

Print ISSN: 2288-4637 / Online ISSN 2288-4645
doi:10.13106/jafeb.2018.vol5.no4.45

Search-based Sentiment and Stock Market Reactions: An Empirical Evidence in Vietnam*

Du D. Nguyen¹, Minh C. Pham²

Received: August 18, 2018 Revised: August 23, 2018 Accepted: September 30, 2018

Abstract

The paper aims to examine relationships between search-based sentiment and stock market reactions in Vietnam. This study constructs an internet search-based measure of sentiment and examines its relationship with Vietnamese stock market returns. The sentiment index is derived from Google Trends' Search Volume Index of financial and economic terms that Vietnamese searched from January 2011 to June 2018. Consistent with prediction from sentiment theories, the study documents significant short-term reversals across three major stock indices. The difference from previous literature is that Vietnam stock market absorbs the contemporaneous decline slower while the subsequent rebound happens within a day. The results of the study suggest that the sentiment-induced effect is mainly driven by pessimism. On the other hand, optimistic investors seem to delay in taking their investment action until the market corrects. The study proposes a unified explanation for our findings based on the overreaction hypothesis of the bearish group and the strategic delay of the optimistic group. The findings of the study contribute to the behavioral finance strand that studies the role of sentiment in emerging financial markets, where noise traders and limits to arbitrage are more obvious. They also encourage the continuous application of search data to explore other investor behaviors in securities markets.

Keywords: Behavioral Finance, Investor Sentiment, Google Search Volume, Emerging Market, Vietnam.

JEL Classification Code: G10, G12, G41.

1. Introduction

The traditional assumption of rationality has failed to explain many anomalies observed in financial markets such as the excess volatility of asset prices and the Mehra-

Prescott equity risk premium puzzle. It is then natural that theoretical models introduce less rational agents (noise traders) into the markets to help clarify the cases. De Long, Shleifer, Summers, and Waldmann (1990)'s model provides theoretical framework to understand how noise traders can push prices away from fundamental values and earn better returns than their rational counterpart. The effect is more likely clear when rational investors face limits to arbitrage. Subsequent empirical studies have relied on this ground to document the significant relationship between market sentiment and short-term swings as well as excessive volatility. This well acknowledged role of market sentiment has triggered a literature strand that focuses on measures of market sentiment, one of which is the internet search behavior of households.

Since the measure has multiple advantages compared to previous measures of market sentiment, a large number of studies has employed it to attest the predictions of sentiment theories. A remarkable example is the FEARS index (Da, Engelberg, & Gao, 2015), which is built upon the Google search volume of negative economic terms by US

* We thank Joseph Engelberg at Univeristy of California at San Diego for providing the detailed construction of the FEARS index. We also thank our colleagues at Hanoi University for valuable comments. We are grateful to Nadav Rabani at Google for providing access to Google Trends, Brian Reis at NYU Volatility Lab for providing the HNX volatility index, and Linh Vu at Hanoi University for great research assistance. Remaining errors and omissions are our own.

1 First Author and Corresponding Author. Lecturer, Faculty of Management and Tourism, Hanoi University, Vietnam [Postal Address: Room 205, Building C, Km 9, Nguyen Trai Road, Thanh Xuan District, Hanoi, Vietnam, 120000] Tel. (+84) (0)936 3212 79, E-mail: dund@hanu.edu.vn

2 Consultant, Kapital AMC Consulting JSC., Vietnam, E-mail: phamcongminh@winvestor.vn

households. The FEARS index is powerful in explaining short-term reversal of returns, temporary market volatility and flows between mutual funds.

Such asset pricing implications are desirable, and they need empirical confirmation across other markets, especially in developing markets like Vietnam. This is because emerging markets are often characterized by a large number of unsophisticated investors who are more prone to irrationality (De Long et al., 1990). Consequently, their sentiment is expected to have major role in explaining the great mispricing in these stock markets. Historical performance of VN index, the largest stock index in Vietnam, exemplifies this view. Starting with a base index of 100 in 2000, it skyrocketed to over 1100 in 2007. A significant crash happened when it plummeted to just about 310 in 2008 and made a record low of 235.5 in 2009. In just two recent months between April and May of 2018, VN index lost more than 18 percent when it fell from its record high of 1204.33. Yet research on the relationship between search-based sentiment and stock market movement has been rare in developing markets for two reasons. The first is the time-consuming nature of collecting data from Google Trends. Because Google only allows the public to download daily data at quarter intervals once at a time, one need to have access to Google Trends to manipulate programming codes for quicker download. Second, the high-frequency nature of sentiment effect creates pressure to have control variables at comparable frequency, which are difficult to obtain in practice.

We overcome the aforementioned obstacles to construct different sentiment indices in Vietnam using the Google Trends' Search Volume Index (SVI). To the best of our knowledge, ours is the first attempt to examine the relationship between search-based sentiment measures and stock market reactions in Vietnam at daily frequency. The use of internet search volume is also our attempt to attest the power of this sentiment proxy in a developing market, where the short-selling regulation imposes a critical limit to arbitrage. Our study differs mainly from that of Da et al. (2015) in which we use both negative and positive economic search terms to address two different aspects of market sentiment: pessimism and optimism. We expect to answer two research questions. The first is whether a search-based sentiment index can have the asset pricing implications for Vietnam stock market, particularly the short-term reversal prediction. The second question is which aspect of sentiment, pessimism or optimism, drives the observed results. We also attempt to provide an explanation for our findings.

Overall, we find that our sentiment index can explain significantly the temporary decrease and the subsequent reversal of returns across three major stock indices in

Vietnam. Further analysis suggests that most of the effect is driven by pessimism, as measured by our separate search-based pessimism index. The optimism index can predict significantly the reversal, which we interpret as the strategic delay from the optimistic investors. We also document the slow-adjusting effect that sentimental and pessimistic investors experience and the delayed response of optimistic group. These findings shed light on understanding the simultaneous reactions of two investor groups and provide avenues for further studies using the internet search data.

2. Literature Review

2.1. Measures of Market Sentiment

While the fact that investor sentiment affects market is widely recognized, the question that how to measure and quantify its effects is debatable (Baker & Wurgler, 2006). Academics often use five types of measures: financial market-based measures, survey-based indices, relevant sources of textual data, non-economic factors, and internet search behavior of individuals and households. To understand while we prefer the last type, we review the weaknesses of the first four ones.

The financial market-based measures rely on frequent market trading data to derive investor sentiment. Close-ended fund discounts is the most commonly used measure among them (Baker & Wurgler, 2006; Lee, Shleifer, & Thaler, 1991; Zweig, 1973). The reason is that since individual investors are primary traders of the close-ended fund shares, the observed discounts are interpreted as negative sentiment factor. Another popular used measure is market liquidity. Baker and Stein (2004) propose a theoretical model that includes a class of irrational investors whose underreacted trading leads to unusually high liquidity and overvalued market. Subsequent empirical studies (Lei, 2005; Liu, 2015) employ different measures of liquidity, such as turnover and dollar volume, to gauge US market sentiment. Other market-based measures are IPO first-day returns and volume, and volatility indices such as the VIX (Baker & Wurgler, 2006; Ljungqvist, Nanda, & Singh, 2006). High one-day returns associated with IPOs are good measurement of investor attention, while volatility indices are considered the market gauges of investor pessimism. The advantage of using market-based measures is their high-frequency availability. However, we cannot ascertain that market-based measures are driven by investor sentiment. Da et al. (2015) assert that they are the equilibrium results of many economic forces and sometimes they work as cycles, rendering them inaccurate measures of sentiment.

The second approach to extract sentiment is survey-based indices. Survey data of consumer confidence are normally employed because of their popularity. Qiu and Welch (2004) use the Michigan Consumer Confidence Index and the UBS/Gallup index of investor optimism to measure US market sentiment. Tran and Nguyen (2013) also use a consumer confidence index in constructing a sentiment measure in Norway and Vietnam. Their study follows Baker and Wurgler (2006)'s in which the sentiment measure is a composite index using both market- and survey-based factors. Brown and Cliff (2004) use survey data from American Association and Investor Intelligence to measure individual and institutional investor sentiment. They believe that survey questions make direct inquiry about individual real attitudes, hence, a better measure of sentiment. However, respondents to survey sensitive questions are not all careful and honest (Singer & Ye, 2013). We also believe that the framing and representativeness biases (Kahneman & Riepe, 1998) make individual response to survey questions inconsistent, thus, less reliable.

The third approach is to analyze relevant sources of textual data to identify existence of sentiment. Because quantitative fundamentals have failed to explain significant movements in stock prices, scholars have turned to qualitative verbal information as a potential explanatory factor. Tetlock (2007) and Tetlock, Saar-Tsechansky, and Macskassy (2008) suggest that negative words written in financial and economic articles by newspapers and media can identify subsequent investor sentiment. Recent research (Ahern & Sosyura, 2015; Dougal, Engelberg, García, & Parsons, 2012) show that sensational financial news and newspaper articles can easily drive stock market price, especially if the news come from reliable sources. Although this method to measure market sentiment is potential, it is often time-consuming due to the wide spanned sources of textual information.

The fourth type of sentiment measure is non-economic factors. Justification for their use is based on the link between investor mood and single events or continuous variables that affect mood. Hirshleifer and Shumway (2003) use morning sunshine based on the psychological association between sunny weather and positive mood. Edmans, García, and Norli (2007) study international soccer results because of their large mood swings on general population. While these psychology-based measures work logically as sentiment indicators, attempts to link them with stock market movements may fall in the path to "dredge for anomalies" (Fama, 1998).

The final approach is to utilize individual and household Internet search. The task focuses on capturing people sentiment by looking at their internet search behavior. We believe this method is superior in several ways. First,

individuals only search if they are genuinely concerned about financial and economic circumstances. A few would search out of their pure interests. In other words, individuals search in an active manner and this behavior reveals their real sentiment. Unlike financial-based measures, the internet search behavior more likely separates individual sentiment and other economic factors. Second, search-based measures have external verification whereas the survey-based ones has not (Da et al., 2015). The advantage lies in their anonymity during the search. This contrasts to the passive and exposed features of the survey-based measures. Third, giant search engines such as Google help mitigate the data collecting and processing burden and cover all financial and economic event search. This practice allows scholars to avoid "dredging for anomalies" (Fama, 1998), where one run through certain event types until significant results appear.

Empirically most economists utilize the Search Volume Index (SVI) provided by Google product, Google Trends. It shows the search history volume of terms scaled by a scalar. Da, Engelberg, and Gao (2011) and Da et al. (2015) use them to construct sentiment indices that measure investor attention and pessimism in the US. Schaul (2013) build a similar index for Dutch market. Habibah, Rajput, and Sadhwani (2017) use both Google SVI and VIX as sentiment index and fear gauge to predict the movement of S&P500. With its advantages emphasized, this paper will use the SVI at daily frequency as a sentiment indicator in Vietnam market.

2.2. SVI and Asset Returns

Researchers have utilized the SVI in recent years mostly to capture individual attention and predict market movements. Da et al. (2011) show that the SVI measures retail investor attention in a timely manner than other measures. In a sample of US Russell 3000 stocks from 2004 to 2008, an increase in the SVI can predict increase in stock prices in the 2-week period and their subsequent reversal. They continue their study by constructing an index named FEARS (Financial and Economic Attitudes Revealed by Search), an index that we construct similarly in this study, to gauge the sentiment in the US stock market and test its relationship with different asset movements (Da et al., 2015). The FEARS index is a composite of searched financial and economic terms that have negative sense or feeling. At daily frequency, increase in the FEARS index is associated with immediate but temporary downward stock returns. The authors find that nearly half of the decrease reverses within the next 2 days. The results are similar for other assets, including different equity portfolios and Treasury securities. Controlling for the possibility of liquidity shocks followed by

major economic events, their results imply that the temporary mispricing is induced by investor sentiment. Our study is most related to theirs but we will highlight differences below.

Most other notable and similar results are obtained in the US market. Joseph, Babajide, Wintoki, and Zhang (2011) rely on weekly online ticker searches to predict abnormal stock returns and trading volumes of US S&P500 companies in the 2005-2008 period. Not only the searches predict reliably both variables, they are also more sensitive and positively related to difficult-to-arbitrage stock returns. Bijl, Kringhaug, Molnár, and Sandvik (2016) employ similar framework and find the negative relationship between the searches and stock returns over their 2008-2013 sample. Their test suggests that an investment strategy based on selling stocks with high search volume and buying infrequent searched stocks is profitable only if transaction costs are not considered. A concern is that both studies use weekly search volume, which does not allow us to examine sudden within-days reaction that one normally observes in the market.

In contrast, using a sample of 189 Japanese companies searched between 2008 and 2011, Takeda and Wakao (2014) only find significant positive correlation between search activity and trading volume but not stock prices. It is worth to note that their sample period contains the financial crisis, so any interpretations must be treated careful. At global level, Chen (2017) use SVI to construct a measure of investor attention toward global benchmark indices. They find a significant negative relationship between the measure and the index returns across countries.

While similar studies have been conducted in emerging markets, most do not use search-based measures, particularly SVI, as the sentiment indicator. Tran and Nguyen (2013) study the predictive power of weekly sentiment measure on Vietnam individual stock returns but use a composite index of financial- and survey-based measures as in Baker and Wurgler (2006). Apart the weaknesses of these measures that we mention, they include VIX, the US S&P100 index option volatility, to their sentiment index. We question its relevance to measure investor sentiment in Vietnam. Unsurprisingly their robustness test suggests that the sentiment indices are not sensitive to VIX for the Vietnam market. We cover these limitations by using SVI as the sentiment measure. Our analysis is also conducted at daily frequency, which will better capture within-days swings of market returns.

We acknowledge that Tran and Nguyen (2013) study the topic at individual firm level and ours at market level. However, our objectives are to focus on testing the short-term reversal hypothesis and identify which aspect of sentiment is more dominant at market level. This is in part

more aligned with Da et al. (2015). Our study differs in two ways. First, while they discard terms or words that have too many gapped missing values (less than 1000 observations), we keep them. This conservative treatment ensures that we are not biased towards the significant search terms. Second, instead of only using terms that have historical negative relationship with market returns, we also use terms that have positive relationship to construct sentiment index. This practice provides further insights into the separate effect that pessimism and optimism have on the market movements.

3. Data and Methodology

3.1. Data

Vietnam Sentiment Index

Since Vietnam sentiment index is the key variable in our study, we take time to explain its construction in detail. We first create a list of economic words that are considered to reveal sentiment towards economic conditions. This 149-word list is taken from Da et al. (2015), which followed the Harvard IV-4 Dictionary and the Lasswell Value Dictionary of words classified as economic and either positive or negative (Tetlock, 2007). Examples are “inflation”, “gold”, “recession”, “poor”. We translate all words into Vietnamese. To address concern that whether the English terms have the same power in Vietnamese, we sample-search a glossary of Vietnamese terms that have similar economic meanings on Google. Results suggest that there exists a homogeneity between two lists where Vietnamese terms such as “gia vang” or “lam phat” return significant search during economic sensitive periods. For further assurance, we input each word into Google Trend, which in turn produces a list of total 1,091 Top-10 related words to the original 149 words. After discarding duplicate words, we have a list of 927 words. For example, the original word “no” has “no xau”, “no cong”, “tra no” as related words that Vietnamese searched the most. We continue by removing non-economic words (e.g. “khung hoang tuoi len 3” (meaning: crisis when reaching 3 years old) is not relevant to the search term “khung hoang” (meaning: crisis)), which results in a list of 179 words. Since Vietnamese has punctuation, we add another set of 179 terms with full punctuation (e.g. “lam phat viet nam” is accompanied by “lạm phát việt nam”, both mean “Vietnam inflation”), making the final set of 358 terms.

We collect from Google Trends daily SVI of all 358 terms for our sample period (01/01/2011 - 06/15/2018). We restrict the search results to Vietnam only. Since we are interested in changes in the term search volume, the daily log SVI change of term j at day t is calculated as:

$$\Delta SVI_{j,t} = \ln(SVI_{j,t}) - \ln(SVI_{j,t-1})$$

Following Da et al. (2015) we adjust the daily log SVI change for outlier effects, seasonality and heteroskedasticity. We perform winsorization at 5% level (2.5% in each tail) for each term. To address seasonality, we regress the winsorized daily log SVI change of each term on dummy variables of weekdays and months. The residuals are kept as the deseasonalized daily SVI change. We standardize each time series following Baker and Wurgler (2006). These adjustments produce the winsorized, deseasonalized and standardized daily log SVI change ($\Delta ASVI$ or $\Delta Adjusted SVI$).

To select terms to be included in the sentiment index, we run expanding rolling backward regressions every month of $\Delta ASVI$ on VN index returns. The results give us closer look at the past relationship between search terms and market returns. Although Tetlock (2007) argue that negative terms in language is more associated with sentiment, our rolling backward regressions suggest the existence of both negative and positive terms. Hence, we pick 20 terms that have the largest negative and 20 terms that have the largest positive t-statistic from the regressions for the sentiment index, i.e. a total of 40 terms are select. The index is computed as follow:

$$Sentiment_t = \sum_{i=1}^{40} \frac{(\Delta ASVI_t)^i}{40}$$

where $(\Delta ASVI_t)^i$ is the $\Delta ASVI_t$ for the search term that has the negative (positive) t-statistic ranked i^{th} , from smallest to largest (largest to smallest) among the terms. To be clear, we start by running the regression from January 1, 2011 to June 30, 2011 for each term. We then rank the t-statistic on the slope coefficients from the most negative to the most positive. We select the 20 most negative t-statistic terms and 20 most positive t-statistic terms to create the sentiment index from July 1, 2011 to July 31, 2011. The sentiment value for day t in this period is the average $\Delta ASVI_t$ of these 40 terms. The next 40 terms are drawn from regressions from January 1, 2011 to July 31, 2011 to compute the sentiment values from August 1, 2011 to August 31, 2011, and so on.

Other Data

We collect daily close prices and trading volume of three major Vietnam stock indices: VN index (Ho Chi Minh Stock Exchange), HNX index (Hanoi Stock Exchange), and Real Estate index. We then calculate log return on each series and use them as dependent variables in our first regression

equation. The augmented Dickey-Fuller test confirms that the time series are stationary.

We continue by describing other independent variables. The first one is market volatility (*annualizedVol*). We obtain a volatility index forecast by a GARCH model on HNX index from the Volatility Lab at New York University's Stern School of Business. Since the model produces daily variance of the index, we use the following formula to calculate annualized volatility:

$$annualizedVol = \sqrt{252 \times daily_variance}$$

We follow Aruoba, Diebold, and Scotti (2009) to construct an index to measure daily change in Vietnam business and macroeconomic conditions. To the best of our knowledge, ours is the first attempt to combine mixed frequency macroeconomic indicators to derive an economic index at daily frequency in Vietnam. We collect data on Vietnam quarterly GDP growth rate, monthly inflation rates, monthly industry production growth rates, and daily VN index returns. The data source is General Statistics Office of Vietnam. We detrend the data to ensure stationarity and employ the dynamic factor framework set by Aruoba et al. (2009) to create the Vietnam business condition index (*vbcChange*). The change in the index is expected to measure innovations following macroeconomic and business movements. If the index increases (decreases), we should expect a better (worse)-than-average macroeconomic condition.

3.2. Methodology

To examine the relationship between market returns and the sentiment index, we run the OLS regressions:

$$Ret_{i,t+k} = \beta_0^S + \beta_1^S Sentiment_t + \sum_m \gamma_m^S ControlVar_{i,t}^m + u_{i,t+k}^S \quad Eq. \quad (1)$$

where $Ret_{i,t+k}$ denotes the log return of asset i at time $t+k$. Our asset i include VN index, HNX index, and Real Estate index. The control variables include lags of different orders (up to 3) of asset returns *annualizedVol* and *vbcChange*. The "S" superscript is short for "Sentiment" since we will refer this equation as the sentiment model.

The sentiment index is the key explanatory variable in our model. We expect to obtain significant temporary (time t) relationship between the index and market returns and subsequent adjustments (time $t+k$). Our k extends to as long as 3 days. Given the dominance of previously found pessimistic search terms, we expect the coefficient sign of β_1^S is negative and subsequent signs would be positive.

Since our sentiment index is built upon both pessimistic and optimistic terms, a natural treatment is to separate and test the impact of each group of terms on market returns. We run two similar equations to equation (1) for this purpose:

$$Ret_{i,t+k} = \beta_0^p + \beta_1^p Pessimism_t + \sum_m \gamma_m^p ControlVar_{i,t}^m + u_{i,t+k}^p Eq. \quad (2)$$

$$Ret_{i,t+k} = \beta_0^o + \beta_1^o Optimism_t + \sum_m \gamma_m^o ControlVar_{i,t}^m + u_{i,t+k}^o Eq. \quad (3)$$

Regressions (2) and (3) use a pessimism index (*Pessimism*) and an optimism index (*Optimism*), respectively. The “P” and “O” superscripts are short for “Pessimism” and “Optimism”. We calculate these indices in similar way to the sentiment index, except that we only use 20 most negative (positive) t-statistic terms for the corresponding index. We are interested in the coefficient signs of β_1^p and β_1^o .

4. Results and Discussion

4.1. Sentiment-related Search Terms

The most concerned variable in our study is the sentiment index and the related pessimism and optimism indices. Since the sentiment is revealed by search terms, we highlight terms that exhibit previous significant relationship with VN index returns over our entire sample period in Table 1. For brevity, we show only terms that are significant at 5 percent level and lower.

Table 1: Significant t-statistic search terms (Jan 1, 2011 - Jun 15, 2018)

No.	Pessimistic terms (VN)	In English	Optimistic terms (VN)	In English
1	vang	gold	tai san tai chinh	financial assets
2	no cong	public debt	đầu tư chứng khoán	securities investment
3	nợ công	public debt	công ty chứng khoán	securities companies
4	công ty thành công	successful companies	bang chung khoan	securities table
5	cách kiếm tiền	how to make money	chi phi	expense
6	tro cap that nghiep	unemployment benefits	dau tu chung khoan	securities investment
7	vàng	gold		
8	lợi nhuận	profits		
9	kiềm chế lạm phát	controlling inflation		

We notice that the pessimistic terms tend to associate with macroeconomic concerns, such as: “vang” (gold), “nợ công” (public debt) or “kiềm chế lạm phát” (controlling inflation) while the optimistic terms are linked to security markets (“đầu tư chứng khoán” (securities investment), “tai san tai chinh” (financial assets)). It is intuitive to think that people who were already participants of the security markets searched out of negative concerns about the macroeconomy. For example, gold is commonly seen as a safe asset in times of economic turbulence and the pessimistic group might consider “flight to safety” under their bearish view. On the other hand, it appears that people who wanted to join the markets searched out of interests in financial investments. The optimistic group might be willing to shift their investment to the stock market; hence they initiate searches on the stock market information. We provide further analysis on these claims in later sections. Another finding is that the number of significant pessimistic terms dominates in all of our rolling periods. This is aligned with Tetlock (2007)’s claim that negative words might be better at capturing sentiment.

4.2. Sentiment Indices and Asset Returns

Market Sentiment Index and Average Asset Returns

We first run the regression (1) to examine the relationship between the search-based sentiment index and the asset returns. The base sentiment index is constructed using 20 most negative and 20 most positive t-statistic terms. Table 2 (Panel A) presents the regression results between VN index returns and the sentiment index with other control variables.

Table 2: Sentiment Index and Asset Returns

Panel A: Sentiment Index and VN Index Returns

	VN Index Returns				
	(1)	(2)	(3)	(4)	(5)
	Ret(t)	Ret(t+1)	Ret[t, t+1]	Ret(t+2)	Ret(t+3)
Sentiment	-1.936e-03	-2.495e-03	-4.561e-03**	3.766e-03**	-2.010e-03
annualizedVol	-4.347e-05	4.592e-05	-3.770e-07	4.559e-05	5.306e-05
vbcChange	7.179e-05	2.021e-04*	2.783e-04*	-7.778e-05	3.146e-04***
Ret(t)		6.166e-02**		7.076e-02***	4.636e-02
Ret(t-1)	6.141e-02**	6.648e-02***	1.318e-01***	4.246e-02*	1.067e-02
Ret(t-2)	6.016e-02**	4.008e-02*	1.038e-01***	5.075e-03	-1.083e-02
Ret(t-3)	3.650e-02	1.802e-03	4.075e-02	-1.698e-02	-7.275e-03
Intercept	1.238e-03*	-4.384e-04	8.774e-04	-4.187e-04	-5.033e-04
Observations	1732	1731	1731	1731	1729
Adjusted R ²	0.0092	0.01124	0.01583	0.007685	0.006046

Panel B: Sentiment Index and HNX Index Returns

	HNX Index Returns			
	Ret(t)	Ret(t+1)	Ret[t, t+1]	Ret(t+2)
Sentiment	-1.025e-03	-2.655e-03	-3.688e-03	4.654e-03***
Control Variables	Yes	Yes	Yes	Yes
Observations	1732	1731	1731	1730
Adjusted R ²	0.01024	0.009136	0.0109	0.01075

Panel C: Sentiment Index and Real Estate Index Returns

	Real Estate Index Returns			
	Ret(t)	Ret(t+1)	Ret[t, t+1]	Ret(t+2)
Sentiment	-2.814e-03	-3.150e-03	-6.215e-03**	6.729e-03***
Control Variables	Yes	Yes	Yes	Yes
Observations	1601	1600	1600	1599
Adjusted R ²	0.01534	0.01319	0.01506	0.009518

Note: *, **, and *** denote significance at 10%, 5%, and 1% level, respectively.

When $k=0$ and $k=1$, the coefficients on *Sentiment* are negative, suggesting that there is a negative relationship between investor sentiment and VN index returns. However, they are not statistically significant. The contemporaneous relationship turns significant (at 5 percent level) when we regress *Sentiment* on the 2-day cumulative returns $Ret[t, t+1]$. Controlling for different lags of VN index returns, *annualizedVol* and *vbcChange*, the model implies that a one standard deviation increase in *Sentiment* (0.1678) predicts a temporary 7 basis-point decrease (0.1678*-0.00456) for the daily VN index returns 2 days following the search. The decline reverses in day $t+2$ as shown by the coefficient on column (4). The significant (at 5 percent level)

and positive coefficient suggests that a one standard deviation increase in *Sentiment* predicts an increase of 6 basis points for the index returns 2 days after the search. The relationship is not significant when we extend k further. In other words, much of the sentiment-induced effect happens within 3 days. It is not a surprise that most of the temporary decrease (7 basis points) reverses (6 basis points) since traditional sentiment theories emphasize the correction following market overreaction. Because our sentiment index is constructed using both negative and positive t-statistic search terms, the result also supports our hypothesis that the negative search reaction dominates that of positive search.

Panels B and C presents the results for HNX index and real estate index returns, respectively. For brevity, we do not report the full results as in the VN index. However, the pattern is similar, except that the sentiment coefficient is not significant on the first 2-day cumulative returns of HNX index. The results show that a one standard deviation increase in *Sentiment* predict nearly 8 basis-point increase for the HNX index returns 2 days after the search. For the real estate index, the effect is a decline of 10 basis points on the 2-day cumulative returns followed by a reversal of 11 basis points. The reversals are significant at 1 percent level. Another implication is that the real estate market, which we use the real estate index as proxy, experiences the most volatility under market sentiment.

Robustness Checks

Following Da et al. (2015), we examine the robustness of our results which are subject to the choice of index construction and additional control variables. We present the results in Table 3. The first concern is the number of search terms to build the sentiment index. A criticism is that increasing the number of terms (i.e. including more terms with less significant t-statistic) could alter our results. Panel A of Table 3 reports the estimation results when we construct the sentiment index using 25 most negative and positive t-statistic terms (i.e. 5 more terms for each group and the total number of search terms used is 50). We obtain the similar pattern in Table 2: a significant temporary decrease in 2-day cumulative returns ($Ret[t, t+1]$) for all 3 stock indices followed by a significant reversal in day $t+2$.

Another concern with our results is the controversial effect of liquidity shock. Campbell, Grossman, and Wang (1993) model shows that liquidity shock can cause short-term reversals in stock returns. In particular, an increase in the expected return is more likely to be associated with a decline in stock price in high-trading volume day than that on low-trading volume day. Empirically, Baker and Wurgler (2006) and Baker and Stein (2004) emphasize the link between sentiment and liquidity by constructing sentiment index using liquidity measures. These prior arguments indicate that liquidity shocks triggered by investor sentiment

could drive our results. To relax the concern, we include a liquidity shock dummy variable into the regression equation. We first compute three 30-day moving average series of trading volume, one for each stock index. We then define three dummy variables: *liquidityShock* equals 1 if the trading volume in day t of an index is greater than its 30-day moving average. The variable is included into the corresponding index regression. Panel B show the regression results and they are consistent with the results from the sentiment model.

Another argument is that extreme market returns could lead to our results because of their reversal nature. Therefore, our *Sentiment* could simply represent the effect from these sources. We address this concern by including in the regressions the decile dummies for the market returns. The dummies are constructed in similar way to those defined in the liquidity shock test, which means we assign dummy variables to each stock index and run corresponding regressions. We report the results in Panel C and find similar pattern as observed in our sentiment model.

We examine the reversal pattern with the holiday effect in Panel D. We collect all holidays over our sample period (including New Year, Lunar New Year and other national holidays) and define a dummy variable that equals 1 for the day before and after the holidays. We find that controlling for the holiday effect does not alter our results.

Finally, we address a concern regarding the real-time predictive power of our models. This stems from the fact that Google Trends does not provide the search volume until one day after the search. For example, the search volume on Monday, June 4, 2018 is not available until sometime at June 5 evening. However, our results are still robust when we use the models to predict the effect of *Sentiment* on day $t+2$ returns. Particularly, we use day- t *Sentiment*, which is observable at late day $t+1$, to predict the open-to-close returns on day $t+2$. Panel E reports the results from these predictive regressions. The coefficients are still positive and significant, albeit smaller in magnitude compared to that of the sentiment models.

Table 3: Robustness Checks

Panel A: Sentiment Index with 50 Total Search Terms

	VN Index Returns		HNX Index Returns		Real Estate Index Returns	
	Ret[t,t+1]	Ret(t+2)	Ret[t, t+1]	Ret(t+2)	Ret[t, t+1]	Ret(t+2)
Sentiment	-5.543e-03**	3.451e-03**	-4.746e-03*	4.394e-03**	-8.052e-03**	6.534e-03***
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1731	1730	1731	1730	1600	1599
Adjusted R ²	0.01631	0.006612	0.01137	0.009597	0.01595	0.007986

Panel B: Liquidity Shock

	VN Index Returns		HNX Index Returns		Real Estate Index Returns	
	Ret[t, t+1]	Ret(t+2)	Ret[t, t+1]	Ret(t+2)	Ret[t, t+1]	Ret(t+2)
Sentiment	-4.439e-03*	3.960e-03**	-3.369e-03	4.861e-03***	-5.552e-03*	6.916e-03***
liquidityShock	2.737e-04	6.336e-04	2.747e-03***	1.292e-03**	-2.225e-03**	-1.509e-03**
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1704	1703	1704	1703	1574	1573
Adjusted R ²	0.01554	0.008534	0.01739	0.01398	0.01573	0.01119

Panel C: Decile Dummies of Market Returns

	VN Index Returns		HNX Index Returns		Real Estate Index Returns	
	Ret[t, t+1]	Ret(t+2)	Ret[t, t+1]	Ret(t+2)	Ret[t, t+1]	Ret(t+2)
Sentiment	-3.238e-03**	3.915e-03**	-2.151e-03	5.040e-03***	-3.654e-03*	6.819e-03***
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1731	1730	1731	1730	1600	1599
Adjusted R ²	0.5148	0.009695	0.4495	0.01082	0.5022	0.01063

Panel D: Holiday Effect

	VN Index Returns		HNX Index Returns		Real Estate Index Returns	
	Ret[t, t+1]	Ret(t+2)	Ret[t, t+1]	Ret(t+2)	Ret[t, t+1]	Ret(t+2)
Sentiment	-4.484e-03**	3.676e-03**	-3.721e-03	4.514e-03***	-5.855e-03*	6.580e-03***
Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1704	1703	1704	1703	1574	1573
Adjusted R ²	0.01532	0.007268	0.01033	0.01049	0.01505	0.009128

Panel E: Real-Time Predictability

	VN Index Returns	HNX Index Returns	Real Estate Index Returns
	Ret(t+2)	Ret(t+2)	Ret(t+2)
Sentiment	3.048e-03**	4.379e-03**	5.391e-03***
Control Variables	Yes	Yes	Yes
Observations	1730	1730	1599
Adjusted R ²	0.006963	0.008043	0.01042

Note: *, **, and *** denote significance at 10%, 5%, and 1% level, respectively.

Overall, our robustness checks provide strong support for the short-term reversal of returns under the effect of investor sentiment. This is consistent with empirical findings in the US market. Yet the US study does not address the question that what aspect of sentiment drives the results. We seek the answer in the next section.

4.3. Market Pessimism Index and Average Asset Returns

We construct the pessimism index using 20 most negative t-statistic search terms and run the regression (2). Table 4 show the results which we present in a similar way to Table 2.

Table 4: Pessimism Index and Stock Market Returns

Panel A: Pessimism Index and VN Index Returns

	VN Index Returns				
	(1)	(2)	(3)	(4)	(5)
	Ret(t)	Ret(t+1)	Ret[t, t+1]	Ret(t+2)	Ret(t+3)
Pessimism	-1.766e-03	-1.563e-03	-3.448e-03**	1.684e-03	-8.580e-04
Control Variables	Yes	Yes	Yes	Yes	Yes
Observations	1733	1732	1732	1731	1729
Adjusted R ²	0.009667	0.01081	0.01592	0.005523	0.0054

Panel B: Pessimism Index and HNX Index Returns

	HNX Index Returns			
	Ret(t)	Ret(t+1)	Ret[t, t+1]	Ret(t+2)
Pessimism	-1.014e-03	-2.328e-03*	-3.345e-03*	2.722e-
Control Variables	Yes	Yes	Yes	Yes
Observations	1732	1731	1731	1730
Adjusted R ²	0.0104	0.009675	0.01157	0.009158

Panel C: Pessimism Index and Real Estate Index Returns

	Real Estate Index Returns			
	Ret(t)	Ret(t+1)	Ret[t, t+1]	Ret(t+2)
Pessimism	-3.664e-03**	-6.400e-04	-4.648e-03**	3.154e-03**
Control Variables	Yes	Yes	Yes	Yes
Observations	1601	1600	1600	1599
Adjusted R ²	0.01752	0.01188	0.01498	0.005538

Note: *, **, and *** denote significance at 10%, 5%, and 1% level, respectively.

Column (3) in Panel A suggests that there is a significant (at 5 percent level) and negative relationship between *Pessimism* and the 2-day cumulative VN index returns. The magnitude is also similar to what we obtain in the sentiment regression: a one standard deviation increase in *Pessimism* predicts a decline of about 7 basis points in the daily VN index return. The coefficient in day $t+2$ turns positive but not significant. For the HNX and the real estate indices, the pattern is clearer and significant: a one standard deviation increase in *Pessimism* corresponds to about 7 and 10 basis-point decreases, respectively. The magnitude resembles that in the sentiment models. On day $t+2$, the significant reversal occurs with an increase of 6 and 7 basis points, respectively. The reversal magnitude is smaller than that compared to what observed in the sentiment model.

Combined with the results from the sentiment model, these findings show that the first 2-day decline in daily returns is mainly driven by the search-based pessimism. It is also evident that short-term reversal exists. In this aspect, our results are consistent with what found in the US market. The difference is that the VN market experience the spreading 2-day effect of pessimism (or FEARS as in Da et al., 2015) and reversal takes place in one day while there is only 1-day decrease followed by subsequent 2-day rebound in the US market.

Since the results resemble closely what we obtain in the sentiment model, we hazard to state that pessimism is the major sentiment that drives Vietnam stock market. This stands contrast to the wide claim that optimism often creates more mispricing than pessimism in times of high sentiment (Stambaugh, Yu, & Yuan, 2012). This conflict opens a natural question about the role of optimism, which we explore in the next section.

4.4. Market Optimism Index and Average Asset Returns

Table 5 reports the results when we run the regression (3) with the optimism index constructed using 20 most positive t-statistic search terms. Across three stock indices, *Optimism* is significantly associated with increase in daily returns at day $t+2$. In particular, a one standard deviation increase in *Optimism* predicts an increase of 5 basis points for VN and HNX indices and 9 basis points for the real estate index. In short, it magnifies the reversal in the stock market on day $t+2$, apart from the similar effect predicted by the pessimism.

Table 5: Optimism Index and Stock Market Returns

Panel A: Optimism Index and VN Index Returns

	VN Index Returns				
	Ret(t)	Ret(t+1)	Ret[t, t+1]	Ret(t+2)	Ret(t+3)
Optimism	-3.583e-04	-1.152e-03	-1.535e-03	2.385e-03**	-1.312e-03
Control Variables	Yes	Yes	Yes	Yes	Yes
Observations	1733	1732	1732	1731	1729
Adjusted R ²	0.008365	0.01035	0.01399	0.006847	0.005853

Panel B: Optimism Index and HNX Index Returns

	HNX Index Returns			
	Ret(t)	Ret(t+1)	Ret[t, t+1]	Ret(t+2)
Optimism	-1.133e-04	-5.831e-04	-7.025e-04	2.331e-03*
Control Variables	Yes	Yes	Yes	Yes
Observations	1732	1731	1731	1730
Adjusted R ²	0.01004	0.007891	0.009623	0.008525

Panel C: Optimism Index and Real Estate Index Returns

	Real Estate Index Returns			
	Ret(t)	Ret(t+1)	Ret[t, t+1]	Ret(t+2)
Optimism	3.820e-04	-2.753e-03	-2.326e-03	4.233e-03***
Control Variables	Yes	Yes	Yes	Yes
Observations	1601	1600	1600	1599
Adjusted R ²	0.01425	0.01382	0.01323	0.007899

Note: *, **, and *** denote significance at 10%, 5%, and 1% level, respectively.

We propose an explanation to unify our findings. Since it is not usual that a person would perform both pessimistic and optimistic search on the same day, there must be two investor groups who search in opposite aspects of sentiment simultaneously. The Vietnamese meaning of the significant t-statistic search terms in Table 1 adds support to this assumption. The pessimistic group, which is arguably

retail investors, overreacts to their market views but later reverses their action. This process is consistent with the prediction of sentiment theories that noise traders can create short-term mispricing. The delayed response of the optimistic group can stem from their hesitance of market uncertainty, so they rather wait to realize market reaction before taking investment action. When the market rebounds, they join in by trading in the same direction. This pattern lends support to the model implications of Chamley (2004) and Chamley and Gale (1994) in which a strategic delay will lead to some form of herding in a short-run investment game. It is possible that the optimistic group follows the impressive rebound after witnessing the negative performance in the previous 2 days. Our results do not support the claim that optimism often creates excessive mispricing in periods of high sentiment (Stambaugh et al., 2012) since there is no statistical evidence that optimism affects the market returns on day t and $t+1$. This is the foundation for our initial claim that the optimistic group could possibly be to-be-investors, who search with an intention to participate in securities markets.

5. Conclusions

Our study sets out to achieve two objectives. The first is to construct an internet search-based measure of sentiment for Vietnam stock market and quantify its effect on the market. The second is to examine what effect that two aspects of sentiment, namely pessimism and optimism, have on the market.

We use Google Trends to collect the volume of financial and economic related terms that Vietnamese searched from January 2011 to June 2018 and build a sentiment index. Consistent with the prediction from sentiment theories, our sentiment index can explain significantly short-term reversal of returns across Vietnamese major stock indices. In particular, we document the gradual and negative reaction under sentiment effect followed by a 1-day rebound. Closer examination implies that the pattern is mainly driven by the pessimism, which we measure using the search-based pessimism index. On the other hand, the optimism index can explain only the later reversal.

Our findings add empirical support to previous literature that documents the overreaction, and subsequently the correction, of bearish investors in developed markets. Another interesting empirical finding is the slow response of investors who search from the optimistic angle. While we rely on the strategic delay to explain this finding, we cannot leave out other possibilities, especially when taking into consideration the differences between developed and developing markets. Further research can look at this

shortage and the simultaneous interlink between the two investor groups to understand their behavior better. Our research also shows that analysis of internet search can help reveal the behavior of market participants in developing markets. Future research can explore it for predictive purposes by using other search engine data.

References

- Ahern, K. R., & Sosyura, D. (2015). Rumor Has It: Sensationalism in Financial Media. *The Review of Financial Studies*, 28(7), 2050-2093. <https://doi.org/10.1093/rfs/hhv006>
- Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real Time Measurement of Business Conditions. *Journal of Business and Economic Statistics*, 27(4), 417-427.
- Baker, M., & Stein, J. C. (2004). Market liquidity as a sentiment indicator. *Journal of Financial Markets*, 7(3), 271-299. <https://doi.org/10.1016/j.finmar.2003.11.005>
- Baker, M., & Wurgler, J. (2006). Investor Sentiment and The Cross-Section of Stock Returns. *Journal of Finance*, 61(4), 1645-1680.
- Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis*, 45, 150-156. <https://doi.org/10.1016/j.irfa.2016.03.015>
- Brown, G. W., & Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of Empirical Finance*, 11(1), 1-27. <https://doi.org/10.1016/j.jempfin.2002.12.001>
- Campbell, J. Y., Grossman, S. J., & Wang, J. (1993). Trading Volume and Serial Correlation in Stock Returns. *The Quarterly Journal of Economics*, 108(4), 905-939. <https://doi.org/10.2307/2118454>
- Chamley, C. (2004). Delays and equilibria with large and small information in social learning. *European Economic Review*, 48(3), 477-501. [https://doi.org/10.1016/S0014-2921\(02\)00315-X](https://doi.org/10.1016/S0014-2921(02)00315-X)
- Chamley, C., & Gale, D. (1994). Information Revelation and Strategic Delay in a Model of Investment. *Econometrica*, 62(5), 1065-1085. <https://doi.org/10.2307/2951507>
- Chen, T. (2017). Investor Attention and Global Stock Returns. *Journal of Behavioral Finance*, 18(3), 358-372. <https://doi.org/10.1080/15427560.2017.1331235>
- Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *The Journal of Finance*, 66(5), 1461-1499.
- Da, Z., Engelberg, J., & Gao, P. (2015). The Sum of All FEARS: Investment Sentiment and Asset Prices. *The Review of Financial Studies*, 28(1), 1-32.
- De Long, J. B., Shleifer, A., Summers, L. H., & Waldmann, R. J. (1990). Noise Trader Risk in Financial Markets. *Journal of Political Economy*, 98(4), 703-738.

- Dougal, C., Engelberg, J., García, D., & Parsons, C. A. (2012). Journalists and the Stock Market. *The Review of Financial Studies*, 25(3), 639-679. <https://doi.org/10.1093/rfs/hhr133>
- Edmans, A., García, D., & Norli, Ø. (2007). Sports Sentiment and Stock Returns. *The Journal of Finance*, 62(4), 1967-1998. <https://doi.org/10.1111/j.1540-6261.2007.01262.x>
- Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3), 283-306. [https://doi.org/10.1016/S0304-405X\(98\)00026-9](https://doi.org/10.1016/S0304-405X(98)00026-9)
- Habibah, U., Rajput, S., & Sadhwani, R. (2017). Stock market return predictability: Google pessimistic sentiments versus fear gauge. *Cogent Economics & Finance*, 5(1), 1390897. <https://doi.org/10.1080/23322039.2017.1390897>
- Hirshleifer, D., & Shumway, T. (2003). Good Day Sunshine: Stock Returns and the Weather. *The Journal of Finance*, 58(3), 1009-1032.
- Joseph, K., Babajide Wintoki, M., & 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. <https://doi.org/10.1016/j.ijforecast.2010.11.001>
- Kahneman, D., & Riepe, M. W. (1998). Aspects of Investor Psychology. *Journal of Portfolio Management*, 24(4), 52-65.
- Lee, C. M. C., Shleifer, A., & Thaler, R. H. (1991). Investor Sentiment and the Closed-End Fund Puzzle. *The Journal of Finance*, 46(1), 75-109. <https://doi.org/10.1111/j.1540-6261.1991.tb03746.x>
- Lei, Y.-C. (2005). *The trading volume trend, investor sentiment, and stock returns*. Louisiana State University, USA: Doctoral dissertation. Retrieved from https://digitalcommons.lsu.edu/cgi/viewcontent.cgi?article=3011&context=gradschool_dissertations
- Liu, S. (2015). Investor Sentiment and Stock Market Liquidity. *Journal of Behavioral Finance*, 16(1), 51-67. <https://doi.org/10.1080/15427560.2015.1000334>
- Ljungqvist, A., Nanda, V., & Singh, R. (2006). Hot Markets, Investor Sentiment, and IPO Pricing. *The Journal of Business*, 79(4), 1667-1702. <https://doi.org/10.1086/503644>
- Qiu, L., & Welch, I. (2004). *Investor Sentiment Measures* (Working Paper No. 10794). National Bureau of Economic Research. <https://doi.org/10.3386/w10794>
- Schau, K. (2013). *Investor Sentiment: The Impact of a Search-Based Measure of Sentiment on Equity Market Returns*. Tillburg, Netherlands: Master thesis. Retrieved from <http://arno.uvt.nl/show.cgi?fid=131016>
- Singer, E., & Ye, C. (2013). The Use and Effects of Incentives in Surveys. *The ANNALS of the American Academy of Political and Social Science*, 645(1), 112-141. <https://doi.org/10.1177/0002716212458082>
- Stambaugh, R. F., Yu, J., & Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2), 288-302. <https://doi.org/10.1016/j.jfineco.2011.12.001>
- Takeda, F., & Wakao, T. (2014). Google search intensity and its relationship with returns and trading volume of Japanese stocks. *Pacific-Basin Finance Journal*, 27, 1-18. <https://doi.org/10.1016/j.pacfin.2014.01.003>
- Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, 62(3), 1139-1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>
- Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance*, 63(3), 1437-1467. <https://doi.org/10.1111/j.1540-6261.2008.01362.x>
- Tran, T. U. T., & Nguyen, N. Y. T. (2013). *The effect of investor sentiment on stocks returns in Norway and Vietnam*. BI Norwegian Business School, Norway: Master thesis. Retrieved from <https://brage.bibsys.no/xmlui/bitstream/handle/11250/95109/Oppgave9.pdf>
- Zweig, M. E. (1973). An Investor Expectations Stock Price Predictive Model Using Closed-End Fund Premiums. *The Journal of Finance*, 28(1), 67-78. <https://doi.org/10.1111/j.1540-6261.1973.tb01346.x>