LAPPEENRANTA-LAHTI UNIVERSITY OF TECHNOLOGY LUT
School of Business and Management
Strategic Finance and Analytics

Master's Thesis:
Timing The Market With Google Trends Search Volume Data

23.06.2020

Tom Helenelund

1st Supervisor: Sheraz Ahmed

2nd Supervisor: Mikael Collan

Tiivistelmä

Tekijä:	Tom Helenelund
Otsikko:	Markkinoiden ajoittaminen Googlen hakusanavolyymien avulle
Akateeminen yksikkö:	LUT School of Business and Management
Maisteriohjelma:	Strategic Finance and Business Analytics
Vuosi:	2020
Pro Gradu:	107 Sivua, 4 Liitettä, 21 Taulukkoa, 10 Kaaviota, 4 Kuviota
Tarkastajat:	Apulaisprofessori, Sheraz Ahmed Professori, Mikael Collan
Hakusanat:	Google, Google-haku, SVI, Sijoittajahuomio, ETF, Hyödyke ETF, Hyödykemarkkinat, Jalometallit

Tämän tutkimuksen tarkoituksena on löytää todisteita siitä, onko markkinoiden ajoitus mahdollista hakusanavolyymien avulla ja kuinka hakukyselypohjainen sijoitusstrategia toimii valitulla ajanjaksolla. Lähestymistapana on analysoida Google-hakumäärien ja jalometalli ETF-hintamuutosten suhdetta Yhdysvaltain markkinoilla vuosina 2015-2020. Tämä tutkimus antaa oman panoksensa keskusteluun siitä, heijastaako hakukäyttäytyminen sijoittajien mielipiteitä vai ei. Huomiota mitataan kahdella erityyppisellä avainsanaryhmällä, joista ensimmäinen lähestymistapa käyttää ETF-tickereitä avainsanoina, kun taas toinen lähestymistapa käyttää yleisempiä markkinoihin liittyviä sanoja. Kaikki hakumäärät haetaan sekä Yhdysvalloista että globaalisti, jotta mahdollistetaan vertailu näiden kahden suoriutumisen välillä. Työssä käytetyt tilastolliset menetelmät ovat OLS-regressio ja paneeliaineiston regressio. Avainsanat ja ETF-rahastot, joille kehitetään yksinkertainen liukuvan keskiarvon kaupankäyntistrategia, valitaan regressioanalyysin tulosten perusteella. Tämän tutkimuksen lopullinen otos sisälsi viiden hyödyke-ETF: n taloudelliset tiedot ja 19 avainsanan hakutiedot sekä Yhdysvalloista että maailmanlaajuisesti.

Kaiken kaikkiaan tulokset osoittavat, että avainsanojen hakumäärien ja vastaavien ETF-tuottojen välillä on yhteys. Näihin suhteisiin perustuvat kaupankäyntistrategiat eivät kuitenkaan tuota merkittävää ylituottoa markkinoihin nähden. Kaikista hakusanoista avainsanan "S&P 500" hakumäärien perusteella rakennetulla strategialla saatiin suurin raakatuotto, 36.6%, joka oli vain 0.7% korkeampi kuin markkinoiden tuotto samalla ajanjaksolla. Tämä osoittaa, että vaikka hakumäärillä on huomattavaa ennustusvoimaa ETF-tuottoja kohtaan, näitä ennustusvoimia ei voida muuntaa kannattaviksi kaupankäyntistrategioiksi. Tämä voi johtua samanaikaisesti esiintyvien eri hakusanojen päinvastaisista vaikutuksista. Lisäksi yksittäisen hakusanan volyymin vaikutus tuottoihin voi olla ajasta riippuva, joten samojen (erilaisten) hakusanojen käyttäminen eri (samoina) aikoina voi olla kannattavampaa kuin yhtenäisen kaupankäyntistrategian käyttäminen koko ajanjakson ajan.

Abstract

Author:	Tom Helenelund
Title:	Timing The Market With Google Trends Search Volume Data
Faculty:	LUT School of Business and Management
Master's program:	Strategic Finance and Business Analytics
Year:	2020
Master's Thesis:	107 Pages, 4 Appendices, 21 Tables, 10 Graphs, 4 Figures
Examiners:	Associate Professor, Sheraz Ahmed Professor, Mikael Collan
Keywords:	Google, Google search, SVI, Investor attention, ETF, Commodity ETF, Commodity market, Precious metal

The purpose of this study is to find evidence on whether market timing is possible using search query-based information and how an investment strategy would perform over the selected time period. The approach is to analyze the relationship between Google search volumes and precious metals ETF price changes in the U.S. market in 2015-2020. Thus, this study makes its own contribution to the debate on whether or not search behavior captures investor sentiment. Attention is measured by two different types of keyword sets. The first approach is to use ETF tickers as keywords, while the second approach uses more general market-related words. All words are retrieved for both U.S. and global search volumes to compare which of these work better as a sentimental proxy. The statistical methods used in the work are OLS regression and panel data regression. The keywords and ETFs for which simple moving average trading strategies are developed are selected on the basis of regression analysis. The final sample of this study included financial data for five commodity ETFs and search query data for 19 keywords from both the U.S. and globally.

Overall, the results show that there is a relationship between search volumes of keywords and the subsequent ETF returns. However, the trading strategies based on these relationships do not produce significant market excess returns. Out of all search words, trading on the basis of search volume of keyword "S&P 500", generated a highest raw return of 36.6% which was only 0.7% higher than the market return during the same time period. This shows that although the search volumes possess significant predictive power for future commodity ETF returns, these predictive powers can hardly be translated into profitable trading strategies. This may be due to the opposite effect of different search words occurring simultaneously. In addition, the impact of individual search word volume on returns could be time-varying therefore using the same (different) search words in different (same) times could be more profitable than using a unified trading strategy over the whole time period.

Acknowledgments

Lappeenranta has been an amazing place to study. I want to thank both the employees and fellow student at Lappeenranta for making this time so special. Special thanks to the people with whom I shared two years on the board of the student organization Enklaavi. These years were highlights of the study period and taught a lot.

I truly believe that Lappeenranta has earned its place among the best universities in Finland thanks to its professional staff and inspiring management. From the employee side I want to especially thank three people. The first of these is Associate Professor Sheraz Ahmed who guided me through this thesis. The second person is Mikael Collan, who is a great professor and teacher. Together with Sheraz Ahmed, they are running an interesting MSF master's program in which I also got to take part. Finally, The third person is the current rector of the university, Juha-Matti Saksa. Juhis is an inspiration that is constantly looking forward and I am glad that I got to know him personally.

Last but not least, I want to thank my loving family and girlfriend who have supported me during my studies and especially during this thesis. I am excited to be graduating and can't wait what the future holds.

Tom Helenelund

Table of Content

Appendices

Tables

Graphs

Figures

1. Introduction

This chapter introduces the reader to the topic and opens up the motivation behind it. The purpose of this Master's Thesis is to study if search query volumes can predict the weekly price movements of commodity-based ETFs. The study focuses on the U.S. ETF market for the period from March 2015 to March 2020. This Master's Thesis uses the researches of Da, Engelberg and Gao (2011) as well as Preis, Moat and Stanley (2013) as a reference study and seeks to implement the approaches of both studies. The first approach uses tickers from selected ETFs to capture the investor sentiment (Da et al., 2011), while the second approach consists of more generic keywords with a similar purpose (Preis et al., 2013). Majority of the existing finance related research that uses search volumes focuses on the U.S. stock market. Compared to the existing literature, this study takes a new perspective by focusing on the precious metals market. Precious metals as an entity have been selected as they are expected to show the nervousness of the market. The goal is to use a time series data of both search query approaches and financial data of precious metal ETFs to run different regressions. Based on these regression results, a moving average trading strategy is then conducted for most promising search words. The analysis methods chosen to be used in this study are simple OLS regression and panel data regression.

This study shows that while the adequacy of Google Trends data has gradually improved, it is still the biggest limiting factor in a study of this type. As a result of data processing, the size of the sample was reduced from 22 precious metal ETFs to only 5 ETFs that met the selection criteria. Both noise and adequacy of data were critical factors leading to the exclusion of ETFs. It should be noted that the size of the ETF is directly related to finding sufficient information. The smaller the ETF in terms of its assets under management (AUM), the less likely it is to meet adequate information requirements. This can affect significantly the reliability of the results. Although the number of ETFs in this study remains small, the assets of the five selected ETFs are 88 billion. The final sample includes both financial data and search query data. Financial sample consists of 5 ETFs, which were also the 5 largest ETFs in the original sample. The search query data includes 19 keywords from the initial 42 search words.

1.1. Motivation

According to Alexa (2020) Google is the number one most visited website in the world. The company has a significant market share among the global search engines (92,71% in 2019), which means that a large part of the search activity is going through Google (Oberlo, 2020). Thus, Google can be seen to represents the search behavior of majority of people that has access to the internet. This makes Google a good choice to be used as a data source in this research. The use of search engine query data has only recently come more popular. Big reason for this was that this information was not previously available to the public. A study using Google data was made possible when the company introduced Google Trends in 2006. This made the search engine data that Google collects accessible for everyone with an access to the internet. It comes as no surprise that this large amount of new data has quickly attracted the interest of researchers in different fields.

Google trends enables accurate and almost real time analyses about the current trends and search behavior. A clear motive for this research is to see if search volumes can be used as sentimental proxy for investment decision. This is accomplished by executing a moving average trading strategy to see how an investment strategy using search volume would actually work. Another motive for this study is to see how commodity ETFs capture investor sentiment compared to individual stocks. Thus, the novelty of this study lies in the use of commodity ETFs as an underlying asset combined with the use of a sentimental proxy in investment decisions.

Google produces 3.5 billion searches daily. Certainly one would think that there is enough volume going through the website. However, there are also some negative aspects associated with Google Trends. The first aspect is the possible noise included to the data. There is such a wide variety of different searches and contexts that it can be seen impossible to exclude all the noise from the data (Choi & Varian, 2011). Another data limitation is that Google Trends limits the availability of its weekly data for more than five years. For longer time periods, the website automatically converts the data into monthly data. The time-series data provided by Google Trends is referred to as SVI, which stands for “Search Volume Index”. This variable is used throughout the thesis and

is referred in the thesis by its abbreviation (SVI). As the term indicates, the variable is reported in index format. This means that the raw search query levels are normalized to give values from 0 to 100 (Choi & Varian, 2009). This allows different search terms to be compared over time, but makes it difficult to combine two time series to create a time series of more than five years. This limits the time period used in this research.

Even if the sufficiency of Google trends related data has increased significantly from the times when Google Trends was first introduced, it is still a major limiting factor. When specifying the search, and looking for a search volume of a specific term with a certain context and limited region and time period, there is still a high probability that the search volume data is not sufficient enough, resulting either a zero or not producing the search information at all. This is considerably larger issue when using company or ETF tickers as search terms, compared to more generic search words. This is because tickers appear more random than generic search words. As a result, generic search words usually have larger search volumes.

From the perspective of this work, there is no existing literature of studies using search term volumes and investor sentiment on commodity based exchange-traded funds. However, inefficiencies in the pricing of exchange-traded funds (ETF) has been studied before. An example of a recent study regarding ETF mispricing is the research by Petajisto, published in 2017. It is possible that investor sentiment has an effect on the pricing of ETFs, and thus, could be one of the reasons to cause mispricing. As a result of researchers reporting evidence of ETF mispricing, new studies have appeared where researchers like Kreis and Licht (2018) study if it is possible to benefit from mispricing of ETF by utilizing different trading strategies.

Preis et al. (2013) write in their research paper that past financial crises are seen to be a result of human errors and complex human behavior. It is an interesting approach to see if search query data can give some warning signs about people's behavior and whether this behavior can be transmitted to commodity ETFs as price changes. According to the same study, by analyzing search terms related to finance, it gives an idea of human behavior and thus gives some idea of stock market movements. Both reference studies (Da et al.,2011 and Preis et al., 2013) are important studies in the financial sector related

to the use of investor sentiment in investment decision making. Both reference studies use Google data as an appropriate proxy. To capture the investor sentiment researchers have started to apply available data from wide range of different sources as news articles, Twitter and Wikipedia. (Tetlock, 2007; Bollen, Mao & Zeng, 2011; Moat, Curme, Avakian, Kenett, Stanley, & Preis, 2013; Challet & Ayed, 2013; Preis, Reith, & Stanley, 2010).

Finance is not the only field in which search query volumes have been used as a sentimental proxy. Good example of this is one of the earlier studies by Ginsberg, Mohebbi, Patel, Brammer, Smolinski and Brilliant (2008), that successfully applied search engine query data to detect influenza epidemics. Other fields, where search volume data has been applied are employment, sales, product pricing and housing market amongst others (Ettredge, Gerdes & Karuga, 2005; Radinsky, Davidovich & Markovitch, 2009; Choi & Varian, 2009; Huang & Penna, 2009; Wu & Brynjolfsson, 2009; Askitas & Zimmermann, 2010; Schmidt & Vosen, 2011; Baker & Fradkin, 2017).

1.2. Research objective, question and hypotheses

The COVID-19 and its consequences to the world economy during spring 2020, shaped the topic of this thesis in a particular direction. The crisis affected the traditional investment instruments such as stocks and indices. At the same time other, so called “safe haven” options raised their head. Even if the time period of this work includes only the early stages of the total effects of COVID-19, these actions have provided an interesting perspective to focus on. The total consequences remain unknown by the time this thesis is published.

During times of instability people get more worried and interested of what is going on in the world. Regarding finance related search activity, this is expected to reflect to the behavior of people in two ways. The first assumption is the expected increase in total search activity during crisis and the second assumption is the growing interest towards alternative investment vehicles. Established mindset among investors is that precious metals are seen to catch increasing attention in times of economic instability. This is

because as a category precious metals are seen to preserve their value well. Based on these relations the objective of this research is to study the effects of search behavior on precious metal ETF market. In other words, can future market movements of commodity ETFs be predicted with help of Google search term data.

Existing finance related literature contains multiple approaches related to search term volumes, where predicting and measuring stock market performance has been in the center of the study. In finance, a common way to measure the attention of individual companies is to use company names or tickers as search terms (Da et al., 2011). Another approach introduced by Preis et al. (2013) is to choose common finance related search terms and study the effect, not to individual stocks, but to a market index. In this research combination of both approaches is used. That said, it means two separate sets of search words are formed. The first set uses the first of the aforementioned approaches, where company tickers act as search terms, and the second set uses the second approach using generic search terms related to underlying assets of chosen ETFs. Both of these sets are compared to the performance of same underlying assets. ETFs enable the use of both methods as they behave similarly to equities and indices. The reasoning behind the forming of these search word sets are described in more detail in the data section of the study.

As its best the topic can offer a new way for measuring market sentiment in the financial markets. This specific topic using ETFs has not been studied yet and, as mentioned above, has been influenced by the financial distress in time of writing. US has been selected as the target market due to big size of the market. Large size in this context refers to a large commodity market and a high number of searches in the country. The research utilizes more recent data compared to the aforementioned reference studies. The Google Trends data has evolved over time by becoming more sufficient. This is expected to reflects to the results. Because Google Trends only allows access to weekly data for a period of five years. To calculate the variables used in the study, the period is shorter than the original five years. Therefore, the final period used in the study is from May 2015 to March 2020.

To have clear objectives in this study some research questions are formed. These questions are also put into a form of hypotheses, which describes better the actual phases

of this research. These questions and hypotheses are built based results of existing research and theoretical background. Few adjustments were made related to the specifics of this particular topic. The questions formed are presented in table 1.

Table 1. Research questions

i.	*Can the investor attention be captured by using search volume data?(main)*
ii.	*Is timing the market possible with Google Trends search volume data?*
iii.	*Can excess return be earned with moving average trading strategy?*

From which separate hypotheses were formed, which are presented in table 2.

Table 2. Research hypotheses

H1:	*Increase in SVI affect changes in the trading volume of the underlying assets.*
H2:	*Increase in SVI has a negative effect on the price of the security (ETF).*
H3:	*The SVI data from US acts as a better predictor than the global SVI data for commodity ETF price changes.*
H4:	*ETF ticker data (DATASET 1) acts as a better predictor than the generic search data (DATASET 2) for commodity ETF price changes.*
H5:	*There are differences in the effects of individual keywords on price changes*
H6:	*Excess return can be earned with a simple moving average trading strategy that takes advantage of SVI values.*

It is assumed that the use of Google search volume data offers some predictive power on ETF market movements. This assumption is done based on earlier result provided by

Preis et al. (2013) and Da et al. (2011). However, the relationship is expected to be weak. All of the hypotheses are examined for both aforementioned approaches. Other preliminary assumptions have also been made on the basis of the existing literature. The first one is that US data is expected to work as a better predictor than global data. This is simply justified by the lower noise level associated with the US data. Another preliminary assumption is that both search volume approaches are expected to have an effect. However, first approach by Da et al. (2011) is expected to yield more promising results. An interesting aspect is how the results differ between the approaches. For the fifth hypothesis, it is assumed that individual keywords have an effect.

The structure of this study follows general guidelines, in which the first chapter introduces the reader to the topic and its objectives. The second chapter contains the theoretical background of the study. The third chapter describes the data section and the fourth chapter focuses on explaining the methodology. Finally, we move on to the findings and conclusions to report the research findings.

1.3. Limitations

In this chapter we introduce limitations concerning the topic. Some of these are based on findings by earlier literature like Challet and Ayed (2013) and some are own observations that have come forward at different stages of the process.

First noticed limitation is related to Google Trends and its features. A factor that affects the selection of the target area is the fact that you can only view either a single country or global view at a time in Google Trends. It would bring some interesting new opportunities to be able the select the areas you want to take under consideration. Also, as an alternative to the global view it would bring new perspective to get your hands on a data for different continents like Europe. Thus, concerning this research the decision was made between the use of only the target region (US) data or both local and the global data.

Other limitation concerning the use of Google Trends are the release interval, and noise of the data. To breakdown these, release interval refers to the fact that Google Trends limits the availability of data. This became a critical factor in this work, since the data is

only available on a weekly level for the duration of five years. For a time period longer than five years Google automatically transforms the data into a monthly form. Combination of two different timeseries is not possible since the data provided by Google Trends is normalized according to the time period, and thus the scaling of different time series is not the same. In other words, Google only enables access to the relative change in search volume for a certain time period, and not actual numerical values of search activity. This affects the research by shortening the time period by two months, so that the variables used in the study can be formed. Similarly Google limits the access for daily data. At the moment, Google allows access for daily data for time period of nine months. However, this is an major upgrade from the previous 90 day access. Still, this means that there is no current means for deeper, more accurate analysis of the attention shifts. Finally, the noisiness of the data is related to the unwanted noise of certain search terms. Usually this occurs when a search term has a double meaning and thus gains higher search volume. This can be prevented up to a point by planning of search terms. For example, when using company names as search terms, it is impossible to know which of the searches are done financial perspective in mind and which of the searches are interest towards the products of the company (Da et al., 2011). In most of the existing studies, this has been dealt with using company tickers as search terms. Tickers are mainly used when acquiring financial information. This bias is closely analyzed during sample formation and opened up in the data chapter.

A bias raised by both Challet and Ayed (2013) and Preis et al. (2013) is a search term related bias. This needs to be taken into consideration, as this study also uses asset related general set of search terms (second approach). According to these studies, it can be assumed that search terms related to finance are more likely to be related to financial performance. This has been neglected by some researchers on the field. This bias can be controlled by choosing random non finance related terms and including them into the search term set. This enables us to know if apparent good performance can be justified with finance related search terms and not just with any search term. Preis et al. (2013) dealt with this by using Google Sets tool to suggest some of the search terms. Google Sets recognizes semantically related words, and thus, these words were not totally random. However, as a result of this step the set of search terms were not arbitrarily chosen. Human brain is wired so that it automatically chooses words that it knows are

related to the recent actions taken place on the financial markets. In this research, similar steps to research by Preis et al. (2013) are taken to tackle this bias.

Another bias raised by multiple studies is the effect of transaction costs. General approach among existing literature is that transaction costs are not included into the study. In this study, however, transaction costs are calculated in moving average calculations as close-to-reality estimates. The impact of transaction costs are acknowledged usually at the stage of methodology and research results and thus are not analyzed further in this section.

2. Theoretical background

This section presents the related theoretical background. The idea is to provide the reader a picture of different theories that have affected the field of finance and brought their own contribution to the development of the topic.

2.1. Development of modern finance

The theoretical background of this topic takes all the way back to the theory of the efficient market hypothesis (EMH) and random walk theory. However, these theories are only opened up superficially, as they are assumed to be familiar for the majority of investors.

Many experts continue to attach importance to the theory of market efficiency and believe that it is impossible to win the market in the long run. According to EMH, the information in the market is transmitted to the market quickly, which should not allow investors to benefit from the use of historical knowledge (Fama, 1970). Thus, investing in indices would be the best way for an investor to succeed in an efficient market. Even if there has been found a lot of evidence about the existence of different anomalies that enable excess returns in the market, these anomalies usually provide opportunities only in the short term. Regardless of this, up to this day investors are divided into two camps, both of which seek to provide evidence for or against the hypothesis of market efficiency. It comes as no surprise that the hypothesis of market efficiency is one of the most widely shared opinion theories in economics (Alexeev & Tapon 2011). EMH is regularly brought up for discussion, as current and new strategies and theories often challenge Fama's hypothesis.

Random walk theory that was popularized by Malkiel (1973) and his book "A Random Walk Down Wall Street" assumes that all price changes of stocks are random and so it is not possible to predict stock performance. The model was rejected later by Lo and Mackinlay (1988). However, it has had a big role in the evolution of finance as it laid the foundation upon which Fama built his theory of efficient market hypothesis.

Another important theory in the development of finance is capital asset pricing model (CAPM) that was introduced to the public already in 1960's. As the name indicates, it is a mathematical formula that can be used in pricing of capital assets and thus calculate a potential rate of return. The asset price is formed by multiplying average market risk premium and company-specific beta and applying them with the risk-free interest rate. Model developed slowly as an outcome of several researchers' working on the model individually. Big contributors working with the model have been Sharpe, Lintner and Mossin.

For almost 20 years CAPM was perceived as an undisputed fact and faced no criticism. Roll (1977) and later De Bondt and Thaler (1985) were among the first ones to raised questions concerning the model and started applying psychological factors in finance research. They found evidence that psychology has some predictive power over investor movements. According to these studies, reasoning behind this, is that investors in general are not rational. In fact, many of the investment decisions are made purely on emotional basis. Since then, many studies have agreed with these findings and deepened the research in the field of behavioral finance.

To conclude this section behavioral finance and efficient market hypothesis are cornerstones in modern asset pricing (Fakhry, 2016). These theories represent in a way opposite views. For this research behavioral finance sets the main theoretical base, and is thus opened below in more detail. This is reasonable, as the assumption behind this topic in general is to see, if it is possible to get some insight about investor behavior in financial markets by using search volumes.

2.2. Behavioral finance

In the history of financial theory, there have been two distinct changes. The first one of these was neoclassical finance that include the aforementioned efficient market hypothesis and CAPM. Current neoclassical finance consists of mathematical models and theories that are created to solve finance related issues. Simple mathematical models are not capable of describing real-life problems accurately and thus the models tend to be complex. These models are almost invariably based on different assumptions. Thus,

financial theory, to a certain extent, simplifies the true reality. For example, perhaps the most common assumption in financial theory is that every individual in the market constantly makes rational decisions. Indeed, financial theory and its models have been created to explain how the ideal investor should operate in the market, rather than to describe how the market actually operates. (Pompian, 2006)

The second milestone is considered to be the behavioral breakthrough that emerged from volatility research in the 1980s. Researchers started to incorporate models based on psychology and behavioral science into financial theory. Since then, the study of behavioral finance has been further accelerated by the discovery of numerous market anomalies, which constantly bring forward new evidence against efficient market hypothesis (Shiller, 2006; Fakhry, 2016; Joo & Durri, 2018). Ricciardi & Simon (2000) presented a simple figure representing how behavioral finance combines relevant fields of science.

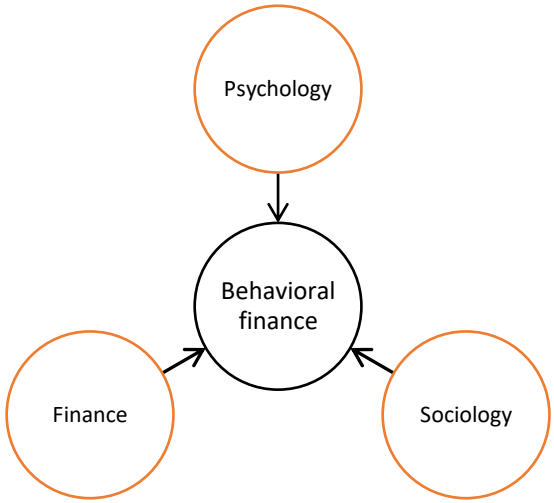


Figure 1. Relevant fields of science concerning behavioral finance (Ricciardi & Simon, 2000)

A direct definition of behavioral finance is the study of psychology behind financial decision making (Joo & Durri, 2018). To separate the traditional efficient market hypothesis and behavioral finance from each other, EMH believes that the stock prices reflects on available information from the market, whereas behavioral finance believes it is actually the behavior to that information by market participants that determines the prices (Fakhry, 2016).

Naturally, the advocates of behavioral finance argue that psychological factors have a significant impact on financial markets. Investors use heuristics as base in their decision making which makes them easily commit cognitive errors (Joo & Durri, 2018). Also, same type of errors committed by market participants are a result of similar heuristics of people. Thus, most decisions by investors are seen to be influenced heavily by different behavioral biases. Because of these biases, investors don't always reach the optimal and rational choice. Also, Hirshleifer (2001) agrees with this by stating "some cognitive tasks are just too hard for any of us".

Bodie, Kane, and Marcus (2005) separates the reasons behind irrational behavior into two. The first one is that investors don't process new information correctly, which leads the market to act on the basis of false conclusions. The second one is that investors manage to process the information correctly but end up still making inconsistent choices. The next chapter seeks to identify the causes of human behavior of investors.

2.2.1. Psychology in the stock market

By nature, people are social creatures and seek to communicate with other people. People are thus easily exposed to social influences. According to Baker and Nofsinger (2002), common sources that people are affected by are different media sources, trends and public opinions from colleagues and friends amongst other. Whereas, from the perspective of individual person, influences are seen to be psychological, which means the decision is usually a result of subconscious processing of information or a result of exposure to human emotions like greed or anger.

According to Nofsinger (2005), trends and overall atmosphere in financial markets, has an impact on decision making. Meaning, in a generally positive atmosphere, financial decision makers are prone to optimism. This optimism itself effects to the rationality of investors as it exposes to the overconfidence. The general negative mood in society, on the other hand, gives rise to pessimism and mistrust among financial decisionmakers. The risk may be exaggerated and the return on investment is estimated to be low. According

to the survey, the mood in society had a direct impact on stock returns. Stock prices tend to go up (down) when there is a general positive (negative) mood in society.

Investor sentiment related general market atmosphere is an interesting and relevant perspective that has been applied from different angles. Among others Edmans, Carcía and Norli (2007) study the impact of sports performance (football, cricket, rugby and basketball) on stock prices. The researchers argued that the final outcome (win or loss) of a sporting event ultimately has an effect on mood of investors. As a result, Edmans et al. (2007) reported evidence that football losses, had a statistically significant negative impact, on stock market prices in the losing country. However, there was no opposite effect found after won matches. Especially, the importance of football and to be more specific football losses was emphasized in the study. In every sport, the phenomenon could be considered statistically significant. The study also found evidence that the phenomenon was particularly strong for small companies. This was explained by the fact that the shares of small companies are largely owned by investors of the losing country.

Another similar study using a different angle was provided by Hirshleifer and Shumway (2003). They study the impact of weather on investors' market behavior. A comprehensive study monitored market behavior in a total of 26 different stock exchanges between 1982 and 1997. The study showed that the amount of sunlight clearly correlated with stock returns. In contrast, stock returns were not dependent on rain or snow. Traditionally, studies working against efficient market hypothesis have been explained by using for example risk premium, randomness, even tampering with research material or using erroneous research methods. However, the results by Hirshleifer and Shumway (2003) are very difficult to misrepresent by any of the above. The explanation can be sought mainly in psychology.

2.2.2. Behavioral biases

It has been discovered that investors are exposed to various different biases. As a concept, behavioral biases can be defined as a systematic error in the judgement of investors. The amount and variety of different biases distinguished by researchers in existing literature is high. However, the categorization of these biases is somewhat problematic. The

behavioral biases can be divided into cognitive and emotional biases, but the categorization has not found a consensus in the research field. (Joo & Durri, 2018; Pompian, 2006). The reasoning behind this is that behavioral finance sees cognitive errors having a bigger effect on investors than behavioral biases (Jureviciene & Jermakova, 2012). In this research it is not intentional to go into all of the different biases. Thus, only part of them have been selected for examination.

Let's first move on to cognitive biases. People tend to overestimate their own abilities. In financial literature this is referred as overconfidence bias. Barber and Odean (2001) study the differences in trading activity between men and women between 1991 and 1996. Surveys shows that men who opened their first investment account traded 45% more than women in the same situation, because they were more confident in their investment ability. For both men and women, trading over-actively reduced annual net income in the form of trading costs. The worst in terms of performance comparison were single men.

Overconfidence is a good example of cognitive bias. It is a state of mind where people have unwarranted belief in their skills regarding cognitive abilities, intuitive reasoning and judgment. (Kumar & Goyal, 2016) Another example related to cognitive biases is the use of reference points. Reference points are points that often indicates a price level, which investors get attached to for some reason. Investors form reference points on the basis of which they buy and sell securities. The purchasing price of a share is probably the most common reference point. These points might play a significant role as the investor is comparing it with the performance of the underlying asset. Thus, the investor sees his or her investment as either successful (positive) or unsuccessful (negative). Investors may also set reference points based, for example, on the highest price in the previous period. Even if the investment decision were not otherwise justified, the rules of sale resulting from reference points guide investors. (Baker & Nofsinger, 2002)

Next to introduce examples of emotional biases. Loss aversion is a well-known behavioral illusion that is categorized in the line of emotional biases (Kahneman & Tversky, 1979). This phenomenon is understood as a meaning that investors appreciate gains and losses of similar magnitude in different ways. By nature, investors try to avoid losses and thus are risk averse (Pompian, 2006; Srivastava, 2012). According to

Huberman (2001) and Srivastava (2012) investors prefer investments that are familiar and safe. Several studies have shown that loss has a psychological effect about twice compared to similar size gain. Avoiding losses very often leads to the maintenance of unprofitable investments and the premature sale of profitable investments (Pompian, 2006). Loss aversion is also closely related to another important bias called disposition effect. Where in general investors are keen to sell the winners and hold the loss-generating asset (Kumar & Goyal, 2015). Empirical evidence of disposition effect has been provided by various researchers as Shefrin and Statman (1985), Frazzini (2006) and Barberis and Xiong (2009).

Another bias related to emotional biases is regret aversion. It is natural for people to avoid decision making or consciously speed it up. This is a result of the fear of making a bad decision. The investor will therefore try to avoid the pain of a possible bad choice beforehand. This phenomenon is not just about loss aversion, since it is not easy to sell a profitable investment without fear of losing even higher returns (Pompian, 2006). Shefrin & Statman (1984) noticed that the phenomenon causes investors to favor large dividend shares, as the certainty of partial compensation in the form of dividends lowers the investor's buying threshold. Investment to a share that generates a high dividend yield will also decrease the selling pressure.

Kumar and Goyal (2015) published a literature review related to behavioral biases in financial decision making. They concentrated on four most relevant biases regarding the topic which according to authors and existing literature were overconfidence, disposition effect, herding and home bias. Both overconfidence and disposition effect are already opened above. However, herding and home bias are still unknown to the reader. As the name implies herding refers to a situation where the rationality of investors suffers as a result of investor imitating the judgement of others. In other words, the decision-making process of investors is influenced by the decisions of large public. Let’s face it, people constantly follow what other people do, and also often imitate them. Lee, Liu, Roll, and Subrahmanyam (2004) state that individual investors are more prone herding than institutional investors. Other important studies regarding the herding bias are published by Grinblatt, Titman and Wermers (1995) and Lakonishok, Shleifer and Vishny (1992). Finally, there is home bias that refers to a situation where investors rather by and hold

domestic investments than foreign securities (Kumar & Goyal, 2015). Initially, the bias was introduced by French and Poterba (1991) and Tesar and Werner (1995) who analyzed the bias in their studies. According to these studies possible reasons behind home bias are transaction costs, investment barriers and information asymmetry among others.

2.2.3. Noise trading

Morgan (1997) categorized decision makers in financial markets into four different groups that are rational arbitrageurs, smart money investors, noise traders and passive buy and hold investors. To break down this classification, a smart money investor referred to institutional and other large private investors. These investors control large amounts of capital which requires extensive information gathering and investing knowledge. The investment decision is rarely made by one person alone. Thus, this group is seen to be highly rational when compared to smaller individual investors that are referred as noise traders. This study focuses on noise traders as a group, as Google search volumes reflect the opinions of this particular group.

As smaller investors, noise traders are often willing to take greater risks in the hope of higher returns. Other critical factors for noise traders are lack of investment knowledge and overconfidence which results to the use of “gut feeling” and personal beliefs to base their investment decisions. Tokic (2007) sums up that these types of investors follow the multitude, by buying during upturns and selling during downturn. They also unconsciously tend to ignore the negative news regarding their portfolio, but react strongly to positive ones, leading ones again to overconfidence. Shleifer and Summers (1990) found out based on psychological tests conducted to test subjects that noise traders use same reasoning, logic and heuristics in decision making and thus end up making same kind of mistakes over and over again. These mistakes clearly indicate a pattern which can lead to false signals, demand shifts and even reflecting to security prices.

Noise traders evoke both positive and negative feelings. However, majority of people think it as a solely negative thing. This is because noise traders act irrationally and thus effects the efficiency of the market. According to Black (1986) this isn’t entirely true. Noise traders at the same time provide more liquidity to the market. That said, the noise

makes a great deal in making financial markets possible. Also, De Long, Shleifer, Summers and Waldman (1990) came to the conclusion that noise trading is not as unambiguous as one would think. The study introduced a new risk called noise trading risk, which was born with noise trading. According to De Long et al. (1990) noise trading risk should be included to the prices of securities. It means that popular stocks favored by noise traders (noise trader stocks) are more volatile and also expected to have stronger reversion back to their long term mean prices.

2.2.4. Sociological factors & investor attention

According to existing literature, investor attention has been measured in numerous ways. Researchers use different indirect proxies to represent the attention. Barber and Odean (2008) used trading volume, media attention and extreme returns as a proxy, Seasholes and Wu (2007) price limits and Chemmanur and Yan (2019) advertising expanses. According to the results of these studies, high-attention shares are more likely to attract the attention of investors. These researches discovered that investors tend to choose shares with high attention levels and only then start applying own selection criterion and preferences. However, in this modern world, these studies use simplified thinking by assuming that high attention levels of one proxy, result to investor attention. Of course it is possible, but in a world where people have easy access to unlimited sources of information, it is unlikely. As a result, the question of what to pay attention to, have become even more critical. Regarding this research Google search term volumes are used, which is an approach first introduced by Da et al. (2011). The assumption behind this is that Internet searches reflect the current interest and attentions of investors. Major proportion of search activity in the target market goes through Google, which makes the webpage eligible for this study.

A question related to investor sentiment that is under discussion is how the sentiment transitions to the underlying asset. This issue is viewed from a finance perspective. The existing literature contains different views on how market sentiment transitions to market price. Preis et al. (2013) show that drops in the financial market are preceded by investor concern. This means, the attention increases prior to market drop. The study also suggest that the aforementioned logic can be used to construct a profitable trading strategy for

query volumes of certain search terms. Another view in existing literature is based on attention theories by Barber and Odean (2008). This aspect battles against the view of Preis et al. (2013), as rise in attention is expected to have a positive effect on underlying asset. The theory is based on assumption that retail investors own a limited number of shares, thus it is more likely that their intentions are towards buying rather than selling. The collective attention of retail investors causes momentary price pressure as retail demand for the underlying asset increases. Klemola (2019), on the other hand, noticed that when predicting stock market performance based on attention shifts, downward market predictions were more apparent than upward prediction. According to Klemola (2019), this was because people have a stronger reaction to negative news and outlook than positive. In most cases, interest in the search word predicts a decline in the underlying asset. Bijl, Kringhaug, Molnár and Sandvik (2016) investigated whether search query volume of a company name can be used to predict stock returns (weekly) for an individual firm. They found that high search term volume predicts low future returns. However, the relationship is too weak to be used as a profitable trading strategy, since the positive gain is offset by transaction costs. Findings of Klemola (2019) and Bijl et al. (2016) are in line with the findings of the reference study by Preis et al. (2013).

As sociological factors in this study, the effect of media and internet, herding or in other words flock behavior and extreme market conditions, are considered. A good example of how sociological factors influence the decision making is the housing bubble in the US. It can be described as a fear of missing out. One of the driving factors behind the financial crisis was discovered to be the fact that people participated in the housing market, simply because they didn’t want to be left out in this great opportunity. Caparrelli, D’Arcangelis and Cassuto (2004) and Zhou and Lai (2009) agrees that this kind of flock behavior is known to lead into extreme market conditions. However, both studies also found this kind of behavior to be related to the market under observation. Market situations behave differently in all markets.

Own contributions to the discussion around sociological factors were brought by Sornette (2003), who especially paid attentions to the effect of network and surroundings of investors to the herding behavior. According to the study, there are two main factors that affect this kind of behavior of investors which are network and media. With network,

Sornette (2003) referred to an open market situation were usually all the transactions done by an individual are known by the network. The network usually consists of fellow investors, family, friends and other random members, which are seen to have an impact to individual's investment activity. During the recent years, the form of a network has changed. The ability to use online chat, public forums and use different messaging services extends the network and provides the ability to influence opinions of other people significantly.

This gives us a good gateway to move on to the effects of media and internet in investor sentiment. We live on an era of social network. People are spending a lot of time in front of a screen, both at work and on free time. Within a few clicks you have the access to endless amount of information. Search engines and especially Google are dominating the internet and it has become a highly important source of information. However, Google as well as the entire internet is full of stimuli designed to influence or appeal to people. On Google, this is reflected in the form of advertising mixed with search results. Another powerful influence of the Internet is the media, which includes e.g. news and social media. Shiller (2005) states that media is a market participant that creates speculation in the market and even generates price movements. The media also plays a big part in building investor incentives, leading the audience in a certain direction. According to Shiller (2005), this is because public unconsciously demands drama to stay entertained. This is why negative news is reported more than positive news. Magazines and television broadcasters try to meet the demand to stay competitive. Roughly half of the topics in national news in US are either sports or finance related. A speculative factor of finance related newsfeed is that it is often related on the current market conditions. This means that if the market is booming the newsfeed is bringing positive news and success stories to boost the public confidence and sentiment. This works vice versa during the downturns. Tetlock (2007) found that it is especially small companies that are prone to be influenced by news. These smaller companies are also the ones that react easier to positive and negative news. Whereas, according to Lucey and Dowling (2005), popular companies in media are considered to be the best investments by investors. They added that, by simply comparing the media coverage of companies, it is possible to discover which companies draw the investor attention.

The effect of media in the economy and politics took a new shape with the election of President Donald Trump. Already during the election campaign Trump started to use Twitter as a tool to communicate with the public. From since, couple other high-profile politicians have joined Trump in the use of Twitter. President Trump has been receiving a lot of criticism and many people claim that the president is pursuing his own interests through his tweets. These tweets have also captured the interest of researchers' in different fields since these tweets are claimed to have a clear impact on, for example, the economy. Juma'H and Alnsour (2018) published a research called: "Using social media analytics: The effect of President Trump's tweets on companies' performance". Event study was conducted to determine if there is a relationship between President Trump's tweets' and daily changes of both targeted companies and major indices. As a result, they found no significant effects caused by Trump's tweets. However, the recent discussions and signs indicate that possible effects of these tweets are possible. Despite the exceptions, these tweets always provoke debate, as well as cross the news threshold, even on an international level.

When writing this thesis during spring 2020, the present market provided another good example of extreme market condition. Corona virus started spreading around the globe affecting also the financial markets. This is mostly a result of media and news coverage received by the topic. The finance related media has been stirring up the audience, which has led to distress among investors. Of course, new infection has arisen, but with news like "Coronavirus Could Infect Two-Thirds of Globe" by Bloomberg, it is no surprise that investors have started to move their assets to more secure options (Lauerman, 2020).

2.3. Previous literature

Using search queries is fairly new form of study within finance. The existing literature has provided some promising results both within finance and in other fields. Similar to other investing strategies, the use of search engine query volumes is a way to challenge the Efficient Market Hypothesis (EMH) by Fama (1970) in the hope of excess returns. EMH is one of the most well-known and argued assumption within finance. However, the approach towards EMH has become more critical and the focus in research has turned more towards on the impact of investor sentiment (Bijl et al., 2016). This is also the main

ideology behind behavioral finance. To capture the investor sentiment researchers have started to apply available data from wide range of different sources as news articles, Twitter, Wikipedia and Google Trends. (Tetlock, 2007; Bollen, Mao & Zeng, 2011; Preis, Reith, & Stanley, 2010; Moat et al., 2013; Challet & Ayed, 2013).

The first researches published, related to search engine query volumes, were not strictly related to the stock market performance. One of the first researches to use tracking of common search terms was published by Ettredge et al. (2005) with a goal of studying macroeconomic factors that have an effect on the unemployment rate in US. Around the same time Cooper, Mallon, Leadbetter, Pollack, Peipins and Jansen (2005) published a research that studied cancer-related topics with the help of search engine volumes. After that the amount of research related to the topic has increased exponentially. Tetlock (2007) studied the role of media in the stock market by utilizing data from Wall Street Journal column. After this came Choi and Varian (2009 & 2011) that followed the steps of Ettredge et al. and expanded the research to multiple fields. They also draw attention to the variation of result accuracy of studies in different fields. The focus of researches by Choi and Varian were on fields related to sales and tourism. Sales in general have also been studied by other researchers. Among others Kulkarni, Kannan and Moe (2012) point out that search terms can indicate the purchasing interest of consumers. The use of search term data has also been expanded to the field of epidemiology by Polgreen, Chen, Pennock and Nelson (2008) and Ginsberg et al. (2008).

In the field of finance, Da et al. (2011), Preis et al. (2010) and Joseph, Wintoki and Zhang (2011) were among the first ones to study Google search query volumes. Da et al. (2011) showed that SVI can be used as a proxy for attention but it differs from earlier forms of attention measurements used. They also pointed out that SVI is a tool to capture specifically retail investors attention and not the attention of institutional investors. This due to the fact that institutional investors use more sophisticated tools, such as Thomson Reuters and Bloomberg terminals. Preis et al. (2010), on the other hand, found that there was no significant correlation between returns and volumes of Google searches for company names. However, what they were able to find was evidence that trading volumes can be predicted by using search query data. Preis et al. (2013) used a different perspective in a latter study where instead of using company names as search words they

used common finance related search terms. With the use of this strategy they found that there is a possibility to get significant positive returns. Over the 7-year time period, a strategy that bought, or sold the market portfolio based on the volume of certain search terms were able to outperform the market index by 310%. Similar positive evidence was presented by Moat et al. (2013), who instead of using Google search query volumes used page view changes of Wikipedia as an input.

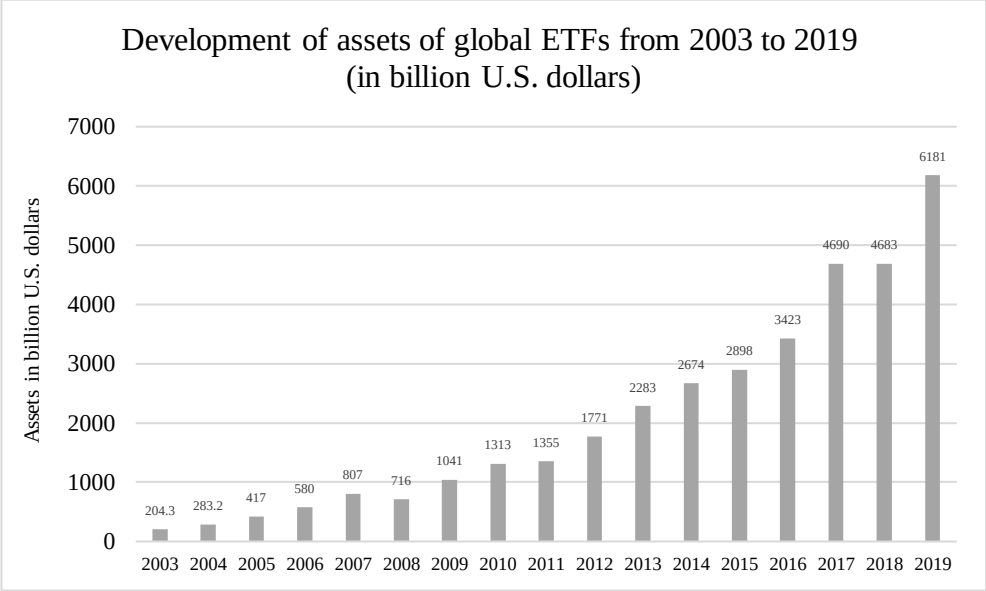
Most of the existing literature introduced above has a focus on the US market (Cooper et al., 2005; Ginsberg et al., 2008; Preis et al., 2010; Choi & Varian, 2011; Challet & Ayed, 2013; Preis et al., 2013; Dimpfl & Jank, 2016 amongst others). The reasons behind this is the size of the market. To be able to utilize the investor sentiment from the market, large dataset is essential part. Bigger search volume going through Google means more data. It is under debate whether the volume of data provided by search engines is enough large to enable the prediction of future returns. Challet and Ayed (2013) discuss these claims in their research "Predicting financial markets with Google Trends and not so random keywords". They also review possible other biases related to the topic. Even if the topic is seen to be bounded to large market regions only, there are some studies that have been conducted using other regions too. Aouadi, Arouri and Teulon (2013) studied evidence from France, providing promising results by showing that in France, Google search volume can be used as a reliable proxy for investor attention. Whereas, Schmeling (2009) provides more international evidence by studying 18 different industrialized countries. In line with the existing literature Schmeling (2009) show that sentiment has a negative predicting power to future stock returns on average. Wuoristo (2012) focused on the UK with an interesting perspective of using both global and UK search volume data. The results show that global search volume data is better for predicting stock price movements than data from the UK. Finally, Klemola (2019) argued that the Google search volume in smaller regions like Finland is not enough large to conduct a study of this kind. Thus, Klemola (2019) also studied the connection between Google searches and future US stock market returns in his doctoral dissertation.

While interpreting the variation of results from existing literature, it is noticeable that many of the results are in line with one other. This applies especially to finance related research that has the main focus in this literature review. Consensus has been found that

search term data has predictability over trading volumes. When predicting stock market performance, downward market predictions were more apparent than upward predictions (Klemola, 2019). This is explained by investors' stronger reaction to negative news than positive. However, there are also topics within the subject on which a strong common ground has not been found. One of these is the discussion about whether the size of the existing data is overall enough large to be used as proxy for investor sentiment. This is a result from a question; can search terms have an actual effect on the market? Barber and Odean (2008) justify this by stating that retail investors possess only limited amount of stocks, as majority of stock are owned and controlled by large institutional investors. However, many researchers have found a relationship between search term volumes and stock market performance but how big is the actual predictive power? The direction of the relationship found is that high search term volume is seen to indicate low future return, and vice versa, low search term volume indicates high future return. However, in most cases the relationship between the two is small like the result provided by Bijl et al. (2016) shows.

2.4. Exchange-traded funds

ETF is the abbreviation for exchange-traded fund, which is commonly acknowledged among investors. The abbreviation is used throughout this study. ETF is an investing vehicle that tracks the performance of an index or another asset and behaves much like stocks and indices. Thus, ETF shares are traded publicly in the secondary market via stock exchange. Most of all, it offers low-cost diversification for investors. ETFs have in a way democratized investing as they allow both small and large investors to access previously inaccessible or cost-prohibitive markets globally. The first generation of ETFs was published in the 1990's. From since, the basic idea has stayed the same. Gastineau (2001) was one of the first to document the creation/redemption process of ETF. The industry has grown significantly over time and is now one of the biggest sectors in financial markets. The graph below describes the development of assets in the field of ETF.



Graph 1. Development of assets of global ETFs 2003-2019 (Statista, 2020)

The first proper ETF is seen to be the S&P 500 Trust ETF (SPDR) released in 1993 by the State Street Global Investors. It was back then released with its original name the Standard and Poor's Depository Receipt (SPDR). This ETF is still actively traded and has become one of the most popular and largest ETF, with assets worth over 260 billion dollars. From one fund, the ETF market grew to almost 1000 funds by the end of 2009. (Fuhr & Kelly, 2009) Currently the ETF market contains over 6000 funds, which are traded globally.

Over the time, wide variety of different ETFs have entered the market. Investors are now provided with a chance to diversify their portfolios between for example different asset classes or geographical location. The variety of ETF based on asset class is opened up in the table 3 below. The table from 2017 shows that most of the ETFs are equity-based funds. However, each asset class category can also be divided into smaller subcategories. Within equity-based ETF category investors can, based on their interest, choose across different characteristics of ETFs, like specific industries, regions or company size. As an example, it is possible to choose an ETF that includes only small-cap companies in US or ETF that contains growth companies in emerging markets.

Table 3. ETF Assets Under Management (AUM) by asset class in 2017 (Maier, 2017)

Asset Class	***Asset ($M)***	***Market share (%)***	***Number of ETFs***
Equity	*2 394 156*	*78 %*	*1219*
Fixed income	*518 698*	*17 %*	*260*
Commodity	*63 410*	*2 %*	*123*
Alternative	*47 410*	*2 %*	*358*
Tax Preferred	*27 634*	*1 %*	*37*
Asset Allocation	*7 273*	*0 %*	*48*
Convertibles	*4 430*	*0 %*	*3*
Total	*3 063 011*	*100 %*	*2048*

The literature review on modern ETF research by Charupat and Miu (2013) divides the existing research into three main aspects which according to authors are ETF pricing efficiency, performance and tracking ability, and the effects of the ETF on its underlying assets. This research assumes that market sentiment affects the prices of ETFs and by using Google search volumes the data would indicate about possible market movements. Thus, if the results of this research show a relationship between market sentiment and ETF prices, it is possible that market sentiment is one of the explanatory factors behind mispricing of ETF.

2.4.1. Commodity market

Like the table 3 above shows, ETFs offer options also for investors with an interest towards commodity markets. In most cases these commodity ETFs invests in actual physical commodities, such as precious metals (e.g. gold and silver), agricultural goods (e.g. corn and cotton), or natural resources (e.g. crude oil and coal). This means that investors who are interested in a commodity market do not have to buy crude oil, for example, but instead can buy a set of contracts that are backed by the commodity. Another reason that makes commodity ETFs popular is that they offer investors exposure to commodities without having to learn how to purchase futures or other types of derivative products. Investment decisions regarding the ETF are part of the role of the ETF's portfolio manager.

ETFs possess unique features compared to more traditional investment vehicles. A commodity ETF can either focus on a single commodity or variety of different commodities. It can also track closely the performance of some commodity index or use derivative contracts as a way to invest in the commodity market. If the commodity ETF is physically backed, it means that ETF physically owns the underlying asset. ETFs can also invest to companies which are producing the underlying asset. These are often referred to as miners ETF. For precious metals such as gold, it is common for an ETF to be physically backed at least up to a certain point. ETFs offer significant opportunities for diversification. By purchasing one share of an ETF, investor gains exposure to all assets underlying the ETF. The London Stock Exchange (2014) and ETF issuer Vanguard (2013), conclude that there are mainly two ways for ETFs to track the performance of their underlying assets: physical and synthetic. The physical method refers to the situation where ETF owns assets that it is tracking. Synthetic tracking refers to a situation where the underlying assets are not physically owned but are monitored using derivative contracts.

Because an ETF tracks another asset, the so called fair value of the ETF is approximately the value of all included assets. In addition, the ETF's NAV should reflect the price at which the ETF trades on the secondary market. If deviation occurs between the price of the ETF and its NAV, ETF can be either under- or overpriced against NAV. In existing literature, these significant deviations are referred to as ETF mispricing’s or pricing inefficiencies. If such a pricing inefficiency occurs, ETFs have unique arbitrage process to help setting prices back to equilibrium (Delcoure & Zhong, 2007). This process is also known as ETF creation/redemption mechanism and is described in figure 2. It works similarly for equity ETFs and commodity ETFs.

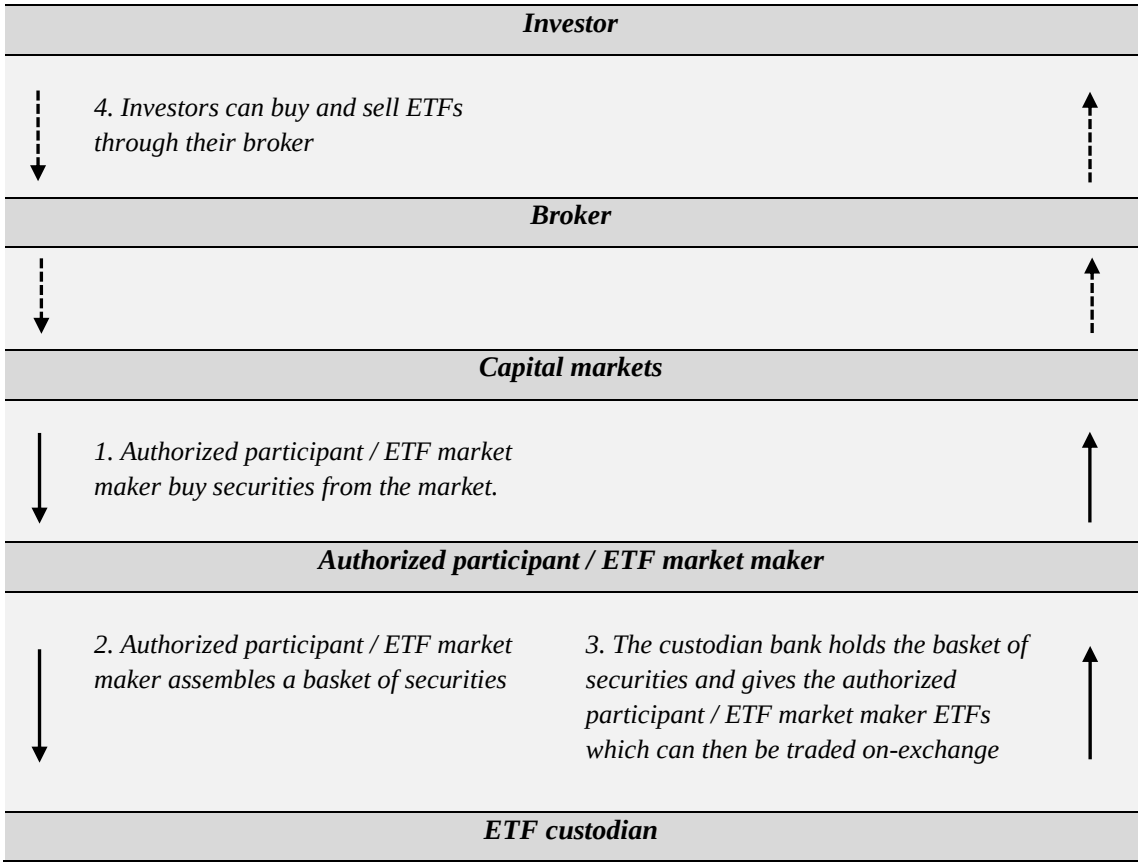


Figure 2. ETF creation/redemption process (London Stock Exchange, 2014)

The different steps of creating and redeeming process are numbered in the figure 2 above. The process starts with an authorized participant (AP) who acts as market makers and liquidity providers and an ETF company that issues the ETF. APs are institutional entities with a lot of buying power which have been granted AP status. Large authorized participants in the U.S are Bank of America (BAC), JPMorgan Chase (JPM), Goldman Sachs (GS), and Morgan Stanley (MS), among others. The process is similar whether the ETF issuer wants to create new shares to meet growing market demand or launch a new product. When an ETF company wants to issue new shares, it is up to the AP to acquire the securities that the ETF wants to hold. The process is repeated daily and starts by ETF issuers publishing a creation basket, which lists the securities it is willing to accept in exchange for creating ETF shares. An AP that is willing to acquire new ETF shares delivers required securities to the ETF issuer in exchange for new ETF shares. Creation and redemption occurs in large quantities of ETF shares which are typically referred as creation or redemption units (usually in blocks of 50 000 shares). In the process, AP may have to pay some fees. When converting ETF shares back into actual assets similar

process occurs but vice versa. In this case the ETF company offers a redemption basket which is a list of securities they are willing to offer in exchange for ETF shares. According to Vanguard (2013) the process is similar for synthetic ETFs, but regarding physically backed ETFs, cash transactions take place instead of physical asset. The redemption and creation processes between ETF issuer and AP is the primary market for ETF shares. Regarding commodity ETFs, the largest ETF operators control a major portion of total investment assets under management (AUM). This group of ETF issuers include Blackrock's iShares, State Street's SPDR, and Vanguard, among others. (Blackrock 2017; Deville 2008; Vanguard 2013)

The secondary market for ETF shares occurs when APs open the created ETF shares for public trading. Trading takes place via stock exchange. Most of the ETF in the original sample of this study are traded on the NYSE Arca. It is an exchange on which both stocks and options are traded. NYSE Arca stands for Archipelago Exchange. When AP has acquired ETF shares it has possibility to trade them in exchange in similar manner than regular shares. Thus, the secondary market attracts the majority of ETFs traders. Because APs are involvement in both primary and secondary markets, they can manage the amount of ETF shares available for public trading. For the same reason, APs also play a major role as maintaining ETF pricing efficiency. (Vanguard 2013; Blackrock 2017)

The price efficiency controlling of APs is referred to as ETF arbitrage. In theoretical perspective it is seen to ensure efficient pricing in the secondary markets. In practice, empirical evidence of ETF mispricing has been found (Petajisto, 2017). Thus, the evidence indicates that arbitrage mechanism may not work as effectively as the theory suggests. The deviation occurs when the market price of an ETF differs from its NAV value. In ETF arbitrage, APs use their role to create and redeem ETF shares to generate arbitrary profits. The actions are dependent on the direction of the spotted deviation. In case of ETF price is higher than its NAV, AP can buy the underlying assets of the ETF and create new ETF shares to be sold on the secondary market in order to earn arbitrary profit. Same works vice versa, when ETF price is lower than its NAV. In this case, AP can buy the ETF shares to be converter back to assets to earn arbitrage profit. There is no obligation by law or other counterparties for APs to address these anomalies. Thus, it is left for APs to determine whether the arbitrage profit is significant enough to take action.

This means that small differences may arise between ETFs and their NAV. (Blackrock 2017; Petajisto 2017)

Depending on the type of ETF chosen, the cost of an ETF can vary widely. ETFs include two types of costs. The first one is the expense ratio, which consists of fund's annual operating expenses. The size of the expense ratio is dependent on whether the fund is actively or passively managed. Most of the ETFs are passively managed. The difference between the two is that while passive ETFs are designed track an index, actively managed ETFs seek to outperform their index through the services of a portfolio manager. The average ETF carries an expense ratio of 0.44%, which means the fund will cost you $4.40 in annual fees for every $1,000 you invest. However, over the past two decades, fund expenses have trended significantly lower providing ETFs with expense ratios as low as 0.03 % per year. The second cost associated with an ETF is the cost of trading. This is determined on the basis of ETF's bid-ask spread. The cost of commodity ETFs is generally higher than the cost of the most common market index ETFs.

Overall commodity market offers interesting approach to investing and a great way to diversify your portfolio. The debate on commodity markets often raises its head in times of economic turmoil. This is because gold is known for its ability to preserve its value. Precious metal ETFs can't go bad or spoil. This study focuses on commodity markets and especially to precious metal ETFs.

3. Data

This chapter introduces the reader to the data used in the study. The first subsection describes how the original sample is formed. The next two subsections describe data and data processing from two sources. The first data source is Thomson Reuters DataStream which is used for collecting financial data. The second source is Google Trends, which is used to collect Google related search volume data. The final sample consists of financial data and SVI data. The financial data in this work that contains information of five ETFs: GLD, IAU, SLV, GDX and GDXJ. Full names are presented in table 8, later in this section. Due to the adequacy of the search query data, the size of the final sample was reduced from the original 22 ETFs to five ETFs, which is a small sample. This will negatively affect the reliability of the results. In addition to the financial data, there is the SVI data, which includes the data related to search words. The total number of search words included to the study was 19 from the original amount of 42 search words. The final sample can be found in appendix 3. Data is retrieved for the time period from March 2015 to March 2020. Next, the entire data filtration process is described.

3.1. Sample

The sample section is divided into two subsections, where sample formation is described for both financial data and search query data.

3.1.1. Sample formation for financial data

From data point of view this study differs widely from the reference study by Da et al. (2011). First of all, this research focuses on ETFs, whereas the reference study uses equities. Unlike individual stocks, ETFs are funds that often contain several assets or even several different asset types. One of the characteristics of ETFs is the potential for diversification. This means there are wide variety of different ETFs focusing on different market regions, company sizes, industries, asset classes, or combinations of all of the above.

The inclusion of different ETFs in this study would mean that several benchmarks are also needed. This is because it is not possible to find a one-fits-all type of index for all the different ETFs. To simplify the research and aim to replicate the reference study, the decision was made to focus on one specific category within the ETF market.

It was the prevailing market conditions during the time of making this thesis that aroused interest in alternative investment options and especially towards an approach using precious metal ETFs. The initial though behind the approach was to see if precious metal ETFs really indicate nervousness on the financial markets. Due to the limited number of precious metals ETFs, the focus of this study is on the US ETF market, which is the largest ETF market in the world.

The total U.S. ETF market today includes more than 2,000 ETFs with assets under management of more than $ 4 trillion in 2019. According to an article by Market Insider, the total ETF market will hit predicted $5.3 trillion in assets at the end of 2020 (Reinicke, 2019). From the total commodity market, which can be invested thought ETF market, precious metals were selected because of their characteristics. The market is centered around a few major ETF providers such as Vanguard, iShares and SPDR. The justification for choosing precious metals is that precious metals form a clear and easily defined entity. Also their ability to attract interest and perform even during economic downturns suited the topic of this research well.

The following figure 3 describes the different type of ETFs, from which the final sample is formed. It was decided that leveraged products, the so-called ETNs, are excluded from the study. Among all commodity ETFs, the strict gold focused ETFs, other precious metal based ETFs and precious metal mining company based ETFs, were chosen. When measured by assets gold based group contains the largest ETFs. Other precious metal ETFs contain for example silver and platinum based ETFs and mining company based ETFs contain companies that produce precious metals. Thus, the last group contains actual equities like Newmont Corporation (Ticker: NEM) and Barrick Gold Corp (Ticker: GOLD).

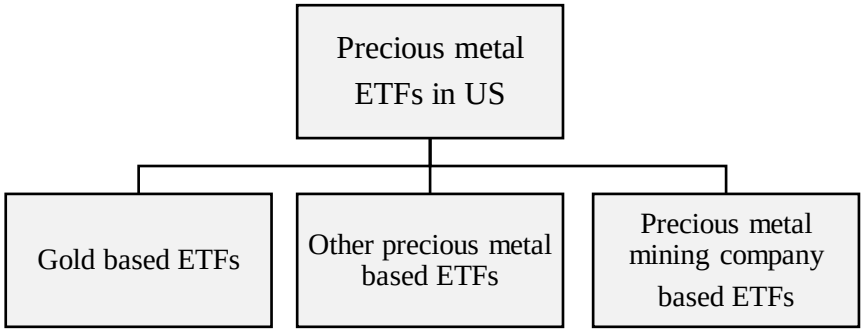
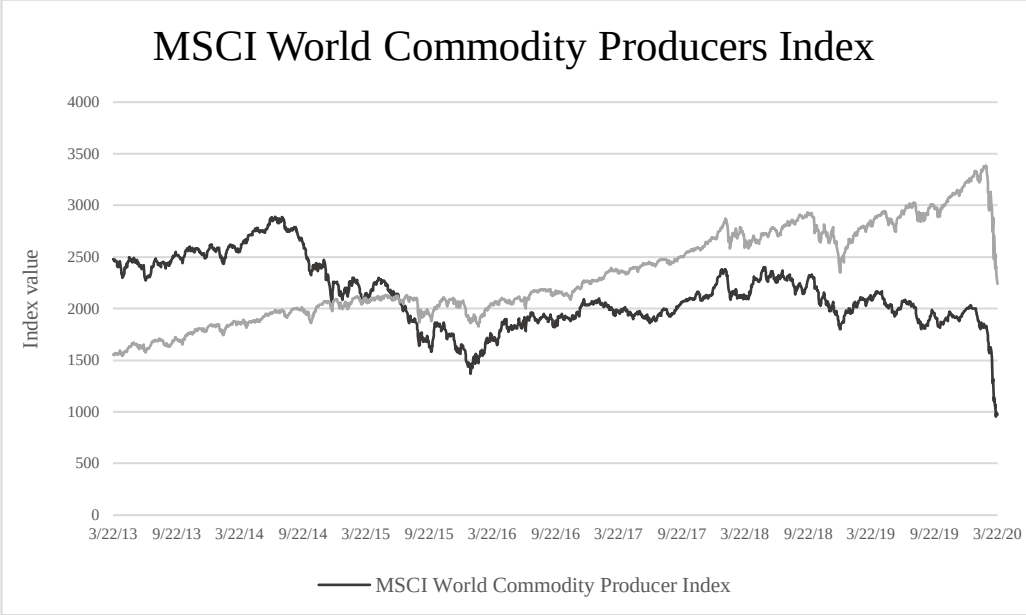


Figure 3. Sample composition

All ETFs belonging to these groups were selected, with some restrictions. The minimum level of assets under management (AUM) of the ETFs included in the study must be $ 10 million. This level is set to ensure that the selected ETFs have enough trading volume. Also the inception date of each ETF was checked and only ETFs which were created before 2015 were included to the study. The time period used in this study was five years from March 2015 to March 2020. The final sample consists of 22 precious metal ETFs from the original amount of 88 commodity ETFs, larger than 10M$ in assets. This sample was formed before integrating Google related SVI data into the review. A list of ETFs selected on the basis of financial information criteria can be found in appendix 1.

To the purposes of this research the MSCI World Commodity Producers Index was selected as the benchmark index. Financial data for the index is available in DataStream. As the study focuses on the US markets, the price index is measured in US dollars ($). The following chart gives a picture of the price development between the comparable commodity sector and the S&P 500.



Graph 2. The performance of benchmark index and S&P 500 Index during 2013-2020 (DataStream).

From the graph 2, it is noticeable that during the whole time period, the broad trend for S&P 500 has been positive, whereas the trend for MSCI World Commodity Producers Index it has been negative. Also, when looking at the negative spikes in the S&P 500, the value of the commodity producer index has also fallen over the same period. Thus, no high opposite market movements are observed for the benchmark index.

3.1.2. Sample formation for search query data

As mentioned in the introduction chapter, this study implements the approaches of both reference studies by Da et al. (2011) and Preis et al. (2013). Regarding the sample formation of search query data it means that the study has two separate sets of search terms that form the final search query data.

The first SVI approach is built upon the reference paper by Da et al. (2011), where financial tickers of ETFs act as search words. Da et al. (2011) justified the use of a ticker as a means to reduce search volume noise. According to the study, when searching with a company ticker, the search is more likely to be finance related and describe investor sentiment. Regarding this research, another reason for choosing ETF ticker, is the usual long length of ETF names. The longer the search term, the narrower the search becomes, which critically influences the search volumes. For example, when searching for an ETF

called "*Aberdeen Standard Physical Swiss Gold Shares ETF*", Google Trends cannot provide results because not enough data matching the query was found. The website show remarkably different results when using the ETF ticker "SGOL" of the same fund.

The process of generating the first set of keywords is precise and straightforward. However, the process of forming the second set involves more uncertainties. The second set of SVI data is built upon the approach of Preis et al. (2013). The search terms consists of more generic keywords, which are defined to be related to the market and underlying assets of ETFs. This means possible search words could be words as "Gold", "Gold ETF", "Financial markets" and "Precious metals", amongst others. Also common market indices "S&P 500" and "VIX" were included to the research as search words. The purpose of adding these terms is to look at how market nervousness as measured by changes in search volumes is transmitted to commodity ETFs. Preis et al. (2013) showed promising results in their study, which focused on measuring the performance of 98 finance related search terms. The reasoning behind incorporating both approaches is to enable the comparison between the two ideologies by Da et al. (2011) and Preis et al. (2013). This comparison forms one of the hypotheses (H4) of the study.

The selection process of search words for the second set was kept simple. To prevent biased search words, it was determined that search words cannot be time-bound to certain period or related to actual performance of the market. Good examples of this type search words are crises and epidemic related words, which are known to cause turmoil during relevant times. To determine suitable keywords, website called Keyword Tool was used, as a help (Keyword Tool, 2020). This follows a procedure, similar to what Preis et al. (2013) used, where the aim is to prevent the formation of biased data. The amount of search terms was limited to 20, to ensure effortless processing and control over the growing amount of data. The total list of search words chosen to the second dataset are seen in appendix 2.

The second approach also introduced class filtering provided by Google Trends. The search volume for each term was taken by filtering only finance related searches. This exposes the data for new biases as the exact procedure behind Google's filtering activity is unknown. At the same time, this step is seen to reduce noise of the data significantly,

which is seen as a bigger problem for this type of research. This filtering action was possible for generic search words, as the overall level of search volumes is high. For the first set, this could not be done because the effect on data adequacy was too large.

In addition to the two approaches described above, data is collected from selected keywords both locally and globally. This is possible by using geographical filtering option provided by Google Trends. This aspect was not taken into account by Da et al. (2011) or Preis et al. (2013). The idea of including both global and local data is to see which data is a better indicator of investor sentiment. Therefore, global data compared to U.S. data as predictors constitute one of the research hypotheses (H3) in this study. Global data generally contains higher search levels than U.S. data. At the same time, global data is also expected to contain more noise.

Global data can be included because interest in investor sentiment is at a general level and not just at the level of U.S. investors. The U.S. financial market is an important market in which people participate around the world. Another reason is that international investors are also able to trade in the U.S. markets without any strict requirements or restrictions.

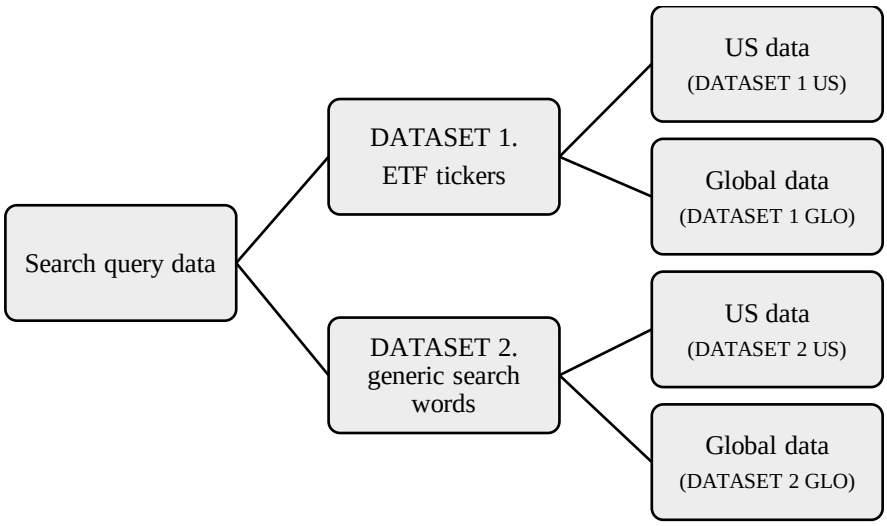


Figure 4. Formation of datasets

Due to the inclusion of different approaches in the study, new variables need to be generated based on the search query data. Later in this study, the search query data generated using the approach by Da et al. (2011) is referred to as DATASET 1. Similarly, the search query data generated using the approach by Preis et al. (2013) is referred to as DATASET 2. The above figure 4 concludes how these different SVI variables are formed in this study. Inclusion of multiple perspectives creates ambiguity in the study, but provides valuable insight of the features of search query data. Later in this study, the U.S. and global versions of different variables are referred to by the abbreviations US and GLO. This means that the U.S. data of the first set of search words containing ETF tickers is referred to as DATASET 1 US. Other three variables are DATASET 1 GLO, DATASET 2 US and DATASET 2 GLO.

3.2. Financial data

The first data source is associated with financial data and is obtained from Thomson Reuters DataStream. The financial data retrieved contains the price development and other certain key figures of underlying assets and benchmark index. All the used variables concerning financial data are listed to the table 4 below.

Table 4. Variable definitions of financial data

Variable	***Definition***
Variables calculated from DataStream data	
Ret (R)	*Weekly return*
Abnormal ret (AR)	*The weekly excess return. Is the part of the return that exceeds the expected return. (Return – (Index return * Company Beta))*
Trade volume (TV)	*Weekly trading volume measured in shares traded per day*
Shares outstanding (SO)	*Weekly amount of shares outstanding*
Turnover (T)	*Weekly turnover (Trade volume/Shares outstanding)*
Abnormal Trade Volume (AT)	*Weekly excess volume. It is the part of trading volume that exceeds the 8-week median.*
Market cap (MC)	*Weekly market capitalization*

As can be seen from the table 4 above, financial data is reviewed on a weekly basis. This is because Google Trends provides only weekly volume changes on a long-term. Thus, the actual data exported from DataStream includes the weekly observations of all underlying assets including the benchmark index. Other variables needed for the calculations are trading volume, market cap and shares outstanding.

The data process revealed a clear difference in the availability of data for commodity ETFs and ordinary stocks. It was noticed by comparing the data availability for both asset types in the DataStream database. This is because ETFs are a much newer form of investment compared to traditional equities. Therefore, more information has been collected on equities over a longer period of time. For the required data listed in table 4, historical values of outstanding shares and market cap were not found for ETFs through the DataStream database.

To obtain the missing data values, a few assumptions must be made. DataStream defines the formula of market value for equity indices as follows:

$$MV_t = \sum_{1}^{n} (P_t \times N_t) \tag{1}$$

Where, number of shares *N* in issue on a day *t,* is multiplied to the price of that share *P* on the same day. *n* represents the number of constituents in the index (Thomson Reuters, 2017). This study assumes that the selected ETFs are seen as a single individual entities. Thus, the market value is calculated in the same way as the market value of an individual stock. This is determined to be the best available option and not far from true values as most ETFs only contain commodities. It is also assumed that shares issued correspond to shares outstanding for the selected ETFs. The total number of shares issued is divided by a thousand (1000) to present the number in thousands. This eliminates the need for separate rounding action. A critical factor is that the variables of each ETF are processed in the same way. To measure the accuracy of the calculated shares issued variable, they were compared to the shares outstanding values found for a few ETFs. As a result, it can

be stated that for these few ETFs the shares issued variable is an accurate indicator of the actual shares outstanding.

3.3. Search query data

The second source of data is Google Trends data, which consists of SVI data for various search terms. This data can be obtained by downloading CSV files straight from the Google Trends website. However, this must be done manually for each keyword.

Table 5. Variable definitions of search query data

Variable	***Definition***
Variables from Google Trends	
$SVI\ DATASET\ 1_{US}$	*Google Trends search frequency (SVI) data on a weekly level for DATASET 1 in US.*
$SVI\ DATASET\ 1_{GLO}$	*Google Trends search frequency (SVI) data on a weekly level for DATASET 1 GLO.*
$SVI\ DATASET\ 2_{US}$	*Google Trends search frequency (SVI) data on a weekly level for DATASET 2 in US.*
$SVI\ DATASET\ 2_{GLO}$	*Google Trends search frequency (SVI) data on a weekly level for DATASET 2 GLO.*
ASVI	*The log SVI of the current week minus the log median of previous 8 weeks (Da et al., 2011)*

As shown in table 5, SVI data is retrieved for all approaches defined in the sample formation section in figure 4. ASVI represents abnormal search volume values calculated using the original retrieved SVI values. Abnormal SVI values are calculated for all four variables separately. The calculation method follows the method used in the reference study by Da et al. (2011) and is described in the methodology section.

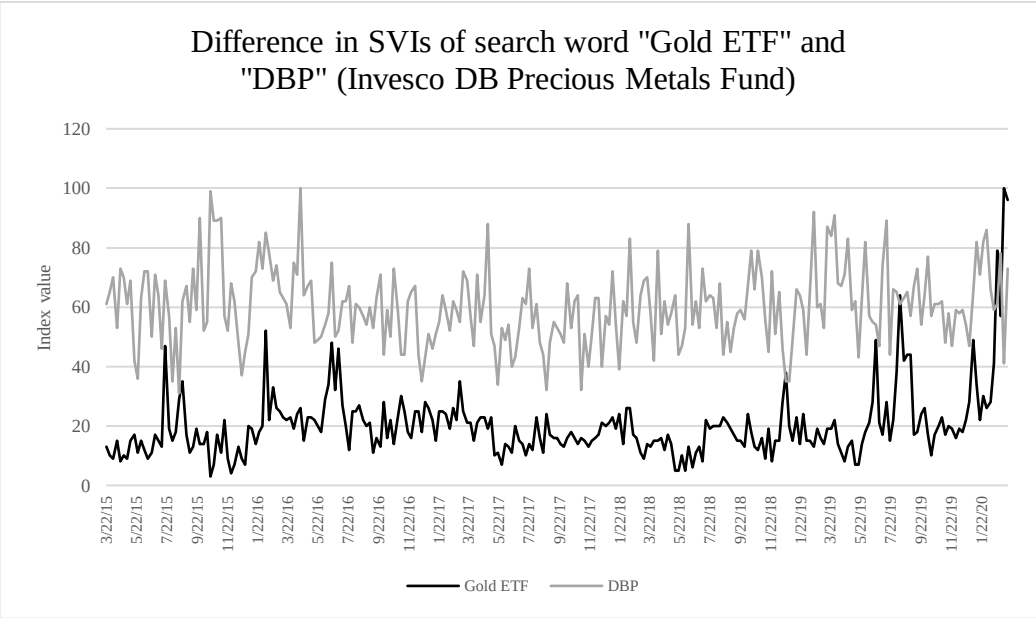
For time periods longer than nine months and shorter than five years Google Trends provides only weekly observations of search term volumes. The search query data is collected for five year time period from March 2015 to March 2020. Google defines week

to start on Monday and end on Sunday. The weekly frequency of Google Trends data, sets the base for the frequency used in this research. The search volume is provided in an index format giving values from 0 to 100. Thus, the data represents the relative change in search volumes for a certain time period, and not actual numerical values of search volumes. Another aspect that needs to be taken into account is that Google only publishes SVI information, if the search word produces sufficient amount of search queries. This is based on the privacy and anonymity requirements of Google.

After checking the availability of financial data for the chosen ETFs, the data collection and sampling process is done regarding SVI data. During the data collection process, the possible noise is checked for each search word individually. After this the decisions was made to either accept or omit the search word from the sample. Both approaches (DATASET 1 & DATASET 2) battle with similar challenges, which are related to the noise and consistency of the data. This has also been a major issue in prior researches, in which a large proportion of the original sample is omitted from the study because of noise and insufficient data. In the research by Da et al. (2011) nearly half of the Russell 3000 stocks ended up being excluded due to the insufficient data.

Finding and omitting noisy tickers and keywords is a critical step in this research. In most cases noise indicates to the existence of alternative meanings for tickers and words. The noise can be dominant or weak, depending on the search activity of the double meaning and the actual search target. The challenge associated with the use of ETF tickers in DATASET 1 is that because the tickers form a somewhat random combination of letters, the likelihood of potential noise and dominant double meanings increase. The variable DATASET 2 probably also contains noise, but due to the nature of the keywords, as they are more common and established, the probability of possible dominant double meanings is lower. The use of more common keywords in DATASET 2 also increases the number of searches on a general level, resulting in more consistent information. To realize this with an example. It is more likely that search word “Gold ETF” represents the investment activity of precious metal ETFs and does not have dominant double meanings than the search word “RING” (iShares MSCI Global Gold Miners ETF) or “DBP” (Invesco DB Precious Metals Fund), which do not resemble financial markets at all.

Same example regarding search words “Gold ETF” and “DBP” is viewed graphically in graph 3 below. “DBP” represent an ETFs called Invesco DB Precious Metals Fund. At first glance, the letter combination DBP seems random, but after a closer analysis of the search volumes, two dominant double meanings can be observed. DBP refers to Development Bank of Philippines (www.dbp.ph) and to a chemical abbreviation called Dibutyl Phthalate. As a result, the search term “DBP” is seen to have a lot of noise and is therefore excluded from the study. This is also reflected in the graph 3 below as a very random movement of the keyword "DBP" compared to the movement of the keyword "Gold price", which is seen to contain significantly less noise.



Graph 3. Difference in SVIs of search word “Gold ETF” and “DBP” (Invesco DB Precious Metals Fund) (Google Trends, 2020)

The process of omitting possible noisy tickers and search words from the sample is related to the use of features provided by Google Trends. The process was performed while retrieving the data for each search term. The features provided by Google Trends for analyzing search volumes are “Top searches” and “Rising searches”. Both of these categories are self-explanatory and associated with the search. ”Top searches” lists the most common word combinations associated with the searched term. Whereas, “Rising searches” highlight the terms related to the topic with significant growth in the given time period. Top searches uses a scale similar to SVI, which gives values from 0 to 100. In

rising searches the values are given as percentages. Uniquely, the term is assigned a value "Breakout", if the growth surpasses 5000 % over the used time period.

Just by looking at SVI values, it is not always self-evident that a keyword has a secondary meaning or meanings. Together these tools "Top searches" and "Rising searches" give an idea of the context associated with the search term. From which, the user can try to interpret if the search term has possible alternative meanings or noise in general. What comes to rising searches, it provides a good way to detect trends and spikes in search volumes and, most importantly, to explain movements and possible outliers in the SVI data for the used time period. To conclude both features, they offer useful insight from the data, which might otherwise go unnoticed.

To make sure to treat each search term in a similar manner, guidelines of data processing actions need to be clarified. In this research the search term is left out if it has a double meaning resulting a score of 100, in the top searches category. If this is not the case, the search term is included to the study. This is a simple and effective procedure also used by other researches as Wuoristo (2012). However, this does not reduce all the noise in the data. Below in the table 6 are two examples search words "PALL" and "GLL" that are omitted from the study by observing both top searches and rising searches.

Table 6. Examples of noisy tickers

Top searches for PALL in US	
Pall Mall	*100*
Cigarettes	*22*
Pall Mall cigarettes	*21*
Pall corporation	*21*
Pall filter	*13*
Pay pall	*11*
Rising searches for PALL in US	
Andrew Taggart	*Breakout*
The Chainsmokers	*Breakout*
Pall Danaher	*Breakout*
Drew Taggart	*Breakout*
Jake Pall	*Breakout*

Top searches for GLL in US	
Bosch	*100*
Bosch laser	*39*
Laser level	*25*
Bosch laser level	*24*
Bosch GLL 30	*15*
Bosch GLL 2	*11*
Rising searches for GLL in US	
Bosch laser level	*Breakout*
Bosch GLL 30	*Breakout*
Bosch GLL 2	*Breakout*
GLL 3-15	*Breakout*
Bosch GLL 3-80	*Breakout*

In table 6 “PALL” clearly indicates to the cigarette brand and not to Aberdeen Standard Physical Palladium Shares ETF. Also search results of “GLL” clearly indicate the volume of a construction equipment by Bosch rather than ProShares UltraShort Gold ETF. On the other hand, in a positive scenario the associated words in top searches and rising searches give a strong indication that the search query is related to the search term. An example of this is shown in table 7 for the keyword "GLD".

Table 7. An example of a positive search scenario

Top searches for GLD in US	
GLD stock	*100*
Gold	*53*
Shop GLD	*50*
GLD price	*39*
Stock price GLD	*27*
the GLD	*25*
GLD ETF	*21*
SLV	*21*

Search word “GLD” represents gold ETF SPDR Gold Shares. It can be seen in table 7 that all top searches are related to finance and to the particular ETF. A value of 100 is given for a search “GLD stock”, which indicates that no dominant double meaning exists and that majority of search activity indeed is related to the SPDR Gold Shares. Thus GLD is included to the study.

The financial data section of this study checked the adequacy of financial information. The decision was made to exclude ETFs with an inception date after 2015. With regard to search query data, the adequacy of the data must also be checked. An aspect that needs to be taken into consideration are the possible missing search queries, which applies especially to less popular keywords. In other words, if Google Trends do not produce SVI data, it means there is not a sufficient amount of search queries for this particular search term. The requirement concerning sufficient data in this research is determined to be one year of sequential data. The data filtering process was explained separately for both

variables DATASET 1 and DATASET 2 due to the different nature of the keywords included.

First, the search query data for DATASET 1 was processed. Of the 22 precious metals ETFs identified in the financial information section of this study, the ticker for each ETF was examined. First, the obvious noisy tickers were removed. Of these 22 ETF tickers, nine were omitted, which means that 13 ETF tickers remained in the sample. The noise of the data was checked by examining both US and global data. Next, it was checked whether the requirements for sufficient information were met. From remaining 13 ETF tickers only five tickers had the required one-year sequential data.

The final set of chosen ETF tickers is gathered into the table 8 below. It should be noted that the size of the ETF is directly related to finding sufficient information. The smaller the ETF in terms of its assets under management (AUM), the less likely it is to meet adequate information requirements. The five ETFs selected for the final sample are also the five largest ETFs from the original sample of 22 ETFs in terms of net assets. Previous studies have addressed similar issues, with the result that a large proportion of original companies have been excluded from the study (Da et al., 2011). This shows that while the adequacy of Google Trends data has gradually improved, it is still the biggest limiting factor in the study.

Table 8. Final ETF tickers (DATASET 1)

Ticker	***Name***
GLD	*SPDR Gold Shares*
IAU	*iShares Gold Trust*
SLV	*iShares Silver Trust*
GDX	*VanEck Vectors Gold Miners ETF*
GDXJ	*VanEck Vectors Junior Gold Miners ETF*

A similar process was performed for DATASET 2. First, of the original 20 keywords contained in the variable, three obvious noisy search terms were omitted. The three noisy search words were “gold”, “silver” and “safe haven”. From the remaining set of words

the sufficiency of data were checked for each search term. With this phase, the keywords "silver ETF", "NYSE Arca" and "VanEck ETF" were rejected. Decisions on the DATASET 2 filtering process were made based on both U.S. and global data. The final set of selected keywords for the variable DATASET 2 is shown in table 9 below.

Table 9. Final generic search words (DATASET 2)

Search term	
Gold price	*Gold ETF*
Precious metals	*Silver price*
ETF	*iShares ETF*
SPDR ETF	*Vanguard ETF*
Financial markets	*Stock market*
VIX	*Commodities*
NYSE	*S&P 500*

In this work, the perspective by Da et al. (2011) is a limiting factor in terms of the number of underlying assets included in the final sample (only 5 ETFs). This is because if the ETF ticker in question, and thus the attention representing the ETF, is not included, then the ETF is not included either. There is no similar issue with the approach of Preis et al. (2013), because if a keyword is excluded from the study, it does not affect the number of ETFs included to the sample in any way. This is because the search words of DATASET 2 do not represent the attention of a specific ETF but seek to represent the attention of the whole market. To maintain a similar starting point for these perspectives and to allow comparability, the ETFs presented in table 8 provide the final sample for both SVI approaches.

4. Methodology

This chapter describes the methodology of this master's thesis. The research itself is conducted by using both Microsoft Excel and coding platform RStudio. This study follows the methodology of researches by Da et al. (2011) and Preis et al. (2013). Parts of both studies are replicated with certain changes in the form of additions and omissions, to make the methodology suitable for this research. This section consists of three subsections. The first part describes the calculation of the required parameters and the process of forming a cross-correlation matrix. The second part presents the regression analysis. Finally, the third and last part describes the moving average calculation used in the study. The moving average is used to implement an investment strategy for the best-performing search terms throughout the review period.

4.1. Different parameters & Cross-correlation matrix

After collecting and processing the data, there are variables that need to be calculated from the collected data. These by nature abnormal values are: abnormal returns, abnormal trading volume and abnormal search volume.

As the data is retrieved on a weekly basis the variables are also calculated on a weekly basis. To enable the calculations of abnormal returns we need to solve the expected returns (*E(R)*). This can be calculated with the equation 2:

$$E(R_{it}) = R_{mt}\beta_{it} \tag{2}$$

Where, R_{mt} represents the returns of the benchmark index *m* in time *t*, and β_{it} represents the beta specific to the ETF *i* in time *t*. For the purposes of this study, the formula means that the weekly returns of the benchmark index MSCI World Commodity Producers Index are multiplied by ETF-specific beta. Beta can be calculated in different ways, the calculation methods and results differ between the different options. The betas for this study have been calculated as moving betas using weekly data for the last two years. This means that the beta changes each week based on the previous 104 weekly observations. Continuous betas are used to ensure beta eligibility for each ETF at different times. After

solving the expected returns (*E(R))*, it can be applied to calculate abnormal returns using the following equation 3:

$$AR_{it} = R_{it} - E(R_{it}) \tag{3}$$

Where, R_{it} is the realized return for ETF *i* at time *t*. Whereas, $E(R_{it})$ is the expected return for time *t*. In this study, abnormal return (AR) is calculated differently than in the reference study by Da et al. (2011). In the reference study, abnormal returns were calculated using characteristic-adjusted returns based on the research by Daniel, Grinblatt, Titman and Wermers (1997). The method used by Da et al. (2011) is especially designed to normalize returns in situation, where the sample size is large and company size, industry and momentum related factors play a major role.

Next is calculated the weekly abnormal volume. This can be calculated by measuring the changes in turnover of the underlying asset. To enable the calculation, we must first calculate the turnover (*T*), which is shown in equation 4 as the ratio of trading volume (*TV*) and shares outstanding (*OS*).

$$T_{it} = \left(\frac{TV_{it}}{OS_{it}}\right) \tag{4}$$

The idea behind the use of outstanding shares is to normalize the trading volumes. This is done because trading volumes differ between ETFs of different sizes. The basic assumption is that the larger the number of shares, the higher the trading volume. Trading volumes used in this research are weekly averages. The abnormal volume calculation is applied from the research of Da et al. (2011). A similar method is also used to calculate the abnormal SVI values in equation 6. The aim of both variables is to find the so called peaks from the time series data under analysis. The equation of abnormal volume (AV) is seen below:

$$AV_{it} = \log T_{it} - \log Med(T_{it} - 1, \dots, T_{it} - 8) \tag{5}$$

Equation 5 measures peaks by comparing the current week turnover to the median of the last eight weeks. Thus, T_{it} represents current week turnover of ETF *i* at time *t*. Whereas, *Med* represents the median turnover of the same ETF for the last eight weeks. Logarithmic values are used in equations 5 and 6 as in the reference study. Abnormal search volume is referred to as *ASVI*. This new variable is created to find attention peaks from search volume data (*SVI*). Variable is formed according to following equation 6:

$$ASVI_{it} = \log SVI_{it} - \log Med(SVI_{it} - 1, \dots, SVI_{it} - 8) \tag{6}$$

The basic logic of equation 6 is similar to equation 5, but turnover is replaced by search volumes (SVI).

4.2. Regression methods & hypotheses testing

To simplify how this research is conducted. The following table 10 summarizes the methods used to test each of the hypotheses of this study.

Table 10. Methods used to test hypotheses

H1	*Increase in SVI affect changes in the trading volume of the underlying assets.*
	↓ *Simple OLS regression*
H2	*Increase in SVI has a negative effect on the price of the security (ETF).*
H3	*The SVI data from US acts as a better predictor than the global SVI data for commodity ETF price changes.*
H4	*ETF ticker data (DATASET 1) acts as a better predictor than the generic search data (DATASET 2) for commodity ETF price changes.*
	↓ *Panel data regression*
H5	*There are differences in the effects of individual keywords on price changes*
	↓ *Simple OLS regression, Robustness checks*

H6	*Excess return can be earned with a simple moving average trading strategy that takes advantage of SVI values.*
	↓ *Moving average calculation*

The first hypotheses (H1) tests the relationship between the trade volume of the underlying asset and SVI of the related search term. Simple OLS regression is used to describe this relation and can be calculated with equation 7.

$$AV_i = \beta_0 + \beta_{1i}ASVI_i + \beta_{2i}AR_i + \beta_{3i}MV_i + \varepsilon_i \tag{7}$$

Where AV_i represents the abnormal trading volumes, $ASVI_i$ abnormal search volumes, AR_i calculated abnormal returns and MV_i the logarithmic market values. In OLS regression formula, abnormal trading volume (*AV*) is the dependent variable and abnormal search volume (*ASVI*), abnormal returns (*AR*) and log market value (*MV*) are independent variables. The last term of the formula ε_i represents the error term of the regression. In OLS regression, the averages of all ETFs for these variables are used to generate a single general-level regression. Regression is performed for two time periods, *t* and *t+1*. Where, *t* measures the effects on same-week basis, while *t+1* measures the possible effects on next week's trading volume. Regarding the first hypothesis (H1) the focus in this equation is in coefficient β_1 for *ASVI*.

The method used for hypotheses two, three and four (H2, H3 & H4) is panel data regression. This is not the same method used in the reference study by Da et al. (2011), in which similar hypotheses were examined using cross-sectional Fama-Macbeth (1973) regression. However, the models are seen to be quite similar in terms of the purpose of the models and the results they produce. In this study, panel data regression was found to suit the used data better and to be easier to understand and conduct.

Panel data regression analyzes the possible effects of abnormal search volumes (ASVI) on abnormal returns (AR) on the underlying asset. Panel data regression is characterized by measuring both firm-fixed (N) and time-fixed (t) effects simultaneously. The model is

seen to be useful, since it gives direct estimates of the marginal effect of the used independent variables. Marginal effects describe how the outcome (dependent variable) changes when independent variable in this case abnormal search volume (ASVI) changes.

Panel data regression has different independent approaches. The two main approaches that this study focuses on are random effects models and fixed effects models. These are the most common models to be tested within panel data regression. From a theoretical point of view, in econometrics, fixed effect model refers to regression, where fixed group means (non-random) are used, whereas in random effect model the group means are a random sample of a population. In practice, the random effect model makes additional assumptions that the fixed effect model does not. If these assumptions are wrong, the random effects model is biased, which is unwanted. However, if these assumptions are correct, the random effects model is more efficient, producing smaller standard errors. The Hausman test is used in this study to compare the results of fixed and random effects to determine which model is used. The Hausman test examines the consistency of estimates (random effects) when compared to a less efficient estimator (fixed effects) which is known to be consistent. If the Hausman test results show differences between fixed effects and random effect models, it means that one of the models (random effects) is inconsistent. When there are differences in results, the test suggests the use of a fixed effect model. However, if the results are the same, it is possible to use either model. (Torres-Reyna, 2007; Bell & Jones, 2015)

Panel data regressions are conducted for five different time periods. The first four weeks make up the first four time periods. Where, the first week measures the effects on same-week basis, the second week the future effects on week t+1, and so on. The duration of each of the first four periods is one week. The fifth time period is formed to present long-term effects and lasts the rest of the one year period, covering weeks 4 to 52. The dependent variable in the panel data regression is abnormal returns. The effects on future abnormal returns in the short term are calculated by using a lead-lag of abnormal returns. Whereas, long-term effects are calculated using a regression in which the future cumulative abnormal returns of weeks 4-52 are used as a dependent variable. The following equation 8 illustrates how the panel data regression is generated:

$$AR_{it} = \beta_0 + \beta_{1i}ASVI_i + \beta_{2i}AV_i + \beta_{3i}MV_i + \varepsilon_i \quad (8)$$

Like earlier presented above in this methodology section, AR_{it} refers to abnormal returns of the underlying asset *i* in a week *t*. The explanatory variables in the panel data regression in equation 8 are $ASVI_i$, which represents abnormal search volumes, AV_i, which refers to abnormal volumes and MV_i, which refers to logarithmic market value.

Regarding hypothesis testing, to accept the second hypothesis (H2) we are looking statistically significant negative values of coefficient β_1 for *ASVI.* Due to how different SVI approaches are formed in this study, the overall effect of the variable ASVI is measured in the form of four different variables. This means that ASVI of all the different SVI variables: DATASET 1 US, DATASET 1 GLO, DATASET 2 US and DATASET 2 GLO are included in the final regression formula. The effects of all four variables combined form the picture of the overall effects of ASVI. This complicates the analysis of general-level results, but allows hypothesis three and four to be considered. Acceptance of hypothesis 3 requires that variables using US data produce more significant results than variables using global data. Similarly, acceptance of hypothesis 4 requires that DATASET 1 produce more promising results than DATASET 2.

For the fifth hypothesis (H5), the review moves to the level of individual keywords and ETFs. This is done by performing a simple OLS regression on each of the five ETFs. As the relationship measured is similar to what panel data regression tests, these additional regressions also serves as robustness checks. The dependent variable is abnormal returns, but individual ETF values are used instead of averages. Abnormal search volumes for individual search words are used as explanatory variables. Of DATASET 1, only a ticker that measures the attention of the ETF in question is selected as an explanatory variable. For DATASET 2, all search words are selected because generic search words do not represent the attention of any particular ETF. Regressions are performed for two time periods, *t* and *t+1*. Where, *t* measures the effects on same-week basis, while *t+1* measures the possible effects on next week's abnormal returns. The regression for each ETF is formed based on the equation 9.

$$AR_i = \beta_0 + \beta_1 ASVI_1 + \beta_2 ASVI_2 + \cdots + \beta_n ASVI_n + \varepsilon_i \quad (9)$$

Where, AR_i represents abnormal volume of ETF i. $ASVI_n$ represent abnormal search volumes of all different individual search words and *n* represent the number of search words. For the purposes of this study, the effects of 15 different search words are measured for each ETF (DATASET 1 = 1 keyword, DATASET 2 = 14 keywords). In hypothesis testing, hypothesis five (H5) is accepted if differences are observed between the effects of individual keywords. Therefore, the focus is on the statistical significance and the relationship between the statistically significant coefficients. Table 11 concludes all generated regressions with respect to the corresponding hypotheses.

Table 11. Regression testing with respect to the corresponding hypotheses

Hypothesis:	***Dependent variable:***	***Explanatory variables:***	***Sought relationship:***
H1:	*Abnormal volume (AV)*	*ASVI for all 4 variables (DATASET 1 US, DATASET 1 GLO, DATASET 2 US and DATASET 2 GLO)*	*Statistical significance*
H2:	*Abnormal returns (AR)*	*ASVI for all 4 variables (DATASET 1 US, DATASET 1 GLO, DATASET 2 US and DATASET 2 GLO)*	*Statistical significance, ASVI coefficients < 0*
H3:	*Abnormal returns (AR)*	*ASVI for all 4 variables (DATASET 1 US, DATASET 1 GLO, DATASET 2 US and DATASET 2 GLO)*	*Statistical significance, ASVI (US) > ASVI (GLO)*
H4:	*Abnormal returns (AR)*	*ASVI for all 4 variables (DATASET 1 US, DATASET 1 GLO, DATASET 2 US and DATASET 2 GLO)*	*Statistical significance, ASVI (DATASET 1) > ASVI (DATASET 2)*
H5:	*Abnormal returns (AR)*	*$ASVI_n$ for all individual search words*	*Statistical significance*

4.3. Moving average calculation

For the sixth and final hypotheses (H6), methodology of the reference study by Preis et al. (2013) is used as a reference. The basic idea of the sixth hypothesis is to measure the performance of market timing strategy using Google Trends search volume data. In other

words, how well a strategy that uses search volumes can perform when applied to historical time-series data.

The existing literature contains different views on how market sentiment transitions to market price. Preis et al. (2013) show that drops in the financial markets are preceded by investor concern. This means, the attention increases prior to market drop. The study suggest that, the aforementioned logic, can be used to construct a profitable trading strategy for query volumes of certain search terms. To determine investments decisions based on the logic by Preis et al. (2013), it means when attention increases, the underlying asset is sold short, and other way around, when attention decreases the asset is bought. Another view in existing literature is based on attention theories by Barber and Odean (2008). This aspect battles against the view of Preis et al. (2013), as rise in attention is expected to have a positive effect on underlying asset. The theory is based on assumption that retail investors own a limited number of shares, thus it is more likely that their intentions are towards buying rather than selling. The collective attention of retail investors causes momentary price pressure as retail demand for the underlying asset increases. To implement the strategy based on attention theories, the underlying asset is bought when attention increases and sold short when it decreases. Klemola (2019), on the other hand, noticed that when predicting stock market performance based on attention shifts, downward market predictions were more apparent than upward prediction. According to Klemola (2019), this was because people have a stronger reaction to negative news and outlook than positive. In most cases, interest in the search word predicts a decline in the underlying asset. Findings of Klemola (2019) are in line with the findings of the reference study by Preis et al. (2013).

The initial assumption regarding the trading rules in this study were made based on reference study by Preis et al. (2013). This means that if the attention increases, the underlying asset is sold short and if the attention decreases, the underlying asset is bought. It is also interesting to see if the regressions of the previous steps show anything about the direction of the potential effects. When dealing with weekly data the actions take place once a week on the first trading day of the week. The positions are always one week (5 trading days) long, after which a new assessment of the direction of the strategy will be made. Google defines a week to start on Monday and to end on Sunday. This means

that positions are taken on Monday of each week and weekly returns are based on the closing price of each Friday.

Preis et al. (2013) base their trading decisions simply to the relative change in search term volume. They compare the calculated moving average to the current search volume at time *t*, to see if it increases or decreased, to know, whether to sell short or buy the underlying asset. The generated indicator represents abnormal search volume. The threshold of abnormal search volume for deciding whether to buy or sell an underlying is zero. Preis et al. (2013) showed the best results using three-week moving averages. Therefore, similar time values are also used in this study.

Another important detail that needs to be determined before implementing a strategy is for which underlying assets the strategy is formed. In the reference study by Preis et al. (2013) the performance of words is compared to only a single asset. This means the strategy is implemented using only one underlying asset. Investing to multiple assets is possible by conducting an equally weighted or value-weighted portfolio of ETFs. However, it would not be optimal in real life as transaction costs, increase with each ETF. It is also not beneficial to select multiple so closely related assets. Therefore, normally one would choose only the best option or an index. In order to follow the methodology of the reference study by Preis et al. (2013) and to simplify the calculations, the decision was made to invest only in a single asset.

To enable the calculations of abnormal search volume factor, first, the moving averages needs to be calculated. The moving averages are set to be the average of three weeks search volumes. This is calculated according to the equation 10.

$$MA_{t-1} = (SVI_{t-1} + SVI_{t-2} + SVI_{t-3})/3 \tag{10}$$

Where, *SVI* represents the search volume for search term during different weeks *t*. In this case, *t-1* is the search volume during last available week, *t-2* the volume of the second-last available week and *t-3* the volume of the third-last available week. The calculated moving averages represent a short-term trend to which the actual number of searches each week is compared. The abnormal search volume factor used in strategy implementation

differs from the ASVI variable used in the regression calculations. Thus, also the name differs. The abnormal search volume used in moving average calculations is formed according to equation 11.

$$ASV_t = SVI_t - MA_{t-1} \tag{11}$$

Similar to the equation 10, *SVI* represents the search volume for search term during different weeks *t*. MA represent the above calculated moving average. To interpret the formula, the average search volume of preceding 3 weeks is deducted from the search volume of the last available week.

The measures taken in relation to the strategy for week *t* are based on the calculated abnormal search volume (*ASV*) from the previous week *t-1*. As mentioned above, the holding period of the position is one week (5 trading days), after which after which the position is sold at Friday's closing price. Trading rules are concluded in the equation 12.

$$Trading\ rules = \begin{cases} if\ ASV_{t-1} > 0, & sell\ short\ at\ time\ t \\ if\ ASV_{t-1} < 0, & buy\ long\ at\ time\ t \end{cases} \tag{12}$$

The equation 12 indicates that when the abnormal search volume increases, the underlying asset is sold short, and vice versa, when the abnormal search volume decreases, the underlying asset is bought.

When the conditions stay the same, meaning abnormal search volume is positive or negative several continues weeks, it is intended that the strategy holds the current position. It is done by selling or buying again the same asset. In real life situation this would not be ideal because of the increasing transaction costs which should be avoided. In this research, the transactions costs are included to the calculation as close-to-reality estimates. Regardless of whether the strategy shows continuous holding periods, a new position is taken each week. Therefore, transaction costs are calculated for each week.

The total cost of an ETF consists of two parts, which are holding costs and transaction costs. The holding costs consists of expense ratio and transaction costs of bid-ask spread.

The implementation process of transaction costs plays an important role in preventing biased estimates. Both the holding cost and transaction cost are calculated for each week. Implementing transaction costs is simple, as they are deducted each week when a trade is made. The size of the transaction costs are determined based on average spread of last 60 days. Holding costs, on the other hand, are expressed as annual costs, which makes it slightly more challenging to allocate them on a weekly basis. Regarding the expense ratio, it is assumed that certain types of ETF positions are held throughout the whole investment period. Thus, the expense ratio is calculated for the whole period (almost five years) and divided between each week of the reference period. Eventually, costs are deducted from the weekly performance of the strategy.

The strategies implemented in this research are chosen based on the results of earlier stages introduced above. Thus, not all search terms and assets are measured. To enable the measurement of the profits of different strategies, each portfolio is set to have an arbitrary value of one at the beginning. Then, the performance of the strategy is calculated for each time step, from which the cumulative returns are calculated. A performance curve for the whole time period is a good way to display and compare results of different strategies. Cumulative return is calculated according to equation 13.

$$Cumulative\ R_t = (1 + R_1) \times (1 + R_2) \times \ldots \times (1 + R_t) - 1 \qquad (13)$$

Where, R represents the returns of the strategy during different weeks t. Strategies are measured for the time period from April 2015 to March 2020. The time period includes a total of 258 weekly time steps. Regarding hypothesis testing, the sixth hypothesis is accepted if the formed strategies are able to earn excess returns during the period under review. Market performance is measured by a buy and hold strategy that invests in an iShares S&P 500 ETF (IVV).

5. Results

This section presents the results of this study. The presentation of the results begins with the general assumptions that emerged during the study, from which we move on to the presentation of the results related to the research hypotheses.

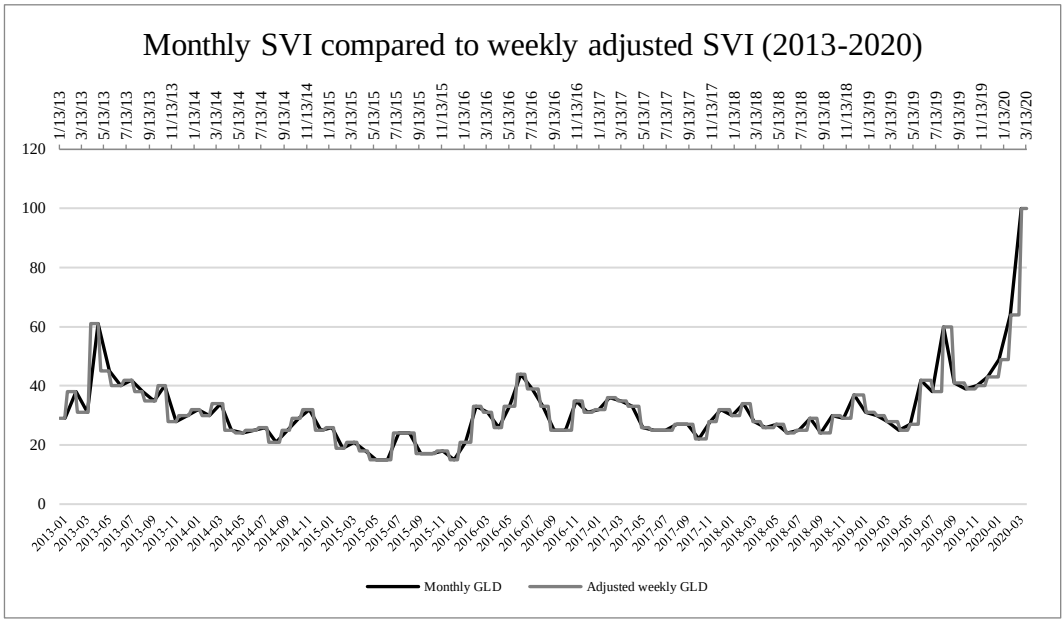
5.1. General findings

At this point, the reader probably has an idea of the data, methodology and objectives of this research. The purpose of this section is to present the findings that have emerged as the study progresses up to this point. First, the findings in this section focus on the limitations found and then describe the used sample in the form of a graphical representation.

The sample formation and data processing were opened in detail in the third section of the study. At this stage, it has become clear that the major limiting factors in this study are data related factors. The biggest factor is the way Google Trends works. More specifically, how Google displays information and allows access to their data. For an unknown reason, Google is restricting access to its data. A few years back it was still possible to get weekly data for a desired period. Now, weekly data is only available for periods of less than five years. This is a clear step backwards. At the same time, Google granted access to daily data for a period of nine months instead of the previous 90 days, which, on the other hand, is a step in the right direction. Restricting data would not be a problem if Google did not normalize data based on the time period retrieved. Normalization helps to compare keywords with each other over a selected time period, but makes it difficult to use Google Trends data in studies with a longer time period. This is because normalization does not allow access to actual search volumes, making it impossible to combine and scale time series accurately.

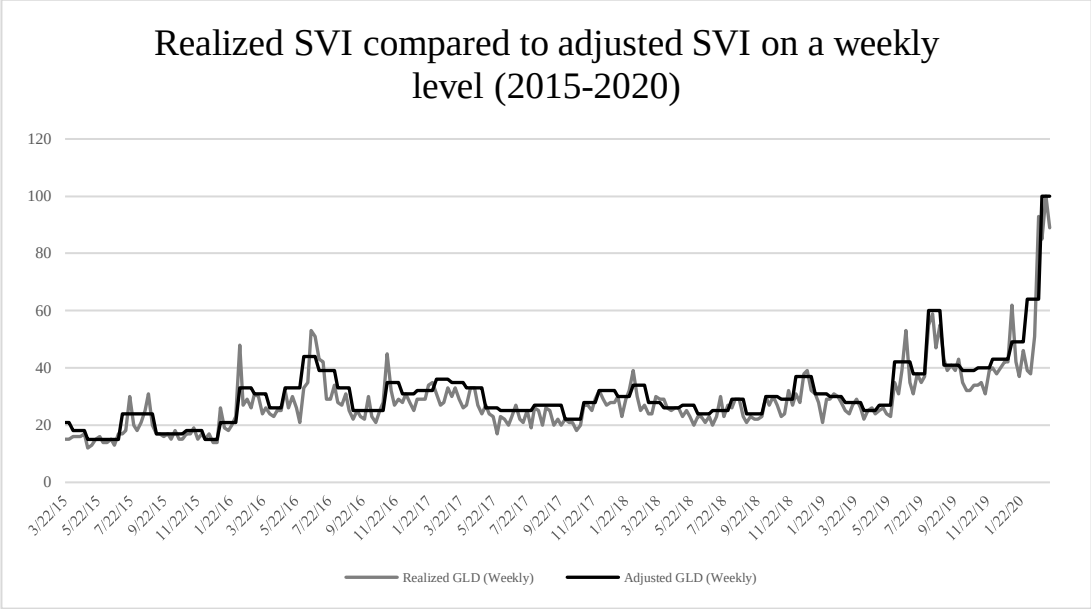
The two graphs 4 and 5 are shown below to see examples of combining two Google Trends time series. The examples are conducted using ETF ticker “GLD” that refers to SPDR Gold Shares ETF. Adjusted values can be calculated using monthly values as a reference point. Monthly values are retrieved for the entire period to give an idea of the

overall scale. So called adjustment factor is used to rescale two separate weekly time series into one larger series. The adjustment factor is calculated by dividing the value of each week in a given month by its monthly value. The initial weekly values are then multiplied by the calculated weekly adjustment factors.



Graph 4. Monthly SVI of search word GLD compared to weekly adjusted SVI of search word GLD

Based on the graph 4, it can be stated that combining two weekly SVI datasets is possible. Adjusted weekly values follows the monthly graph nicely. However, the graph 4 above does not express the whole truth, as the original monthly values represent the average of the weekly values. This means monthly values and original weekly values can differ significantly. Therefore, the adjusted weekly search volumes should be compared with the actual realized weekly values, as shown in the graph 5 below.

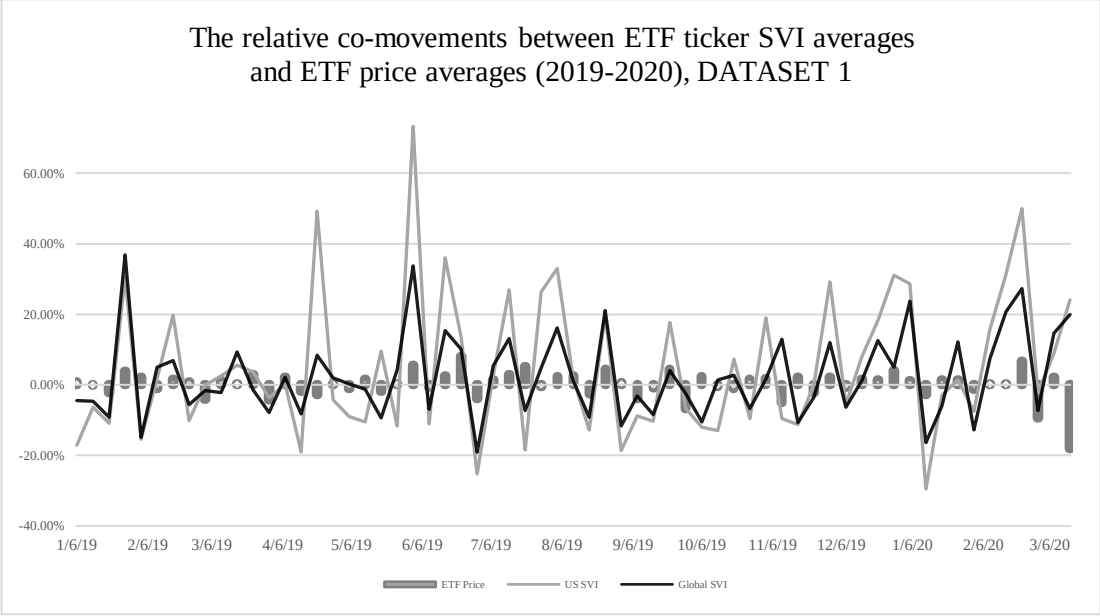


Graph 5. Realized SVI of search word GLD compared to weekly adjusted SVI of search word GLD

From graph 5, it is possible to see that the adjusted weekly values do not reflect the weekly changes in SVI, but show a flattened curve compared to the original realized weekly SVI. After careful analysis, the decision was made to use only five years of SVI data. Factors influencing the decision were the deterioration in the accuracy of the SVI data and the large amount of work required. Using a five-year period, the data can be accessed directly without the need for separate processing, which should be done for dozens of keywords. From the original research plan, this shortens the time period by eight weeks to enable the calculation of SVI-related variables.

To get a preliminary insight of the relationship between SVI and ETF, a graphical overview is performed for both SVI approaches represented by variables DATASET 1 and DATASET 2. These are performed using simple averages of the variables. Graph 6 shows the relationship between DATASET 1 and ETF returns, whereas graph 7 shows the relationship between DATASET 2 and ETF returns. The following graphs 6 and 7 also provide information about the relationship between the US and global data regarding both variables. As a reminder, “DATASET 1” refers to SVI data that contains chosen ETF tickers and “DATASET 2” refers to SVI data for chosen generic keywords. At first glance, there is no strong link between realized ETF returns and SVI values. In fact, the co-movements appear to be quite random and do not show any patterns recognizable to the human eye. Both variables show strong search spikes, but there does not appear to be

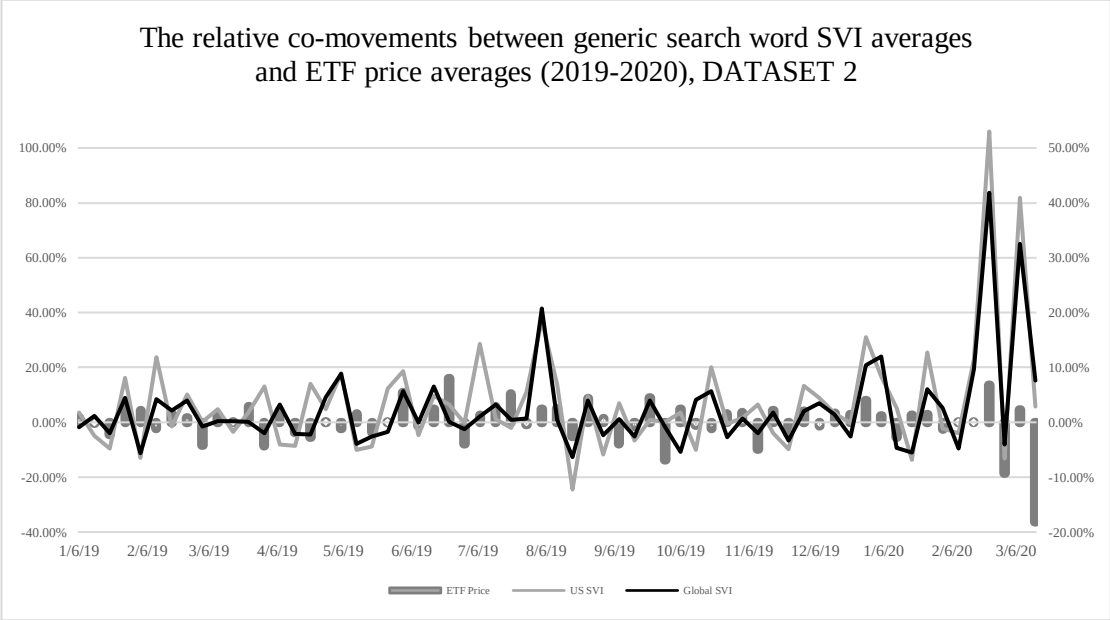
a corresponding ETF price volatility. To address skewness and kurtosis later in the study, logarithmic values are used in regression testing.



Graph 6. The relative co-movements between averages of DATASET 1 and ETF prices for 2019-2020

Graph 6 shows that changes in global SVI are often more moderate than changes in US SVI. A similar relationship is also noticed in graph 7 below. The differences are not large but clearly noticeable. There is also some differences in the curves between US and global SVIs. However, the two follow each other surprisingly consistently.

One reason behind the slightly flattened curve and lower variance of global SVIs, is the larger volumes behind global data. These general higher levels of search volumes prevent radical changes from happening in the SVI. Whereas, in US alone the volume remains lower, which gives more weight to for example current trends to have an impact on attention. However, the United States cannot be categorized as a small market. It is highly likely, that by choosing a smaller market region, the difference in the curves would become even more stronger. In general, higher search volumes are seen to include higher noise levels, which also contributes to reducing variance.



Graph 7. The relative co-movements between averages of DATASET 2 and ETF prices for 2019-2020

In graph 7 the left vertical axis represents the percentage change in SVI variables and the right vertical axis represents the percentage variation in ETF returns. When comparing the percentage changes between DATASET 1 and DATASET 2, it is worth mentioning that the graphs differ quite significantly. During the ongoing COVID-19 crisis in 2020, a clear peak in search volumes was expected, as shown in graph 7. In graph 6, the changes in search volumes are more moderate. Thus, graphs 6 and 7 show well the differences in search volumes between the use of more general keywords and the use of ETF tickers. However, based on these findings, it is still difficult to draw any conclusions as to which approach better reflects the investor sentiment.

5.2. Cross-correlation matrix

The first statistical test conducted in this study is a cross-correlation matrix. Table 12 shows the possible correlation of the different variables with each other. The calculation of the correlation matrix is kept as simple and straightforward as possible. Therefore, no lag values are used, but values are calculated on same-week basis. Because most measured variables have multiple components, averages of these variables are used. The calculation includes a total of 13 variables based on data from five precious metals ETFs and different SVI variables. The time period used is from March 2015 to March 2020.

Even if it's on a general level, the table below gives a good indicative picture of the relations between different participants included in the study. It also enables comparison of results between this study and the preference study by DA et al. (2011). The first thing to notice is that there is a similarity in the correlation between realized returns of ETFs and other measured variables. For both studies, the ratio between returns and other variables is surprisingly low. As can be seen from the table 12 below, the strongest correlation between ETF returns is between the variable ASVI for DATASET 1 US, giving a value of 0.171. As the name implies this variable describes abnormal search volumes of the first approach (DATASET 1) using US data. In addition to the research by Da et al. (2011), this study has an additional SVI approach consisting of generic search words. The difference between the two based on the correlation matrix is that the first variable DATASET 1 has a positive relationship to returns, while the second variable DATASET 2 has a negative relationship. Since the correlation is so widespread between different variables, it means price changes are influenced not only by one, but by several factors. A similar effect is shown for abnormal returns, where the highest correlation is also between abnormal search volumes of DATASET 1.

In comparison between the US and global data, the correlation between different SVI-based variables is strong. This suggests that at least the averages of U.S. and global data follow each other quite consistently. The high correlation between US and global data may also indicate about successful noise reduction during data processing. Large differences, on the other hand, could mean that some unwanted search data is included in the data. To interpret the table below, a small difference in the correlations is detected. For the DATASET 1, US data have a stronger correlation with other variables. This is vice versa for the second dataset, where stronger correlation is shown for the global data. Based on these observations no significant differences in results are expected between the U.S. and global data. Greater differences in results are expected between DATASET 1 and DATASET 2, as they represent different approaches.

Finally, abnormal volumes appear to have a stronger correlation with SVI variables than ETF returns. Additional information is needed to better understand the relationships before drawing any conclusions. This is done by using the simple OLS regression and panel data regression in the following sections of the work. Due to the nature of

DATASET 2, it is possible that different words have different effects on the underlying asset. As a result, the use of averages can lead to offsetting effect and false assumptions about certain words. This may be the reason why the variables related to DATASET 2 show a relatively weaker correlation between returns and abnormal returns.

Table 12. Cross-correlation matrix of different variables

This table presents the correlation matrix of 13 variables used in this study. The data consists of 253 weekly observations from May 2015 to March 2020. To interpret the variables in the table, ASVI refers to the calculated abnormal search volumes. DATASET 1 and DATASET 2, on the other hand, represent the different SVI approaches defined in the data section of this study. DATASET 1 contains ETF tickers as search words, whereas, DATASET 2 uses generic search words. The finance related variables are formed based on the sample that consist of all 5 ETFs shown in the table 8. The number in parentheses after each variable represents numbers on the horizontal axis.

	1.	*2.*	*3.*	*4.*	*5.*	*6.*	*7.*	*8.*	*9.*	*10.*	*11.*	*12.*	*13.*
ETF returns (1)	*1*												
Abnormal return (2)	*0,958*	*1*											
ASVI for DATASET 1 US (3)	*0,171*	*0,217*	*1*										
ASVI for DATASET 1 GLO (4)	*0,162*	*0,207*	*0,878*	*1*									
DATASET 1 US (Log) (5)	*0,103*	*0,124*	*0,615*	*0,556*	*1*								
DATASET 1 GLO (Log) (6)	*0,089*	*0,100*	*0,473*	*0,551*	*0,934*	*1*							
ASVI for DATASET 2 US (7)	*-0,096*	*-0,004*	*0,686*	*0,697*	*0,456*	*0,412*	*1*						
ASVI for DATASET 2 GLO (8)	*-0,105*	*-0,009*	*0,684*	*0,734*	*0,494*	*0,460*	*0,934*	*1*					
DATASET 2 US (Log) (9)	*-0,078*	*0,006*	*0,638*	*0,648*	*0,605*	*0,587*	*0,873*	*0,842*	*1*				
DATASET 2 GLO (Log) (10)	*-0,094*	*-0,023*	*0,509*	*0,557*	*0,602*	*0,630*	*0,700*	*0,754*	*0,898*	*1*			
Abnormal volume (11)	*-0,045*	*-0,031*	*0,243*	*0,267*	*0,152*	*0,135*	*0,217*	*0,224*	*0,291*	*0,225*	*1*		
Turnover (Log) (12)	*0,065*	*0,095*	*0,205*	*0,245*	*0,318*	*0,290*	*0,089*	*0,142*	*0,303*	*0,293*	*0,710*	*1*	
Market value (Log) (13)	*0,030*	*0,010*	*-0,013*	*0,007*	*0,551*	*0,661*	*0,074*	*0,094*	*0,282*	*0,512*	*-0,080*	*0,017*	*1*

5.3. Regression results

After reviewing the correlation between the different variables, the study moves forward to presenting the regression results. This study uses two different types of regressions, simple OLS regression and panel data regression, both of which form their own subsections. In addition, a third subsection is formed, which seeks to gain a deeper picture of the different effects by performing additional regressions from the perspective of individual ETFs. This subsection is called robustness checks because the regressions performed measure a similar relationship on an individual level as the panel data regression on a general level.

5.3.1. OLS Regression

The aim of simple ordinary least squares (OLS) regression is to measure the possible statistically significant relation between SVI and trading volumes of an ETF. This section covers the first hypothesis (H1) of this research. Based on the earlier research the expectation is that SVI has some explanatory power over trading volume. This means that an increase in the number of searches on a keyword would affect the change in the trading volume of that ETF. The regression was conducted by using coding language R and the results are presented in table 13 below.

The regression formula was constructed based on the formula presented in the methodology chapter. Regression uses abnormal trading volume calculated in the methodology section as a dependent variable. Explanatory variables are abnormal search volume (ASVI), abnormal returns (AR) and logarithmic market value (MV). To track the reference study of by Da et al. (2011) as closely as possible, the regressions are calculated for two time periods, week t and week $t+1$. By adding a lag of one week to the dependent variable, the results show possible effects of explanatory variables for both the current and the previous week. Therefore, the time period $t+1$ represents the possible predictive power of used explanatory variables.

The first regression was conducted using a broad perspective, which means all variables included to the regression are averages. Thus, for example, abnormal return (AR) used in

the regression represents the average of abnormal returns of all ETFs, measured separately at each time point. The time period used in OLS regression is from May 2015 to March 2020 and contains 253 weekly time steps. In regression testing, due to how the SVI data is formed in this study, the overall effect of the variable ASVI is measured in the form of four variables. This means that ASVI of all the different SVI variables DATASET 1 US, DATASET 1 GLO, DATASET 2 US and DATASET 2 GLO are included in the regression formula. The effects of all four variables combined form a picture of the overall effects of ASVI.

Table 13. Simple OLS Regression for the whole sample

Table 13 reports the results of the simple OLS regression. The dependent variable is the abnormal trading volume at times t and t+1. Independent variables are abnormal search volume (ASVI), abnormal returns (AR) and logarithmic market value (MV). The regression has four different abnormal search volume variables 3, 4, 5, and 6. These variables represent different SVI approaches, as illustrated in figure 4. The regression is conducted for all five ETFs and time period is from May 2015 to March 2020. The regression contains 253 observations for each variable (7 variables), resulting in a total of 1771 observations.

OLS Regression – General level *Significance levels: '***' 0.01, '**' 0.05, '*' 0.10*		
Coefficients:	*t*	*t+1*
1. Abnormal returns (AR)	*-0.6054*	*-0.8448 **
2. Market value (MV) (Log)	*-0.1757*	*-0.1351*
3. ASVI for DATASET 1 US	*0.0861*	*0.4483*
4. ASVI for DATASET 1 GLO	*1.0276*	*1.1218 **
5. ASVI for DATASET 2 US	*0.0514*	*0.1012*
6. ASVI for DATASET 2 GLO	*0.0855*	*0.0987*
R-squared	*0.08711*	*0.176*

Table 13 presents the results of simple OLS regression. The regression includes a total of six variables, from which four variables (3, 4, 5 and 6) were different ASVI-related variables. The results do not show statistical significance on any of the explanatory variables on the same-week basis. During t+1, the only statistically significant effects observed are abnormal search volumes (ASVI) for DATASET 1 GLO and abnormal returns (AR). Both variables show an effect with a significance level of 10 %. Thus, the result in table 13 suggest that AR and ASVI have some predictive power in the following week's abnormal trading volumes. However, the relationship found is weak.

For the reliability of the results, it is good that the coefficients of the regression results are parallel for both time periods. Regarding abnormal returns, the relationship is negative, indicating that as the abnormal return increases, the trading volume is expected to decrease. On the other hand, the relationship of all ASVI variables is positive indicating that higher ASVI increases trade volume.

When search volumes are expected to represent a shift in attention in the market, the relationship between ASVI variables and abnormal trading volume is logical. When the attention increases also the trading volume is expected to increase. However, the negative relationship between abnormal returns and abnormal trading volume is not as logical. It also challenges Barber's and Odean's (2009) price pressure hypothesis.

When all observations are taken into account, the explanatory variables explain the changes in trading volume better for the following week than for the current week. Although both time periods produce low R-square values, the values of period t + 1 are more than twice as large as the values of period t. However, because the results are so weak, no conclusions can be drawn about the superiority of different perspectives or datasets. The aim of the first hypothesis was also not to compare different perspectives, but to find possible effects at a general level. Based on the reasoning above the first hypothesis (H1) is accepted conditionally and it is concluded that SVI has a weak effect on trading volume.

5.3.2. Panel data regression

This section approaches hypothesis two, three and four (H2, H3 & H4) by conducting a panel data regression. As the name suggests, the method is used to run a regression for panel-type data. In the context of this work, panel data refers to a dataset that contains observations of multiple resources at several different time series points. Another advantage of the model is that it handles large sample sizes and potential lags well.

The second hypothesis (H2) focuses on finding possible relationships between abnormal returns and abnormal SVI (ASVI). In Hypothesis three and four (H3 & H4), the focus

shifts to a comparison between the US and global data and to differences between the results of DATASET 1 and DATASET 2.

As with simple OLS regression, the panel data regression was constructed based on the regression formula presented in the methodology chapter. In the formula, abnormal returns are used as a dependent variable and abnormal trading volumes (AV), logarithmic market values (MV), and abnormal search volumes (ASVI) are used as explanatory variables. To enable the testing of the hypothesis three and four, similar to OLS regression calculation, there are four different ASVI variables in the panel data regression. These variables represent abnormal search volumes of DATASET 1 US, DATASET 1 GLO, DATASET 2 US and DATASET 2 GLO. The effects of all four variables combined form a picture of the overall effects of abnormal search volume (ASVI).

Panel data regression offers a variety of models from which the user can select the most appropriate one. These models act as different approaches in determining the specifics, how you want the regression to run. For example, in R, the “plm” function used to calculate panel data regression includes six different approaches, which are “pooling”, “within”, “between”, “random”, “fd”, or “ht”. In this research, both fixed effects (in R: within) model and random effects model were conducted. These were then compared to each other by performing a Hausman test. Results of Hausman test are presented in the table 14 below. As a result, fixed effects model was chosen to be used in this research because according to the Hausman test random effects model is inconsistent. A general threshold value for accepting the hypothesis related to the Hausman test is 0.05, which means that if p-value is greater than 0.05 the random effects model is consistent and efficient.

Table 14. Results of Hausman test

Hausman test
chisq = 22.67 *p-value = 1.195e-05*
Alternative hypothesis: one model is inconsistent

Five different time periods are used to measure the predictive power of future abnormal returns in both short and long term. As the table 13 shows, the first time period is same-week basis (1), the second (2), third (3) and fourth (4) time periods measure effects on future abnormal returns on a short-term, and finally, the fifth (5) time period measures effects on cumulative long term returns. The sample of panel data regression includes all five precious metal ETFs: GLD, IAU, SLV, GDX and GDXJ. The regression is conducted for time period from May, 2015 to March, 2020.

Table 15. Panel data regression for the whole sample (Coefficients in percentages)

This table reports the results of panel data regression. The dependent variable is the abnormal returns (in percentages) during the first four weeks and during the weeks 5-52. Independent variables are abnormal search volume (ASVI), abnormal returns (AR) and logarithmic market value (MV). The regression has four different abnormal search volume variables 3, 4, 5, and 6. These variables represent different SVI approaches, as illustrated in figure 4. The regression is conducted for all five ETFs and time period is from May 2015 to March 2020. The regression contains a total of 13915 observations.

Panel data regression – fixed effects model

*Significance levels: '***' 0.01, '**' 0.05, '*' 0.10*

Coefficients	*Week t* (1)	*Week t+1* (2)	*Week t+2* (3)	*Week t+3* (4)	*Weeks 4-52* (5)
1. Abnormal volume (AV)	*-0.31076*	*-0.30231*	*0.25070*	*0.15721*	*0.012947*
2. Market value (MV)(Log)	*0.62135*	*-2.68375 ****	*-2.72961 ****	*-2.73874 ****	*-0.564622 ****
3. ASVI for DATASET 1 US	*4.51956 ****	*3.28709 ****	*0.36315*	*-0.51978*	*-0.059457*
4. ASVI for DATASET 1 Glo	*7.14297 ****	*3.68702 **	*0.61147*	*4.11132 ***	*0.058454*
5. ASVI for DATASET 2 US	*-0.68636*	*-2.04149*	*-7.45283 ***	*5.39690 **	*0.040528*
6. ASVI for DATASET 2 Glo	*-6.65006 **	*-8.34317 ***	*4.05693*	*-8.22454 ***	*0.022351*
R-squared	*0.050659*	*0.046204*	*0.019979*	*0.01759*	*0.19974*

When comparing the ASVI of DATASET 1 and DATASET 2, a factor that is good to keep in mind is the difference in the nature of the keywords. The first variable (DATASET 1) contains ETF tickers of included ETFs, which allow the search word to be allocated to each ETF separately. Meaning search word "GLD" represents the search activity of SPDR Gold Shares. The second variable (DATASET 2), contains general keywords, which do not represent the search activity of any particular ETF. This causes

problems with panel data, as the search volume cannot be allocated to the corresponding ETF. This has been solved by using the averages of the DATASET 2 search volumes for each ETF. This may affect the accuracy of the results for the second data set.

For panel data regression, abnormal returns are expressed as a percentage. This means that the coefficients are also presented as a percentage change. By explaining this with an example from the table 15 above, it seems that variable 3 representing *ASVI for DATASET 1 US* has a strong positive effect of ~ 4.52 percent on the same-week basis and ~ 3.29 percent on the next weeks abnormal returns. The percentage change in abnormal returns happens in relation to every one-standard-deviation increase in variable 3. During both time periods (*t and t+1*) the effects are significant at a 1% significance level. This is consistent with the results found by Da et al. (2011), who also found statistically significant effects between ASVI and abnormal returns at weeks t + 1 and t + 2. Da et al. (2011) also showed a strong price reversal in the coming weeks. However, similar significant opposite relationships are not seen in table 15 for variables 3 and 4 related to ASVI for DATASET 1.

According to the second hypothesis, the relationship between abnormal search volumes and abnormal returns is expected to be negative. This hypothesis was set based on the relationship found in the reference study by Preis et al. (2013). The results shown in table 15 show two types of results. The DATASET 1 variables show strong positive effects that are consistent over time periods t and t + 1. DATASET 2 variables, on the other hand, show weaker negative but somewhat mixed results. Thus, the observed effects of variables 3 and 4 contradict the second hypothesis. As stronger performers, the effect of these variables also leads to the rejection of the second hypothesis (H2). The opposite effects of the two variables DATASET 1 and DATASET 2 make it difficult to interpret general-level effects.

The third hypothesis (H3) tests whether U.S. SVI data act as better predictors than global SVI data. For this hypothesis, the effects of variables 3 and 5 are compared with the effects of variables 4 and 6. Table 15 shows that global data produces stronger effects. However, the difference between the two is not large, but sufficient to reject the third hypothesis. Finally, the fourth hypothesis (H4) tests if ETF ticker data (DATASET 1)

acts as a better predictor than the generic search data (DATASET 2). The null hypothesis is set so that DATASET 1 is a better predictor. Although the effects are opposite for DATASET 1 and DATASET 2, hypothesis 4 is accepted, stating that variables 3 and 4 representing the ASVI of DATASET 1 have stronger and more consistent statistically significant effects. Both hypotheses three and four also take into account the results of robustness tests presented in the next section of the work.

What comes to other results logarithmic market value shows statistically significant predicting power over abnormal returns both short- and long-term. The effect of log market value is negative, meaning when log market value increases, abnormal returns decreases. Differing from the reference study, the results of this research show no statistically significant effects on abnormal trading volume. Whereas, Da et al. (2011) show that abnormal turnover has an effect during the first two weeks $t+1$ and $t+2$, with a significance level of 5 %. Finally, the values of R-square are small, this was to be expected from the reference study by Da et al. (2011).

5.3.3. Robustness checks

The goal of this section is to obtain additional information regarding individual effects of search words to test the fifth hypothesis (H5). The hypothesis (H5) tests, whether some of the search terms have stronger effects on results than others. This offers interesting insight regarding DATASET 2 with generic search words, as it has been earlier measured only by using mean values. Individual-level comparison using OLS regression uses the same approach as panel data regression and thus this section also serves as a robustness check for panel data regression results.

A simple OLS regression is performed on each of the five ETFs. Like in the panel data regression the dependent variable is abnormal returns. However, ETF-specific variables are used instead of averages. Abnormal search volumes for individual search words are used as explanatory variables. There are a total of 15 different search words for each ETF that are measured (DATASET 1 = 1 keyword, DATASET 2 = 14 keywords). Regressions are performed for two time periods, t and $t+1$. Where, t measures the effects on same-week basis, while $t+1$ measures the possible effects on next week's abnormal returns.

In hypothesis testing, hypothesis five (H5) is accepted if differences are observed between the effects of individual keywords. Therefore, the focus is on statistically significant variables and their ASVI coefficients. Due to the large number of words measured, only statistically significant variables are included to the tables 16, 17, 18, 19 and 20. Similar to panel data regression abnormal returns are presented in percentages, and thus coefficient factors are also showing in percentage change. The first table represents results of SPDR Gold Shares ETF (GLD).

Table 16. Results for SPDR Gold Shares ETF (GLD)

This table reports the results of individual level regression for SPDR Gold Shares (GLD). The dependent variable is the abnormal returns as percentages at times t and t+1. Independent variables are abnormal search volumes (ASVI) of all search terms included to this research. Out of all 15 search words measured, table below presents only the variables that have statistically significant values. Time period used to test these effects is from May 2015 to March 2020. The regression contains 253 observations for each variable (16 variables), resulting in a total of 4048 observations.

GLD				
*Significance levels: '***' 0.01, '**' 0.05, '*' 0.10*				
	US		*Glo*	
Coefficients	*t*	*t+1*	*t*	*t+1*
GLD (ETF ticker)	*6.5606 ****	*1.8440*	*10.4393 ****	*2.3844*
Gold price	*-3.7447*	*0.2975*	*-8.1425 ****	*-2.4567*
Gold ETF	*1.1301*	*0.3892*	*4.1192 ****	*3.1527 **
Silver price	*2.6327*	*-2.1108*	*-1.0159*	*-4.7985 ***
Precious metals	*-2.9719 **	*2.4246*	*2.5762*	*4.0004 **
S&P 500	*-1.2695*	*-2.2826*	*-4.4442*	*-5.6702 ***
VIX	*-0.8979*	*1.6660*	*-0.9258*	*2.0124*
Stock market	*0.9976*	*-0.5402*	*4.8555 ***	*-0.0963*
R-squared	*0.0946*	*0.0684*	*0.1633*	*0.1090*

Of the 15 keywords tested in the regressions shown in tables 16-20, only 8 words (ETF ticker, Gold price, Gold ETF, Silver price, Precious metals, S&P 500, VIX, and Stock market) have any statistical significance. Only results for these keywords are shown in the tables. This means that results of other 7 keyword have been omitted from the tables. These words are "Financial markets", "ETF", "iShares ETF", "Vanguard ETF", "SPDR ETF", "Commodities" and "NYSE".

Already straight away from the first table 16 for GLD, it can be stated that global data seems to have more predicting power than US data. There is more statistically significant variables as well as higher R-squared values. From the words presented in the table 16,

ETF ticker “GLD” and search words “Gold ETF”, “Precious metals” and “Stock market” seem to have positive effects, whereas search words “Gold price”, “Silver price” and “S&P 500” has negative relationship on abnormal returns. Next table 17 displays results of iShares Gold Trust (IAU).

Table 17. Results of iShares Gold Trust (IAU)

This table reports the results of individual level regression for iShares Gold Trust (IAU). The dependent variable is the abnormal returns at times t and t+1. Independent variables are abnormal search volumes (ASVI) of all search terms included to this research. Out of all 15 search words measured, table below presents only the variables that have statistically significant values. Time period used to test these effects is from May 2015 to March 2020. The regression contains 253 observations for each variable (16 variables), resulting in a total of 4048 observations.

IAU				
*Significance levels: ‘***’ 0.01, ‘**’ 0.05, ‘*’ 0.10*				
	US		*Glo*	
Coefficients	*t*	*t+1*	*t*	*t+1*
IAU (ETF ticker)	*1.9521 ** *	*0.0893*	*4.2546*	*-1.7252*
Gold price	*-1.7205*	*1.1812*	*-8.4010 *** *	*-2.3286*
Gold ETF	*1.8981 ** *	*0.6810*	*6.0690 *** *	*3.6903 ** *
Silver price	*2.7031*	*-2.0877*	*-0.5441*	*-5.0203 ** *
Precious metals	*-2.5373*	*2.4951*	*3.2362*	*4.2421 ** *
S&P 500	*-1.4205*	*-2.1122*	*-3.2641*	*-5.8093 ** *
VIX	*-0.5330*	*1.7540*	*-0.8839*	*2.0595*
Stock market	*1.4579*	*-0.5314*	*5.3017 ** *	*0.0773*
R-squared	*0.07403*	*0.06488*	*0.1386*	*0.1082*

Like the GLD results in table 16, the results in table 17 show that using global data results in larger number of statistically significant variables and higher R-square values. For both GLD and IAU search word “Gold ETF” is the only word to show statistical significance for both time periods *t* and *t+1*. A similar division between keyword effects is also repeated in table 17. The results of the iShares Silver Trust (SLV) ETF are shown in Table 18 below.

Table 18. Results of iShares Silver Trust (SLV)

This table reports the results of individual level regression for iShares Silver Trust (SLV). The dependent variable is the abnormal returns at times t and t+1. Independent variables are abnormal search volumes (ASVI) of all search terms included to this research. Out of all 15 search words measured, table below presents only the variables that have statistically significant values. Time period used to test these effects is from May 2015 to March 2020. The regression contains 253 observations for each variable (16 variables), resulting in a total of 4048 observations.

SLV				
*Significance levels: '***' 0.01, '**' 0.05, '*' 0.10*				
	US		*Glo*	
Coefficients	*t*	*t+1*	*t*	*t+1*
SLV (ETF ticker)	*8.4864 ***	*11.6979 ****	*16.8931 ****	*4.5706*
Gold price	*-3.9320*	*0.81117*	*-12.7589 ****	*1.4580*
Gold ETF	*2.2922 **	*-0.7765*	*6.8536 ****	*0.7348*
Silver price	*2.9103*	*-9.9305 ***	*2.0053*	*-4.8722*
Precious metals	*-4.8512 **	*3.4413*	*1.0566*	*4.4161*
S&P 500	*-3.5891*	*-6.2989*	*-7.3696 **	*-7.4746*
VIX	*-0.6027*	*2.9085*	*-0.2004*	*4.7435 ***
Stock market	*1.4192*	*0.3845*	*8.7734 ***	*-4.4970*
R-squared	*0.08517*	*0.1023*	*0.1371*	*0.08384*

Regarding SLV results in table 18, strongest predicting power is shown for the US data at *t+1*. Strong effects are shown for the ETF ticker (SLV) and search word "Silver price". For ETF ticker the relationship discovered is positive similar to findings of reference study by Da et al. (2011). On the other hand, keywords related to metal prices, such as " Gold price" and "Silver price", show a negative relationship.

Table 19. Results of VanEck Vectors Gold Miners ETF (GDX)

This table reports the results of individual level regression for VanEck Vectors Gold Miners ETF (GDX). The dependent variable is the abnormal returns at times t and t+1. Independent variables are abnormal search volumes (ASVI) of all search terms included to this research. Out of all 15 search words measured, table below presents only the variables that have statistically significant values. Time period used to test these effects is from May 2015 to March 2020. The regression contains 253 observations for each variable (16 variables), resulting in a total of 4048 observations.

GDX				
*Significance levels: '***' 0.01, '**' 0.05, '*' 0.10*				
	US		*Glo*	
Coefficients	*t*	*t+1*	*t*	*t+1*
GDX (ETF ticker)	*7.3232 **	*5.9526*	*12.8083 ***	*8.0207*
Gold price	*-3.1544*	*5.6932*	*-17.0748 ****	*1.1354*
Gold ETF	*3.4799 **	*0.0582*	*9.4340 ***	*3.4096*
Silver price	*4.7479*	*-9.8565*	*-4.6035*	*-15.6933 ****
Precious metals	*-2.9599*	*3.5389*	*10.3127 ***	*7.5057*
S&P 500	*3.7450*	*-2.0510*	*-4.5646*	*-13.9765 **
VIX	*-0.2873*	*5.6704*	*-1.9597*	*3.3306*
Stock market	*-2.0529*	*-6.7636*	*6.4408*	*0.2797*
R-squared	*0.07581*	*0.08245*	*0.1209*	*0.1078*

GDX and GDXJ are both gold miner ETFs and thus differ from the other ETFs in the sample. Instead of physical commodities, gold mining companies include shareholdings in large precious metal producers. Equities are seen to be more volatile than commodities. This is one explanatory factor for the relatively strong effects of GDX and GDXJ in tables 19 and 20.

Table 20. Results of VanEck Vectors Jr Gold Miners ETF (GDXJ)

This table reports the results of individual level regression for VanEck Vectors Jr Gold Miners ETF (GDXJ). The dependent variable is the abnormal returns at times t and t+1. Independent variables are abnormal search volumes (ASVI) of all search terms included to this research. Out of all 15 search words measured, table below presents only the variables that have statistically significant values. Time period used to test these effects is from May 2015 to March 2020. The regression contains 253 observations for each variable (16 variables), resulting in a total of 4048 observations.

GDXJ				
*Significance levels: '***' 0.01, '**' 0.05, '*' 0.10*				
	US		*Glo*	
Coefficients	*t*	*t+1*	*t*	*t+1*
GDXJ (ETF ticker)	*10.6189 ****	*8.7121 ****	*10.9750 ****	*9.7662 ****
Gold price	*-3.8643*	*4.0841*	*-14.8413 ***	*4.3051*
Gold ETF	*2.6432*	*-0.5334*	*8.0224 **	*1.1191*
Silver price	*4.9507*	*-10.2779*	*-6.1695*	*-14.8189 ***
Precious metals	*-3.1696*	*5.6198*	*10.2120 **	*4.9164*
S&P 500	*-5.4848*	*-10.7499*	*-8.3120*	*-17.5798 ***
VIX	*0.0265*	*8.1234 ***	*-1.1694*	*6.4518 **
Stock market	*2.7739*	*-6.1218*	*7.5314*	*0.2797*
R-squared	*0.1156*	*0.1365*	*0.1286*	*0.1347*

To analyze the results from the perspective of the fifth hypothesis, the hypothesis is accepted and it is stated that clearly part of the words are more likely to have an effect than others. For tables 16-20, this shows that the same 8 keywords show statistical significance. Thus, it is possible to rate these search words based on their predictive capabilities into so called "strong" and "weak" search words. Where, strong keywords are prone to have an effect and weak words have no effects at all.

Another interesting finding is that some of the words under analysis have strong negative relationships, while some have strong positive. For example, "Gold price" has strong negative effect and "Gold ETF" a strong positive. Some words follow a very rational logic, but for some search words, it is difficult to know in advance how the keywords will work. Because of these positive and negative effects, the use of averages can lead to false assumptions about some search words.

There are similarities in results between the individual level approach (robustness checks) and panel data regression. Both tests found evidence of a positive relationship between ETF tickers (DATASET 1) and abnormal returns on ETFs. For the general keywords in DATASET 2, the robustness checks show both negative and positive effects of the

keywords. The mixed effects of the different search words in tables 16-20, explain the somewhat mixed results of the panel data regression shown in table 15. Tables 16-20 also show that the observed negative effects tend to be stronger than the positive ones. This is also shown in table 15 in the panel data regression results as occasional negative effects for variables 5 and 6.

Preis et al. (2013) state the found relationship to be negative. This means that as search volume increases, the price of the security is expected to fall. Based on the regression results found in this study this applies only to some individual search words, but not for the majority of search words. This also confirms the correct decision to reject hypothesis two and three and accept the fourth hypothesis. Research is continued to test the sixth hypothesis and to determine whether the moving average calculations support the results presented so far.

5.4. Moving average calculations

This final subchapter presents the results of moving average calculations and aims to answers the final hypothesis six (H6). This part of the study was built on the basis of the reference study by Preis et al. (2013). The goal is to see, whether it is possible to earn excess returns by timing the market with Google search volumes. This allows one to see how the actual investment strategy would work over the time period used.

Preis et al. (2013) suggest that certain search words can be used to anticipate market movements. What they did, was conducted an investments strategy for individual search terms, where investment decisions were made based on moving averages of search volumes. The study included altogether 98 finance related search words. They then compared the performance of each search term to see which of the search terms performed the best. As a result, Preis et al. (2013) got the performance of all words for the whole time period. They claimed that search word “debt” performed the best and that a single-word-strategy was able reach a profit of 326 % between 2004 and 2011. During the same time period a buy-and-hold strategy supposedly made only a profit of 16 %. Based on these results, the initial assumption made regarding the sixth hypothesis is that it is possible to earn excess returns with the market timing strategy.

Similar process to the research by Preis et al. (2013), was also conducted in this research. The relationship found in the reference study by Preis et al. (2013) was also used to implement the strategies of this study. Preis et al. (2013) compared the performance of different search words to a single asset. To imitate the reference study and to simplify the calculations, the decision was made to invest in only one asset. Out from all ETFs included in the research GLD is by far the largest one. When analyzing the nature of all ETFs included to the research, it was noticed that stock-based ETFs are highly volatile compared to actual physically backed precious metal ETFs. Thus, a large physically backed ETF is expected to be a better proxy for measuring the performance of the overall precious metals market than a riskier and more volatile option. Another observation is that GLD and IAU follow each other very closely. Based on the aforementioned reasoning, formed strategies are implemented only for SPDR Gold Shares (GLD). This provides a good cross-section of commodity market performance.

Since the strategy is only generated for the SPDR Gold Share ETF, the keywords chosen are the ones that show statistically significant effects on abnormal returns on the underlying asset in table 16. These words are the ETF ticker “GLD” and search words “Gold price”, “Gold ETF”, “Silver price”, “Precious metals”, “S&P 500” and “Stock market”. The cumulative return of each search word is compared to two different buy-and-hold strategies. The first one is buy-and-hold strategy for SPDR Gold Shares and the second one is buy-and-hold strategy for iShares S&P 500 Index ETF (IVV). The measured excess return is determined in relation to the market performance represented by the second buy-and-hold strategy formed for IVV. The buy-and-hold strategy for GLD is designed to provide an indicator of how the ETF is performing over the time period used.

Table 21 describes all the formed strategies. In this study, the relationship on which strategies are constructed is determined based on the reference study by Preis et al. (2013). Thus, all strategies follow the same trading rules regardless of the relationship found in the regression results. In addition to search volume data, strategy itself, and trading rules, table 21 contains the relationship found in the regression results. The relationship of

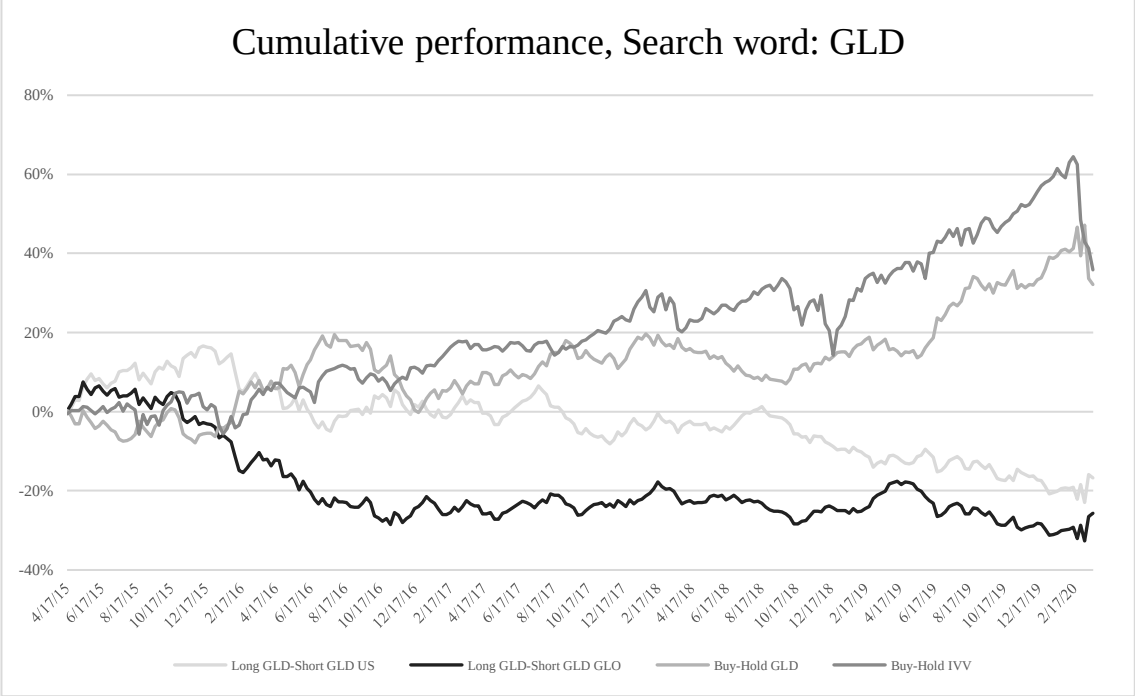
regression results is formed based on the results in table 16. This is presented to determine if the performance of different strategies is in line with the regression results.

Table 21. Strategy formation in moving average calculations

SVI Data	***Strategies***	***Relationship (regression)***	***Trading rules***
1.GLD (ETF ticker)	*Long GLD-Short GLD*	*positive*	*ASVI < 0 → buy long* *ASVI > 0 → sell short*
2. Gold ETF	*Long GLD-Short GLD*	*positive*	*ASVI < 0 → buy long* *ASVI > 0 → sell short*
3. Gold price	*Long GLD-Short GLD*	*negative*	*ASVI < 0 → buy long* *ASVI > 0 → sell short*
4. Silver price	*Long GLD-Short GLD*	*negative*	*ASVI < 0 → buy long* *ASVI > 0 → sell short*
5. Precious metals	*Long GLD-Short GLD*	*positive*	*ASVI < 0 → buy long* *ASVI > 0 → sell short*
6. S&P 500	*Long GLD-Short GLD*	*negative*	*ASVI < 0 → buy long* *ASVI > 0 → sell short*
7. Stock market	*Long GLD-Short GLD*	*positive*	*ASVI < 0 → buy long* *ASVI > 0 → sell short*

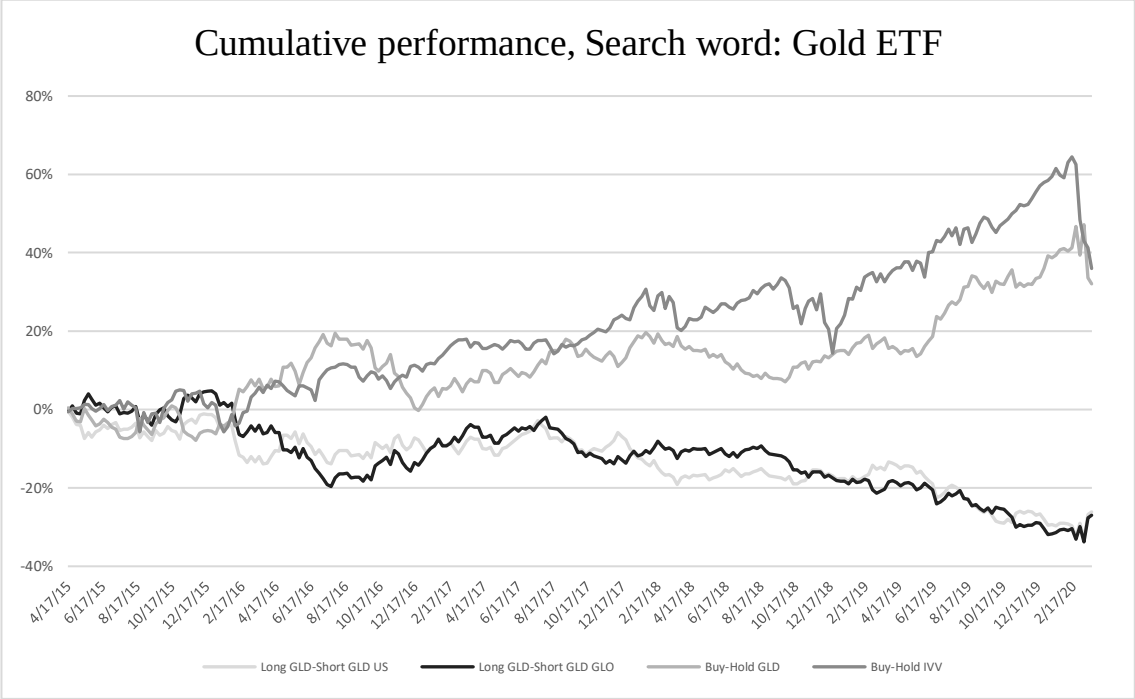
Table 21 shows that the strategies formed are long GLD – short GLD. This means that if the abnormal search volume (ASVI) of the search word decreases, strategy buys the asset (GLD) and if the ASVI of the search word increases, strategy sells short the underlying asset. With regard to hypothesis testing, the sixth hypothesis expects that the strategies formed on the basis of table 21 win the market (buy-and-hold IVV) over the measured time period.

Of all the strategies generated, only the results for the keywords “GLD”, “Gold ETF” and “S&P 500” are shown in graphs 8, 9 and 10 below. These three words, showed the strongest statistically significant results and a clear positive or negative relationship in the regression results. The graphical presentation for the rest of the words is shown in appendix 4. The time period under review is from April 2015 to March 2020. The time period under consideration gives an interesting aspect as the market has just witnessed a significant market crash in yearly 2020. It is promising if search volumes indicate some type of warning signs and the strategy is able dodge this hazard. Graph 8 below shows a graphical representation of the strategy generated using the search volume of ETF ticker “GLD”.



Graph 8. Cumulative performance of strategies using search word GLD

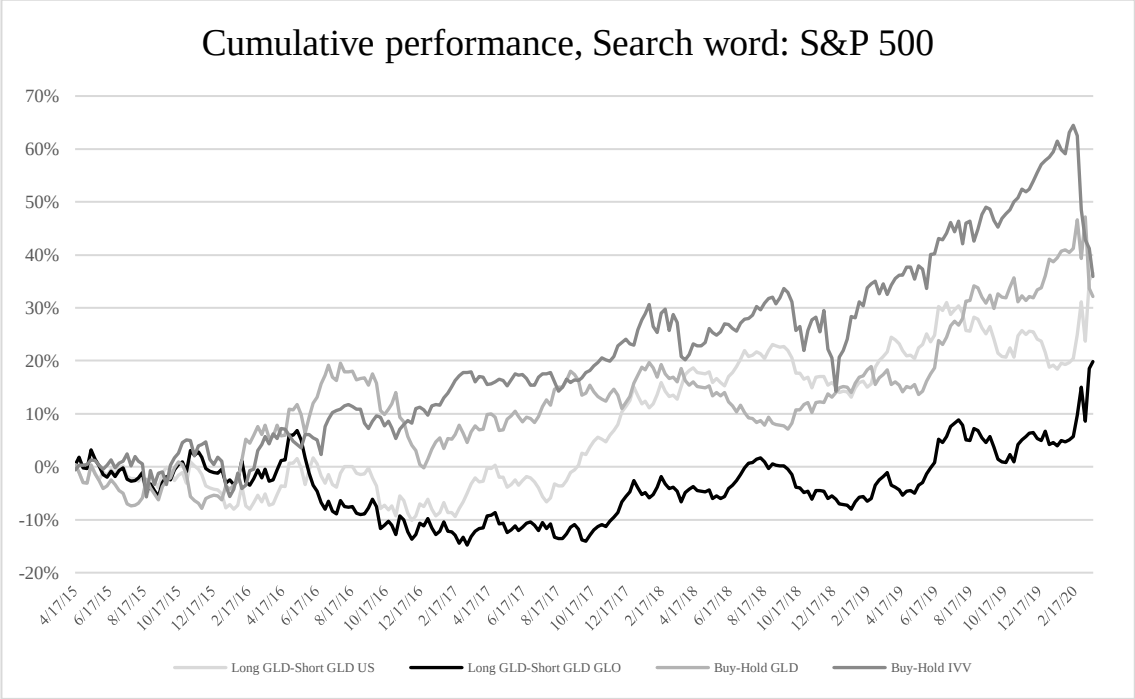
The graph 8 represents the strategies related with the search word “GLD”. The buy-and-hold strategy for SPDR Gold Shares returns a profit of 32.2% for the whole time period, whereas the buy-and-hold strategy for iShares S&P 500 Index ETF returns a profit of 35.9%. The recent market crash is eroding the market returns. This is easily noticed in the graph by looking at the performance of iShares S&P 500 Index (IVV). A similar strong effect is not seen for SPDR Gold Shares. This is a great example of the ability of gold to maintain its value. After reducing transaction costs both actively traded strategies perform significantly weaker than markets, generating negative returns. The strategy using US data generated -16.8% and the strategy with global data generated -25.7% during the whole time period. The following graph 9 shows the results for the strategy generated using the search volumes for the word "Gold ETF".



Graph 9. Cumulative performance of strategies using search word Gold ETF

Similar to the graph 8, after reducing transaction costs both actively traded strategies perform significantly weaker than markets, generating negative returns.

According to the regression results, both search words “GLD” and “Gold ETF” have a strong positive statistically significant effects on abnormal returns on SPDR Gold Shares. A positive relationship indicates that as search volume increases, abnormal returns are also expected to increase. As the trading rules were constructed using the opposite view of the reference study by Preis et al. (2013), these strategies are expected to perform poorly. As a result, the strategies in tables 8 and 9 are consistent with the regression results. This also means that these strategies would work better if the trading rules were set according to the regression results. As the performance of the strategies is only measured by trading rules built on the basis of the reference study by Preis et al. (2013), this remains a mere assumption.



Graph 10. Cumulative performance of strategies using search word S&P 500

Finally, graph 10 shows the results of a strategy that uses the search volume of the keyword "S&P 500" to make investment decisions. Unlike the strategies considered in graphs 8 and 9, the regression results show a negative relation between keyword S&P 500 and SPDR Gold Shares ETF returns. This means that Preis et al. (2013) and the regression results show a similar relationship. This is also seen as better performance of the active strategies in graph 10. The total return on the long GLD - short GLD US strategy during the time period was 36.6%. With a market return of 35.9% over the same period, this means that the strategy barely wins the market by earning a 0.7% excess return. At its highest, the buy-and-hold strategy representing market performance, was up 64.5%. Active trading strategies are at their peak at the end of the review in March 2020.

As mentioned at the beginning of this section, the initial assumption of the last hypothesis six is that excess returns are possible through an investment strategy that takes advantage of abnormal search volumes. Based on the results shown in graphs 8, 9, and 10 above, the hypothesis is rejected. In general, strategies related to SPDR Gold Shares (GLD) do not appear to offer significant added value, especially when using the approach by Preis et al. (2013). The return on a gold ETF is generally lower than market performance in the long run. The only strategy that managed to beat the market during the period under

review is a strategy that leverages S&P 500 search results to make investment decisions. However, the excess returns are so small that it is not profitable to run the strategy in real life.

The recommendation made based on the results of this research is that gold is a good safe haven investment and a good way to diversify your portfolio to less risky assets. However, it is not an investment that is expected to generate large profits in the short term. Therefore, precious metal ETFs are rather recommended in hedging and diversification purposes. As can be seen from graphs 8-10, the buy-and-hold strategy for SPDR Gold Shares works almost as well over the time period as the market index ETF iShares S&P 500. Therefore, a longer-term investment instead of an active trading strategy may be a better option for precious metals ETFs.

As the study progressed, understanding increased with respect to the keywords used. General keywords work differently than ETF tickers. This is because ticker search volumes all have a similar relationship to the underlying asset. However, for generic search words the direction of the relationship is dependent on the search word itself and on the underlying asset. Based on the results found in this section, although search volumes have significant predictive power for future commodity ETF returns, these predictive forces can hardly be translated into profitable trading strategies. This may be due to the opposite effect of different search words occurring simultaneously. In addition, the impact of individual search word volume on returns could be time-varying therefore using the same (different) search words in different (same) times could be more profitable than using a unified trading strategy over the whole time period.

6. Conclusions

The goal of this section is to summarize the results and draw a common conclusion based on all results. The research is concluded by answering all hypothesis one by one and answering the research questions introduced in the beginning of the thesis. The chapter consists of two sub-sections from which the first one forms the actual conclusions and the second one describes possible further research.

6.1. Concluding the research

The purpose of this study was to find evidence on whether market timing is possible using search query-based information and how an actual investment strategy would perform over the selected time period. In this study the chosen approach is to study the link between Google search volumes and precious metal ETFs in the US market from May 2015 to March 2020. Precious metals were chosen as the underlying asset as they are expected to be a good indicator of market nervousness. This nervousness would show as increasing attention in Google search volumes of used keywords and possibly give some early warning signs of market movements. The novelty of this study lies in the use of a sentimental proxy in investment decisions for commodity ETFs. The debate of the use of search query volumes as a sentimental factor, in general, is related to whether search behavior captures the investor sentiment or not. However, many of the studies presented throughout this thesis, have found search query data to be a reliable proxy for investor sentiment and produce promising results (Da et al., 2011; Aouadi et al., 2013; Preis et al., 2013; Klemola, 2019). There is also a discussion about the direction of the possible relationship between search volumes and security prices. Representatives of both positive and negative perspectives can be found (Preis et al., 2011; Barber & Odean, 2008). Therefore, this research uses a slightly different approach to the existing research by focusing on commodity ETFs. The relationship set based on the research by Preis et al. (2013) was expected to be negative.

The study used two quite different approaches as a reference study to examine the link between search volumes and ETF prices. From both reference studies by Da et al. (2011) and by Preis et al. (2013), an own approach was implemented. As a result, both ETF

tickers and general predefined keywords were used as search words in the study. The inclusion of both approaches made it possible to compare the possible effects of both approaches. As limiting factors this study shows that while the adequacy of Google Trends data has gradually improved, it is still the biggest limiting factor in a study of this type. The final sample of this research included only 5 ETFs. Both noise and sufficiency of data were critical factors leading to the exclusion of ETFs. It should be noted that the size of the ETF is directly related to finding sufficient information. The smaller the ETF in terms of its assets under management (AUM), the less likely it is to meet adequate information requirements. This can affect significantly the reliability of the results. Although the number of ETFs in this study remains small, the assets under management of the five selected ETFs are 88 billion. Another limiting factor of this study is availability of data in Google Trends. Only maximum of five years of weekly data can be accessed and exported from the website. If access to data greater than five years is required, Google Trends automatically converts the data into monthly data that is not dense enough in most financial studies. For this study, it was decided to settle for a period of five years and not to start merging and scaling Google's normalized data. The final time period ended up being two months (8 weeks) shorter than the original five-year period.

There are a total of six hypotheses in this study. The empirical analysis were conducted to test each of these hypotheses. This was done by focusing on three different entities. The first hypothesis forms the first entity and tested the relationship between search volumes and the trading volumes of the underlying assets. This is tested by a simple OLS regression. According to the null hypothesis, increase in SVI affect the changes in the trading volume of the underlying assets. The findings of the first hypothesis (H1) show a weak positive relationship between abnormal search volume and next week’s trading volume. The occurred effect is significant on a 10% significance level. The hypothesis was eventually accepted only conditionally because the relationship was so weak and inconsistent.

The second entity tested hypotheses two, three, four and five. The focus of this entity was to test the relationship between search volumes and returns of the underlying assets. This was tested by conducting statistical methods such as panel data regression and OLS regression. Like stated above the expectation was to find a negative relationship between

the search volume and returns. However, the hypothesis H2 was rejected as the overall relationship found was positive. This means that an increase in attention, measured in weekly search volumes, has a positive effect on ETF prices rather than a negative effect. This found relationship is consistent with attention theories of Barber and Odean (2008) and the results of Da et al. (2011). However, unlike the results by Da et al. (2011), the results of this study did not show a price reversal in the long run. Other hypothesis in this entity (H3, H4, and H5) focused on comparing different approaches to gain deeper insight into potential impacts. First the US and global data were compared as predictors. In the context of this research the use of global data provided stronger statistically significant results compared to the use of US data. Next, the performance of the two different types of keywords, both ETF tickers and general predefined search words, were compared. Between the two keyword types, ETF tickers showed stronger and more consistent effects in terms of statistical significance and directions of effects. Finally, an individual approach was performed, which was possible due to the small sample size. The individual-level approach taken in H5 tested if there were differences in the effects of the keywords. At the same time the approach acted as a robustness check for results of panel data regression. It was found that the same 8 words, out of the total 15 words tested for each ETF, showed statistically significant results between different ETFs. Thus, the keywords used could be categorized into so-called strong and weak words. In addition, it was found that the direction of the relationship varies between words. When using average search volumes, this means that the data contains opposite, offsetting effects. This can be seen as the reason for somewhat mixed results of the general keywords in panel data regression. Robustness checks were seen to be consistent with the panel data regression results.

The final entity consist of hypothesis six and moving average calculations. The goal of H6 was to test if excess returns can be earned with a simple moving average trading strategy that takes advantage of SVI values. The strategy was conducted using only the statistically significant keywords for SPDR Gold Shares ETF presented in table 16. Trading rules were based on the expected relationship (negative) used in the second hypothesis (Preis et al., 2013). In this research the hypothesis was rejected and stated that the trading strategy using commodity ETFs do not offer any significant added value. It can also be noted that the relationships found in the moving average calculations were

similar to the relationships found in the regression results. This means that when using predefined trading rules based on reference study, the strategies work relatively better when the relationship observed in the regression results was negative. On the other hand, if the ratio observed in the regression results was positive, the performance of the strategy was lower. Out of all search words, trading on the basis of search volume of keyword "S&P 500", generated a highest raw return of 36.6% which was only 0.7% higher than the market return during the same time period. This shows that although the search volumes possess significant predictive power for future commodity ETF returns, these predictive powers can hardly be translated into profitable trading strategies. This may be due to the opposite effect of different search words occurring simultaneously. In addition, the impact of individual search word volume on returns could be time-varying therefore using the same (different) search words in different (same) times could be more profitable than using a unified trading strategy over the whole time period.

Results are concluded by shortly answering the research questions. The first one is, can the investor attention be captured by using search volume data? The answer to this questions is yes. Especially, in the same week and the following week, the results show consistency for some variables in terms of effect size, direction of effects, and statistical significance of the coefficients. The assumption, large precious metal ETFs indicate market nervousness, is explained with a relationship between abnormal returns of precious metal ETFs and market related search words. The second question holds, is timing the market possible with Google Trends search volume data? The results indicate that, yes, it is possible at some level and with some specific keywords. Certain search terms are seen to offer insight on whether the ETF prices are moving up or down over short-term ($t+1$). Market timing can also be time-dependent for a particular keyword. Like in the reference study by Da at al. (2011) and Preis et al. (2013), the R-squared values remain low also in this research. This indicates the detected effect are there but the relationship is weak. This and the small sample size affect the reliability of the results and should be kept in mind. The final research question is, can excess returns be made by using search volume data? The answer to this question must be no, because the strategies developed work poorly with the pre-defined trading rules. Timing the market with Google Trends data and commodity ETFs does not provide significant enough profits to build a profitable investment strategy. Gold as an investment is also relatively stable, which is

reflected in moderate price changes. Thus, as an investment, gold is better for hedging and diversification.

6.2. Possible ideas for further research

Google Trends data offers numerous opportunities. As described in the theoretical background section, it has already been applied in many fields, such as the medical industry. Even if the Google Trends data has become more and more sufficient it still has its shortcomings. Hopefully, the trend of open data sources continue to develop and Google continues to further open the access to their databases also in the future. This would enable more short-term and long-term possibilities using diverse perspectives.

Based on this research it would be interesting to continue the study to test the relationship of general search words on a market index ETF. A good example would be how attention to the VIX index affects, for example, the price development of the iShares S&P 500 Index ETF (IVV). In the end, VIX index is also known as the so-called fear factor. Another perspective would be to measure how the increase in gold attention measured in search volumes performs relative to price changes of large indices.

Google is not the only company that has been opening their databases for public. The most common way nowadays is an open application programming interface (API), where external users have access to company data. Existing literature has already taken advantage of data by Wikipedia, YouTube, social media platforms as Twitter and other search machines. To name few other examples of corporations with open API's, Spotify, UPS, PayPal and Mastercard, all belong to this group. All these companies are major operators in their own fields and have massive databases. Imagine the information that could be obtained by researching people's credit card behavior and what it could potentially show about investor confidence.

Using different data sources together or comparing them offers another interesting approach. This would mean that data would be collected from multiple sources, such as Yahoo and Google. Also the use of company tickers or generic search words can still be expanded in different market areas as the volume of data increases. What comes to the

approach by Preis et al. (2013), it offers endless possibilities related to keyword selection. If conducting a research using generic search words, it would be interesting to modify the structure by first eliminating non-effective search words from the sample and then run new set of results to see how the possible effects change. Also, grouping statistically significant positive or negative words into new variables and testing the potential effects of these new variables would also provide an interesting perspective.

References

Alexa (2020) The top 500 sites on the web. [www.document] [Accessed 20.01.2020] Available: https://www.alexa.com/topsites

Alexeev, V. & Tapon, F. (2011) Testing weak form efficiency on the Toronto Stock Exchange. Journal of Empirical Finance, Vol. 18, pp. 661-691.

Aouadi, A., Arouri, M. & Teulon, F. (2013) Investor attention and stock market activity: Evidence from France. Economic Modelling, Vol. 35, pp. 674-681.

Askitas, N. & Zimmermann, K. (2009) Google Econometrics and Unemployment Forecasting. Applied Economics Quarterly, Vol. 55, pp. 107-120.

Baker, H.K. & Nofsinger, J.R. (2002) Psychological Biases of Investors. Financial Services Review, Vol. 11, pp. 97-116.

Baker, S. & Fradkin, A. (2017) The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data. The Review of Economics and Statistics, Vol. 99, pp. 756-768.

Barber, B.M. & Odean, T. (2001) Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment. The Quarterly Journal of Economics, Vol. 116, pp. 261-292.

Barber, M.B. & Odean, T. (2008) All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, Review of Financial Studies, Vol. 21, pp. 785-818.

Barberis, N. & Xiong, W. (2009) What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation. Journal of Finance, Vol. 64, pp. 751-784.

Bell, A. & Jones, K. (2015) Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. Political Science Research and Methods, Vol. 3, pp. 133-153.

Bijl, L., Kringhaug, G., Molnár, P. & Sandvik, E. (2016) Google searches and stock returns. International Review of Financial Analysis, Vol. 45, pp. 150-156.

Black, F. (1986) Noise. Journal of Finance, Vol. 41, pp. 529–543.

Blackrock. (2017). A primer on ETF primary trading and the role of authorized participants. [www.document] [Accessed 05.06.2020] Available: https://www.blackrock.com/corporate/en-at/literature/whitepaper/viewpoint-etf-primary-trading-role-of-authorized-participants-march-2017.pdf. [Accessed 10.2.2018]

Bodie, Z., Kane, A. & Marcus, A.J. (2005) Investments. 6th ed. Boston (MA): McGraw-Hill.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, Vol. 2, pp. 1-8.

Caparrelli, F., D'Arcangelis, A. & Cassuto, A. (2004) Herding in the Italian stock market: A case of behavioral finance. The Journal of Behavioral Finance, Vol. 5, pp. 222-230.

Challet, D. & Ayed, A.B.H. (2013) Predicting financial markets with Google Trends and not so random keywords. Cornell University, arXiv:1307.4643, pp. 1-9.

Charupat, N. & Miu, P. (2013) Recent developments in exchange-traded fund literature. Managerial Finance, Vol. 39, pp. 427-443.

Chemmanur, T. & Yan, A. (2019) Advertising, Attention, and Stock Returns. The Quarterly Journal of Finance, Vol. 9, pp. 1-51.

Choi H. & Varian H. (2009) Predicting the Present with Google Trends. Technical Report, Google Inc.

Choi H. & Varian H. (2011) Predicting the Present with Google Trends. Technical Report, Google Inc.

Cooper, C.P., Mallon, K.P., Leadbetter, S., Pollack, L.A., Peipins, L.A. & Jansen, J. (2005) Cancer Internet Search Activity on a Major Search Engine, United States 2001-2003. Journal of Medical Internet Research, Vol. 7, pp. 1-10.

Da, Z., Engelberg, J. & Gao, P. (2011) In Search of Attention. The Journal Of Finance, Vol. 66, pp. 1461-1499.

Daniel, K.D., Grinblatt, M., Titman, S. & Wermers, R. (1997) Measuring mutual fund performance with characteristic-based benchmarks. Journal of Finance, Vol. 52, pp. 1035-1058.

De Bondt, W.F.M. & Thaler, R. (1985) Does The Stock Market Overreact? The Journal of Finance, Vol. 40, pp. 793-805.

De Long, J., Shleifer, A., Summers, L. & Waldman, R. (1990) Noise trader risk in financial markets. The Journal of Political Economy, Vol. 98, pp. 703-738.

Delcoure, N. and Zhong, M. (2007) On the premiums of iShares. Journal of Empirical Finance, Vol. 14, pp. 168-195.

Deville, L. 2008. Exchange traded funds: History, trading and research. IDEAS Working Paper Series from RePEc, pp. 1-37.

Dimpfl, T. & Jank, S. (2016) Can Internet Search Queries Help to Predict Stock Market Volatility? European Financial Management, Vol. 22, pp. 171-192.

Edmans, A., Carcía, D. & Norli, Ø. (2007) Sport Sentiment and Stock Returns. The Journal of Finance, Vol. 62, pp. 1967–1998.

Ettredge, M., Gerdes, J. & Karuga, G. (2005) Using web-based search data to predict macroeconomic statistics. Communications of the ACM, Vol. 48, pp. 87-92.

Fakhry, B. (2016) A Literature Review of Behavioural Finance. Journal of Economics Library, Vol. 3, pp. 458-465.

Fama, E.F. (1970) 'Efficient Capital Markets : A Review of Theory and Empirical'. The Journal of Finance, Vol. 25, pp. 383-417.

Fama, E.F. & Macbeth, J.D. (1973) Risk return and equilibrium: Empirical tests. Journal of Political Economy, Vol. 81, pp. 607-636.

Frazzini, A. (2006) The Disposition Effect and Underreaction to News. Journal of Finance, Vol. 61, pp. 2017-2046.

French, K. & Poterba, J. (1991) Investor Diversification and International Equity Markets. The American Economic Review, Vol. 81, pp. 222-226.

Fuhr, D. & Kelly, S. (2009) Contrasting the development of ETF markets. Global Pensions, pp. 21.

Gastineau, G.L. (2001) Exchange-Traded Funds: An Introduction. The Journal of Portfolio Management, Vol. 27, pp. 88-96.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. & Brilliant, L. (2008) Detecting influenza epidemics using search engine query data. Nature, Vol. 457, pp. 1012-1015.

Google (2020) Google: Finance. [www.document] [Accessed 20.01.2020] Available: https://www.google.com

Google Trends (2020) Google Trends. [www.document] [Accessed 21.01.2020] Available: https://trends.google.com/trends/

Grinblatt, M., Titman, S. & Wermers, R. (1995) Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. The American Economic Review, Vol. 85, pp. 1088-1105.

Hirshleifer, D. (2001) Investor Psychology and Asset Pricing. Journal of Finance, Vol. 56, pp. 1533-1597.

Hirshleifer, D. & Shumway T. (2003) Good Day Sunshine: Stock Returns and the Weather. The Journal of Finance, Vol. 58, pp. 1009–1032.

Huang, H. & Penna, D. (2009) Constructing Consumer Sentiment Index for U.S. Using Google Searches. Working Papers 2009-26, University of Alberta, Department of Economics, pp. 1-20.

Huberman, G. (2001) Familiarity Breeds Investment. The Review of Financial Studies, Vol. 14, pp. 659-680.

Joo, B. & Durri, K. (2018) Comprehensive review of literature on behavioral finance. Indian Journal of Commerce & Management Studies, Vol. 6, pp. 11-19.

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, Vol. 27, pp. 1116-1127.

Juma'H, A. & Alnsour, Y. (2018) Using social media analytics: The effect of President Trump's tweets on companies' performance. Accounting and Management Information Systems, Vol. 17, pp. 100-121.

Jureviciene, D. & Jermakova, K. (2012) The Impact of Individuals' Financial Behaviour on Investment Decisions. Electronic International Interdisciplinary Conference, pp. 242-250.

Kahneman, D. & Tversky, A. (1979) Prospect Theory: An Analysis of Decision Under Risk. Econometrica, Vol. 47, pp. 263-291.

Keyword Tool (2020) Find Great Keywords Using Google Autocomplete. [www.document] [Accessed 27.03.2020] Available: https://keywordtool.io

Klemola, A. (2019) Essays on irrational investors' behavioral biases and pricing efficiency. Doctoral Dissertation. Faculty of the University of Vaasa. Available: http://urn.fi/URN:ISBN:978-952-476-886-3

Kreis, Y. & Licht, J.W. (2018) Trading on ETF mispricings. Managerial Finance, Vol. 44, pp. 357-373.

Kulkarni, G., Kannan, P. & Moe, W. (2012) Using online search data to forecast new product sales. Decision Support Systems, Vol. 52, pp. 604-611.

Kumar, S. & Goyal, N. (2015) Behavioural biases in investment decision making - a systematic literature review. Qualitative Research in Financial Markets, Vol. 7, pp. 88-108.

Kumar, S. & Goyal, N. (2016) Evidence on rationality and behavioural biases in investment decision making. Qualitative Research in Financial Markets, Vol. 8, pp. 270-287.

Lakonishok, J., Shleifer, A. & Vishny, R.W. (1992) The impact of institutional trading on stock prices. Journal of Financial Economics, Vol. 32, pp. 23-43.

Lauerman, J. (2020) Coronavirus Could Infect Two-Thirds of Globe, Research Shows. Bloomberg. [www.document] [Accessed 25.02.2020] Available: www.bloomberg.com/news/articles/2020-02-13/coronavirus-could-infect-two-thirds-of-globe-researcher-says

Lee, Y., Liu, Y., Roll, R. & Subrahmanyam, A. (2004) Order Imbalances and Market Efficiency: Evidence from the Taiwan Stock Exchange. Journal of Financial and Quantitative Analysis, Vol. 39, pp. 327-341.

Lo, A.W. & Mackinlay, A.C. (1988) Stock Market Prices do not Follow Random Walks: Evidence from a Simple Specification Test. The Review of Financial Studies, Vol. 1, pp. 41-66.

London Stock Exchange. (2014). ETFs for private investors. [www.document] [Accessed 05.06.2020] Available: https://www.lseg.com/sites/default/files/content/documents/ETFs%20for%20Private%20In

Lucey, B.M. & Dowling, M. (2005) The Role of Feelings in Investor Decision-Making. Journal of Economic Surveys, Vol. 19, pp. 211-237.

Maier, J. (2017) Exploring ETFs. Global X Management Company LLC. [www.document] [Accessed 16.03.2020] Available: https://content.globalxfunds.com/hubfs/eXploringETFs-4Q17.pdf?_ga=2.122884099.965975501.1584103382-1473235805.1584103382

Malkiel, B.G. & Rosti, J. (2007) A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing. Sattuman kauppaa Wall Streetillä. Helsinki, Talentum.

Moat, H.S., Curme, C., Avakian, A., Kenett, D.Y., Stanley, H.E. & Preis, T. (2013) Quantifying Wikipedia Usage Patterns Before Stock Market Moves. Scientific Reports, Vol. 3, pp. 1-5.

Morgan, K. (1997) Do noise traders influence stock prices? Journal of Money, Credit, and Banking, Vol. 29, pp. 351-363.

Nasdaq (2020) Indices. [www.document] [Accessed 01.02.2020] Available: http://www.nasdaqomxnordic.com/indeksit

Nadig, D. (2010) Understanding ETF Daily Data. ETF.com [www.document] [Accessed 30.03.2020] Available: https://www.etf.com/sections/blog/7766-understanding-etf-daily-data.html

Nofsinger, J.R. (2005) Social Mood and Financial Economics. The Journal of Behavioral Finance, Vol. 6, pp. 144–160.

Oberlo (2020) Search Engine Market Share in 2019. [www.document] [Accessed 21.01.2020] Available: https://www.oberlo.com/statistics/search-engine-market-share

Petajisto, A. (2017) Inefficiencies in the Pricing of Exchange-Traded Funds. Financial Analysts Journal, Vol. 73, pp. 24-54.

Polgreen, P.M., Chen, Y., Pennock, D.M. & Nelson, F.D. (2008) Using internet searches for influenza surveillance. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, Vol. 47, pp. 1443-1448.

Pompian, M.M. (2006) Behavioral Finance and Wealth Management. USA: John Wiley and Sons.

Preis, T., Moat, H.S. & Stanley, H.E. (2013) Quantifying Trading Behavior in Financial Markets Using Google Trends. Scientific Reports, Vol. 3, pp. 1-6.

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. Series A, Mathematical, physical, and engineering sciences, Vol. 368, pp. 5707-5719.

Radinsky, K., Davidovich, S. & Markovitch, S. (2008) Predicting the News of Tomorrow Using Patterns in Web Search Queries. 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, pp. 363-367.

Reinicke, C. (2019) The ETF market will hit $50 trillion by 2030, Bank of America says. Market Insider. [www.document] [Accessed 25.03.2020] Available: https://markets.businessinsider.com/news/stocks/etf-market-grow-50-trillion-assets-2030-bank-america-passive-2019-12-1028763048

Ricciardi, V. & Simon, H. (2000) What is behavioral finance? The Business, Education and Technology Journal, Vol. 2, pp. 26–34.

Roll, R. (1977) A critique of the asset pricing theory's tests Part I: On past and potential testability of the theory. Journal of Financial Economics, Vol. 4, pp. 129-176.

Schmeling, M. (2009) Investor sentiment and stock returns: Some international evidence. Journal of Empirical Finance, Vol. 16, pp. 394-408.

Schmidt, T. & Vosen, S. (2011) Forecasting private consumption: Survey-based indicators vs. Google trends. Journal of Forecasting, Vol. 30, pp. 565-578.

Seasholes, M.S. & Wu, G. (2007) Predictable behavior, profits, and attention. Journal of Empirical Finance, Vol. 14, pp. 590-610.

Shane, F. (2005) Cognitive Reflection and Decision Making. Journal of Economic Perspectives, Vol. 19, pp. 25-42.

Shefrin, H. & Statman M. (1984) Explaining Investor Preference for Cash Dividends. The Journal of Financial Economics, Vol. 13, pp. 253-282.

Shefrin, H. & Statman, M. (1985) The Disposition to Sell Winners Too Early and Ride Losers Too Long: Theory and Evidence. Journal of Finance, Vol. 40, pp. 777-790.

Shiller, R.J. (2005) Irrational exuberance. Princeton University Press, New Jersey.

Shiller, R.J. (2006) Tools for Financial Innovation: Neoclassical Versus Behavioral Finance. The Financial Review, Vol. 41, pp. 1-9.

Shleifer, A. & Summers, L.H. (1990) The noise trader approach to finance. Journal of economic perspectives, Vol.. 4, pp. 19-33.

Sornette, D. (2003) Why stock markets crash: Critical events in complex financial systems. Princeton (NJ): Princeton University Press.

Srivastava, S. (2012) Rational Investment Decisions of Irrational Household Investors. Global Research Analysis, Vol. 1, pp. 59-61.

Statista (2018) Number of search engine user in the United States from 2014 to 2020 (in millions) [www.document] [Accessed 21.01.2020] Available: https://www.statista.com/statistics/253795/number-of-search-engine-users-in-the-united- states/

Statista (2020) Worldwide ETF assets under management 2003-2019. [www.document] [Accessed 13.03.2020] Available: https://www.statista.com/statistics/224579/worldwide-etf-assets-under-management-since-1997/

Tesar, L.L. & Werner, I.M. (1995) Home bias and high turnover. Journal of International Money and Finance, Vol. 14, pp. 467-492.

Tetlock, P.C. (2007) Giving Content to Investor Sentiment: The Role of Media in the Stock Market. Journal of Finance, Vol. 62, pp. 1139-1168.

The Economist (2011) US Confidence indicators. [www.document] [Accessed 14.04.2020] Available: http://www.economist.com/blogs/dailychart/2011/08/us-confidence-indicators

Thomson Reuters (2017) DataStream Global Equity Indices – User Guide. [www.document] [Accessed 31.03.2020] Available: http://www.datastream.jp/wp/wp-content/uploads/2017/02/DatastreamGlobalEquityIndicesUGissue05.pdf

Tokic, D. (2007) Semiconductor stocks: Signaling a bear market? Journal of Corporate Accounting & Finance, Vol. 18, pp. 3-7.

Torres-Reyna, O. (2007) Panel Data Analysis Fixed and Random Effects using Stata. Data & Statistical Services, Princeton University. [www.document] [Accessed 08.06.2020] Available: http://dss.princeton.edu/training/

Vanguard. (2013). Understanding synthetic ETFs. [www.document] [Accessed 05.06.2020] Available: https://pressroom.vanguard.com/content/nonindexed/6.14.2013_Understanding_Synthetic_

Varian H.R. (2014) Big Data: New Tricks for Econometrics. Journal of Economic Perspectives. Vol. 28, pp. 3-28.

Wu, L. & Brynjolfsson, E. (2009) The future of prediction: How google searches foreshadow housing prices and quantities. 30th International Conference on Information Systems, pp. 1-14.

Wuoristo, L. (2012) Can you google the future? A study on the predictive power of Google Trends on company shares in the UK. Aalto University, Department of Finance. Available: http://urn.fi/URN:NBN:fi:aalto-201301211213

Zhou, R. & Lai, R. (2009) Herding and information based trading. The Journal of Empirical Finance, Vol. 16, pp. 388-393.

Appendix

Appendix 1. The list of ETFs after financial data processing

Ticker:	***Name:***	***Inception date***	***Net asset***
GLD	*SPDR Gold Shares*	*2004-11-18*	*45,99B*
IAU	*iShares Gold Trust*	*2005-01-21*	*19,55B*
GDX	*VanEck Vectors Gold Miners ETF*	*2006-05-16*	*11,67B*
SLV	*iShares Silver Trust*	*2006-04-21*	*6,32B*
GDXJ	*VanEck Vectors Junior Gold Miners ETF*	*2009-11-10*	*4,42B*
SGOL	*Aberdeen Standard Physical Swiss Gold Shares ETF*	*2009-09-09*	*1,5B*
PPLT	*Aberdeen Standard Platinum Shares ETF*	*2010-01-08*	*694,99M*
GLTR	*Aberdeen Standard Physical Precious Metals Basket Shares ETF*	*2010-10-22*	*584,36M*
PALL	*Aberdeen Standard Physical Palladium Shares ETF*	*2010-01-08*	*460,07M*
SIVR	*Aberdeen Standard Physical Silver Shares ETF*	*2009-07-24*	*415,25M*
SIL	*Global X Silver Miners ETF*	*2010-04-19*	*472,39M*
RING	*iShares MSCI Global Gold Miners ETF*	*2012-01-31*	*381,97M*
OUNZ	*Van Eck Merk Gold Trust*	*2014-05-16*	*208,71M*
SGDM	*Sprott Gold Miners ETF*	*2014-07-14*	*174,23M*
UGL	*ProShares Ultra Gold*	*2008-12-01*	*136,28M*
DBP	*Invesco DB Precious Metals Fund*	*2007-01-05*	*133,51M*
SILJ	*ETFMG Prime Junior Silver Miners ETF*	*2012-11-28*	*131,22M*
DGL	*Invesco DB Gold Fund*	*2007-01-05*	*129,78M*
SLVP	*iShares MSCI Global Silver and Metals Miners ETF*	*2012-01-31*	*85,84M*
GOEX	*Global X Gold Explorers ETF*	*2010-11-03*	*38,28M*
GLL	*ProShares UltraShort Gold*	*2008-12-01*	*17,31M*
DBS	*Invesco DB Silver Fund*	*2007-01-05*	*14,09M*

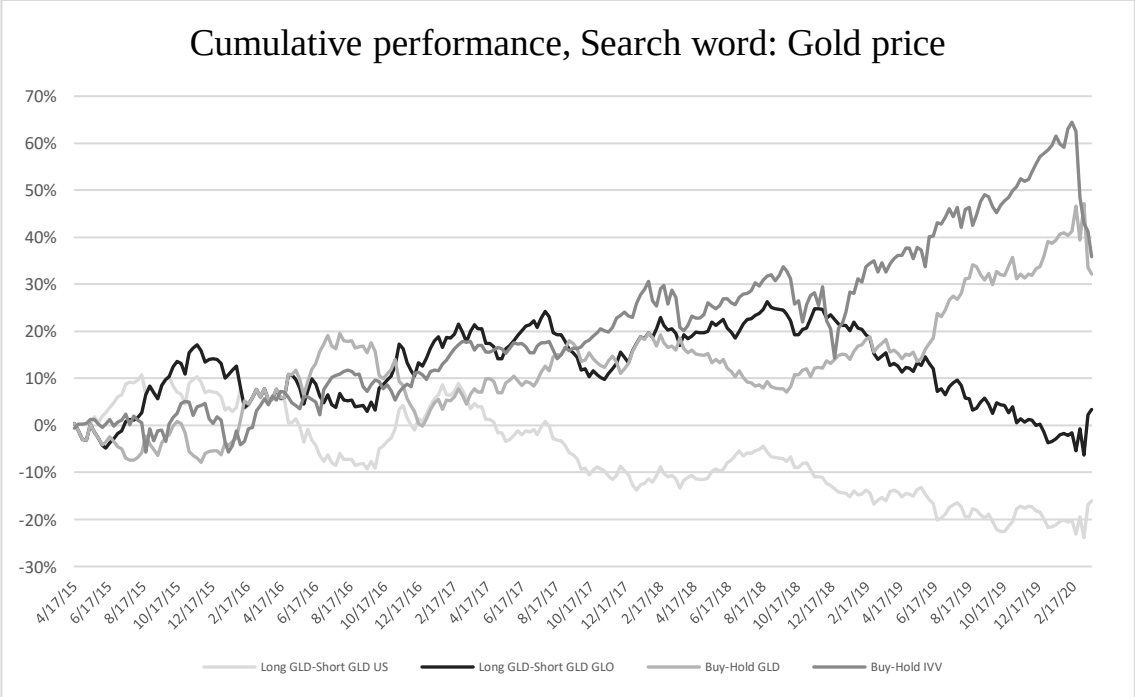
Appendix 2. The original list of search words chosen to second dataset (20 pcs)

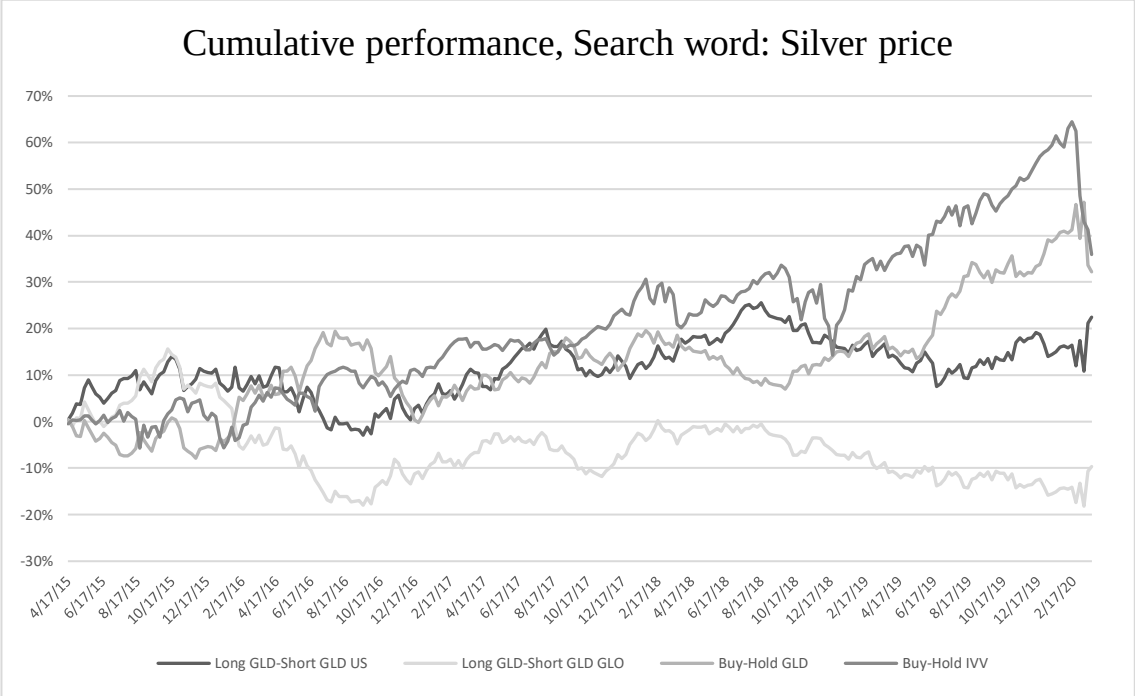
Gold	*Gold price*
Gold ETF	*Silver*
Silver price	*Silver ETF*
Precious metals	*S&P 500*
VIX	*Safe haven*
Stock market	*Financial markets*
ETF	*Commodities*
iShares ETF	*Vanguard ETF*
SPDR ETF	*VanEck ETF*
NYSE	*NYSEArca*

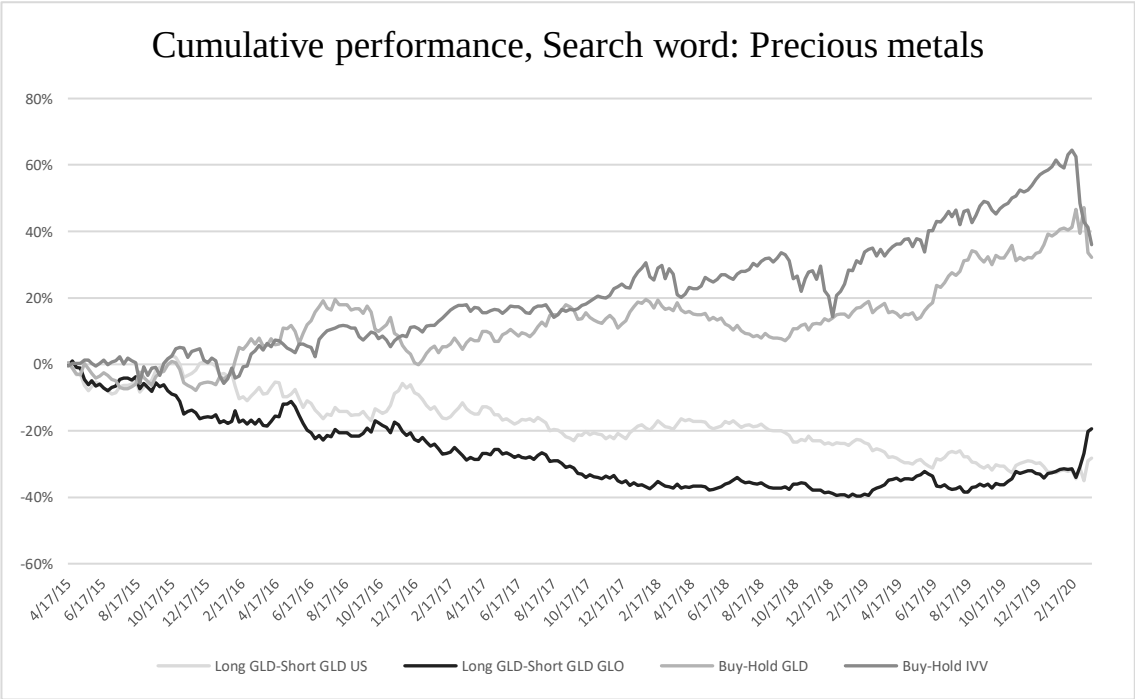
Appendix 3. Final sample

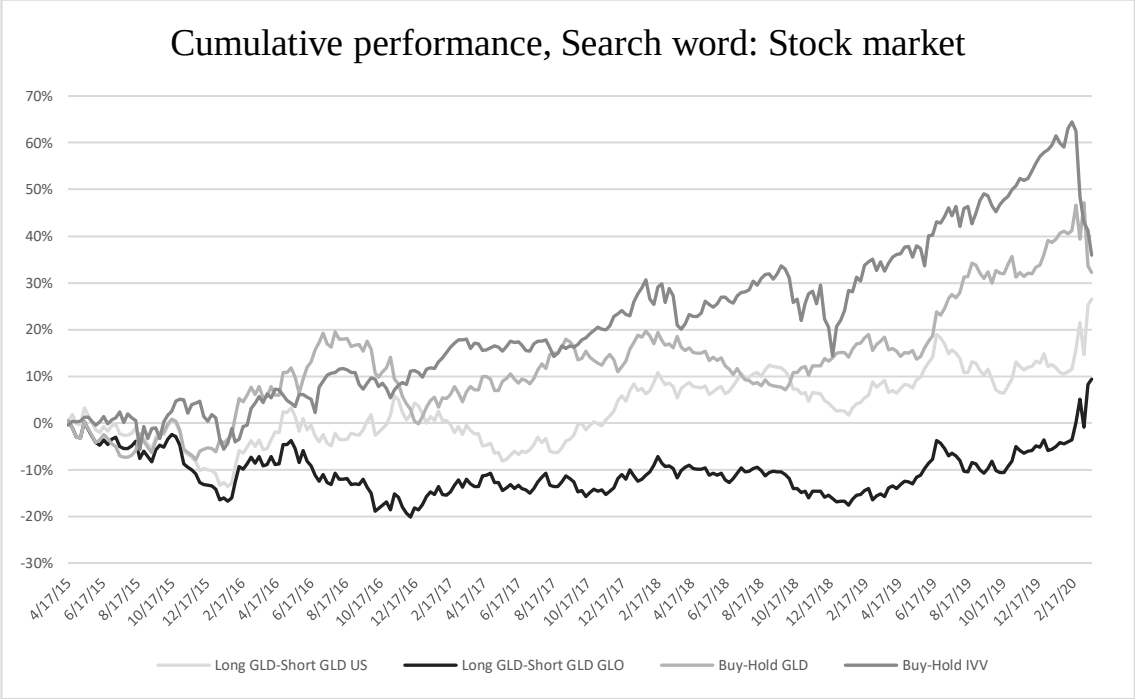
Financial data		*Search query data*	
GLD	*SPDR Gold Shares*	*Gold price*	*Financial markets*
IAU	*iShares Gold Trust*	*Gold ETF*	*Commodities*
SLV	*iShares Silver Trust*	*Silver price*	*ETF*
GDX	*VanEck Vectors Gold Miners ETF*	*Precious metals*	*iShares ETF*
GDXJ	*VanEck Vectors Junior Gold Miners ETF*	*S&P 500*	*Vanguard ETF*
		VIX	*SPDR ETF*
		Stock market	*NYSE*
		GLD	*IAU*
		SLV	*GDX*
		GDXJ	

Appendix 4. Graphical presentation for search terms Gold price, Silver price, Precious metals and Stock market




Cumulative performance, Search word: Silver price
70%
60%
50%
40%
30%
20%
10%
0%
-10%
-20%
-30%
4/17/15
6/17/15
8/17/15
10/17/15
12/17/15
2/17/16
4/17/16
6/17/16
8/17/16
10/17/16
12/17/16
2/17/17
4/17/17
6/17/17
8/17/17
10/17/17
12/17/17
2/17/18
4/17/18
6/17/18
8/17/18
10/17/18
12/17/18
2/17/19
4/17/19
6/17/19
8/17/19
10/17/19
12/17/19
2/17/20
Long GLD-Short GLD US
Long GLD-Short GLD GLO
Buy-Hold GLD
Buy-Hold IVV


Cumulative performance, Search word: Precious metals
80%
60%
40%
20%
0%
-20%
-40%
-60%
4/17/15
6/17/15
8/17/15
10/17/15
12/17/15
2/17/16
4/17/16
6/17/16
8/17/16
10/17/16
12/17/16
2/17/17
4/17/17
6/17/17
8/17/17
10/17/17
12/17/17
2/17/18
4/17/18
6/17/18
8/17/18
10/17/18
12/17/18
2/17/19
4/17/19
6/17/19
8/17/19
10/17/19
12/17/19
2/17/20
Long GLD-Short GLD US
Long GLD-Short GLD GLO
Buy-Hold GLD
Buy-Hold IVV


Cumulative performance, Search word: Stock market
70%
60%
50%
40%
30%
20%
10%
0%
-10%
-20%
-30%
4/17/15
6/17/15
8/17/15
10/17/15
12/17/15
2/17/16
4/17/16
6/17/16
8/17/16
10/17/16
12/17/16
2/17/17
4/17/17
6/17/17
8/17/17
10/17/17
12/17/17
2/17/18
4/17/18
6/17/18
8/17/18
10/17/18
12/17/18
2/17/19
4/17/19
6/17/19
8/17/19
10/17/19
12/17/19
2/17/20
Long GLD-Short GLD US
Long GLD-Short GLD GLO
Buy-Hold GLD
Buy-Hold IVV