

Atenção do Investidor: O Volume de Buscas no Google é Capaz de Prever os Retornos de Ações?

Claudia Yoshinaga¹

claudia.yoshinaga@fgv.br |  0000-0002-7756-6083

Fabio Rocco¹

f.bouroc@gmail.com |  0000-0003-0151-7231

RESUMO

Este artigo investiga o papel da atenção do investidor na previsão de retornos futuros de ações brasileiras usando o Volume de Buscas no Google (GSV, em inglês). Testamos se variações defasadas do GSV explicam retornos anormais em uma amostra de 57 ações do Ibovespa. Dados semanais de pesquisas feitas no Google Brasil de 2014 a 2018 foram analisados para verificar a relação. De forma semelhante a pesquisas anteriores no mercado americano, descobrimos que aumentos no GSV são seguidos por menores retornos anormais futuros. Além disso, mostramos que quanto mais negociada uma ação, maior o efeito. Isso é consistente com a hipótese de que uma maior atenção do investidor individual leva a menores retornos subsequentes, sugerindo que uma maior popularidade leva as ações a se desviarem de seu valor fundamental.

PALAVRAS-CHAVE

Investimentos, Retornos anormais, Mercados ineficientes, Finanças comportamentais, Anomalia de preços

¹Fundação Getúlio Vargas (FGV EAESP),
São Paulo, SP, Brasil

Recebido: 11/11/2019.
Revisado: 11/02/2020.
Aceito: 29/02/2020.
Publicado Online em: 24/08/2020.
DOI: <http://dx.doi.org/10.15728/bbr.2020.17.5.3>



1. INTRODUÇÃO

Muitos estudos mostram que atrair a atenção do investidor tem um importante impacto nas características do mercado financeiro, incluindo liquidez, diversidade de propriedade (Grullon, Kanatas, & Weston, 2004), retornos (Barber & Odean, 2007) e volatilidade (Andrei & Hasler, 2014). Uma citação comumente atribuída a Joseph Kennedy, logo antes da quebra de Wall Street em 1929, vincula a popularidade de ações à previsibilidade dos retornos: “Você sabe que é hora de vender quando engraxates lhe dão dicas sobre ações”. Seu argumento era de que o aumento da popularidade das ações até o ponto de serem negociadas por investidores inexperientes supostamente indicaria uma iminente queda em seus preços. A popularidade de um investimento pode estar relacionada à variável conhecida como atenção do investidor, a qual tem sido estudada por Barber e Odean (2007); Challet e Ayed (2013); Preis, Moat e Stanley (2013); Preis, Reith e Stanley (2010); e, mais recentemente, Bijl, Kringhaug, Molnár e Sandvik (2016) e Mayer (2018).

Entender a relação entre a atenção do investidor e os retornos futuros das ações é fundamental para estratégias de investimento. Se for possível prever o impacto de um aumento ou diminuição da atenção do investidor em retornos futuros, negociações bem-sucedidas podem explorar essa anomalia de mercado. Nossa proposta neste artigo é avaliar se o GSV pode prever retornos de ações no mercado brasileiro. Como em Bijl et al. (2016), nossa hipótese é que um aumento no GSV para uma ação durante certa semana é seguido por retornos mais baixos nas semanas seguintes. Para analisar isso, estimamos um painel longo dos log-retornos semanais anormais ajustados de 57 ações do Ibovespa contra defasagens de variações no GSV de janeiro de 2014 a dezembro de 2018, controlando por diversos fatores de mercado e características das empresas. Estimamos usando FGLS para corrigir problemas de autocorrelação. Também incluímos um termo de interação no modelo para verificar se a atenção do investidor poderia afetar as ações diferentemente, de acordo com o volume de negociação, algo que é novidade nesta literatura.

Em nossa análise principal, encontramos que um aumento de um ponto percentual no GSV em uma semana é seguido por uma diminuição de 0,001 pontos percentuais nos retornos nas semanas seguintes. As próximas três defasagens também são negativas e estatisticamente significantes em conjunto, acumulando uma diminuição de 0,0035 p.p.. Em seguida, avaliamos se os resultados são sensíveis para ações que são negociadas com mais e menos frequência. Achamos que quanto mais frequentemente as ações são negociadas, mais sensíveis estas são às mudanças na atenção do investidor. Enquanto as defasagens do GSV para ações no menor quartil de volume de negociação da amostra têm um efeito cumulativo negativo para retornos futuros de 0,0027 p.p., o efeito cumulativo é de 0,0045 para as ações no terceiro quartil.

Nossos resultados contribuem tanto para o debate existente sobre anomalias de preço, atenção do investidor e o poder preditivo do GSV, quanto fornecem informação aos investidores brasileiros, que podem explorar a anomalia de preços. A principal motivação deste artigo é examinar a atenção do investidor e seu impacto nos retornos das ações no mercado brasileiro; essa informação está atualmente faltando na literatura. Para isso, investigamos a previsibilidade dos retornos das ações no Brasil analisando a série temporal da atenção do investidor disponível para o mercado brasileiro usando dados semanais até 2018. Nós contribuimos para a pesquisa acadêmica em finanças comportamentais ao testar o vínculo entre a atenção do investidor e o retorno de ações individualmente. Tomamos o GSV como uma medida da atenção do investidor uma vez que indivíduos usualmente coletam informações usando a internet. O Google lidera entre todos os mecanismos de pesquisa disponíveis na internet, com aproximadamente 90% de participação de mercado. Especificamente, em relação à ferramenta de GSV, este artigo testa a previsibilidade do retorno de ações em um novo ambiente, seguindo a metodologia inicialmente proposta por Preis et al. (2010), e posteriormente melhorada por Bijl et al. (2016) e Challet e Ayed (2013),

todos os quais estudam os mercados americanos. Dito isso, este artigo é o primeiro a analisar o GSV e retornos de ações para ações brasileiras.

Por fim, sob a perspectiva de um investidor, esta investigação oferece uma visão relevante da eficácia de estratégias baseadas na atenção do investidor, usando GSV. Considerando a peculiaridade do mercado de ações brasileiro comparado aos de países desenvolvidos, este estudo oferece evidência local para investidores brasileiros em relação a outros países com maiores níveis de negociação com investidores individuais. Em 2019, a B3 alcançou um recorde histórico de 1,5 milhão de indivíduos registrados para negociar ações. Isso representa 0,7% da população do país; por outro lado, esses investidores são responsáveis por quase 20% do volume de negociação (B3, 2019). Para ilustrar essa grande diferença, nos Estados Unidos, 52% dos adultos investem em ações (Gallup, 2016); essa é a mínima histórica contra um pico de 65% em 2007, medido pouco antes da crise do *subprime*. Considerando que os mercados brasileiros alcançaram atualmente o menor nível de taxa de juros da história econômica brasileira, espera-se que os mercados acionários recebam mais atenção e investimento de investidores individuais.

2. REVISÃO DE LITERATURA

É importante mencionar que, na comparação de estudos sobre o tema, não apenas a definição de atenção do investidor muda, como também as ferramentas usadas para quantificar a variável atenção em si. A atenção dos investidores pode ser medida por diferentes *proxies*, como despesas com propaganda (Grullon, Kanatas, & Weston, 2004), cobertura dos analistas (Liu, Wu, & Chiang, 2014), preço máximo das últimas 52 semanas (Li & Yu, 2012), e interesse da mídia (Fang & Peress, 2009).

Notadamente, a crescente disponibilidade de *big data* na *internet* contribuiu engajando mais pesquisas no tema e fornecendo, portanto, mais evidência empírica para muitos mercados diferentes. Para ilustrar a relevância da atenção do investidor, pode-se recorrer aos recentes fenômenos do Bitcoin e trabalhar na mesma direção de Kristoufek (2013). Ao longo de 2017, a correlação entre buscas mundiais no Google pela palavra “bitcoin” e o preço dessa criptomoeda cotada em dólar foi aproximadamente 0,95. Analisando a dinâmica recente de evolução do preço do bitcoin, Kristoufek (2013) identificou a dependência do preço da criptomoeda em relação ao sentimento do investidor (“*investor sentiment*”), observando que os picos de preço geralmente ocorrem em sincronia com o interesse geral do investidor no ativo. Além de criptomoedas, Mayer (2018) demonstra que anúncios sem conteúdo relacionado ao valor fundamental da ação exibidos em finais de torneios de futebol universitários altamente assistidas também eram um meio poderoso de mudar os preços das ações. Desse modo, a atenção do investidor emerge como um elemento-chave para entender verdadeiramente os movimentos de preço, *momentum* e a dinâmica de reversão à média em mercados reais.

Novas ferramentas de pesquisa realmente ajudaram a aprofundar nosso entendimento sobre os impactos potenciais da atenção do investidor. A ferramenta *Google Analytics* permite a extração de séries temporais do GSV, em que volumes de busca por palavras ou expressões específicas podem ser estudados ao longo do tempo. Isso abre uma nova janela para estudar a previsibilidade de preço de ações, uma vez que as funções de atividade do Google são, talvez, uma das melhores *proxies* para mensurar a atenção dos investidores até o momento. Essa informação pode ser delimitada em termos de localização geográfica e período de tempo, e pode ser coletada ao longo do tempo para formar uma série temporal. Como resultado, um crescente número de estudos passou a usar essa ferramenta de pesquisa específica (ver, por exemplo, Bijl et al., 2016; Challet & Ayed, 2013; Da, Engelberg, & Gao, 2011; Kristoufek, 2013; Mayer, 2018; Preis et al., 2013; Preis et al., 2010)

para mensurar a atenção do investidor, ou sentimento, pelo GSV. Ainda assim, apesar da ampla gama de possíveis estudos baseados no GSV, tais como a sua relação com a atividade econômica, desemprego, e flutuações gerais dos índices, bem como giro e volume das ações, qualquer relação entre o GSV e o retorno dos preços das ações apresenta uma anomalia de preço significativa que desafia o modelo de mercado eficiente baseado em passeio aleatório de Fama (1970).

Da, Engelberg e Gao (2011) apresentaram pela primeira vez a frequência de buscas no Google como uma medida do interesse do investidor. Lou (2014) e Chemmanur e Yan (2019) exploraram o tema da atenção do investidor examinando o impacto da propaganda nos preços das ações. Mayer (2018) usou campanhas de publicidade das empresas veiculadas em torneios de futebol universitário, medindo sua popularidade por quantidade de telespectadores, permitindo uma análise de sensibilidade dos preços das ações. Tipicamente, uma empresa age como a maior patrocinadora de um torneio universitário, e, por sua vez, ganha exposição na televisão por meio de anúncios e comerciais no jogo. Usando índices de audiência para os torneios como *proxies* de atenção e analisando o impacto de diferentes níveis de atenção nos preços das empresas patrocinadoras, o autor conclui que a atenção do investidor, especificamente a atenção do *noise trader*, causa retornos anormais em ações de 1,51% em 1-5 dias e, posteriormente, uma reversão quase completa de 1,09% na semana seguinte (Mayer, 2018).

Barber e Odean (2007) discutem que a atenção dos investidores individuais é um recurso escasso e, dadas as milhares de opções disponíveis aos investidores, existe um claro problema de busca. Tal problema é agravado pelo fato de que investidores individuais tendem a ser essencialmente compradores, usualmente mantendo posições compradas e apenas vendendo ações que já possuem (e vendendo a descoberto com muito menos frequência que investidores institucionais). Os resultados dos autores coincidem com essas conjecturas, pois identificam que investidores individuais compram orientados pela atenção, medida por diferentes eventos que atraem a atenção, tais como exposição a notícias, negociação anormal, e retornos anormais nos volumes de negociação de curto prazo das ações.

Inerentemente, os estudos de GSV são relativamente novos para as finanças devido à natureza da ferramenta usada, a qual só se tornou disponível e confiável após 2008. Embora outros estudos relacionados tenham sido realizados ao redor do mundo, por exemplo, vinculando o volume de negociação ao volume de buscas, três estudos-chave olharam especificamente para o poder de previsão de preço do GSV. Primeiramente, Preis et al. (2010) tentaram definir a relação entre preço e GSV. Embora tenha sido constatado que o GSV não tinha poder preditivo estatisticamente significativo para retornos de ações, ele parecia ter forte correlação com o volume de negociação. Posteriormente, Preis et al. (2013) testaram se uma estratégia de negociação baseada em comprar e vender ações de acordo com certas consultas de pesquisa ao Google conseguiria de fato superar o mercado. Os resultados indicaram que as consultas do Google foram realmente capazes de prever o retorno das ações, superando o índice de mercado em 310% nos sete anos testados. Challet e Ayed (2013) trabalharam na mesma direção de Preis et al. (2013), embora com menos objetivos específicos, procurando descobrir se a informação do GSV é robusta o suficiente para prever movimentos de mercado. Entretanto, uma das contribuições mais relevantes do estudo é sua análise metodológica e a examinação de potenciais vieses, especificamente aqueles relacionados a *backtesting*.

O primeiro tipo de viés identificado é o viés de ferramenta, no qual é fácil encontrar previsibilidade em dados antigos sobre mercados financeiros utilizando ferramentas modernas (com forte poder computacional, especialmente analisando grandes conjuntos de dados). Por essa razão, não é metodologicamente recomendado fazer *backtest* com dados antigos utilizando a capacidade de ferramentas modernas. O segundo tipo de viés, chamado viés de dados, refere-se à real

disponibilidade de dados e sua significância relativa em diferentes momentos. Por exemplo, Challet e Ayed (2013) apontam que os dados de GSV eram atualizados apenas esporadicamente antes de agosto de 2008, e isso significa que qualquer estudo antes desse limite lida automaticamente com diferentes qualidades de dados no mesmo conjunto de dados. Em outra nota, os autores afirmam a existência de um problema de revisão de dados inerente, que, embora coloque todos os dados disponíveis em uma escala comparável, pode atrapalhar uma série temporal. Isso impacta o GSV significativamente, dado que em 2012 o método de normalização usado pelo Google mudou para o que é atualmente um sistema de índice de 100 pontos.

Finalmente, o *Google Trends* – e sua natureza ainda experimental – sinaliza a necessidade de futuras medidas para assegurar que vieses não impactem significativamente o experimento. A escolha da consulta por palavra-chave surge como provavelmente a decisão mais significativa do ponto de vista experimental e traz uma série de problemas. Ao testar o GSV em um contexto financeiro, o ideal seria distinguir (e isolar) a ação ou empresa das perspectivas do investidor e do consumidor. Procurar pelo *ticker* da ação na bolsa pode ser uma solução razoável, embora remova consultas significativas de investidores que estavam usando o nome real da empresa, por exemplo, para acessar seu site de relações com investidores. Ainda assim, há um problema com mudanças nos *tickers*, as quais devem ser consideradas em casos de fusões, alterações de segmentos de listagem, e uma multidão de outras razões. Em resumo, Challet e Ayed (2013) contribuem ao adicionar uma análise robusta de vieses à literatura enquanto também defendem a tese de que existem conclusões aleatórias significativas (i.e. falso-positivas) a serem tiradas do GSV.

Por fim, Bijl et al. (2016) baseiam-se nos artigos citados anteriormente para testar a previsibilidade de preço do GSV usando uma metodologia ligeiramente diferente. A previsibilidade do retorno das ações é testada semanalmente para empresas individuais usando o GSV para o mercado de ações americano (empresas listadas no índice S&P500) de 2007 a 2013; eles concluem que um alto GSV prevê baixos retornos futuros subsequentes. A conclusão é também testada em um contexto prático, usando uma estratégia *long-short* que confirma o efeito não ser forte o suficiente para ser lucrativo após contabilizar os custos de transação. Metodologicamente, os autores admitem a falta de confiabilidade dos dados do *Google Trends* anteriores a 2008, mas mesmo assim usam dados de 2007 para estimar betas em janelas móveis de 52 semanas usados no modelo CAPM para testar retornos anormais. Além disso, eles optam por restringir os dados regionalmente (somente EUA) para reduzir o risco de buscas não relacionadas a investidores, o que intuitivamente deve ser mais comum fora dos EUA. Isso é considerado menos problemático no Brasil, dada a relativamente alta concentração de investidores internacionais na B3, representando aproximadamente 27% do volume negociado em 2018 a partir de abril (B3, 2018). No entanto, Bijl et al. (2016) destacam que as consultas mundiais do GSV têm maior probabilidade de serem “falso-positivas” uma vez que as chances de palavras sobrepostas com diferentes significados aumentam consideravelmente. Na escolha da palavra-chave, os autores usam os nomes das empresas em vez dos *tickers*, dado que muitos *tickers* são também abreviações de empresas, que ainda capturam pesquisas indesejadas de clientes de varejo e de outros não investidores. Para o processamento de dados, os autores adicionaram várias outras variáveis à análise de regressão por razões de robustez, incluindo o efeito de janeiro, volatilidade das ações e o logaritmo do volume sem tendência de negociação (Bijl et al., 2016).

3. DADOS E METODOLOGIA

A análise cobre o período entre janeiro de 2014 e dezembro de 2018. Usamos as 57 ações (54 empresas) que compunham o Índice Ibovespa a partir de janeiro de 2018 e para as quais havia os dados necessários no período. A decisão de incluir apenas empresas pertencentes ao Ibovespa foi

feita para garantir que todas as ações tivessem um volume de negociação diária relativamente alto e assegurar que os preços se ajustassem dinamicamente devido às informações de mercado. Em outras palavras, não seria desejável incluir ações de empresas que não negociam com frequência e, portanto, não reagem às notícias de mercado, GSV, e outras variáveis de mercado, por dias seguidos. Além disso, considerando que o Ibovespa, assim como o S&P 500, concentra a parte mais relevante das ações da B3 (Castro et al., 2019), essas são também as ações com maior probabilidade de ter dados significativos de GSV disponíveis, como sugerido por Bijl et al. (2016).

Os cinco anos de negociação de 2014-2018 refletem diferentes períodos de política macroeconômica brasileira e desempenho econômico. Em 2014, o crescimento do Produto Interno Bruto (PIB) observado estava alinhado com os pares de mercados emergentes do Brasil; nos últimos dois anos, o PIB encolheu aproximadamente 3% com uma ligeira recuperação refletida em um leve crescimento de 1% em 2017 (Bacen, 2018). Esse período também abrange uma ampla gama de políticas fiscais e monetárias. Isso, entretanto, não se apresenta como uma limitação significativa do estudo, pois ele não é focado em determinado setor. Dado que focamos em empresas do Ibovespa, que são as ações mais líquidas e relevantes representando o movimento de mercado, deve ser aproximadamente neutro aos ciclos macroeconômicos. Curiosamente, conforme demonstrado pelos estudos de *momentum* de Asness, Moskowitz e Pedersen (2013) e de Daniel e Moskowitz (2016), testes de anomalia são empiricamente sensíveis a crises, tornando o intervalo de tempo selecionado conveniente para esse reexame.

Para examinar o valor preditivo do GSV nos retornos de ações, estimamos um modelo de regressão em painel em que cada log-retorno em excesso de mercado do ativo i (r_i^e) é explicado por suas cinco primeiras defasagens, além da primeira à quinta defasagem da variação logarítmica em seu GSV (gsv_i), mais a primeira à quinta defasagem de volumes de negociação (vol_i), assim como os retornos e volatilidade do GSV no médio prazo (σ_i^{mid} e gsv_i^{mid}) e longo prazo (σ_i^{long} e gsv_i^{long}), além dos tamanhos das empresas e efeitos do setor (10 classificações de setor GICS), como descrito na Equação (1):

$$r_i^e = \beta_0 + \sum_{\ell=1}^5 \rho_{\ell} r_{i,t-\ell}^e + \sum_{\ell=1}^5 \theta_{\ell} gsv_{i,t-\ell} + \sum_{\ell=1}^5 \lambda_{\ell} vol_{i,t-\ell} + \beta_1 \sigma_{it}^{mid} + \beta_2 \sigma_{it}^{long} + \beta_3 \sigma_{it}^{long} + \beta_4 \sigma_{it}^{long} + \beta_5 size_{it} + \sum_{j=1}^9 \alpha_j Industry_j + \varepsilon \quad (1)$$

A motivação para a adição de cinco defasagens de retorno anormal, de GSV, e de volume negociado é dupla. Primeiramente, isso permite que o modelo capture um conjunto maior de dados, diminuindo potencialmente o impacto de qualquer imprecisão de dados, como se o GSV de um determinado período na verdade representasse volumes de períodos anteriores. Em segundo lugar, as defasagens também permitem que o mercado elimine qualquer anomalia de preço, como efeitos de *momentum*.

O painel é formado por 57 ações ao longo de 261 semanas, criando uma regressão em painel longo; isso requer adicionar características assintóticas de séries temporais à estimativa, em contraste com os painéis curtos tradicionais, quando se presume que o número de períodos de tempo seja fixo e o número de indivíduos cresce indefinidamente. A correlação ao longo do tempo para uma dada ação pode ser incorporada ao modelo usando um modelo autorregressivo de médias móveis (ARMA) para os erros, permitindo os parâmetros diferenciarem-se entre as ações (Cameron & Trivedi, 2005). Se os erros forem heteroscedásticos e o parâmetro de autocorrelação também diferir entre as ações, podemos estimar regressões de séries temporais separadas para cada ação usando todos os períodos para produzir parâmetros ARMA consistentes. Entretanto, como explicado em Cameron e Trivedi (2005), ao utilizar Método dos Mínimos Quadrados

Generalizado Factível (FGLS de *Feasible Generalized Least Squares* em inglês), nós podemos obter estimativas consistentes para esses parâmetros utilizando todas as observações NT permitindo a heteroscedasticidade, correlação específica de ações ao longo do tempo e entre ações.

Todas as variáveis são medidas semanalmente. Retornos anormais ajustados ao mercado são mensurados como na Equação (2), em que r_{it} são os retornos logarítmicos semanais para cada ativo, r_{mt} é o retorno logarítmico de mercado, e β é o beta móvel de um ano de cada ativo contra o Ibovespa:

$$r_{it}^e = r_{it} + \beta r_{mt} \quad (2)$$

As séries de volumes de negociação tiveram sua tendência removida e foram transformadas em uma função logarítmica para atenuar picos e vales em todo o conjunto de dados, uma vez que o volume pode ser altamente volátil. Essa metodologia foi baseada em Campbell, Grossman e Wang (1993) e Bijl et al. (2016):

$$vol_{it} = \log(\text{volume})_{it} - \frac{1}{12} \sum_{s=t-11}^t \log(\text{volume})_s \quad (3)$$

A volatilidade dos retornos foi calculada diariamente usando janelas móveis de cinco e 20 dias de negociação para volatilidade a médio e longo prazo, referentes aos períodos de uma semana e um mês de negociação, respectivamente. Posteriormente, selecionamos as medidas em intervalos semanais para parear com as outras variáveis. Para a volatilidade do GSV, uma vez que possuímos apenas dados semanais originais, a volatilidade de médio e longo prazo foi calculada usando janelas móveis de quatro e 12 semanas, referentes aos períodos de um e três meses, respectivamente. Temos:

$$\sigma r_{is}^{mid} = \sqrt{\sum_{s=1}^5 (r - \bar{r})_{is}^2} \quad e \quad \sigma r_{is}^{long} = \sqrt{\sum_{s=1}^{20} (r - \bar{r})_{is}^2} \quad (4)$$

$$\sigma gsv_{it}^{mid} = \sqrt{\sum_{t=1}^5 (gsv - \overline{gsv})_{it}^2} \quad e \quad \sigma gsv_{it}^{long} = \sqrt{\sum_{t=1}^{12} (gsv - \overline{gsv})_{it}^2} \quad (5)$$

Finalmente, o tamanho é medido como o logaritmo do valor de mercado total para cada ação a cada semana, e os efeitos de setor são capturados por 9 *dummies* para cada uma das classificações GICS menos uma.

Todas as variáveis, exceto o GSV, foram extraídas da plataforma Economatica e cotadas em reais (BRL) quando aplicável. A variável GSV foi obtida diretamente da ferramenta *Google Trends* via pacote “gtrendsR” da linguagem de programação R, o qual inclui diversas opções de pesquisa que devem ser cuidadosamente consideradas e parametrizadas. Primeiramente, é importante entender como o Google apresenta seus dados de GSV, como destacado por Challet e Ayed (2013). O GSV não apresenta um número absoluto de pesquisas por palavra-chave, mas um valor padronizado que varia de 0 a 100. Nessa metodologia, 100 representa o maior valor da série, que depende tanto da série temporal escolhida, quanto das outras palavras-chave selecionadas. O Google permite até cinco pesquisas simultâneas de palavras-chave, mas os dados de GSV padronizam a série de cada palavra-chave com base nas outras quatro seleções. Isso pode ser um problema se pesquisarmos simultaneamente as palavras-chave “futebol” e “CMIG4” no Brasil, por exemplo.

Uma vez que o volume de buscas nominais por “futebol” é muito maior que por “CMIG4”, este último parece ser nulo comparado ao primeiro. Portanto, palavras-chave com um maior volume de buscas anulam qualquer variação em palavras-chave com um menor volume quando ambas são pesquisadas simultaneamente para capturar o GSV; pesquisar mais de uma palavra-chave pode distorcer severamente dados históricos.

Outra opção relevante no *Google Trends* é o escopo geográfico dos dados de GSV. Dada a relevante participação de investidores estrangeiros no mercado de ações brasileiro, poderíamos ter levantado dados das consultas mundiais. Entretanto, considerando Bijl et al. (2016) e os dados preliminares poluídos coletados na opção mundial, escolhemos restringir os dados de GSV para consultas brasileiras para limitar o risco de sobrepor o significado de palavras ao redor do mundo (Bijl et al., 2016). Finalmente, os dados de GSV são publicados aos domingos e, portanto, são apenas disponíveis semanalmente; assim, a segunda-feira é o primeiro dia de negociação para os mercados reagirem quando usarem o GSV. Essa é a razão principal para usarmos dados semanais para todas as variáveis.

Em seguida, talvez a decisão mais crucial quanto à coleta de dados no *Google Trends* está nas palavras-chave que escolhemos. Enquanto Bijl et al. (2016) escolheram pesquisar pelos nomes das empresas, argumentando que os *tickers* são muito similares a suas abreviações, nós buscamos por *tickers* para evitar ruído nos dados causados por não investidores simplesmente procurando por produtos os serviços da empresa. Isso é fácil de imaginar para empresas de varejo, como “Lojas Renner”, “Lojas Americanas”, e “Banco Itaú”, os quais possuem milhões de clientes no Brasil. Dado que investidores individuais representam apenas uma pequena fração da população brasileira, pesquisar os nomes da empresa pode estar enviesado para refletir a atenção do cliente em vez da atenção do investidor.

Portanto, buscamos pelos *tickers* de cada ação separadamente para todo o período e restringimos os resultados ao Brasil. Uma vez que os dados são padronizados para cada pesquisa, em que o menor valor é zero, e o maior é 100, usar o volume de buscas (tanto bruto quanto na forma logarítmica) poderia distorcer os dados, pois 100 para um *ticker* não tem o mesmo significado de 100 para outro. Isso significa que não podemos comparar volumes de pesquisas de diferentes expressões (*tickers*, neste caso), e o número absoluto de consultas não está disponível. Essa é a principal razão pela qual adotamos uma medida de GSV em log para capturar o aumento e diminuição das buscas de uma semana até outra. Logo, focamos na variação nos volumes de buscas de uma semana para outra. Além disso, volume de pesquisas tendem a crescer ao longo do tempo; usar essa variação evita resultados de regressão espúrios devido à natureza não estacionária da série de GSV.

4. RESULTADOS

A Figura 1 mostra a evolução comparativa dos preços de mercado e o GSV para um portfólio ponderado por valor composto pelas ações da amostra. Na Figura 1, a tendência ascendente para as duas amostras é clara, embora os preços atinjam um mínimo no começo de 2016, e os GSVs sejam muito pequenos durante 2014.

A Figura 2 mostra uma comparação entre os retornos de mercado do portfólio ponderado por valor e a variação logarítmica do GSV. Nenhuma das séries parece apresentar tendências, e o GSV é muito mais volátil que os retornos. Finalmente, a Figura 3 compara o volume de negociação com o GSV. Apenas recentemente as duas séries parecem seguir tendências similares, mostrando um leve aumento e uma maior variação nas últimas semanas de 2018; isso ocorre especialmente para a série do GSV.

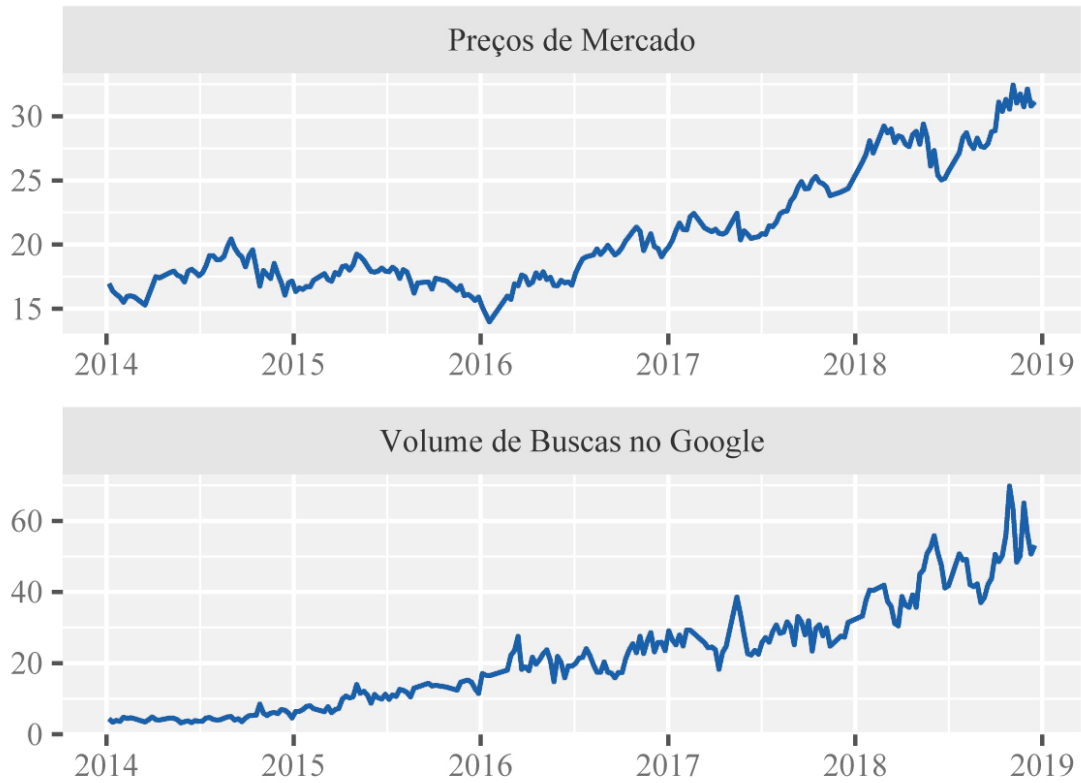


Figura 1. Preços de Mercado Versus GSV
Fonte: Dados de pesquisa (2019).

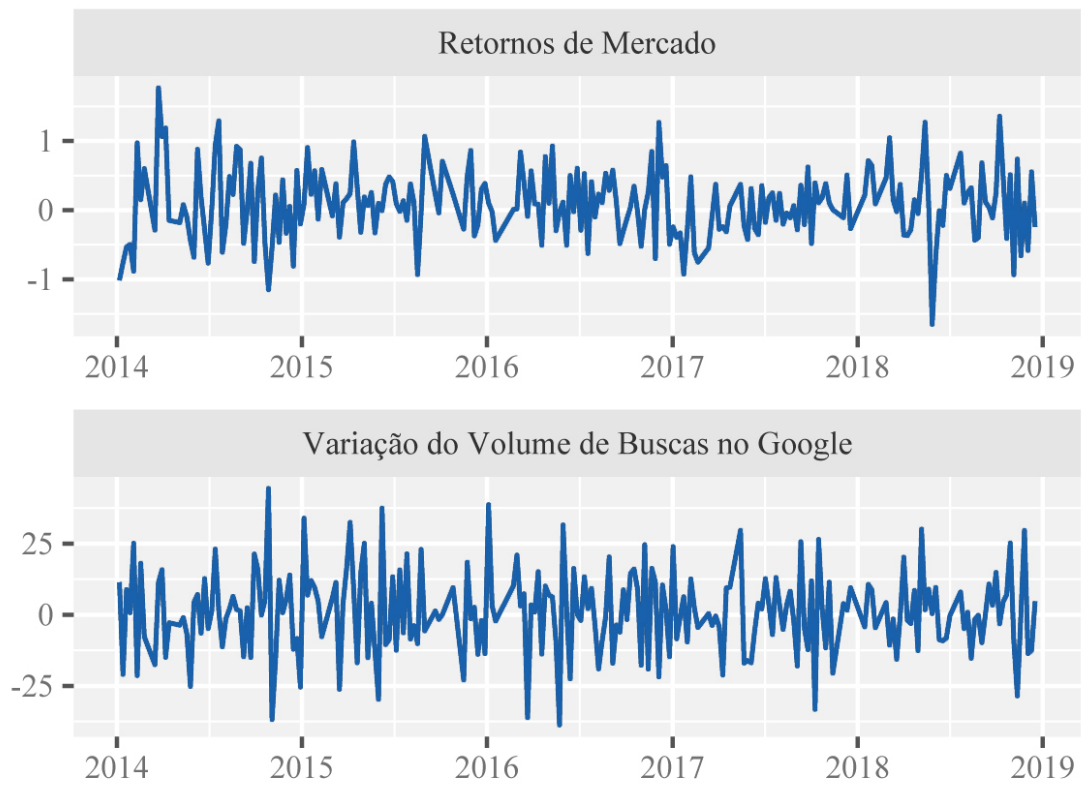


Figura 2. Retornos de Mercado Versus Variação do GSV
Fonte: Dados de pesquisa (2019).

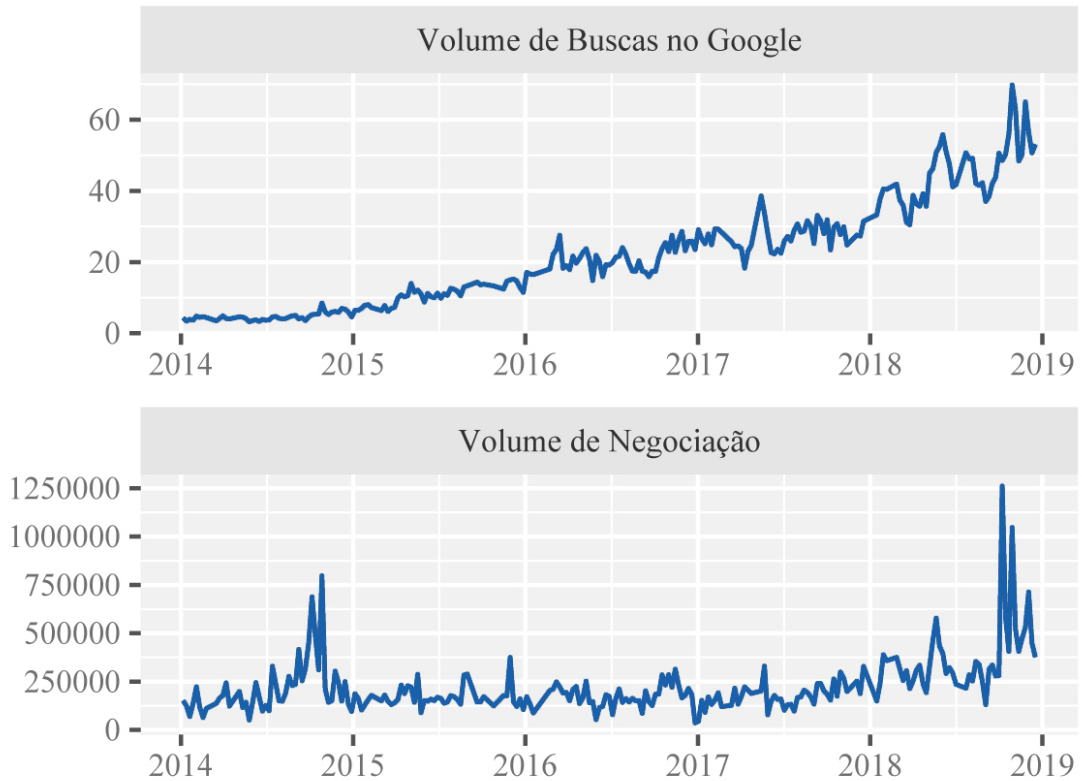


Figura 3. Volume de Mercado Versus GSV

Fonte: Dados de pesquisa (2019).

A Tabela 1 mostra as estatísticas descritivas para as variáveis da regressão. Embora muito próximos a zero, os retornos anormais de mercado semanais são negativos, o mínimo é próximo a -50% (JBSS3 em maio de 2017) e o máximo é próximo a 50% (MGLU3 em dezembro de 2015). A redução (aumento) mais dramática no GSV, mais de 400%, é da EQTL3 em setembro de 2016 (CPFE3 em março de 2018). Em geral, o GSV é muito mais volátil que o retorno, como também se pode ver na Figura 2, e as empresas são geralmente muito grandes.

Tabela 1
Estatísticas Gerais

Variável	Estatística	Valor
Retornos Anormais de Mercado (%)	Mín.	-48,834
	Média	-0,032
	Máx.	46,015
Variação do GSV (%)	Mín.	-412,713
	Média	0,926
	Máx.	411,087
Volume de Negociação	Mín.	-2,106
	Média	0,028
	Máx.	2,981
Volatilidade no médio prazo dos retornos	Mín.	0,005
	Média	3,956
	Máx.	39,537

Tabela 1
Cont.

Variável	Estatística	Valor
Volatilidade no longo prazo dos retornos	Mín.	0,296
	Média	9,299
	Máx.	47,868
Volatilidade no médio prazo do GSV	Mín.	0
	Média	109,770
	Máx.	689,500
Volatilidade no longo prazo do GSV	Mín.	0
	Média	216,600
	Máx.	969,600
Tamanho	Mín.	12,060
	Média	16,760
	Máx.	19,770

Fonte: Dados de pesquisa (2019).

A Figura 4 mostra a distribuição da amostra entre os 10 setores. Materiais básicos têm o maior número de observações, seguido por utilidade pública. Em seguida, vem o setor financeiro, seguido por bens de consumo e bens de consumo básico. Tecnologia da informação e telecomunicações têm os menores números de observações.

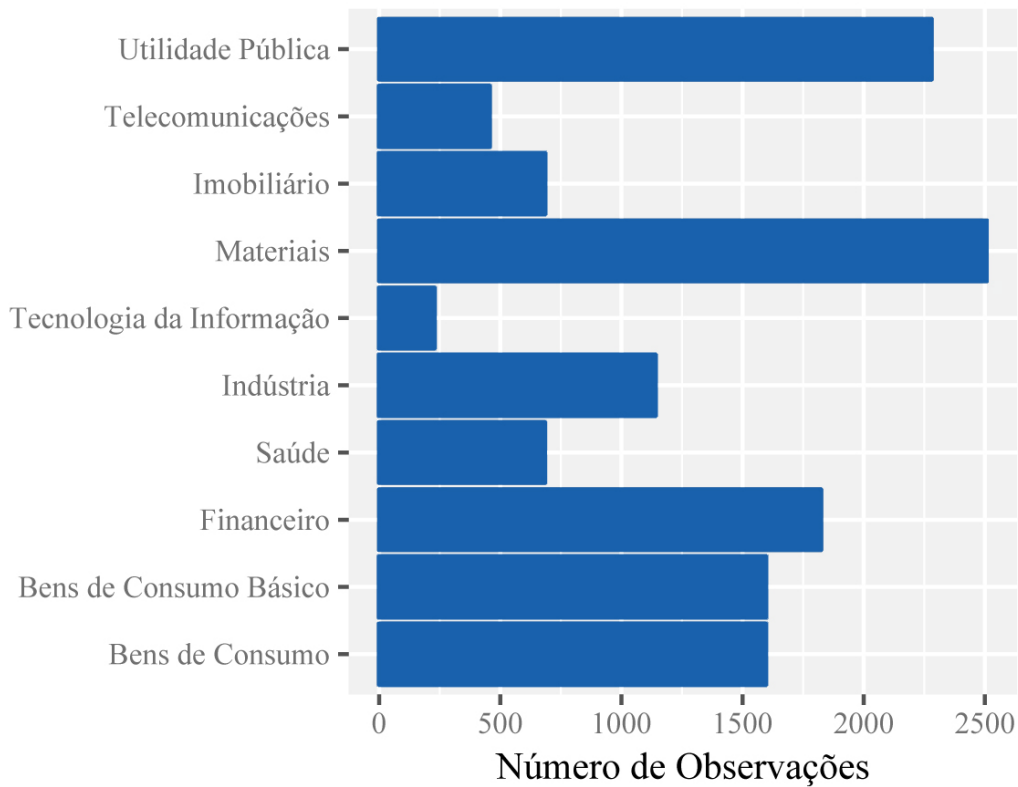


Figura 4. Distribuição da Amostra por Setor
Fonte: Dados de pesquisa (2019).

Finalmente, a Tabela 2 mostra a matriz de correlação das variáveis, incluindo preços das ações e o nível do GSV. A correlação contemporânea entre retornos anormais de mercado e a variação do GSV é próxima a zero, mas os preços e o nível de GSV têm uma correlação de pouco mais de 10% na amostra. A variação logarítmica do GSV é correlacionada principalmente com seu nível, volume de negociação e volatilidade. Empresas maiores têm maiores níveis de GSV, mas sua variação é menor. Empresas maiores também têm menor volatilidade dos retornos.

Tabela 2

Matriz de Correlação

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) Preço	1	0.042	0.119	-0.004	0.034	-0.168	-0.224	-0.015	-0.022	0.225
(2) Retorno Anormal		1	0.012	-0.002	0.092	0.008	0.011	0.007	0.004	0.027
(3) GSV			1	0.168	0.123	0.094	0.102	-0.297	-0.362	0.144
(4) Var. GSV				1	0.045	-0.034	-0.031	0.024	0.012	-0.002
(5) Volume					1	0.194	0.077	-0.006	-0.015	0.021
(6) Vol. MP Ret.						1	0.671	-0.088	-0.111	-0.111
(7) Vol. LP Ret.							1	-0.125	-0.164	-0.156
(8) Vol. MP GSV								1	0.816	-0.238
(9) Vol. LP GSV									1	-0.290
(10) Tamanho										1

Fonte: Dados de pesquisa (2019).

A Tabela 3 mostra os resultados da estimação do FGLS para a Equação (1). O primeiro e o quinto componentes autorregressivos dos retornos são estatisticamente significantes (o teste F mostra que as cinco defasagens são, em conjunto, estatisticamente significantes a 1%). As três primeiras defasagens da variação logarítmica do GSV são negativas e estatisticamente significantes. Todas as defasagens, exceto a quinta, são negativas. Conjuntamente, os primeiros quatro coeficientes negativos também são significantes em 1% de acordo com o teste F. Embora estatisticamente significativa, o efeito da atenção do investidor mensurada pelo GSV é muito pequena. Considerando apenas a primeira defasagem, um aumento de um ponto percentual no GSV em uma semana é seguido por uma redução de 0,001 ponto percentual nos retornos na semana seguinte. Considerando as quatro defasagens negativas, o efeito soma 0,0035 p.p.. Em relação às variáveis de controle, o tamanho da empresa também é um forte componente na previsão de retornos (empresas maiores da amostra têm retornos mais altos). As volatilidades dos retornos e do GSV não são estatisticamente significantes, e o volume de negociação é significativo apenas para a primeira defasagem ao nível de 10%.

Esses resultados estão alinhados com Bijl et al. (2016), que também encontraram efeitos negativos do GSV em retornos anormais futuros. Portanto, o efeito da atenção do investidor no mercado brasileiro é similar ao dos Estados Unidos, apesar da vasta diferença de desenvolvimento entre os dois mercados. Primeiramente, como já relatado, investidores de varejo são muito mais escassos no Brasil, onde taxas de juros historicamente altas e prêmios negativos de risco de mercado levaram investidores individuais a produtos simples de renda fixa, como títulos do governo. Como resultado, é mais difícil para indivíduos causarem pressão significativa nos preços das ações, que tendem a acompanhar a demanda institucional dos investidores (nacional e internacionalmente), de acordo com as leis de oferta e demanda. Isso sugere um poder preditivo reduzido da atenção de investidores individuais, conforme medido por consultas na *internet*, pois seus volumes são

Tabela 3.
Resultados da Estimativa

	Variável Dependente: r_i^e
Intercept	-1,7349*** (0,5142)
r_{t-1}^e	-0,0272*** (0,0088)
r_{t-2}^e	-0,0112 (0,0088)
r_{t-3}^e	-0,0090 (0,0087)
r_{t-4}^e	0,0100 (0,0087)
r_{t-5}^e	0,0249*** (0,0087)
gsv_{t-1}	-0,0009*** (0,0003)
gsv_{t-2}	-0,0013*** (0,0004)
gsv_{t-3}	-0,0009*** (0,0004)
gsv_{t-4}	-0,0006 (0,0004)
gsv_{t-5}	0,0001 (0,0003)
vol_{t-1}	0,0656 (0,0416)
vol_{t-2}	0,0731* (0,0423)
vol_{t-3}	0,0329 (0,0424)
vol_{t-4}	-0,0090 (0,0421)
vol_{t-5}	-0,0262 (0,0412)
σ_{it}^{long}	0,0112 (0,0082)
σ_{it}^{mid}	0,0043 (0,0131)
σgsv_{it}^{long}	0,0002 (0,0002)
σgsv_{it}^{mid}	0,0002 (0,0003)
Tamanho	0,0990*** (0,0283)
Efeitos fixos de setor	Sim
# ações; # semanas; #observações	57, 228, 12,991
Estatística de Wald	66,4408***

Nota: *p<0,1; **p<0,05; ***p<0,01

Fonte: Dados de pesquisa (2019).

direcionados principalmente por investidores individuais de baixo volume e não investidores institucionais que individualmente conduzem grandes volumes de negociação. No entanto, o *noise trading* ainda parece ser capaz de desempenhar um papel na explicação dos retornos de mercado.

Como um teste de robustez, decidimos expandir a análise do impacto do GSV nas ações brasileiras. Nossa hipótese era de que a atenção do investidor poderia impactar diferentemente os retornos das ações. Para esse propósito, avaliamos se empresas mais frequentemente negociadas eram mais ou menos sensíveis a variações na atenção do investidor. Para fazer isso, expandimos a Equação (1) para incluir interações do GSV com o volume de negociação. Esse termo de interação é uma novidade na literatura do GSV e atenção do investidor. Se as interações são conjuntamente negativas (positivas) e estatisticamente significantes, o efeito negativo do GSV é reforçado (minimizado) quanto mais negociada for uma ação. A Tabela 4 apresenta os resultados das estimações FGLS para esse modelo.

Tabela 4

Resultados da Estimativa: Interações com Volume de Negociação

	Variável Dependente: r_i^e
Intercept	-1,7530*** (0,5134)
r_{t-1}^e	-0,0274*** (0,0088)
r_{t-2}^e	-0,0118 (0,0088)
r_{t-3}^e	-0,0098 (0,0087)
r_{t-4}^e	0,0106 (0,0087)
r_{t-5}^e	0,0241*** (0,0087)
gsv_{t-1}	-0,0009*** (0,0003)
gsv_{t-2}	-0,0013*** (0,0004)
gsv_{t-3}	-0,0008*** (0,0004)
gsv_{t-4}	-0,0006 (0,0004)
gsv_{t-5}	0,0002 (0,0003)
vol_{t-1}	0,0653 (0,0418)
vol_{t-2}	0,0703* (0,0425)
vol_{t-3}	0,0349 (0,0425)
vol_{t-4}	-0,0129 (0,0423)
vol_{t-5}	-0,0201 (0,0414)

Tabela 4
Cont.

	Variável Dependente: r_i^e
σr_{it}^{long}	0,0122 (0,0083)
σr_{it}^{mid}	0,0033 (0,0132)
σgsv_{it}^{long}	0,0002 (0,0002)
σgsv_{it}^{mid}	0,0001 (0,0003)
Size	0,1001*** (0,0283)
$gsv_{t-1} * vol_{t-1}$	0,0000 (0,0005)
$gsv_{t-1} * vol_{t-2}$	-0,0010* (0,0005)
$gsv_{t-1} * vol_{t-3}$	-0,0004 (0,0005)
$gsv_{t-1} * vol_{t-4}$	0,0001 (0,0005)
$gsv_{t-1} * vol_{t-5}$	-0,0016*** (0,0005)
Efeitos fixos do setor	Sim
# ações; # semanas; #observações	57, 228, 12,991
Estatística de Wald	79,654***

Nota: *p<0,1; **p<0,05; ***p<0,01

Fonte: Dados de pesquisa (2019).

Os resultados básicos da Tabela 3 ainda são válidos. As defasagens do GSV são negativas e, em conjunto, estatisticamente significantes em 1%. Entretanto, enquanto o volume de negociação não é significativo sozinho (o teste F não mostra significância conjunta para as cinco defasagens), as interações com o GSV são, embora muito pequenas, negativas, indicando que as ações mais negociadas tendem a ser mais vulneráveis a mudanças na atenção do investidor, medida por consultas na *internet*. Embora não tenhamos considerado o volume na análise, um aumento de um p.p. no GSV acumula uma redução de 0,0035 p.p. nos retornos, o efeito é de apenas 0,0027 para ações no primeiro quartil de volume de negociação e de 0,0045 para ações no terceiro quartil. Portanto, quanto mais negociadas forem as empresas, mais sensíveis são a mudanças na atenção do investidor. Isso é consistente com a ideia de que a atenção do investidor é mais concentrada em empresas mais negociadas, implicando que investidores dedicam uma grande parte de seu tempo pesquisando ações que já estão sob destaque no mercado.

5. OBSERVAÇÕES FINAIS

O objetivo deste artigo era analisar o papel da atenção do investidor para prever futuros retornos de mercado. Seguindo pesquisas anteriores (ver, e.g., Bijl et al., 2016; Challet & Ayed, 2013; Kristoufek, 2013; Mayer, 2018; Preis et al., 2013; Preis et al., 2010), usamos consultas na *internet* no GSV como *proxy* para a atenção do investidor individual. Estimamos uma regressão

dos retornos de 57 ações contra a variação do GSV da primeira à quinta semana anterior, controlando para características de mercado e de firmas.

Consistente com a pesquisa anterior no mercado dos Estados Unidos (Bijl et al., 2016), descobrimos que uma maior atenção dos investidores medida por aumentos no GSV é seguida por retornos mais baixos, e isso é reforçado por volumes de negociação mais altos. Isso é consistente com as ideias iniciais de J. Kennedy nos anos 1920: que a popularidade pode indicar que um ativo perdeu aderência com seu valor fundamental, muito antes da avaliação científica do impacto de *noise traders* (e.g., De Long, Shleifer, & Summers, 1990).

Esta pesquisa colabora com o debate sobre anomalias nos preços de mercado e contribui especificamente para a avaliação de um mercado emergente, uma vez que a pesquisa convencional se concentra nos mercados desenvolvidos, principalmente nos Estados Unidos. Além disso, nossos resultados oferecem *insights* sobre possíveis estratégias de investimento baseadas em seguir as variações do GSV. Futuras pesquisas poderiam focar nessa questão, testando a rentabilidade de estratégias de negociação projetadas para se beneficiar de dadas informações sobre o GSV, por meio de um backtesting de um portfólio *long-short*, por exemplo.

Finalmente, uma linha comum tecida ao longo deste artigo é que a tecnologia deve permanecer relevante no campo da atenção e do sentimento do investidor; isso impactará a qualidade e disponibilidade de ferramentas usadas para medir esses conceitos complexos e ainda abstratos. Nesse sentido, novos estudos podem muito bem estar mais bem posicionados para confirmar, expandir ou desafiar as conclusões aqui apresentadas e contribuir para um ciclo virtuoso em que a tecnologia aprofunda nossa compreensão de complexos fenômenos comportamentais.

REFERÊNCIAS

- Andrei, D., & Hasler, M. (2014). Investor attention and stock market volatility. *The Review of Financial Studies*, 28(1), 33-72.
- Asness, C. S., Moskowitz, T. J., & Pedersen, L. H. (2013). Value and momentum everywhere. *The Journal of Finance*, 68(3), 929-985.
- B3. (2019). Participação dos investidores. <http://www.bmfbovespa.com.br/pt-br/servicos/market-data/consultas/mercado-a-vista/participacao-dos-investidores/volume-total/>
- Bacen. (2018). Sistema de expectativas do mercado. <https://www3.bcb.gov.br/expectativas/publico/consulta/serieestatisticas>
- Barber, B. M., & Odean, T. (2007). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The 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.
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.
- Campbell, J. Y., Grossman, S. J., & Wang, J. (1993). Trading volume and serial correlation in stock returns. *The Quarterly Journal of Economics*, 108(4), 905-939.
- Castro, F. H., Eid Jr, W., Santana, V. F., & Yoshinaga, C. E. (2019). Fifty-year History of the Ibovespa. *Revista Brasileira de Finanças*, 17(3), 47-65.
- Challet, D., & Ayed, A. B. H. (2013). Predicting financial markets with Google Trends and not so random keywords. *arXiv preprint*, 1307.4643. <https://arxiv.org/pdf/1307.4643.pdf>

- Chemmanur, T. J., & Yan, A. (2019). Advertising, attention, and stock returns. *Quarterly Journal of Finance*, 9(3), 1-51.
- Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5), 1461-1499.
- Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2), 221-247.
- De Long, B., Shleifer, A., & Summers, L. (1990). Noise trader risk in financial markets. *The Journal of Political Economy*, 98(4), 703-738.
- Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- Fang, L., & Peress, J. (2009). Media coverage and the cross section of stock returns. *The Journal of Finance*, 64(5), 2023-2052.
- Gallup (2016). Just Over Half of American Own Stocks, Matching a Record Low. <https://news.gallup.com/poll/190883/half-americans-own-stocks-matching-record-low.aspx>
- Grullon, G., Kanatas, G., & Weston, J. P. (2004). Advertising, breadth of ownership, and liquidity. *The Review of Financial Studies*, 17(2), 439-461.
- Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific reports*, 3, 3415.
- Li, J., & Yu, J. (2012). Investor attention, psychological anchors, and stock return predictability. *Journal of Financial Economics*, 104(2), 401-419.
- Lou, D. (2014). Attracting investor attention through advertising. *The Review of Financial Studies*, 27(6), 1797-1829.
- Mayer, E. J. (2018). Advertising, investor attention, and stock prices: Evidence from a natural experiment. <http://dx.doi.org/10.2139/ssrn.2468181>
- Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, 3, 1684.
- 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: Mathematical, Physical and Engineering Sciences*, 368(1933), 5707-5719.

CONTRIBUIÇÕES DE AUTORIA

Os autores contribuíram igualmente para este manuscrito.

AGRADECIMENTOS

Os autores gostariam de agradecer à Verônica Santana pela excelente assistência em pesquisa e à Nathália Gil pelo auxílio com a tradução do artigo para o português. Agradecemos ao Editor (Felipe Ramos), ao Editor Associado (Luiz Felipe Girão) e aos dois revisores anônimos pelas valiosas contribuições e sugestões. Os autores, entretanto, se responsabilizam pelos eventuais erros remanescentes. Claudia Yoshinaga agradece o suporte financeiro do FGV EAESP Pesquisa.

CONFLITOS DE INTERESSE

Os autores não têm conflitos de interesse.

CONFORMIDADE COM PADRÕES ÉTICOS

Este artigo não contém estudos com participantes humanos ou animais realizados por qualquer um dos autores.