Master's thesis

2020

Amanda Borge Byrkjeland, Mette Liset

NTNU
Norwegian University of
Science and Technology
Faculty of Economics and Management
Department of Industrial Economics and Technology
Management

Master's thesis

NTNU
Norwegian University of
Science and Technology

Amanda Borge Byrkjeland
Mette Liset

Predicting stock returns using Google Trends

June 2020

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of
Science and Technology

Predicting stock returns using Google Trends

Amanda Borge Byrkjeland
Mette Liset

Industrial Economics and Technology Management
Submission date: June 2020
Supervisor: Peter Molnár

Norwegian University of Science and Technology
Department of Industrial Economics and Technology Management

Preface

This Master's thesis examines the predictability of stock returns using Google Trends. The thesis concludes our Master of Science in Industrial Economics and Technology Management within Financial Engineering at the Norwegian University of Science and Technology (NTNU) in the spring of 2020.

We would like to thank our supervisor Peter Molnár, Associate Professor at the Norwegian University of Science and Technology, for helpful guidance and constructive feedback. In addition, we thank Zhi Da, Professor of Finance at the University of Notre Dame, for sharing with us parts of the dataset used in the highly acknowledged paper ”In Search of Attention” (Da et al., 2011).

Trondheim, June 18, 2020
Amanda Borge Byrkjeland, Mette Liset

Abstract

Some research investigating the relationship between Google search volume and stock returns finds that increased search volume predicts higher returns, while other papers draw the opposite conclusion. We reinvestigate this relationship using Fama-Macbeth cross-sectional regressions for the Russell 3000 companies with the use of either stock ticker or company name as Google search keyword while controlling for several other variables such as the number of analysts following the company. We find a positive relationship between search volume and stock return in the period from 2004 to 2008, and a negative relationship in the period from 2009 to 2019. While searches for the stock ticker predict returns better than searches for company name from 2004 to 2008, the opposite is true from 2009 to 2019. We evaluate the economic significance of our results by a trading strategy built upon the same Fama-Macbeth cross-sectional regressions. A trading strategy where we buy the 50% stocks with the highest predicted abnormal return and short the 50% stocks with the lowest predicted abnormal return delivers a yearly abnormal return of 11.3% after accounting for transaction costs, while a similar strategy based on buying the 5% of stocks with the highest and shorting the 5% with the lowest predicted returns delivers an impressive abnormal return of 30.6% after transaction costs.

Sammendrag

Blant forskere som undersøker forholdet mellom Googles søkevolum og aksjeavkastning, finner noen at økt søkevolum spår høyere avkastning, mens andre trekker den motsatte konklusjonen. Vi undersøker dette forholdet ved å bruke Fama-Macbeth tverrsnittsregresjoner for Russell 3000-selskapene ved bruk av enten aksjetikker eller selskapsnavn som Google-søkeord, samtidig som vi kontrollerer for flere andre variabler som antall analytikere som følger selskapet. Vi finner et positivt forhold mellom søkevolum og aksjeavkastning i perioden 2004 til 2008, og et negativt forhold i perioden 2009 til 2019. Søk etter aksjetikker er bedre til å forutsi avkastning enn søk etter selskapsnavn fra 2004 til 2008, mens det motsatte gjelder fra 2009 til 2019. Vi vurderer den økonomiske betydningen av resultatene våre med en tradingstrategi bygd på den samme Fama-Macbeth tverrsnittsregresjonen. En tradingstrategi der vi kjøper de 50% av aksjene med høyest antatt unormal avkastning og shorter de 50% av aksjene med lavest antatt unormal avkastning, gir en årlig unormal avkastning på 11,3% etter inkludering av transaksjonskostnader, mens en lignende strategi basert på å kjøpe de 5% av aksjene med høyest og shorte de 5% av aksjene med lavest antatt avkastning gir en imponerende unormal avkastning på 30,6% etter transaksjonskostnader.

Table of Contents

Preface **i**

Abstract **i**

Sammendrag **i**

Table of Contents **iii**

List of Tables **iv**

List of Figures **v**

1 Introduction **1**

2 Data **5**
2.1 Search volume variables 7
2.2 Abnormal return 8
2.3 Volatility 8
2.4 Abnormal turnover 9
2.5 Market capitalization 10
2.6 Advertising expense/sales 10
2.7 Number of analysts 10
2.8 Stationarity 10
2.9 Summary statistics 11

3 Methodology **13**
3.1 Fama-Macbeth cross-sectional regression 14

4 Results **15**
4.1 Predicting stock return 15
4.2 Comparing stock ticker and company name as search keyword 19
4.3 Predicting volatility and turnover 22

5 Trading strategies **27**
5.1 The value of Google Trends 28
5.2 Training windows 29

5.3 Risk exposure 30
5.4 Trading costs 31
5.5 Trading only stocks with extreme predicted returns 32
5.6 Long or short portfolio? 33

6 Conclusion **35**

Bibliography **36**

Appendix **41**
A Predicting volatility and turnover with stock ticker 41

List of Tables

2 Data

2.1 Definitions of financial and search volume variables 6
2.2 Correlation matrix for all variables 11

4 Results

4.1 FM regression for AbnReturn with AbnormalTickerSVI, 2004-2008 . . . 17
4.2 FM regression for AbnReturn with AbnormalTickerSVI, 2009-2019 . . . 18
4.3 FM regression for AbnReturn with AbnormalTickerSVI and AbnormalNameSVI, 2004-2008 20
4.4 FM regression for AbnReturn with AbnormalTickerSVI and AbnormalNameSVI, 2009-2019 21
4.5 FM regression for Volatility with AbnormalNameSVI, 2004-2008 23
4.6 FM regression for Volatility with AbnormalNameSVI, 2009-2019 24
4.7 FM regression for AbnTurnover with AbnormalNameSVI, 2004-2008 . . 25
4.8 FM regression for AbnTurnover with AbnormalNameSVI, 2009-2019 . . 26

5 Trading strategies

5.1 Return of strategies with different sets of regressors. 28
5.2 Exposure to risk factors 30
5.3 Yearly average return adjusted for trading costs 31
5.4 Alpha, volatility, and Sharpe ratio for different thresholds 32
5.5 Comparing long and short portfolios 33

Appendix

A1.1 FM regression for Volatility with AbnormalTickerSVI, 2004-2008 42
A1.2 FM regression for Volatility with AbnormalTickerSVI, 2009-2019 43
A1.3 FM regression for AbnTurnover with AbnormalTickerSVI, 2004-2008 . . 44
A1.4 FM regression for AbnTurnover with AbnormalTickerSVI, 2009-2019 . . 45

List of Figures

2 Data

2.1 Plot of scaled Google search volume 8

5 Trading strategies

5.1 Cumulative return of trading strategies comparing different training windows. 29

5.2 Cumulative return for portfolios with different thresholds 33

5.3 Cumulative return for long and short only portfolios with different thresholds 34

Chapter 1

Introduction

For a long time, researchers have acknowledged that stock markets are driven not only by rational investors (Baker and Wurgler, 2007). Researchers in behavioral finance has been working to augment the standard asset-pricing model to include investor sentiment. Investor sentiment is broadly defined as ”a belief about future cash flows and investment risks, that is not justified by the facts at hand” (Baker and Wurgler, 2007). However, in the past, investor sentiment could only be studied indirectly, as direct measures were hard to come by.

One of the most popular proxies for investor sentiment is investor attention. Direct measures of investor attention are also difficult to obtain, but well known indirect measures for investor attention are broadly researched. This includes extreme returns (Barber and Odean, 2007), trading volume (Barber and Odean, 2007; Gervais et al., 2001; Hou et al., 2009), news and headlines (Barber and Odean, 2007; Yuan et al., 2008), advertising expense (Chemmanur and Yan, 2019; Grullon et al., 2004; Lou, 2014), and price limits (Seasholes and Wu, 2007). However, a news article or advertisement does not guarantee attention. Information supply is rapidly growing, while attention is a scarce resource.

The increased popularity and technical advances of online services has allowed researchers to access several direct measures of investor attention, such as search engine volume and website traffic. After Google made search data publicly available through Google Trends in 2008, search volume has become a popular proxy for investor attention. Internet users usually use a search engine to collect information, and Google continues to be the favorite. Indeed, as of May 2020, Google accounted for 88.2% of all search queries performed in the United States (Statscounter, 2020). If somebody searches for a term on Google, they are undoubtedly paying attention to it; thus the changes in the search volume can be used as evidence of the changes in the attention.

As a result of more comprehensive data made available by Google, the number of empirical studies investigating the relationship between Google search volume and the stock market performance has increased in recent years, see Da et al. (2011); Vlastakis and Markellos (2012); Bijl et al. (2016). Da et al. (2011) find that increased search volume predicts higher stock returns in the next two weeks, and Joseph et al. (2011) find that, over a weekly horizon, Google searches for company ticker predict stock returns. On the other hand, Bijl et al. (2016) find that increased Google search volume predicts negative returns. Challet and Ayed (2014) and Kim et al. (2019) find that Google search volumes are unable to predict future returns. The inconsistent findings might be caused by research on different periods, samples consisting of companies from different indices, or different keywords used for measuring search volume for a company. While Da et al. (2011) are using data from January 2004 to June 2008 on the stock ticker for the Russell 3000 companies, Bijl et al. (2016) are using data from January 2008 to December 2013 on the company names for the S&P 500 companies.

In research on Google search volume and financial markets both stock ticker and company name are frequently used as search keyword. Da et al. (2011), Joseph et al. (2011), Pancada (2017), Kristoufek (2013), Ding and Hou (2015) and Baker and Wurgler (2007) use searches for stock ticker as a proxy for investor attention. Da et al. (2011) conclude that searches for ticker capture the attention of people in search of financial information about a given stock. Joseph et al. (2011) state that the effort required to process the results of a ticker query is only worthwhile for someone who is seriously considering an investment decision. Baker (2016) argue that when searching for company information, ”entering the entire company name will generate interest data that are not exclusively a result of earnings expectations”, and they use stock ticker ”to control for this possible contemporaneous interest result”. Kristoufek (2013) use both the ticker symbol alone and the combination of the word ”stock” and the ticker symbol to ensure that the searched term is not misinterpreted as the ticker symbol. Both Challet and Ayed (2014) and Kim et al. (2019) use a combination of ticker and company name.

Bijl et al. (2016), Preis et al. (2010), Vlastakis and Markellos (2012), Bank et al. (2011) and Moussa et al. (2017) use searches for company name as a proxy for company attention. Bank et al. (2011) believe that the use of searches for company name will capture the extent of attention the company is receiving from a much broader, and potentially relevant audience. Vlastakis and Markellos (2012) state that the use of company name measures investor attention related to the company in general, rather than only to the stock, in addition to avoiding the problems associated with tickers having generic meanings. Moussa et al. (2017) justify the choice by stating that market participants tend to type the stock name because it is easier and simpler since stock tickers are not very known by people. Bijl et al. (2016) conclude that company name searches have a stronger relationship to stock market returns than ticker searches.

The use of company name requires a lot more data cleaning than using stock ticker. As far as we know, the largest sample used with company name is S&P 500 (Bijl et al., 2016; Vlastakis and Markellos, 2012). Both Da et al. (2011) and Bank et al. (2011) conclude that the relationship between search volume and stock return is stronger for smaller companies, so by extending samples also to include smaller companies we can expect stronger results.

To find out whether the inconsistent results for predicting stock returns are due to time periods, size of companies, search keywords, or maybe all of them, we start by studying the Russell 3000 companies looking at the same period as Da et al. (2011) from 2004 to 2008. We then study the same companies in the period from 2009 to 2019. The results show that increased search volume has a positive impact on stock return before 2008 in line with Da et al. (2011), while the results after 2008 show that search volume has negative predictive power, in line with Bijl et al. (2016). We study both search volume for stock ticker and company name in both periods to see if any of them can outperform the other. For the period after 2008 searches for company name perform better than searches for stock ticker.

In addition to studies investigating the relationship between Google search volume and stock returns, several papers study the predictive power of Google search volume for stock volatility and trading volume. Vlastakis and Markellos (2012) find that search volume has a positive association with volatility and trading volume at the individual stock level. Fink and Johann (2014) find that volatility and trading volume of stocks increase on days with high search volume, and Aouadi et al. (2013) conclude that higher stock-specific search volume leads to higher volume, but has mixed impact on volatility. Bank et al. (2011) find a positive predictive relationship between increased search volume and trading volume, and Preis et al. (2010) find that increasing transaction volumes of stocks coincide with an increasing search volume and vice versa. We investigate these relationships and find a negative relationship between search volume and volatility and a positive relationship between search volume and trading volume.

To test the economic significance of our results, we create a trading strategy based on Google search volume. In earlier research, Kristoufek (2013) create a trading strategy based on Google search volumes that beats the Dow Jones index, and Bijl et al. (2016) create a trading strategy based on Google search volume that is profitable without transaction costs. We create a market neutral portfolio where we buy the top 50% stocks with highest predicted abnormal return and short sell the bottom 50% with lowest predicted abnormal return. We test for different number of weeks with training data and conclude that 52 weeks of training gives the best performance. After inclusion of trading costs, our best portfolio gives an abnormal return of 11.3%. We also construct portfolios only consisting of stocks with extreme high or low predicted abnormal return to confirm that our model is able to predict extreme returns. By constructing long only and short only portfolios we conclude that our prediction of extreme positive abnormal return is more accurate than the prediction of extreme negative abnormal return.

The rest of the thesis is organized as follows: chapter 2 describes data collection and preprocessing, chapter 3 contains the methodology, and chapter 4 presents our results. We use Fama-Macbeth cross-sectional regressions to investigate the relationships between search volume and individual stock performance. Chapter 5 contains an analysis of the different trading strategies, while chapter 6 summarizes our key findings.

Chapter 2

Data

We have gathered data on the Russell 3000 companies. Following Da et al. (2011), we collect price data, trading volume, advertising expenses, sales, number of analysts following the company, the number of shares outstanding, and Google search volumes. Table 2.1 shows an overview of all variables. Like Da et al. (2011), to eliminate survivorship bias and the impact of index addition and deletion, we examine all stocks ever included in the index during our sampling period. We extend the period used by Da et al. (2011) with 11 years, now containing data from January 2004 to December 2019. Different availability of the different variables leaves us with an unbalanced dataset.

We collect search volume from the Google Trends webpage. Google Trends is a service by Google that offers users the ability to visualize the relative popularity of a keyword over time, as well as the opportunity to compare the popularity of one keyword with another. The data is not presented in absolute numbers; rather, it is scaled from 0 to 100, where 100 represents the maximum popularity during the time period chosen. Each data point is also divided by the total searches of the geography and time range it represents to remove time effects. This output from Google Trends is called the search volume index (SVI).

For small periods, trends data can be collected daily, but if there are not enough searches for a search term, Google Trends will return a zero value for that term's SVI. We are, therefore, using weekly financial data and search volume data to collect enough data for searches on tickers and company names for as many of the companies as possible. Google Trends is only showing data on a weekly level for up to five years at a time, and earlier research is mainly limited to either five year sampling period or monthly search volume (see e.g., Bui and Nguyen (2019)). To overcome this limitation, we are downloading data in periods of 4 years at a time with 20 weeks overlapping. We are then scaling the data using a ratio calculated by using the data from the 20 weeks overlap. See Figure 2.1 for a visualization of the scaling across time.

Daily financial data for the companies are obtained from Wharton Research Data Services (WRDS) and Thomson Reuters Eikon. Specifically, we collect daily open, close, high, low, cumulative factor to adjust price and share volume for each company from The Center for Research in Security Prices (CRSP) through WRDS. From Compustat through WRDS, we collect advertising expenses and sales, and from the I/B/E/S Database through Thomson Reuters Eikon, we collect the number of analysts covering each company. We also obtain weekly values of Fama-French's three factors from French's online data library.

We use weeks starting on Sunday for SVI and weeks starting the first trading day in each week when calculating the financial variables. We focus on the time period from 2004 to 2019 due to the data available from Google (SVI).

Variable	**Definition**	**Source**
$AbnormalTickerSVI$	The logarithm of current week aggregate search frequency from Google Trends based on stock ticker, minus the logarithm of the median search frequency from the last eight weeks, minus the average across all companies.	Google Trends
$AbnormalNameSVI$	The logarithm of current week aggregate search frequency from Google Trends based on company name, minus the logarithm of the median search frequency from the last eight weeks, minus the average across all companies.	Google Trends
$AbnReturn$	Weekly actual stock return minus the expected return from Fama French 3-factor model.	CRSP
$Volatility$	Volatility estimated using the Garman and Klass (1980) volatility estimator.	CRSP
$AbnTurnover$	Weekly trading volume over shares outstanding minus the median of weekly trading volume over shares the last eight weeks.	CRSP
$MarketCap$	The logarithm of share price multiplied by number of shares outstanding.	CRSP
$NoAnalysts$	The logarithm of 1 + the number of analysts covering the company.	I/B/E/S
$XadSales$	The ratio of advertising expense over sales, from the previous fiscal year.	Compustat
$MC * AbnormalTickerSVI$	$MarketCap$ multiplied with $AbnormalTickerSVI$.	CRSP/ Google Trends
$MC * AbnormalNameSVI$	$MarketCap$ multiplied with $AbnormalNameSVI$.	CRSP/ Google Trends

Table 2.1: Variables definition for Google Trends variables and financial variables calculated.

2.1 Search volume variables

When measuring investor attention using Google search volume, an important decision is what to use as search keywords. We use both the company's stock ticker (e.g., AAPL is the company ticker for Apple) and the company name as search keyword to see if one of them is better.

Da et al. (2011) argue that searches for company name is a bad proxy of attention and that it is better to use the company's ticker. There are several concerns with using ticker as a measure of investor attention, but it is still frequently used. Pancada (2017) give us three main reasons why stock tickers should be used as search keywords over company names. First, the ticker is a unique identifier and, therefore, avoids the issues with multiple reference names. Second, only people interested in financial information would search for the ticker, and third, the ticker is easy to obtain from a search engine or the news. A problem with using stock ticker is that some companies have tickers with alternative meanings and some of the companies have one or two-letter stock tickers with generic meaning such as "C" (Citigroup Inc) or "CA" (Carrefour SA).

Vlastakis and Markellos (2012) use company name as search keyword for two main reasons. First, using the company name avoids the problem with many tickers having alternative or generic meaning. Second, search volume for the company name is a better measure of investor attention related to the firm in general rather than only to the stock. We obtain the company names used as search keywords by following the method of Vlastakis and Markellos (2012). We use Google Trends to compare the full company name to other variations known to us (including abbreviations), and then choose the keyword with the largest search volume.

For both the stock ticker and the company name, we calculate the abnormal search volume, $AbnormalTickerSVI$ and $AbnormalNameSVI$. We start by subtracting the median SVI for the last eight weeks for each company i:

$$\begin{aligned} RawAbnormalTypeSVI_{i,t} &= log(TypeSVI_{i,t}) \\ &\quad - log[Med(TypeSVI_{i,t-8}, ..., TypeSVI_{i,t-1})] \end{aligned} \tag{2.1}$$

where $Type$ is a placeholder for either $Ticker$ or $Name$.

Following Da et al. (2011) we cross-sectionally demean the $RawAbnormalTypeSVI$ by subtracting the week's average abnormal SVI across all companies, as shown in Equation 2.2.

$$\begin{aligned} AbnormalTypeSVI_{i,t} &= RawAbnormalTypeSVI_{i,t} \\ &\quad - Avg_i(RawAbnoralTypeSVI_{i,t}) \end{aligned} \tag{2.2}$$

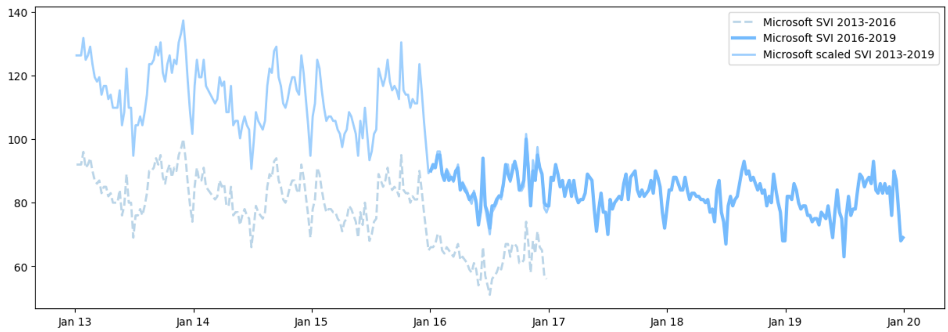

Figure 2.1: Search volume index (SVI) for "Microsoft" from 2013-2016 and 2016-2019. The thin blue line show the search volume after scaling.

2.2 Abnormal return

When calculating abnormal return, we are using the Fama–French three-factor model to obtain weekly expected return. We calculate the firm-specific Fama-French betas by running a linear regression with a rolling window of 2 years (104 weeks), with the three factors; market return, $\beta_{MKT-R_f,t}$, small minus big, $\beta_{SMB,t}$, and high minus low, $\beta_{HML,t}$, as regressors. The linear models are estimated using the following equation:

$$\begin{aligned} RawReturn_t = \alpha + R_{R_f,t} + \beta_{MKT-R_f,t} * R_{MKT-R_f,t} \\ + \beta_{SMB,t} * R_{SMB,t} + \beta_{HML,t} * R_{HML,t} \end{aligned} \tag{2.3}$$

where $RawReturn_t$ is calculated by $log(\frac{O_{t+1}}{O_t})$, where O_t is the adjusted open price for first trading day in week t.

Expected return is then given by:

$$\begin{aligned} ExpReturn_t = R_{R_f,t} + \beta_{MKT-R_f,t} * R_{MKT-R_f,t} \\ + \beta_{SMB,t} * R_{SMB,t} + \beta_{HML,t} * R_{HML,t} \end{aligned} \tag{2.4}$$

We then detract the expected log returns from the actual log returns to obtain weekly abnormal returns.

$$AbnReturn_t = RawReturn_{t+1} - ExpReturn_{t+1} \tag{2.5}$$

2.3 Volatility

We use the Garman and Klass (1980) volatility estimator adjusted for opening jumps. In Molnár (2012) this estimator is recognized as the best range-based volatility estimator for the purpose of standardizing returns. The following formula is used to estimate daily variance for day d:

$$\sigma_d^2 = \frac{1}{2}(h_d - l_d)^2 - (2log(2) - 1)c_d^2 - j_d^2 \tag{2.6}$$

with:

$$\begin{aligned} c_d &= log(close_d) - log(open_d), \\ l_d &= log(low_d) - log(open_d), \\ h_d &= log(high_d) - log(open_d), \\ j_d &= log(aopen_d) - log(aclose_{d-1}), \end{aligned} \tag{2.7}$$

Weekly variance for week t is calculated as:

$$\sigma_t^2 = \sum_{d \in t} \sigma_d^2 \tag{2.8}$$

Finally, weekly volatility for week t is calculated as:

$$Volatility_t = \sqrt{\sigma_t^2} \tag{2.9}$$

where $high_d$ and low_d are the highest and lowest realized price on day d. The opening and closing price on the given day are defined as $open_d$ and $close_d$, and $aopen_d$ $(aclose_d)$ is the adjusted opening (closing) price.

2.4 Abnormal turnover

To measure abnormal trading volume, we start by calculating share turnover, $Turnover$, using Equation 2.10.

$$Turnover_t = \frac{Volume_t}{ShrOut_t}, \tag{2.10}$$

where $ShrOut_t$ is the company's total number of outstanding shares in week t and $Volume_t$ is the total number of shares traded this week. We then calculate abnormal trading volume:

$$AbnTurnover_t = log(Turnover_t) - log[Med(Turnover_{t-8}, ..., Turnover_{t-1})], \tag{2.11}$$

where $log[Med(Turnover_{t-8}, ..., Turnover_{t-1})]$ is the logarithm of the median for the previous eight weeks.

2.5 Market capitalization

$MarketCap$ is calculated by taking the logarithm of the market capitalization for each company. We find the market capitalization by multiplying the open price, O_t, of the stock by the company's total number of outstanding shares, $ShrOut_t$.

$$MarketCap_t = log(O_t * ShrOut_t) \tag{2.12}$$

2.6 Advertising expense/sales

The advertising-expense-over-sales-ratio is calculated using data from the previous fiscal year sourced from Compustat. Following Da et al. (2011), we set advertisement expense, $AdvertisingExpense$, to zero if it is not reported. For instance, Compustat does not report advertisement expenses for utility companies. The weekly advertising-expense-over-sales-ratio equals the yearly ratio and is given by:

$$XadSales_t = \frac{AdvertisingExpense_{t-1}}{Sales_{t-1}} \tag{2.13}$$

2.7 Number of analysts

The number of analysts reported by I/B/E/S for each company is the number of analysts who provide an earnings per share estimate for the next financial year for this company. We use this number to calculate $NoAnalysts$:

$$NoAnalysts_t = log(1 + Number\ of\ analysts_t) \tag{2.14}$$

2.8 Stationarity

The log-median transformation for calculating abnormal search values are done to remove possible trends from the Google search volume data. We do this to generate stationary time-series to avoid that variables are associated but not causally related. After this transformation, we test for stationarity using a Fisher type unit root test for panel data by using the built-in Stata command *xtunittest fisher*. The Fisher type test is using the augmented Dickey-Fuller test on each panel and allows unbalanced panel data. Rejection of the null hypothesis indicates stationarity. The tests indicate stationarity for all financial variables and search volume variables after the log-median transformation.

2.9 Summary statistics

To make regression coefficients easily comparable, we standardize all variables to have zero mean and a standard deviation of one for each company. Correlation coefficients between the variables can be seen in Table 2.2. We follow the same method as Da et al. (2011) when calculating the correlation. First, we calculate correlations individually for each company, and then we average the results across all companies. We do this for the time period from 2004 to 2019 at a weekly frequency.

From Table 2.2, we see that in general, the correlations between the search volume variables and the other variables are low. The correlation between $AbnormalTickerSVI$ and $AbnormalNameSVI$ is 3.1%. The low correlation shows that people may search for ticker and company name with a different motivation. Both extreme returns and trading volume are used as proxies for investor attention, but their correlation with each other is 3.5%. Both abnormal return and abnormal turnover have a low correlation with the two search volume variables. As stated in Da et al. (2011), the low correlation may be due to the fact that both returns and turnover are equilibrium outcomes that are functions of many economic factors in addition to investor attention.

	AbnormalTickerSVI	AbnormalNameSVI	AbnReturn	Volatility	AbnTurnover	MarketCap	NoAnalysts
AbnormalNameSVI	0.031						
AbnReturn	0.009	0.009					
Volatility	0.049	0.053	0.079				
AbnTurnover	0.071	0.081	0.035	0.402			
MarketCap	0.001	0.008	-0.126	-0.265	-0.005		
NoAnalysts	0.002	0.004	0.021	-0.072	-0.004	0.266	
XadSales	0.001	-0.003	0.000	0.009	0.000	-0.051	-0.036

Table 2.2: Correlation matrix for the variables included in the dataset. The correlation is found by first calculating correlations individually for each company, and then averaging the results across all companies.

Chapter 3

Methodology

We study if Google search volume can predict individual stock performance. Following Da et al. (2011), we first run Fama and MacBeth (1973) cross-sectional regression for the Russell 3000 dataset on the period from 2004 to 2008. We set abnormal return as the dependent variable, and include the abnormal search volume variable, $AbnormalTickerSVI$, and the other attention measures as independent variables. To see if the findings from Da et al. (2011) hold in the present time, we run the same models on the period from 2009 to 2019.

Both searches for ticker and searches for company name have been used in research to study the relationships between search volume and financial markets. Therefore, we also run the Fama-Macbeth models comparing the performance of the two search volume variables, which are based on different keyword choices. We do this for both time periods.

We also study if search volume can predict volatility and turnover and whether these effects have changed over time. We do this by running the same Fama-Macbeth models with volatility and turnover as dependent variables, for both time periods.

For all model specifications, we run both a simple model, including only the search volume variables and a model, including other control variables. This is to make sure that the relationships we may find between search volume variables and financial markets are not only due to the inclusion of specific control variables. In addition, the inclusion of control variables will allow us to compare the performance of the search volume variables to the other attention measures.

3.1 Fama-Macbeth cross-sectional regression

The Fama and MacBeth (1973) cross-sectional regression is a two-step procedure. The first step involves estimation of one cross-sectional regression for each time period, and the second step involves calculating the average of the coefficients from the T cross-sectional regressions. We use the cross-sectional regression specifications shown in Equation 3.1 and Equation 3.3, and calculate the time-average as shown in Equation 3.2 and Equation 3.4 to get the Fama-Macbeth coefficient estimates.

Equation 3.1-3.2 represents the simple models only including the search volume variables, while Equation 3.3-3.4 also includes other control variables. In the specifications $dependent$ is a placeholder for either $AbnReturn$, $Volatility$ or $AbnTurnover$, and $Type$ is placeholder for $Ticker$ or $Name$. We run the models from one week lag to five weeks lag between the dependent and independent variables, and u indicates the lag in the specific model. For abnormal return, instead of regressing the return five weeks ahead, we regress the abnormal return from five to 52 weeks ahead on the independent variables. For volatility and turnover, calculating the abnormal value from five to 52 weeks ahead, would not make sense in the same way, as there is no expected volatility or turnover to compare with.

$$\begin{aligned} dependent_t = c_{0,t} &+ c_{1,t} AbnormalTypeSVI_{t-u} \\ &+ c_{2,t} MC * AbnormalTypeSVI_{t-u} \end{aligned} \tag{3.1}$$

$$\hat{c_j} = \frac{1}{T}\sum_{t=1}^{T} \hat{c_{j,t}} \qquad \text{for } j = 0, 1, 2 \tag{3.2}$$

$$\begin{aligned} dependent_t = c_{0,t} &+ c_{1,t} AbnormalTypeSVI_{t-u} \\ &+ c_{2,t} MC * AbnormalTypeSVI_{t-u} \\ &+ c_{3,t} MarketCap_{t-u} + c_{4,t} Absolute\ AbnReturn_{t-u} \\ &+ c_{5,t} XadSales_{t-u} + c_{6,t} NoAnalysts_{t-u} \\ &+ c_{7,t} AbnTurnover_{t-u}{}^{1} \end{aligned} \tag{3.3}$$

$$\hat{c_j} = \frac{1}{T}\sum_{t=1}^{T} \hat{c_{j,t}} \qquad \text{for } j = 0, 1, ..., 7 \tag{3.4}$$

[1]Not included for models with $AbnTurnover$ as dependent variable

Chapter 4

Results

In this chapter, we first present the results from the Fama-Macbeth regressions trying to replicate Da et al. (2011) on predicting abnormal return using the Russell 3000 companies in the time period from 2004 to 2008. Then we present the results from the same models run on the time period from 2009 to 2019. Third, we present the results comparing ticker and company name as search volume variables. Lastly, we present the results from the same models predicting volatility and turnover.

Our results are indeed in line with Da et al. (2011), showing that increased search volume predicts increased abnormal return the next weeks, in the period from 2004 to 2008. Studying the same in the period from 2009 to 2019 shows us that increased search volume here predicts decreased abnormal return the next weeks. When comparing ticker and company name as keyword, our results show that ticker is the strongest predictor in the first time period, while company name is the strongest in the second time period. We find a negative relationship between search volume and volatility, and a positive relationship between search volume and turnover.

Since all variables are standardized, all reported coefficients are standardized, and therefore they can be compared across models.

4.1 Predicting stock return

Overall, we find even stronger relationships between the search volume variables and abnormal return than Da et al. (2011). Comparing the results in the two time periods, we see that searches for ticker had positive predictive power for the next weeks' return in 2004-2008, while negative predictive power in 2009-2019. Both coefficients and significance levels are similar for the simple models and the models including control variables, indicating that our results are strong and not just valid for a specific combination of control variables.

As seen in Table 4.1, $AbnormalTickerSVI$ has positive predictive power for $AbnReturn$ at time horizons from one week ahead to four weeks ahead, in the time period from 2004 to 2008. On the other hand, the search volume variable weighted by $MarketCap$ has negative predictive power for $AbnReturn$. Interpreting these results as shown in Equation 4.2, rather than the obvious interpretation shown in Equation 4.1, indicates that company size matters, and that the positive relationship between search volume and return is strongest for the smaller companies. This is visible through the opposite signs of the two coefficients, and knowing that $MarketCap$, due to standardization, is positive for the biggest companies and negative for the smallest companies.

$$
\begin{aligned}
AbnReturn_t = {} & c_{1,t} AbnormalTickerSVI_{t-u} \\
& + c_{2,t} MC * AbnormalTickerSVI_{t-u}
\end{aligned}
\tag{4.1}
$$

$$
AbnReturn_t = (c_{1,t} + c_{2,t} MarketCap) AbnormalTickerSVI_{t-u} \tag{4.2}
$$

The results in Table 4.1 are in line with Da et al. (2011), which finds that an increase in searches for ticker predicts higher stock returns for the next weeks and an eventual reversal within the year, shown by the negative coefficient for $AbnormalTickerSVI$ for week 5-52. In contrast with Da et al. (2011), we get significant results also for predicting return three weeks, four weeks and one year ahead.

The same models for the time period from 2009 to 2019 give us different results. In contrast with the positive relationship between search volume and abnormal return in the period from 2004 to 2008, the results in Table 4.2 indicate that increased search volume leads to decreased return the next weeks in the period from 2009 to 2019. Interpreting the signs and absolute values of the coefficients for the two search volume variables in the same manner as in Equation 4.2, these results are also strongest for the smaller companies. Higher R^2 values and coefficients combined with strong significance indicates that it is easier to forecast long term returns than short term returns.

	Dependent variable: AbnReturn									
	Week 1		Week 2		Week 3		Week 4		Week 5-52	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalTickerSVI$	0.141***	0.156***	0.130***	0.157***	0.132***	0.154***	0.135***	0.150***	-0.194***	-0.200***
	(0.0152)	(0.0171)	(0.0143)	(0.0167)	(0.0142)	(0.0186)	(0.0145)	(0.0178)	(0.0165)	(0.0231)
$MC * AbnormalTickerSVI$	-0.143***	-0.155***	-0.132***	-0.157***	-0.134***	-0.153***	-0.137***	-0.152***	0.191***	0.197***
	(0.0156)	(0.0174)	(0.0147)	(0.0168)	(0.0145)	(0.0188)	(0.0150)	(0.0179)	(0.0167)	(0.0233)
$MarketCap$		-0.0616***		-0.0561***		-0.0531***		-0.0503***		-0.0376***
		(0.00382)		(0.00368)		(0.00396)		(0.00381)		(0.00149)
$Absolute\ AbnReturn$		0.0460***		0.0803***		0.0652***		0.0717***		0.0221***
		(0.00769)		(0.00740)		(0.00662)		(0.00775)		(0.00273)
$XadSales$		-0.00328**		-0.00295*		-0.00246		-0.00287*		-0.00145*
		(0.00163)		(0.00165)		(0.00167)		(0.00164)		(0.000837)
$NoAnalysts$		-0.00254		-0.00361*		-0.00449**		-0.00583***		-0.0109***
		(0.00182)		(0.00184)		(0.00182)		(0.00182)		(0.000738)
$AbnTurnover$		0.00634**		0.000920		-0.00309		-0.00238		0.00105
		(0.00283)		(0.00257)		(0.00271)		(0.00262)		(0.00115)
N	787,188	228,274	783,508	227,197	779,824	226,118	776,142	225,033	767,171	224,266
R^2	0.004	0.041	0.003	0.042	0.003	0.037	0.003	0.041	0.012	0.046
adj. R^2	0.003	0.035	0.003	0.035	0.003	0.031	0.003	0.034	0.012	0.040

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.1: Predictive model for $AbnReturn$ with $AbnormalTickerSVI$ in the period from 2004 to 2008, estimated using Fama-Macbeth regression. The dependent variable is the abnormal return during the first 4 weeks and during weeks 5 to 52. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

	Dependent variable: AbnReturn									
	Week 1		Week 2		Week 3		Week 4		Week 5-52	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalTickerSVI$	-0.108***	-0.0969***	-0.111***	-0.0819***	-0.113***	-0.0925***	-0.0989***	-0.0837***	-0.293***	-0.276***
	(0.00822)	(0.0106)	(0.00804)	(0.0103)	(0.00847)	(0.0109)	(0.00827)	(0.0111)	(0.0161)	(0.0174)
$MC * AbnormalTickerSVI$	0.107***	0.0946***	0.110***	0.0812***	0.111***	0.0921***	0.0968***	0.0812***	0.291***	0.275***
	(0.00821)	(0.0105)	(0.00797)	(0.0102)	(0.00847)	(0.0108)	(0.00824)	(0.0110)	(0.0161)	(0.0173)
$MarketCap$		-0.108***		-0.104***		-0.104***		-0.103***		-0.199***
		(0.00335)		(0.00343)		(0.00339)		(0.00343)		(0.00263)
$Absolute\ AbnReturn$		0.0252***		0.0481***		0.0493***		0.0500***		0.0666***
		(0.00554)		(0.00511)		(0.00497)		(0.00537)		(0.00400)
$XadSales$		-0.00132		-0.00123		-0.00134		-0.00115		-0.00266***
		(0.00124)		(0.00125)		(0.00127)		(0.00123)		(0.000650)
$NoAnalysts$		0.00251		0.00179		0.00135		0.000952		0.00678***
		(0.00176)		(0.00175)		(0.00173)		(0.00173)		(0.00176)
$AbnTurnover$		0.00427**		0.00192		-0.0000696		-0.00105		-0.00809***
		(0.00210)		(0.00194)		(0.00190)		(0.00191)		(0.00119)
N	1,657,206	595,613	1,652,769	594,151	1,648,342	592,681	1,643,926	591,205	1,577,274	570,192
R^2	0.003	0.048	0.003	0.046	0.003	0.046	0.003	0.047	0.014	0.132
adj. R^2	0.002	0.043	0.002	0.041	0.002	0.040	0.002	0.041	0.014	0.126

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.2: Predictive model for $AbnReturn$ with $AbnormalTickerSVI$ in the period from 2009 to 2019, estimated using Fama-Macbeth regression. The dependent variable is the abnormal return during the first 4 weeks and during weeks 5 to 52. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

4.2 Comparing stock ticker and company name as search keyword

Table 4.3 and Table 4.4 show that while ticker outperforms company name in the period from 2004 to 2008 with higher coefficients and R^2 values, company name performs better than ticker in the period from 2009 to 2019. The coefficients for ticker and company name have the same sign respectively in the first and the second time period, indicating that both of them serve as a measure of investor attention, and the interpretation is the same for company name as for stock ticker (see section 4.1).

Since 2004 the amount of information available for individual investors has increased dramatically. The information base does no longer consist only of basic financial information but now includes news in general newspapers, government publications, and company websites, among others. This information is available for individual investors through Google if searched for the company name (combined with other terms), while searches for ticker gives the investor mainly financial information. At the same time, the number of searches from consumers has increased dramatically, with the number of Google searches increasing by a factor of 40 since 2004 (Internet Live Stats, 2020). This contributes to an increased number of searches for company names, which, in turn, is a good measure for the overall attention of a company. Both the fact that more information is available for the individual investors from a larger variety of sources and increased use of Google from consumers might explain why search volume for company name now is a better predictor of return than before 2008.

	Dependent variable: AbnReturn									
	Week 1		Week 2		Week 3		Week 4		Week 5-52	
	Ticker	Name	Ticker	Name	Ticker	Name	Ticker	Name	Ticker	Name
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalTypeSVI$	0.156***	0.0660***	0.157***	0.0965***	0.154***	0.105***	0.150***	0.0951***	-0.200***	-0.134***
	(0.0171)	(0.0202)	(0.0167)	(0.0196)	(0.0186)	(0.0198)	(0.0178)	(0.0193)	(0.0231)	(0.0226)
$MC * AbnormalTypeSVI$	-0.155***	-0.0701***	-0.157***	-0.101***	-0.153***	-0.108***	-0.152***	-0.101***	0.197***	0.132***
	(0.0174)	(0.0198)	(0.0168)	(0.0196)	(0.0188)	(0.0197)	(0.0179)	(0.0194)	(0.0233)	(0.0224)
$MarketCap$	-0.0616***	-0.0611***	-0.0561***	-0.0561***	-0.0531***	-0.0528***	-0.0503***	-0.0498***	-0.0376***	-0.0348***
	(0.00382)	(0.00381)	(0.00368)	(0.00367)	(0.00396)	(0.00392)	(0.00381)	(0.00381)	(0.00149)	(0.00134)
$Absolute\ AbnReturn$	0.0460***	0.0443***	0.0803***	0.0771***	0.0652***	0.0616***	0.0717***	0.0691***	0.0221***	0.0214***
	(0.00769)	(0.00771)	(0.00740)	(0.00738)	(0.00662)	(0.00658)	(0.00775)	(0.00783)	(0.00273)	(0.00256)
$XadSales$	-0.00328**	-0.00431**	-0.00295*	-0.00394**	-0.00246	-0.00365**	-0.00287*	-0.00348**	-0.00145*	0.00103
	(0.00163)	(0.00170)	(0.00165)	(0.00173)	(0.00167)	(0.00175)	(0.00164)	(0.00175)	(0.000837)	(0.000949)
$NoAnalysts$	-0.00254	-0.00242	-0.00361*	-0.00351*	-0.00449**	-0.00427**	-0.00583***	-0.00569***	-0.0109***	-0.00930***
	(0.00182)	(0.00190)	(0.00184)	(0.00189)	(0.00182)	(0.00186)	(0.00182)	(0.00188)	(0.000738)	(0.000641)
$AbnTurnover$	0.00634**	0.00691**	0.000920	0.00170	-0.00309	-0.00296	-0.00238	-0.00339	0.00105	0.000887
	(0.00283)	(0.00289)	(0.00257)	(0.00266)	(0.00271)	(0.00272)	(0.00262)	(0.00262)	(0.00115)	(0.00111)
N	228,274	225,751	227,197	224,697	226,118	223,640	225,033	222,578	224,266	221,159
R^2	0.041	0.040	0.042	0.041	0.037	0.035	0.041	0.039	0.046	0.039
adj. R^2	0.035	0.034	0.035	0.034	0.031	0.029	0.034	0.032	0.040	0.032

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.3: Predictive model for $AbnReturn$ comparing $AbnormalTickerSVI$ and $AbnormalNameSVI$ in the period from 2004 to 2008, estimated using Fama-Macbeth regression. The dependent variable is the abnormal return during the first 4 weeks and during weeks 5 to 52. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

	Dependent variable: AbnReturn									
	Week 1		Week 2		Week 3		Week 4		Week 5-52	
	Ticker	Name	Ticker	Name	Ticker	Name	Ticker	Name	Ticker	Name
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalTypeSVI$	-0.0969***	-0.150***	-0.0819***	-0.145***	-0.0925***	-0.150***	-0.0837***	-0.146***	-0.276***	-0.352***
	(0.0106)	(0.0105)	(0.0103)	(0.0103)	(0.0109)	(0.0105)	(0.0111)	(0.00998)	(0.0174)	(0.0152)
$MC * AbnormalTypeSVI$	0.0946***	0.150***	0.0812***	0.145***	0.0921***	0.147***	0.0812***	0.144***	0.275***	0.345***
	(0.0105)	(0.0103)	(0.0102)	(0.00995)	(0.0108)	(0.0103)	(0.0110)	(0.00980)	(0.0173)	(0.0148)
$MarketCap$	-0.108***	-0.111***	-0.104***	-0.107***	-0.104***	-0.107***	-0.103***	-0.106***	-0.199***	-0.201***
	(0.00335)	(0.00335)	(0.00343)	(0.00342)	(0.00339)	(0.00340)	(0.00343)	(0.00343)	(0.00263)	(0.00276)
$Absolute\ AbnReturn$	0.0252***	0.0249***	0.0481***	0.0475***	0.0493***	0.0489***	0.0500***	0.0494***	0.0666***	0.0665***
	(0.00554)	(0.00555)	(0.00511)	(0.00516)	(0.00497)	(0.00498)	(0.00537)	(0.00541)	(0.00400)	(0.00400)
$XadSales$	-0.00132	-0.00161	-0.00123	-0.00142	-0.00134	-0.00155	-0.00115	-0.00121	-0.00266***	0.000184
	(0.00124)	(0.00125)	(0.00125)	(0.00127)	(0.00127)	(0.00128)	(0.00123)	(0.00124)	(0.000650)	(0.000703)
$NoAnalysts$	0.00251	0.00227	0.00179	0.00147	0.00135	0.00110	0.000952	0.000437	0.00678***	0.00494***
	(0.00176)	(0.00179)	(0.00175)	(0.00178)	(0.00173)	(0.00177)	(0.00173)	(0.00176)	(0.00176)	(0.00175)
$AbnTurnover$	0.00427**	0.00318	0.00192	0.00149	-0.0000696	0.000110	-0.00105	-0.000759	-0.00809***	-0.00844***
	(0.00210)	(0.00212)	(0.00194)	(0.00194)	(0.00190)	(0.00190)	(0.00191)	(0.00193)	(0.00119)	(0.00121)
N	595,613	576,562	594,151	575,144	592,681	573,719	591,205	572,288	570,192	551,897
R^2	0.048	0.049	0.046	0.048	0.046	0.047	0.047	0.048	0.132	0.133
adj. R^2	0.043	0.044	0.041	0.042	0.040	0.041	0.041	0.042	0.126	0.127

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.4: Predictive model for $AbnReturn$ comparing $AbnormalTickerSVI$ and $AbnormalNameSVI$ in the period from 2009 to 2019, estimated using Fama-Macbeth regression. The dependent variable is the abnormal return during the first 4 weeks and during weeks 5 to 52. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

4.3 Predicting volatility and turnover

From both Table 4.5 and Table 4.6, we see that searches for company name have negative predictive power for volatility, and that searches for company name weighted by $MarketCap$ has positive coefficients. This indicates that higher search volumes are followed by lower volatility the next weeks and that this relationship is strongest for the smaller companies. Earlier research, including Fink and Johann (2014), find a positive relationship between search volume and volatility, but the relationship was strongest for the large stocks. Aouadi et al. (2013) conclude that higher search volume has mixed impact on volatility. Inconsistent findings in earlier research may be due to the fact that attention increases volatility by incorporating more information into the prices, while it also decreases volatility by reducing uncertainty, as stated in Aouadi et al. (2013). For the smaller companies, which the investors are less exposed to information about, increased attention would likely reduce uncertainty more than for the bigger companies, where there is a lot of information available. Higher coefficients in Table 4.5 indicates a stronger negative relationship between search volume variables and volatility in the period 2004-2008 than 2009-2019, which may be due to even less information being available about the smaller companies before 2008 than after.

The results in Table 4.7 and Table 4.8 show that increased search volume for company name leads to higher turnover the following three weeks, in line with Vlastakis and Markellos (2012) and Preis et al. (2010). This holds in both time periods but is stronger in the period from 2009 to 2019, which may be due to more investors being online using Google to gather information. The results in Table 4.8 show negative coefficients for $AbnormalNameSVI$ for week 5, indicating that this effect is gone after 3-4 weeks and that turnover is decreasing back to the normal level. The negative coefficient for $MC * AbnormalNameSVI$ in Table 4.7 and Table 4.8 might indicate that for smaller companies the effect is slightly stronger. The coefficients for the search volume variables in Table 4.7 and Table 4.8 are more significant for the simple models, indicating that some of the control variables are better predictors of turnover.

We also ran the prediction models for volatility and turnover using $AbnormalTickerSVI$ as search volume variable. The results are similar, but less significant than the results from the models using $AbnormalNameSVI$. The results from the models using $AbnormalTickerSVI$ can be found in appendix A.

	Dependent variable: Volatility									
	Week 1		Week 2		Week 3		Week 4		Week 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalNameSVI$	-0.0658***	-0.141***	-0.0647***	-0.0940***	-0.0736***	-0.145***	-0.0702***	-0.111***	-0.0750***	-0.133***
	(0.0133)	(0.0221)	(0.0120)	(0.0207)	(0.0116)	(0.0201)	(0.0127)	(0.0206)	(0.0127)	(0.0219)
$MC * AbnormalNameSVI$	0.0720***	0.140***	0.0704***	0.0971***	0.0783***	0.145***	0.0738***	0.112***	0.0783***	0.133***
	(0.0133)	(0.0219)	(0.0118)	(0.0204)	(0.0115)	(0.0197)	(0.0124)	(0.0204)	(0.0125)	(0.0215)
$MarketCap$		-0.0697***		-0.0679***		-0.0646***		-0.0611***		-0.0592***
		(0.00454)		(0.00464)		(0.00461)		(0.00460)		(0.00453)
$Absolute\ AbnReturn$		-0.00626		-0.0248***		-0.0224***		-0.0233***		-0.0192***
		(0.00548)		(0.00477)		(0.00451)		(0.00468)		(0.00440)
$XadSales$		-0.00188		-0.00193		-0.00143		-0.000879		-0.00116
		(0.00179)		(0.00176)		(0.00176)		(0.00179)		(0.00187)
$NoAnalysts$		-0.00216		-0.00293		-0.00416**		-0.00442**		-0.00483**
		(0.00190)		(0.00186)		(0.00187)		(0.00187)		(0.00187)
$AbnTurnover$		0.0380***		0.00679*		0.00204		-0.00267		-0.00224
		(0.00440)		(0.00378)		(0.00395)		(0.00357)		(0.00339)
N	819647	225122	815384	224072	811135	223050	806949	222022	802812	220995
R^2	0.002	0.036	0.001	0.031	0.001	0.029	0.001	0.027	0.002	0.027
adj. R^2	0.001	0.029	0.001	0.024	0.001	0.022	0.001	0.021	0.001	0.020

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.5: Predictive model for $Volatility$ with $AbnormalNameSVI$ in the period from 2004 to 2008, estimated using Fama-Macbeth regression. The dependent variable is the volatility during the following 5 weeks. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

	Dependent variable: Volatility									
	Week 1		Week 2		Week 3		Week 4		Week 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalNameSVI$	-0.0276***	-0.0433***	-0.0297***	-0.0418***	-0.0219***	-0.0272***	-0.0251***	-0.0461***	-0.0236***	-0.0486***
	(0.00812)	(0.0107)	(0.00743)	(0.0101)	(0.00720)	(0.0104)	(0.00764)	(0.0102)	(0.00741)	(0.0101)
$MC * AbnormalNameSVI$	0.0373***	0.0491***	0.0323***	0.0464***	0.0248***	0.0309***	0.0264***	0.0472***	0.0247***	0.0499***
	(0.00826)	(0.0107)	(0.00758)	(0.0100)	(0.00740)	(0.0102)	(0.00779)	(0.0101)	(0.00757)	(0.0101)
$MarketCap$		-0.174***		-0.174***		-0.171***		-0.168***		-0.166***
		(0.00372)		(0.00367)		(0.00360)		(0.00352)		(0.00345)
$Absolute\_AbnReturn$		-0.00703**		-0.0234***		-0.0228***		-0.0251***		-0.0218***
		(0.00332)		(0.00286)		(0.00274)		(0.00281)		(0.00290)
$XadSales$		-0.00272**		-0.00279**		-0.00252**		-0.00316***		-0.00334***
		(0.00108)		(0.00109)		(0.00109)		(0.00109)		(0.00107)
$NoAnalysts$		-0.000670		-0.000507		-0.000831		-0.000767		-0.00123
		(0.00159)		(0.00159)		(0.00153)		(0.00153)		(0.00150)
$AbnTurnover$		0.0439***		-0.00271		-0.0111***		-0.0132***		-0.0140***
		(0.00259)		(0.00235)		(0.00204)		(0.00208)		(0.00227)
N	1674280	576249	1669257	574830	1664285	573465	1659255	572089	1654321	570713
R^2	0.002	0.060	0.002	0.054	0.002	0.052	0.002	0.051	0.002	0.050
adj. R^2	0.002	0.054	0.002	0.048	0.002	0.046	0.002	0.045	0.002	0.045

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.6: Predictive model for $Volatility$ with $AbnormalNameSVI$ in the period from 2009 to 2019, estimated using Fama-Macbeth regression. The dependent variable is the volatility during the following 5 weeks. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

	Dependent variable: AbnTurnover									
	Week 1		Week 2		Week 3		Week 4		Week 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalNameSVI$	0.0252***	0.0141***	0.0173***	0.0131***	0.0101***	0.00800**	0.00593**	0.00645*	0.00271	0.00112
	(0.00309)	(0.00376)	(0.00292)	(0.00407)	(0.00286)	(0.00365)	(0.00249)	(0.00346)	(0.00270)	(0.00386)
$MC * AbnormalNameSVI$	-0.000307**	0.0000539	-0.000300***	-0.000183	-0.000170	0.0000248	0.00000143	-4.48e-10	-0.0000170	-0.0000124
	(0.000120)	(0.000165)	(0.000113)	(0.000172)	(0.000112)	(0.000170)	(0.0000977)	(0.000159)	(0.000105)	(0.000180)
$MarketCap$		-0.00334		0.000520		0.00510		0.00962**		0.0121***
		(0.00372)		(0.00385)		(0.00390)		(0.00401)		(0.00405)
$Absolute\ AbnReturn$		0.00930		-0.0123**		-0.0120**		-0.0105**		-0.00558
		(0.00634)		(0.00527)		(0.00478)		(0.00493)		(0.00479)
$XadSales$		-0.00141		-0.000742		-0.000818		-0.000493		-0.0000545
		(0.00224)		(0.00227)		(0.00231)		(0.00234)		(0.00233)
$NoAnalysts$		-0.00944***		-0.00928***		-0.0105***		-0.0105***		-0.0103***
		(0.00270)		(0.00262)		(0.00269)		(0.00268)		(0.00269)
N	835,833	225,807	831,641	224,794	827,469	223,775	823,291	222,751	819,127	221,724
R^2	0.001	0.014	0.001	0.012	0.001	0.012	0.001	0.012	0.001	0.012
adj. R^2	0.001	0.00	0.001	0.007	0.001	0.007	0.001	0.007	0.001	0.007

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.7: Predictive model for $AbnTurnover$ with $AbnormalNameSVI$ in the period from 2004 to 2008, estimated using Fama-Macbeth regression. The dependent variable is the turnover during the following 5 weeks. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

	Dependent variable: AbnTurnover									
	Week 1		Week 2		Week 3		Week 4		Week 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalNameSVI$	0.0420***	0.0361***	0.0220***	0.0194***	0.00932***	0.00931***	0.000105	-0.000778	-0.00692***	-0.00400*
	(0.00206)	(0.00230)	(0.00192)	(0.00230)	(0.00183)	(0.00231)	(0.00169)	(0.00226)	(0.00165)	(0.00225)
$MC * AbnormalNameSVI$	-0.000727***	-0.000489***	-0.000419***	-0.000397***	-0.0000820	-0.0000210	0.000102	0.000171	0.000106	0.0000737
	(0.000112)	(0.000145)	(0.000115)	(0.000153)	(0.000116)	(0.000152)	(0.000109)	(0.000149)	(0.000106)	(0.000155)
$MarketCap$		-0.00136		-0.00136		0.0000157		0.00252		0.00481
		(0.00322)		(0.00325)		(0.00327)		(0.00327)		(0.00327)
$Absolute\ AbnReturn$		0.0134***		-0.00758***		-0.0151***		-0.0126***		-0.0133***
		(0.00363)		(0.00291)		(0.00270)		(0.00265)		(0.00271)
$XadSales$		-0.000625		-0.000673		-0.000131		-0.000270		-0.000105
		(0.00130)		(0.00131)		(0.00131)		(0.00129)		(0.00130)
$NoAnalysts$		-0.00634***		-0.00610***		-0.00607***		-0.00599***		-0.00638***
		(0.00181)		(0.00181)		(0.00181)		(0.00179)		(0.00176)
N	1,691,391	576,638	1,686,639	575,281	1,681,879	573,918	1,677,125	572,547	1,672,405	571,171
R^2	0.002	0.017	0.002	0.014	0.001	0.013	0.001	0.013	0.001	0.013
adj. R^2	0.002	0.012	0.001	0.009	0.001	0.008	0.001	0.008	0.001	0.008

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 4.8: Predictive model for $AbnTurnover$ with $AbnormalNameSVI$ in the period from 2009 to 2019, estimated using Fama-Macbeth regression. The dependent variable is the turnover during the following 5 weeks. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

Chapter 5

Trading strategies

In this chapter, we evaluate various trading strategies to test the economic significance of our results. We start trading in 2006 to allow for up to two years of training for the first prediction. All companies from the earlier mentioned Russell 3000 dataset with more than one year of available data for both search volume and financial variables are included. We are only using search volume for the company name, as it performed better than stock ticker in the previous analysis. We use the training data to predict standardized abnormal returns for the following week using Fama MacBeth cross-sectional regression. By standardizing the returns, we prevent the companies with a high variance from being weighted more. We use the forecast of abnormal return to construct an equally-weighted portfolio where we buy the top 50% stocks with the highest predicted abnormal return and short the bottom 50% with the lowest predicted abnormal return. We hold this portfolio for one week before rebalancing. We use a rolling training window of 52 weeks, meaning we only use data from the 52 most recent weeks for training when predicting next week's abnormal return.

We calculate the return of the strategy by Equation 5.1:

$$Return_{portfolio,t} = \frac{Return_{long,t} - Return_{short,t}}{2} \tag{5.1}$$

Following Equation 5.1, the short position is treated as a capital investment giving the opposite return of a long position in the same stock. If the long position has 6% return and the short position has 10% return, the portfolio return becomes -2%.

With an equal value of short positions and long positions, the portfolio is free to buy. The expected return for the respectively random portfolio is 0% since the return of both the long portfolio and the short portfolio is expected to follow the market. The return for our portfolios is likely to be excess return since the portfolio is constructed as market neutral. Later we will confirm this, but it is essential to know when interpreting the results.

5.1 The value of Google Trends

We start by comparing three different portfolios with three different sets of variables as regressors. We calculate both the average yearly return and the abnormal return, α, using the CAPM model. The three portfolios are all constructed using 52 weeks of training.

Regressors	Yearly α	Yearly avg. return
$AbnormalNameSVI$, $MC * AbnormalNameSVI$	3.6%	5.7%
$MarketCap$, $Absolute\ AbnReturn$, $XadSales$, $NoAnalysts$, $AbnTurnover$	9.4%	11.4%
$AbnormalNameSVI$, $MC * AbnormalNameSVI$, $MarketCap$, $Absolute\ AbnReturn$, $XadSales$, $NoAnalysts$, $AbnTurnover$	10.9%	12.3%

Table 5.1: Average yearly return and yearly abnormal return, α, for portfolios constructed with different sets of variables as regressors.

As seen in Table 5.1 the first portfolio, only including Google search volume combined with market capitalization as regressors, delivers positive alpha. This demonstrates that search volume for company name is a relevant indicator of future abnormal returns that the market has not fully incorporated into its expectations. The five other variables are performing better than Google search volume alone, when they are combined, but by adding Google search volume we see that the alpha increases from 9.4% per year to 10.9%. This indicates that Google search volume can contribute to the prediction beyond the other variables.

Going forward we will use all the variables as regressors as this gives the best results.

5.2 Training windows

To see how the size of the training window affects the predictions, we create five different portfolios, one for each of the training window sizes 2, 12, 26, 52, and 104 weeks. Only stocks with enough training data are included in the evaluation. The cumulative returns for all of the five portfolios are plotted together in Figure 5.1.

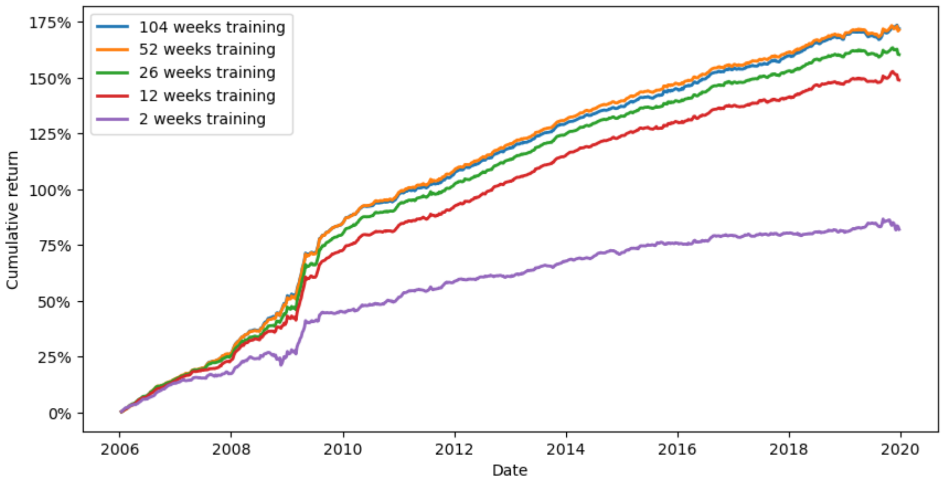

Figure 5.1: Cumulative return for different training windows. In all strategies, we buy the top 50% of stocks with the highest predicted abnormal return and short sell the bottom 50%.

We draw two main conclusions from Figure 5.1. First, we can see that the strategy is outperforming the equivalent random strategy with an expected return of 0% for all training window sizes. This indicates that search volume combined with other variables are relevant predictors of future abnormal returns that the market has not fully incorporated into its expectations. Second, strategies with larger training window perform better than those with lower training window up to 52 weeks, increasing dramatically between 2 and 12 weeks. With too low window size, the model will not capture the full pattern and is therefore too sensitive to time effects. Too large window size will make the model less adaptable to changes in the relationship between abnormal return and the given predictors. In previous sections, we have proven that this relationship has changed after 2008. Too large training window will also reduce the number of stocks with enough available training data and therefore limit the number of potential stocks to buy and sell.

5.3 Risk exposure

As mentions before, our portfolios should be market neutral by construction. The strategy is, however, not random, so the stock selection might have resulted in loading of other risk factors. In chapter 4, we have shown that search volume is a better predictor for small companies, so our portfolios could be exposed to the small minus big factor. We check for several relevant factors and calculate the abnormal returns, α, based on these. We are using both the CAPM model, the Fama-French three-factor model, and the Fama-French five-factor model. The results are presented in Table 5.2.

Trading strategy	Avg. yearly return	Yearly α	Mkt-R$_f$	SMB	HML	RMW	CMA
	11.3%	10.4%	0.02***				
50-50 window=104	11.3%	10.9%	0.00	0.03***	0.06***		
	11.3%	10.9%	0.00	0.03***	0.07***	0.02*	-0.03*
	12.3%	10.4%	0.02***				
50-50 window=52	12.3%	10.9%	0.00	0.03***	0.06***		
	12.3%	10.4%	0.00	0.03***	0.06***	0.03***	-0.03*
	11.5%	9.4%	0.02***				
50-50 window=26	11.5%	9.9%	0.00	0.03***	0.06***		
	11.5%	9.9%	0.00	0.03***	0.06***	0.03**	-0.02
	10.6%	8.8%	0.02***				
50-50 window=12	10.6%	8.8%	0.00	0.03***	0.05***		
	10.6%	8.8%	0.00	0.03***	0.05***	0.03***	-0.01
	5.9%	4.7%	0.00				
50-50 window=2	5.9%	4.7%	0.00	0.02*	0.03***		
	5.9%	4.7%	0.00	0.02**	0.03***	0.02	0.00

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5.2: Load of risk factors for portfolios created with different size of trading window.

For all of the portfolios, the factor loading is low. For market risk, the beta is close to zero. This is expected since the portfolios should be close to market neutral by construction. For *small minus big*, *high minus low*, and *robust minus weak* the factors are significant and positive. The positive beta for *small minus big* means that our portfolios are weighted toward small-cap stocks, which is expected since the market capitalization is included as one of the predictors, and the results in chapter 4 showed that search volume is a better predictor of abnormal return for small-caps. Despite positive loading, all the portfolios deliver large alphas/abnormal returns after accounting for the most common risk factors.

5.4 Trading costs

We will now check if the trading strategies are still profitable after accounting for trading costs. Trading costs consist mainly of transaction costs, market-maker spread and market impact. Market impact is irrelevant assuming the trades to be too small to influence the market and will be ignored going forward.

We estimate the transaction fees by observing the fees of online brokers. At Interactive Brokers (2020) they offer an account which charges \$0.005 per share, while Speed Trader (2020) offers accounts with a flat fee of \$ 4.95 per trade, independent of number of shares. By assuming a minimum of 100 shares per trade, the average cost will be no more than \$ 0.005 per share for both brokers. The average share price for the stocks in our dataset is \$ 30.7, resulting in an average transaction fee of 0.01%.

Since the efficient bid-ask spread cannot be observed directly, it is harder to determine. Olbryś and Mursztyn (2019) investigates the effective bid/ask spread and find a mean of 0.13% while Bijl et al. (2016) uses a bid-ask spread of 0.08%. Ball and Chordia (2001) finds an average quoted spread for large-stocks of 0.2%. We apply half of a bid-ask spread of 0.2%, giving us a double one-way transaction cost of 0.22% (from a long position to a short position, or opposite). The results can be seen in Table 5.3.

Trading strategy	Avg. yearly return including trading costs	Avg. yearly return without trading costs	Percentage of portfolio traded each week
104 weeks training	11.28%	12.29%	10.4%
52 weeks training	11.12%	12.25%	11.6%
26 weeks training	10.11%	11.46%	13.7%
12 weeks training	8.93%	10.64%	17.3%
2 weeks training	2.34%	5.86%	33.6%

Table 5.3: Yearly average return after adjusted for trading costs for portfolios created with different size of training window.

As expected, the turnover is highest for the portfolios with the lowest training window, resulting in higher total trading costs. Trading costs decrease the yearly performance by 1.0 - 3.5 percent points, removing a substantial amount of the excess return. The simple, equally-weighted portfolio consisting of the same stocks gives an average yearly return for the same period of 9.63%. Compared to the results in Table 5.3, only the portfolios with 26 weeks of training or more are generating positive returns after accounting for trading costs. To comparison, Bijl et al. (2016) construct a profitable portfolio based on search volume when the transaction cost is not taken into account. However, their strategy underperforms the equally weighted strategy by approximately 1% per year after the inclusion of transaction costs. However, it should be mentioned that our trading strategy is based on a different prediction model than Bijl et al. (2016); our strategy is based on Fama-MacBeth regression, while their strategy is based on a panel data regression with fixed effects.

5.5 Trading only stocks with extreme predicted returns

Until now, we have used the same trading strategy with different training windows. The 50-50 portfolios investigated so far includes all stocks and therefore illustrate how the prediction works for both extreme and average abnormal return predictions. It is also easy to evaluate the performance since the idiosyncratic risk is minimized. Still, the strategy can be modified to maximize returns. When the model includes all stocks, the portfolio will consist of a large part of stocks with predicted abnormal returns close to average. Even if the predictions are correct, these stocks will not contribute to the abnormal portfolio return. They will, at the same time, generate higher trading costs since a small change in the prediction from one week to another will, in many cases, result in the stock moving from the long portfolio to the short portfolio. A way to increase the portfolio return could be to only buy or sell the stocks with more extreme predicted abnormal returns, and take no position in the stocks which have a predicted abnormal return close to the median.

In this section, we investigate what happens if we change the buy/sell threshold to X%. We test the strategy for a threshold of 20%, 10%, and 5%, and compare it with the 50% threshold. We are using a training window of 52 weeks since this training window provided sufficiently good results in previous sections. The results are given in Table 5.4.

From Table 5.4, we see that the abnormal return, α, increases with lower thresholds. This indicates that our model can predict both normal and extreme abnormal returns. The portfolio volatility is also increasing with lower thresholds. After adjusting for trading costs, the highest Sharpe ratio is found for the portfolio with a long position in the top 10% stocks and a short position in the bottom 10% stocks.

Thresholds	Including trading costs			Without trading costs			Percentage of portfolio
	Yearly α	σ	Sharpe ratio	Yearly α	σ	Sharpe ratio	traded each week
Long-short 50%	8.8%	3.4%	2.74	10.9%	3.4%	3.11	11.6%
Long-short 20%	15.1%	6.2%	2.50	19.7%	6.2%	3.20	40.6%
Long-short 10%	23.9%	8.6%	2.79	29.6%	8.6%	3.47	54.2%
Long-short 5%	30.6%	11.6%	2.65	37.8%	11.6%	3.28	68.0%

Table 5.4: Yearly alpha, volatility and Sharpe ratio for a trading strategy buying a long position in the X% of stocks with highest predicted abnormal return, and selling a short position in the X% of stocks with lowest predicted abnormal return.

A plot of cumulative return for the four thresholds can be seen in Figure 5.2. When lowering the threshold, the returns are consistently improving. We can also see that the returns are moving very similarly for all the portfolios. Since the only stocks represented in all the portfolios are the ones with an extreme predicted abnormal return, we can conclude that these are the ones shaping the portfolio return. This confirms that the predictions for extreme abnormal returns are reliable.

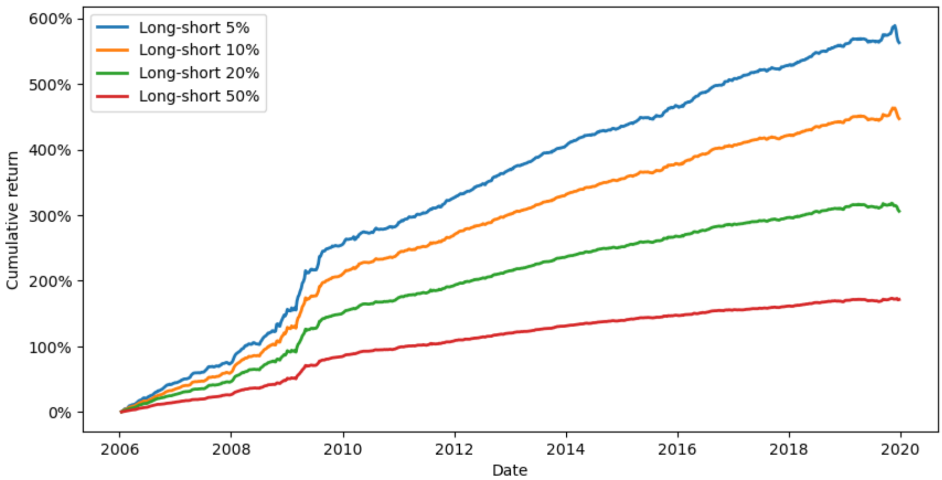

Figure 5.2: Cumulative return for portfolios constructed with different thresholds. A threshold of X% means buying a long position in the X% of stocks with highest predicted abnormal return, and selling a short position in the X% of stocks with lowest predicted abnormal return. All portfolios are using 52 weeks of training data.

5.6 Long or short portfolio?

To see whether the abnormal returns in the previous sections come from the long portfolio or the short portfolio, or a combination, we will in this section investigate the long and short portfolios separately. We look at portfolios consisting of top 20%, 10%, and 5% long, and portfolios consisting of bottom 20%, 10%, and 5% short. The results are shown in Table 5.5.

Trading strategy	Avg. yearly return	σ	Yearly α	Mkt-R$_f$	SMB	HML	RMW	CMA
Long top 5%	61.2%	29.8%	61.5%	0.23***	0.21***	0.55***	-0.04	-0.54***
Long top 10%	48.6%	26.8%	48.1%	0.22***	0.20***	0.47***	-0.04	-0.47***
Long top 20%	35.3%	24.5%	33.7%	0.22***	0.18***	0.42***	0.00	-0.41***
Long-short 50%	12.3%	3.4%	10.9%	0.00	0.03***	0.06***	0.03***	-0.03*
Short bottom 20%	3.4%	20.7%	7.3%	-0.22***	-0.07	-0.17***	0.08	0.29***
Short bottom 10%	9.4%	21.2%	13.5%	-0.22***	-0.07	-0.14***	0.09	0.30***
Short bottom 5%	13.1%	22.0%	17.1%	-0.21***	-0.06	-0.12***	0.11	0.29***

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 5.5: Average yearly return, volatility (σ), yearly abnormal return (α) and load of risk factors for different long and short only strategies. Long top X% means we buy a long position for the X% of stocks with highest predicted abnormal return.

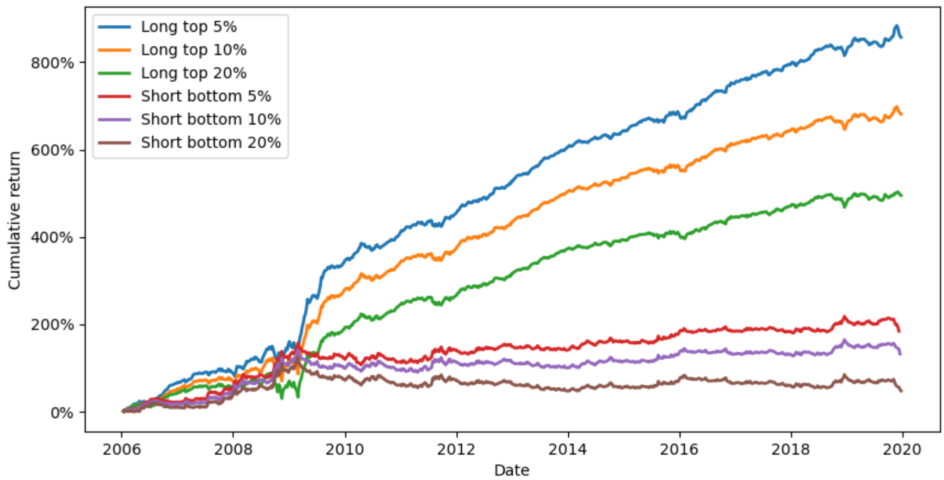

Figure 5.3: Cumulative return for long and short only portfolios constructed with different thresholds. Long top x% means buying a long position in the x% of stocks with the highest predicted abnormal return, while Short bottom x% means selling a short position in the x% of stocks with the lowest predicted abnormal return. All portfolios are using 52 weeks of training data.

Positive alpha for all of the portfolios in Table 5.5 tells us that we can predict both stocks with future positive abnormal returns and stocks with future negative abnormal returns. The increasing alpha when the threshold decrease for both the long and the short portfolios indicates that our model is able to predict both extreme negative abnormal return and extreme positive abnormal return. The abnormal return for the long portfolios are higher than for the short portfolios telling us that we are better at predicting extreme positive abnormal returns than extreme negative abnormal returns.

As expected, the volatilities, σ, are multiple times higher than the volatilities for the combined long-short portfolios from section 5.5. A plot of cumulative return for all the long and short only portfolios are shown in Figure 5.3.

Chapter 6

Conclusion

There has been increased interest in investigating the relationship between Google search volume and individual stock performance. Early findings concluded that Google searches can predict returns, while other papers come to the opposite conclusion. We study the findings of Da et al. (2011) to see if we could get similar results and if these results are present also after 2008. In line with Da et al. (2011), we find that more searches for ticker predict increased return the next weeks in the period from 2004 to 2008. However, we find that after 2008 higher search volume predicts decreased return.

Different studies use different search keywords to measure company-level investor attention, and both stock tickers and company names are frequently used. We study whether company names or stock tickers perform best, and if there are any differences across time. The results show that searches for company name are a better predictor of abnormal return than searches for stock ticker after 2008.

To test the economic significance of our results we create a trading strategy where we buy the top 50% stocks with highest predicted abnormal return and short the bottom 50% with lowest predicted abnormal return. We test for different training windows and conclude that 52 weeks of training gives the best performance. Accounting for trading costs decrease the yearly performance by 1.0-3.5 percent points. After inclusion of trading costs only the portfolios with 26 weeks of training or more outperform the market, with our best portfolio giving an yearly abnormal return of 11.3%. To make sure the return is not created by risk loading, we check the returns against known risk factors by estimating the CAPM model, the Fama-French three-factor model, and the Fama-French five-factor model. Despite of positive loading on some of the factors, all the portfolios deliver large abnormal returns after accounting for the most common risk factors.

By constructing portfolios with lower threshold where we only buy the top X% stocks and a short position in the bottom X% stocks we confirm that our model is able to predict both normal and extreme return. This is shown by increased alpha with lower threshold.

The best result is given for the portfolio with a long position in the top 10% stocks and an equal sized short position in the bottom 10% stocks. After accounting for transaction costs this portfolio delivers a Sharpe ratio of 2.79. We further investigate whether the alpha come from the long or short part of the portfolio. Both long only and short only portfolios are delivering positive alpha for all the thresholds, indicating that our model is able to predict both positive and negative extreme abnormal return. The alphas for the long portfolios are larger than those for the short portfolios indicating that we are better at predicting positive abnormal return than negative.

For further research on the field of Google Trends and stock market performance, we would suggest to use SVI for Google's concept id for the different companies as a proxy for investor attention. Concept id as a newly introduced feature from Google Trends, grouping all keywords and translations relevant to a specific concept (e.g. a company) together. Using concept id will avoid the issues with both ticker and company name as search keywords, but will require a huge manual job if done for a big dataset like Russell 3000 companies.

Bibliography

Aouadi, A., Arouri, M., Teulon, F., 2013. Investor attention and stock market activity: Evidence from france. Economic Modelling 35, 674–681.

Baker, J., 2016. Forecasting volume and price impact of earnings surprises using google insights. The International Journal of Business and Finance Research 10, 53–62.

Baker, M., Wurgler, J., 2007. Investor sentiment in the stock market. Journal of economic perspectives 21, 129–152.

Ball, C.A., Chordia, T., 2001. True spreads and equilibrium prices. The Journal of Finance 56, 1801–1835.

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, 239.

Barber, B.M., Odean, T., 2007. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. The review of financial studies 21, 785–818.

Bijl, L., Kringhaug, G., Molnár, P., Sandvik, E., 2016. Google searches and stock returns. International Review of Financial Analysis 45, 150–156.

Bui, V.X., Nguyen, H.T., 2019. Stock market activity and google trends: the case of a developing economy. Journal of Economics and Development .

Challet, D., Ayed, A.B.H., 2014. Do google trend data contain more predictability than price returns? arXiv preprint arXiv:1403.1715 .

Chemmanur, T.J., Yan, A., 2019. Advertising, attention, and stock returns. Quarterly Journal of Finance , 1950009.

Da, Z., Engelberg, J., Gao, P., 2011. In search of attention. The Journal of Finance 66, 1461–1499.

Ding, R., Hou, W., 2015. Retail investor attention and stock liquidity. Journal of International Financial Markets, Institutions and Money 37, 12–26.

Fama, E.F., MacBeth, J.D., 1973. Risk, return, and equilibrium: Empirical tests. Journal of political economy 81, 607–636.

Fink, C., Johann, T., 2014. May i have your attention, please: The market microstructure of investor attention. Please: The Market Microstructure of Investor Attention (September 17, 2014) .

Garman, M.B., Klass, M.J., 1980. On the estimation of security price volatilities from historical data. Journal of business , 67–78.

Gervais, S., Kaniel, R., Mingelgrin, D.H., 2001. The high-volume return premium. The Journal of Finance 56, 877–919.

Grullon, G., Kanatas, G., Weston, J.P., 2004. Advertising, breadth of ownership, and liquidity. The Review of Financial Studies 17, 439–461.

Hou, K., Xiong, W., Peng, L., 2009. A tale of two anomalies: The implications of investor attention for price and earnings momentum. Available at SSRN 976394 .

Interactive Brokers, 2020. Commissions - fixed. URL: https://www.interactivebrokers.com/en/index.php?f=1590&p=stocks1. [Accessed: 11-June-2020].

Internet Live Stats, 2020. Google search statistics. URL: https://www.internetlivestats.com/google-search-statistics/. [Accessed: 4-June-2020].

Joseph, K., Wintoki, M.B., Zhang, Z., 2011. Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. International Journal of Forecasting 27, 1116–1127.

Kim, N., Lučivjanská, K., Molnár, P., Villa, R., 2019. Google searches and stock market activity: Evidence from norway. Finance Research Letters 28, 208–220.

Kristoufek, L., 2013. Can google trends search queries contribute to risk diversification? Scientific reports 3, 2713.

Lou, D., 2014. Attracting investor attention through advertising. The Review of Financial Studies 27, 1797–1829.

Molnár, P., 2012. Properties of range-based volatility estimators. International Review of Financial Analysis 23, 20–29.

Moussa, F., Delhoumi, E., Ouda, O.B., 2017. Stock return and volatility reactions to information demand and supply. Research in International Business and Finance 39, 54–67.

Olbryś, J., Mursztyn, M., 2019. Alternative estimators for the effective spread derived from high-frequency data, in: Tarczyński, W., Nermend, K. (Eds.), Effective Investments on Capital Markets, Springer International Publishing, Cham. pp. 177–188.

Pancada, J.T., 2017. Google search volume as a proxy of investor attention: are previous findings robust? Master's thesis. BI Norwegian Business School.

Preis, T., Reith, D., Stanley, H.E., 2010. Complex dynamics of our economic life on different scales: insights from search engine query data. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 368, 5707–5719.

Seasholes, M.S., Wu, G., 2007. Predictable behavior, profits, and attention. Journal of Empirical Finance 14, 590–610.

Speed Trader, 2020. Commission fees. URL: https://speedtrader.com/commission-fees/. [Accessed: 11-June-2020].

Statscounter, 2020. Search Engine Market Share. URL: http://gs.statcounter.com/search-engine-market-share/all/united-states-of-america. [Accessed: 4-June-2020].

Vlastakis, N., Markellos, R.N., 2012. Information demand and stock market volatility. Journal of Banking & Finance 36, 1808–1821.

Yuan, Y., et al., 2008. Attention and trading. Unpublished Working Paper. University of Pennsylvania .

Appendix

A Predicting volatility and turnover with stock ticker

Table A1.1 and Table A1.2 show similar patterns as Table 4.5 and Table 4.6, indicating a negative relationship between search volume and volatility. Table A1.3 and Table A1.4 show a positive relationship between searches for ticker and turnover. However, the results are less significant than the results from the models using searches for company name to predict turnover, shown in Table 4.7 and Table 4.8.

	Dependent variable: Volatility									
	Week 1		Week 2		Week 3		Week 4		Week 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalTickerSVI$	-0.0878***	-0.104***	-0.0842***	-0.128***	-0.102***	-0.173***	-0.0922***	-0.178***	-0.0934***	-0.157***
	(0.0113)	(0.0182)	(0.0105)	(0.0190)	(0.0105)	(0.0205)	(0.0102)	(0.0210)	(0.0101)	(0.0223)
$MC * AbnormalTickerSVI$	0.0950***	0.109***	0.0894***	0.132***	0.106***	0.174***	0.0949***	0.179***	0.0964***	0.157***
	(0.0115)	(0.0182)	(0.0107)	(0.0190)	(0.0106)	(0.0204)	(0.0104)	(0.0207)	(0.0101)	(0.0220)
$MarketCap$		-0.0718***		-0.0703***		-0.0670***		-0.0635***		-0.0614***
		(0.00459)		(0.00467)		(0.00467)		(0.00464)		(0.00457)
$Absolute\ AbnReturn$		-0.00557		-0.0240***		-0.0226***		-0.0220***		-0.0186***
		(0.00545)		(0.00468)		(0.00449)		(0.00463)		(0.00439)
$XadSales$		-0.00188		-0.00179		-0.00158		-0.000771		-0.00127
		(0.00167)		(0.00167)		(0.00166)		(0.00168)		(0.00174)
$NoAnalysts$		-0.00282		-0.00390**		-0.00480**		-0.00521***		-0.00554***
		(0.00195)		(0.00191)		(0.00190)		(0.00192)		(0.00196)
$AbnTurnover$		0.0374***		0.00719*		0.00327		-0.000776		-0.000642
		(0.00432)		(0.00379)		(0.00393)		(0.00363)		(0.00340)
N	849,122	227,625	844,668	226,554	840,200	225,515	835,902	224,468	831,589	223,421
R^2	0.002	0.037	0.002	0.032	0.002	0.031	0.002	0.029	0.002	0.029
adj. R^2	0.002	0.030	0.002	0.025	0.002	0.024	0.001	0.023	0.002	0.022

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A1.1: Predictive model for $Volatility$ with $AbnormalTickerSVI$ in the period from 2004 to 2008, estimated using Fama-Macbeth regression. The dependent variable is the turnover during the following 5 weeks. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

	Dependent variable: Volatility									
	Week 1		Week 2		Week 3		Week 4		Week 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalTickerSVI$	-0.0195**	-0.0301***	-0.0178**	-0.0370***	-0.0189**	-0.0381***	-0.0136*	-0.0337***	-0.000914*	-0.000514
	(0.00793)	(0.0105)	(0.00771)	(0.0109)	(0.00749)	(0.0107)	(0.00777)	(0.0100)	(0.000494)	(0.000478)
$MC * AbnormalTickerSVI$	0.0299***	0.0334***	0.0208***	0.0391***	0.0216***	0.0389***	0.0163**	0.0338***	0.000901*	0.000472
	(0.00825)	(0.0106)	(0.00802)	(0.0109)	(0.00788)	(0.0107)	(0.00806)	(0.0101)	(0.000501)	(0.000476)
$MarketCap$		-0.176***		-0.175***		-0.172***		-0.170***		-0.00518***
		(0.00370)		(0.00364)		(0.00358)		(0.00349)		(0.000249)
$Absolute\ AbnReturn$		-0.00778**		-0.0247***		-0.0240***		-0.0263***		-0.00127***
		(0.00340)		(0.00280)		(0.00269)		(0.00275)		(0.000181)
$XadSales$		-0.00199*		-0.00215**		-0.00177*		-0.00235**		-0.000156***
		(0.00104)		(0.00104)		(0.00105)		(0.00105)		(0.0000416)
$NoAnalysts$		0.000245		0.000310		-0.0000926		-0.0000361		-0.0000620
		(0.00157)		(0.00155)		(0.00150)		(0.00150)		(0.000103)
$AbnTurnover$		0.0436***		-0.00236		-0.0104***		-0.0124***		-0.000354**
		(0.00255)		(0.00230)		(0.00201)		(0.00204)		(0.000145)
N	1,756,472	595,292	1,751,249	593,830	1,746,036	592,426	1,740,891	591,010	1,735,730	589,597
R^2	0.002	0.060	0.002	0.054	0.002	0.052	0.002	0.051	0.003	0.064
adj. R^2	0.002	0.055	0.002	0.049	0.002	0.047	0.002	0.046	0.003	0.059

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A1.2: Predictive model for $Volatility$ with $AbnormalTickerSVI$ in the period from 2009 to 2019, estimated using Fama-Macbeth regression. The dependent variable is the turnover during the following 5 weeks. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

	Dependent variable: AbnTurnover									
	Week 1		Week 2		Week 3		Week 4		Week 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalTickerSVI$	0.0215**	0.0452**	0.0159	0.0119	0.00604	-0.0498**	-0.00685	-0.0692***	-0.0117	-0.0369
	(0.00939)	(0.0210)	(0.00984)	(0.0223)	(0.00898)	(0.0219)	(0.00913)	(0.0216)	(0.00942)	(0.0229)
$MC * AbnormalTickerSVI$	-0.00357	-0.0295	-0.00788	-0.00902	-0.00355	0.0501**	0.00490	0.0647***	0.00778	0.0314
	(0.00921)	(0.0209)	(0.00974)	(0.0218)	(0.00888)	(0.0215)	(0.00896)	(0.0216)	(0.00932)	(0.0229)
$MarketCap$		-0.00338		0.000154		0.00451		0.00882**		0.0115***
		(0.00376)		(0.00383)		(0.00391)		(0.00399)		(0.00404)
$Absolute\ AbnReturn$		0.00884		-0.0129**		-0.0119**		-0.00996**		-0.00625
		(0.00625)		(0.00528)		(0.00480)		(0.00493)		(0.00491)
$XadSales$		-0.000794		-0.000147		-0.000352		0.0000698		0.000522
		(0.00216)		(0.00220)		(0.00222)		(0.00223)		(0.00224)
$NoAnalysts$		-0.00904***		-0.00934***		-0.0101***		-0.0102***		-0.00963***
		(0.00266)		(0.00262)		(0.00270)		(0.00272)		(0.00274)
N	867,483	228,331	863,124	227,297	858,791	226,258	854,458	225,213	850,126	224,166
R^2	0.001	0.014	0.001	0.012	0.001	0.012	0.001	0.013	0.001	0.013
adj. R^2	0.001	0.008	0.001	0.007	0.001	0.007	0.001	0.007	0.001	0.007

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A1.3: Predictive model for $AbnTurnover$ with $AbnormalTickerSVI$ in the period from 2004 to 2008, estimated using Fama-Macbeth regression. The dependent variable is the turnover during the following 5 weeks. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.

	Dependent variable: AbnTurnover									
	Week 1		Week 2		Week 3		Week 4		Week 5	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$AbnormalTickerSVI$	0.0195***	0.00476	0.0194***	0.0123	0.0136**	0.0118	0.0197***	0.0237**	0.0113**	0.0189*
	(0.00573)	(0.0102)	(0.00574)	(0.00994)	(0.00553)	(0.00997)	(0.00526)	(0.00956)	(0.00560)	(0.00975)
$MC * AbnormalTickerSVI$	0.0153***	0.0279***	-0.00427	0.00139	-0.00874	-0.00983	-0.0223***	-0.0288***	-0.0203***	-0.0290***
	(0.00564)	(0.0101)	(0.00576)	(0.00980)	(0.00554)	(0.00988)	(0.00525)	(0.00948)	(0.00560)	(0.00963)
$MarketCap$		-0.00166		-0.00164		-0.000125		0.00234		0.00488
		(0.00321)		(0.00327)		(0.00331)		(0.00330)		(0.00329)
$Absolute\ AbnReturn$		0.0132***		-0.00763***		-0.0150***		-0.0119***		-0.0127***
		(0.00363)		(0.00294)		(0.00271)		(0.00266)		(0.00276)
$XadSales$		-0.000514		-0.000573		-0.00000854		-0.000127		-0.0000835
		(0.00126)		(0.00126)		(0.00126)		(0.00124)		(0.00124)
$NoAnalysts$		-0.00690***		-0.00678***		-0.00697***		-0.00708***		-0.00759***
		(0.00182)		(0.00182)		(0.00183)		(0.00180)		(0.00177)
N	1,776,839	595,698	1,771,915	594,306	1,766,985	592,907	1,762,060	591,500	1,757,170	590,088
R^2	0.003	0.017	0.001	0.014	0.001	0.013	0.001	0.013	0.001	0.013

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A1.4: Predictive model for $AbnTurnover$ with $AbnormalTickerSVI$ in the period from 2009 to 2019, estimated using Fama-Macbeth regression. The dependent variable is the turnover during the following 5 weeks. Independent variables are defined in Table 2.1. Model specifications are given by Equation 3.1-3.4. Each column lists the coefficients (standard errors) and significance levels for the variables included in the respective model.