ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
BSc Economics & Business economics
Specialization Financial Economics

A new trading strategy: Is it profitable?

The effects of investor attention on stock returns on the Dutch stock market

Author:	L Appel
Student number:	509714
Thesis supervisor:	Dr. J.J.G. Lemmen
Second reader:	PhD Candidate O. Commandeur
Finish date:	July 2021

PREFACE AND ACKNOWLEDGEMENTS

I would like to thank dr. Jan Lemmen for supervising my thesis. His fast feedback helped me a lot and made me write a better thesis. His shared knowledge and guidance proved to be of great value to me. I would also like to thank the ESE for the last three years of education and my family and friends for supporting me throughout these years.

NON-PLAGIARISM STATEMENT
By submitting this thesis the author declares to have written this thesis completely by himself/herself, and not to have used sources or resources other than the ones mentioned. All sources used, quotes and citations that were literally taken from publications, or that were in close accordance with the meaning of those publications, are indicated as such.

COPYRIGHT STATEMENT
The author has copyright of this thesis, but also acknowledges the intellectual copyright of contributions made by the thesis supervisor, which may include important research ideas and data. Author and thesis supervisor will have made clear agreements about issues such as confidentiality.

ABSTRACT

This study examines the effects of internet search volume on stock returns, trading volume and volatility in the period 2016-2021 on the Dutch stock market. This paper tests whether institutional and retail investors are influenced differently by search queries, whether the influence is bigger in some industries and if trading in stocks with the highest search volume could lead to a profitable trading strategy. The study relies on panel data mainly from Google, Yahoo! Finance and Bloomberg's Terminal. Multiple Fama-MacBeth regressions are performed. The findings show that abnormal search volume has a positive effect on stock returns, while standardized search volume has a negative effect. There is no clear relationship between volatility and search volume. The results support the hypothesis that retail investors are more likely to be influenced than institutional investors. Stocks, of companies that people are more familiar with, are more likely to be affected by search volume. Going long in a portfolio with the most searched stocks and short in a portfolio with the least searched stocks, can lead up to positive returns of 19%.

Keywords: Investor attention, search volume, stock returns, trading strategy, Fama-MacBeth regression.
JEL classifications: G10, G11, G41

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1 Introduction

1.1 Internet and financial markets

The past decade (social) media and its usage has become more important to investors and traders. For example, think about apps like Robin Hood that make investing accessible for everyone. The internet is influencing financial markets through many channels. People can gather tons of information about firms in just a few mouse clicks. Newspapers and academics are no longer the only information holders, nowadays you can find (almost) everything about financial markets online. The internet has become a new source of information and knowledge. This digital environment provides us with new measures of consumer search behavior, that used to be unavailable. This new measure also opens the door for new studies.

Da et al. (2009) where the first ones to pick up on this. They wanted to study if search volume is a measure of investor attention and if it influences stock prices. Their research focused on Google searches and concluded that search volume is a good proxy for investor attention. They also concluded that returns will be positive in the first two weeks but will go back afterwards. After them, many economists started to research this new measure of investor attention. Bank et al. (2011) also concluded that search volume can lead to positive returns in the short run. Joseph et al. (2011) even created a trading strategy based on these earlier studies. They went long in stocks with high level of search volume and shorted stocks with low level of search volume. This trading strategy gives positive returns unless transaction costs are included according to Joseph et al. (2011). Bijl et al. (2016) used a more recent dataset and found the exact opposite. This contrary result raises questions like, is this strategy time changing? Or does it depend on other factors? A lot of time has passed since Bijl et al. (2016) publicized their results; therefore, it is possible to examine if the strategy is time dependent. It is not yet clear if the trading strategy is profitable or can be made profitable when using a longer holding period to avoid transaction costs. Besides that, almost all existing literature is about the US stock market. There has been little research about other stock markets, for example in Europe or Australia. This leaves the question, if the results would be the same in other markets, still open. Although a lot of research has been done, there are still a lot of unresolved issues as mentioned before. This paper will try to fill those blanks.

1.2 Research question

This paper will examine further what the influence of investor attention on stock prices is. This is extremely relevant, because it could lead to a new trading strategy that has not enough academic foundation yet. The existing literature will be complemented. Some earlier results will (probably) be confirmed and new hypotheses and ideas will be tested. This research is thus an extension of the existing studies. This paper will research the effects of online searches on stock prices, returns, trading volume and the volatility of stocks. The main question in this paper is:

Does investor attention affect stock returns, trading volume and volatility? If so, can this lead to a profitable trading strategy?

1.3 Setting

Search volume is widely being accepted as a new and direct measure of investor attention. Da et al. (2009) where the ones to introduce search intensity as a proxy for investor attention. Bank et al. (2011), Joseph et al. (2011), Vlastakis & Markellos (2012), Takeda & Wakao (2014), Bijl et al. (2016), Tang and Zhu (2017), and Yang et al. (2020) all used search intensity as a proxy for investor attention. Most of them made use of Google Trends, which mostly captures the attention of retail investors according to Da et al. (2009) and Bank et al. (2011). This paper will try to grasp the attention of sophisticated investors as well. Sophisticated investors are more likely to use exclusive data and in-house information systems. When a bank or pension fund wants to make new investments, it is unusual for them to just Google a stock name. They will dive deeper into a company, before making investment decisions. Retail investors usually invest smaller amounts of money than institutional investors, which explains why institutional investors have a higher incentive to gather a lot of information. Besides that, institutional investors also have more resources to research stocks and possible returns. To capture the attention of more sophisticated investors, the Bloomberg Terminal is being used. Bloomberg's Terminal monitors and analyses real-time financial market data and their yearly $24,000 fee makes it very exclusive.

When measuring investor attention this paper will look at search intensity. This is how often people search for information about companies/stocks. People can search a company name or use its ticker symbol. Every publicly traded firm has a ticker symbol, a unique code to made trading easier. Da et al. (2009) and Joseph et al. (2011) looked at search volume of ticker symbols, while Preis et al. (2010) and Bank et al. (2011) looked at company names instead. To capture as much investor attention as possible, both will be used. This has never been done before and thus will exceed the existing literature.

The sample will contain the three biggest Dutch stock market indexes. The Amsterdam Exchange index (AEX) includes 25 companies. This index is composed of the 25 largest market-capitalization weighted companies. The AMX, Amsterdam Midkap Index, contains the following 25 largest companies. The Amsterdam Small Cap Index (AScX) contains the 51-75 biggest companies. The firms used where all listed on one of these exchanges in April 2021. See table 2 for the unique ticker code of every company listed on these exchanges. Earlier research has mainly focused on the US stock market (Da et al. (2009), Preis et al. (2010), Joseph et al. (2011), Vlastakis & Markellos (2012) and Bijl et al. (2016)). Although some economists have studied other markets, no research has yet been done about the Dutch stock market. Bank et al. (2011) researched the German stock market, Takeda & Wakao (2014) and Adachi et al. (2017) the Japanese market and Zhang & Wang (2015) and Yang et al. (2020) the Chinese market. Researching a new market also adds value compared to the existing literature.

This thesis has the following structure. First, all relevant literature will be discussed. The four hypotheses will be introduced based on earlier research outcomes. After that, the data will be reported. The sample and size will be motivated, the sources of data will be revealed, and some statistics will be provided. Chapter 4 will provide the methodology per hypothesis. The next chapter will focus on the empirical results followed up by the conclusions that can be drawn from the results. The last chapter will discuss the outcome of this thesis.

Table 1. Ticker symbols per firm.
Every public traded firm has a unique Ticker symbol. The table contains all the tickers of the firms registered on the AEX, AMX and AScX indexes.

Exchange	**Company name**	**Ticker**
AEX	Adyen NV	ADYEN
	Aegon NV	AGN
	Akzo Nobel NV	AKZA
	ArcelorMittal SA	MT
	ASM International NV	ASM
	ASML Holding NV	ASML
	ASR Nederland NV	ASRN
	BE Semiconductor Industries	BESI
	Heineken HV	HEIA
	IMCD NV	IMCD
	ING Groep NV	INGA
	Just Eat Takeaway	TKWY
	Koninklijkse Ahold Delhaize NV	AD
	Koninklijkse DSM NV	DSM
	Koninklijke KPN NV	KPN
	Koninklijke Philips NV	PHIA
	NN Group NV	NN
	Prosus NV	PRX
	Randstad NV	RAND
	RELX PLC	REN
	Royal Dutch Shell PLC	RDSA
	Signify NV	LIGHT
	Unibail-Rodamco-Westfield	URW
	Unilever NV	UNA
	Wolters Kluwer NV	WKL
AMX	AALBERTS NV	AALB

	ABN Amro Bank NV	ABN
	Air France-KLM	AF
	Alfen	ALFEN
	AMG	AMG
	Aperam	APAM
	Arcadis	ARCAD
	Basic-Fit	BFIT
	Boskalis	BOKA
	Corbion	CRBN
	Eurocommercial Properties	ECMPA
	Fagron	FAGR
	Flow Traders	FLOW
	Fugro	FUR
	Galapagos	GLPG
	GrandVision	GVNV
	Intertrust	INTER
	JDE Peet's	JDEP
	OCI	OCI
	Pharming	PHARM
	PostNL	PNL
	SBM Offshore	SBMO
	TKH Goup	TWEKA
	Vopak	VPK
	WDP	WDP
AScX	Accell Group	ACCEL
	Accsys Technologies	AXS
	Amsterdam Commodities	ACOMO
	AFC Ajax	AJAX
	Avantium	AVTX
	B&S Group	BSGR
	Koninklijke BAM Groep	BAMNB
	Brunel International	BRNL
	CM.com	CMCOM
	ForFarmers	FFARM
	Heijmans	HEIJM
	Hunter Douglas	HDG
	Kendrion	KENDR
	Kiadis Pharma	KDS
	Lucas Bols	BOLS
	Nedap	NEDAP
	NSI	NSI
	Ordina	ORDI
	SIF Holding	SIFG
	Sligro Food Group	SLIGR
	TomTom	TOM2
	Van Lanschot Kempen	VLK
	Vastned Retail	VASTN
	Vivoryon Therapeutics	VVY
	Wereldhave	WHA

CHAPTER 2 Literature review

2.1 Introduction

The way financial markets and (social) media interact is changing and developing over time. Beatty & Smith (1987) found that a consumer's search for information precedes his or her decision. People are making decisions based on the information given to them. With much information being held online, it is more accessible and easier than ever, to find whatever you're looking for. Investors can find information in just seconds. With the internet and media becoming more accessible, it is also becoming more important for financial markets. People gather information faster about whether to buy or sell a stock. The internet gives a new measure of investor attention. A lot of research has been done, trying to measure investor attention. Most of these studies have focused on online search volume of either company names or tickers. The internet and its impact on financial markets keeps developing and growing, which makes this subject so extremely relevant.

Earlier research has mainly focused on the search volume of either a company's name or its ticker by Google. Da et al. (2009), Joseph et al. (2011), Engelberg et al. (2012) and Yang et al. (2020) found that mostly individual retail investors use the Google search engine. These studies therefore focus on individual retail investors. Institutional/sophisticated investors often make more use of in-house databases or highly regarded databases like Bloomberg's Terminal. Bloomberg's Terminal also has information on search volume, which will be used. This paper will thus exceed the existing literature by not only looking at Google searches. Bijl et al. (2016) concluded that the relationship between Google search volume and stock returns changes over time. So that adds even more value to this research. Besides that, looking at the time lag between Google searches and changes in stock returns can be interesting as well. This paper will focus on the short run. The first hypothesis will discuss the time lag.

Also interesting is the research of Vozlyublennaia (2014). She investigated if past returns indicate a nature of information that investors pay attention to. The results showed that there is a significant interaction between lagged returns and attention. This suggests that attention can alter the predictability of returns. According to Vozlyublennaia (2014) this would improve the market efficiency. Tantaopas et al. (2016) confirm this. According to them, attention effect is good for the market, because attention leads to more informative decision making by investors and faster information processing. This leads to more efficient markets. This makes the subject relevant from a different perspective, namely market efficiency.

2.2. Measures

2.2.1 Company name or ticker

As mentioned before, retail investors and institutional investors differ in their way of gathering information. But that is not the only contrary. Retail investors are more likely to search a company's name, while institutional investors are more likely to search on a company's unique ticker. Because of this, both measures are included.

However, it is important to keep in mind, that the search of a company's name can have different meanings. For example, when searching for "ABN AMRO" people may want to transfer money or look at their savings rather than informing themselves on the (stock) value of the bank. Or for instance, when people search "Heineken" they can either be interested in buying a beer or getting to know more about the company. When people search a ticker symbol, it is almost certain they are looking for investment information.

It is important to note that the companies KPN, ASML, ASM and Adyen have tickers (almost) identical to their names. We also see this for other companies registered at the AMX or AScX index. This makes the line between a ticker search or company name search very thin. This is important to keep in mind. In addition to that, it is also important to remember that a search is more valuable for buyers than for sellers. This was found by Joseph et al. (2011). Their explanation suggests that people can only sell the stocks they already own and therefore research is more valuable when buying a stock.

2.2.2 Direct measure of investor attention

Google searches and Bloomberg terminal information are both direct measures of consumer attention. This is more reliable than indirect measures as used before. Indirect measures such as advertising expenses used by Hou et al. (2009) or newspaper headlines used by Grullon et al. (2004) are not precise enough. For example, it is not sure how much people an actual newspaper headline reaches. This new method of measuring investor attention will therefore be used. It is widely accepted as a measure of investor attention. Da et al. (2009), Bank et al. (2011) and Joseph et al. (2011) all concluded that it is a good proxy.

2.3 Hypotheses

2.3.1 Higher stock returns?

A lot of research has been done about whether Google search volume is a good proxy for investor attention, what its influence on stock prices is and if it can lead to a profitable trading strategy. It all started when Barber and Odean (2008) introduced the attention-grabbing theory, which confirmed that individual investors are net buyers of attention-grabbing stocks. By attention-grabbing, they meant stocks that are in the news, or experience high abnormal trading volume or have extreme one-day returns. They explain this attention-driven buying because of the difficulty investors undergo when searching for a stock to buy. Investors do not face the same problem when selling, because they can only sell the stocks, they already own. Investors only buy the stocks that caught their attention according to the theory. Da et al. (2009) and Tang & Zhu (2017) also confirmed this attention-grabbing theory. They studied this topic even further.

Da et al. (2009) investigated whether search volume is a good proxy for investor attention and what the effects on stock prices are. They use the weekly Search Volume Index (SVI) by Google Trends. The SVI is the number of searches for a term scaled by its time-series average. They use stock ticker symbols as a measure to capture a particular stock interest. They used all Russell 3000 stocks for their studies from 2004 to 2008.

Their results show that following an increase in SVI a stock has positive returns for the first two weeks. But this result is only found among smaller stocks, because the price pressure is bigger for those stocks. This outperformance stops after two weeks (and is reversed within the first year). They find that when a retail investor pays more attention to a stock, they are more likely to buy it and push up its price temporarily. This concludes that search volume has a positive effect on stock prices in the short run. This is in line with hypothesis 1:

Higher search volume will lead to higher stock returns in the short run.

The short run will be defined as the first two weeks following the strong search increase. This is based on Da et al. (2009), who found positive returns for the first two weeks. Tang and Zhu (2017) found abnormal returns, but only for day zero.

This paper will not only research the effects of search volume on stock returns, but also on trading volume and volatility. Preis et al. (2010) and Bank et al. (2011) researched this as well. Preis et al. (2010) researched if there is a relation between weekly search volume and financial market fluctuations. They also used information from Google. They studied the weekly search volume data from 2004 until 2010 for all S&P 500 company names. Their findings show that the trading volume of stocks is correlated to its search volume. Increasing search volume overlaps with high transaction volume, which reflects the attractiveness of a stock. They conclude that news, search volume and market movements are closely related.

Bank et al. (2011) also look at the effects Google search volume has on financial markets. They focus on the effects on trading activity and stock liquidity. Their research is limited to all German stocks from 2004 to 2010. They look at company names normalized provided by Google Insights. Their results show that an increase in internet search volume is related to higher trading activity, improved stock liquidity and leads to higher future returns in the short run. The Google search volume captures the attention of uninformed investors, which results in a reduced information asymmetry, improved liquidity, and short-term buying pressure.

Yang et al. (2020) recognize that investor attention is a factor that affects stock returns and trading volume. By looking at Chinese listed firms from 2011 until 2018 they documented investor attention and its effect. Their results showed that abnormal search volume is positively associated with stock returns, but with a complete reversal in the next period. Abnormal search volume also has a positive relation with trading volume but does not reserve. They conclude that search volume only affects stock returns and trading volume in the short run, which is in line with hypothesis 1. Tang and Zhu (2017) came to the same conclusion: investor attention leads to positive abnormal returns, but only in the short run. They found positive abnormal returns the same day of the surge in investor attention. The day after, this abnormal return disappeared or even reversed.

Tang and Zhu (2017) also studied the impact of investor attention on financial markets. They took a closer look at the effects of a surge in investor attention on the returns of ADRs. They used Google search volume as an indicator for investor attention as well. They found evidence that a surge in investor attention leads to a positive abnormal return, but that this will disappear or reverse short after.

2.3.2 Retail investors versus institutional investors

Da et al. (2009) also investigated if Google search volume is a direct measure for investor attention. They found that it is a good proxy. They also examine whose attention SVI captures and found that it is mostly the attention of individual/retail investors. They also tested the attention theory of Barber and Odean (2008), their findings strongly support this theory.

Bank et al. (2011) also conclude that Google search volume captures the attention of uninformed investors. As stated before, this paper will also use information on institutional investor attention. This raises an interesting question, which is formulated in hypothesis 2:

Institutional investors are less likely to be influenced by the media.

The underlying assumption is that institutional investors are less likely to base their choices on online information. More sophisticated investors have a lot of experience and will be more likely to identify a temporary price increase. They are also less likely to take that much risk, because for them there is a lot at stake. Therefore, they are less likely to be influenced by irrational information.

2.3.3 Different Industries

An interesting question that comes up when studying this subject is, whether there are differences between industries. There are a lot of different industries present on the AEX, AMX and AScX. For example, ASML Holding NV is a tech company, Just Eat Takeaway sells consumer goods, while ABN AMRO Bank provides financial services. The idea is that people are more familiar with companies they encounter in their everyday life and thus know more about these companies through general knowledge. This makes that people are more likely to investigate unknown companies. These companies are mostly from the technological or industrial industries. Because people know less about it, they are most certainly more influenced by information surrounding these industries. This theory led to the following hypothesis:

The media influence is bigger in technology and industrial markets.

2.3.4 Momentum strategy

Joseph et al. (2011) use online ticker searches to forecast stock returns and trading volumes. They also find that online search volume from Google Trends is a good measure of investor attention. They look at the weekly search volume of all S&P 500 firms' tickers over the period 2005-2008. They look at both normalized and scaled search volume. They sort the firms in 5 different portfolio's each week. Quantile 1 contains the firms with the lowest search intensity, while quantile 5 holds the firms with the highest search intensity. Every portfolio is regressed of daily returns on the three Fama-French factors.

The results show that going long on the high search intensities and short on the low search intensities gives a daily abnormal return of 0.028%, but because of transaction cost this strategy is not profitable. What can be learned from this paper is that the momentum investing can result in positive abnormal returns, but that transaction cost can be disrupting. An important lesson learned from Joseph et al. (2011) and Bijl et al. (2016), is that the holding period should be longer (than one week), for the strategy to be possibly profitable.

Bijl et al. (2016) find the exact opposite of Joseph et al. (2011). They use a more recent time frame, namely 2008-2013. Their results show that a reversed momentum strategy is profitability, this phenomenal is also called the overreaction hypothesis. The overreaction hypothesis finds that the winners start losing and the losers start winning. The strategy sells stocks with high search volume (short) and buys stocks with low search volume (long). The same principle holds for the transaction costs, taking these into account no longer leads to profits. They conclude that the relationship between search volume and returns is changing over time and thus will this topic remain interesting as media and data flows keeps developing. This leads us to the last hypothesis:

Going long on a portfolio containing the most searched stocks and going short on a portfolio containing the least searched stocks, gives positive returns.

2.4 Summary

As found by Da et al. (2009), Bank et al. (2011) and Joseph et al. (2011) it is clear to say that online ticker searches are a valid measure for investor attention. The literature also finds support for the theory that retail investors make use of more simple search engines like Google Trends and institutional investors make use of more sophisticated databases.

Stock returns are expected to rise following an increase in investor attention according to Da et al. (2009), Bank et al. (2011), Adachi et al. 2017, Tang & Zhu (2017) and Yang et al (2020). Contrary to all other researchers, Adachi et al. (2017) found the possibility for stock returns to stay positive in the long run.

Trading volume is also expected to rise with investor attention. The more attention a stock is getting, either good or bad, the more people start trading (Preis et al. 2010: Bank et al. 2011: Yang et al. 2020: Takeda & Wakao 2014). Vlastakis & Markellos (2012) zoom in on volatility and search intensity. Their results show that demand for information is significantly positively related to historical and implied volatility. Their results remain the same, even after controlling for market return and information supply. They also found that information demand rises during times of higher returns and when people have a higher level of risk aversion. Table 2 gives a summary of all the important literature.

Given that financial markets and investor attention will keep changing and developing, this topic will remain an important economic subject. This makes the outcomes of earlier research questionable. Besides that, no research has looked at the search volume of more sophisticated investors. This paper will thus add value to the existing literature, by also including institutional investors' attention. All existing literature has focused on the US markets, with exceptions for Germany, China and Japan. This paper will examine the Dutch stock markets, AEX, AMX and AScX.

This paper thus contributes to the existing literature by studying a new market, including the attention of more sophisticated investors, looking at more recent years and introducing a trading strategy with longer holding periods.

Table 2. A literature overview.

The table displays the most import and related literature. Column 1 contains the names of all writers and the year the paper was (first) published. The second column gives information on the data being used. The next column describes the methodology used. The fourth column shows the results found by the writers. The last column sums the results up and concludes.

Author(s) (Publication year)	Data	Methodology	Results	Conclusion
Gervais, Kaniel & Mingelgrin (2001)	NYSE Stocks. 1963-1996. Daily and weekly data.	Portfolios are formed. Zero investment portfolio: long in high-volume stocks and short in low-volume stocks. Reference return portfolio: weights are not equal but adjusted in a way that net investment is zero.	Returns per day and size firm, ranging from 0.08 to 0.96 percent per dollar. There is a significant positive relation between trading volume and future returns.	Stocks experiencing abnormal high (low) trading volume over a short period (a day or week) tend to appreciate (depreciate) in the following month. This high-volume premium is due to a stock's visibility. High-volume return premium lasts at least 20 trading days but could last up to 100 trading days.
Da, Engelberg & Gao (2009)	Google Trends weekly search volume index (SVI & ASVI) for all Russell 3000 stocks. 2004 – 2008. Tickers of firms. ASVI is the log of SVI during week minus log of medium SVI during previous 8 weeks.	Vector autoregression (VAR)	Positive returns during the first two weeks following an increase in SVI, but only among smaller stocks. Week 1: 18.742 Week 2: 14.904 From week 3 onwards not significant. Outperformance stops after two weeks (and is reversed within first year).	SVI captures the active attention of retail investors. The results support the hypothesis of Barber and Odean (2008). We see this happening especially in the context of IPOs.
Preis, Reith and Eugene Stanley (2010)	Weekly search volume (SVI) data of Google on S&P 500 firms. Company names. 2004 – 2010.	Linear autocorrelation regression and linear cross-sectional regression	The trading volumes of the S&P500 index are correlated to their company name search volumes.	Transaction volumes increases together with search volume. Conclusion that news, search volume and market movements are closely related.
Bank, Larch & Peter (2011)	Google search volume by Google Insights. Company names. A term is relative (normalized). German stocks and thus also only German internet	Portfolios based on internet search volume (3 quantiles). Perform regression(s). Then form a zero-investment strategy (long in portfolio with the largest	A firm's name is significantly (0.34) related to trading activity and thus to investor recognition. A large signed change in search volume is	Increase in internet search volume is related to higher trading activity, improved stock liquidity and leads to higher future returns in the short run. The Google

	users. Weekly and monthly data.	signed change and short in smallest signed change). The returns are regressed on the market model (CAPM), Fama and French three-factor model, and Carhart four-factor model.	associated with a small signed change in illiquidity and vice versa. Positive short-run relationship between Google search volume and the future stock returns.	search volume captures the attention of uninformed investors, which results in a reduced information asymmetry, improved liquidity and short-term buying pressure.
Joseph, Babajide Wintoki & Zhang (2011)	Online ticker search from Google Trends. S&P500 firms. Weekly data. Data normalized and scaled. Stock returns, volume data and volatility from the CRSP Database.	Every week firms are sorted into 5 quintiles. Portfolios are regressed of returns on the three Fama and French factors. The abnormal trading volume is the difference between trading volume on a given day and its average over the entire sample period.	A portfolio that is long on the high search intensities and short on the low search intensities gives a daily abnormal return of 0.0280%. But not profitable because of trading cost of rebalancing every week.	A ticket search is more valuable for somebody who is considering a "buy" decision (rather than a "sell"). Mostly captures individual investors.
Vlastakis & Markellos (2012)	Weekly search volume (SVI) from Google Trend. 30 largest stocks on NYSE and NASDAQ.	Volatility (realized in a week) is summing up squared returns for each week.	Market demand information has a significant positive effect (0.5786) on historical volatility, implied volatility, and trading volume. Also, information demand increases during periods of higher returns.	Information demand increases together with risk aversion.
Moat, Curme, Avakian, Kenett, Stanley & Preis (2013)	How often Wikipedia pages have been viewed. Data revision history of Wikipedia. They looked at page usage as well as page changes.	Regressions	Researched if the usage of Wikipedia influences the decisions made on the stock markets.	Evidence of increases in the page views relating to stock market falls.
Takeda & Wakao (2014)	The relation between online search intensity (SVI & ASVI) and stock returns and trading volume. Search volume from Google using the Nikkei 225 stocks. 2008-	Returns are based on Fama-French three-factor model. They make 4 quantiles, based on search intensity. A multivariate analysis.	Strong positive correlation (0.192) between search volume and trading volume and weak positive correlation (0.094) between search volume and stock returns.	An increase in search intensity also leads to an increase in trading activity. High investor attention also leads to high trading volume, but chance of increase in stock price is low.

	2011.			
Vozlyublennaia (2014)	Investor attention (SVI) is measured by Google search. Dow Jonges Industrial Average, NASDAQ and S&P 500.	Regressions. Test on autocorrelation about past returns determination of attention on future returns.	There is a short-term change in returns following increase in attention. On the other hand, a shock in returns leads to long-term change in attention.	Past returns determine the impact of attention on future returns. There is interaction between lagged returns and investor attention.
Zhang & Wang (2015)	Focus on the performance of ChiNext market. 2011-2012. Baidu index for search volume (SVI).	Unit root tests and multiple regressions	Measure if investor attention has influence on a stock. Find a positive price pressure in the short run (0.0073).	Investor attention can affect the stock market.
Bijl, Kringhaug, Molnár & Sandvik (2016)	Google Trends and Wharton Research Data Services (WRDS). S&P500 index company names. 2008-2013. Weekly data. Standardized the data.	A regression Excess returns are returns minus beta times market returns from daily stock returns.	Shows that the Google search volumes lead to negative returns. Thus, high levels of Google search volume predict low future (excess) returns. This is significant, but their impact is small (-0.0007).	Overreaction hypothesis. This strategy is profitable unless transaction costs are considered. Concludes that the relationship between Google search volume and stock returns changes over time.
Tantaopas, Padungsaksawasdi & Treepongkaruna (2016)	Relationship between investor attention and stock returns, volatility & trading volume in Asia-Pacific stock markets. Volume Index of Google (SVI).	Granger causality tests with exogenous macroeconomic variables suggested by Chen et al. (1986). Vector Autoregression (VAR).	Investigate impact of attention on market efficiency from three distinct perspectives: return predictability, volatility predictability, and trading volume predictability. Stationarity for all variables.	Attention is good for market overall because it promotes efficiency. Volatility and stock returns are predictable, but trading volume is not.
Adachi, Masuda & Takeda (2017)	Research about the relationship between investor attention (SVI & ASVI) and stock prices in Japan. Startup stocks Japan. Firms listed on mothers and JASDAQ. 2014-2016.	Estimate stock returns based on the Fama-French three factor model.	There is a positive relation between stock returns and search intensity. This relation tends to be larger for startup firms. (Mothers: -0.007 JASDAQ: 1.437)	Contrary to other research, they find the possibility for an increase in stock returns to stay in the long run.

Colaco, Cesari & Hegde (2017)	Google search volume (SVI) to measure retail investor attention and its influence on IPOs.	Univariate regressions and multivariate tests.	Retail investor attention following IPO filing but prior to pricing is positively related to the initial valuation. Prior: 1.499. After:0.721	The results are robust to alternative matching methods. retail investor attention has an important role in the early stages of IPO valuation.
Tang & Zhu (2017)	Study if a surge in investor attention leads to security price changes. SVI from Google Trends for ADRs (American Depositary Receipts). 2004-2015. Ticker symbols. CRSP.	ASVI is the logarithm of the search volume index for firm i on day t minus the logarithm of the median SVI during the previous ten trading days. Panel data and the Carhart four-factor model for returns.	A surge in investor attention is associated with a positive abnormal return (0.074) the same day, but this disappears or even reverses soon after.	The results apply for both developing countries as developed countries. They also find evidence for the attention theory of Barber and Odean (2008).
Yang, Ma, Wang & Wang (2020)	"Baidu Index" (retail investors) Weekly search volume index. 2011-2018. ASVI.	Regression with Fama-French three-factor CAPM. Also add control variables like financial analysts' ratings and market indices in previous weeks.	Abnormal SVI has positive association with stock returns (0.104), but this is reserved in next period. Also, positive association with trading volume (0.365), but this becomes weaker after two weeks. Investor attention affects returns and trading volume in short term.	Results search volume indicator works better capturing attention of less sophisticated investors. Investor attention is mostly driven by earnings announcements, management forecasts, financial analysts following, mergers and acquisitions and dividend payout.

CHAPTER 3 Data

3.1 Introduction

This section will discuss the data collection process. The collected data will also be addressed. The three main data sources are: Google Search data, Bloomberg Search data and Yahoo! Finance. The Google Search data is obtained from Google Trends. The Bloomberg Search data from Bloomberg's Terminal, which is accessible for Erasmus Rotterdam University students. The information on stock prices and trading volume comes from Yahoo! Finance.

3.2 Google Search data

3.2.1 Google Search data

Google search data will be used as a proxy for investor attention. As mentioned before, search inquiries are widely accepted as a measure of investor attention. Da et al. (2009), Bank et al. (2011) and Joseph et al. (2011) showed that online searches are a good proxy. For example, when searching a company ticker, a person is unquestionably paying attention to this company. Search volume is a direct measure of attention, leaving indirect measures such as newspaper headlines founded by Hou et al. (2009) and advertising expenses found by Grullon et al. (2004) behind.

Why Google? Google is the online search engine with the biggest market share (Johnson 2021). By 92% of all online searches Google is used, which makes the results of the search engine a good representation of all online searches. Besides that, Google's data base goes back to 2004 and features almost all countries.

3.2.2 Google Trends

Google is the market leading search engine, according to their website (2015) their mission is to "Organize the world's information and make it universally accessible and useful". Google Trends is a service provided by Google that analyzes the popularity of top search queries in Google Search across various regions, countries and languages.

When searching a keyword in Google Trends, the sample period and included countries must be chosen. This paper uses a time period of 5 years. Bijl et al. (2016) also use a period of 5 years (2008-2013). The data in this paper will be more recent, namely from the first of May 2016 until the first of May 2021. Google Trends gives weekly data. The search queries are measured worldwide. The interest over time will be shown. The interest is relative to itself, meaning that every observation has a value of 0 to 100. Important to note is that the value does not represent the number of times a term is searched, but instead it is relative to its own value. The value 100 represents the highest search level a search term has had in the past five years. The value 50 means that the word was half as popular as its peak. The value 0 means to little information is available. The output from Google Trends is called Search Volume Index (SVI). Earlier research uses either, SVI, Abnormal Search Volume Index (ASVI) or standardized SVI. Preis et al. (2010), Vlastakis & Markellos (2012), Vozlyublennaia (2014), Tantaopas et al. (2016) and Tang & Zhu (2017) make use of the Search Volume Index (SVI). The Abnormal Search Volume Index is defined by Da et al. (2009) as

$$ASVit = ln(SVit) - ln[med\ (SVit\text{-}1\ \ldots,\ SVit\text{-}4)] \quad (3.1)$$

Where ln(SVit) is the natural logarithm of SV for stock I on day t and ln[med (SVit-1 …, SVit-4)] is the natural logarithm of the median value of SV during the prior two weeks. Da et al. (2009), Takeda & Wakao (2014), Adachi et al. (2017) and Yang et al. (2020) all use the AVSI.
The standardized SVI is search volume divided by the standard deviation. Bank et al. (2011), Joseph et al. (2011) and Bijl et al. (2016) use this method. Both abnormal search volume and standardized search volume will be investigated, to see which works best. “Normal” SVI is too biased because some companies have a standard higher search level than others. Because of this, SVI is excluded.

In figure 1 we see the search queries in Google Search for the term “HEIA”, which is the ticker symbol of Heineken NV. The peak is reached in September 2016. The search term reaches its low in May 2016, June 2017 and October 2020. As mentioned before, this paper will not only look at ticker symbols but also at company names to grab the attention of less sophisticated investors as well. Figure 2 show the number of times “Heineken” or “heineken” has been searched. Again, the values are relative to the peak. The peak is in January 2019. The graphs show that the ticker symbol and company name do not reach their peaks at the same time.

What caused the peak in investor attention in September 2016 and January 2019?
In September 2016 Heineken started a global partnership with Formula1, the organization behind the highest class of international car racing. Besides that, multiple innovations like the wild lager beers and Heineken Light were launched. The stock price is fluctuating a lot around this time. The peak in January 2019 was about a rapport of the Bank Morgan Stanley. They published that Heineken was overrated. The Bank said that the prices of raw materials would go up and forecasted slow growth in beer markets. After this, the stock value of Heineken decreased with 2.5 precent. These are both examples of how investor attention can affect stock prices.

As already noticed, some company names are almost identical to their tickers. This applies to the companies Adyen, ASM, ASML, DSM, KPN and NN listed on the AEX. For most companies the issue was solved by using adjectives. For example, for KPN “Koninklijke KPN” is being used and for NN “NN Group”. Only the company names of Adyen and ASML are still identical to their ticker symbols. For the AMX, Alfen and OCI have the exact same ticker as company name. The same holds for NSI. All these companies have identical search volume for their ticker and their name. Therefore, the companies with tickers ADYEN, ASML, ALFEN, OCI and NSI will be excluded from the sample. Information on the stock prices of the company Kiadis Pharma (KDS) was not available and therefore this company will be excluded as well. Not all stocks had already gone public by May 2016. This was the case for ASM, ASRN, TWKY, PRX, LIGHT, UNA, BFIT, JDEP, AVTX, BSGR and FFARM. All these stocks only have data from their share issuance onwards. Sometimes the search volume had the value 0. This is changed to 0,0000001 to make the natural logarithm calculations possible.

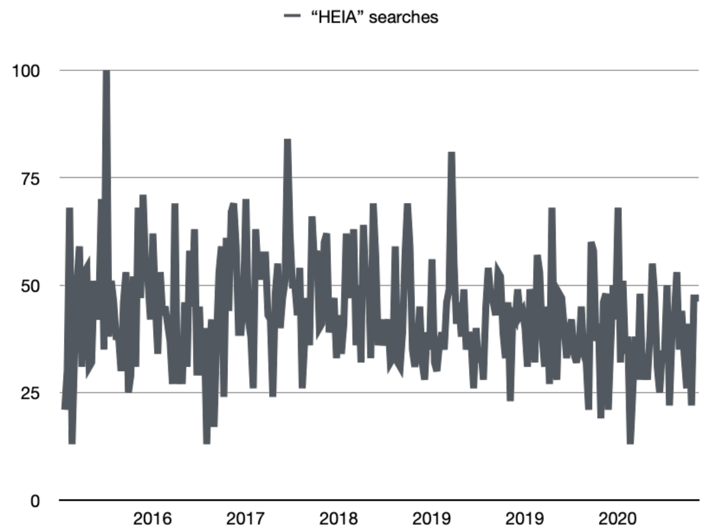

Figure 1. A graphical summary of “HEIA”.

The figure shows how often HEIA was searched relative to its own peak. HEIA is the ticker symbol of Heineken NV. The graph shows the search queries in Google Search from May 2016 – May 2021 worldwide.

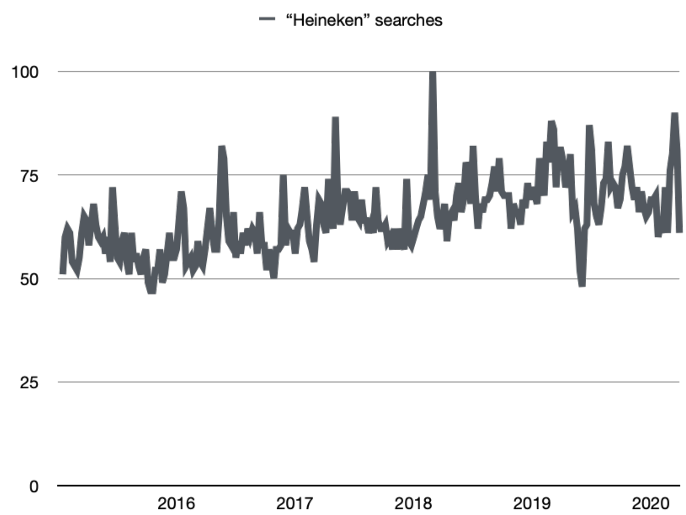

Figure 2. A graphical summary of “Heineken”.

The graph shows the search volume of the company name Heineken. The values on the y axis are relative to the peak of 100. This peak was in January 2019 and means that the term Heineken was searched the most on that day in the last five years.

3.3 Bloomberg Terminal

3.3.1 Bloomberg Terminal

The Bloomberg Terminal is a financial data provider. It is used to access, compile and analyze information. The data service has new feeds, allows for chatting and facilitates the placement of financial transactions.

Why Bloomberg? The Bloomberg Terminal is, together with Thomsom Reuters Eikon, the most popular platform. Bloomberg is the most expensive tool. A subscription costs $24,000 per year, which makes it very exclusive. Currently, there are around 325,000 subscribers. These subscribers typically are large institutional investors, portfolio managers and financial analysts. These are exactly the investors we are trying to get a hold on.

3.3.2 Bloomberg Data

Bloomberg has a lot of different functions. The “TREN” function will be used in this paper. This function can be used on weekly or monthly bases. The weekly data will be collected since the Google Trends data is weekly as well. Bloomberg’s data does not make a difference between a company’s name or ticker. The “TREN” function shows the securities with the most activities surrounding news and media. The news reader activity displays companies ranked by news searches and story readership. The social volume rank shows the companies ranked by social publication volume. The social velocity displays securities with the most recent Bloomberg Insights triggered by a surge in social publication volume. The latter is the most interesting function. The social velocity shows the securities with the highest social velocity in the last 365 days. The social velocity is the rate of change of engagement with content on social media, think about tweets, a like or a comment. This social engagement is more than just a headline, it shows people’s reaction and urge to share with others. The social velocity shows the effect an increase in social volume in Bloomberg’s Terminal has had on companies trading volume and price. The data gives the highest social velocity per day with the corresponding company name. The data is in percentages. Because the data is only available for one year and only covers the stocks with the highest social velocity, the data from Bloomberg is not sufficient for all regressions.

3.4 Yahoo! Finance

3.4.1 Yahoo! Finance

Yahoo! Finance is a media property of Yahoo! Yahoo! is a web services provider. They have different products, like a web portal, search engine and finance page. Yahoo! Finance provides financial news and data like stock prices, trading volume and sustainability rankings. The data on stock prices and trading volume will be used. The stock price is the current price that a share of a company is worth. Trading volume refers to the number of shares traded.

3.4.2 Data on stock prices and trading volume

The data on stock prices is collected from Yahoo! Finance. For every day the opening price, close price, adjusted close prices and trading volume are given. The adjusted close price is the price adjusted for both dividends and splits. This paper will use the adjusted close price to calculate the returns. The sample period is May 2016 – May 2021. The data consist of all 75 stocks listed on the AEX, AMX and AScX stock markets. Figure 3 shows the stock price development of Heineken for the past five years. Since Google Trends gives weekly data, the information on stock prices will also be collected on a weekly basis. In 2017 and again in 2019, we see strong increases of the stock's price. In the beginning of 2020, we see a large drop in the stock price, due to the corona pandemic outbreak.

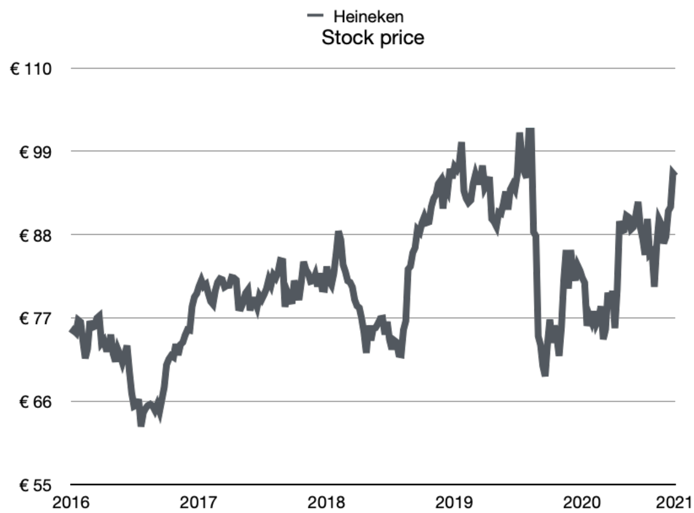

Figure 3. Stock prices Heineken.

The figure shows the development of the stock price of Heineken NV. The data starts in May 2016 and ends in May 2021. Between 2018 and 2020 we see large fluctuations. Currently the stock price is recovering from its large drop following the corona pandemic outbreak.

Figure 4 presents the trading volume of Heineken over the past five years. The sample period is May 2016 – May 2021 again. The trading volume is the number of shares traded. We see the trading volume varying from around one million to ten million per day. A high level of trading volume can either be good or bad, people can rapidly sell their stocks or buy new once. There is a high level of trading volume in September 2017. This is around the same time that prices went up. We also see a peak in the beginning of 2020. Again, in the month March, during the dramatic price decline as seen in figure 3. Both peaks show a relation between the stock prices and trading volume of Heineken, which is not surprising. If both are influenced by investor attention will be researched later.

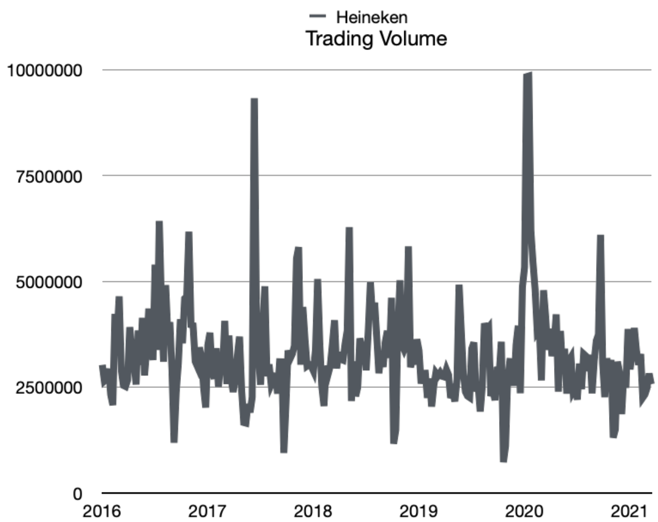

Figure 4. Trading Volume Heineken.

The graph shows the trading volume of Heineken stocks. The trading volume is the number of shares traded per day. The number is between one million and ten million shares a day. The total amount of shares outstanding is around 570 million.

3.5 Three factor model and STOXX Europe 600

To calculate the returns, the Fama and French (1993) factors are necessary. The risk-free rate of return, the total market portfolio return, the size premium and value premium are needed. The data can be collected from the website of Fama and French.

The STOXX Europe 600 is a stock index of European stocks, that represents large, mid and small capitalization companies. The index covers 17 European countries, including the Netherlands. The STOXX 600 will be used as a benchmark. The benchmark will be used as a measure of abnormal returns. The returns generated with the introduced trading strategy, being stocks with high search intensity and selling stocks with low search intensity, will be compared to returns of the STOXX 600.

The data is obtained from investing.com. Investing.com is one of the top three financial websites, providing information about stocks, futures and more.

Two control variables will be added: market capitalization and the interaction term between search volume and market capitalization. The data on market capitalization comes from the annual financial reports published by the companies themselves. Missing data is complemented using data from Morningstar, an American financial service.

CHAPTER 4 Methodology

4.1 Testing hypothesis

To formulate an answer on the main question, multiple hypotheses have been drafted. All hypotheses will be tested and thereafter an answer on the research question will be formulated. As mentioned before, this paper will research the effects of online searches on stock prices, returns, trading volume and the volatility of stocks. All hypotheses will be tested using various regressions. The regressions will be performed using Stata. Every hypothesis will be treated separately in the next sections. The empirical results will be discussed in the next chapter.

4.2 Demeaning and standardizing

Hypothesis one assumes that higher search volume will lead to higher future stock returns in the short run. To test this hypothesis, a Fama-MacBeth cross-sectional regression will be performed. This regression will show the relationship between search volume and future stock returns for the Dutch stock market. Da et al. (2009) also used this regression to examine the relation between abnormal search volume and stock returns. Two control variables are added to the regression as well. Adding control variables controls against omitted-variable bias. The first control variable is the logarithm of market capitalization. Market capitalization and stock returns are related to each other, leaving market capitalization out would lead to endogeneity (omitted-variable bias). Market capitalization is measured as the total amount of shares outstanding multiplied by the average market price of one share in that year. The logarithm of market capitalization is used to simplify and because it is more convenient. The second control variable is the interaction term effect between market capitalization and search volume. The interaction term effect tells us how the two independent variables work together to impact the dependent variable. The interaction term effect gives a better understanding of the variability in the dependent variable. Both the interaction term between market capitalization and abnormal search volume and the interaction term between market capitalization and standardized search volume are added.

This research will, as explained before, examine both abnormal and standardized search volume. Abnormal search volume means the natural logarithm of SV minus the natural logarithm of the median value of SV during the prior two weeks. We will also standardize the variables, meaning that the standard deviation is one. This can be done by dividing the difference between an observation minus the mean divided by the standard deviation. Both demeaning and standardizing are used to make the results easier to interpret. The results can be interpreted as follows: a one standard deviation increase in X leads to a … increase or decline in Y, with X being the independent variable and Y being the dependent variable. Unfortunately, Bloomberg does not hold search volume information on all stocks. Therefore, for this first test only Google search volume is being used.

The returns are obtained by the regression of the daily time series of returns on three factors from Fama and French (1993). The three factors are the excess return on the market, the return difference between portfolios of small and big stocks (SMB) and the return difference between portfolios of high and low book-to-market stocks (HML). The Fama and French Three-Factor Model is an asset pricing model that expands the CAPM. The Fama and French Three-Factors have found to explain cross-sectional differences in stock returns and are therefore being used. Joseph et al. (2011), Takeda and Wakao (2014), Adachi et al. (2017) and Yang et al. (2020) all used the Fama and French factors. The following regression is used to obtain the returns:

$$R - Rft = a + Bm*(Rmt - Rft) + Bs*SMBt + Bh*HMLt + et \quad (4.1)$$

This paper will also examine the relationship between volatility and stock returns. Volatility will be measured as the square root of the sum of squared daily returns. We assume that the mean return is zero in comparison to the standard deviation. Poon and Granger (2003) conclude that this assumption makes the estimates of volatility more precise. The following expression gives the formula for volatility:

$$\sigma_{w,t} = \sqrt{\sum_{i=1}^{n} r_d^2} \quad (4.2)$$

N is the number of trading days in the corresponding month. Volatility will be compared among the stocks to see if there are differences between stocks with high or low search intensity.

The results will be checked for robustness. The sample will be split in 2 different periods. The first 2.5 years and the last 2.5 years will be regressed separately to check the results. Also, noisy tickers or company names will be removed, to see whether that makes a difference. The "noisy" tickers are being removed, to see if that changes the results. The companies that are excluded based on their tickers are: AD, RAND, LIGHT, AF, FLOW, FUR, INTER. All these words have other or general meanings, for example "AD" is also the name of a Dutch newspaper and "RAND" means edge. There are also some companies excluded based on their name. These are companies that provide services or products that are so common that these names are mostly searched for other purposes than trading. The company names that are removed for the robustness checks are: ASR, Heineken, ING, Just Eat Takeaway, KPN, Randstad, Shell, ABN, Basic-Fit, PostNL and AJAX. For instance, "AJAX" is a soccer club that can be searched by fans to check statistics or buy tickets to the stadium and a cleaning soap as well. Or "Randstad" is also an area in the Netherlands.

4.3 The investors

Hypothesis 2 is as follows: Institutional investors are less likely to be influenced by the media. To test this hypothesis, investor attention must be split. The Fama-MacBeth regression will focus on the effects of Google search volume on future stock returns, and on the effects of Bloomberg searches on future stock returns. The Google searches are a measure of retail investor attention, while the Bloomberg searches are used as a measure of institutional investors. To see if there is a difference between the investors, the coefficients must be compared. The Bloomberg information comes from the social velocity ("TREN" function). The data from Bloomberg only covers the last 365 days. Besides that, it only gives data for the stocks experiencing the highest change of social engagement. Not all stocks are thus covered by the Bloomberg data. This is important to keep in mind when interpreting the results.

To check whether the results are plausible, the same comparison will be done for the tickers and company names. Under the assumption that institutional investors are more likely to search tickers and retail investors are more likely to search a company's name, the regressions will be executed. This time, the regression will focus on the difference both search terms have on future stock returns. The ticker searches will be regressed on future stock returns and company name searches will be regressed on future stock returns. Again, the coefficients will be compared to test if institutional investors are less likely to be influenced.

4.4 Splitting industries

Before concluding whether different industries make a difference for the influence of investor attention, it is necessary to split all companies in industries. There are 16 different industries: Financial Services, Insurances, Chemicals, Industrial Metals, Technology/Electronics, Foods and Drinks, Oil & Gas Producers, Real Estate/Properties, Land Reclamation, Construction & Materials, Lifestyle, Transport, Pharmaceuticals, Services, Goods (consumer and household) and Agriculture. All companies are assigned to a sector. Some sectors contain of just one company, their results will be shown but not used to draw conclusions. These are the sectors Land Reclamation (Bokalis), Lifestyle (Ajax), Services (Brunell) and Agriculture (ForFarmers). The industries must be split up in order to make a distinction between industries people are more and less familiar with. To the category daily encounter industry, we count: Financial Services, Insurances, Technology/Electronics, Foods and Drinks, Lifestyle, Transport, Services and Goods. The residual industries are the ones less familiar.

To test whether search intensity has a different effect per industry, multiple regressions will be run. The returns will be estimated again, this time per industry. The Fama-MacBeth regressions will be performed again. The two control variables, market capitalization and the interaction term, are added again. This is done to avoid omitted-variable bias. Comparing the coefficients of the regressions will give you the results.

Table 3. Industry overview.

All companies listed on the Dutch stock markets AEX, AMX and AScX are divided in different sectors. There are 16 different sectors: Financial Services, Insurances, Chemicals, Industrial Metals, Technology/Electronics, Foods and Drinks, Oil & Gas Producers, Real Estate/Properties, Land Reclamation, Construction & Materials, Lifestyle, Transport, Services and Agriculture.

Company	Ticker	Exchange	Sector
Adyen NV	ADYEN	AEX	Financial services
Aegon NV	AGN	AEX	Life Insurances
Akzo Nobel NV	AGZA	AEX	Chemicals
ArcelorMittal SA	MT	AEX	Industrial Metals
ASML Holding NV	ASML	AEX	Technology/Electronic
ASR Nederland NV	ASRN	AEX	Nonlife Insurance
BE Semiconductor Industries	BESI	AEX	Technology/Electronics
Heineken NV	HEIA	AEX	Foods and Drinks
IMCD NV	IMCD	AEX	Chemicals
ING Groep NV	INGA	AEX	Financial services
Just Eat Takeaway NV	TKWY	AEX	Foods and Drinks
Ahold Delhaize NV	AD	AEX	Foods and Drinks
DSM NV	DSM	AEX	Pharmaceuticals
KPN NV	KPN	AEX	Consumer Goods
Philips NV	PHIA	AEX	Technology/Electronics
NN Group NV	NN	AEX	Nonlife Insurance
Prosus NV	PRX	AEX	Technology/Electronics
Randstad NV	RAND	AEX	Consumer Goods
RELX PLC	REN	AEX	Consumer Goods
Royal Dutch Shell PLC	RDSA	AEX	Oil & Gas Producers
Signify NV	LIGHT	AEX	Construction & Materials
Unibail-Rodamco-Westfield	URW	AEX	Real Estate/properties
Unilever NV	UNA	AEX	Household Goods
Wolters Kluwers NV	WKL	AEX	Financial Services
Aalberts NV	AALB	AMX	Industrial Engineering
ABN AMRO NV	ABN	AMX	Financial Services
AIR France-KLM	AF	AMX	Transport

Alfen	ALFEN	AMX	Transport
Advanced Metallurgical Groep	AMG	AMX	Industrial Metals
Aperam	APAM	AMX	Industrial Metals
Basic-Fit	BFIT	AMX	Consumer Goods
Boskalis	BOKA	AMX	Land Reclamation
Corbion	CRBN	AMX	Technology/Electronics
Eurocommercial Properties NV	ECMPA	AMX	Real Estate/properties
Fagron	FAGR	AMX	Pharmaceuticals
Flow Traders	FLOW	AMX	Financial Services
Fugro NV	FUR	AMX	Oil & Gas Producers
Galapagos NV	GLPG	AMX	Pharmaceuticals
Grandvision	GVNV	AMX	Pharmaceuticals
Intertrust	INTER	AMX	Financial Services
JDE Peet's	JDEP	AMX	Foods and Drinks
OCI	OCI	AMX	Construction & Materials
Pharming	PHARM	AMX	Pharmaceuticals
PostNL	PNL	AMX	Consumer Goods
SBM Offshore	SBMO	AMX	Oil & Gas Producers
TKH Groep	TWEKA	AMX	Technology/Electronics
Vopak	VPK	AMX	Consumer Goods
WDP	WDP	AMX	Real Estate/properties
Accell Group	ACCEL	AScX	Transport
Accsys Technologies	AXS	AScX	Construction & Materials
Amsterdam Commodities	ACOMO	AScX	Food and Drinks
AFC Ajax	AJAX	AScX	Lifestyle
Avantium	AVTX	AScX	Chemicals
B&S Group	BSGR	AScX	Consumer Goods
Koninklijke BAM Groep	BAMNB	AScX	Construction & Materials
Brunel International	BRNL	AScX	Services
CM.com	CMCOM	AScX	Technology/Electronics

ForFarmers	FFARM	AScX	Agriculture
Heijmans	HEIJM	AScX	Construction & Materials
Hunter Douglas	HDG	AScX	Construction & Materials
Kendrion	KENDR	AScX	Transport
Kiadis Pharma	KDS	AScX	Pharmaceuticals
Lucas Bols	BOLS	AScX	Foods and Drinks
Nedap	NEDAP	AScX	Technology/Electronics
NSI	NSI	AScX	Real Estate/Properties
Ordina	ORDI	AScX	Technology/Electronics
SIF Holding	SIFG	AScX	Oil & Gas Procedures
Sligro Food Group	SLIGR	AScX	Foods and Drinks
TomTom	TOM2	AScX	Transport
Van Lanschot Kempen	VLK	AScX	Financial Services
Vastned Retail	VASTN	AScX	Transport
Vivoryon Therapeutics	VVY	AScX	Pharmaceuticals
Wereldhave	WHA	AScX	Real Estate/Properties

4.5 New trading strategy

The final hypothesis focusses on a new trading strategy. All firms will be split up in three portfolios. The 25 most searched companies will form portfolio one. The second portfolio will contain the next 25 most searched companies, while the third holds the 25 least searched companies. The holding periods will vary between one month, 3 months and 6 months. Based one earlier research of Joseph et al. (2011) and Bijl et al. (2016) it is known that a shorter holding period will not lead to profitability. They both concluded that a holding period of one week was too short, because rebalancing every week led to high transaction costs. The information period will be one month (for all holding periods). The excel sort function will be used to build the portfolios.

A potential investor will go long in portfolio one and short portfolio three. To see if this leads to a profitable trading strategy, the returns will be measured. The returns will be measured (again) using the Fama and French (1993) three factor model. This includes the excess return on the market, the return difference between the small and big stocks and the return difference between the portfolios of high and low book-to-market stocks. Returns will be obtained by the following expression:

$$R - Rft = a + Bm*(Rmt - Rft) + Bs*SMBt + Bh*HMLt + et \quad (4.3)$$

The returns will be compared to a benchmark. The STOXX Europe 600 index will be used as benchmark, considering that this is one of the most established benchmarks in Europe. This paper will also investigate whether high search volume strengthens the momentum effects and vice versa. Vozlyublennaia (2014) found that past returns determine the impact of attention on the future returns. So, researching this means the thought of winners winning even more and losers losing even more. This can be concluded from the autocorrelation.

With the three portfolios being formed, we can also test to see if there is a difference between abnormal trading volume. The portfolios are composed based on search intensity. For every company the abnormal trading volume is measured as follows:

$$(AV) = (Vit - Vi,avg) / Vi,avg \quad (4.4)$$

This means the trading volume for firm i on day t minus the average daily volume over the entire sample divided by the average volume over the entire sample. Next, the abnormal trading volume per portfolio will be summed. This makes it possible to find an association between the search intensity and abnormal trading volume.

4.6 Remaining tests

Lastly, we test for simultaneity and whether the past value(s) of a stock's returns have influence on the present value. Vozlyblennaia (2014) found that past returns influence the attention on future returns. Her research concluded that winners win even more, and losers lose even more. The theory of Vozlyblennaia (2014) can be tested based on autocorrelation.

Simultaneity is a source of endogeneity. Simultaneity would mean that companies with higher returns attract more attention, and more attention would lead to higher returns. This can be tested using a granger causality test. The variables in the test will be total abnormal search volume and stock returns.

4.7 Summary

This section discussed all hypotheses testing. Most hypotheses will be tested using regressions. The Fama-MacBeth cross-sectional regression will be used to test most of them. The returns will be measured using the Fama and French (1993) three factor model. The regressions will be performed using Stata. The momentum strategy splits the companies in three portfolios. A potential investor will go long in portfolio one (highest search volume) and short portfolio three (lowest search volume). The empirical results will be discussed in the next chapter.

CHAPTER 5 Empirical results

5.1 Introduction

Multiple Fama-MacBeth regressions have been performed. Besides that, a trading strategy was constructed and executed. The results will now be discussed. With the results, we hope to find answers to the hypotheses and of course the main question.

5.2 Hypothesis one

The available data is panel data. Therefore, a Fama-MacBeth regression is performed to test whether search volume has influence on stock returns. To measure search volume, only data from Google Trends is used. The data obtained from Bloomberg was not sufficient. The data comes from all companies on the Dutch stock markets (AEX, AMX and AScX). For every company there is weekly data starting 16 May 2016 until 26 April 2021. For all companies and dates, the returns, abnormal search volume (based on name and ticker), standardized search volume (based on name and ticker), market capitalization and the three Fama and French (1993) are collected and measured.

The purpose of the regression is to measure the effects of search volume on stock returns. Therefore, the dependent variable is stock returns. The independent variables are the abnormal search volume, standardized search volume and the three Fama and French (1993) factors. The logarithm of the market capitalization and interaction term between this logarithm and search volume are added as control variables. Newey-West standard errors are used, they are robust to both autocorrelation as well as heteroskedasticity. Besides that, Newey-West standard errors are often used by economists when performing a Fama-MacBeth regression. For example, Da et al. (2009) used Newey-West standard errors as well. The standard errors are computed using the Newey-West (1987) formula with five lags. The number of lags was calculated using the rule of thumb of Stock and Watson (2011). The rule of thumb is the time (T) to the power of a third multiplied by 0.75. The time is 52 weeks multiplied by 5 years, which gives us 5 lags.

5.2.1 Search volume and returns

The results are shown in Table 4. The results are measured for the short run, defined as the first two weeks following the increase/decrease in search volume. Column 1 holds the results for the ASVI and column 2 for the SSVI. The results can be interpreted as follows: A one-standard-deviation increase in ASVI leads to an increase of about 1130% in returns. This means that higher abnormal search volume also leads to higher future stock returns. For SSVI, this is a decrease of 374%, meaning that higher standard search volume leads to lower future stock returns. Unfortunately, none of the coefficients is significant. However, we can clearly see that the ticker search has a positive effect on stock returns, while name search has a negative effect. This will be discussed further in the next section.

Table 4. The price pressure.

The table shows the results of a Fama-MacBeth regression. The regressions test the influence of search volume on stock returns. The dependent variable is stock returns. The independent variable search volume. Search volume is measured abnormally and standardized. Column 1 holds the results for abnormal search volume and column 2 for standardized search volume. Market capitalization is added as control variable, as well as the interaction effect between the market capitalization and search volume. Standard errors are computed using the Newey-West (1987) formula with five lags.
* Significant at 1% level, ** significant at 5% level, *** significant at 10% level.

	(1)	(2)
Marketcap (LOG)	-781.879	-487.216
	(743.940)	(465.120)
Marketcap (LOG) *	-42.528	
ASVI ticker	(44.453)	
Marketcap (LOG) *	10.393	
ASVI name	(10.998)	
Marketcap (LOG) *		-781.075
SSVI ticker		(701.593)
Marketcap (LOG)*		833.095
SSVI name		(763.430)
ASVI Ticker	1371.378	
	(1429.221)	
ASVI Name	-241.964	
	(320.463)	
SSVI Ticker		25582.25
		(23036.87)
SSVI Name		-25956.92
		(23818.11)

Rmt – Rft	-838.636 (806.634)	-714.388 (683.000)
SMBt	-1233.954 (1217.763)	-861.262 (873.229)
HMLt	-310.0952 (281.009)	31.675 (25.118)
Constant	25415.06 (24987.144)	16620.57 (16238.01)
Observations	14870	14874
R squared	0.1387	0.1668

5.2.2 Search volume and volatility

The first hypothesis also focuses on the effects search volume has on volatility. To see whether volatility depends on search volume, the volatility of all stocks was measured. This is compared to their total (abnormal) search volume to draw conclusions. An overview of the data can be found in the appendix. In order to make it easier to draw conclusions, the results are presented in a graph. See figure 5. As you can see, there is no clear relationship between volatility and search volume.

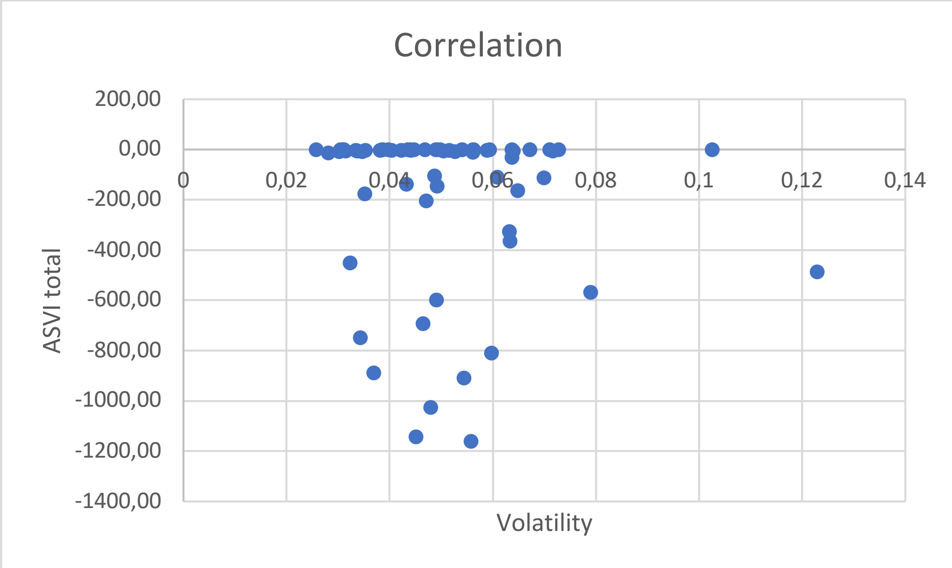

Figure 5. Correlation between volatility and search volume.

The x-axis is volatility and y-axis is total ASVI. The search volume total varies from around 0 tot 1200. Volatility lies between 0.02 and 0.13. The plot shows the correlation between volatility and search volume.

5.2.3 Robustness

The previous results are checked for robustness. The sample is split up in two time periods and the companies with the corresponding tickers: AD, RAND, LIGHT, AF, FLOW, FUR, INTER, ASRN, HEIA, INGA, TWKY, KPN, RDSA, ABN, BFIT, PNL and AJAX are removed. The reason behind this is already explained in section 4.2. The time periods are 1 May 2016 – 31 October 2018 and 1 November 2018 – 1 May 2021. The results for the first period (1 May 2016 – 31 October 2018) are presented in table 5 and the results for the second period (1 November 2018 – 1 May 2021) in table 6. Again, column 1 gives the results for ASVI, and column 2 for SSVI.

The results are interpreted the same as before. A one-standard-deviation increase in ASVI leads to a decrease of about 5168%. An increase in SSVI now leads to an increase in stock returns, by 30647%. In the sample we saw the opposite, ASVI led to higher stock returns and SSVI to lower stock returns. The results for the second period are much smaller dan the sample and first period results. There is no clear explanation for this. A one-standard-deviation increase in ASVI leads to a decrease of 0.1175% in the second period. A one-standard-deviation increase in SSVI leads to an effect of almost zero. In the first period, the effects of a company's ticker searches are positive, while the company's name searches are negative. For the second period, we see the same results for the ASVI, but it reverses for SSVI.

Table 5. Robustness first period

The table gives the relationship between stock returns and search volume. The table holds the results for the robustness check. The sample is split in two different time periods. The first time period is presented in this table. Some companies are removed from the sample. Column 1 gives the results for ASVI, and column 2 for SSVI. Again, market capitalization and the interaction term are added as control variables. Standard errors are computed using the Newey-West (1987) formula with five lags.
* Significant at 1% level, ** significant at 5% level, *** significant at 10% level.

	(1)	(2)
Marketcap (LOG)	-2200.278	-1468.58
	(2015.773)	(1462.423)
Marketcap (LOG) * ASVI ticker	-88.086	
	(208.083)	
Marketcap (LOG) * ASVI name	275.101	
	(196.120)	
Marketcap (LOG) * SSVI ticker		-2562.886
		(2380.010)
Marketcap (LOG)* SSVI name		1613.335
		(1463.98)
ASVI Ticker	2828.428	
	(3480.111)	
ASVI Name	-7996.261	
	(5783.94)	
SSVI Ticker		80896.5
		(75085.06)
SSVI Name		-50249.57
		(45486.53)
Rmt – Rft	-1733.652	-1241.932
	(1694.801)	(1159.799)

SMBt	-1964.698	-1653.418
	(2063.311)	(1676.761)
HMLt	-2239.182	239.998
	(2068.001)	(173.36)
Constant	70725.5	44740.01
	(66091.90)	(47602.59)
Observations	5296	5298
R squared	0.1853	0.2372

Table 6. Robustness second period.

The table holds the results for the second time period of the robustness check. For the robustness check some companies are removed. Again, we measure the relationship between stock returns and search volume. Column 1 gives the results for ASVI, and column 2 for SSVI. Standard errors are computed using the Newey-West (1987) formula with five lags.
* Significant at 1% level, ** significant at 5% level, *** significant at 10% level.

	(1)	(2)
Marketcap (LOG)	-0.0090	-0.0080
	(0.0001)	(0.0083)
Marketcap (LOG) * ASVI ticker	-0.0003	
	(0.0003)	
Marketcap (LOG) * ASVI name	0.0042	
	(0.0036)	
Marketcap (LOG) * SSVI ticker		0.0105
		(0.0082)
Marketcap (LOG)* SSVI name		-0.0038
		(0.0057)

ASVI Ticker	0.0068	
	(0.0088)	
ASVI Name	-0.1243	
	(0.1124)	
SSVI Ticker		-0.3458
		(0.2698)
SSVI Name		0.1240
		(0.1688)
Rmt – Rft	-0.0015	0.0023
	(0.0149)	(0.0160)
SMBt	0.0026	-0.0000
	(0.0131)	(0.0135)
HMLt	0.0197	0.0210
	(0.0131)	(0.0180)
Constant	0.2219	0.1911
	(0.2681)	(0.2819)
Observations	5602	5604
R squared	0.2005	0.2287

5.3 Hypothesis two

The second hypothesis focusses on the differences between investors. Assuming that there are two types of investors, institutional investors and retail investors, the differences between their impressionability is measured. The assumption holds that institutional investors are more likely to search tickers and retail investors company names. Besides that, it is also assumed that institutional investors make use of more complex databases like Bloomberg's Terminal, while retail investors use simpler data like Google.

The previous results already made distinctions between ticker and name searches. The results in table 4 clearly show that the ticker searches have a positive effect on the returns, while name searches have a negative effect. This shows that an increase in ticker searches make returns go up, while an increase in name searches make returns go down. This doesn't say much about how big the influence is. Therefore, we need to perform a regression again.

Again, a Fama-Macbeth regression is performed. Standard errors are computed using the Newey-West (1987) formula with five lags. The regression measures the influence of both social velocity and abnormal search volume on stock returns. Social velocity (Bloomberg) represents the institutional investors and abnormal search volume (Google) represents the retail investors. Institutional investors are more likely to use Bloomberg, while retail investors use Google Trends more often. Bloomberg's Terminal provides data on the social velocity. This is the change of (social) engagement of stocks, and the corresponding change in price and volume. It only provides data of the highest engagement stocks, not all stocks. Therefore, the results could be biased. The companies with the highest engagement are ArcelorMittal, Shell, Air-France and Unilever. The dependent variable is stock returns, while the independent variables are social engagement and abnormal name searches (ASVI name).

The results can be found in table 7. The effects of social engagement on stock returns are significant. An increase in social engagement leads to a very small decline in stock returns. Abnormal search volume has a small positive effect on stock returns. From these results we can derive that Bloomberg searches have a small negative influence on stock returns, while Google's abnormal search volume has a (slightly) larger positive effect on stock returns.

Table 7. Bloomberg and Google.

The table presents a Fama-MacBeth regression. In the first column we look at the effects between social engagement and stock returns (Bloomberg). The dependent variable is stock returns, and the independent variables are social engagement (Bloomberg) and abnormal search volume (Google). Standard errors are computed using the Newey-West (1987) formula with five lags.
* Significant at 1% level, ** significant at 5% level, *** significant at 10% level.

	Stock returns
Social engagement	-0.0001 (0.0000)***
ASVI	0.0017 (0.0018)
Rmt – Rft	0.0001 (0.0001)
SMBt	0.0003 (0.0003)
HMLt	-0.0004 (0.0002)*
Constant	0.0001 (0.0003)
Observations	89
R squared	1

5.4 Hypothesis three

The third hypothesis focusses on the differences between industries. Are some industry stocks influenced more by search volume? The hypothesis presumes that people are more likely to search information on relatively unfamiliar companies and that these company stocks are therefore more influenced by search volume. For example, industries that people are less familiar with, are chemicals, land reclamation and oil & gas producers.

The sample contains a total of 16 different industries. Only Google data is used since Bloomberg's data is insufficient. In total, 16 different Fama-MacBeth regressions were performed to test this hypothesis. For the analysis, the Abnormal Search Volume index (ASVI) is used. This is the name and ticker Google abnormal search combined. Again, the Fama and French (1993) three factors are included. Also, the log of the market capitalization and the interaction effect with de ASVI are used as control variables. Standard errors are computed using the Newey-West (1987) formula with five lags.

The results are displayed in table 8. The first column contains the results for the first industry, the financial services, the second the results for the insurance industry, the third for the chemicals industry and so on. The coefficients can be interpreted as follows: a one-standard-deviation increase in ASVI leads to a negative return change of 2.8% of stocks in the financial service industry. To test the hypothesis, we must distinguish between companies' consumers are more and less familiar with. People have a lot of encounters in life with industries like food and drinks, goods (consumer and household), insurances and financial services. They are less familiar with industries like chemicals, industrial metals and construction & materials.

Comparing all coefficients, the technology & electronics industry is influenced the most in a positive way, and the foods & drinks industry in a negative way. A one-standard-deviation increase in ASVI leads to a positive increase of 18.2% for technology & electronics market. For the foods & drink sector the decrease is around 8.3%. Some variables were omitted because the sample is too small. This happens when an industry consists of only one company. This is for example the case for land reclamation, this industry only covers Boskalis. These results are shown, but not used to draw further conclusions.

Table 8. Industry differences.

The table shows the results of 16 Fama-MacBeth regressions. Every regression measures the influence of abnormal search volume on stock returns. Every column holds the results for a different industry. Standard errors are computed using the Newey-West (1987) formula with five lags.
* Significant at 1% level, ** significant at 5% level, *** significant at 10% level.

	Financial Services (1)	Insurances (2)	Chemicals (3)	Industrial metals (4)	Technology & Electronics (5)	Food & Drinks (6)
ASVI	-0.0279 (0.021)	0	-0.0001 (0.0001)	0.0421 (0.420)	0.1821 (0.1678)	-0.0825 (0.1498)
Log Marketcap	0.0028 (0.0013)*	0.0000 (0.0001)	0.0001 (0.0001)	-0.00146 (0.0019)	-0.0031 (0.0031)	0.0006 (0.0028)
Log Marketcap * ASVI	0.0015 (0.0017)	0.0000 (0.0002)	-0.0000 (0.0002)	-0.0031 (0.001)**	-0.0062 (0.0068)	0.0023 (0.0043)
Rmt - Rft	-0.0041 (0.0183)	0.0007 (0.0006)	0.0000 (0.0004)	0.0016 (0.0287)	0.0219 (0.0133)*	-0.0177 (0.0320)
SMBt	-0.0435 (0.0383)	0.0001 (0.0006)	-0.0002 (0.0004)	0.0511 (0.0264)*	-0.0001 (0.0227)	-0.3462 (0.3252)
HMLt	-0.1390 (0.1644)	0.0007 (0.0005)	0.0003 (0.0003)	0.0066 (0.0187)	0.0161 (0.0122)	0.3101 (0.3237)
Constant	-0.0408 (0.0419)	0	0	-0.0328 (0.0172)*	0.1136 (0.0827)	-0.0703 (0.1149)
Observations	1295	765	576	826	1691	1548

	Oil & Gas (7)	Real Estate (8)	Land reclamation (9)	Construction & Materials (10)	Lifestyle (11)	Transport (12)
ASVI	0.0005 (0.0002)**	-0.0009 (0.0013)	0	-0.0015 (0.0008)*	0	0.0648 (0.0531)

Log Marketcap	-0.0000	-0.0001	-0.0000	0.0000	-0.0001	0.0007
	(0.0001)	(0.0001)	(0.0000)	(0.0004)	(0.0001)	(0.0033)
Log Marketcap * SVI	-0.0002	0.0000	0.0000	0.0003	0.0001	-0.0014
	(0.0002)	(0.0001)	(0.0000)	0.0001)	(0.0007)	(0.0016)
Rmt – Rft	0.0007	-0.0016	0	0.0005	0	-0.0018
	(0.0029)	(0.0014)		(0.0028)		(0.0303)
SMBt	-0.0019	-0.0000	0	0.0036	0	0.0016
	(0.0019)	(0.0001)		(0.0033)		(0.0326)
HMLt	-0.0008	-0.0004	0	-0.0004	0	0.0451
	(0.0012)	(0.0012)		(0.003721)		(0.0347)
Constant	0.0008	0	0	0.0121	0	0.0447
	(0.0015)			(0.0112)		(0.0629)
Observations	879	899	258	943	69	1295

	Pharmaceutical (13)	Services (14)	Goods (15)	Agriculture (16)
ASVI	-0.0252	0	-0.0046	0
	(0.0355)		(0.0058)	
Log Marketcap	-0.0045	0	-0.0063	0
	(0.0018)**		(0.0033)**	
Log Marketcap *ASVI	0.0007	0.0000	-0.0040	0
	(0.0011)	(0.0000)	(0.0026)	
Rmt – Rft	-4344.82	0	0.0043	0
	(4051)		(0.0018)	
SMBt	-11357.9	0	-0.0012	0
	(11360)		(0.005)	
HMLt	-6232.08	0	-0.0007	0
	(6342)		(0.0034)	
Constant	14348.55	0	0.0068	0
	(13825)		(0.0042)	
Observations	1240	259	773	259

5.5 Hypothesis four

Lastly, the trading strategy is performed. Three different holding periods are tested, one month, three months and six months. The first holding period is one month. For example, the first portfolios are formed on 30 March 2016. The information period is one month for all portfolios. Every month new portfolios are formed, based on their abnormal search volume. Portfolio one contains the 25 most searched companies and portfolio three the 25 least searched. For each portfolio, the daily returns are regressed on the three Fama and French (1993) factors. Robust standard errors are used because this controls against heteroskedasticity. Now it is possible to use robust standard errors because the regression is linear. The results are reported in Table 9. The results show that the most searched portfolio gives positive returns, and the least searched portfolio gives negative returns. The strategy that goes long in portfolio one and short in portfolio three gives a daily abnormal return of 0.0724%. This would mean about 18.8% annually (measuring a 5-day trading week). Unfortunately, all coefficients are insignificant.

Table 9. Holding period one month.

The table gives the results for a momentum trading strategy with an information and holding period of both one month. The returns are regressed on the three Fama and French (1993) factors: the excess return on the market (Rm – Rf), the return difference between small and big stocks portfolios (SMB) and return difference between high and low book-to-market stocks portfolios (HML). Portfolio one holds the stocks with the highest search volume, while portfolio three holds the stocks with the lowest search volume. Robust standard errors are used.
* Significant at 1% level, ** significant at 5% level, *** significant at 10% level.

Portfolio	α	Rmt - Rft	SMBt	HMLt	R squared
One	0.0611	-0.0393	0.0936	-0.0649	0.0953
	(0.0981)	(0.0642)	(0.0100)	(0.0830)	
Three	-0.0113	-0.0573	0.0554	-0.0228	0.0837
	(0.0889)	(0.0477)	(0.0829)	(0.0955)	
One minus	0.0724	0.0180	0.0382	-0.0421	0.0482
three	(0.0653)	(0.0361)	(0.0621)	(0.0599)	

Next, the same is done, but for a holding period of three months. The results are shown in table 10. Both the returns of portfolios one and three are negative. The strategy gives a daily abnormal return of -0.0408%. Again, the coefficients are insignificant.

Table 10. Holding period three months.

The table shows the results for the trading strategy with an information period of one month and a holding period of 3 months. The returns are regressed on the three Fama and French (1993) factors: the excess return on the market (Rm – Rf), the return difference between small and big stocks portfolios (SMB) and return difference between high and low book-to-market stocks portfolios (HML). Portfolio one holds the stocks with the highest search volume, while portfolio three holds the stocks with the lowest search volume. Robust standard errors are used.
* Significant at 1% level, ** significant at 5% level, *** significant at 10% level.

Portfolio	α	Rmt - Rft	SMBt	HMLt	R squared
One	-0.0938 (0.2187)	-0.1886 (0.1484)	0.1722 (0.1841)	-0.0780 (0.1791)	0.0641
Three	-0.0530 (0.1901)	-0.1127 (0.1308)	0.0577 (0.1570)	-0.0916 (0.1469)	0.040
One minus Three	-0.0408 (0.1561)	-0.0758 (0.1669)	0.1145 (0.1643)	0.1465 (0.1157)	0.027

Lastly, the holding period of 6 months is tested. There are again three portfolios formed based on search volume. The information period remains one month but now the portfolios are hold for 6 months. The trading strategy gives a daily abnormal return of 0.0216% as shown in table 11. This would mean about 5.6% annually (measuring a 5-day trading week). Unfortunately, all coefficients are insignificant again.

Table 11. Holding period six months.

The returns for a trading strategy with 6 months holding period and one month formation period are represented. The returns are regressed on the three Fama and French (1993) factors: the excess return on the market (Rm – Rf), the return difference between small and big stocks portfolios (SMB) and return difference between high and low book-to-market stocks portfolios (HML). Portfolio one holds the stocks with the highest search volume, while portfolio three holds the stocks with the lowest search volume. Robust standard errors are used.
* Significant at 1% level, ** significant at 5% level, *** significant at 10% level.

Portfolio	α	Rmt – Rft	SMBt	HMLt	R squared
One	0.0506 (0.1202)	0.1970 (0.1215)	-0.0823 (0.0948)	0.0590 (0.1175)	0.2276
Three	0.0290 (0.1215)	0.0673 (0.1037)	-0.0495 (0.1048)	0.1083 (0.1711)	0.0967
One minus Three	0.0216 (0.1245)	0.1297 (0.0736)	-0.0328 (0.1339)	-0.0492 (0.1201)	0.2815

Comparing the returns to the STOXX600 gives an indication of how well the strategy is performing. The STOXX600 has had an average annual return of 6.94% in the last 5 years. The strategy with a holding period of one month highly outperforms the STOXX600. The other two holdings period do not. This shows that the strategy can only be valuable when used with the shortest holding period (one month).

Another interesting question is, is there a relationship between search and trading volume. Three portfolios are formed based on search intensity. For every portfolio, we sum the average abnormal trading volume and median abnormal trading volume. Portfolio 1 holds the stocks with the highest search volume and portfolio 3 the stocks with the lowest search volume. The results are presented in table 12. The results show no clear relationship. However, the trading volume of the most searched stocks is twice as little as the trading volume of the least searched stocks.

Table 12. Trading and search volume.

The first portfolio holds the stocks with the most search activity, while the third portfolio holds the stocks with the least search activity. The table shows average trading volume and median trading volume per portfolio. The second column shows average trading volume. The third column shows median trading volume.

Portfolio	Average trading volume	Median trading volume
1	-0.4743	-0.8925
2	-0.1174	-0.7189
3	-0.1222	-0.7928

5.6 Testing the past

We have seen a lot of variables being of influence on stock returns, but what is the influence of past stock returns on the present? Vozlyublennaia (2014) found that past returns determine the impact of attention on the future returns. So, researching this means the thought of winners winning even more and losers losing even more. This can be concluded from the autocorrelation. Autocorrelation measures the relationship between a variable's current value and its past values. This paper uses a lag of one week. The autocorrelation between this week's returns and last week's returns is measured using STATA. The autocorrelation is -0.0000714. This means that last week's returns have a small negative effect on this week's returns. The theory of Vozlyublennaia (2014), about winners winning even more and losers losing even more, cannot be confirmed.

A granger causality test is performed to test for simultaneity. Total abnormal search volume and stock returns are used as variables. The results are shown in table 13. The p-value of 0.947 shows that the lagged value of total abnormal search volume does not granger cause stock returns. The p-value of 0.897 shows that the lagged value of stock returns does not granger cause total abnormal search volume. The results shows that there is no simultaneity. However, a granger causality test does not test causality perfectly. The results can better be described as variables forecasting each other. We therefore cannot completely rule out simultaneity.

Table 13. Testing simultaneity.

The table holds the results of a granger causality test. The test is used to detect simultaneity. The results can be interpreted as follows: the excluded variable granger causes the equation variable if the p-value is equal or smaller then 0.05.

Equation	Excluded	P-value
Returns	Total abnormal search volume	0.947
Total abnormal search volume	Returns	0.897

CHAPTER 6 Conclusion

Unfortunately, most of the coefficients are insignificant. There are multiple possible reasons for this. Further research could try adding more control variables. A potential control variable could be advertising expenses. Unfortunately, no reliable information was available on this. Other potential control variables are a news event dummy, abnormal turnover and retail trading, measured as the 5-day trading volume divided by the total trading volume of the previous month. Not including these control variables, could lead to omitted-variable bias. Omitted-variable bias is one of the sources of endogeneity. The other ones are measurement error and simultaneity. A measurement error means that variables are not measured correctly. All data comes from reliable data sources and all data is processed using software and computer programs. Therefore, a measurement error is unlikely, but cannot be ruled out.

The last source of endogeneity is simultaneity. Simultaneity means that not only the independent variable has influence on the dependent variable, but also the other way around. In this study that would mean, companies with higher returns attracting more attention and more attention leading to higher returns. This is called circular reasoning. A granger causality test is performed to test for simultaneity. The results shows that there is no simultaneity. However, a granger causality test does not test causality perfectly. We therefore cannot completely rule out simultaneity. Selection bias could also be a problem when performing regressions. This paper investigated stocks from the Dutch stock market only. The Dutch stock market is not perfectly representative for all stocks, but the differences are probably not that big. If there is selection bias, it won’t make that much of a difference.

Future research could use different standard errors. A possibility could be using robust standard errors because they control for heteroskedasticity. Unfortunately, it was not possible to combine a Fama-MacBeth regression with robust standard errors in STATA. Therefore, this study chose Newey-West standard errors. Another option could have been, using white standard errors. However, white standard errors often lead to more insignificance and sometimes even accept a hypothesis that should have been rejected. Yet another recommendation for future research could be, trying to lag the independent variable by one period. This could help fix the endogeneity problem.

As said before, most of the coefficients are insignificant. Nevertheless, the results will be interpreted to draw conclusions. All hypotheses will be discussed separately. The first hypothesis suggested that higher search volume will lead to higher stock returns in the short run. The short run was defined as the first two weeks. The findings show that abnormal search volume has a positive effect on stock returns, of 1130%. Standardized search volume however has a negative effect, of 374%. Both percentages are abnormally high and low.

When checked for robustness, different results come up. In the first period (May 2016 – November 2018) abnormal search volume has a negative effect on stock returns and standardized search volume a positive effect. Here we see the exact opposite. In the second period (November 2018 – May 2021) the effects for abnormal search volume are again negative, while for standardized search volume close to zero. Because of these abnormal percentages and differences, it is hard to draw conclusions. The most obvious conclusion however is that abnormal search volume has a positive effect, while standardized search volume has a negative effect on stock returns in the short run. Hypothesis one can thus be accepted as far as abnormal search volume concerns. As regards to standardized search volume, the first hypothesis must be rejected.

The second hypothesis focusses on different investors. Are institutional investors less likely to be influenced? From the first regression, shown in table 4, we can derive that ticker searches have a positive effect on returns, while name searches have a negative effect on returns. Although it doesn't tell us much about who is influenced more, it does tell us that ticker searches are related to a price increase, while name searches are related to a price decline. Assuming that institutional investors are more likely to use ticker searches, we can conclude that they often search for stocks to buy them. On the other hand, retail investors are more likely to search company names and are therefore more likely to sell stocks after doing online investigation.

In table 7, both social engagement (Bloomberg) and abnormal search volume (Google) are regressed on stock returns. The results show that Bloomberg searches barely have any effect on stock returns and thus institutional investors are not that sensitive for media attention. On the contrary, retail investors are influenced (positively) by media. The second hypothesis can be accepted, institutional investors are less likely to be influenced.

The third hypothesis assumes that people value the content of media/news more, when they are less familiar with the industry. The industries are split up in two different categories to make a distinction between more and less familiar industries. The results show that positively the technology & electronics industry and the transport industry have the highest percentage. Abnormal search volume leads to a 18% increase in stock returns in the technology and electronics market. For the transport market, the increase is about 6%. Both industries fall under the daily encounter category. The industries with the most negative influence are the food & drinks industry and the financial services industry. Abnormal search volume about companies in the foods & drinks market leads to an 8% decrease in stock returns, while search volume for the financial services industry leads to a 2% decrease. Both industries are again part of the daily encounter category. The results clearly show that companies in industries people are more familiar with, are also more influenced by attention. The third hypothesis must be rejected.

The fourth, and final hypothesis tries to establish a new trading strategy. The strategy goes long in the portfolio holding the most searched stocks, while going short in the portfolio with the least searched stocks. Three different holdings periods are studied. The information period is always one month. The strategy with a holding period of one month gives a 19% annual return. The holding period of three months gives negative returns. Lastly, the holding period of six months gives an annual return of 5,6%. The one month holding period is the only strategy that outperforms the 7% annual return of the STOXX600. This strategy can thus be even more profitable than the benchmark. This profitability will decrease a lot, when taking trading cost into account. People can indeed use the new trading strategy to generate positive returns, but when taking transaction cost into account these returns will not be (much) higher than the benchmark. Trading costs vary between 0.75% to 1.5% per transaction. Every portfolio holds approximately 20 stocks. Around 70-80% of the stocks needs to be rebalanced every time when forming new portfolios. These transaction costs thus cause a large drop in the positive returns generated. The fourth hypothesis thus dependent on transaction cost. Overall, we can say that the trading strategy can lead to positive returns, which means that the fourth hypothesis can be accepted.

Now that all hypotheses have been discussed, we can formulate an answer to the main question:

Does investor attention affect stock returns, trading volume and volatility? If so, can this lead to a profitable trading strategy?

We can clearly say that investor attention influences stock returns. Abnormal search volume leads to positive returns, while standardized search volume leads to negative returns. The relationship between volatility and investor attention is not clear. The same holds for the relationship between trading volume and search volume. The results do show that the most searched stocks have an average trading volume twice as little as the least searched stocks, but no real conclusions can be drawn. Investor attention can be used to form a profitable trading strategy. The one month holding period strategy even outperforms the benchmark. When taking transaction costs into account, these positive returns may not be so promising anymore.

REFERENCES

Adachi, Y., Masuda, M., & Takeda, F. (2017). Google search intensity and its relationship to returns and liquidity of Japanese startup stocks. *Pacific-Basin Finance Journal*, 46, 243-257.

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management*, 25(3), 239-264.

Barber, M.B., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies*, 21(2), 785-818.

Beatty, S.E., & Smith, S.M. (1987). External Search Effort: An Investigation Across Several Product Categories. *Journal of Consumer Research,* 14(1), 83-95.

Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google Searches and Stock Returns. International Review of Financial Analysis, 45(C), 150-156.

Blitz, D., Huisman, R., Swinkels, L., & Van Vliet, P. (2020). Media attention and the volatility effect, Finance Research Letters, 36(10), 101317.

Bradley, P. (2010). Be where the conversations are: The critical importance of social media. *Business Information Review,* 27(4), 248-252.

Cesta, L. (2021). Social stock: how social media and influencers affect financial markets. *Boston University New Service.*

Colaco, H.M.J., Cesari, A.D., & Hegde, S.P. (2017). Retail Investor Attention and IPO Valuation. *European Financial Management*, 23(4), 691-727.

Cooper, M. (1999). Filter rules based on price and volume in individual security overreaction. *Review of Financial Studies*, 12(4), 901-935.

Cooper, M., Gutierrez Jr., R.C., & Hameed, A. (2004). Market states and momentum. *Journal of Finance*, 59(3), 1345-1365.

Da, Z., Engelberg, J., & Gao, P. (2009). In Search of Attention. *The Journal of Finance*, 66(5), 1461-1499.

Du, D., & Watkins, B. (2007). When competing momentum hypotheses really do not compete: How the sources of momentum profits change through time. *Journal of Economics and Business*, 59(2), 130-143.

Engelberg, J., Sasseville, C., & Williams, J. (2012). Market Madness? The Case of "Mad Money". *Management Science*, 58(2), 351-364.

Fama, E.F., and James D. MacBeth. (1973). Risk return and equilibrium: Empirical tests, *Journal of Political Economy,* 81, 607–636.

Fama, E.F., & French, K.R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.

French, K. R. (2021). *Current Research Returns*. Dartmouth data library.

https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Gervais, S., Kaniel., R., & Mingelgrin, D.H. (2001). The High-Volume Return Premium.

Journal of Finance, 56(3), 877-919.

Google About. (2015). Google. Geraadpleegd op 20 mei 2021, van

https://about.google/intl/ALL_nl/

Granger, C.W.J., & Poon, S. (2003). Forecasting Volatility in Financial Markets: A Review. *Journal of Economic Literature*, 41(2), 478-539.

Grullon, G., Kanatas, G., & Weston, J.P. (2004). Advertising, Breadth of Ownership, and Liquidty. *The Review of Financial Studies*, 17(2), 439-461.

Hou, K., Xiong, W., & Peng, L. (2009). A Tale of Two Anomalies: The Implications of Investor Attention for Price en Earnings Momentum. *SSRN Electronic Journal.*

Johnson, J. (2021, 12 March). Global market share of search engines 2010-2021. *Statista.* https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/#:~:text=Ever%20since%20the%20introduction%20of,share%20as%20of%20February%202021.

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.

Jun, S., Yoo, H.S., & Choi, S. (2018). Ten Years of Research Change Using Google Trends: From the Perspective of Big Data Utilizations and Applications. *Technological Forecasting and Social Change*, 130, 69-87.

Moat, H.S., Curme, C., Avakian, A., Kenett, D.Y., Eugene Stanley, H., & Preis, T. (2013). Quantifying Wikipedia Usage Patterns Before Stock Market Moves. *Scientific Reports*, 3(1801).

Morningstar Essentials. (2021). Morningstar. https://www.morningstar.nl/nl/

Newey, W., & West, K. (1987). A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, 55(3), 703-708.

Preis, T., Reith, D., & Eugene Stanley, H. (2010). Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions of the Real Society,* 368(1933), 5707-5719.

Scott, J., Stumpp, M., & Xu, P. (2003). News, Not Trading Volume, Builds Momentum. *Financial Analysts Journal*, 59(2), 45-54.

Stock, J., & Watson, M. (2011). *Introduction to Econometrics* (3rd ed.). Addison-Wesley.

Investing.com. (2016–2021, May). *STOXX 600 (STOXX)* [STOXX Europe 600 returns]. https://nl.investing.com/indices/stoxx-600-historical-data

Tantaopas, P., Padungsaksawasdi, C., & Treepongkaruna, S. (2016). Attention effect via internet search intensity in Asia-Pacific stock markets. *Pacific-Basin Finance Journal*, 38, 107-124.

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.

Tang, W., & Zhu, L. (2017). How security prices respond to a surge in investor attention: Evidence from Google Search of ADRs. *Global Finance Journal*, 33, 38-50.

Vlastakis, N., & Markellos, R. (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.

Yang, D., Ma, T., Wang, Y., & Wang, G. (2020). Does Investor Attention Affect Stock Trading and Returns? Evidence from Publicly Listed Firms in China. *Journal of Behavioral Finance.*

Zhang, B., & Wang, Y. (2015). Limited attention of individual investors and stock performance: Evidence from the ChiNext market. *Economic Modelling*, 50 (C), 94-104

APPENDIX A

Table 14. Data on volatility.

The data used for hypothesis one on volatility. The first column holds the ticker code, the second the volatility and the third the total abnormal search volume.

Ticker	Volatility	Total ASVI
Aegon	0.0588	-2.7268
Akza	0.0369	-886.8540
MT	0.0637	-0.1630
ASM	13.4158	0.1976
ASRN	0.0450	-1142.6686
BESI	0.0561	0.5932
HEIA	0.0302	-6.97586
IMCD	0.0347	-7.3841
INGA	0.0541	-0.6006
TWKY	0.0557	-1158.8261
AD	0.0311	-0.4100
DSM	0.0305	-0.6693
KPN	0.0387	-0.4609
PHIA	0.0337	-3.7233
NN	0.0397	-0.0879
PRX	0.0492	0.2235
RAND	0.0435	-0.5579
REN	0.0311	-1.0560
RDSA	0.0441	-2.9210
LIGHT	0.0499	0.09999
URW	0.0717	-4.8368
UNA	0.0257	0.13406
WKL	0.0281	-12.98571
AALB	0.0479	-1025.1176
ABN	0.0562	-0.4250
AF	0.0711	-0.2029
AMG	0.0672	0.2305
APAM	0.0516	-1.9691
ARCAD	0.0561	-9.8977
BFIT	0.0639	-3.6473
BOKA	0.0447	0.4111

CRBN	0.0385	1.2572
ECMPA	0.0632	-324.8207
FAGR	0.0433	-136.0091
FLOW	0.0490	0.4030
FUR	0.0728	-0.2002
GLGP	0.0608	-108.8820
GNVN	0.0324	-449.2256
INTER	0.0403	-1.5219
JDEP	0.0351	-176.4984
PHARM	0.0440	-0.8400
PNL	0.0309	-1.7317
SBMO	0.0464	-691.75315
TWEKA	0.0486	-102.87642
VPK	0.0311	-1.7189
WDP	0.0314	-4.5799
ACCEL	0.0505	-5.6716
AXS	0.1026	-0.7051
ACOMO	0.0308	-0.0684
AJAX	0.0337	-5.3772
AVTX	0.0790	-567.3096
BSGR	0.0648	-162.9285
BAMNB	0.0637	-29.7643
BRNL	0.0491	-598.7670
CMCOM	0.0634	-363.0534
FFARM	0.0343	-748.5425
HEIJM	0.0597	-807.6191
HDG	0.0335	-2.5995
KENDR	0.0526	-8.6907
BOLS	0.0353	-1.1009
NEDAP	0.0381	-3.2256
ORDI	0.0540	-0.7322
SIFG	0.0544	-907.6943
SLIGR	0.0492	-145.3030
TOMTOM	0.0699	-112.5133
VLK	0.0423	-1.4331
VASTN	0.0470	-204.0111

VVY	0.1229	-485.0963
WHA	0.0594	0.5732