



REAd. Revista Eletrônica de Administração (Porto Alegre)

ISSN: 1980-4164

ISSN: 1413-2311

Escola de Administração da UFRGS

Pereira, Matheus Machado de; Rosa, Thais Gomez da; Bender, Reisoli
**INFLUÊNCIA DO GOOGLE TRENDS EM AÇÕES LISTADAS NA BOLSA DE
VALORES BRASILEIRA: EVIDÊNCIAS A PARTIR DA MODELAGEM PVAR1**

REAd. Revista Eletrônica de Administração (Porto Alegre),
vol. 26, núm. 3, 2020, Setembro-Dezembro, pp. 796-818
Escola de Administração da UFRGS

DOI: <https://doi.org/10.1590/1413-2311.303.101823>

Disponível em: <https://www.redalyc.org/articulo.oa?id=401172030011>

- Como citar este artigo
- Número completo
- Mais informações do artigo
- Site da revista em redalyc.org



Sistema de Informação Científica Redalyc

Rede de Revistas Científicas da América Latina e do Caribe, Espanha e Portugal

Sem fins lucrativos acadêmica projeto, desenvolvido no âmbito da iniciativa
acesso aberto

INFLUÊNCIA DO GOOGLE TRENDS EM AÇÕES LISTADAS NA BOLSA DE VALORES BRASILEIRA: EVIDÊNCIAS A PARTIR DA MODELAGEM PVAR¹

Matheus Machado de Pereira²

Thais Gomez da Rosa³

Reisoli Bender Filho⁴

<http://dx.doi.org/10.1590/1413-2311.303.101823>

RESUMO

O trabalho propôs analisar como as pesquisas no buscador *Google* influenciam o retorno, a volatilidade e o volume de negociações das ações que compõem o índice Ibovespa, considerando o período entre 2015 e 2020. Para isso aplicou-se a modelagem de Pannel de Vetores Autorregressivos (PVAR). Os resultados do volume histórico de pesquisas do nome da empresa e do *ticker* apresentam relação bidirecional com o desvio padrão dos retornos, da volatilidade e do volume negociado, sugerindo que a demanda de informação do investidor é atendida, em parte, por pesquisas em buscadores, efeito que também é observado no aumento do volume de negociações, após um choque no volume histórico de pesquisas do *ticker*. Essas evidências apontam à eficiência do mercado, pelo menos, para situações semanais, em que há a possibilidade de o investidor pesquisar e compreender cada nova situação do mercado. Entretanto, o acompanhamento da empresa não garante, pelo menos no longo prazo, que os retornos sejam maiores, determinando que a Hipótese do Mercado Eficiente, na versão semi-forte, seja indiretamente observada, pelo aumento de negociações sem a devida alteração no

¹ Recebido em 9/4/2020, aceito em 30/10/2020.

² Universidade Federal de Santa Maria. Centro de Ciências Sociais e Humanas. Santa Maria – RS (Brasil); <http://orcid.org/0000-0003-2854-1856>; mateus.mpereira@hotmail.com.

³ Universidade Federal de Santa Maria. Centro de Ciências Sociais e Humanas. Santa Maria – RS (Brasil); <http://orcid.org/0000-0001-6774-351X>; thaisgomezr@gmail.com.

⁴ Universidade Federal de Santa Maria. Centro de Ciências Sociais e Humanas. Santa Maria – RS (Brasil); <http://orcid.org/0000-0002-1019-4414>; reisolibender@yahoo.com.br.

retorno. Para tanto, a utilização do *Google Trends* pode, em alguma medida, melhorar a acurácia de modelos de previsão que busquem prever o retorno, a volatilidade e o volume de ações.

Palavras-chave: *Google Trends*, Mercado de ações, Índice Ibovespa. PVAR.

INFLUENCE OF GOOGLE TRENDS ON SHARES LISTED ON THE BRAZILIAN STOCKS EXCHANGE: EVIDENCE FROM PVAR MODELING

797

The proposed work analyzed how searches on the Google search engine influence the return, volatility and volume of the shares that make up the Ibovespa index, considering the period between 2015 and 2020. For this, the Auto-regressive Vector Panel (PVAR) modeling was applied. The results implies that the historical search volume for the company name and the ticker show a bidirectional relationship with the standard deviation of returns, volatility and traded volume, suggesting that the investor's demand for information is met, in part, by search engines, which that there is also increase in the traded volume, after a shock, in historical volume of search on the ticker. This evidence points to market efficiency, at least, for weekly situations, in which there is a possibility for the investor to research and understand each new market situation. However, the company's follow-up does not guarantee, at least in the long run, determining that the Efficient Market Hypothesis, in the semi-strong version, is indirectly obtained, by increasing the volume traded without altering the return in long run. Therefore, the use of Google Trends can, to some extent, improve the accuracy of forecasting models that seek to predict the return, volatility and volume of shares in short term.

Keywords: Google Trends, Stock markets, Ibovespa index. PVAR.

INFLUENCIA DE *GOOGLE TRENDS* SOBRE LAS ACCIONES QUE COTIZAN EN LA BOLSA DE VALORES BRASILEÑA: PRUEBAS DEL MODELADO PVAR

El trabajo propuso analizar cómo las búsquedas en Google influyen en la rentabilidad, volatilidad y volumen de negociación de las acciones que componen el índice Ibovespa, considerando el período comprendido entre 2015 y 2020. Para ello, el Panel de Vector Auto-regresivo (PVAR) fue aplicado. Los resultados del volumen de búsqueda histórico para el nombre de la empresa y el *ticker* muestran una relación bidireccional con la desviación estándar de los rendimientos, la volatilidad y el volumen negociado, lo que sugiere que la demanda de información del inversor se satisface, en parte, mediante búsquedas en los motores de búsqueda.

efecto que también se observa en el aumento del volumen de operaciones, luego de un choque en el volumen histórico de búsquedas del *ticker*. Esta evidencia apunta a la eficiencia del mercado, al menos, para situaciones semanales, en las que existe la posibilidad de que el inversor investigue y comprenda cada nueva situación del mercado. Sin embargo, el seguimiento de la empresa no garantiza, al menos en el largo plazo, que los retornos sean mayores, determinando que la Hipótesis de Mercado Eficiente, en la versión semi-fuerte, se observe indirectamente, al incrementar las operaciones sin el debido cambio en el retorno. Con este fin, el uso de *Google Trends* puede, hasta cierto punto, mejorar la precisión de los modelos de pronóstico que buscan predecir el rendimiento, la volatilidad y el volumen de acciones.

Palabras clave: Google Trends. Bolsa de valores. Índice Ibovespa. PVAR

INTRODUÇÃO

O preço como sinalizador do mercado é um conceito conhecido a décadas, podendo ser encontrado no trabalho de Roberts (1967) e tendo como ideia principal que o preço reflete todas as informações possíveis de um referente ativo. Logo, sendo o ativo precificado corretamente ele serve como indicador para o investidor, conceito comumente denominado de Hipótese do Mercado Eficiente (HME).

Essa hipótese ampara-se no argumento de que todas informações relevantes são incorporadas, de forma imediata e correta, aos preços dos ativos financeiros, de forma que a melhor estimativa para o preço de um título é seu preço atual (ver Belo e Brasil, 2006). Entretanto, Fama (1970) ressalta que a validade da HME depende de certas condições, caso de as informações estarem disponíveis a todos os agentes; da ausência de custos de transação e; de que todos os agentes possuam expectativas homogêneas.

Essas condições ampliaram as discussões acerca dos determinantes da precificação dos ativos financeiros e a HME. Linha seguida, anos mais tarde, por Fama (1970 e 1991) quando sugere ser impossível para o investidor obter lucros anormais no longo prazo. Para tanto desenvolveu três versões da referida hipótese: i) fraca, ii) semi-forte, e, iii) forte. A forma primeira postula que os preços das ações refletem as informações dos preços passados. A segunda versão postula que os preços refletem seu comportamento passado e ainda refletem as informações disponíveis ou públicas da empresa.

A terceira e mais restritiva das hipóteses considera que os preços refletem toda informação pública e privilegiada, assim sendo, os preços partem para um novo equilíbrio, o

qual reflete toda e qualquer informação, sendo que nenhum tipo de informação permite ao investidor obter lucros extraordinários, pois essa informação também já está refletida no preço do ativo. Uma forma fácil de compreender a HME é pensar que o comportamento dos preços é um *random walk* (passeio aleatório).

Os pressupostos da HME são um pouco restritivos, pois baseiam-se em: i) não deve haver custos de transação; ii) as informações no mercado devem ser divulgadas simetricamente; iii) os participantes do mercado são capazes de compreender essas informações, portanto, apesar de consolidada, essa teoria também é largamente criticada no mercado (Campbell e Shiller, 1988; Haugen e Jorion, 1996 e; Wouters, 2006 entre outros), sobremaneira por utilizar determinadas “anomalias” que contrapõem a eficiência do mercado, conforme Naseer e Tariq (2015).

Naseer e Tariq (2015) fazem uma revisão da literatura sobre anomalias relacionadas a sazonalidade ou ainda preço da ação, um exemplo é do comportamento sazonal do retorno das ações é que no mês de janeiro o retorno é temporariamente maior que em outros meses, ou que o final do mês apresenta retornos maiores que início do mês; sobre o valor da ação, o efeito sobre reação em que há uma reação exagerada sobre notícias, bem como notícias recentes são mais impactantes que notícias antigas, surpresas positivas levam um choque lento nas ações, enquanto surpresas negativas levam uma queda rápida nos preços das ações. Anomalias como essas que segundo Naseer e Tariq (2015), não poderiam atuar num cenário de mercado eficiente.

Na proposição de Neto (2006), as anomalias são resultado da interpretação dos investidores, que faz parte da eficiência de mercado. Ainda, Neto (2006) argumenta que há confusão em pesquisas que buscam rejeitar a HME apresentando evidências de irracionalidade, mas para Fama (1970) o mercado preserva a eficiência na presença de investidores irracionais.

Essas divergências tem resultado na busca e na incorporação de novos instrumentos nos modelos de previsão de preços, os quais têm se mostrado eficazes para analisar o comportamento humano como também na realização de previsões, caso dos denominados *Big Data*. Nesta linha, destacam-se os mecanismos de pesquisa que vem realizando o armazenamento de informações sobre o que indivíduos pesquisam na internet, mecanismos esses que podem ser usados para prever diversas tendências. Considerando esta estrutura de informações, um crescente número de trabalhos tem usado essa busca (pesquisa) como *proxy* da atenção do investidor, caso dos trabalhos de Challet e Ayed (2014) e de Hu *et al.* (2018).

Este instrumento tem operacionalidade desde 2004 a partir do buscador da Google que armazena e disponibiliza o volume histórico de pesquisas⁵ (VHP), esse indicador transforma em série temporal um ou mais termos pesquisados no buscador. Sobre isso, Dzielinski (2012) aponta que 70% do tráfego de pesquisas é realizado a partir do buscador da Google. Por essa razão, o indicador pode apresentar propriedades interessantes, como apresentar tendências, especialmente de curto prazo, ou ainda servir como *proxy* da atenção do investidor, pois pesquisas sobre o *ticker*⁶ da empresa ou até o nome da empresa pode simbolizar o interesse desse investidor e sua demanda de informação por esse ativo.

Este indicador já foi usado em estudos à previsão de indicadores econômicos (ver Choi e Varian, 2012) que desenvolveram exemplos para vendas e viagens, o VHP já foi usado para a previsão de índices de bolsa de valores (ver Heiberger, 2015; Hu *et al.*, 2018), Heiberger (2015) considera o *Google trends* um bom indicador de más notícias ao usar o VHP para prever o Standart e Poor's, enquanto Hu *et al.* (2018) Dow Jones e S&P. Ao passo que inúmeros trabalhos usaram para analisar a relação com retornos (ver Dzielinski, 2012; Loughlin e Harnisch 2013; Perlin *et al.*, 2016; Bijl *et al.*, 2016; Challet e Ayed, 2014 e Ramos Ribeiro e Perlin, 2017), Dzielinski (2012) considera que quanto maior incerteza, maior demanda de informação, portanto o VHP carrega essa incerteza, já Perlin *et al.* (2016) considera que o VHP tem um impacto positivo, mas pequeno.

Há trabalhos também para a volatilidade (ver Perlin *et al.*, 2016 e Ramos, Ribeiro e Perlin, 2017), em que Ramos, Ribeiro e Perlin (2017) encontram relações bidirecionais com o VHP. Além deste instrumento, há também trabalhos que usam *tweets* para compreender o humor do mercado (ver Bollen *et al.*, 2011) ou que combinam *tweets* com VHP, conforme Loughlin e Harnisch (2013), os autores ainda afirmam que os *tweets* são melhores indicadores que o *Google trends*. Assim como pesquisas focadas na demanda de informação (ver Vlastakis e Markellos, 2012, Moussa, Delhoumi e Ouda, 2012 e Wang, 2012) em que o Google e buscadores atendem à demanda de informação dos investidores.

De forma geral, esses trabalhos inovam em metodologia, período, frequência e base dos dados, objetivos e mercados. Todavia, embora crescente, esta estrutura analítica ainda encontra vasto espaço para o mercado brasileiro, dado que apenas um estudo se propôs a analisar o mercado doméstico, o de Ramos, Ribeiro e Perlin (2017), a partir do impacto de VHP no índice Ibovespa utilizando termos específicos. Apesar de importante e robusto, o trabalho tem caráter

⁵ Volume Histórico de Pesquisas é como ficam armazenadas as pesquisas realizadas em buscadores na *internet*, como por exemplo o da Google.

⁶ *Ticker* é o código de uma ação listada na bolsa de valores.

abrangente, uma vez que tem objetivo exploratório e relaciona apenas as buscas de dois *tickers* (PETR4 e VALE5) com o Ibovespa.

Além de encontrar causalidade de Granger entre o VHP e retorno do Ibovespa, o trabalho de Ramos, Ribeiro e Perlin (2017) também compara estratégia *buy and hold* contra estratégias de *trading* que usando o VHP para prever retornos, nessa estratégia quando prevê-se que o retorno da ação será positivo, compra-se, quando prevê-se que o retorno será negativo, vende-se papel. Nessa comparação a estratégia de *trading* com o VHP demonstrou-se vencedora.

A partir dessas discussões, o trabalho busca ampliar o conjunto de evidências acerca do uso destas informações na precificação dos ativos ao analisar como as pesquisas no buscador *Google* sobre o nome da empresa e *ticker*, influenciam o retorno, a volatilidade e o volume de negociações das ações que compõem o índice Ibovespa, considerando o período entre 2015 e 2020 com dados semanais. Dessa forma, mesmo que o foco da pesquisa seja a relação de curto prazo.

Para isso utiliza-se um *Panel Vector Autoregression* (PVAR), metodologia que avança em relação ao tradicional método VAR ao permitir explicar a especificidade no nível das variáveis, introduzindo os efeitos fixos e isolando a resposta dos choques ao mesmo tempo em que permite a heterogeneidade não observada, conforme Love e Ariss (2013).

Ainda, tal método não encontra aplicação tanto para dados de informações de buscas quanto para o mercado acionário brasileiro. A partir disto, busca-se estruturar evidências que possibilitem um melhor entendimento do comportamento do investidor, especificamente a demanda de informações, no mercado acionário brasileiro.

A pesquisa parte da hipótese que se a demanda de informação é atendida por buscadores como o *Google* as variáveis devem ter uma relação de Granger Causalidade, ou seja, o VHP do *ticker* e nome da empresa devem Granger Causar retorno, volatilidade e volume de negociações, se a hipótese é falsa a Hipótese de Mercado Eficiente (HME) pode ser vista indiretamente pois informações não influenciam retornos, se essa hipótese é verdadeira o impacto dessa relação poderá ser visto a partir da Função Resposta ao Impulso, ainda se esse impulso é dissipado a relação pode indicar que, pois significa que acompanhar a empresa não gera retornos superiores no longo prazo, conforme a HME.

De forma simples, os resultados apontam que usar o VHP com nome da empresa e *ticker* da empresa em modelos de previsão pode gerar melhor acurácia, uma vez que as variáveis têm

relação de Granger Causalidade, entretanto essa relação é de curto prazo, em que os choques simulados se dissipam rapidamente.

Estruturalmente, além desta exposição preambular, o artigo está estruturado em outras quatro seções. Na segunda apresentam-se as evidências empíricas; na terceira a estrutura metodológica; na quarta os resultados e discussões e; na última, delineiam-se as principais conclusões.

1 EVIDÊNCIAS EMPÍRICAS: TRABALHOS RELACIONADOS

Cronologicamente, um dos primeiros trabalhos relacionando o VHP fornecido pelo *Google Trends*, com indicadores econômicos, foi o trabalho de Choi e Varian (2012), que tem carácter informativo sobre como usar a ferramenta, provando a eficácia da ferramenta para melhorar previsões desde a venda de veículos, o desemprego, como também sobre a confiança do consumidor, a partir de diferentes metodologias. Com isso, o trabalho demonstra a flexibilidade e a aplicabilidade do banco de dados nos diversos setores, abrindo caminho para futuros trabalhos.

Dentre os trabalhos aplicados, destaca-se o Bollen *et al.* (2011) que relata uma correlação positiva entre os *tweets* e o indicador da atividade industrial do *Dow Jones* a partir uma análise de sentimentos e do teste de causalidade de Granger, encontrando resultados que apontam que os *tweets* melhoram a acurácia de modelos preditivos do fechamento do *Dow Jones Industrial Average*. Loughlin e Harnisch (2013) ao evoluir na análise, usam os *tweets* e o volume de buscas a partir do *Google Trends* para prever o retorno de empresas específicas. Os resultados sugerem que o VHP não é um bom instrumento, mas os *tweets* se mostraram significativos.

Os primeiros a usarem explicitamente o *Google Trends* como *proxy* para a demanda de informação são Vlastakis e Markellos (2012), os quais a partir de modelos heterocedásticos encontraram ligação entre a volatilidade e o volume de negociações de ações da bolsa norte-americana NASDAQ. Moussa, Delhoumi e Ouda (2012) que também usaram modelos heterocedásticos, ainda apontaram a importância do conceito de demanda de informação e uso do *Google Trends* para estratégias de investimento que usam modelos preditivos, pois a volatilidade sofre um impacto maior.

Dzielinski (2012) afirma que a frequência de pesquisas em buscadores tem influência no retorno agregado das ações e na volatilidade. Para o autor, as vantagens do uso do *Google Trends* é a alta frequência dos dados e que eles são gerados de forma espontânea. O modelo usado foi a regressão linear, do retorno contra a diferença do índice de buscas e a diferença de

outros índices de mercado e o retorno, encontrando resultados de que a maior incerteza na economia eleva a demanda de informação, elevando com isso o índice de pesquisas do *Google Trends*.

Preis, Moat e Stanley (2013) relacionaram 98 termos de pesquisa a operações compradas e vendidas. O resultado encontrado indica que é possível montar uma estratégia de investimentos baseadas nas pesquisas do buscador. Nesta linha, o *Google Trends* poderia ser capaz de indicar as tendências de curto prazo do mercado acionário.

Sobre essas pesquisas, Challet e Ayed (2014) alertam sobre alguns detalhes que podem comprometer os resultados, como o viés de escolha, desde o grupo de palavras que pode levar a conclusões espúrias. Para evitar esse problema, os autores usam o nome da empresa e o *ticker*, período de análise que influencia nos resultados como também a frequência semanal dos dados. Os autores, inclusive, sugerem que modelos lineares não são adequados à previsão do mercado, por isso usam *Support Vector Machines* (SVM). O trabalho conclui que os preços são fracamente dependentes das buscas no *Google*, o que demonstra que as próprias buscas são sensíveis a eventos externos, tais como más notícias.

Heiberger (2015) ao analisar a influência do volume de pesquisas no índice norte-americano *Standart and Poor's 100*, a partir do uso de regressões lineares, relacionando linearmente termos específicos e o índice *Standart and Poor's 100*, encontrou resultados diferentes, concluindo que o VHP disponibilizado pelo *Google Trends*, funciona adequadamente como indicador de “más notícias”, dado que os indivíduos usam o buscador em situações negativas economicamente. Assim, a atenção coletiva segue a tendência dos jornais, contribuindo para reações exageradas em situações de más notícias.

Bijl *et al.* (2016) buscaram também no VHP do *Google Trends* uma forma de prever o retorno do mercado a partir de modelos lineares. O trabalho conclui que o volume de pesquisas no buscador *Google* é significativo, mas seu impacto é pequeno nos retornos, além disso os autores apontam a dificuldade em generalizar a pesquisa, já que a relação entre o volume de pesquisas e retorno mudam ao longo do tempo.

O trabalho de Perlin, *et al.* (2016) ao selecionar um grupo específico de palavras pesquisadas no *Google*, baseado em quatro livros de investimento, analisaram sua relação com os índices. O trabalho desenvolve três modelos que buscam avaliar o impacto das pesquisas no retorno e na volatilidade de quatro índices a partir de Vetores Autorregressivos (VAR). Os resultados apontaram que buscas por “*stock*” tem impacto significativo no retorno e na

volatilidade, resultado esse que pode indicar que os investidores pesquisam informações sobre o mercado antes de investir.

Já Hu *et al.* (2018), Ahmed *et al.* (2017) e Lee, Cha e Jang (2017) desenvolveram estudos a partir de Redes Neurais Artificiais em diferentes variáveis, sendo elas: preço de abertura do Dow Jones e S&P; preço de abertura, fechamento e volume da Bolsa do Paquistão, Shanghai e Índice da Coreia do Sul, respectivamente. Os trabalhos de Hu *et al.* (2018) e Ahmed *et al.* (2017) concluíram que os dados do *Google Trends* melhoram significativamente os modelos de previsão, enquanto que Lee, Chaa e Jang (2017) encontrou resultados diferentes para a bolsa coreana, com o índice do *Google Trends* não melhorando a qualidade da previsão.

Conquanto a literatura internacional esteja crescendo, para o mercado brasileiro, até o momento, encontra-se unicamente o estudo desenvolvido por Ramos, Ribeiro e Perlin (2017), os quais estudaram a influência do volume de pesquisas no *Google* no índice Ibovespa, especificamente sobre as variáveis volume de negociação, retorno e volatilidade, considerando o período entre 2007 e 2014. Para isso, foram aplicadas as metodologias VAR e causalidade de Granger. Os resultados apontaram que há relação Granger causal bidirecional entre as pesquisas e o índice Ibovespa; ainda, foi verificado que a estratégia de *trading*, usando os termos específicos, apresentou desempenho melhor que a estratégia *buy and hold*.

Apesar da ampla quantidade de estudos acerca do impacto da atenção do investidor, os trabalhos realizados não chegam a um consenso sobre qual metodologia empregar, o impacto real das buscas na *internet*, qual grupo de palavras seriam as melhores *proxies*, e até a melhor frequência dos dados para se analisar o fenômeno. Todavia, destaca-se que a maioria dos trabalhos analisam os retornos de empresas específicas, a partir de diferentes metodologias como Redes Neurais Artificiais (RNA), *Support Vector Machines* (SVM), séries temporais e regressões lineares, enquanto que a frequência mais usada ainda é semanal e há uma preferência por buscas com o nome da empresa.

2 MATERIAL E MÉTODOS

2.1 DADOS

Para entender a demanda de informação dos investidores são coletadas duas variáveis coletadas do *Google Trends*: i) *ticker* da empresa e ii) nome da empresa, as quais são definidas como *ttp* e *tep*, respectivamente, nos modelos empíricos. Essas variáveis servem como *proxy* à demanda de informação, pois além da vantagem da simplicidade, são resultados da ação voluntária dos investidores, sem sofrer intervenção alguma. Os dados não são brutos, pois o

buscador de pesquisas normaliza o número de pesquisas realizadas na semana dividindo o número de pesquisas pelo valor máximo no intervalo, passando para um intervalo entre 0 e 100. A frequência semanal para o VHP do *Google Trends* tem início no domingo e termina no sábado. Ainda, à análise proposta não será considerado limitação geográfica que a fonte de dados permite, devido a importância do investidor estrangeiro na bolsa brasileira. No *software R* os dados podem ser importados a partir do pacote *Gtrends* da fonte *Google Trends*.

Os dados referentes ao retorno, a volatilidade e ao volume de negociações, definidas como *rap*, *volap* e *vp*, são coletados do Yahoo Finanças, mas diferente das anteriores, as ações são negociadas apenas em dias úteis, com a semana começando na segunda-feira e terminando na sexta-feira. A amostra das empresas foi escolhida a partir da composição teórica do índice Ibovespa, que é reavaliada quadrimestralmente, sendo composta pelas empresas mais líquidas da bolsa de valores, isso corresponde à 80% das negociações (B3, 2020). Como o índice inclui ações ordinárias, preferenciais e *units*, isso gera um total de 73 *tickers* dentro do índice Ibovespa.

Todavia, a amostra analisada contém somente 66 *tickers*, sendo que foram excluídas as informações das empresas AZUL4, CRBF3, HAPV3, GNDI3, BRDDT3, por serem empresas cuja *Inicial Public Ofert* (IPO) foi realizado depois de janeiro de 2015, e os *tickers* da BPAC11 e da BRFS3 por problemas no *Google Trends*. No *software R*, os dados sobre as variáveis de mercado podem ser coletados a partir do pacote *BatchGetSymbols*. Este conjunto de informações foi coletado para o período entre 2015 e 2020, com frequência semanal, totalizando 17490 observações por variável.

O retorno das ações é frequentemente a variável alvo das análises financeiras que utilizam modelos econométricos, de forma que essa foi a variável definida como endógena no primeiro modelo (modelo 1). A volatilidade constitui-se em outra medida largamente utilizada, nesse caso, a volatilidade histórica, sendo essa variável utilizada no segundo modelo (modelo 2). Por fim, no terceiro modelo (modelo 3), utilizou-se a medida referente ao volume de negociações, que consiste no número de negociações da semana do ativo em questão. Essas três variáveis tem suas definições expressas nas expressões de (1) a (3):

$$Retorno_{it} = \frac{P_{it}}{P_{it-1}} \quad (1)$$

$$Volatilidade_{it} = \left(\frac{P_{it}}{P_{it-1}} \right)^2 \quad (2)$$

$$Volume_{it} = \sum_{i=1}^{n=5} negociações\ diárias \quad (3)$$

Nas estimações, essas variáveis são identificadas como *rap* para o retorno, *volap* à volatilidade e *vp* para o volume de negociações. Ainda, as variáveis foram padronizadas, ou seja, de cada observação se subtrai a média e divide pelo desvio padrão, essa transformação é interessante pois a interpretação passa a ser em desvios padrão, utilizando a definição em (4):

$$Y_p = \frac{y_i - \hat{y}}{\sigma} \quad (4)$$

É importante salientar que Moretin (2011) define a equação em (1) como a forma mais simples de retorno, sendo a taxa de variação o resultado entre o preço inicial e o final, conforme definido na equação (1). O pacote *BatchGetSymbols*, desenvolvido por Perlin (2018), ao gerar os retornos, usa essa definição, embora ela possa subestimar o retorno.

2.2 MODELO TEÓRICO

O modelo *Panel Vector Autoregression* (PVAR), conforme encontrado em Love e Zicchino (2006), possui a mesma estrutura do tradicional do *Vector Autoregressive Models* (VAR), ao assumir que todas as variáveis são consideradas endógenas e interdependentes, entretanto, adiciona a dimensão de seção cruzada (*cross section*) (Canova e Ciccarelli, 2013). A partir de Abrigo e Love (2015), o modelo PVAR com efeitos fixos é representado por:

$$Y_{it} = Y_{it-1}A_1 + \dots + Y_{it-p+1}A_{p-1} + Y_{it-p}A_p + X_{it}B + u_{it} + e_{it} \quad (5)$$

com $i \in \{1, 2, \dots, N\}$ e $t \in \{1, 2, \dots, T\}$

em que Y_{it} é o vetor de variáveis dependente ($1 \times k$); X_{it} é o vetor de variáveis exógenas ($1 \times l$); u_i é vetor de efeitos fixos ($1 \times k$) e e_{it} o erro idiossincrático ($1 \times k$). As matrizes A_1 ; A_{p-1} , A_p ($k \times k$) e B ($l \times k$) correspondem aos parâmetros a serem estimados. Em relação aos erros, assume-se que eles são não serialmente correlacionados e seguem $E[e_{it}] = 0$, $E[e'_{it}e_{it}] = \Sigma$ e $E[e'_{it}e_{is}] = 0$ para todo $t > s$.

Para evitar a obtenção de coeficientes viesados em função da correlação entre os efeitos fixos e os regressores, utiliza-se o procedimento de Helmert, seguindo Love e Zicchino (2006). Este procedimento preserva a ortogonalidade entre as variáveis transformadas e os regressores defasados, permitindo usar os regressores defasados como instrumentos e estimar a equação em (5) pelo método de momentos generalizados (GMM), proposto por Arellano e Bover (1995).

Como critério de adequação, o modelo PVAR precisa atender a condição de estabilidade, de forma que o módulo da matriz \bar{A} deve ser menor que um. Isso permite que o PVAR seja invertível, bem como possuir um vetor de ordem infinita de média móvel (VMA), o que garante a interpretação das funções impulso-resposta e à decomposição da variância dos erros (ver Abrigo e Love, 2015).

As funções impulso-resposta (IR) Φ_i são obtidas reescrevendo o modelo apresentado em (5) como um vetor média móvel, onde Φ_i são os parâmetros VMA:

$$\Phi_i = \begin{cases} I_k, i = 0 \\ \sum_{j=1}^i \Phi_{t-j} A_j, i = 1, 2, \dots, N \end{cases} \quad (6)$$

E a decomposição da variância dos erros pode ser expressa por:

$$Y_{it+h} - E[Y_{it+h}] = \sum_{i=0}^{h-1} e_{1(t+h-i)} \Phi_i \quad (7)$$

em que Y_{it+h} é o vetor observado no tempo $t + h$ e $E[Y_{it+h}]$ é o vetor de previsão no tempo t .

Porém, assim como na modelagem de séries temporais ou em dados em painel, inicialmente é preciso avaliar a estrutura das séries utilizadas. Para essa finalidade, aplica-se o teste de raiz unitária para Painel-VAR *Cross-sectionally Augmented Dickey-Fuller* (CADF), proposto por Pesaran (2007). O teste propõe um procedimento para eliminar a dependência cruzada das séries pela inclusão das médias *cross-section* dos termos defasados e das primeiras diferenças das séries individuais ao teste ADF padrão (PESARAN, 2007). A formulação CADF segue a especificação em (8):

$$\Delta y_{it} = a_i + b_i y_{i,t-1} + c_i \bar{y}_{t-1} + d_i \Delta \bar{y}_t + u_{it} \quad (8)$$

cujas hipóteses nulas são $H_0: b_i = 0$ para todo i .

Seguindo os testes à definição do modelo aplicado, a seleção do PVAR foi realizada a partir da escolha da defasagem ótima, seguindo Andrews e Lu (2001), que propuseram o critério de seleção do modelo (MMSC), baseado na estatística J de Hansen (1982) que considera restrições sobre-identificadas. Conforme Abrigo e Love (2015), esta proposta é análoga aos diversos critérios de seleção comumente utilizados, caso do Critério de Informação de Akaike

(AIC), Critério de Informação Bayesiano (BIC) e Critério de Informação de Hannan-Quinn (HQIC).

Formalmente, o critério proposto por Andrews a Lu (2001) selecionada o par de vetores (p, q) que minimiza, conforme equações de (9) a (11):

$$MMSC_{BIC,n}(k, p, q) = J_n(k^2 p, k^2 q) - (|q| - |p|)k^2 \ln n \quad (9)$$

$$MMSC_{AIC,n}(k, p, q) = J_n(k^2 p, k^2 q) - 2k^2(|q| - |p|) \quad (10)$$

$$MMSC_{HQIC,n}(p, q) = J_n(k^2 p, k^2 q) - Rk^2(|q| - |p|)\ln \ln n, \quad R > 2 \quad (11)$$

em que $J_n(k, p, q)$ são as estatísticas J de restrição sobre-identificadas para k -variações do PVAR de ordem p e condições baseadas em q defasagens das variáveis dependentes. Assim, a seleção de ordem ótima é definida pelo menor critério de seleção.

Considerando os dados e o modelo teórico, estimaram-se três formulações (modelos empíricos) utilizando a equação exposta em (5) e definidas como modelos 1, 2 e 3, respectivamente:

Modelo 1: vetor de variáveis dependentes $Y_{it} = rap_{it}$ e vetor de variáveis exógenas $X_{it} = ttp_{it}; tep_{it}$. Assim o primeiro modelo 1 avalia a relação entre retornos e o VHP do *ticker* e nome da empresa.

Modelo 2: vetor de variáveis dependentes $Y_{it} = volap_{it}$ e e vetor de variáveis exógenas $X_{it} = ttp_{it}; tep_{it}$. Dessa forma o modelo 2 avalia a relação entre a volatilidade e o VHP do *ticker* e nome da empresa.

Modelo 3: vetor de variáveis dependentes $Y_{it} = vp_{it}$ e e vetor de variáveis exógenas $X_{it} = ttp_{it}; tep_{it}$ e rap_{it} . Então o modelo 3 avalia a relação entre o volume de negociações e o VHP do *ticker* e nome da empresa.

Nos modelos, as variáveis são identificadas como ttp o VHP do *ticker*, tep o VHP da empresa, rap o retorno, $volap$ a volatilidade e vp o volume. Essa modelagem evidencia a relação entre o VHP do *ticker*, da empresa e as variáveis de mercado. Resumidamente, a informação gerada pelo PVAR, de acordo com as funções impulso-resposta, fornece o impacto das pesquisas do investidor nas variáveis de mercado e a aplicabilidade em modelos de previsão.

Complementando a proposta metodológica, foi aplicado o teste de Causalidade de Granger que é realizado a partir de um teste de restrição de Wald, aplicado nos parâmetros estimado A_{it} e B_{it} pelo modelo exposto na equação em (1). Analiticamente, há causalidade de Granger entre Y e X se nem todos betas e gamas forem iguais a zero. Assim, se apenas os A_{it}

são significativos e B_{it} são iguais a zero, a Granger causalidade é unidirecional, já se ambos A_{it} e B_{it} são diferentes de zero, a Granger causalidade é bidirecional.

A partir da análise da metodologia implementada, a significância das variáveis e a estabilidade dos modelos fornecem informações sobre o comportamento do investidor, em específico sobre a demanda de informação e alteração no desvio padrão. As funções impulso-resposta (ou Função Resposta ao Impulso), indicam se essa relação é positiva e se o choque perde efeito no longo prazo, esse último resultado condiz com o efeito predito pela HME. Em relação a Granger Causalidade, caso ela seja significativa, pode-se dizer que as variáveis auxiliam na previsão uma da outra, em termos de precedência temporal, sendo esse resultado importante para a aplicação de métodos preditivos.

3 RESULTADOS

À estimação dos modelos inicialmente identificou-se a estrutura das séries e a forma como elas se relacionam no longo prazo. Para a definição da ordem de integração aplicou-se o teste de raiz unitária *Cross-sectionally Augmented Dickey-Fuller* (CADF), estimado com três *lags* e sem tendência. Os resultados encontram-se na Tabela 1 e indicam que todas as variáveis foram estacionárias em nível.

Tabela 1 – Resultados do teste de raiz unitária CADF, de Pesaran

Variável	Estatística	P-value
<i>Trends ticker (ttp)</i>	-4,070***	0,0000
<i>Trends empresa (tep)</i>	-5,423***	0,0000
Retorno (<i>rap</i>)	-8,055***	0,0000
Volatilidade (<i>volap</i>)	- 6,400***	0,0000
Volume (<i>vp</i>)	-5,638***	0,0000

***, ** e * indica que a hipótese nula é rejeitada ao nível de significância de 1%, 5% e 10% respectivamente. O STATA executa a partir da função `pescadf`

Na etapa seguinte foi definido o número ótimo de *lags* a partir dos critérios de informação (MBIC, MAIC e MQIC) propostos por Andrews e Lu (2001). Os resultados apontam que o modelo com três *lags* ajusta-se melhor (ver Tabela 2), definição utilizada à estimação dos três modelos propostos.

Tabela 2 – Resultados do teste de Andrews e Lu para seleção de *lags*

Modelo 1				Modelo 2			Modelo 3		
<i>Lag</i>	MBIC	MAIC	MQIC	MBIC	MAIC	MQIC	MBIC	MAIC	MQIC
1	406,12	615,38	546,40	364,96	574,22	505,24	659,97	869,05	800,08
2	286,84	426,34	380,36	140,38	279,89	233,90	505,47	644,98	598,99
3	208,63	278,38	255,39	105,265	175,01	152,02	238,37	308,13	285,13
	*	*	*	*	*	*	*	*	*

Ainda, como forma de garantir a confiabilidade das estimativas dos modelos, procedeu-se o teste de estabilidade, conforme resultados expostos na Tabela 3. Como todos módulos dos autovalores foram estritamente menores que a unidade nas três formulações propostas, o PVAR é invertível e possui representação de média móvel de um vetor de ordem infinita, fornecendo interpretação às funções impulso-resposta (ver Abrigo e Love, 2015).

Tabela 3 – Resultados dos testes de estabilidade do PVAR

Modelo 1	Modelo 2	Modelo 3
0,9689	0,9622	0,9170
0,7878	0,7851	0,7774
0,4761	0,4957	0,6532
0,4761	0,4865	0,4449
0,4375	0,4580	0,4449
0,4375	0,4580	0,4269
0,2825	0,4444	0,4269
0,2451	0,4444	0,4057
0,2451	0,0295	0,4057

O STATA executa a partir da função `pvarstable`.

Definida a estrutura das variáveis e dos modelos, estimou-se as funções Resposta ao Impulso (FIR) para os três modelos, respectivamente, conforme resultados apresentados nas Figuras 1, 2 e 3. Nas estimações, as variáveis são identificadas por *ttp* o VHP do *ticker*, *tep* o VHP da empresa, *rap* o retorno, *volap* a volatilidade e *vp* o volume, seguindo as definições apresentadas na metodologia.

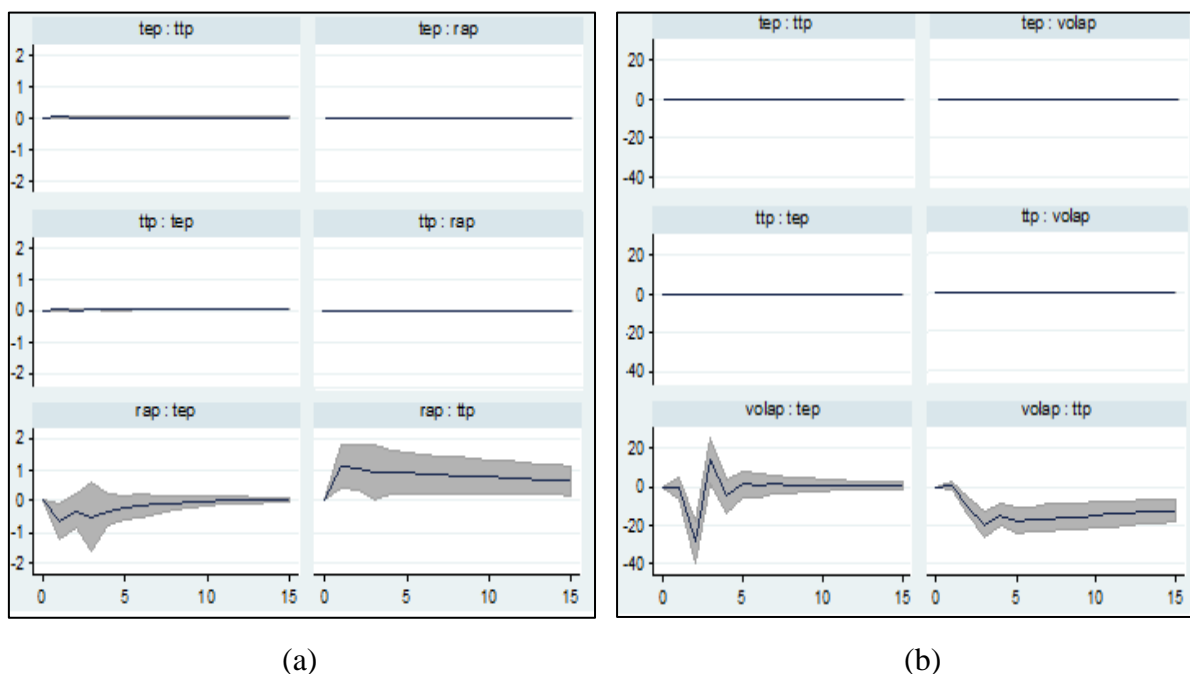
Verifica-se que o impacto de um choque nos Volume Histórico de Pesquisa (VHP) das variáveis coletadas no *Google Trends* apresenta impacto muito pequeno nas variáveis retorno e volatilidade, enquanto o choque na variável volume (*vp*) é dissipado rapidamente. Os achados

indiretamente concordam com a hipótese do mercado eficiente (HME) na sua versão semi-forte, ou ainda com a proposta de Neto (2006), quando as informações públicas não refletem nos retornos dado que os agentes logo precificam a nova informação, pelo menos para as relações semanais; entretanto, ainda não elimina a existência de possíveis impactos no curtíssimo prazo.

Sob a óptica da demanda de informação, um choque no desvio padrão do retorno eleva o desvio padrão do VHP *ticker*(*ttp*). Este resultado demonstra o comportamento do investidor, uma vez que, indicando que alterações nos retornos chamam a atenção do investidor e estimulam o investidor a acompanhar o *ticker* da respectiva ação, ainda esse choque não se dissipa no curto prazo, por outro lado empresas que não alteram retornos se mostram menos procuradas dada a estabilidade dos retornos. Esse resultado é observado no modelo 1.

Já, o choque no desvio padrão da volatilidade tem o efeito contrário, reduzindo o desvio padrão no VHP *ticker*(*ttp*), resultado indica que empresas voláteis são acompanhadas de forma estável, pois reduz o desvio padrão do VHP *ticker*, entretanto o efeito no VHP com o nome da empresa é ambíguo. Assim função impulso resposta corrobora o achado de Challet e Ayed (2014), em que a própria natureza da pesquisa é imprevisível, dado que um retorno mais elevado pode influenciar a pesquisa. Esse resultado é observado no modelo 2.

Figura 1 – Resultado das funções Resposta ao Impulso para o Modelo 1 e 2.



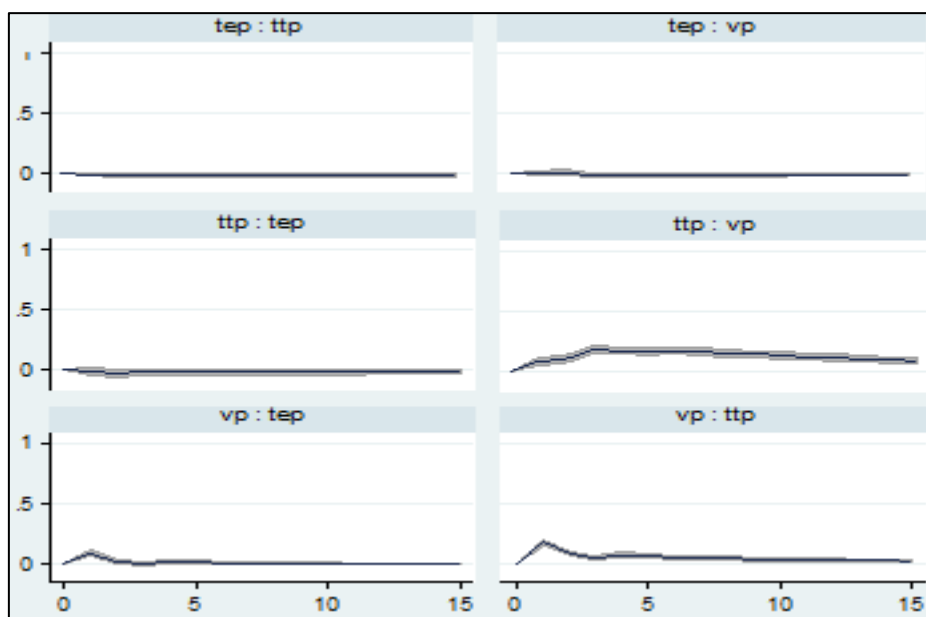
O STATA executa a partir da função *pvarirf*.

Os achados da Função Resposta ao Impulso alinham-se também com as evidências encontradas por Loughlin e Harnisch (2013), Challet e Ayed (2014), Bijl *et al* (2016) e Lee

Chaa e Jang (2017) quando retratam que as pesquisas no *Google Trends* não influenciam as variáveis e especialmente é relacionado com os resultados de Bijl *et al.* (2016), em que o impacto é estatisticamente significativo, porém pequeno. Resultado que reforça a hipótese de que a relação entre o VHP e os indicadores de mercado se alteram ao longo do tempo, o que dificulta as previsões como também a mensuração do impacto das pesquisas mesmo que a relação seja verdadeira.

Porém, estes resultados divergem daqueles encontrados por Campbell e Shiller (1988); Haugen e Jorion (1996) e Wouters (2006), dado de que o impacto no retorno padronizado são pequenos economicamente, sugerindo a ausência de anomalias sobre informação, embora não se descarta a possível presença por outro teste. Outro fator que diferencia os estudos reside no fato do uso dos buscadores como *Google Trends* serem bem mais recentes que as pesquisas elaboradas pelos respectivos trabalhos, outro fator é o impacto do buscador resulta em um melhor acesso à informação do que um pressuposto da HME.

Figura 3 – Resultado das funções Resposta ao Impulso para o Modelo 3



Fonte: O STATA executa a partir da função pvarirf

Sobre o terceiro modelo que analisa o volume de negociações, o choque do VHP *ticker* sugere que o processo de investimento pode ser influenciado pelas pesquisas em buscadores, indicando que uma das etapas do processo decisório é a pesquisa. Corroborando os achados de Vlastakis e Markellos (2012), que evidenciaram que as buscas tem impacto no volume de negociações. O resultado pode ser relevante para modelos preditivos que usam previsão, de

acordo com os achados de Moussa, Delhoumi e Ouda (2012). Essa relação reforça a hipótese de que o *Google Trends* pode ser usado como *proxy* para a demanda de informação do investidor, manifestando-se no volume de negociações.

Complementando a análise, aplicou-se o teste de causalidade de Granger, o qual indica que apenas a variável retorno não Granger causa VHP empresa, considerando o nível de significância de 5%. No Modelo 2, o VHP empresa (*tep*) não Granger causa volatilidade (*volap*) enquanto que no Modelo 3, o VHP empresa não Granger causa o VHP *ticker*. Estes resultados sugerem que, em termos de previsão, buscar o VHP do *ticker* é mais indicado que o VHP da empresa, possivelmente pela relação mais direta com o mercado. Ainda, o nome da empresa pode ser resultado de buscas por clientes, não investidores que objetivam conhecer a empresa ou mesmo navegar no site ou algum outro motivo não sendo exclusivamente por investidores. Outro fator que pode influenciar os resultados é de que os dados são semanais.

Essas evidências complementam os achados de Ramos, Ribeiro e Perlin (2017), quando encontraram que o VHP de termos coletados no *Google Trends* está relacionado as variáveis de mercado (retorno, volatilidade e volume), já que Granger-Causa retornos, volatilidade e volume do Ibovespa de forma bidirecional. Igualmente seguem a mesma linha dos resultados encontrados por Perlin *et al.* (2016) de que alguns termos pesquisados apresentam precedência temporal aos retornos e a volatilidade.

Esta bidirecionalidade da relação de causalidade de Granger é importante pois a pesquisa sobre a empresa (*tep*) ou o *ticker* (*ttp*) pode levar a uma melhor precificação na semana, ou auxiliar na relação entre a empresa e o investidor, ou ainda, os retornos maiores da empresa podem chamar a atenção do investidor, que passa a pesquisar mais sobre a empresa, realizando buscas no *ticker*, elevando o volume de negociações e a volatilidade.

Tabela 4 – Resultados do teste de causalidade de Granger

Equação	Excluída	χ^2
Modelo 1		
Retorno (<i>rap</i>)	VHP <i>ticker</i> (<i>ttp</i>)	12,578***
	VHP empresa (<i>tep</i>)	20,490***
	Modelo	29,815***
VHP <i>ticker</i> (<i>ttp</i>)	Retorno (<i>rap</i>)	10,734***
	VHP empresa (<i>tep</i>)	9,274**
	Modelo	21,453***
VHP empresa (<i>tep</i>)	Retorno (<i>rap</i>)	7,757*
	VHP <i>ticker</i> (<i>ttp</i>)	10,287**

	Modelo	15,384**
Modelo 2		
Volatilidade (<i>volap</i>)	VHP <i>ticker</i> (<i>ttp</i>)	18,865***
	VHP empresa (<i>tep</i>)	0,337
	Modelo	19,601***
VHP <i>ticker</i> (<i>ttp</i>)	Volatilidade (<i>volap</i>)	42,188***
	VHP empresa (<i>tep</i>)	28,546***
	Modelo	2494,369***
VHP empresa (<i>tep</i>)	Volatilidade (<i>volap</i>)	10,671**
	VHP <i>ticker</i>	13,177***
	Modelo	18,427***
Modelo 3		
Volume (<i>vp</i>)	VHP <i>ticker</i>	185,192***
	VHP empresa (<i>tep</i>)	8,211**
	Modelo	189,655***
VHP <i>ticker</i> (<i>ttp</i>)	Volume (<i>vp</i>)	194,940***
	VHP empresa (<i>tep</i>)	4,615
	Modelo	196,663 ***
VHP empresa (<i>tep</i>)	Volume (<i>vp</i>)	57,693***
	VHP <i>ticker</i> (<i>ttp</i>)	6,725**
	Modelo	71,595***

***, ** e * indica que a hipótese nula é rejeitada ao nível de significância de 1%, 5% e 10% respectivamente. O STATA executa a partir da função *pvargranger*

Os resultados também associam-se, ainda que indiretamente, com os achados de Bollen *et al.* (2016) que evidenciou relação Granger causal entre as variáveis de mercado e o VHP do *Google*, mas acompanhando um índice de mercado (DJIA). Da mesma forma, os trabalhos de Dzielinski (2012) e Heiberger (2015) encontraram que termos específicos coletados no *Google Trends* podem prever as variáveis de mercado, como retorno, volatilidade e volume. Ainda, Hu *et al.* (2018) demonstraram que o VHP do *Google Trends* melhora a acurácia das estimativas. Todavia, mesmo não é consenso entre os trabalhos quanto ao grau de previsibilidade das variáveis de mercado e o poder de previsão do VHP do *Google Trends*, os trabalhos indicam que pode ser benéfico usar o *Google Trends* para informações.

Vale salientar que a relação explorada por Dzielinski (2012) e Heiberger (2015) é linear, bem como o uso de termos específicos que dificultam uma generalização, como também lembrar que os autores avaliaram o impacto em índices, diferente dos resultados apresentados que avaliam o impacto nas empresas listadas. Já Bollen *et al.* (2016) e Hu *et al.* (2018) combinam metodologias avançadas de *machine learning* e séries temporais.

De outro lado, os resultados do teste de causalidade de Granger parecem discordar com aqueles encontrados por Loughlin e Harnisch (2013), Challet e Ayed (2014) e Lee, Cha e Jang (2017), os quais discordam da qualidade do indicador para modelos de previsão. Especificamente, Loughlin e Harnisch (2013) utilizam dados do *Twitter* ao VHP gerado pelo *Google Trends*, enquanto que Challet e Ayed (2014) consideram o VHP do *Google Trends* instável e igualmente suscetível as notícias. Ressalta-se que, dentre esses estudos, apenas o de Challet e Ayed (2014) usou dados semanais. Ainda, faz-se importante lembrar que as pesquisas focadas no curtíssimo prazo, frequentemente avaliam os preços de abertura e de fechamento.

CONCLUSÕES

O trabalho buscou compreender a demanda de informação do investidor a partir da utilização de informações do *Google Trends* associado ao retorno, a volatilidade e ao volume de negociações semanais, entre 2015 e 2020. Para isso, utilizou-se da modelagem de Painel de Vetores Autorregressivos (PVAR). A análise realizada em desvios padrão apontou à eficiência do mercado, pelo menos, para situações semanais, em que há a possibilidade de o investidor pesquisar e compreender cada nova situação do mercado.

Ainda, o fato de os resultados do Volume Histórico de Pesquisas (VHP) do nome da empresa e do *ticker* apresentarem relação bidirecional de Granger Causalidade com o desvio padrão dos retornos, da volatilidade e do volume negociado, sugere que a demanda de informação do investidor é atendida, em parte, por pesquisas em buscadores como o Google, efeito que também é observado no aumento do volume de negociações, após um choque no volume histórico de pesquisas do *ticker*. Logo, modelos de previsão podem ganhar acurácia incluindo informações de buscadores, embora esse impacto seja, em termos de magnitude, pequeno e de curto prazo, incapaz de gerar elevar os retornos no longo prazo.

Também é de se esperar que empresas com maiores retornos atraiam a atenção do investidor, que passa a acompanhar a empresa por um período superior que o curto prazo, para em um horizonte futuro, tomar a decisão de investimento, situação que amplia a compreensão de como a demanda de informação do investidor é formada. Entretanto, o acompanhamento da empresa não garante, pelo menos no longo prazo, que os retornos sejam maiores, determinando que a HME na versão semi-forte seja indiretamente observada, pelo aumento de negociações sem a devida alteração no retorno. Por outro lado, a maior volatilidade tende a reduzir a atenção do investidor.

Destaca-se que, simplificar a demanda de informação de investidores às duas *proxies*: *ticker* e nome da empresa, per si só é insuficiente e não atende toda a demanda de informação do investidor, sendo que seu impacto ainda é pequeno. Portanto, ao considerar modelos de previsão ainda é importante incorporar informações macro e microeconômicas como também setoriais, regionais e de outros investimentos. Cabe salientar que o trabalho ao incorporar novas informações aos modelos, expande as condições de análise. Todavia, os resultados são para uma proposta específica, ao passo que faz-se importante enriquecer o debate com um conjunto maior de evidências empíricas, as quais englobam a utilização de diferentes frequências de dados, metodologias, amostra de palavras e propostas.

REFERÊNCIAS

- ABRIGO, M. R. M.; LOVE, I. Estimation of Panel Vector Autoregression in Stata: a Package of Programs. **The Stata Journal**, v.16, n.3, p.778-804, 2016.
- ANDREWS, D. W. K; LU, B. Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. **Journal of Econometrics**, v.101, p.123-164. 2001.
- BACHELIER. **Annales scientifiques de l'É.N.S.** 3^e série, tome 17 (1900), p. 21-86. Disponível em: <http://www.numdam.org/item?id=ASENS_1900_3_17__21_0>. Acesso 20 fev. 2020.
- BELO, N. M.; BRASIL, H. G. Assimetria informacional e eficiência semiforte do mercado. **Revista de Administração de Empresas**, v.46, p.48-57. Edição especial, nov/dez. 2006.
- B3, Índice **Ibovespa**. Disponível em <http://www.b3.com.br/pt_br/market-data-e-indices/indices/indices-amplos/ibovespa.htm>. Acesso em: 02 mar 2020.
- BIJL, L, KRINGHAUG, G, MOLNÁR, P, SANDVIK, E. Google searches and stock returns. **International Review of Financial Analysis**, v.45, p.150-156. 2016.
- BOLLEN, J.; MAO, H.; ZENG, X. Twitter mood predicts the stock market. **Journal of Computational Science**, n.2, n.1, p.1-8. 2011.
- CAMPBELL, J, Y. SHILLER, R, J. Stock prices, earnings and expected dividends. **Journal of Finance**, v.18, n.3, p.661-676. 1988.
- CHALLET, D.; AYED, A. Predicting financial markets with Google Trends and not so random keywords. **Quantitative Finance**. 2014. Disponível em: <https://arxiv.org/pdf/1307.4643.pdf>. Acesso em: 01. mar. 2020.

- CHOI, H.; VARIAN, H. R. Predicting the Present with Google Trends. **Economic Record**, v.88, n. 1, p.2-9. 2012.
- DZIELINSKI, M. Measuring economic uncertainty and its impact on the stock Market. **Finance Research Letters**, v.9. n.3 p.167-175. 2012.
- FAMA, E. F. Efficient markets: a review of theory and empirical work. **Journal of finance**. v.25, n.2, p.383-417. 1970.
- FAMA, E. F. Efficient capital markets II. **The Journal of Finance**, v.46, n.5, p.1575-1617, 1991.
- HAUGEN, R. A; JORION, P. **Financial Analysts Journal**, v.52, n.1, p.27-31. 1996.
- HU. H; TANG. L; ZHANG. S; WANG. H. Predicting the direction of stock markets using optimized neural networks with Google Trends. **Neurocomputing**, v.285, n.12, p.188-195. 2018.
- HANSEN, L. P. Large sample properties of generalized method of moments estimators. **The Econometric Society**, v.50, n.4, p.1029-1054. 1982.
- HEIBERGER, R. H. Collective Attention and Stock Prices: Evidence from Google Trends Data on Standard and Poor's 100. **PLoS ONE**, v.10, n.8. 2015.
- JOSEPH, K.; WINTOKI, M. B.; ZHANG, Z. Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. Elsevier: **International Journal of Forecasting**, v.27 n.4. p.1116-1127. 2011.
- LEE, P, S. CHA, M. JANG, H. Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets. **PLoS ONE**. v. 12, n.11. 2017.
- LOUGHLIN, C.; HARNISCH, E. The Viability of StockTwits and Google Trends to Predict the Stock Market. **StockTwits.com**, 2013. Disponível em: <https://stocktwits.com/research/Viability-of-StockTwits-and-Google-Trends-Loughlin_Harnisch.pdf>.
- LOVE, I.; ARISS, R. T. Macro-financial linkages in Egypt: A panel analysis of economic shocks and loan portfolio quality. **International Monetary Fund: Middle East Center for Economics and Finance**. Working Paper n.271, p.158-181. 2013.
- LOVE, I. ZICCHINO, L. Financial development and dynamic investment behavior: Evidence from panel VAR. **The Quarterly Review of Economics and Finance**, v.46, n.2, p.190-210. 2006.

- MASSICOTTE, P. EDDELBUETTEL, D. gtrendsR: Perform and Display Google Trends Queries. R package version 1.4.7. 2020. Disponível em <<https://ithub.com/PMassicotte/gtrendsR>>
- MORETIN, P. **Econometria Financeira – um curso em séries temporais financeiras**. 2. Ed. – São Paulo: Edgar Blucher, 2011.
- MOUSSA, F. DELHOUMI, E. OUDA, O.B. Stock return and volatility reactions information demand and supply. **International Businnes and Finance**, v.9, n.1, p.54-67. 2012.
- NETO, J. W. F. **A hipótese da eficiência de mercado e as finanças comportamentais: evidências empíricas no mercado acionário brasileiro e uma proposta teórica integrativa**. Dissertação (mestrado) – Universidade Federal de Pernambuco. CCSA. Mestrado em Administração. 2006.
- PERLIN, M. **Processamento e Análise de Dados Financeiros e Econômicos com o R**. Self Published. Edição do Kindle. 2018.
- PERLIN, M. BatchGetSymbols: Downloads and Organizes Financial Data for Multiple Tickers. R package version 2.5.7. 2020.
- PERLIN, M. S; CALDEIRA, J.F; SANTOS. A. A. P; PONTUSCHKA. M. Can We Predict the Financial Markets Based on Google's Search Queries? **Journal of Forecasting**, v.36 n.4, p. 454-467. 2016.
- PESARAN, M. H. A simple panel unit root test in the presence of cross-section dependence. **Journal of Applied Econometrics**, v.22, n.2, p.265-312. 2007.
- PREIS, T; MOAT, H. S.; STANLEY, H.S. Quantifying Trading Behavior in Financial Markets Using Google Trends. **Scientific Reports**, v.3, p.1684. 2013.
- RAMOS, H. P.; RIBEIRO.K. K. M.; PERLIN, M. S. The forecasting power of internet search queries in the Brazilian financial Market. **Mackenzie Management Review**, v. 19 n.2. p.184-210. 2017.
- ROBERTS, H. V. **Statistical versus clinical prediction of the stock market**. Unpublished work presented in the Conference of Securities Price Analysis. Chicago, 1967.
- WANG, Y. **Media and google: The impact of supply and demand for information on stock returns**. Available at SSRN 2180409, 2012.
- VLASTAKIS, N.; MARKELLOS, R. N. Information demand and stock market volatility. **Journal of Banking & Finance**, v.36, n.6, p.1808-1821, 2012.
- WOUTERS, T. Style investing: behavioral explanations of stock market anomalies. **University of Groninguen**. s.n., p.227. 2006.