2020

Google Search Sentiment and Sector ETF Performance

Andrew J. Williamson
Colby College

Google Search Sentiment

And Sector ETF Performance

Drew Williamson

Colby College

Economics Department

Abstract

This paper utilizes data from Google searches in an attempt to utilize online investor sentiment as a predictor of sector exchange traded fund (ETF) performance. The paper tests the assumptions of the Efficient Market Hypothesis that all known information about a stock is incorporated into the price of the stock. With the emergence of ETFs as a popular form of investment for casual investors, there is a possibility that these investors may use Google as a way to collect information about potential stock picks. Thus, this paper investigates the association between online search interest and excess ETF returns by collecting data using Google's Trends search functionality to calculate investor sentiment for sector ETFs over a five-year time span. Empirical results from this paper suggest that Google search interest has no association with excess returns, supporting the theory associated with the Efficient Market Hypothesis.

Acknowledgments

I would like to extend a sincere thank you to my advisor Professor Randy Nelson for his guidance throughout my final year at Colby College. His assistance with all my questions was extremely helpful throughout the thesis writing process. I struggled with data and methodology issues throughout the year and he was always willing to meet and walk through the challenges that I faced.

In addition, I would like to thank Professor Lindsey Novak for her guidance on the research writing process throughout the year. Her passion for research and her students was clear as she provided extremely useful guidance throughout the entire research process.

Lastly, I would like to thank all of the professors that challenged me throughout my time at Colby and inspired me to think critically about all problems that I faced.

I. Introduction

Most investors utilize the stock market as a way to diversify their investment portfolios and to provide relatively consistent returns over time. With the invention of the Internet, investors are increasingly exposed to more information about companies and their performance. Many investors argue over different investment strategies and pricing theories that provide the best way to model stock movement and pricing. A common pricing model, the Capital Asset Pricing Model (CAPM) suggests that the expected returns of a stock are based on the stock's riskiness relative to the stock market as a whole. The CAPM explains systematic, or 'un-diversifiable' risk associated with specific stocks. However, as more and more investors are able to access information about stocks through Google and alternative forms of media, stocks may begin to be susceptible to price changes based on new information found from online investment research.

In addition to the CAPM model, the Efficient Market Hypothesis is another common theory that states that consistent alpha generation is an impossible endeavor. In the context of investing, alpha is the performance of an investment in comparison to a market index that represents the market as a whole. The Efficient Market Hypothesis essentially states that share prices for any stock reflect all known information about the stock. Thus, attempting to purchase or sell an undervalued or overvalued stock is not feasible. If the Efficient Market Hypothesis is true, investors should only passively invest in the stock market and make no attempt at actively managing a stock-based asset portfolio. This indicates that generating excess return through active management trading strategies is not a lucrative endeavor.

A new area of research involves attempts to understand how the information investors find online impacts their investment decisions. There is potential to analyze online search patterns to investigate if they have predictive power for asset returns. Newfound insight into how Google search interest and search sentiment could provide an additional way to understand asset price movement and insight into how to further evaluate systematic risks associated with investing in the stock market. Leveraging Google search data through Google Trends could provide some additional information about current investor sentiment about a particular stock or collection of stocks. If an investor can utilize patterns within Google Trends data to predict stock movement or develop trading strategies based on this data, hypotheses such as the Efficient Market Hypothesis and models such as the CAPM would be tested.

This paper addresses the relationship between Google searches on Exchange Traded Fund (ETF) performance across different industry sectors over the last five years. An exchange-traded fund is a type of security that includes a collection of securities that typically tracks an underlying index. Typically an ETF consists of multiple different types of stocks, bonds or a combination of the two. Specifically, I investigate how Google search sentiment corresponds to asset price movements in Vanguard sector ETFs. I choose to investigate Vanguard sector ETFs as these ETFs consist of only stock equity assets. Using Google's Google Trends database, I investigate the historical search interest and sentiment for a variety of stocks within sector ETFs to evaluate the relationship of search sentiment on ETF returns. I use data on some of Vanguard's sector ETFs and corresponding Google Trends data on the search sentiment among the top stocks within sector ETFs. To test the Efficient Market Hypothesis, I perform empirical research on the predictive power of online search sentiment on ETF performance compared to the overall stock market as a whole. Understanding how Google search sentiment influences asset returns in comparison to the overall market would provide insight into new ways investors could model and predict potential asset price movements. Such strategies would mitigate investment risks and provide opportunities to create trading strategies leveraging Google Trends data to generate consistent excess returns over time.

My contribution to existing literature is three-fold. First, this research will provide insight into asset pricing theories, such as the CAPM and how investors can better understand systematic risks associated with investors finding readily available information on stocks based on Google searches. Second, the research may provide insight into potential trading patterns and strategies that may be useful for institutional investors. Third, my research will investigate how search sentiment impacts different economic sectors, allowing an investigation into varying correlations between price movement and search sentiment across different asset sectors.

Next, I will discuss with more detail the prior research on pricing theories and how online information impacts asset behavior and trading patterns. After, I will discuss the data involved in my research and my empirical regression model that produced my results. To conclude, I will discuss my main findings and the further research that can be done to further understanding of asset pricing and its association with Internet search interest and sentiment.

II. Literature Review

Predicting stock price returns is a topic widely researched and the topic continues to be essential to finance and investor theory. However, research shows that predicting stock returns based on current information is a difficult task. Research even shows that common stock price forecasting metrics such as short rates, term spreads, and dividend or earnings to price ratios provide minimal predictive power for stocks traded on international markets (Hjalmarsson, Erik 2005). Even traditional models, such as the Capital Asset Pricing Model (CAPM), do not sufficiently explain expected returns of an asset (Fama and French 1996). More recent research suggests that asset price predictability relies on predicting asset volatility. Evidence suggests that if one can effectively predict the sign of a stock's future volatility, then asset returns are predictable (Christoffersen and Diebold 2006). Thus, most new research relies on the idea that indicators that may provide insight into when unpredictable volatility may occur can identify unpredictable stock returns.

A lot of research has investigated the association between online activity and corresponding stock price movements in domestic and international markets. Specifically, research using Twitter activity, or more importantly, Google search activity is a new area of study for many finance researchers. The majority of existing research investigates the association between online interest in a stock and its corresponding behavior. For example, research on how social media activity influences stock trading volume on thirty stocks in the Dow Jones Industrial Average Index provides insight into how online activity relates to large-capitalization stock behavior. Results from research analyzing Twitter-Tweet sentiment shows significant correlations between abnormal stock returns and Twitter activity during peak Tweet volume (Ranco et al. 2015).

Research has also shown significant associations between Google search interest and stock behavior in international stock exchanges. The majority of research on foreign markets; however, suggests that Google Trends data and search interest is only associated with increases in a stock's trading volume. A study on stocks traded on the Japanese exchange attempted to use Google Trends data to predict future stock prices. The results suggest that higher Google search volumes are correlated with increases in trading volume (Takeda and Wakao 2014). Takeda and Wakao conclude that they do not believe increases in search interest cause increases in stock prices. A similar study was conducted on stocks traded on Norway's Oslo Stock Exchange. The results are consistent with prior research, finding little evidence for Google search data providing price-predicting power and strong evidence that increases in Google search interest is associated with increased trading volume (Kim et al. 2019). Similar research was done in Germany, finding the same results on stocks traded on German markets (Bank, Larch, and Peter 2011). Bank, Larch, and Peter speculate that search volume measures mainly the interest of non-institutional investors and likely has minimal impacts on future stock returns. In similar studies, other researchers conclude that data from Google Trends provides no additional information about future stock performance, rather it is roughly equivalent to current price return information (Challet and Ayed 2014).

In contrast, other research suggests significant correlations between Google Trends data and corresponding stock returns. Some studies use Google Trends to create a measure of investor sentiment. Research using Google Trends to create a measure of French investor sentiment reveals some evidence of an association between Google search sentiment and short-term predictability of stock price (Beer, Hervé, and Zouaoui 2013). Similar research suggests Google Trends data can provide investors with insight into the current state of the economy (Preis, Moat, and Stanley 2013). Some research finds strong evidence that surges in investor interest is associated with same-day abnormal returns (Tang and Zhu 2017). However, Tang and Zhu find that these abnormal returns disappear or reverse after they are discovered.

Additionally, Google Trends data can provide investors with potential information about warning signs in the stock market, suggesting that increases in search interest occur prior to stock market declines (Preis, Moat, and Stanley 2013). Using these results, other research suggests a trading strategy of buying stocks with relatively low Google search interest and selling stocks with high Google search interest (Bijl et al. 2016). This strategy proved to be profitable; however, taking into account transaction costs renders the trading strategy ineffective. Interestingly, some research uses the results of research from Christoffersen and Diebold, and suggests that investors can create profitable investment strategies using Google Trends data to predict future volatilities (Chronopoulos, Papadimitriou, and Vlastakis 2018).

Such contrasting research results suggest a need for further investigation of the use of Google Trends as a way to predict asset price. Next, I will discuss my data and how I use Google Trends data to capture investor sentiment to predict sector ETF price movement.

III. Data

I collect data from three main sources: Yahoo Finance historical databases, Google Trends, and the Federal Reserve Bank of St. Louis. The dataset is a weekly time series dataset, spanning over the last five years from October 5th 2014 through September 15th 2019. I choose to analyze Vanguard sector Exchange Traded Funds (ETF) to analyze the effect of Google search sentiment on sector ETF performance. These sector ETFs include the following sectors: consumer discretionary, consumer staples, financials, information technology, communication services, energy, health care, industrials, materials, real estate, and utilities. Table 1 indicates which sector ETFs are included in the final dataset. Due to Google Trends data limitations, calculations of sentiment score were not robust for the energy, health care, industrials, materials, real estate, and utilities sectors. Many of the companies held within these ETFs tend to have less interest in regards to the number of Google Trends searches to buy or sell the stock. Thus, using my methodology, many of the companies have minimal or no related Google Trends data. In addition to Vanguard sector ETFs, an SPDR Gold ETF is included in the dataset to analyze the predictive power of Google search sentiment on a commodity-tracking ETF. The SPDR Gold ETF tracks the price of the gold bullion less the trust's expenses. Vanguard sector ETF historical data and SPDR Gold ETF historical data comes from Yahoo Finance databases and other control variables such as the risk free rate of return come from the Federal Reserve Bank of St. Louis. In this research, the rate on the three-month United States Treasury bill is used for the risk free rate of return.

Vanguard Sector ETFs

Name	Ticker
Communication Services	VOX*
Consumer Discretionary	VCR*
Consumer Staples	VDC*
Energy	VDE
Financials	VFH*
Health Care	VHT
Industrials	VIS
Information Technology	VGT*
Materials	VAW
Real Estate	VNQ
Utilities	VPU

* (ETFs included in dataset)

Table 1

In addition, sector search sentiment and individual stock searches data were collected through Google Trends. Google Trends tracks online search interest over time. In order to capture search sentiment for a sector ETF, I collect individual stock search sentiment for at least the top 50 percent of weighted assets within each sector ETF.

Figure 2 below shows the top weighted stocks within each sector ETF and their corresponding tickers.

VOX
10 Largest holdings

Company Name	Ticker	Weight
Alphabet inc.	GOOG	22.70%
Facebook inc.	FB	15.30%
Verizon	VZ	7.10%
AT&T	T	4.70%
Walt Disney	DIS	4.70%
Comcast	CMCSA	4.40%
Netflix	NFLX	4.00%
Charter Communications	CHTR	2.50%
Activision Blizzard	ATVI	1.50%
T-Mobile	TMUS	1.40%
Total		68.30%

VDC
10 Largest holdings

Company Name	Ticker	Weight
Procter & Gamble	PG	14.30%
Coca Cola	KO	10.00%
PepsiCo	PEP	8.40%
Walmart	WMT	7.40%
Philp Morris International	PM	5.90%
Costco Wholesale	COST	4.90%
Mondelez International	MDLZ	4.20%
Altria Group	MO	4.00%
Colgate-Palmolive Co.	CL	3.10%
Kimberly-Clark Co.	KMB	2.50%
Total		64.70%

VFH
Largest holdings

Company Name	Ticker	Weight
JPMorgan	JPM	9.58%
Bank of America	BAC	7.10%
Berkshire Hathaway Class B	BRK.B	6.76%
Wells Fargo	WFC	5.28%
Citigroup	C	4.22%
American Express	AXP	2.25%
Us Bancorp	USB	2.19%
Goldman Sachs	GS	1.84%
Chubb Ltd.	CB	1.77%
CME Group Inc.	CME	1.75%
PNC Fincancial Services	PNC	1.65%
S&P Global Inc.	SPGI	1.54%
Morgan Stanley	MS	1.44%
BlackRock Inc.	BLK	1.40%
Charles Schwab	SCHW	1.31%
Total		50.07%

VCR
10 Largest holdings

Company Name	Ticker	Weight
Amazon	AMZN	21.80%
Home Depot	HD	7.70%
McDonald's	MCD	5.10%
Starbucks	SBUX	3.80%
Nike	NKE	3.50%
Booking Holdings	BKNG	2.70%
Lowes Cos. Inc.	LOW	2.60%
TJX Cos. Inc.	TJX	2.20%
General Motors	GM	1.60%
Target Corp.	TGT	1.40%
Total		52.40%

VGT
10 Largest holdings

Company Name	Ticker	Weight
Apple Inc.	AAPL	15.60%
Microsoft Corp.	MSFT	15.40%
Visa Inc.	V	4.50%
Mastercard Inc.	MA	3.90%
Cisco Systems Inc.	CSCO	3.80%
Intel Corp.	INTC	3.50%
Adobe Inc.	ADBE	2.30%
Oracle Corp.	ORCL	2.20%
International Business Machines (	IBM	2.00%
PayPal Holdings Inc.	PYPL	1.90%
Total		55.10%

Figure 1: ETF Holdings

A specific Google Trends search methodology was used to capture both bullish and bearish search sentiment along with individual stock tickers, or company name if needed. I used a search of "Buy [ticker] stock" to capture bullish stock search sentiment and "Sell [ticker] stock" to capture bearish search sentiment. For example, one of the Google Trends searches to obtain data for the communications sector is: "buy FB stock" and "sell FB stock" to capture both bearish and bullish search sentiment for Facebook stock over the last five years. Due to Google Trends data restrictions, this search methodology often resulted in a data restriction error. If individual searches do not occur with sufficient interest, Google Trends does not report the search data. Thus, for some searches, a search of "Buy [company name] stock" or "Sell [company name] stock" was used to capture either bearish or bullish search sentiment. I utilize the search methodology involving the company's ticker first in order to ensure that the majority of search results in Google Trends pertain to the company's performance on publicly traded markets, and to avoid any arbitrary search results due to people searching for any products the company may buy or sell. Due to Google Trends data limitations, search sentiment data was not available for the energy, health care, industrials, materials, real estate, and utilities sector ETFs.

By using individual stock search sentiments, I find both positive (bullish) and negative (bearish) search sentiment for each sector. By multiplying each stock's weight within the sector ETF by the sector's positive search interest, I am able to calculate a sector's overall weighted bullish search sentiment for each week. The same follows for the total negative search sentiment. I keep the positive and negative search sentiment separate in the calculation of sentiment and in my empirical model, as some investing theory states that investors react more strongly to negative news about a stock (Tversky and Kahneman 1992). For this reason, I do not use a net sentiment index in my model. I hypothesize that positive search sentiment scores indicate that investors are more likely to purchase the stock, thus driving the overall price for the sector ETF up. The same theory follows for negative search sentiments. Below is a table of summary statistics for all positive and negative search sentiments in addition to information about sector ETF average returns and trading volume.

Summary Statistics		
VARIABLES	Mean	Standard Error
VCR Positive Sentiment	10.943	0.28
VCR Negative Sentiment	9.793	0.32
VCR Weekly Returns (%)	0.153	0.12
VCR Weekly Excess Return (%)	0.002	0.056
VCR Trading Volume	516,970.270	22753.17
VDC Positive Sentiment	13.393	0.30
VDC Negative Sentiment	12.509	0.25
VDC Weekly Returns (%)	0.120	0.10
VDC Weekly Excess Return (%)	-0.031	0.008
VDC Trading Volume	646,998.842	22163.95
VFH Positive Sentiment	9.660	0.19
VFH Negative Sentiment	7.270	0.19
VFH Weekly Returns (%)	0.157	0.14
VFH Weekly Excess Return (%)	0.005	0.076
VFH Trading Volume	3114736.680	118194.16
VGT Positive Sentiment	15.455	0.29
VGT Negative Sentiment	13.007	0.34
VGT Weekly Returns (%)	0.278	0.14
VGT Weekly Excess Return (%)	0.126	0.063
VGT Trading Volume	2381508.494	77424.20
VOX Positive Sentiment	15.643	0.42
VOX Negative Sentiment	12.623	0.42
VOX Weekly Returns (%)	-0.008	0.13
VOX Weekly Excess Return (%)	-0.160	0.090
VOX Trading Volume	732,133.591	39074.23
GOLD Postive Sentiment	65.656	0.56
GOLD Negative Sentiment	38.328	0.40
GOLD Weekly Return (%)	0.118	0.12
GOLD Weekly Excess Return (%)	-0.033	0.164
GOLD Trading Volume	39234123.166	936103.64
Observations	259	259

In addition to the sentiments, risk free rate (3 month Treasury Bill) is included in the dataset for control purposes in regressions.

Next, I will discuss the empirical model I used to analyze my data to produce my results.

IV. Methodology

I use an ordinary least squares regression methodology to estimate my results. The positive and negative sector ETF sentiments, and a one-week lagged value of sector ETF sentiment, for each sector are the main explanatory variables. The response variable in this model is excess returns for the ETF. Weekly returns for each ETF are calculated as the percent change from weekly ETF opening price to the closing price of the ETF at the end of the trading week. Excess returns are then calculated by subtracting the S&P500 weekly returns from each sector's weekly returns. This is the main response variable as it will provide information on whether positive or negative sentiment can provide any predictive power on whether the ETF outperforms the market in any particular trading week. The primary equation I estimate is as follows:

$$ETFexcessreturns_{ct} = \beta_0 + \beta_1 positivesentiment_{c,t} + \beta_2 negative\ sentiment_{c,t} + \beta_3 positivesentiment_{c,t-1} + \beta_4 negativesentiment_{c,t-1} + \beta_5 riskfreerate_t + \varepsilon$$

In this equation, the idea is that the sentiment score for each sector, c, may affect the ETF's returns in the same week or potentially returns in the next week. The risk free rate of return is included as a control variable since it is the rate of return in which investors expect from an investment with zero risk. In my model, the 3-month Treasury Bill is used as the United States government has never defaulted and the 3 month is the only real asset with minimal to nearly zero interest rate or inflation risk. I use robust standard errors to account for heteroskedasticty without affecting the regression estimations. The results of the regression are reported in the tables of the results section. In addition to using sector ETFs I also run the same regressions on a SPDR gold ETF that tracks the underlying price of gold in attempts to understand how Google search sentiment affects the price of a popular commodity often viewed as an alternative asset to stocks. The results from these regressions are also reported in the results section below.

In addition, to the model above, I also run a model utilizing the percent change in investor sentiment to estimate the association of momentum behind changes in positive or negative search sentiment. The results of this regression are reported in column two of my regression table.

Despite my best effort, my dataset is not entirely ideal. There are many factors that could contribute to creating an ideal dataset. First, Google Trends data is limited and I was only able to construct a sector sentiment score for five out of the eleven Vanguard sector ETFs. If Google Trends provided data for my search methodology for all stocks within each sector ETF, my dataset would provide more information about how all sectors are affected by Google search sentiment. Second, my calculations of sector sentiment scores are not fully robust and do not capture the sentiment for all stocks held in each sector ETF. For example, most sentiment scores capture Google search sentiment for only the top 50 percent weighted stocks in each ETF. An ideal dataset and calculation of sentiment score would include search results from all stocks within all sectors. Third, my dataset would be significantly better if Google Trends could specify searches specifically related to stock information. For instance, most stock tickers are common abbreviations and it is possible that some Google searches using the methodology I used could provide results that do not influence the searcher to buy or sell the stock. In addition, there are likely other search terms that investors use to get information about stocks that are not captured by my dataset. While creating a more robust calculation of sentiment scores is possible, it likely would involve massive amounts of data collection and would likely complicate the calculation of the sentiment score. Fourth, while it is necessary to separate positive and negative search sentiment in the dataset, it does not account for the relative magnitudes of the two values. For example, if negative search sentiment is relatively large, an increase in positive search sentiment may not necessarily impact excess returns. Finally, Google Trends data reports zero search interest for a term when there is not enough interest to make reporting the data a worthwhile endeavor. This could occur if only searches for a term occur, but the search interest reported as zero. Thus, there is likely measurement error in my calculations of sentiment score when Google Trends reports a search interest score of zero.

In addition, there is always concern of reverse causality in my regression results. There is a possibility that increased Google searches may be a result of changes in underlying stock performance. Thus, it is possible that increased ETF returns may affect the number of Google searches in the same week. However, it is unlikely that reverse causality occurs for the lagged values of sentiment score. These results are consistent to existing literature as well. While some prior literature suggests that some Google Trends data can be utilized to actively manage stocks to generate positive returns, it also suggests that these positive returns are often offset by transaction costs (Bijl et al. 2016). The results of previous literature also investigate whether Google Trends data can be used to generate positive returns; whereas, I investigate if the data can be leveraged to produce excess returns which is a more difficult task.

V. Results

The results of my regressions for all five sector ETFs and the SPDR Gold-tracking ETF are shown in the tables below. For the most part, my results are not what I expect. Contrary to my original hypothesis that Google Trends data may provide investors with information on how to create profitable trading strategies in which they can actively produce excess returns, the results of my model find no significant correlation between investor search sentiment and excess returns. This suggests that the majority of information that investors may receive from Google searches includes all current pricing information. If this is the case, investors are not able to utilize Google search results in general to leverage trading strategies that produce excess return. This suggests that market equilibrium agrees with the Efficient Market Hypothesis. Despite having no significant correlations with investor search sentiment and excess returns, the results are consistent with existing theory.

These results also potentially highlight discrepancies between institutional and casual investors. In many cases, large institutional investors will likely utilize Bloomberg terminals to gather real-time data on stocks to incorporate into their stock pricing models. Whereas, casual investors attempting to utilize Google search results for pricing information may only have access to data that is dated in the form of news articles or buy or sell suggestions. Bloomberg terminals are costly and expensive to use over long periods of time making them largely inaccessible to casual investors. This suggests that institutional investors may often have better access to up-to-date information over the casual investor attempting to utilize Google Trends data to create profitable trading strategies.

Consumer Disc.	(1)	(2)
VARIABLES	Excess Return	Excess Return
Pos. Sentiment (t)	0.023	
	(0.017)	
Neg. Sentiment (t)	0.007	
	(0.014)	
Pos. Sentiment (t-1)	0.018	
	(0.016)	
Neg. Sentiment (t-1)	-0.031*	
	(0.016)	
Risk Free Rate	-0.042	-0.011
	(0.076)	(0.071)
% Change Pos. Sent.		0.001
		(0.063)
% Change Neg. Sent.		0.038
		(0.038)
Constant	-0.166	0.006
	(0.175)	(0.091)
Observations	258	258
R-squared	0.030	0.002

Consumer Staples	(1)	(2)
VARIABLES	Excess Return	Excess Return
Pos. Sentiment (t)	0.032	
	(0.021)	
Neg. Sentiment (t)	-0.012	
	(0.021)	
Pos. Sentiment (t-1)	0.009	
	(0.019)	
Neg. Sentiment (t-1)	-0.016	
	(0.021)	
Risk Free Rate	0.037	0.013
	(0.098)	(0.099)
% Change Pos. Sent.		-0.057
		(0.099)
% Change Neg. Sent.		0.141
		(0.143)
Constant	-0.277	-0.062
	(0.380)	(0.131)
Observations	258	258
R-squared	0.015	0.004

Comm. Service	(1)	(2)
VARIABLES	Excess Return	Excess Return
Pos. Sentiment	-0.004	
	(0.015)	
Neg. Sentimen	-0.006	
	(0.014)	
Pos. Sentiment	0.001	
	(0.015)	
Neg. Sentimen	-0.011	
	(0.015)	
Risk Free Rate	0.074	0.026
	(0.103)	(0.096)
% Change Pos.		-0.092
		(0.086)
% Change Neg.		0.110
		(0.072)
Constant	0.039	-0.192
	(0.296)	(0.151)
Observations	258	257
R-squared	0.005	0.009

Financials	(1)	(2)
VARIABLES	Excess Return	Excess Return
Pos. Sentiment (t)	0.008	
	(0.026)	
Neg. Sentiment (t)	0.002	
	(0.026)	
Pos. Sentiment (t-1)	0.027	
	(0.027)	
Neg. Sentiment (t-1)	0.011	
	(0.022)	
Risk Free Rate	-0.098	-0.054
	(0.095)	(0.091)
% Change Pos. Sent.		0.002
		(0.075)
% Change Neg. Sent.		-0.028
		(0.017)
Constant	-0.345	0.069
	(0.291)	(0.110)
Observations	258	257
R-squared	0.009	0.004

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Information Tech.	(1)	(2)
VARIABLES	Excess Return	Excess Return
Pos. Sentiment (t)	-0.011	
	(0.018)	
Neg. Sentiment (t)	0.017	
	(0.014)	
Pos. Sentiment (t-1)	-0.019	
	(0.014)	
Neg. Sentiment (t-1)	0.015	
	(0.012)	
Risk Free Rate	0.038	0.065
	(0.079)	(0.070)
% Change Pos. Sent.		0.186
		(0.141)
% Change Neg. Sent.		-0.043
		(0.066)
Constant	0.144	0.065
	(0.300)	(0.089)
Observations	258	258
R-squared	0.017	0.009

Gold	(1)	(2)
VARIABLES	Excess Return	Excess Return
Pos. Sentiment	0.007	
	(0.032)	
Neg. Sentimen	0.010	
	(0.029)	
Pos. Sentiment	0.017	
	(0.032)	
Neg. Sentimen	-0.025	
	(0.034)	
Risk Free Rate	0.225	0.094
	(0.240)	(0.203)
% Change Pos.		-0.043
		(2.151)
% Change Neg.		0.126
		(1.117)
Constant	-1.230	-0.147
	(1.759)	(0.268)
Observations	258	258
R-squared	0.006	0.001

In addition to my results supporting existing investing theory, the magnitude of the coefficients on investor search sentiment are quite small. For example, the largest coefficient on positive search sentiment is for the consumer staples sector. The magnitude of the coefficient suggests that for a one unit increase in positive investor search sentiment for the sector, excess return increases by 0.032 percent holding all else equal in the model. While a one-unit change in investor search sentiment is small, the resulting change in excess returns would be quite small even for a five to ten unit change in investor search sentiment. For an investor looking to actively manage a portfolio by leveraging investor search sentiment, the excess returns gained from utilizing search sentiment as an indicator of when to buy or sell sector ETFs it is likely that the transaction costs of purchasing or selling the ETF shares would outweigh the excess return generated.

These results are consistent to existing literature as well. While some prior literature suggests that some Google Trends data can be utilized to actively manage stocks to generate positive returns, it also suggests that these positive returns are often offset by transaction costs (Bijl et al. 2016). The results of previous literature also investigate whether Google Trends data can be used to generate positive returns; whereas, I investigate if the data can be leveraged to produce excess returns which is a more difficult task. In addition, ETFs in general also provide a more diversified set of stocks to investors. This makes them less susceptible to massive price changes from panic buying or selling. Such diversification might also allow the ETF to perform more similar to the S&P 500 overtime, making the task of utilizing Google Trends data to generate excess returns even more difficult.

VI. Conclusion

The results of the empirical model are largely inconclusive. This indicates that Google Trends data may not be able to provide investors with information to outperform the market. While not having any significant results does not provide any information about how to create a trading strategy with Google Trends data, it is consistent with the Efficient Market Hypothesis. This would indicate that Google Trends data is priced into stock pricing and that investors are not able to utilize the data to predict price movement or create a trading strategy that produces consistent excess returns over time.

VII. References

Bank, Matthias, Martin Larch, and Georg Peter. 2011. "Google Search Volume and Its Influence on Liquidity and Returns of German Stocks." *Financial Markets and Portfolio Management* 25 (3): 239–64. https://doi.org/10.1007/s11408-011-0165-y.

Beer, Francisca, Fabrice Hervé, and Mohamed Zouaoui. 2013. "Is Big Brother Watching Us? Google, Investor Sentiment and the Stock Market." *Post-Print*.

Bijl, Laurens, Glenn Kringhaug, Peter Molnár, and Eirik Sandvik. 2016. "Google Searches and Stock Returns." *International Review of Financial Analysis* 45 (May): 150–56. https://doi.org/10.1016/j.irfa.2016.03.015.

Challet, Damien, and Ahmed Bel Hadj Ayed. 2014. "Do Google Trend Data Contain More Predictability than Price Returns?" *Papers*.

Christoffersen, Peter F., and Francis X. Diebold. 2006. "Financial Asset Returns, Direction-of-Change Forecasting, and Volatility Dynamics." *Management Science* 52 (8): 1273–87. https://doi.org/10.1287/mnsc.1060.0520.

Chronopoulos, Dimitris K., Fotios I. Papadimitriou, and Nikolaos Vlastakis. 2018. "Information Demand and Stock Return Predictability." *Journal of International Money and Finance* 80 (February): 59–74. https://doi.org/10.1016/j.jimonfin.2017.10.001.

Fama, Eugene F, and Kenneth R French. 1996. "The CAPM Is Wanted, Dead or Alive."

Journal of Finance 51 (5): 1947–58.

Hjalmarsson, and Erik. 2005. “On the Predictability of Global Stock Returns.” *Working Papers in Economics*, February.

Kim, Neri, Katarína Lučivjanská, Peter Molnár, and Roviel Villa. 2019. “Google Searches and Stock Market Activity: Evidence from Norway.” *Finance Research Letters* 28 (March): 208–20. https://doi.org/10.1016/j.frl.2018.05.003.

Preis, Tobias, Helen Susannah Moat, and H. Eugene Stanley. 2013. “Quantifying Trading Behavior in Financial Markets Using Google Trends.” *Scientific Reports* 3 (1): 1684. https://doi.org/10.1038/srep01684.

Ranco, Gabriele, Darko Aleksovski, Guido Caldarelli, Miha Grčar, and Igor Mozetič. 2015. “The Effects of Twitter Sentiment on Stock Price Returns.” Edited by Tobias Preis. *PLOS ONE* 10 (9): e0138441. https://doi.org/10.1371/journal.pone.0138441.

Takeda, Fumiko, and Takumi Wakao. 2014. “Google Search Intensity and Its Relationship with Returns and Trading Volume of Japanese Stocks.” *Pacific Basin Finance Journal* 27 (1): 1–18. https://doi.org/10.1016/j.pacfin.2014.01.003.

Tang, Wenbin, and Lili Zhu. 2017. “How Security Prices Respond to a Surge in Investor Attention: Evidence from Google Search of ADRs.” *Global Finance Journal* 33 (May): 38–50. https://doi.org/10.1016/j.gfj.2016.09.001.

Tversky, Amos, and Daniel Kahneman. 1992. “Advances in Prospect Theory: Cumulative Representation of Uncertainty.” *Journal of Risk and Uncertainty* 5 (4): 297–323. https://doi.org/10.1007/BF00122574.