

THE FORECASTING POWER OF INTERNET SEARCH QUERIES IN THE BRAZILIAN FINANCIAL MARKET



HENRIQUE PINTO RAMOS

Master's Degree in Business Administration from the School of Management,
Federal University of Rio Grande do Sul (UFRGS).
PhD student in Business Administration from the School of Management,
Federal University of Rio Grande do Sul (UFRGS).
Rua Washington Luiz, 855, Centro Histórico, Porto Alegre – RS – Brasil – CEP 90010-460
E-mail: henrique.ramos@ufrgs.br

KADJA KATHERINE MENDES RIBEIRO

Master's Degree in Business Administration from the School of Management,
Federal University of Rio Grande do Sul (UFRGS).
11605 Haynes Bridge Road, Alpharetta – Georgia – United States – ZIP CODE 30005
E-mail: kadjamendes@gmail.com

MARCELO SCHERER PERLIN

PhD in Finance from the International Capital Market Centre,
University of Reading.
Assistant Professor at the School of Management,
Federal University of Rio Grande do Sul (UFRGS).
Rua Washington Luiz, 855, Centro Histórico, Porto Alegre – RS – Brasil – CEP 90010-460
E-mail: marcelo.perlin@ufrgs.br

ABSTRACT

Purpose: To analyze the predictability of Google's search queries in the Brazilian financial market.

Originality/gap/relevance/implications: Despite a growing foreign literature using Google's search query data, there is no acknowledgement of work on this area in Brazil. An application to the Brazilian financial market shows new sources of information about market movements and may contribute to researchers and practitioners to understand how changes in specific search queries affect the market.

Key methodological aspects: Following previous studies, we estimate VAR models and Granger causality tests to investigate the effects over three variables in both stock and fixed income markets: traded volume, return and volatility. Following this procedure, we verify both the hypothesis of financial variables being affected by search queries, as well as the opposite relationship. Weekly data from Google's search queries and financial markets was gathered for the period between 2007 and 2014.

Summary of key results: The existence of a predictive effect between search query data and financial variables, particularly in the stock market, is evident. However, this result was not robust in all cases studied. It is noteworthy that, for the inverse relationship, i.e. financial market impacting search queries on Google, strong evidence of a causal relationship has been found. A trading strategy based on this type of data yielded higher returns than the defined benchmarks.

Key considerations/conclusions: A significant relationship between Google's search query data and the financial market has been discovered. Results provide a new source of information that affects the Brazilian financial market.

KEYWORDS

Google Trends. Investor attention. Market efficiency. Market microstructure. VAR Models.

JEL: G10, G11, G15

1. INTRODUCTION

One of the main functions of financial markets is to channel capital into productive activities that demand resources in the economy. This simple transferring function promotes the country's economic development by improving competitiveness and employment (Shiller, 2013). For this resource application to be efficient, both investors as invested agents emit and demand information regarding these investments. One of the main issues in Finance literature is to understand how this information affects stakeholders and, consequently, impacts asset prices set in financial markets (Fama, 1965; 1970). To understand the explaining potential of information has become a challenge for research given the absence of any dataset covering investor's behavior and decision making process.

Given the advances in IT, data acquisition on search query data in popular searching engines, such as Google, has become available. In current days, it is possible to extract a time series of search query data for the terms "*buy stocks*", for example. This is a relevant source of data, since online search may indicate an ongoing situation, event or a population's bias in many fields. For instance, a significant increase in search volume about flu may indicate the occurrence of a disease outbreak in a specific region (Carneiro & Mylonakis, 2009; Polgreen, Chen, Pennock, Nelson, & Weinstein, 2008). Likewise, a raise in online search queries of given types of cars may precede future sales (Kristoufek, 2013).

In Finance, Google's search query data might indicate individual's bias for trading in financial markets or a systemic increase of investor's attention (Da, Engelberg, & Gao, 2011). Both effects can be indicators of future investor's behavior. A rise of searches for the term "*buy stocks*" can be understood as a predictive signal of a systematic incoming of buy orders from investors and a rise in asset prices in the stock market. The use of these signals preceding behavior of financial market is relevant since it may prove useful in constructing portfolios, forecasting financial crisis and, in general, be helpful to understand which factors impact prices of financial contracts.

Within this scenario, this paper aims to understand whether different search terms present forecasting ability over the Brazilian financial market. Following recent literature (Joseph, Wintoki, & Zhang, 2011; Vozlyublenaiia, 2014), search query data from indices names, stock tickers¹ and words

¹ Tickers are codes for identification of stocks and other financial instruments. For example, the preferred stocks of Vale do Rio Doce hold the ticker VALE5.

related to the fixed income market were used. These datasets were used to explain three important variables in Finance: future return, future volatility and future trading volume. Given the dependency of Brazil's financial market with the international one, this work innovates by testing a contagion effect in which Google's searches originated across other countries, different from Brazil, may affect the local market.

The results using Granger causality tests exhibit a predictive relation both in search query volume from Google causing changes in the financial variables chosen as in the opposite relation. Search query data on terms such as "Bovespa" originated both in Brazil as in USA precedes both positive returns in Ibovespa index as a raise in its volatility. Seemingly results are found for the search query data on the ticker for preferred stocks of Petrobras. Although the effects are less robust for the fixed income market, results show that terms related to this market affects the DI Over rate. In this work, we show first evidence that Google search data may predict future returns, changes in volatility and traded volume. A trading strategy using information to forecast Ibovespa index was tested in order to expose the informational potential of this type of data. The strategy has outperformed naive strategies for capital allocation. Based on these results, one can understand how investor's attention affects the financial market and vice-versa. This study brings evidences for classical issues on financial literature such as return predictability and, in a broader sense, the efficient markets hypothesis. Moreover, there is no knowledge of other studies approaching this issue in an emerging market such as Brazil, neither employing the models presented by Perlin, Caldeira, Santos and Pontuschka (2016) and Vozlyublennaia (2014) for the fixed income market².

Aiming a better understanding, the study is organized as follows: current literature related to the subject is presented. Next, the methodology is exposed. Then, we proceed to the results and a trading strategy is developed based on these results. The last section ends the paper with the final considerations.

2. LITERATURE REVIEW

Intending a better understanding of objectives of this paper, the theoretical framework presents studies employing search query data with distinct

² The only study found using Google Trends data for emerging markets is the paper of Carrière-Swallow and Felipe (2013), which analyses car sales behavior in Chile.



applications and papers in which this framework was used in the Finance context.

2.1. Internet searches and miscellaneous applications

Several studies have the objective of forecasting short-term economic indicators based on data from Google Trends. The examples include car sales, unemployment rates, consumer confidence, inflation and disease outbreaks.

The study of Ettredge, Gerdes and Karuga (2005) was pioneer in using Google Trends data to support the analysis and forecasting of macroeconomic datasets. The authors have analyzed unemployment series in the United States and concluded that Google Trends' series are related to unemployment during the sample period (77 weeks). Thereby, it is suggested that this type of input may help predict macroeconomic variables. This study motivated the work of Choi and Varian (2012), who tested in-sample forecasting ability of data originated from Google Trends related to consumption indexes, unemployment insurance benefits and consumer confidence. Applying simple econometric models, the authors show that estimations using Google Trends data outperformed in over twenty percent the predicting ability in comparison to estimations using different datasets. Guzman (2011) uses search query data from Google to predict United States inflation. The model using Google data has presented a lower out-of-sample forecasting error than other indicators used in the literature. Li, Shang, Wang and Ma (2015) employ a MIDAS (Mixed Data Sampling) model to predict Chinese inflation using data combining different search query terms. Following this line of studies, the work of Seabold and Coppola (2015) investigated the possibility of using search query volume to predict food prices and consumer goods price series in Central America. The authors found significant results for the markets in Costa Rica, El Salvador and Honduras.

Other studies have been developed within the objective of testing the predictive capability of Google Trends data towards disease outbreaks. According to Carneiro and Mylonakis (2009), information regarding flu provided by Google Flu Trends platform can detect regional surges faster (up to seven to ten days before) than the Centers for Disease Control and Prevention (CDC). Polgreen *et al.* (2008) used search query volume extracted from Yahoo and showed evidence confirming the possibility of predicting infectious diseases. Ginsberg, Mohebbi, Patel, Brammer, Smolinski and Brilliant (2009) have conducted another similar research using data provided by Google. The study has estimated weekly flu activity in Unites States prece-



ding one up to three weeks in comparison to the benchmarking. The work on the efficiency of detection and previous communication of dengue in endemic countries of Chan, Sahai, Conrad and Brownstein (2011) has concluded that Google Dengue Trends is capable of predicting and tracking dengue fever activity in Brazil, Bolivia, Singapore, India and Indonesia. These evidence shows that despite not being forecasting tools, Google Flu Trends and Google Dengue Trends can be used as surveillance systems that supplies indications for tracking both flu and dengue fever trends in real time. The rapid acquisition of this information can be important, since data provided by official sources may be disclosed within a certain delay.

2.2. Internet searches and Finance

A growth is evident in academic works relating Google Trends data and financial markets (Bijl, Kringhaug, Molnár, & Sandvik, 2016; Da *et al.*, 2011; Da, Engelberg, & Gao, 2015; Vlastakis & Markellos, 2012). By investigating changes in Google's search query volume in terms related to finance, Preis, Moat and Stanley (2013) find patterns that may be read as alert signals from financial transactions in the stock market. During the sample period (2004 to 2011), the results found by the authors show that Google Trends' data not only reflect actual behavior of stock markets but also may anticipate future trends. The study concludes that Google Trends' data can be employed for the construction of profitable trading strategies.

Heiberger (2015) studied the use of Google's search query volume as an indicator of bad news for companies listed on *Standard and Poor's 100* index, which measures the performance of the 100 larger companies in terms of market capitalization in the United States. His results support the use of an investment strategy which exhibits larger returns in times of market turmoil and extensive losses for other market agents. Similar to this work, Vozlyublenniaia (2014) uses Google Trends data as a proxy³ for investor's attention to financial markets indexes and commodities, such as oil and gold. Employing simple VAR (Vector Autoregressive) models, the author's results confirm the hypothesis that investor's attention has predictive relevance in both return and volatility of the indexes analyzed. Presenting similar results, Joseph *et al.* (2011) show evidences that search query data from tickers of 475 American companies has forecasting power over their respective abnormal returns and traded volumes. Kristoufek (2013) builds portfolio

³ A proxy is a variable supposed to have high correlation to a variable of interest that is not possible to measure.

lios with inversely proportional weights to the search query volumes from Google Trends and shows that portfolios formed within this strategy presents lower volatility regarding equally weighted portfolios. In an indirect manner, these studies approach the efficient market hypothesis: if there is any information which may help to predict stock's returns and it is not embedded in asset prices (in our case, the information regarding online search about financial assets), one can consider this fact as a rejection of the efficient market hypothesis.

Linking the use of Google's search query data from flu to financial markets, McTier, Tse and Wald (2013) estimate a set of regressions in order to verify the impact of flu in distinct financial variables from United States. The authors have found that a surge in flu is related to decreasing returns, lower volatility, lower traded volume and an increase in bid-ask spread⁴. Return shrinkage is assigned to a lower liquidity of the assets (measured by traded volume), as being priced with a discount, as well as a decrease in economic activity due to flu effects. The results reported occur with a larger magnitude when data from New York is used, city on which two of the largest stock exchanges in the world are based (NYSE and NASDAQ).

Overall, previous work reported that search query data on financial assets could anticipate stock market movements (Joseph *et al.*, 2011; Vozlyublennaia, 2014). This study contributes to this research field once these evidences are unknown to the Brazilian literature. Besides, there is no knowledge of work relating search data originated from other countries and linking the fixed income market with search query data from Google. Based on the results presented, we intend to take a first step in research in this field of studies in the Brazilian market.

3. METHODOLOGICAL PROCEDURES

This work is clearly of empirical nature since quantitative data obtained from the stock exchange and from the Google Trends platform are used. Moreover, as this is the first study analyzing this issue in Brazil, it can be accounted as an exploratory work. In order to keep the content coherent and easy to follow, the next subsections expose the type of data employed and the econometric models used for estimations.

⁴ This term refers to the difference between the buying price of an asset and its respective selling price.

3.1. Data

Since 2006, Google provides free of charge access to the Google Trends tool⁵. By inserting a term/word for search and a determined geographical location the website supplies information regarding the frequency of queries in the search engine related to determined term or word. If there is sufficient volume of data, the tool provides weekly or monthly information. Data is normalized in such manner to be contained in the interval between 0 and 100. To reach this relative frequency, each nominal value for a specific interval is divided by the maximum value in the same period (Choi & Varian, 2012).

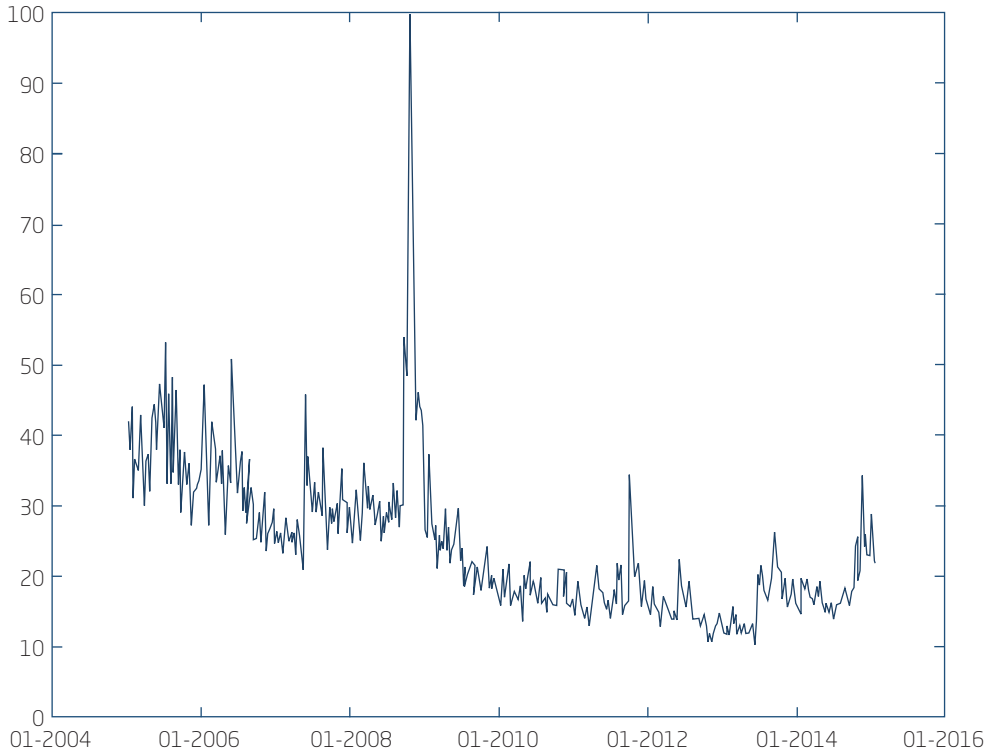
The platform calculates volumes of searches for each word based on all uses of it. For example, in the search query data for the word 'volatility', it will also be included search data for 'volatility stocks', 'what is volatility', among other uses. Concerning the possibility of information generated being redundant, words with a low level of search and repeated queries from same users are not included in the calculation of values provided by the tool. An example of output from Google Trends is presented in Graph 1. A plot referring to level of search queries from the word 'dólar' (Portuguese writing for 'dollar') originated from Brazil shows increases and decreases in the search volume along the series. In this example, there is a higher frequency of search at the end of 2008, specifically at October. This behavior may be associated to the world financial crisis. During this period, there was a rise in the US Dollar/Brazilian Real quotation and a shift to less risky investments, phenomenon known as flight-to-quality (Beber, Brandt, & Kavajecz, 2009). The example elucidates how this type of data can reflect events and real phenomena in financial markets.

In order to achieve the objectives of this paper, information related to weekly search data on Google Trends will be used. All tickers of stocks composing Ibovespa index were searched in Google Trends platform although it was used only data regarding assets in which weekly data was available (PETR4, VALE5 and the word *Bovespa*). It has also been searched terms related to the fixed income market, such as *Selic*, *Taxa CDI*, *Tesouro Direto* and *Renda Fixa*. The choice of representative terms/words for the fixed income market was based in the intensity in which these terms reproduce the related markets. It is understood that terms such as these are important for the investor interested in the market, besides serving as a reference for comparison among alternatives of investments that are not traded in the financial market, such as private projects, fixed assets investment and real options.

⁵ www.Google.com/trends

(Graph 1)

EXAMPLE OF SERIES OUTPUT FROM QUERY ON GTRENDS



Source: Elaborated by the authors.

The relationship between search volume data originating from Google and financial market variables will be tested. For the stock market, the Ibovespa index will be used as proxy. And, for the fixed income market, the DI Over rate provided by CETIP is used as proxy⁶. Similar to choice of terms, the proxy definition has also been due to the representativeness of the market in which they are set on. A widely used rate for the fixed income market is the DI Over rate, while the Ibovespa index is built by the theoretical portfolio composed by stocks with high liquidity and companies with large market value.

⁶ The DI Over rate reports the interest rate of daily average of interbank deposits, disregarding operations between the same financial groups.

Both stock and fixed income markets will have their respective returns, volatilities and traded volumes related to the search query levels from Google. This method was chosen since these variables have already been related to search query data from Google in studies conducted outside Brazil (Arditi, Yechiam, & Zahavi, 2015; Bordino, Battiston, Caldarelli, Cristelli, Ukkonen, & Weber, 2012; Perlin *et al.*, 2016; Vozlyublennaia, 2014). These measures will be used both as dependent variables as regressors in the econometric framework. The variables were firstly regressed against dummies for each month of the year and the respective residuals were used as main variables. The definition of these series is given by the following equations:

$$Volatility_t = \sqrt{\frac{\sum_{j=1}^{nDays_t} (R_i - E(R_i))^2}{nDays_t}} \quad (1)$$

$$Return_t = \frac{\sum_{j=1}^{nDays_t} R_i}{nDays_t} \quad (2)$$

$$Volume_t = \frac{nDays_t^{-1} \sum_{j=1}^{nDays_t} Vol_i}{100.000} \quad (3)$$

The sample period spans from 2007 to 2014. Financial data used was collected in a daily frequency and aggregated in a weekly basis, the frequency of the Google Trends dataset. Equations 1 and 2 present daily log-returns between every day at week t . The sum of these returns is divided by the number of days at week t , which may be different from five due to possible market closings (example: holidays). The volatility is referred to the t week, which is analogous to the measure of return. Average volume is the medium traded volume at week t , scaled by 100,000.

In order to better organize the work, the analysis is split into two market: the Brazilian stock market and the fixed income market. In both markets, search levels for the word “Bovespa” originated from United States and the word “Nasdaq” originated from Brazil were included. These variables aim to test the hypothesis that Brazilian investors could be paying attention to international stock exchanges (such as Nasdaq) and the hypothesis that foreign investors could be monitoring the Brazilian market. As data on search queries involve other terms associated to the word searched, it could occur some type of noise or a potential bias in the values provided by Google.

gle. However, it is believed that such distortions do not present loss in interpretation of results, once the terms and words searched (tickers, name of an index or terms for fixed income) are specific words. Thus, it is expected that investors who might be searching for these words will not be searching for other information rather than regarding to financial markets. In accordance to this view, Joseph *et al.* (2001) allege that the search for a term so specific probably will be linked to a present or future investment decision and not entirely at random. Nevertheless, it is notable to comment that a search for codes such as *PETR4* may not be related to buying or selling decisions, but merely being realized for academic purposes, speculation or, simple curiosity. Even so, these motives still cover financial subjects and investor's attention to the respective stock.

3.2. Estimated models

The purpose of this paper is to quantify the predictive effect of Google Trends' search data over the Brazilian market. In order to attain this, a structural VAR model was estimated to measure not only the impact of online search data on the financial market, but also the inverse effect (Perlin *et al.*, 2016). For each model, the following variables will be used: difference of volatility ($\Delta Volat_t$), returns (R_t), and the difference of traded volume (ΔVol_t) named as γ_t in the equations bellow.

$$y_t = \alpha_1 + \sum_{p=1}^{MaxLag} \beta_p y_{t-p} + \sum_{p=1}^{MaxLag} \lambda_p \Delta Trends_{t-p}^* + \varepsilon_{1,t} \quad (4)$$

$$\Delta Trends_t^* = \alpha_2 + \sum_{p=1}^{MaxLag} \gamma_p \Delta Trends_{t-p}^* + \sum_{p=1}^{MaxLag} \varphi_p y_{t-p} + \varepsilon_{2,t} \quad (5)$$

VAR model estimations allow the researcher to identify the impact of lagged regressors in the current value of vector variables, as well as the inverse relation. Equations 4 and 5 exemplify the case that both lagged financial variables impact their own values at time t as lagged variables referring to Google's search data. The opposite relation is also estimated: it is analyzed how lagged values of Google's search data influences their own current values, as well as the impact from financial variables.

The variable $Trends_t^*$ refers to Google Trends' deseasonalized series. It is defined as the residuals from the regression $GTrends_t = \alpha + \sum_{\kappa=1}^{11} \varphi_{\kappa} D_{\kappa,t} + \varepsilon_t$, which $D_{\kappa,t}$ are dummies assuming value one for each month of the year, excluding January. Lag definition for the VAR models was based on the

sequential Likelihood Ratio test described by Lütkepohl (2005). Finally, Granger causality tests are performed using the models presented so far. This method consists in testing whether the coefficients λ_p at equation 4 and the coefficients ϕ_p at equation 5 are jointly different from zero for each equation. Exemplifying, the test informs if including the variable $\Delta Trends^*$ (see equation 4) with up to $t-p$ lags results in better predictions than estimates excluding this regressor. If these coefficient are jointly different from zero, it is said that the variable $\Delta Trends^*$ Granger-causes y_t . An analogous logic may be used for equation 5. To maintain concision, the list of all estimations is reported in the appendix of this paper.

4. RESULTS

In this section, results are presented for the Granger causality tests based on VAR models introduced previously. Results for the stock markets and for the fixed income market are available. Following Perlin *et al.* (2016) e Vozlyublenniaia (2014), it will be reported only the sum of coefficients λ_p and ϕ_p instead of the specific value for each lag. This is justified, since our concern is about the long run relation between Google's search data and the dependent variables, not the effect of a specific lag. Coefficient significance is calculated based on Granger causality test, as described in the previous section. The list of financial assets and weekly descriptive statistics are detailed on Table 1.

(Table 1)
DESCRIPTIVE STATISTICS

Asset	Mean (weekly)	Median (weekly)	Standard Deviation (weekly)	Mean Traded Volume (weekly)
PETR4	-0.063%	-0.049%	1.148%	R\$ 26.283.650,74
VALE5	-0.033%	0.017%	0.988%	R\$ 16.365.770,30
BVSP	0.005%	0.050%	0.799%	R\$ 10.693.626,22
DI Over	0.000%	0.000%	0.001%	R\$ 3.648.725,55

Source: Elaborated by the authors.

Table 2 exhibits estimations for the stock market. In Panel A, coefficients for traded volume and search levels for the assets are reported. A

positive and significant impact (at a significance level of 5%) is verified on online searches for the Petrobras ticker in its traded volume, revealing that an increase of attention to this asset precedes larger transactions (in volume). Search data for the term *Renda Fixa* Granger-causes an increase in traded volume for the Ibovespa index with a positive and significant coefficient (at 1% level). This result may be associated to a shift between the stock market and the fixed income market. In recent years, Brazilian economic scenario has been marked by a decrease in economic output and an increase in interest rates. Such conditions draw attention to indexed investments, turning investors to migrate their resources to less risky assets. Analyzing the coefficients for the second equation in VAR model, it is noteworthy that a larger traded volume diminishes search for the assets in all estimations.

Table 2 shows results for Granger Causality test estimations from equations (4) and (5) using variables related to the stock market. Sums of coefficients λp e φp are presented, as well as their respective p-value in parenthesis. Symbols ***, ** and * account for p-values significant at 1%, 5% and 10% level for the rejection of the null hypothesis $\lambda_1 = \lambda_2 = \dots = \lambda_p = 0$ ou $\varphi_1 = \varphi_2 = \dots = \varphi_p = 0$. Sample period spans from 2007 to 2014 using weekly data.

(Table 2)

CAUSALITY TESTS FOR GOOGLE TRENDS AND STOCK MARKETS

Panel A - Results for Volume						
Search on GTrends (location)	Asset	Sum λp (Trends → Volume)	MaxLag	Sum φp (Volume → Trends)	MaxLag	# Note
PETR4(BR)	PETR4	1.056** (0.015)	5	-1.348*** (0.000)	5	373
VALE5(BR)	VALE5	0.612 (0.532)	5	-0.543*** (0.001)	5	382
BVSP(BR)	BVSP	-0.648 (0.886)	5	-0.439*** (0.000)	5	417
BVSP (US)	BVSP	1.076 (0.813)	5	-0.692*** (0.000)	5	412
NASDAQ (BR)	BVSP	0.043 (0.695)	5	-0.646*** (0.000)	5	387

(continue)

(Table 2 (continuation))

CAUSALITY TESTS FOR GOOGLE TRENDS AND STOCK MARKETS

Panel A - Results for Volume						
Search on GTrends (location)	Asset	Sum λp (Trends → Volume)	MaxLag	Sum φp (Volume → Trends)	MaxLag	# Note
Taxa CDI (BR)	BVSP	-0.037 (0.595)	5	-0.415*** (0.000)	5	355
Selic (BR)	BVSP	-1.819 (0.130)	5	-1.811*** (0.000)	5	417
Tesouro Direto (BR)	BVSP	-1.106 (0.336)	5	-1.690*** (0.000)	5	417
Renda Fixa (BR)	BVSP	3.175*** (0.000)	5	-1.172*** (0.000)	5	409

Panel B - Results for Return						
Search on GTrends (location)	Asset	Sum λp (Trends → Return)	MaxLag	Sum $\varphi \$/$ (Return → Trends)	MaxLag	# Note
PETR4(BR)	PETR4	0.283 (0.633)	5	0.954*** (0.000)	5	373
VALE5(BR)	VALE5	0.167 (0.956)	5	0.892*** (0.000)	5	382
BVSP(BR)	BVSP (points)	0.030*** (0.000)	5	0.964*** (0.000)	5	417
BVSP (US)	BVSP (points)	0.144*** (0.000)	5	0.960*** (0.000)	5	412
NASDAQ (BR)	BVSP (points)	0.585** (0.012)	5	0.756*** (0.000)	5	387
Taxa CDI (BR)	BVSP (points)	0.003 (0.959)	5	0.836*** (0.000)	5	355
Selic (BR)	BVSP (points)	0.114 (0.389)	5	0.832*** (0.000)	5	417
Tesouro Direto (BR)	BVSP (points)	-0.097 (1.000)	5	0.850*** (0.000)	5	417

(continue)

(Table 2 (conclusion))**CAUSALITY TESTS FOR GOOGLE TRENDS AND STOCK MARKETS**

Panel B - Results for Return						
Search on GTrends (location)	Asset	Sum λp (Trends \rightarrow Return)	MaxLag	Sum ϕS (Return \rightarrow Trends)	MaxLag	# Note
Renda Fixa (BR)	BVSP (points)	0.259 (0.214)	5	0.817*** (0.000)	5	409
Panel C - Results for Volatility						
Search on GTrends (location)	Asset	Sum λp (Trends \rightarrow Volatility)	MaxLag	Sum ϕS (Volatility \rightarrow Trends)	MaxLag	# Note
PETR4(BR)	PETR4	1.525** (0.016)	5	-1.393*** (0.000)	5	373
VALE5(BR)	VALE5	1.609 (0.828)	5	-0.517*** (0.003)	5	382
BVSP(BR)	BVSP (points)	9.403*** (0.000)	5	-0.421*** (0.000)	5	417
BVSP (US)	BVSP (points)	8.343*** (0.000)	5	-0.782*** (0.000)	5	412
NASDAQ (BR)	BVSP (points)	2.968* (0.066)	5	-0.665*** (0.000)	5	387
Taxa CDI (BR)	BVSP (points)	-0.307 (0.246)	5	-0.406*** (0.000)	5	355
Selic (BR)	BVSP (points)	-1.413 (0.382)	5	-1.796*** (0.000)	5	417
Tesouro Direto (BR)	BVSP (points)	-1.788 (0.966)	5	-1.673*** (0.000)	5	417
Renda Fixa (BR)	BVSP (points)	0.708 (0.150)	5	-1.050*** (0.000)	5	409

Source: Elaborated by the authors.

More intuitive relationships are presented for assets returns, Panel B. The information in Table 2 shows that searching for terms such as *Bovespa*, both originated from Brazil as in the United States, precedes positive returns in the Brazilian stock market. This result is possibly connected to a higher attention

from investors to this market. It is hypothesized that searches for information regarding American investors precedes buying action, raising the respective return (Joseph *et al.*, 2011). An analogous logic may be conducted to the search level for Nasdaq Granger-causing an increase in the Brazilian stock market. Other estimations using Google's search levels do not present significant results. By analyzing Brazilian stock returns Granger-causing search for assets in Google, a positive and significant relation (at 1%) is shown. When stock and index returns increase, there is evidence that investors search more information upon these assets, probably driven by news concerning this positive return. This behavior occurs even with terms related to fixed income, what may be explained by investors comparing returns between markets.

The third dependent variable (volatility) is reported in Panel C. It is noteworthy that a higher level of online searches in Google for the Petrobras preferred stock ticker raises volatility of this asset, what is corresponding with previous results in the literature. Similar results occur for the level of search for the Brazilian stock exchange (with origin both in Brazil as in United States) and Ibovespa's volatility. Along with the results of Panel B, it is emphasized the evidence that investors search for information regarding the asset before buying, pushing prices up and raising volatility. Analyzing the second VAR equation, negative coefficients are reported to be Granger-causing assets' online search. In the case of assets presenting a higher risk, search for these stocks and related terms fall, what may be assigned to risk aversion by part of the investors.

Using the DI Over rate as a proxy for the fixed income market, Panel A from Table 3 reports the sum of coefficients for Granger causality tests related for the volume traded in this market. It is notable that the search for the term *Taxa CDI* has a negative impact in interbank deposits contracts volume, while the search for terms such as *Tesouro Direto* presents a positive impact. The latter coefficient may be justified by the rise in popularity of fixed income investments in the last years of sample period: Brazil's main interest rate, SELIC, had its lower value during 2012 reaching 7.15% and rose to 11.65% at the end of 2014, last year of the sample⁷. This growth in interest rates promotes an increase in applications indexed to DI rates and, consequently, to the volume of investments in *Tesouro Direto* platform.

Table 3 shows results for Granger Causality test estimations from equations (4) and (5) using variables related to the fixed income market. Sums of coefficients λp e ϕp are presented as well as their respective p-value in paren-

⁷ Source: Central Bank of Brazil.

thesis. Symbols ***, ** and * account for p-values significant at 1%, 5% and 10% level for the rejection of the null hypothesis $\lambda_1 = \lambda_2 = \dots = \lambda_p = 0$ ou $\varphi_1 = \varphi_2 = \dots = \varphi_p = 0$. Sample period spans from 2007 to 2014 using weekly data.

(Table 3)

CAUSALITY TESTS FOR GOOGLE TRENDS AND FIXED INCOME MARKETS

Panel A - Results for Volume						
Search on GTrends (location)	Asset	Sum λ_p (Trends \rightarrow Volume)	MaxLag	Sum φ_p (Volume \rightarrow Trends)	MaxLag	# Note
BVSP(BR)	DI Over	-0.306 (0.493)	5	-0.451*** (0.000)	5	417
BVSP (US)	DI Over	-0.433 (0.673)	5	-0.699*** (0.000)	5	412
NASDAQ (BR)	DI Over	0.021 (0.840)	5	-0.623*** (0.000)	5	387
Taxa CDI	DI Over	-0.195*** (0.009)	5	-0.412*** (0.000)	5	355
Selic	DI Over	0.253 (0.346)	5	-1.784*** (0.000)	5	417
Tesouro Direto	DI Over	0.388** (0.036)	5	-1.665*** (0.000)	5	417
Renda Fixa	DI Over	-0.009 (0.728)	5	-1.026*** (0.000)	5	409

Panel B - Results for Return						
Search on GTrends (location)	Asset	Sum λ_p (Trends \rightarrow Return)	MaxLag	Sum φ_p (Return \rightarrow Trends)	MaxLag	# Note
BVSP(BR)	DI Over	-0.000 (0.970)	5	0.964*** (0.000)	5	417
BVSP (US)	DI Over	-0.000 (0.765)	5	0.957*** (0.000)	5	412
NASDAQ (BR)	DI Over	-0.000 (0.274)	5	0.761*** (0.000)	5	387

(continue)

(Table 3 (conclusion))

CAUSALITY TESTS FOR GOOGLE TRENDS AND FIXED INCOME MARKETS

Panel B - Results for Return						
Search on GTrends (location)	Asset	Sum λp (Trends \rightarrow Return)	MaxLag	Sum $\phi \delta$ (Return \rightarrow Trends)	MaxLag	# Note
Taxa CDI	DI Over	-0.000*** (0.000)	5	0.833*** (0.000)	5	355
Selic	DI Over	0.000 (0.295)	5	0.833*** (0.000)	5	417
Tesouro Direto	DI Over	-0.000 (0.998)	5	0.851*** (0.000)	5	417
Renda Fixa	DI Over	-0.000 (0.620)	5	0.815*** (0.000)	5	409

Panel C - Results for Volatility						
Search on GTrends (location)	Asset	Sum λp (Trends \rightarrow Volatility)	MaxLag	Sum $\phi \delta$ (Volatility \rightarrow Trends)	MaxLag	# Note
BVSP(BR)	DI Over	-0.001* (0.087)	5	-0.453*** (0.000)	5	417
BVSP (US)	DI Over	-0.001 (0.409)	5	-0.674*** (0.000)	5	412
NASDAQ (BR)	DI Over	-0.000* (0.072)	5	-0.605*** (0.000)	5	387
Taxa CDI	DI Over	0.001** (0.048)	5	-0.349*** (0.000)	5	355
Selic	DI Over	-0.002** (0.015)	5	-1.657*** (0.000)	5	417
Tesouro Direto	DI Over	-0.001 (0.228)	5	-1.657*** (0.000)	5	417
Renda Fixa	DI Over	-0.000** (0.015)	5	-1.003*** (0.000)	5	409

Source: Elaborated by the authors.

In Panel B (returns), the only significant effect reported is for the term *Taxa CDI*, in which an increase in this type of search had a negative impact

in the return of DI Over rate. According to Vozlyublennaia (2014), investor's attention ought to be different for positive or negative changes. Although the negative coefficient seems implausible, this result may be related to the selectivity of attention: by acknowledging negative information regarding CDI, investors should search for information in Google about expectations for the fixed income market. Once the CDI rate is not affected directly by the retail investors or by a specific trader, but is determined by the interbank lending market, the negative coefficient is likely related to previous expectations being realized. In accordance to this, negative coefficients for the second VAR equation (returns Granger-causing search for financial terms) evidences that after positive returns, investor attention for terms related to the stock market and the fixed income market are both increased.

Panel C of Table 3 reports different signaling for terms from online searches affecting the DI Over volatility. At a significance level of 10%, search for *Bovespa* and *Nasdaq* terms, both originated from Brazil, reduces DI rate volatility. This result may be related to investor leaving the fixed income market and entering in the stock market. However, coefficients for terms *Renda Fixa* and *Selic* are negative, while the coefficient related to *Taxa CDI* is positive. In the second VAR equation, it is notable that all terms of online searches are affected negatively by the higher volatility in the DI market, exhibiting coefficients significant at the 1% level.

5. TRADING STRATEGIES

Based on previous results, a trading strategy was built using Google's data related to the stock market. Once the results of Table 2 show a positive and significant time dependency between the returns from the Ibovespa index and the search levels related to some of the terms searched, this information may lead to positive returns through a trading strategy.

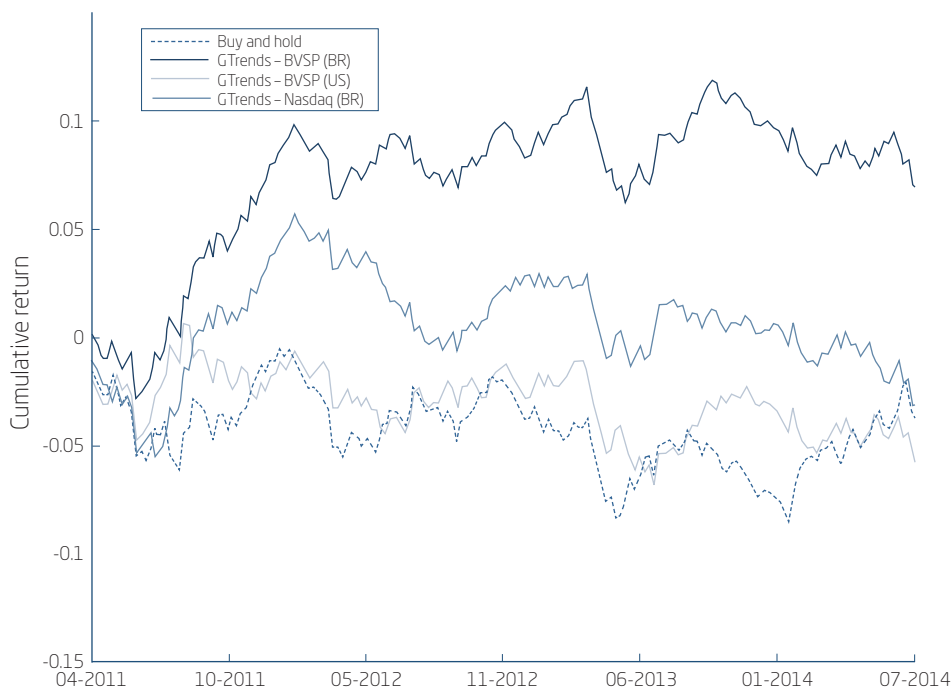
Following Preis *et al.* (2013), a simple market timing strategy was employed: sample period was split in two, one for modeling and other for prediction. In the modeling period (2007 to 2010), the VAR model described in section 4 was estimated for weekly returns from the Ibovespa index. Since the interest remains in the forecasting returns of market indexes, the coefficients for the first VAR equation, which estimates the weekly returns as a function of online searches, are used. Thus, for each week the coefficients are used to forecast the Ibovespa index in the period $t+l$. If the forecasting value is positive, a long position in the index is simulated and, in

case it is negative, a short position is simulated. The total return of this strategy, its volatility and respective Sharpe Ratio are evaluated for the weekly predictions over the 2011-2014 time period.

Graph 2 exhibits cumulative returns for this strategy using three distinct predictors of Ibovespa’s behavior in comparison to a buy and hold strategy, in which the investor holds the assets in his portfolio since the beginning of the period until the end. In our strategy, forecasts were made using the search level for terms *Bovespa* both with origin in Brazil as in United States and the word *Nasdaq* originating from Brazil. These variables are reported to have positive and significant coefficients at the estimation of VAR models presented in Table 2. It is noteworthy that the trading rule using search level of *Bovespa* with origin in Brazil has offered cumulative profits superior to other strategies.

(Graph 2)

CUMULATIVE RETURN OF TRADING STRATEGIES



Source: Elaborated by the authors.

In order to the detail efficiency of this procedure, metrics from the buy and hold strategy and a strategy in which the trading signal was randomly

generated from a uniform distribution with values between -1 and 1 were compared. When a positive value is simulated, a long position for the asset is registered. The same logic applies for short positions. Using a strategy where the trading signal is generated based on a uniform distribution will assure the operation to be at random or, without any reason based in the Finance context. This procedure was repeated 10,000 times and the means are reported in Table 4 together with other statistics from trading strategies.

Table 4 exhibits results for different investment strategies for the Ibovespa index during the forecasting period from 2011 to 2014. A comparison between strategies was conducted using a naive strategy (buy and hold), a strategy which the buying signal was given by simulations of an uniform distribution with values between -1 and 1 and, a strategy in which the buying or selling of the index is originated from predictions derived from equation (4).

TABLE 4
RESULT FOR TRADING STRATEGY

	Buy and Hold	Uniform	Bovespa (BR)	Bovespa (US)	Nasdaq (BR)
Total Return	-6.925%	-0.105%	5.232%	-6.508%	-7.831%
Volatility	0.614%	0.613%	0.614%	0.614%	0.614%
Sharpe Ratio	-5.372%	-0.082%	4.056%	-5.048%	-6.077%
Modeled Note	200	200	200	200	200
Predicted Note	210	210	210	210	210

Source: Elaborated by the authors.

Only the strategy using predictions based on search data for the term *Bovespa* originating from Brazil yields a positive return, outperforming other strategies. A Sharpe Ratio higher than zero reflects this. The results are consistent with Graph 2, which shows the strategy based on searches over *Bovespa* (BR) exhibiting superior performance over other strategies. As the strategies trade every week in which data from Ibovespa is available, returns only vary in signal based on a long or short position.

6. FINAL CONSIDERATIONS

The amount of studies analyzing the impact of search volume data in financial markets are growing (Choi & Varian, 2012; Da *et al.*, 2011, 2015;

Joseph *et al.*, 2011; Vozlyublennai, 2014). In this first study using Brazilian data, methodology employed in international literature was adjusted to the Brazilian case. Our results are similar to previous work (Joseph *et al.*, 2011; Vozlyublennai, 2014), in which search level for companies' tickers and market indexes affect financial variables such as return, volatility and traded volume in the stock market.

In this paper, evidence was found that the Ibovespa index is impacted by the respective search level both in Brazil as in United States. The reverse relation is also significant: changes in financial variables affect the level of searches for these assets, showing that investor's attention is drawn according to variations in return, volatility and traded volume.

The results evidence that yes, it is possible to explain the Brazilian financial market based on search data from Google. Results show this effect to be stronger in the stock market. Although traders and market makers should use sophisticated platforms to acquire information, it is believed that local and foreign retail investors may use Google to obtain information, primarily in a buying decision (Joseph *et al.*, 2011). This fact illustrates an opportunity for price forecasting. Section 5 shows a trading strategy based on the search data for the word *Bovespa* originating from Brazil as outperforming naive strategies. With this, we explore the applicability of studies in this strand.

Finally, it is necessary to account for the fact that not all the results in Granger Causality tests have shown significance for all tickers and indexes analyzed, what brings the need of later studies, mainly in the theoretical area. Thus, the refinement of trading strategies based on data from Google Trends is also suggested, in order to account for market frictions such as trading costs.

O PODER PREDITIVO DE PESQUISAS NA INTERNET SOBRE O MERCADO FINANCEIRO BRASILEIRO

RESUMO

Objetivo: Analisar a capacidade preditiva de pesquisas no Google sobre o mercado financeiro brasileiro.

Originalidade/lacuna/relevância/implicações: Apesar de uma crescente literatura estrangeira utilizando dados sobre pesquisas oriundas no Google, não se tem conhecimento de trabalhos desta natureza no Brasil. A aplicação no mercado financeiro evidencia novas fontes de informação

acerca do movimento dos mercados e pode contribuir para pesquisadores e praticantes compreenderem esta dinâmica.

Principais aspectos metodológicos: Foram estimados testes de Causalidade de Granger para investigar os efeitos em três variáveis dos mercados de renda acionário e de renda fixa: volume, retorno e volatilidade. Testam-se as hipóteses de que tanto o nível de pesquisas afeta as três variáveis financeiras quanto a relação contrária. Foram usados dados semanais de pesquisas do Google Trends e dos mercados financeiros entre o período de 2007 a 2014.

Síntese dos principais resultados: Evidencia-se a existência de um efeito preditivo entre os níveis de pesquisas e as variáveis financeiras, principalmente no mercado de renda variável. Todavia, este resultado não foi robusto em todos os casos analisados. Destaca-se que, para a relação inversa, isto é, o mercado financeiro impactando o nível de pesquisas no Google, encontrou-se forte evidência de uma relação causal. O uso de uma estratégia de negociação baseada neste tipo de dados gerou retornos maiores do que os *benchmarks* definidos.

Principais considerações/conclusões: O estudo revelou uma relação significativa entre o nível de pesquisas no Google e o mercado financeiro. Os resultados oferecem uma nova fonte de informação que afeta o mercado financeiro do Brasil.

PALAVRAS-CHAVE

Google Trends. Atenção do investidor. Eficiência de mercado. Microestrutura de mercado. Modelos VAR.

JEL: G10, G11, G15

EL PODER PREDICTIVO DE LAS CONSULTAS DE BÚSQUEDA EN INTERNET SOBRE LO MERCADO FINANCIERO DE BRASIL

RESUMEN

Objetivo: El análisis de la capacidad predictiva de las búsquedas de Google en el mercado financiero brasileño.

Originalidad/laguna/relevancia/implicaciones: A pesar de una extensa literatura internacional en la investigación utilizando datos procedentes de Google, en Brasil no se tiene conocimiento de estudios de esta naturaleza. La aplicación muestra nuevas fuentes de información sobre el movimiento de los mercados y puede contribuir a profesionales comprender mejor esta dinámica.

Principales aspectos metodológicos: Utilizando testes de causalidad de Granger se investigaron los efectos de tres variables de los mercados de valores y de renta fija: volumen, rentabilidad y volatilidad. De este modo, las hipótesis se prueban que tanto el nivel de la investigación afecta a las tres variables financieras como la relación opuesta. Fue utilizados datos semanales de las encuestas de Google Trends y los mercados financieros durante 2007-2014.

Síntesis de los principales resultados: La existencia de un efecto predictivo entre los niveles de investigación y las variables financieras, en particular en el mercado de valores es evidente. Pero este resultado no era robusto en todos los casos analizados. Es de destacarse, para la relación inversa, los mercados financieros impactando búsquedas en Google, hemos encontrado una fuerte evidencia de relación causal. Una estrategia de negociación basada en este tipo de datos genera una mayor rentabilidad que benchmarkings definidos.

Principales consideraciones/conclusiones: El estudio encontró una relación significativa entre el nivel de investigación en Google y en el mercado financiero. Los resultados proporcionan una nueva fuente de información que afecta al mercado de Brasil.

PALABRAS CLAVE

Google Trends. Atención de los inversores. Eficiencia del mercado. Micro estructura del mercado. Modelos VAR.

JEL: G10, G11, G15

REFERENCES

Arditi, E., Yechiam, E., & Zahavi, G. (2015). Association between stock market gains and losses and Google searches. *Plos One*, 10(10): e0141354. DOI: <http://dx.doi.org/10.1371/journal.pone.0141354>.

- Beber, A., Brandt, M. W., & Kavajecz, K. A. (2009). Flight-to-quality or flight-to-liquidity? Evidence from the euro-area bond market. *Review of Financial Studies*, 22(3), 925-957.
- Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis*, 45(5), 150-156.
- Bordino, I., Battiston, S., Caldarelli, G., Cristelli, M., Ukkonen, A., & Weber, I. (2012). Web search queries can predict stock market volumes. *Plos One*, 7(7): e40014. DOI: <http://dx.doi.org/10.1371/journal.pone.0040014>.
- Carneiro, H. A., & Mylonakis, E. (2009). Google trends: a web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases*, 49(10), 1557-1564.
- Carrière-Swallow, Y., & Felipe, L. (2013). Nowcasting with Google trends in an emerging market. *Journal of Forecasting*, 32(4), 289-298.
- Chan, E. H., Sahai, V., Conrad, C., & Brownstein, J. S. (2011). Using web search query data to monitor dengue epidemics: a new model for neglected tropical disease surveillance. *Plos Negl Trop Dis*, 5(5): e1206. DOI: <http://dx.doi.org/10.1371/journal.pntd.0001206>.
- Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *Economic Record*, 88(s1), 2-9.
- Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The Journal of Finance*, 66(5), p. 1461-1499.
- Da, Z., Engelberg, J., & Gao, P. (2015). The sum of all fears investor sentiment and asset prices. *Review of Financial Studies*, 28(1), 1-32.
- Ettredge, M., Gerdes, J., & Karuga, G. (2005). Using web-based search data to predict macroeconomic statistics. *Communications of the ACM*, 48(11), 87-92.
- Fama, E. F. (1965). The behavior of stock-market prices. *Journal of business*, 38(1), 34-105.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383-417.
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014.
- Guzman, G. (2011). Internet search behavior as an economic forecasting tool: The case of inflation expectations. *The Journal of Economic and Social Measurement*, 36(3), 119-167.

- Heiberger, R. H. (2015). Collective attention and stock prices: evidence from Google trends data on standard and poor's 100. *Plos One*, 10(8): e0135311. DOI: <http://dx.doi.org/10.1371/journal.pone.0135311>.
- Joseph, K., Wintoki, M. B., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: evidence from online search. *International Journal of Forecasting*, 27(4), 1116-1127.
- Kristoufek, L. (2013). Can Google trends search queries contribute to risk diversification? *Scientific Reports*, 3, Article 2713. DOI: <http://dx.doi.org/10.1038/srep02713>.
- Li, X., Shang, W., Wang, S., & Ma, J. (2015). A midas modeling framework for Chinese inflation index forecast incorporating Google search data. *Electronic Commerce Research and Applications*, 14(2), 112-125.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. New York: Springer Science & Business Media.
- McTier, B. C., Tse, Y., & Wald, J. K. (2013). Do stock markets catch the flu?. *Journal of Financial and Quantitative Analysis*, 48(03), 979-1000.
- Perlin, M. S., Caldeira, J. F., Santos, A. A. P., & Pontuschka, M. (2016). Can we predict the financial markets based on Google's search queries?. *Journal of Forecasting*, 35(7), 592-612. DOI: <http://dx.doi.org/10.1002/for.2446>.
- Polgreen, P. M., Chen, Y., Pennock, D. M., Nelson, F. D., & Weinstein, R. A. (2008). Using internet searches for influenza surveillance. *Clinical Infectious Diseases*, 47(11), 1443-1448.
- Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying trading behavior in financial markets using Google trends. *Scientific Reports*, 3, Article 1684. DOI: <http://dx.doi.org/10.1038/srep01684>.
- Seabold, S., & Coppola, A. (2015). Nowcasting prices using Google trends: an application to Central America. *World Bank Policy Research Working Paper*, 1(1), Report 7398.
- Shiller, R. J. (2013). *Finance and the good society*. Princeton: Princeton University Press.
- Vlastakis, N., & Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance*, 36(6), 1808-1821.
- Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking & Finance*, 41(C), 17-35.

(APPENDIX I)

(Table 5)

LIST OF ESTIMATIONS FOR VAR MODELS

Stock Market	
Search on GTrends (location)	Asset
PETR4 (BR)	PETR4
VALE5(BR)	VALE5
BVSP(BR)	BVSP
BVSP (US)	BVSP
NASDAQ (BR)	BVSP
Taxa CDI	BVSP
Selic	BVSP
Tesouro Direto	BVSP
Renda Fixa	BVSP
Fixed Income Market	
Search on GTrends (location)	Asset
BVSP(BR)	DI Over
BVSP (US)	DI Over
NASDAQ (BR)	DI Over
Taxa CDI	DI Over
Selic	DI Over
Tesouro Direto	DI Over
Renda Fixa	DI Over

Source: Elaborated by the authors