Advances in Economics, Business and Management Research, volume 192
Proceedings of the Seventh Padang International Conference On Economics Education, Economics, Business and Management, Accounting and Entrepreneurship (PICEEBA 2021)

Google Search, News, and Stock Market Return in Indonesian Stock Market

Aditya Firmansyah[1,*] Zaafri Ananto Husodo[1]

[1] *Universitas Indonesia*
[*] *Corresponding author. Email:j.aditya.f@gmail.com*

ABSTRACT
This study investigates the effect of internet search and news headline on explaining abnormal return of 425 non-financial companies listed on Indonesia Stock Exchange from 2015 to 2019. Investor attention is measured by internet search volume from Google Trend. Market news is gathered from free online news publisher, Bisnis Indonesia. Panel data regression results show that while the probability is not high, daily internet search and news can consistently explain today's abnormal return and can be associated with higher abnormal return; this effect can be better observed on smaller companies and when company's ticker code is not ambiguous.

Keywords: *Asset Pricing, Investor Attention, Google Search, News Headline, Abnormal Stock Return.*

1. INTRODUCTION

Information is an important resource because it is the basis of decision-making process. In stock market, within Efficient Market Hypothesis framework [7] assumes investors gather correct information immediately and quickly affect stock price. However, active attention is an exhaustible resource that is used to process information [10]. Since not all information are able to be processed, it is important to be aware of what information is consumed by investors that affect actions and ultimately will be reflected in stock price.

Attention-induced price pressure hypothesis [2] explains that before buying a stock, investors first need to be aware of the stock. If there is an attention-grabbing stock, investors will be more likely to consider and purchase it over stocks that are unknown to them. This behavior generates pressure to attention-grabbing stocks that temporarily drives the price. Investors do not see similar behavior on selling because generally investors can only sell stocks they already own, especially on market where short selling is prohibited.

There are several proxies for investor attention shown by [2] such as extreme one-day return, high abnormal trading volume, and news. However, those variables are all indirect proxies of investor attention. In this era of internet and abundant information, [5] shows that internet search frequency is a direct measure of attention that is able to promptly catch investor attention.

Attention-driven purchases are more likely to come from less sophisticated investors, as professional investors usually spend more time and resources to monitor wider range of stocks [2]. This is in line with internet search proxy that is more likely to catch attention of less sophisticated retail investors that utilize free resources such as Google search engine [5].

Internet search volume has been used previously in various research around the world with inconclusive result. In the United States stock market, [3] and [5] show positive and negative short-term effect; in Norwegian stock market, [11] does not find relation between internet search and stock return but find increased trading activity on weeks with high attention; [14] finds positive association of internet search and stock return in Japanese stock market; [12] finds negative impact of internet search to stock return in Philippines, Thailand, and Vietnam, positive in Malaysia, and insignificant in Indonesia during 2009 – 2016 period.

The inconclusive result of how GSV affects stock return might be related to investor demographic and information shown to the investors as a result of their search. As seen on Figure 1, when investors search for a publicly traded company stock code, Google will return the background of the company, latest price, and possibly top news headlines. The existence of news has shown to affect stock price [2], [5]. In Indonesia, [6] shows how domestic investors generate bigger profits than foreign investors in intramonth and intraday trading by taking advantages of short-term information; perhaps caused by physical, linguistic, or cultural distance.

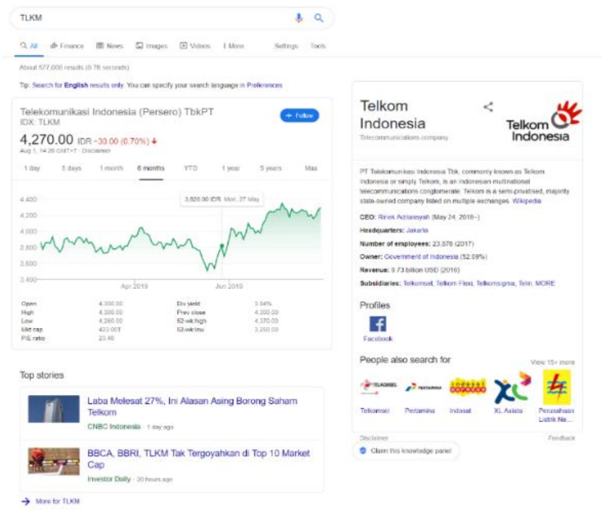

Figure 1 Example search result of stock ticker code on Google search engine

Based on Indonesia Central Securities Depository (KSEI) data of investor demographic[1], there are 2.48 million single investor identifications registered at the end of 2019 in which 39.4% has education lower than undergraduate degree, and 70.7% has income per annum of 100 million Rupiah or less. This number becomes even bigger at the end of 2020. This data leads us to believe that majority of retail investors in Indonesia to be less sophisticated and more likely to use free services such as Google to search for information as opposed to paid subscription services. Data from World Bank also shows big growth in internet users in Indonesia[2] that leads to more online information consumption and possibly strengthen the effect.

Based on the previous shortcoming and inconclusive results, this research contributes by examining the effect of investor attention on stock return in Indonesia, especially in a shorter timeframe where the attention effect has not diminished and on a more recent period, where there are more individual investors, possibly enhancing the effect. The effect is found to be consistent on several measures, however the probability is quite low.

The next sections of the paper are organized as follows. Section 2 shows information on variables and methodology. Section 3 shows result and discussion. Section 4 concludes.

2. METHOD

This research uses all eligible public companies registered in Indonesia Stock Exchange (IDX) during the period of 2015 – 2019. In order to prevent survivorship bias, all companies that are registered between 2015 and 2019 will be included in the sample if the observations are sufficient.

This research uses Fama and French Three Factor Model (FF3F) [8] to calculate abnormal return. The requirements for the chosen stocks are as follows. Non-financial companies; has price data, market cap data, and book value data of 2 years prior the inclusion of the stock into portfolio; has positive equity value. Companies that have negative equity during portfolio selection will be excluded from sample for the year but might be included in the next selection if the equity becomes positive.

2.1. Abnormal Return

The FF3F factors are calculated based on the equal weighted average of the 6 portfolio groups return. Stocks are grouped into portfolios based on company's size and book-to-market ratio. Portfolio return is calculated using weighted average method of the return of the companies in the portfolio. Abnormal return model of FF3F is shown on Equation (1).

$$\mathrm{r} - \mathrm{r_f} = \alpha + \hat{\beta}_{\mathrm{m}}(\mathrm{r_m} - \mathrm{r_f}) + \hat{\beta}_{\mathrm{smb}}\mathrm{SMB} + \hat{\beta}_{\mathrm{hml}}\mathrm{HML} + \varepsilon \qquad (1)$$

where r is daily log return of stock, r_f is the risk-free rate of return, α is abnormal return, β is coefficient of the variable, r_m is market return, SMB is size factor of FF3F, HML is book-to-market ratio factor of FF3F, and ε is the error term.

Since the frequency of return is calculated on daily basis, the risk-free rate referenced in this research is rate that also reflect the money market in daily basis; Indonesia Overnight Index Average (IndONIA) and Jakarta Interbank Offered Rate (JIBOR). IndoNIA and JIBOR are money market benchmark rate calculated everyday by Bank of Indonesia. IndONIA was first published in 2017 and is expected to replace JIBOR. In this research, IndONIA rate is used when available, and JIBOR is used for sample on older period.

$$\mathrm{AR} = \mathrm{r} - \mathrm{r_f} - \hat{\beta}_{\mathrm{m}}(\mathrm{r_m} - \mathrm{r_f}) - \hat{\beta}_{\mathrm{smb}}\mathrm{SMB} - \hat{\beta}_{\mathrm{hml}}\mathrm{HML} \qquad (2)$$

Daily β estimation for each variable is calculated using rolling regression based on Equation (1). Abnormal return is then calculated based on Equation (2).

2.2. Google Search Volume

Google Trend provides the number of searches done in Google search engine in a relative value of 0 to 100. This value, Google Search Volume (GSV), reflect the

---

[1] Source: KSEI (https://www.ksei.co.id/publications/Kaleidoskop-2020) retrieved February 25, 2021

[2] Source: World Bank (https://data.worldbank.org/indicator/IT.NET.USER.ZS?locations=ID) retrieved February 17, 2021

value of search frequency of the keyword at one point in time compared to the other observations within a defined period. If a keyword is rarely searched for the whole period of the requested query, Google Trend might return an error and no data is given.

2.2.1. Keywords

Investors can use various kind of words to get information; registered company name, product name, brand name, company's CEO name, abbreviated company name, or stock ticker code. [5] uses company name and stock ticker code as keywords for the research to compare but uses stock ticker code for most of the research except on Initial Public Offering event since the stock code is not yet determined or known to the public.

When someone search for a company name, the reason might be to look up company's address or finding job opening. Motivation for searching product or brand name might be finding product information or reviews. This research uses stock ticker code in order to get the value of search that is related to investing.

There is a fraction of keywords between the ticker code that is similar to Indonesian word, English word, unrelated famous instance, or geographic location (e.g., ROTI, FISH, KONI, BALI). These keywords are manually flagged as ambiguous tickers because the result of the search would not be about the public company but about the other subject and thus filling the data with noises. About 8.7% of the ticker codes are identified as ambiguous ticker in the sample.

2.2.2. Source and Category

Google Trend provides the option of selecting the source of search based on geographic location. In order to better capture the attention of the less sophisticated investors in Indonesia, this research uses the country of the stock market as the source. [13], [3], and [11] shows that the usage of origin country as the source of search data yields better result.

Besides country source, there is also an option of 26 categories, including "Finance" category. Both [3] and [11] shows that this category filter does not improve result and most of the times resulting in a lot of 0 values. This might indicate that majority of users search Google just by typing in the search box without choosing category of search. Thus, this research will focus on data from the default filter, "All Categories".

2.2.3. Value and Scaling

One of the limitations in getting GSV from Google Trend is that the scale of the result is based on the period of search. If the period of interest is below 4 hours, the scale of result will be in minute. However, if the search query is a bigger timeframe, the result will also be in bigger scale. Google Trend will return hourly, daily, weekly, and monthly data based on the search time limit of 1 week, 9 months, 5 years, and above 5 years, respectively. Similar to [4], this research uses lower frequency score as a master scale for higher frequency data as stated in Equation (3), where $\mathrm{GSV}_{\mathrm{t(daily)}}$ is the value of GSV of day t obtained by doing a query of 1 month and $\mathrm{GSV}_{\mathrm{t(monthly)}}$ is the value of GSV of month t obtained by doing a query of more than 5 years.

$$\mathrm{GSV}_{\mathrm{t}} = \mathrm{GSV}_{\mathrm{t(daily)}} \times \frac{\mathrm{GSV}_{\mathrm{t(monthly)}}}{100} \tag{3}$$

Since the value is not an absolute search number, following [3] and [11], GSV is then standardized (SGSV) using Equation (4) in order to make it comparable between companies. A change of SGSV is a change of its GSV based on the average and standard deviation for the past 365 days.

$$\mathrm{SGSV}_{\mathrm{t}} = \frac{\mathrm{GSV}_{\mathrm{t}} - \frac{1}{365}\sum_{\mathrm{i=1}}^{365}\mathrm{GSV}_{\mathrm{t-i}}}{\sigma_{\mathrm{GSV,t}}} \tag{4}$$

where $\mathrm{GSV}_{\mathrm{t}}$ is the scaled value of search on day t obtained from Equation (3), $\sigma_{\mathrm{GSV,t}}$ is the standard deviation of GSV one year prior to day t.

This research also uses standardization method of GSV from [5] to calculate Abnormal Google Search Value (AGSV) stated on Equation (5) as robustness test. Due to the nature of the calculation of logarithmic value, AGSV cannot process GSV with value of zero, resulting in removal of all GSV observations with 0 value.

$$\mathrm{AGSV}_{\mathrm{t}} = \log(\mathrm{GSV}_{\mathrm{t}}) \\ -\log[\mathrm{Median}(\mathrm{GSV}_{\mathrm{t-1}}, \ldots, \mathrm{GSV}_{\mathrm{t-n}}) \tag{5}$$

In order to be able to calculate SGSV and AGSV from January 2015, this research retrieved GSV data one year earlier from sample data, from January 2014 to December 2019, to calculate the average and standard deviation of the previous year.

Depending on the time of data retrieval, Google Trend might give a slightly different result due to the random sampling that is applied at Google Trend. The GSV data retrieval on this research was done over a course of few weeks in 2020. The data was unable to be retrieved in one day due to the limitation of number of downloads from Google Trend. However, [5] calculated the variation on Google Trend data and found that the data have more than 97% correlation and might not give any significance difference.

Table 1 Descriptive Statistic

	Obs.	Mean	Median	Std. Dev.	Min	Max	Skewness	Kurtosis
AR	464,753	-0.0004	-0.0002	0.0337	-1.991	2.1157	0.6162	80.8159
SGSV	399,899	0.1498	-0.2196	1.2062	-8.653	70.287	3.7303	70.6113
AGSV	201,451	0.0370	0.0224	0.4586	-2.9767	4.4034	0.3144	5.6755
Ln(1+News)	467,656	0.0237	0.0000	0.1382	0.0000	2.4849	6.1972	44.2144
Ln(1+Chunky)	467,656	0.0100	0.0000	0.0875	0.0000	1.9459	9.1242	91.3681

2.3. News Headline

Another proxy of attention-catching stock is news headline [2]. In order to calculate the attention of less sophisticated investors, this research also uses market news headlines from one of the largest market news publishers in Indonesia, Bisnis Indonesia (https://market.bisnis.com/) that is free, updated several times in a day, and easily accessible using the internet.

Based on [5], this research uses count of news headline that mention one of the ticker codes (News) and Chunky News (Chunky) that captures coverage of news headline that consists of multiple ticker code and dummy variables of news and chunky news. This variable will show if investors will be affected on more coverage (News), event (Chunky News), or occurrence of news (Dummy News and Dummy Chunky News).

3. RESULT AND DISCUSSION

Over the five-year period, there are 425 unique companies in the sample that makes up 65% – 70% of the total market cap of IDX. Descriptive statistic is

Table 2 Panel regression result of abnormal return as dependent variable. Panel A reports on models with SGSV as proxy of internet search. Panel B reports on models with AGSV as proxy of internet search. Significance of 0.1%, 1%, and 5% levels are marked by symbols ***, **, and *, respectively.

Panel A

	Model (1)	Model (2)	Model (3)	Model (4)
AR_{t-1}	-0.069***	-0.0689***	-0.069***	-0.0689***
$SGSV_t$	0.0005***	0.0005***	0.0005***	0.0005***
Ln(1 + $News_t$)	0.0061***			
Ln(1 + $Chunky_t$)		0.0071***		
News $Dummy_t$			0.0049***	
Chunky $Dummy_t$				0.0054***
Intercept	-0.0006***	-0.0005***	-0.0006***	-0.0006***
Observations	303,399	303,399	303,399	303,399
R^2	0.0054	0.0052	0.0054	0.0052

Panel B

	Model (5)	Model (6)	Model (7)	Model (8)
AR_{t-1}	-0.0361**	-0.0359**	-0.0361**	-0.0359**
$AGSV_t$	0.0008**	0.0009***	0.0008**	0.0009***
Ln(1 + News) $_t$	0.0069***			
Ln(1 + Chunky) $_t$		0.0076***		
News $Dummy_t$			0.0059***	
Chunky $Dummy_t$				0.0059***
Intercept	-0.0004***	-0.0003***	-0.0004***	-0.0003***
Observations	152,990	152,990	152,990	152,990
R^2	0.0021	0.0016	0.0021	0.0016

shown on Table 1. AR has negative mean which shows that the stocks underperform for about 4 basis point. For most variables, the data has high kurtosis value. There are extreme values on both tail for the variables, however this research uses all of the data including the extreme values.

Following [3] and [11], this research implements fixed effect panel regression. Hausman test [9] is used to check the significance of random effect compared to fixed effect and the result shows that fixed effect model is preferred on all models. This research uses Arellano method [1] to control for autocorrelation and heteroskedasticity. The results below are calculated using robust standard error.

The result of the regression can be seen on Panel A of Table 2. It shows that SGSV coefficient is positive and significant in explaining abnormal return. This result supports that attention pressures investors into buying stock and push the price. News headline also contributes to the result with positive significant coefficient. AGSV regression result can be seen on

Table 3 Correlations

	AR	**SGSV**	**AGSV**	**News**
AR	1.00			
SGSV	0.02	1.00		
AGSV	0.01	0.66	1.00	
News	0.03	0.06	0.09	1.00
Chunky News	0.02	0.05	0.07	0.70

Panel B of Table 2. The values are consistent to be positive between variation of the models. Intercept is the average value of individual effects.

The correlation between variables can be seen on Table 3. AR is not closely correlated with all of the variables. SGSV and AGSV are highly correlated because both of the variables use the same data but different standardization. News and Chunky News are also highly correlated because both of them represent similar data.

[5] mentioned that the effect can be examined better

Table 4 Panel regression result of abnormal return as dependent variable. Panel A shows regression result of smaller size companies. Panel B shows regression result of bigger size companies. Significance of 0.1%, 1%, and 5% levels are marked by symbols ***, **, and *, respectively.

Panel A – Small Companies

	Model (9)	Model (10)	Model (11)	Model (12)
AR_{t-1}	-0.0976***	-0.0973***	-0.0975***	-0.0973***
$SGSV_t$	0.0006**	0.0006**	0.0006**	0.0006**
Ln(1 + $News_t$)	0.0287***			
Ln(1 + $Chunky_t$)		0.0631***		
News $Dummy_t$			0.0211***	
Chunky $Dummy_t$				0.0435***
Intercept	-0.0006***	-0.0005***	-0.0006***	-0.0005***
Observations	144.067	144.067	144.067	144.067
R^2	0.012	0.0118	0.0118	0.0117

Panel B – Big Companies

	Model (13)	Model (14)	Model (15)	Model (16)
AR_{t-1}	-0.0224*	-0.0225*	-0.0224*	-0.0225*
$SGSV_t$	0.0005***	0.0005***	0.0005***	0.0005***
Ln(1 + $News_t$)	0.0028***			
Ln(1 + $Chunky_t$)		0.0032***		
News $Dummy_t$			0.0022***	
Chunky $Dummy_t$				0.0024***
Intercept	-0.0006***	-0.0006***	-0.0006***	-0.0006***
Observations	159.332	159.332	159.332	159.332
R^2	0.0011	0.001	0.0011	0.001

on smaller companies, where smaller funds can make a bigger price impact. The result on Table 4 shows that the effect can be better examined on smaller companies, with higher coefficient value and higher goodness of fit.

As previously mentioned, there are several tickers that are ambiguous and might not return company information when entered into Google search engine. The result on Table 5 shows that on sample without ambiguous tickers still maintain positive and significant coefficient both for internet search and news variable. Result for sample with only ambiguous tickers results in insignificant negative SGSV coefficient, but positive significant news variable. This shows that keywords search matter and might reflect the intention of investors to buy a stock. On the contrary, ambiguous tickers, arguably shows search result that are not related to company, does not reflect investors' attention of a stock. News variable still shows to be significant because the news is sourced from categories that focus on stock market, which means that the news headlines are related with the company.

Table 5 Panel regression result of abnormal return as dependent variable separated based on ambiguous ticker code. Significance of 0.1%, 1%, and 5% levels are marked by symbols ***, **, and *, respectively.

	Normal Keywords	Ambiguous Keywords
AR_{t-1}	-0.0703***	-0.05874*
$SGSV_t$	0.0006***	-0.00003
$\ln(1 + News_t)$	0.0059***	0.00963*
Intercept	-0.0006***	-0.00079***
Observations	274,379	29,020
R^2	0.0057	0.0038

This research shows that in daily timeframe, an increase in internet search can be associated with positive abnormal return. Previously in Indonesia, [12] shows that in a monthly timeframe, internet search shows to be insignificant in explaining stock return. This might be related to how retail investors' attention in Indonesia is short-lived, similar to how [6] shows domestic investors' advantage in short term trading. This research adds value by adding examples how investor attention pressures stock market, as previously shown by [2], [3], [5], [12], and [14].

On the internet, there are various ways investors can gather information. This research limitedly captures attention that are shown in internet search and news headline of selected news platform. Future research may find it advantageous to examine bigger data sets from various source of news and public opinion, including public forum and social media. Positive or negative sentiment of the text might also give better insight on how it affects stock market. This research uses FF3F model for abnormal return that does not apply to financial companies; different methodology of calculating abnormal return might shows additional information related to financial industry.

This result shows that there is still an abnormal return caused by behavioral bias of investors; however, this mispricing effect might be gone when most investors are more resourceful and absorb all information related to the fundamental value of the companies.

4. CONCLUSIONS

Google Trend provides sizeable information on number of internet searches performed by less sophisticated individual investors. This research examines the effect of investor's attention proxied by Google Search Volume and news headline to stock abnormal return. Positive and significant coefficient value is found for both GSV and news variables; this shows the effect of pressure of investor attention. While the impact and probability are small, the result is consistent on all models and robustness test with different standardization calculation, company size, and normal tickers. The effect is found to be stronger on smaller companies and GSV effect is not found in companies with ambiguous ticker code.

REFERENCES

[1] M. Arellano, "Computing Robust Standard Errors for Within-Groups Estimators," *Oxford Bulletin of Economics and Statistics*, vol. 49, no. 4, pp. 431–34, 1987.

[2] B. M. Barber and T. Odean, "All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors," The Review of Financial Studies, vol. 21, no. 2, pp. 785–818, Apr. 2008, doi: 10.1093/rfs/hhm079.

[3] L. Bijl, G. Kringhaug, P. Molnár, and E. Sandvik, "Google searches and stock returns," International Review of Financial Analysis, vol. 45, pp. 150–156, May 2016, doi: 10.1016/j.irfa.2016.03.015.

[4] D. K. Chronopoulos, F. I. Papadimitriou, and N. Vlastakis, "Information demand and stock return predictability," Journal of International Money and Finance, vol. 80, pp. 59–74, Feb. 2018, doi: 10.1016/j.jimonfin.2017.10.001.

[5] Z. Da, J. Engelberg, and P. Gao, "In Search of Attention," The Journal of Finance, vol. 66, no. 5,

pp. 1461–1499, 2011, doi: 10.1111/j.1540-6261.2011.01679.x.

[6] T. Dvořák, "Do Domestic Investors Have an Information Advantage? Evidence from Indonesia," The Journal of Finance, vol. 60, no. 2, pp. 817–839, 2005.

[7] E. F. Fama, "Efficient Capital Markets: A Review of Theory and Empirical Work," The Journal of Finance, vol. 25, no. 2, pp. 383–417, 1970, doi: 10.2307/2325486.

[8] E. F. Fama and K. R. French, "Common risk factors in the returns on stocks and bonds," Journal of Financial Economics, vol. 33, no. 1, pp. 3–56, Feb. 1993, doi: 10.1016/0304-405X(93)90023-5.

[9] J. A. Hausman, "Specification Tests in Econometrics," Econometrica, vol. 46, no. 6, pp. 1251–1271, 1978, doi: 10.2307/1913827.

[10] D. Kahneman, (1973). "Attention and effort," Englewood Cliffs, N.J: Prentice-Hall, 1973.

[11] N. Kim, K. Lučivjanská, P. Molnár, and R. Villa, "Google searches and stock market activity: Evidence from Norway," Finance Research Letters, vol. 28, pp. 208–220, Mar. 2019, doi: 10.1016/j.frl.2018.05.003.

[12] C. P. Nguyen, C. Schinckus, and T. V. Hong Nguyen, "Google search and stock returns in emerging markets," Borsa Istanbul Review, Jul. 2019, doi: 10.1016/j.bir.2019.07.001.

[13] T. Preis, H. S. Moat, and H. E. Stanley, "Quantifying Trading Behavior in Financial Markets Using Google Trends," Scientific Reports, vol. 3, p. 1684, Apr. 2013, doi: 10.1038/srep01684.

[14] F. Takeda and T. Wakao, "Google search intensity and its relationship with returns and trading volume of Japanese stocks," Pacific-Basin Finance Journal, vol. 27, pp. 1–18, Apr. 2014, doi: 10.1016/j.pacfin.2014.01.003.