo modelo VAR(6) a partir de séries diárias logarítmicas desde 2004 ao primeiro trimestre de 2016. Os gráficos destacam em área sombreada um intervalo de confiança de 95%. A resposta ao impulso foi simulada por 100 dias.



5. Considerações finais

Neste estudo, investigamos no mercado brasileiro a hipótese de pressão dos preços induzida por excesso de atenção. Atenção é requisito para sentimento e sentimento exacerbado pode levar a uma alocação sub-ótima da atenção, que, por sua vez, resulta em vieses de comportamento (Kahneman, 1973).

Se níveis superiores de atenção podem indicar sentimento exacerbado, espera-se que a mensuração do grau de atenção de grupos de investidores com influência relevante, como é o caso dos *noise traders*, possa contribuir para entender melhor a volatilidade dos mercados. Utilizamos como medida de atenção dos *noise traders* o volume de buscas realizadas no popular *site* de buscas Google. A capacidade dessa medida para esse propósito apresentou resultados convincentes nos estudos de Dimpfl e Jank (2012) e Da et al. (2011).

Adotamos o Ibovespa para o trabalho pela sua popularidade como índice representativo do mercado de ações brasileiro. Além disso, a motivação para a busca por índices de mercado tende a ser menos ambígua do que a por empresas individuais. Optamos pela palavra-chave “Ibovespa” para obter os volumes de busca pela sua relevância e por melhor simbolizar a busca especificamente pelo índice. Palavras-chave similares apresentaram séries de volume de buscas bastante correlacionadas com a da “Ibovespa”.

Após modelar a volatilidade diária do índice, verificamos o poder de determinação de valores defasados do volume de buscas sobre essa volatilidade, controlando pela quantidade de transações e pelas autocorrelações das séries.

Os resultados indicaram que, de fato, há indução de volatilidade adicional nos momentos antecedidos por maior atenção. Essa relação mostrou-se bastante acentuada na defasagem de um dia. Trata-se de uma evidência também observada nos Estados Unidos por Dimpfl e Jank (2012), a despeito das diferenças entre os mercados em termos de grau de concentração dos ativos e de influência de investidores individuais.

Os resultados também suportam a ideia de que maior volume de transações desperta a atenção dos investidores individuais e determina aumento na variância do índice. Isso corrobora com estudos que sugerem o uso do volume de transações como medida do grau de atenção (Barber & Odean, 2008; Da et al., 2011), mas é oposto ao encontrado em outros mercados por Brooks (1998) e Dimpfl e Jank (2012).

Os achados vão ao encontro da hipótese de pesquisa de que parte deste trabalho: nos momentos de atenção exacerbada, o comportamento dos investidores individuais

induz volatilidade adicional no mercado de ações brasileiro. Como no modelo multiagentes descrito por Lux e Marchesi (1999), uma atuação mais intensa dos *noise traders* muda a dinâmica de interação entre os agentes e dá aos preços maior instabilidade.

Trabalhos complementares podem abordar a capacidade preditiva do grau de atenção sobre a atividade em outros mercados, como o monetário, de câmbio e até de trabalho. O estudo de sua influência sobre retornos anormais no curto prazo, especialmente em eventos de IPO (oferta inicial pública de ações), também apresenta potencial para interessantes achados.

Referências

- Andreassen, P., Kraus, S. (1990). Judgmental Extrapolation and the Saliency of Change. *Journal of Forecasting*, 9, 347-372.
- Barber, B., & Odean, T. (2008). All that Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies*, 21(2), 785-818.
- Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google Searches and Stock Returns. *International Review of Financial Analysis*, 45, 150-156.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 21, 307-328.
- Brooks, C. (1998). Predicting Stock Index Volatility: Can Market Volume Help? *Journal of Forecasting*, 17(1), 59-80.
- Carhart, M. M. (1997). On Persistence in Mutual Fund Performance. *Journal of Finance*. 52(1), 57-82.
- Challet, D., & Ahmed, B. H. A. (2013). *Predicting Financial Markets with Google Trends and Not so Random Keywords*. arXiv preprint arXiv:1307.4643.
- Choi, H., & Varian, H., (2009). *Predicting Initial Claims for Unemployment Benefits*, Relatório Técnico. Google Inc.
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *The Journal of Finance*, 66, 1461-1499.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4), 703-738.
- Dimpfl, T., & Jank, S. (2016). Can Internet Search Queries Help to Predict Stock Market Volatility? *European Financial Management*, 22, 2, 171-192.
- Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation. *Econometrica*, 50, 987-1007.
- Fama, E., & French, K. (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*. 33(1), 3-56.

- Gillings, M. R., Hilbert, M., & Kemp, D. J. (2016). Information in the Biosphere: Biological and Digital Worlds. *Trends in Ecology & Evolution*, 31(3), 180-189.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting Influenza Epidemics Using Search Engine Query Data. *Nature*, 457, 1012-1014.
- Gomes, P. & Taamouti, A. (2016). In Search of the Determinants of European Asset Market Comovements. *International Review of Economics and Finance*, 44, 103-117.
- Huddart, S. J., Lang, M. H., & Yetman, M. (2012). Volume and Price Patterns Around a Stock's 52-Week Highs and Lows: Theory and Evidence. *Management Science*, 55(1), 16-31.
- Johansson, E. (2014). *Creating Daily Search Volume Data from Weekly and Daily Data*. Disponível em erikjohansson.blogspot.com.
- Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kristoufek, L. (2013). Can Google Trends Search Queries Contribute to Risk Diversification? *Scientific Reports*, 3.
- Lux, T., & Marchesi, M. (1999). Scaling and Criticality in a Stochastic Multi-agent Model of a Financial Market. *Nature*, 397, 498-500.
- Ministério do Planejamento, Desenvolvimento e Gestão (2015). *Mercado de Capital Brasileiro está em 12º no Ranking Mundial*. Disponível em www.planejamento.gov.br/assuntos/assuntos-economicos/noticias/mercado-de-capital-brasileiro-esta-em-12o-no-ranking-mundial.
- Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*, 3.
- Preis, T., Reith, D., & Stanley, H. E. (2010). Complex Dynamics of Our Economic Life on Different Scales: Insights from Search Engine Query Data. *Philosophical Transactions of the Royal Society A*, 368, 5707-19.
- Tversky, A., & Kahneman, D. (1974). Judgment Under Uncertainty: Heuristics and Biases. *Science*, 185, 1124-31.
- Shleifer, A., & Summers, L. (1990). The Noise Trader Approach to Finance. *Journal of Economic Perspective*, 4, 2, 19-33.
- Statista (2014). *Popular Online Search Engines in Brazil as of May 2014, based on Market Share*. Disponível em www.statista.com/statistics/309652/brazil-market-share-search-engine.