INVESTOR ATTENTION AND STOCK PERFORMANCE:
A SEARCH ENGINE OPTIMIZATION APPROACH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF APPLIED MATHEMATICS
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

ABDULLAH EFE GÜL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
FINANCIAL MATHEMATICS

FEBRUARY 2022

Approval of the thesis:

INVESTOR ATTENTION AND STOCK PERFORMANCE: A SEARCH ENGINE OPTIMIZATION APPROACH

submitted by **ABDULLAH EFE GÜL** in partial fulfillment of the requirements for the degree of **Master of Science in Department of Financial Mathematics, Middle East Technical University** by,

Prof. Dr. A. Sevtap Kestel
Dean, Graduate School of **Applied Mathematics** __________

Prof. Dr. Ali Devin Sezer
Head of Department, **Financial Mathematics** __________

Assoc. Prof. Dr. Seza Danışoğlu
Supervisor, **Business Administration**, **METU** __________

Examining Committee Members:

Prof. Dr. A. Sevtap Kestel
Actuarial Sciences, IAM, METU __________

Assoc. Prof. Dr. Seza Danışoğlu
Business Administration, METU __________

Prof. Dr. Zehra Nuray Güner
Business Administration, METU __________

Dr. Hande Ayaydın Hacıömeroğlu
Business Administration, METU __________

Prof. Dr. Burak Günalp
Economics, Çankaya University __________

Date: __________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: ABDULLAH EFE GÜL

Signature :

ABSTRACT

INVESTOR ATTENTION AND STOCK PERFORMANCE:
A SEARCH ENGINE OPTIMIZATION APPROACH

Gül, Abdullah Efe
M.S., Department of Financial Mathematics
Supervisor: Assoc. Prof. Dr. Seza Danışoğlu

February 2022, 67 pages

This thesis proposes two new measures of investor attention: Search Traffic (ST) and Click Per Search (CPS). These two measures as well as the commonly used Google Search Volume Index (SVI) measure are constructed using a search engine optimization and the number of keywords is optimized while measuring SVI. ST is measured based on financial website URLs without using any search keyword and is a direct measure of investor attention. The relationships between investor attention and stock market activities consisting of return and volatility are investigated for the Dow Jones Index (DJI) and its constituent stocks. The study provides robust evidence that attention has significant and asymmetric impact on index returns as well as excess returns. It has significant and negative influence on returns under bearish conditions while significant and positive effect during bullish conditions. Attention is also a significant driver of both index and stock volatility such that volatility increases following an increase in attention. In addition, investors respond to price reversals more quickly compared to positive index returns. Observations on CPS suggest that the more investors search for a financial keyword, the less they click on financial websites per searched keyword.

Keywords: Returns, Volatility, Investor attention, Search Engine Optimization

ÖZ

YATIRIMCI İLGİSİ VE HİSSE SENEDİ PERFORMANSI:
ARAMA MOTORU OPTİMİZASYONU YAKLAŞIMI

Gül, Abdullah Efe
Yüksek Lisans, Finansal Matematik Bölümü
Tez Yöneticisi: Doç. Dr. Seza Danışoğlu

Şubat 2022, 67 sayfa

Bu tez, yatırımcı ilgisi ölçümü için iki yeni yöntem önermektedir: Arama Trafiği (ST) ve Arama Başına Tıklama (CPS). Bu iki ölçü ve yaygın olarak kullanılan Google SVI ölçüsü, bir arama motoru optimizasyonu kullanılarak oluşturulur ve SVI ölçülürken anahtar kelime sayısı optimize edilir. ST, herhangi bir arama anahtar sözcüğü kullanılmadan finansal web sitesi URL'lerine dayalı olarak ölçülür ve yatırımcı ilgisinin doğrudan bir ölçüsüdür. Yatırımcı ilgisi ile getiri ve oynaklıktan oluşan hisse senedi piyasası faaliyetleri arasındaki ilişkiler Dow Jones Endeksi ve endeksin bireysel hisse senetleri için araştırılmıştır. Bu çalışma, ilginin hem fazla getiriler hem endeks getirileri üzerinde belirgin asimetrik etkilere sahip olduğuna dair güçlü kanıt sunmaktadır. Düşüş piyasasında getiriler üzerinde belirgin ve negatif, yükseliş piyasasında ise belirgin ve pozitif bir etkiye sahiptir. İlgi aynı zamanda hem endeks hem de hisse senedi oynaklığının önemli bir itici gücüdür, öyle ki ilgideki bir artışın ardından oynaklık artar. Ayrıca, yatırımcılar pozitif endeks getirilerine kıyasla fiyat dönüşlerine daha hızlı tepki verirler. CPS ile ilgili gözlemler, yatırımcılar bir finansal anahtar kelimeyi ne kadar çok ararsa, aranan anahtar kelime başına o kadar az finansal web sitelerine tıkladığını göstermektedir.

Anahtar Kelimeler: Getiriler, Oynaklık, Yatırımcı İlgisi, Arama Motoru Optimizasyon

To My Family and Future Wife…

ACKNOWLEDGMENTS

I would like to express my very great appreciation to my thesis supervisor Assoc. Prof. Dr. Seza Danışoğlu for her precious guidance and valuable suggestions throughout conducting my thesis. She is always supportive and patient during the preparation of this thesis. Her self-discipline and tremendous knowledge have encouraged me in all the time of my thesis and daily life. I am grateful to her for giving me the freedom to develop my ideas during the development of this thesis.

I wish to extend my warmest thanks to research assistants Meral Şimşek and Özge Tekin for their friendly supports during my M.S. education.

I would like to express my sincere gratitude to my family for their continuous support as always in my life. Their support and love have always motivated me to achieve my goals during my entire life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CPS	Click Per Search
DJI	Dow Jones Index
DJIA	Dow Jones Industrial Average
ER	Excess Return
GMM	Generalized Method of Moments
NYSE	New York Stock Exchange
OLS	Ordinary Least Squares
QR	Quantile Regression
R	Return
SEO	Search Engine Optimization
SPCS	Standardized Click Per Search
SST	Standardized Search Traffic
SSV	Standardized Search Volume
SSVI	Standardized Search Volume Index
ST	Search Traffic
SV	Search Volume
SVI	Search Volume Index
U.S.	United States
URL	Uniform Resource Locator
VIX	CBOE S&P 500 Volatility Index
VLM	Detrended Log Volume
VXD	CBOE DJIA Volatility Index

CHAPTER 1

INTRODUCTION

Psychological evidence shows that investors may behave irrationally while making a financial decision since they have limited resources and are influenced by cognitive biases. Kahneman [18] argues that attention can be defined with effort and introduces a model in which attention can be improved as a result of conscious focus since it is a limited cognitive resource. In behavioral finance, attention is present when retail or institutional investors pay attention to any information about an asset. Behavioral finance researchers analyze the cognitive bias mentioned in Kahneman [18] in order to observe the relationship between investor attention and stock markets. Most studies conclude that investor attention has significant potential to forecast future stock market returns and volatilities. However, these results violate the efficient market hypothesis [10] which argues that all available information is embedded in stock market prices and the forecast of stock market movements is not possible since prices follow a random path as new information arises [9].

In this study, the primary aim is to observe investor behavior under different market conditions and investigate the relationship between retail investor attention and stock market movements. The leading measure of retail investor attention in the literature is the Google Search Volume Index (SVI) introduced by Da, Engelberg, and Gao [8]. Following Da, Engelberg, and Gao [8]; Joseph, Wintoki, and Zhang [17], Vozlyublennaia [33], Klemola, Nikkinen, and Peltomäki [19], Bijl et al. [4], Chen [6], Padungsaksawasdi, Treepongkaruna, and Brooks [27], Reyes [29], and

Swamy and Dharani [31] measure attention using SVI to examine the relationship between investor attention and stock market activities.

This thesis contributes to the literature by offering Search Engine Optimization (SEO) as a new measure of investor attention. This alternative measure overcomes the ambiguity of keyword selection which plays an important role in the calculation of Google SVI. With the SEO measure, the proxy for investor attention is optimized by capturing more relevant keywords in the calculation of SVI. In addition to SEO, the thesis introduces two other proxies for investor attention: Click Per Search (CPS) and Search Traffic (ST). CPS is defined as the average number of clicks in websites following the search for a target keyword. ST represents the number of visitors who reach the target website based on the search results. The ST measure is an alternative to SVI because both two proxies measure the same type of attention with different methods. However, since ST is calculated directly based on a website's URL without using any keywords, it is superior to SVI in capturing attention. The CPS measure is different from ST and SVI because it is negatively correlated with search volume since people seem to have fewer clicks per keyword when they search for a financial keyword. SVI, ST, and CPS are calculated to measure investor attention for the Dow Jones Index (DJI) as well as its constituent stocks.

Prior studies mostly use regression methods, which are extensions of the Ordinary Least Squares (OLS) method, in order to test whether investor attention has any predictive ability for stock market returns and volatility. OLS-based methods can be misleading for estimating the comprehensive relationship among observations since the techniques take into account only the conditional mean of the response variable. The Quantile Regression (QR) method, introduced by Koenker and Bassett [20], makes it possible to examine the relationships across different conditional quantiles of the dependent variable and it does not have the same strong assumptions about observations as OLS. Also, QR is flexible and not sensitive to outliers across distributions of variables, unlike OLS-based methods. Thus, since stock prices do not follow a normal distribution, this thesis uses the QR as a more suitable method. The Generalized Method of Moments (GMM) is employed as well in order to estimate the relationship between investor attention and returns since GMM does not have restrictive assumptions regarding the distribution of variables. The control variables used in the baseline models are (i) trading volume [2, 7, 12], (ii) volatility [14, 27], and (iii) VXD, that is, the version of the S&P 500 VIX for DJI [1, 24, 26]. The dependent variable is excess stock returns following Bijl et al. [4] and Swamy and Dharani [31]. The response variable is index returns [19, 33] for DJI.

Results of the study show that there is a significant relationship between investor attention and stock market activities. In addition, retail attention has an asymmetric impact on returns and excess returns in different market environments. Index and excess returns significantly increase under bullish conditions while they decrease under bearish conditions following an increase in attention, which is consistent with retail investor herding behavior. Furthermore, attention is a significant determinant of index and individual stock volatility, and it is associated with uncertainty. There is evidence that an increase in attention predicts higher future volatility in stock returns. It is also shown that when daily returns are analyzed, investors seem to pay more attention to negative index returns, and when weekly returns are analyzed, positive index returns attract more attention from investors.

The remainder of the thesis is organized as follows. Chapter 2 presents a literature review on investor attention. Chapter 3 describes the data and how investor attention is measured. Chapter 4 provides baseline models and estimation methodology. Chapter 5 provides and discusses the empirical results, and Chapter 6 concludes the thesis.

CHAPTER 2

LITERATURE REVIEW

2.1. Overview of Investor Attention Studies

In the literature, the main objective of studies on investor attention is to analyze the relationship between investor attention and asset price movements. This relation is mainly about investigating how investor attention affects the asset prices. In addition, it involves investigating how market activities affect investor attention. The first query focuses on forecasting future asset returns using an investor attention measure where investor attention is measured based on some market activity, such as trading volume and volatility. The second query focuses on understanding whether returns, trading volume and volatility have an influence on investor attention.

In early research, Merton [25] proposes a model of stock market equilibrium with incomplete information, and he argues that firm value increases with the degree of investor recognition of the firm and thus future returns decrease with increased investor recognition. Wang [34] shows that while institutional investor sentiment can be used to forecast future stock market movements, small investor sentiment does not considerably impact future stock returns. Likewise, Wang, Keswani, and Taylor [35] examine whether investor sentiment helps to forecast the return volatility and find that lagged returns lead to higher volatility. Barber and Odean [2] suggest that stock price movements are directly related to investor attention and divide the attention process into two different decision-making states, which are buying and selling a stock. They conclude that individual investors have limited attention while buying stocks due to difficulties in obtaining large amounts of stock information. However, there is no such problem while selling stocks as owned stocks are certain.

2.2. Proxies of Investor Attention

In the literature, investor attention is measured in several ways and these measure types are generally determined according to whether the investor is less sophisticated (retail investor) or more sophisticated (institutional investor). Generally, the simplest measure of investor attention is the trading volume regardless of investor type since it is easy to observe and understand trading volume. Chordia and Swaminathan [7] analyze the pattern between stock returns and trading volume as a proxy of investor sentiment and find that trading volume is significantly related to stock returns. Similarly, Gervais, Kaniel, and Mingelgrin [12] examine whether there is a link between trading volume and future stock prices. They find that trading volume impacts future stock prices and show that increase in trading volume implies an increase in investor attention.

Although different attention proxies are suggested in the literature, several studies provide evidence that Google's Search Volume Index (SVI) can be used as a measure of retail investor attention. In fact, the Da, Engelberg, and Gao [8] study is the first to use Google SVI as a measure of investor attention. They find that investor attention can be used to predict future stock returns and high SVI forecasts positive returns in the next two weeks. Likewise, Joseph, Wintoki, and Zhang [17] use ticker searches of the stocks included in S&P 500 as an investor attention measure in the calculation of SVI. They examine the usability of the SVI to forecast abnormal returns and trading volume, and find that SVI has a predictive quality depending on whether the stock is less or more volatile.

For measuring institutional investor attention, news searches and reading activity on Bloomberg terminals are typically used as proxies. For instance, Ben-Rephael, Da, and Israelsen [3] develop an attention measure for sophisticated investors using the user profile search function (PEOP) on Bloomberg for specific stocks and find that institutional investors are more attentive to major news events. In addition, Li et al. [21] use Electronic Data Gathering and Retrieval (EDGAR) log files to measure institutional investor attention and analyze how sophisticated investors affect the incorporation of information into stock prices.

There are other studies in the literature offering different proxies for investor attention. For example, investor attention is gauged from excess returns [2], Wikipedia activity [1, 11], abnormal trading volume [2, 23], and asset-specific tweets [22].

2.3. Recent Studies on Investor Attention Using Google SVI

Several studies use the SVI proxy to examine the effect of attention by studying either market index returns [6, 16, 19, 27, 33], or common stock returns [1, 15, 29, 31, 32], or cryptocurrency returns [30, 36].

These studies also use different methodologies. For instance, Padungsaksawasdi, Treepongkaruna, and Brooks [27] use a panel VAR model estimated by GMM and study developing and developed country stock market indexes. They conclude that attention improves the predictability of future stock market volatility and higher search frequency predicts a more volatile market. Herwartz and Xu [14] also use a VAR model to analyze the U.S. stock market index (DJI), the German stock market index (DAX), and the U.K. stock market index (FTSE 100), and document that attention has an instantaneous influence on market volatility. On the other hand, Audrino, Sigrist, and Ballinari [1] employ a heterogeneous autoregressive (HAR) model, measure attention by SVI and Wikipedia activity and analyze the returns on individual U.S. companies listed on the New York Stock Exchange (NYSE) or Nasdaq in addition to the returns on the Dow Jones Index. They find that attention is a significant predictor of future stock volatilities. Furthermore, Vozlyublennaia [33] uses Granger Causality tests and Vector Autoregression (VAR) models and shows that investor attention is a significant predictor of future returns on the DJI, S&P 500, and NASDAQ indexes. Likewise, Klemola, Nikkinen, and Peltomäki [19] estimate VAR models for S&P 500 index returns in order to investigate the effect of attention for different market conditions. They find that investors are prone to pay attention to price reversals. Similarly, Chen [6] examines different country stock market indexes with VAR models and finds that high Google search volumes predict negative returns. On the other hand, Bijl et al. [4] apply panel data regressions to S&P 500 firms and find that, contrary to Da, Engelberg, and Gao [8], and Joseph, Wintoki, and Zhang [17] results, excess returns decrease following an increase in investor attention. Swamy and Dharani [31] employ the quantile regression methodology to analyze the returns of the NIFTY 50 companies and show that attention has an asymmetric impact at the upper and lower quantiles of excess returns. Similarly, Hsieh, Chan, and Wang [15] use pooled OLS and panel regressions and demonstrate the asymmetric effect of attention during bull and bear markets for the returns of individual stocks traded on the Taiwan stock market.

CHAPTER 3

DATA

3.1. Sample Description

Dow Jones Index (DJI) is a stock index that tracks the prices of the largest market capitalization industrial companies whose shares are traded on the New York Stock Exchange (NYSE). DJI is one of the most widely followed stock market indexes in the world and is used as an indicator of the U.S. economy's performance. In fact, Audrino, Sigrist, and Ballinari [1] state that search activities for the keywords "stock market" and "Dow Jones" are strongly correlated with a coefficient of 0.88, suggesting that DJI is perceived to reflect a considerable portion of the U.S. stock market activities. In this thesis, the sample for analyzing the relationship between investor attention and stock performance is selected as the DJI itself as well as its constituent stocks. The daily closing prices and daily trading volumes of the DJI and its constituent stocks, and the daily closing prices of the DJIA volatility index (VXD) are obtained from the Thomson Reuters database for the period between September 2015 and August 2020. Since constituent stocks change over time, added and dropped stocks over the selected period are excluded. Hence, 27 stocks (out of the regularly included 30) remain in the sample. Table 3.1 lists the companies included in the DJI over the sample period.

Table 3.1: Constituents of DJI included in the sample

Ticker Symbol	Company
AAPL	Apple Inc.
AXP	American Express Company
BA	The Boeing Company
CAT	Caterpillar Inc.
CSCO	Cisco Systems Inc.
CVX	Chevron Corporation
DIS	The Walt Disney Company
GS	The Goldman Sachs Group Inc.
HD	The Home Depot Inc.
IBM	International Business Machines Corporation
INTC	Intel Corporation
JNJ	Johnson & Johnson
JPM	JPMorgan Chase & Co.
KO	The Coca-Cola Company
MCD	McDonald's Corporation
MMM	3M Company
MRK	Merck & Co. Inc.
MSFT	Microsoft Corporation
NKE	Nike Inc.
PFE	Pfizer Inc.
PG	The Procter & Gamble Company
TRV	The Travelers Companies Inc.
UNH	UnitedHealth Group Inc.
V	Visa Inc.
VZ	Verizon Communications Inc.
WMT	Walmart Inc.
XOM	Exxon Mobil Corporation

3.2 Construction of Return, Volatility, and Volume Variables

A logarithmic return is calculated to represent the rate of return on the DJI, individual stocks, and VXD:

$$R_t = log\left(\frac{P_t}{P_{t-1}}\right)$$

In this calculation, P_t refers to the closing price of either the DJI, or an individual constituent stock, or the VXD on day t.

Joseph, Wintoki, and Zhang [17], Bijl et al. [4], and Swamy and Dharani [31] show that excess returns are significantly related to retail investor attention. In this study, excess returns are calculated for the constituent stocks in order to capture the unsystematic change in the price of each individual stock:

$$ER_{i,t} = R_{i,t} - R_{Dow,t}$$

In this calculation, $R_{i,t}$ is the logarithmic return of stock i at time t and $R_{Dow,t}$ is the logarithmic return of DJI at time t.

Since return volatility is shown to have a relationship with investor attention [14, 27], volatility for the DJI and individual stock returns is calculated as the standard deviation of returns:

$$Volatility = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(R_i - \bar{R})^2} * \sqrt{m}$$

In this equation, R_i is defined as before and m is the number of working days in a week (month) when calculating the volatility of weekly (monthly) returns. .

Trading volume is another variable that is shown to be influenced by investor attention [2, 7, 12]. Therefore, a detrended log volume, denoted as VLM, is calculated to represent trading volume, with a trend of the average of past 3 months following Bijl et al. [4], and Swamy and Dharani [31].

$$VLM_t = \log(TV_t) - \frac{1}{k}\sum_{i=t-(k-1)}^{t}\log(TV_i)$$

In this equation, TV is trading volume and k is 3, 12, or the number of working days for the past 3 months, when calculating monthly, weekly and daily trading volume, respectively. Also, please note that results are similar when raw trading volume is included in the model.

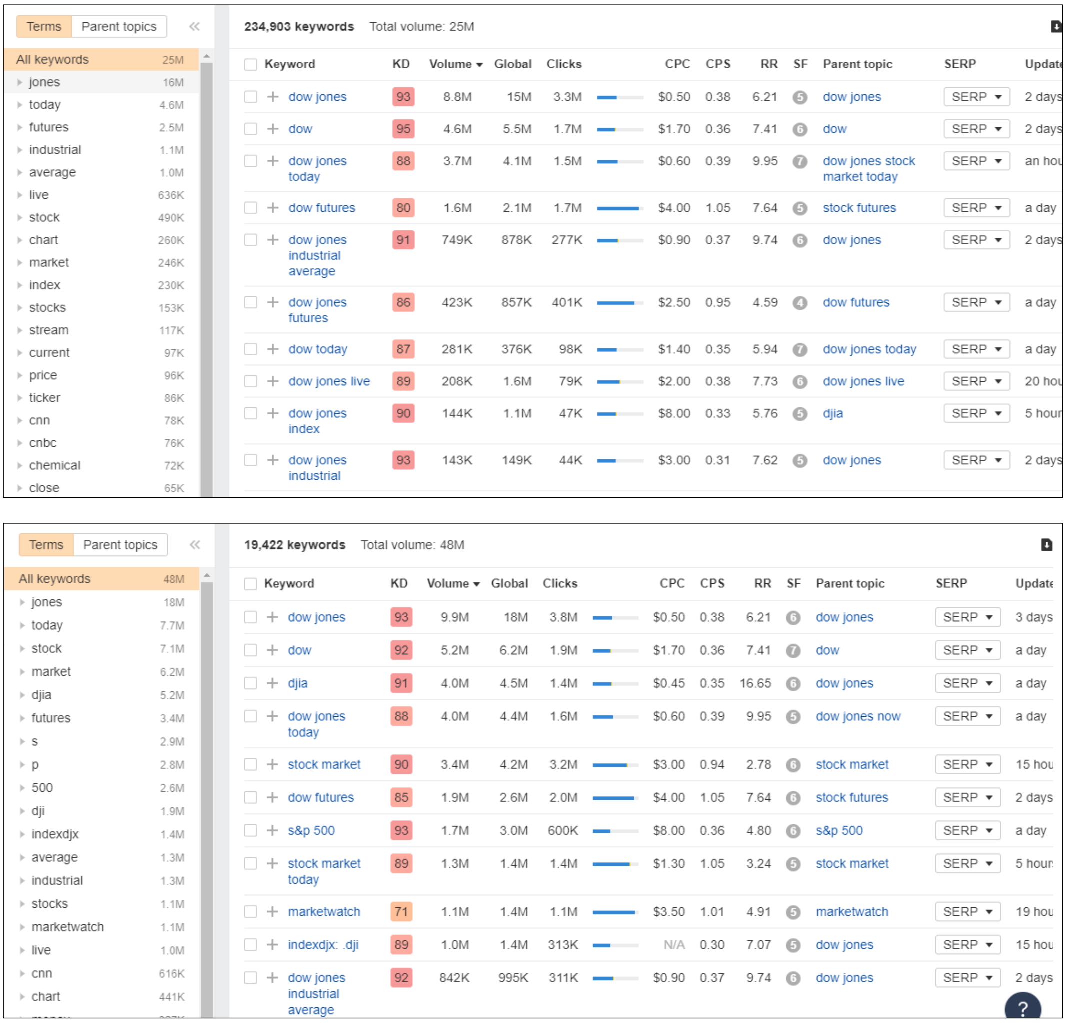
3.3. Construction of Investor Attention Measures

In the literature, retail attention is mostly gauged with Google SVI that is obtained for arbitrary keywords that are related to the stock or index of interest. The need to select these "relevant keywords" is the most important weakness of the SVI measure since the basis for choice is ambiguous. For instance, when obtaining the Google SVI for a stock, whether the company name, or the stock's ticker, or the company’s famous product, or some other obvious identifier of the company should be selected as the relevant keyword is not certain. As an alternative to the Google SVI, this thesis examines whether investor attention can be captured more precisely by performing a Search Engine Optimization (SEO) which is a website tool that helps to optimize a website or webpage to improve its traffic volume from web searches. SEO offers high functionality since there are several ways to optimize a website by using tools such as a keyword explorer or a website explorer. A keyword explorer allows to obtain the full list of possible keywords searched on Google according to desired target, and this way makes it possible to capture a broader measure of attention by using the SVI. Alternatively, a website explorer provides a measure of Search Traffic that is directly measured based on the website URLs, and this way may help to overcome the ambiguity of measuring attention with keywords. Moreover, SEO also provides the Click Per Search data which is the ratio of clicks to search volume for any keyword. The website www.ahrefs.com is used to carry out the search engine optimization for all different measures of investor attention.

3.3.1. Google SVI

As a first method, investor attention is measured by the Google SVI. Da, Engelberg, and Gao [8] introduce the Google SVI as a direct measure of investor attention to investigate the relationship between individual stock returns and retail investor attention. Google Trends provides SVI relative to the highest search volume, implying that a value of 100 implies the highest popularity and a value of 50 implies half of that popularity for the selected period. Since SVI provides relative data, the SVI for different keywords for the same stock or index cannot be summed up directly. In contrast, SEO provides raw data that makes it possible to add together different keywords for the same stock or index, resulting in a broader attention measure. Since Google trends offers data on a weekly basis for a period of up to five years, the sample period for constructing the SEO-based SVI is selected to be between September 2015 and August 2020.

The keyword explorer enables to capture all possible keywords related to the DJI and its constituent stocks searched on Google by providing the list of all keywords that have same terms with the target keyword and also keyword ideas that may be associated with the target keyword. For example, the list of some keywords that are related to the DJI is provided in Figure 3.1. Note that the list includes keywords that are directly related to the DJI, such as “dow jones” or “dow” as well as keywords that are not directly related to the DJI, such as “s&p 500” or “marketwatch”. These unrelated keywords are not included while measuring attention since they are not a direct measure for DJI. The list of keywords used for constructing the attention measure for the DJI is reported in the Appendix.



Terms | Parent topics

All keywords	25M
jones	16M
today	4.6M
futures	2.5M
industrial	1.1M
average	1.0M
live	636K
stock	490K
chart	260K
market	246K
index	230K
stocks	153K
stream	117K
current	97K
price	96K
ticker	86K
cnn	78K
cnbc	76K
chemical	72K
close	65K

234,903 keywords Total volume: 25M

Keyword	KD	Volume	Global	Clicks	CPC	CPS	RR	SF	Parent topic	SERP	Update
dow jones	93	8.8M	15M	3.3M	$0.50	0.38	6.21	5	dow jones	SERP	2 days
dow	95	4.6M	5.5M	1.7M	$1.70	0.36	7.41	6	dow	SERP	2 days
dow jones today	88	3.7M	4.1M	1.5M	$0.60	0.39	9.95	7	dow jones stock market today	SERP	an hou
dow futures	80	1.6M	2.1M	1.7M	$4.00	1.05	7.64	5	stock futures	SERP	a day
dow jones industrial average	91	749K	878K	277K	$0.90	0.37	9.74	6	dow jones	SERP	2 days
dow jones futures	86	423K	857K	401K	$2.50	0.95	4.59	4	dow futures	SERP	a day
dow today	87	281K	376K	98K	$1.40	0.35	5.94	7	dow jones today	SERP	a day
dow jones live	89	208K	1.6M	79K	$2.00	0.38	7.73	6	dow jones live	SERP	20 hou
dow jones index	90	144K	1.1M	47K	$8.00	0.33	5.76	5	djia	SERP	5 hour
dow jones industrial	93	143K	149K	44K	$3.00	0.31	7.62	5	dow jones	SERP	2 days

Terms | Parent topics

All keywords	48M
jones	18M
today	7.7M
stock	7.1M
market	6.2M
djia	5.2M
futures	3.4M
s	2.9M
p	2.8M
500	2.6M
dji	1.9M
indexdjx	1.4M
average	1.3M
industrial	1.3M
stocks	1.1M
marketwatch	1.1M
live	1.0M
cnn	616K
chart	441K

19,422 keywords Total volume: 48M

Keyword	KD	Volume	Global	Clicks	CPC	CPS	RR	SF	Parent topic	SERP	Update
dow jones	93	9.9M	18M	3.8M	$0.50	0.38	6.21	6	dow jones	SERP	3 days
dow	92	5.2M	6.2M	1.9M	$1.70	0.36	7.41	7	dow	SERP	a day
djia	91	4.0M	4.5M	1.4M	$0.45	0.35	16.65	6	dow jones	SERP	a day
dow jones today	88	4.0M	4.4M	1.6M	$0.60	0.39	9.95	5	dow jones now	SERP	a day
stock market	90	3.4M	4.2M	3.2M	$3.00	0.94	2.78	6	stock market	SERP	15 hou
dow futures	85	1.9M	2.6M	2.0M	$4.00	1.05	7.64	6	stock futures	SERP	2 days
s&p 500	93	1.7M	3.0M	600K	$8.00	0.36	4.80	6	s&p 500	SERP	a day
stock market today	89	1.3M	1.4M	1.4M	$1.30	1.05	3.24	5	stock market	SERP	5 hour
marketwatch	71	1.1M	1.4M	1.1M	$3.50	1.01	4.91	5	marketwatch	SERP	19 hou
indexdjx: .dji	89	1.0M	1.4M	313K	N/A	0.30	7.07	5	dow jones	SERP	15 hou
dow jones industrial average	92	842K	995K	311K	$0.90	0.37	9.74	6	dow jones	SERP	2 days

Figure 3.1: The list of some keywords related to the Dow Jones Index. The top half of the figure presents keywords that have common words with the target keyword. The bottom half of the figure shows the keywords that are associated with the target keyword.

When constructing the attention measure for individual firms, following Da, Engelberg, and Gao [8] keywords that are more frequently used to represent something other than the company itself are not included: the best example for such a case is the keyword "apple" being excluded from the attention measure constructed for the Apple Company since the word "apple" is very likely to represent searches that are not related to investing. For the individual stocks, not only the stock tickers but also other possible keywords related to stock are collected to construct the most comprehensive proxy of investor attention. The typical choice of keywords would include the company name including the word "stock", its ticker symbol and variations of the company name and the ticker. For example, for the Apple Company, the keywords are "apple stock", "apple stocks", "apple stock price", "apple stock price today", "aapl", "aapl stock", "aapl stock price", and "aapl stock price today". The keywords for the other stocks are selected in a similar fashion. During data collection, it was observed that some stocks have fewer keywords due to lack of data. For instance, the keywords for the Pfizer Company are "pfizer stock", "pfizer stock price", "pfe", "pfe stock", "pfe stock price", and "pfe stock price today" so there are no "pfizer stocks" and "pfizer stock price today" since these combinations exhibited very low search volume. Moreover, keywords are collected in the manner described above based on worldwide searches for individual stocks and DJI.

Consequently, the choice of keywords is not arbitrary because all related possible keywords searched are taken into consideration while measuring investor attention.

3.3.2. Search Traffic

As a second method, investor attention is measured based on the search traffic at a particular financial website URL without using any keyword. The website explorer provides search traffic information for any website domain or URL. Search traffic refers to how much traffic the target website gets by users who reach the website from search results. As such, search traffic provides an alternative measure of investor attention by making it possible to calculate the change in search traffic for a particular website. "Yahoo Finance" is chosen as the particular website domain because its ranking on Google is top 5 for almost all individual stocks as well as the DJI keywords searched on Google. For instance, Figure 3.2 shows how Yahoo Finance is the most frequently visited website for the target keyword “MMM stock”.

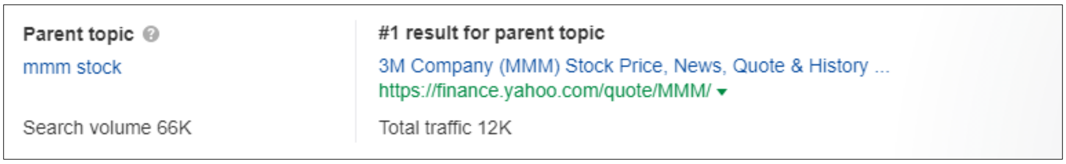


Figure 3.2: Rank of a website for a target keyword. The figure shows the ranking of “Yahoo Finance” for the keyword “MMM stock” and it has a rank of 1.

The URL format for all individual stocks as well as the index is www.finance.yahoo.com/quote/ with an extension of the stock ticker. For example, the URL for the stock “IBM” is www.finance.yahoo.com/quote/IBM and its output is presented in Figure 3.3.

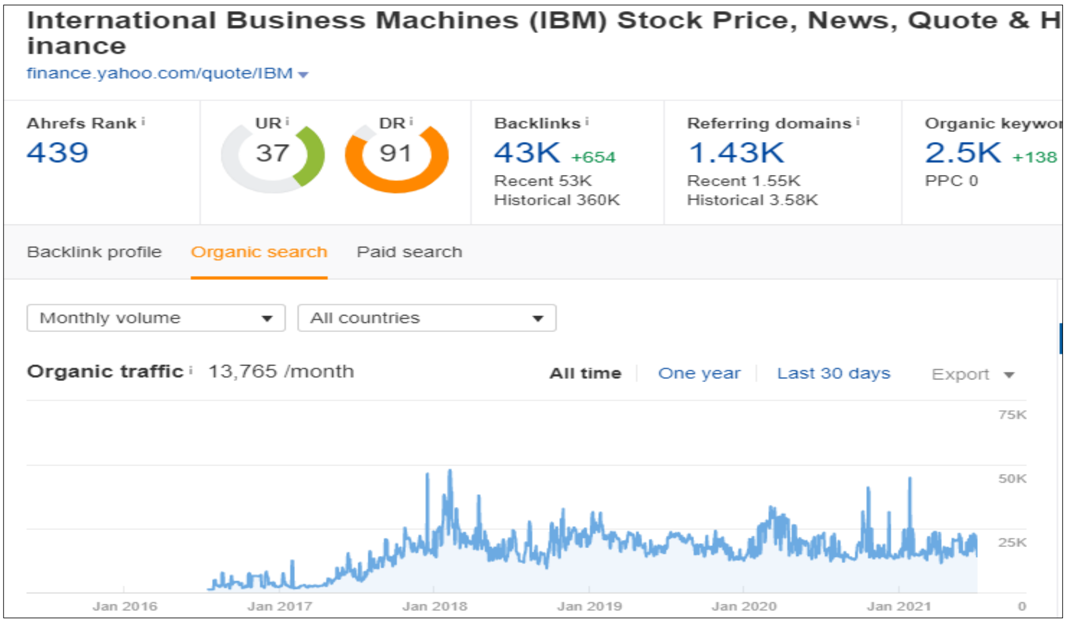


Figure 3.3: Search Traffic. The graph plots the daily search traffic of the website URL, www.finance.yahoo.com/quote/IBM for the stock “IBM”.

With the search traffic measure, a financial website’s URL is enough to gauge how much attention the index or a constituent stock receives during a given period without the need to use the keyword explorer. Search traffic has the largest frequency among the three attention proxies as it is available on a daily basis. Search traffic data for the website Yahoo Finance is available beginning on July 18, 2016. Hence, search traffic data are collected between July 18, 2016 and August 28, 2020 on a daily basis.

3.3.3. Click Per Search

As a third method, Ahref's SEO provides Click Per Search (CPS) and Search Volume (SV) data on a monthly basis. Search volume is the average number of searches in a month on Google for a target keyword. CPS is the ratio of clicks to search volume and it shows the average number of clicks in websites following the search for a target keyword. For example, Figure 3.4 shows SV and CPS data for the keyword "intc" for the stock "Intel", and "dow jones today" for Dow Jones Index, respectively.

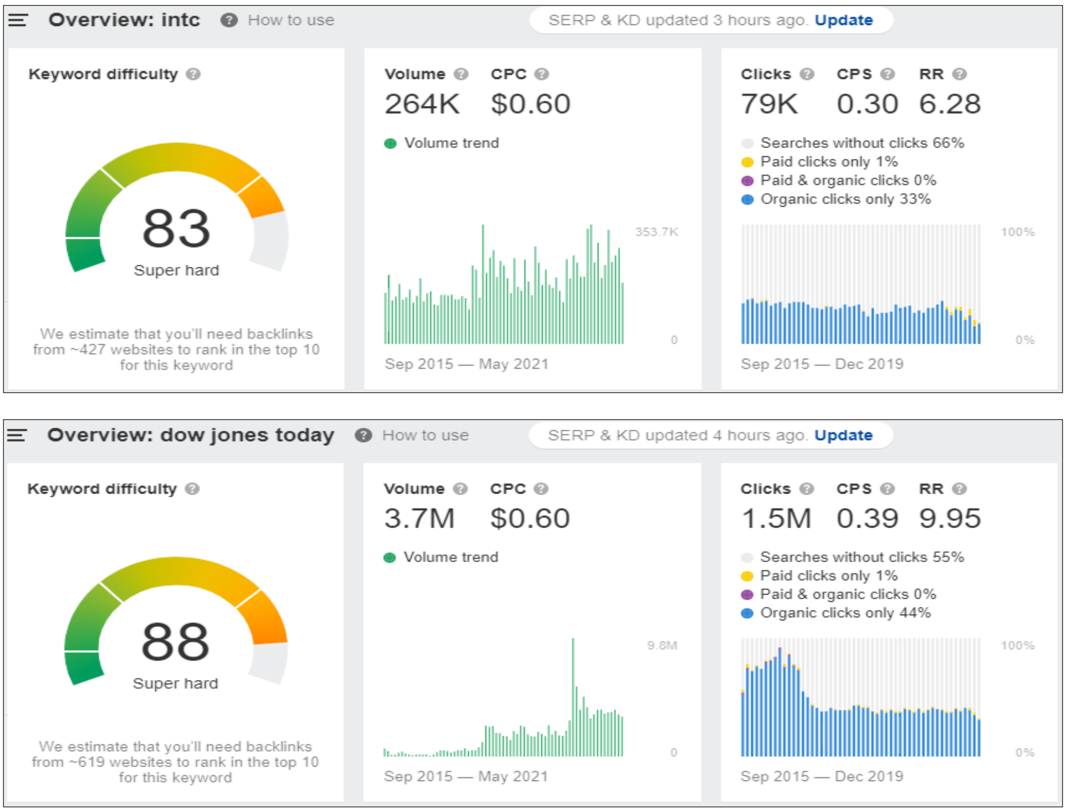


Figure 3.4: Search Volume and Click Per Search. The figures presents the output of monthly SV and CPS of the target keyword "intc" for the stock "INTEL" and "dow jones today" for DJI, respectively.

For some keywords, majority of the searches do not receive any clicks at any website. The most likely reason is that Google shows the summary price chart when one searches a target keyword related to an individual stock or the index. Figure 3.5 demonstrates an example for the keyword "dow jones today". It is very likely that some people observe the price chart that is automatically shown by Google and do not need to click any other website after the initial search. Still, it may be plausible to expect that especially retail investors may need to obtain more information about the target stock or index by clicking on financial websites after their initial search for the keywords on Google. Therefore, CPS is taken as a proxy of investor attention to observe the impact of reaching additional or further information about a stock. The CPS data are available between September 2015 and December 2019 on monthly basis. Please note that, for consistency, the keywords used in obtaining the CPS data are exactly the same as the used for the SVI measure.

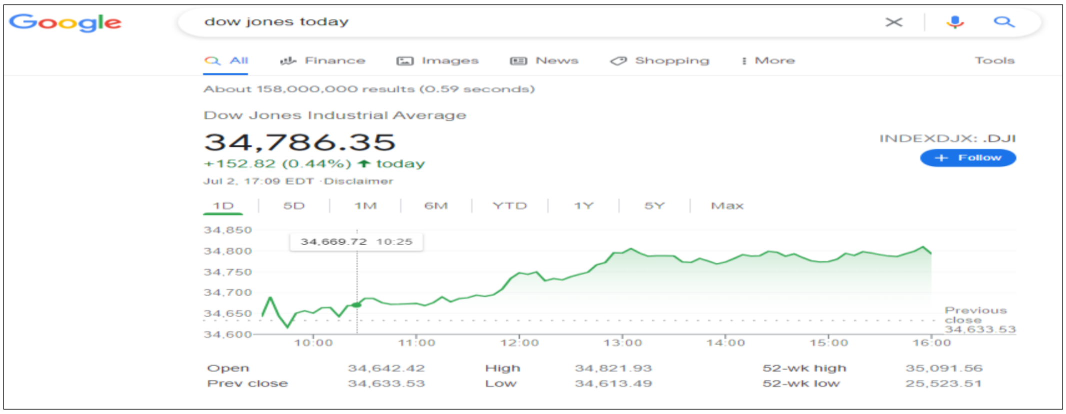


Figure 3.5: Output of a Google search for the target keyword "dow jones today"

3.4. Summary of Proxies and Standardizing the Attention Measures

Following Bijl et al. [4], and Swamy and Dharani [31] attention measures are standardized in order to make the measures calculated for the individual stocks versus the index more comparable across the different proxies:

$$Standardized\ Attention_t = \frac{Attention_t - \frac{1}{n}\sum_{i=1}^{n} Attention_i}{\sigma_{Attention}}$$

In this equation, n is the number of months, weeks, or working days for monthly, weekly, and daily proxies, respectively, and $\sigma_{Attention}$ is the standard deviation of the attention series over the entire sample period. The standardized measures are denoted by adding the capital letter “S” in front of the initial abbreviations used for the proxies summarized in Table 3.2: SSVI, SST, SCPS , and SSV.

Table 3.2 summarizes the attention proxies used in the models.

Table 3.2: Summary of proxies

SEO is based on www.ahrefs.com for all proxies.

	Search Volume Index	Search Traffic	Click Per Search	Search Volume
Designation	SVI	ST	CPS	SV
Frequency	Weekly	Daily	Monthly	Monthly
Measure	Keyword	Website URL	Keyword	Keyword
Source	SEO and Google Trends	SEO	SEO	SEO

CHAPTER 4

METHODOLOGY

4.1. Baseline Models

The baseline regression models' equations are constructed following the literature. Chordia and Swaminathan [7], and Gervais, Kaniel, and Mingelgrin [12] conclude that trading volumes are significantly associated with the stock market movements. Barber and Odean [2] use trading volume as an indirect measure for investor attention and state that investor attention, trading volume, and stock returns are interrelated. Furthermore, Da, Engelberg, and Gao [8], and Joseph, Wintoki, and Zhang [17] show that investor attention has a potential to forecast future stock returns. Bijl et al. [4], and Swamy and Dharani [31] show that investor attention can be used as a trading strategy to predict future excess returns of stocks. Padungsaksawasdi, Treepongkaruna, and Brooks [27] find that investor attention is related to the returns, volatility, and trading volume. Likewise, Herwartz and Xu [14] find that investor attention has an influence on both volatilities and trading volumes. Moreover, Mittnik, Robinzonov, and Spindler [26] state that the implied volatility index VIX is a main risk driver of volatility and it has a significant influence on future volatility. Audrino, Sigrist, and Ballinari [1] document that both investor attention and sentiment are significant determinants to predict future volatilities and conclude that the implied volatility index VIX is one of the most relevant economic variables to forecast stock market volatility. In addition, Mbanga, Darrat, and Park [24] use the VIX to measure investor sentiment and find that investor attention statistically affects investor sentiment.

The implied volatility index VIX is based on S&P 500 index, which allows investors to identify the level of risk, fear, or stress in the market while making investment decisions. Since this thesis examines the investor attention for the Dow Jones Index (DJI), VXD, which is the implied volatility index for the DJI, is used to measure investor sentiment.

The baseline regression equation for the DJI models where the index return is the dependent variable is written as follows:

$$R_t = \beta_0 + \beta_1 Standardized\ Attention_t + \beta_2 VLM_t + \beta_3 Volatility_t + \beta_4 VXD_t$$

In this equation, R_t is the log return on the DJI at time t, $Standardized\ Attention_t$ is either the SSVI, SST , SCPS or SSV at time t, VLM_t is the detrended log volume of DJI at time t, $Volatility_t$ is the volatility of DJI return at time t, and VXD_t is the log change in the DJIA volatility index at time t.

Similarly, the baseline regression equation for the individual stock models where the constituent stock's excess return is the dependent variable is written as follows:

$$ER_{t,i} = \beta_{0,i} + \beta_{1,i} Standardized\ Attention_{t,i} + \beta_{2,i} VLM_{t,i} + \beta_{3,i} Volatility_{t,i} + \beta_{4,i} VXD_t$$

where $ER_{t,i}$ is the excess return of stock i at time t, $Standardized\ Attention_{t,i}$ is either the SSVI, SST, SCPS or SSV for the individual stock at time t, $VLM_{t,i}$ is the detrended log volume of stock i at time t, and $Volatility_{t,i}$ is the volatility of stock i returns at time t, and VXD_t is the log change in the DJIA volatility index at time t.

Please note that the frequency of the response variables as well as explanatory variables in the above models is changed in accordance with the frequency of the calculated attention measure.

4.2. Estimation Methodology

For estimating the models presented above, the traditional Ordinary Least Square (OLS) methodology could provide misleading results especially when the data have heavy-tailed distributions since OLS is based on the conditional mean as a measure of central location. Thus, OLS becomes inappropriate when the distribution of observations is not Gaussian as a result of being heavily affected by outliers.

Alternatively, the Quantile Regression (QR), introduced by Koenker and Bassett [20], is not sensitive to outliers or heavily skewed distributions. It is based on the premise that changing a data value that is below (above) the p^{th} sample quantile to some other value below (above) the p^{th} sample quantile does not affect the value of the p^{th} sample quantile. QR is more informative and efficient if variables used in the model do not exhibit the characteristics of the normal distribution. Also, while OLS uses merely conditional mean modelling for measures of central tendency, QR is flexible such that it can be modelled by any desired measure of location. It makes it possible to investigate the connection between the response variable and explanatory variables for different population segments with off-central locations by focusing on the upper and lower tails of the distribution of the response variable. It also allows to capture the change in the effect of a single covariate on the response variable across the subsequent quantiles, which implies that it helps to interpret the scale and shape shifts over the quantiles. Consequently, the QR method is more appropriate to interpret the relationship among variables that have asymmetric distributions. QR makes it possible to examine the relationship between attention and returns for different return quantiles as well as under different market conditions due to the method's flexibility of location measures.

Quantile regression is a type of linear regression in the study of the linear relationship between a response variable and explanatory variables by specifying the τ^{th} quantile,

$$Q(\tau \mid X_i, \beta(\tau)) = X_i^T \beta(\tau),$$

where X is the vector of explanatory variables and $\beta(\tau)$ is the vector of coefficients associated with the τ^{th} quantile for some value of $\tau \epsilon (0, 1)$.

Quantile regression uses the least-absolute-distance estimation method by minimizing the average (weighted) sum of the positive and negative residuals. The QR estimator is,

$$\hat{\beta}_n(\tau) = \ \text{argmin}_{\beta(\tau) \in R^p} \sum_{i=1}^{n} \rho_\tau \left(Y_i - X_i^T \beta(\tau)\right),$$

where Y_i is the vector of the response variable and ρ_τ is a linear loss function with an error u defined as,

$$\rho_\tau(\text{u}) = \ \tau \max(u, 0) + (1 - \ \tau)\max(-u, 0)$$

Robustness is crucial while studying highly skewed distributions. Since the asymptotic standard error has restrictions due to the assumption of i.i.d. standard errors, it may be misleading to perform hypothesis testing and to estimate confidence intervals. For this reason, the bootstrap method derived from a Monte-Carlo simulation by drawing samples of size n with replacement from actual observed data is performed and bootstrapped standard errors are calculated for all models in order to capture scale and shape shifts.

Each regression equation is estimated for the quantiles between 0.10 and 0.90 with an increment of 0.05 and coefficients across the quantiles in this range are presented in the figures included in the Results and Discussion chapter. Also, the 0.10, 0.50 and 0.90 quantiles are chosen for the lower tail, central, and upper tail measures, respectively and regression results are presented for these three quantile categories. The 0.10 and 0.90 quantiles represent extreme negative and positive return segments, for either the index return or the excess return series. Hence, the 0.10 conditional quantile is used to analyze the relationship between index return (or, excess return of individual stocks) and investor attention under bearish market conditions. On the contrary, the 0.90 conditional quantile is used to observe the same relationship under bullish market conditions.

In order to understand whether the model parameter estimates actually change for the extreme values of the dependent variable, equivalence of coefficients across quantiles is tested. In other words, the Wald slope equality test is employed for the following null hypotheses:

$$H_0: \beta_1(\tau_{0.10}) = \beta_1(\tau_{0.50}), \qquad H_0: \beta_1(\tau_{0.50}) = \beta_1(\tau_{0.90})$$

$$H_0: \beta_1(\tau_{0.10}) = \beta_1(\tau_{0.90})$$

In these equations, β_1 is the coefficient of the attention proxy for the models where the dependent variable is either the return or volatility, and it is the coefficient of the return and excess return variable for the models where the dependent variable is attention.

In addition, the Generalized Method of Moments (GMM) is employed for estimating the relationship between attention and returns based on the mean estimates of the response variables. Unlike OLS, the GMM methodology does not make strong assumptions about the distributions of observations. Therefore, if variables do not follow a normal distribution, GMM is more suitable than OLS to estimate the conditional mean of parameters. The GMM estimator is,

$$\hat{\theta} = \operatorname{argmin}_{\theta \in \vartheta} \hat{g}(\theta)^T \widehat{W} \hat{g}(\theta),$$

where $\widehat{W}$ is some positive semi-definite matrix and $\hat{g}(\theta)$ is the sample average of the population moments with the vector of observation Y_i defined as,

$$\hat{g}(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(Y_i, \theta)$$

Robust standard errors enabling the GMM moment conditions to be heteroskedastic while assuming they are not correlated across observations are calculated to obtain unbiased standard errors so that more reliable hypothesis tests can be performed.

CHAPTER 5

RESULTS AND DISCUSSION

5.1. Google SVI

In this section, estimation results are provided when investor attention is measured by the Google SVI. It should be remembered that when obtaining the Google SVI data, prior research uses stock tickers, firm names, or financial keywords related to the stock market, covering only a particular fragment of attention. The Search Engine Optimization (SEO) method used in this thesis makes it possible to collect search volume information for all possible keywords related to individual stocks or the index and captures a broader measure of investor attention. For this reason, it can be argued that the proxy of retail investor attention used in the thesis is superior to the measure used in the prior literature since the SEO-based SVI is more inclusive and keyword selection is not arbitrary. Table 5.1 reports the descriptive statistics of index variables to observe basic features of observations used in the SVI models for the entire sample period. Figure 5.1 provides both the box plot and histogram of index variables to summarize their distribution, central tendency, and variability visually. The correlation matrix is reported for multicollinearity purposes in Table 5.2.

As seen from Figure 5.1, variables have significant outliers that do not follow the relationship for majority of the observations. Attention and volatility variables have considerably right-skewed distributions. The null hypothesis that observations are from a normal distribution is rejected at the 1% significance level for all variables using the Jarque-Bera test, as seen in Table 5.1.

Table 5.1: Descriptive statistics for SVI model variables

This table shows the descriptive statistics of weekly DJI return, SSVI, VLM, Volatility, and VXD. Jarque-Bera test is also reported to check normality. The estimation period is from September 2015 to August 2020.

	Return	SSVI	VLM	Volatility	VXD
Mean	0.002	0.000	0.025	0.019	-0.0002
Median	0.003	-0.238	-0.004	0.013	-0.020
Standard Deviation	0.027	1.000	0.266	0.021	0.152
Kurtosis	14.381	15.210	3.606	37.230	4.502
Skewness	-1.643	3.325	1.113	4.989	1.078
Range	0.311	7.330	1.915	0.219	1.345
Minimum	-0.190	-0.819	-0.632	0.001	-0.465
Maximum	0.121	6.512	1.283	0.220	0.880
Jarque-Bera	2273	2888	188	15557	260
Observations	261	261	261	261	261

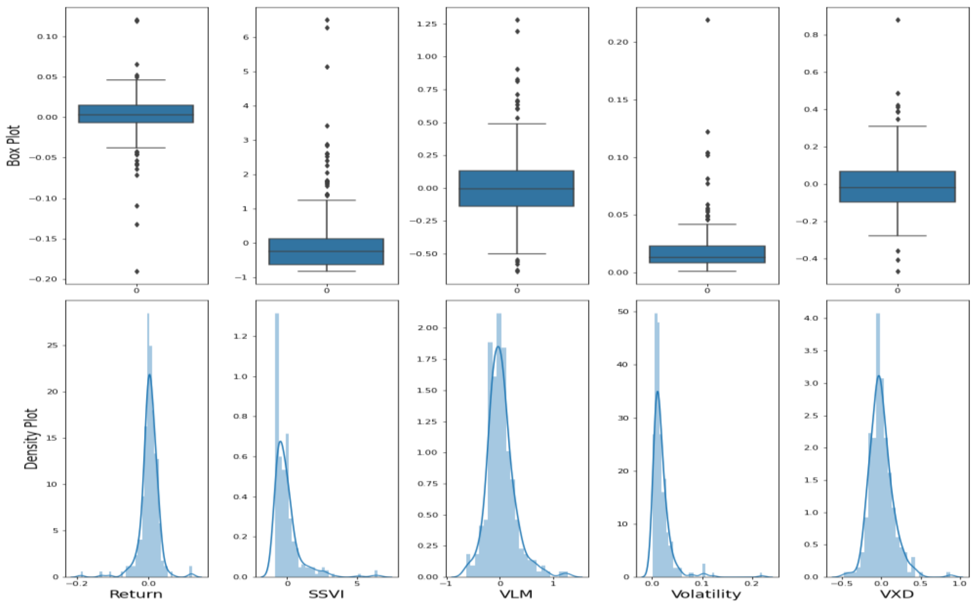


Figure 5.1: Box plot and histogram of index variables used in SVI models. The figure shows the box plot and histogram of each variable at a weekly frequency. Density plots are also placed on the histograms. The sample period is from September 2015 to August 2020.

Table 5.2: Correlation matrix for SVI model variables

This table shows correlations among index variables at weekly frequency. The variables that do not have a time subscript represent the whole sample period. Lagged variables used in the models are also reported. $Return_{t-2}$ defines the two weeks lagged return and VXD_{t-1} defines the one week lagged VXD.

	Return	SSVI	VLM	Volatility	VXD	$Return_{t-2}$
SSVI	-0.224					
VLM	-0.239	0.25				

(*continued*)

Table 5.2–*Continued*

	Return	SSVI	VLM	Volatility	VXD	$Return_{t-2}$
Volatility	-0.374	0.813	0.344			
VXD	-0.599	0.095	0.156	0.256		
$Return_{t-2}$	0.073	-0.223	-0.101	-0.404	0.075	
VXD_{t-1}	-0.057	0.16	0.28	0.198	-0.102	-0.244

5.1.1. Impact of SVI on Index Returns

Based on the correlations presented in Table 5.2, it is observed that SSVI and volatility have a very strong positive linear association. Since the variable of interest in this model is the SVI, the Volatility variable is dropped from baseline regression model to avoid multicollinearity problems. The following regression model is estimated to test the relationship between DJI returns and investor attention, which is proxied by the Search Volume Index (SVI):

$$R_t = \beta_0 + \beta_1 SSVI_t + \beta_2 VLM_t + \beta_3 VXD_t$$

Generalized Method of Moments (GMM) results for the conditional mean estimation and Quantile Regression results for the 0.10, 0.50 (median), and 0.90 conditional quantiles of the dependent variable DJI return are reported in Table 5.3.

Investor attention (SVI) and DJI returns do not have statistically significant relationship according to the GMM estimations. On the other hand, they have a significant relationship at the 1% significance level for all three quantile levels and the effect of SVI on the DJI return is asymmetric between upper and lower tails such that the estimated coefficients are -0.011 and 0.015 for the 0.10 and 0.90 quantiles, respectively. The most likely reason why attention does not have a significant influence on return at the conditional mean is that upper and lower tail effects offset each other. For this reason, interpreting solely the conditional mean estimation would be misleading to comment on all possible states of the market. For the 0.10 quantile, an increase in investor attention is associated with a decrease in the lower tails of the DJI return, which correspond to negative extreme returns. This finding suggests that investor attention further lowers the DJI return during bearish market conditions. On the other hand, investor attention has a positive influence on the upper tails of the DJI return, which correspond to positive extreme returns, implying that it affects the DJI return positively under bullish conditions. Also, the absolute value of the significant coefficient is larger for the 0.90 quartile, suggesting a somewhat stronger relationship between investor attention and index returns during bullish market periods. These results are consistent with the herding behavior of retail investors such that while they search for the DJI-related keywords, investors give buy orders if index returns are positive and give sell orders if index returns are negative. This finding is also consistent with Hsieh, Chan, and Wang [15]. Moreover, the pseudo R^2 decreases as the quantile level increases, implying that lower tail models explain more variation in the DJI returns compared to upper tail models. Also, note that there is robust evidence that investor sentiment, measured by the VXD and index returns are significantly and negatively related.

Table 5.3: Impact of SVI on DJI returns

The coefficient values are marked with significance levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Robust and bootstrapped standard errors are in parentheses for the conditional mean and quantile parameters, respectively. R^2 represents the pseudo R^2 for quantile regressions.

		Quantiles		
	Mean	0.10	0.50	0.90
Intercept	**0.002***	**-0.015*****	**0.002****	**0.022*****
	(0.001)	(0.001)	(0.001)	(0.001)
SSVI	-0.004	**-0.011*****	**-0.003*****	**0.015*****
	(0.005)	(0.002)	(0.001)	(0.002)
VLM	**-0.012****	**-0.015*****	-0.004	0.008
	(0.006)	(0.004)	(0.003)	(0.006)
VXD	**-0.101*****	**-0.098*****	**-0.092*****	**-0.080*****
	(0.008)	(0.009)	(0.006)	(0.007)
R^2	0.399	0.532	0.302	0.265

5.1.2. Impact of SVI on Index Volatility

According to Mbanga, Darrat, and Park [24], implied volatility can be used to measure the future volatility of the stock market. Hence, the one week lagged implied volatility, VXD_{t-1}, is included in the volatility model to account for this effect. The following regression model is estimated in order to examine the influence of investor attention on the index volatility. Results are presented in Table 5.4.

$$Volatility_t = \beta_0 + \beta_1 SSVI_t + \beta_2 VLM_t + \beta_3 VXD_{t-1} + \beta_4 R_t$$

The estimated SSVI coefficients are positive and significant at the 1% significance level for the conditional mean and the three quantile levels, suggesting that there is a symmetric and positive relationship between investor attention and volatility. In other words, information gathering by investors leads to higher market volatility, regardless of the existing level of volatility in the market. These findings are consistent with Padungsaksawasdi, Treepongkaruna, and Brooks [27]. The impact change of attention is greater at higher quantiles, implying that investor attention affects volatility with the uncertainty. This is consistent with Pastor and Veronesi [28], and Hautsch and Hess [13] who argue that investors pay more attention to new information if uncertainty about the stock market increases, leading to higher volatility in the market. Furthermore, the model's goodness of fit increases with the quantiles and is highest when it is estimated for the conditional mean.

Table 5.4: Impact of SVI on DJI volatility

The coefficient values are marked with significance levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Robust and bootstrapped standard errors are in parentheses for the conditional mean and quantile parameters, respectively. R^2 represents the pseudo R^2 for quantile regressions.

		Quantiles		
	Mean	0.10	0.50	0.90
Intercept	**0.019*****	**0.009*****	**0.017*****	**0.033*****
	(0.001)	(0.000)	(0.001)	(0.001)
SSVI	**0.016*****	**0.008*****	**0.012*****	**0.021*****
	(0.002)	(0.000)	(0.001)	(0.002)
VLM	**0.008****	0.002	0.002	**0.014*****
	(0.003)	(0.001)	(0.003)	(0.005)
VXD_{t-1}	0.006	**0.006****	**0.009****	0.003
	(0.007)	(0.002)	(0.004)	(0.012)
Return	**-0.136****	**-0.131*****	**-0.153*****	**-0.206*****
	(0.064)	(0.008)	(0.024)	(0.067)
R^2	0.715	0.238	0.316	0.549

5.1.3. Impact of Index Returns on SVI

It is also plausible to expect that lagged returns have an influence on retail investor attention [33]. In order to test whether this is supported by the data, the following regression model is estimated where investor attention is the dependent variable:

$$SSVI_t = \beta_0 + \beta_1 R_{t-2} + \beta_2 VLM_t + \beta_3 VXD_t + \beta_4 Volatility_t$$

Estimation results are presented in Table 5.5. The table reports the results where the two-week lagged returns are included in the model. It should be noted that returns with a one-week lag or lags higher than two weeks do have a significant effect on investor attention. Also, when the model is estimated with the contemporaneous returns, the coefficients for the return variable are 2.843***, 3.504**, and 1.290 for 0.10, 0.50 and 0.90 quantiles, respectively. Results in Table 5.5 show that past two weeks' returns have a significant and positive impact on investor attention for all three quantiles although the effect decreases in magnitude for the 0.10 quantile and median compared to the contemporaneous model. The GMM model also estimates a significant and positive coefficient suggesting a symmetric relationship across all models. In short, search frequency increases following an increase in return with a greater impact at higher quantiles, implying that positive index returns of past two weeks attract retail investors more compared to negative returns. Furthermore, it can be said that returns significantly affect retail attention with a long-term effect, which is consistent with the Vozlyublennaia [33] study. The model explains only 4.6% of the variation in SVI for the 0.10 quantile model but the pseudo R^2 amplifies considerably as the quantile level increases.

Table 5.5: Impact of DJI returns on SVI

The coefficient values are marked with significance levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Robust and bootstrapped standard errors are in parentheses for the conditional mean and quantile parameters, respectively. R^2 represents the pseudo R^2 for quantile regressions.

		Quantiles		
	Mean	0.10	0.50	0.90
Intercept	**-0.850*****	**-0.839*****	**-0.799*****	**-0.577*****
	(0.072)	(0.026)	(0.051)	(0.082)
$Return_{t-2}$	**5.994*****	**1.149***	**3.003****	**8.920*****
	(2.154)	(0.650)	(1.422)	(2.392)
VLM	-0.114	0.049	-0.132	-0.142
	(0.137)	(0.111)	(0.139)	(0.177)
VXD	**-1.021*****	**-0.291***	**-0.522****	**-2.044*****
	(0.296)	(0.173)	(0.241)	(0.377)
Volatility	**44.863*****	**12.303*****	**41.035*****	**67.460*****
	(4.248)	(0.757)	(2.003)	(3.360)
R^2	0.705	0.046	0.331	0.606

5.1.4. Visualization of the SVI-DJI Models

Figure 5.2 presents the plots for the observations and fitted lines for the focus variables while keeping the other explanatory variables constant. The asymmetric impact of attention on the index return is clearly seen between the upper and lower tails. On the other hand, the influence of return on retail investor attention is weaker than the impact of attention on return and volatility.

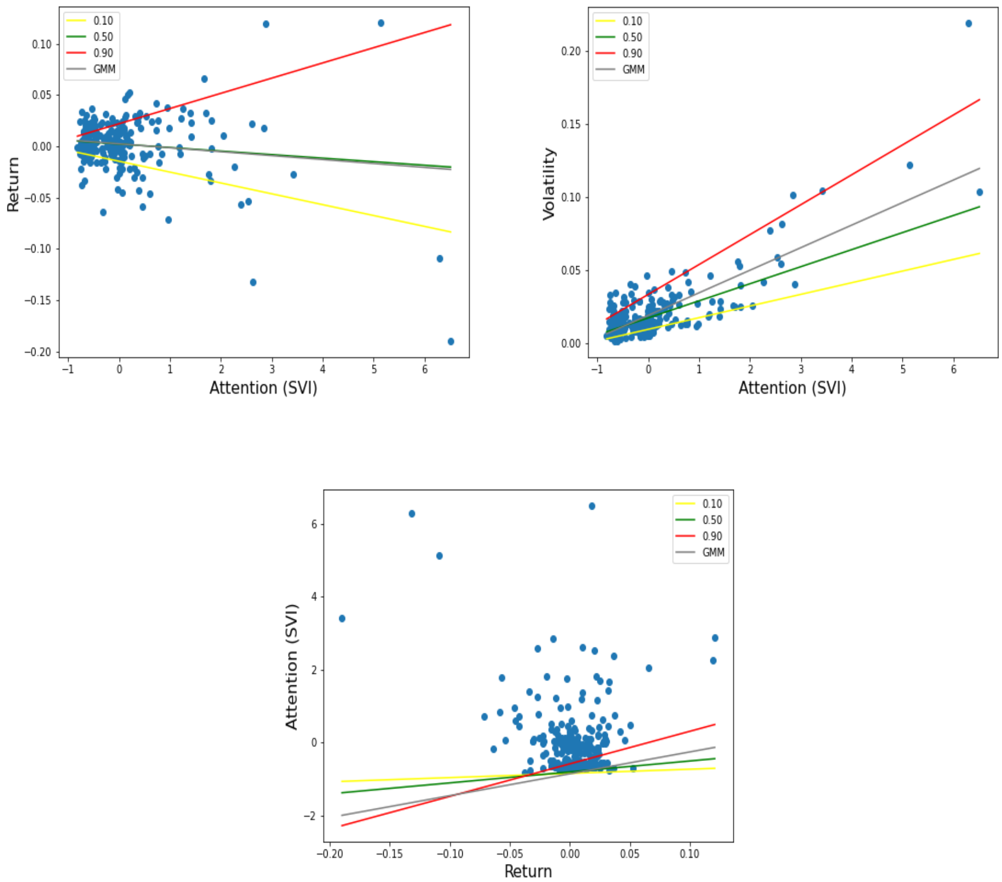


Figure 5.2: Scatter and fitted SVI-DJI model plots. The figure shows the scatter plot and fitted lines for the estimated quantile regression models. Red lines represent 0.90, green lines represent 0.50, yellow lines represent 0.10 quantile regressions, and grey lines represent the GMM model fits as seen in the legends.

The coefficients of the focus independent variables are also plotted to observe both the location and shape shifts across quantiles in Figure 5.3. For the dependent variable return, the slopes at upper tails are steeper than those at lower tails, leading to larger scale shifts at the upper tails. For the response variable volatility, the coefficient steadily increases at lower tails and accelerates at upper tails and the impact of attention dramatically increases above the 0.75 quantile. The confidence envelope never crosses the horizontal 0 line, implying that the estimated coefficients are significant at all quantile levels. Coefficient changes are always nonnegative across the quantiles for both of the dependent variables return and volatility. By contrast, both positive and negative shifts are observed across the quantiles for the dependent variable attention.

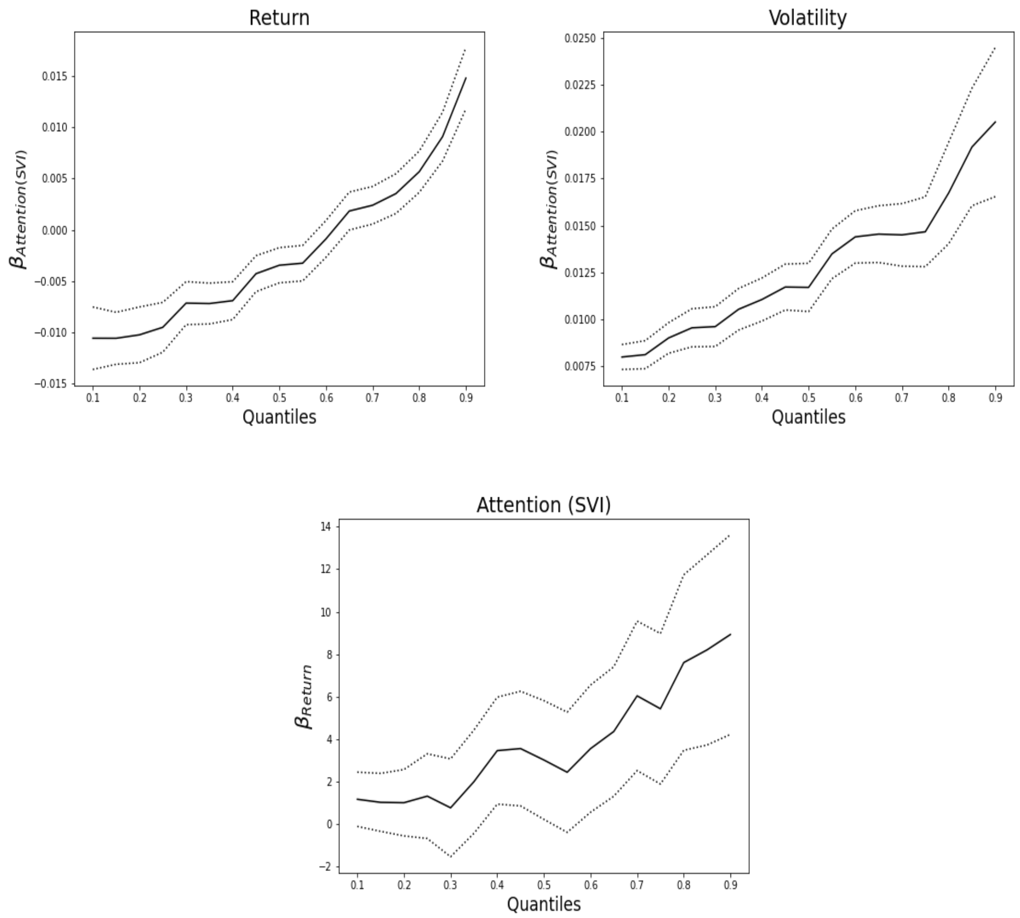


Figure 5.3: Coefficients across quantiles for SVI-DJI models. The figure shows coefficient changes of SVI-DJI models over the quantiles, ranging from 0.10 to 0.90 with their 5% confidence intervals. Titles of graphs indicate the dependent variable of the model.

5.1.5. Robustness for SVI-DJI Models

For the aforementioned models, the results of the slope equality tests are reported in Table 5.6. For the dependent variable return, the equivalence of the attention effect cannot be rejected at the 10% significance level between the 0.10 quantile and the median. Yet, the difference is significant between the median and the 0.90 quantile as well as between the 0.10 and 0.90 quantiles. The significant differences show that retail attention has an asymmetric impact between the upper and lower tails of the DJI return. For the dependent variable volatility, the impact of attention differs across all tested quantiles. Therefore, while the direction of the relationship between retail attention and DJI volatility is symmetric, the magnitude of the relationship is larger for the higher quantiles. For the dependent variable attention, the impact of the return variable is not statistically different between the 0.10 quantile and the median; however, it significantly varies between the median and the 0.90 quantile at the 10% significance level as well as between the 0.10 and 0.90 quantiles at the 5% significance level.

Table 5.6: Slope equality tests for the SVI-DJI models

The *p*-values of Wald tests are reported for slope equality test. *, **, and *** indicate that the null hypothesis is rejected at the 10%, 5%, and 1% significance levels, respectively. (1), (2) and (3) represent null hypotheses $\beta_1(\tau_{0.10}) = \beta_1(\tau_{0.50})$, $\beta_1(\tau_{0.50}) = \beta_1(\tau_{0.90})$ and $\beta_1(\tau_{0.10}) = \beta_1(\tau_{0.90})$, respectively.

	(1)	(2)	(3)
Dependent variable: Return Coefficient: SSVI	0.196	**0.000*** **	**0.001*** **
Dependent variable: Volatility Coefficient: SSVI	**0.009*** **	**0.002*** **	**0.000*** **
Dependent variable: SSVI Coefficient: Return	0.368	**0.061***	**0.020** **

5.1.6. Impact of SVI on Excess Returns of Constituent Stocks

The following regression model is estimated to test the impact of SVI on excess returns of the DJI's constituent stocks and results are reported in Table 5.7.

$$ER_{t,i} = \beta_{0,i} + \beta_{1,i}SSVI_{t,i} + \beta_{2,i}VLM_{t,i} + \beta_{3,i}VXD_t$$

All significant coefficients are negative for 22 stocks (out of 27) at the 0.10 quantile level and positive for 24 stocks (out of 27) at the 0.90 quantile level. Retail attention has an asymmetric effect under different market conditions such that excess return increases during the period of stock's good performance, but it decreases under stock's bad performance following an increase in search frequency. These results are consistent with the index return results. Interestingly, GMM estimates insignificant coefficients for all but three of the stocks. Thus, it can be said that retail investor attention has similar effects on the excess returns of individual stocks as well as the return on the market index but the magnitude of the effect differs across individual stocks. The finding of an asymmetric effect is consistent with Swamy and Dharani [31]. The 0.90 quantile results are consistent with Da, Engelberg, and Gao [8], and Joseph, Wintoki, and Zhang [17], but contradict with Bijl et al. [4] and vice versa is true for the 0.10 quantile results. Since the estimates of upper and lower tails cancel each other out in terms of sign and magnitude, it is not possible to observe a significant relationship between investor attention and returns when the estimations are carried out for the mean of excess returns. Note that the coefficients of determination are smaller than those of the DJI return models, suggesting that the model explains a smaller portion of the variation in individual stocks' excess returns compared to the DJI return.

Table 5.7: Impact of SVI on excess return of individual stocks

Coeff represents the estimated coefficient of the SSVI variable in the conditional mean and different quantile models where the dependent variable is the excess return on the constituent stocks. ***, **, * indicate levels of significance at 1%, 5%, and 10%, respectively. R^2 represents pseudo R^2 for quantile models.

			Quantiles					
	Mean		0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
AAPL	**0.004***	0.046	-0.001	0.107	**0.002***	0.011	**0.014*****	0.216
AXP	-0.001	0.037	**-0.007*****	0.246	**-0.002*****	0.012	**0.008*****	0.188
BA	-0.002	0.021	**-0.025*****	0.291	**-0.005*****	0.022	**0.018*****	0.152
CAT	-0.002	0.076	**-0.006*****	0.164	-0.002	0.049	**0.009*****	0.155
CSCO	-0.001	0.027	**-0.014*****	0.166	**0.003*****	0.015	**0.007*****	0.188
CVX	0.001	0.003	**-0.005*****	0.051	**0.001***	0.011	**0.007*****	0.121
DIS	0.001	0.008	**-0.002*****	0.061	0.000	0.008	**0.004*****	0.072
GS	0.001	0.039	-0.001	0.046	0.001	0.027	0.002	0.081
HD	0.001	0.017	-0.002	0.094	0.001	0.011	**0.005*****	0.099
IBM	0.000	0.026	**-0.010*****	0.287	**-0.002***	0.011	**0.009*****	0.162
INTC	-0.003	0.025	**-0.012*****	0.156	-0.002	0.021	**0.008*****	0.106
JNJ	0.001	0.064	**-0.008*****	0.126	0.000	0.058	**0.006*****	0.231
JPM	-0.001	0.025	**-0.004*****	0.134	-0.001	0.009	**0.003*****	0.108
KO	-0.001	0.070	**-0.002*****	0.084	**-0.001***	0.049	**0.003*****	0.132

(*continued*)

Table 5.7—*Continued*

	Mean		Quantiles					
			0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
MCD	-0.001	0.084	-0.001	0.111	-0.002	0.056	**0.004*****	0.120
MMM	**0.008*****	0.280	**-0.005*****	0.117	**0.005*****	0.070	**0.014*****	0.417
MRK	0.001	0.044	**-0.005*****	0.088	0.001	0.034	**0.007*****	0.161
MSFT	**0.002***	0.035	**-0.004*****	0.045	**0.002*****	0.022	**0.005*****	0.179
NKE	-0.001	0.015	0.001	0.069	-0.001	0.011	0.000	0.081
PFE	0.000	0.009	**-0.002*****	0.103	**0.001*****	0.013	**0.003*****	0.128
PG	0.002	0.114	**-0.005*****	0.096	0.001	0.076	**0.007*****	0.190
TRV	-0.002	0.021	**-0.006*****	0.054	**-0.002***	0.017	0.000	0.011
UNH	0.001	0.016	**-0.012*****	0.041	0.000	0.002	**0.012*****	0.222
V	0.000	0.019	**-0.002*****	0.093	0.000	0.010	**0.004*****	0.078
VZ	0.001	0.085	**-0.004****	0.129	**-0.002***	0.038	**0.005****	0.171
WMT	0.002	0.066	**-0.008*****	0.208	0.000	0.057	**0.010*****	0.260
XOM	0.000	0.016	**-0.003*****	0.159	0.000	0.008	**0.003*****	0.071

5.1.7. Impact of SVI on Stock Return Volatilities

The following regression model is estimated to observe the impact of investor attention on the volatility of individual stocks and the findings are reported in Table 5.8.

$$Volatility_{t,i} = \beta_{0,i} + \beta_{1,i}SSVI_{t,i} + \beta_{2,i}VLM_{t,i} + \beta_{3,i}VXD_{t-1} + \beta_{4,i}ER_{t,i}$$

The results are very similar to those of the DJI model and all predictions are nonnegative. The coefficient for SSVI is insignificant for only four stocks in the 0.10 model and one stock in the 0.90 model. For the models with central measures, only two and three stocks are insignificant for the mean and median estimations, respectively. Thus, volatility is strongly associated with retail investor attention across individual stocks, which is consistent with the Audrino, Sigrist and Ballinari [1] results. Similar to the DJI volatility results, the impact of attention on volatility is positive and its effect increases as the quantile increases for individual stock return volatilities, which is consistent with Pastor and Veronesi [28], and Hautsch and Hess [13] studies arguing that investors are more interested in new information when uncertainty is high, leading to an increase in search intensity and further volatility in returns. Furthermore, the explanatory power of the model increases with the quantiles.

Table 5.8: Impact of SVI on the volatility of individual stocks

Coeff represents the estimated coefficient of the SSVI variable in the conditional mean and different quantile models where the dependent variable is volatility. ***, **, * indicate levels of significance at 1%, 5%, and 10%, respectively. R^2 represents pseudo R^2 for quantile models.

			Quantiles					
	Mean		0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
AAPL	**0.009*****	0.491	**0.003*****	0.129	**0.009*****	0.240	**0.017*****	0.432
AXP	**0.010*****	0.549	**0.002*****	0.068	**0.006*****	0.223	**0.012*****	0.481
BA	**0.015*****	0.610	**0.007*****	0.170	**0.013*****	0.303	**0.024*****	0.543
CAT	**0.011*****	0.341	0.001	0.048	**0.008*****	0.144	**0.018*****	0.291
CSCO	**0.010*****	0.414	**0.003*****	0.077	**0.008*****	0.148	**0.012*****	0.367
CVX	**0.011*****	0.659	**0.004*****	0.157	**0.009*****	0.247	**0.017*****	0.536
DIS	**0.008*****	0.634	**0.004*****	0.207	**0.006*****	0.265	**0.010*****	0.474
GS	**0.006*****	0.224	0.001	0.067	0.002	0.091	**0.010*****	0.192
HD	**0.008*****	0.463	**0.002*****	0.045	**0.004*****	0.158	**0.011*****	0.373
IBM	**0.012*****	0.463	**0.004*****	0.065	**0.008*****	0.203	**0.018*****	0.406
INTC	**0.008*****	0.330	**0.004*****	0.083	**0.005*****	0.145	**0.017*****	0.319
JNJ	**0.006*****	0.443	**0.001*****	0.027	**0.003*****	0.104	**0.010*****	0.375
JPM	**0.007*****	0.517	**0.002*****	0.103	**0.005*****	0.159	**0.009*****	0.436
KO	**0.004*****	0.587	**0.002*****	0.067	**0.003*****	0.235	**0.006*****	0.463
MCD	**0.011*****	0.523	**0.002*****	0.054	**0.006*****	0.147	**0.016*****	0.397
MMM	**0.008*****	0.377	**0.003*****	0.098	**0.007*****	0.157	**0.012*****	0.320
MRK	**0.005*****	0.300	**0.001*****	0.033	**0.003*****	0.109	**0.007*****	0.284
MSFT	**0.005*****	0.485	**0.002*****	0.113	**0.004*****	0.191	**0.007*****	0.385
NKE	0.002	0.270	0.002	0.057	0.001	0.083	**0.027*****	0.291
PFE	**0.001*****	0.328	**0.001*****	0.062	**0.001*****	0.129	**0.005*****	0.342
PG	**0.009*****	0.429	**0.002*****	0.043	**0.005*****	0.106	**0.013*****	0.327
TRV	0.000	0.161	0.000	0.055	0.000	0.047	0.004	0.112
UNH	**0.007*****	0.285	**0.003*****	0.057	**0.007*****	0.120	**0.012*****	0.240
V	**0.007*****	0.439	**0.001*****	0.086	**0.004*****	0.132	**0.011*****	0.351
VZ	**0.005*****	0.414	**0.002*****	0.063	**0.003*****	0.120	**0.006*****	0.309
WMT	**0.006*****	0.548	**0.002*****	0.112	**0.003*****	0.190	**0.008*****	0.415
XOM	**0.005*****	0.681	**0.002*****	0.148	**0.005*****	0.311	**0.006*****	0.566

5.1.8. Impact of Excess Returns on SVI

The following regression model is estimated to test whether excess return has a significant impact on retail investor attention when the relationship is analyzed for individual stocks.

$$SSVI_{t,i} = \beta_{0,i} + \beta_{1,i}ER_{t,i} + \beta_{2,i}VLM_{t,i} + \beta_{3,i}VXD_t + \beta_{4,i}Volatility_{t,i}$$

Estimation results are presented in Table 5.9. Multiple tests are performed for the explanatory variable ER with different lags but it is not possible to show that individual stocks' excess returns have a significant influence on retail investor attention. Since the relationship is significant for the index returns, this finding suggests that investor attention increases especially when the overall market demonstrates bullish tendencies but not necessarily when an individual stock has high positive returns.

Table 5.9: Impact of excess returns on SVI

Coeff represents the estimated coefficient for the excess return (ER) variable in the conditional mean and different quantile models where the dependent variable is SSVI. ***, **, * indicate levels of significance at 1%, 5%, and 10%, respectively. R^2 represents pseudo R^2 for quantile models.

			Quantiles					
	Mean		0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R	Coeff	R^2
AAPL	**6.508****	0.518	**2.798*****	0.218	**4.527****	0.308	8.002	0.434
AXP	-7.734	0.506	2.737	0.114	2.456	0.199	-9.269	0.486
BA	-2.508	0.580	-0.321	0.040	**-4.015*****	0.273	**-9.313*****	0.537
CAT	-0.992	0.396	0.302	0.083	-1.794	0.173	-0.228	0.376
CSCO	-1.224	0.343	-0.231	0.072	0.955	0.160	-6.758	0.324
CVX	**9.158****	0.633	**3.565***	0.131	4.821	0.284	6.411	0.507
DIS	-2.662	0.602	0.889	0.087	4.106	0.202	-6.903	0.525
GS	2.007	0.125	0.819	0.061	3.130	0.060	**8.718***	0.107
HD	4.915	0.382	**10.211*****	0.064	0.069	0.121	2.209	0.312
IBM	0.852	0.480	**-3.400***	0.124	-1.507	0.212	-2.094	0.432
INTC	**-5.856***	0.256	-1.122	0.067	**-7.110*****	0.154	-2.693	0.214
JNJ	9.956	0.360	-1.482	0.034	**6.886****	0.130	17.450	0.264
JPM	-8.346	0.427	2.320	0.021	-1.317	0.116	-10.178	0.423
KO	7.701	0.496	6.780	0.059	6.558	0.142	-1.063	0.452
MCD	-4.524	0.493	0.065	0.061	0.282	0.186	-7.011	0.404
MMM	**-8.143****	0.387	2.125	0.082	**-7.427****	0.168	-11.075	0.331
MRK	1.103	0.176	-0.831	0.027	3.163	0.085	-1.825	0.168
MSFT	7.936	0.322	2.916	0.041	**13.130*****	0.107	16.926	0.331
NKE	-1.252	0.016	-0.439	0.014	**0.732****	0.085	-0.949	0.132
PFE	4.846	0.133	**7.511*****	0.029	**5.411***	0.085	-2.673	0.191
PG	2.334	0.390	**-5.609***	0.103	2.153	0.164	10.819	0.312
TRV	-3.640	0.075	**-6.554****	0.046	-2.684	0.035	1.778	0.097
UNH	1.054	0.313	-1.352	0.095	2.043	0.153	1.872	0.251
V	2.890	0.372	-2.378	0.026	**13.448****	0.099	1.042	0.361

(*continued*)

Table 5.9–*Continued*

	Mean		Quantiles					
			0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R	Coeff	R^2
VZ	-1.459	0.333	-4.067	0.039	**-4.993***	0.098	-0.329	0.318
WMT	-2.159	0.366	1.681	0.036	-1.901	0.112	-9.388	0.342
XOM	-1.900	0.650	-2.897	0.127	-6.534	0.185	-1.950	0.594

5.1.9 Robustness for SVI-Stock Models

The results of slope equality tests for the individual stock return models are reported in Table 5.10. As seen from the table, the retail attention's impact on excess return and volatility significantly differs across the quantiles for the majority of individual stocks. Therefore, test results imply that attention has an asymmetric influence on excess returns in terms of the direction of the relationship while it has a symmetric effect on volatility with the magnitude getting larger towards higher quantiles.

Please note that the slope equality tests for the model where the dependent variable is attention are not reported since excess return does not have a significant effect on attention.

Table 5.10: Slope equality tests for the SVI-Stock models

The *p*-values of Wald tests are reported for slope equality test. *, **, and *** indicate that the null hypothesis is rejected at the 10%, 5%, and 1% significance levels, respectively. (1), (2) and (3) represent null hypotheses $\beta_1(\tau_{0.10}) = \beta_1(\tau_{0.50})$, $\beta_1(\tau_{0.50}) = \beta_1(\tau_{0.90})$ and $\beta_1(\tau_{0.10}) = \beta_1(\tau_{0.90})$, respectively.

	Dependent variable: Excess return Coefficient: SSVI			Dependent variable: Volatility Coefficient: SSVI		
	(1)	(2)	(3)	(1)	(2)	(3)
AAPL	0.479	**0.002*** **	**0.007*** **	**0.000*** **	**0.043** **	**0.001*** **
AXP	**0.043** **	**0.000*** **	**0.000*** **	**0.000*** **	**0.036** **	**0.001*** **
BA	**0.002*** **	**0.011** **	**0.000*** **	**0.007*** **	**0.001*** **	**0.000*** **
CAT	0.234	**0.021** **	**0.006*** **	**0.000*** **	**0.000*** **	**0.000*** **
CSCO	**0.000*** **	**0.072* **	**0.000*** **	**0.001*** **	0.275	**0.019** **
CVX	**0.033** **	**0.004*** **	**0.001*** **	**0.000*** **	**0.000*** **	**0.000*** **
DIS	**0.023** **	**0.016** **	**0.002*** **	**0.001*** **	**0.030** **	**0.001*** **
GS	0.258	0.747	0.309	0.576	**0.045** **	**0.040** **
HD	0.143	**0.029** **	**0.011** **	**0.019** **	**0.001*** **	**0.000*** **
IBM	**0.001*** **	**0.000*** **	**0.000*** **	**0.001*** **	**0.000*** **	**0.000*** **

(*continued*)

Table 5.10–*Continued*

	Dependent variable: Excess return Coefficient: SSVI			Dependent variable: Volatility Coefficient: SSVI		
	(1)	(2)	(3)	(1)	(2)	(3)
INTC	**0.040****	**0.015****	**0.001*****	0.295	**0.005*****	**0.002*****
JNJ	**0.002*****	**0.004*****	**0.000*****	0.158	**0.002*****	**0.001*****
JPM	**0.074***	**0.054***	**0.005*****	**0.001*****	**0.087***	**0.003*****
KO	0.374	**0.000*****	**0.001*****	**0.011****	**0.078***	**0.009*****
MCD	0.762	**0.002*****	0.157	**0.000*****	**0.001*****	**0.000*****
MMM	**0.001*****	**0.046****	**0.000*****	**0.001*****	**0.000*****	**0.000*****
MRK	**0.002*****	**0.006*****	**0.000*****	**0.041****	**0.014****	**0.002*****
MSFT	**0.005*****	**0.034****	**0.000*****	**0.008*****	**0.034****	**0.002*****
NKE	0.918	0.975	0.983	0.939	0.383	0.432
PFE	0.302	0.211	0.152	0.340	**0.000*****	**0.000*****
PG	**0.038****	**0.003*****	**0.000*****	**0.002*****	**0.001*****	**0.000*****
TRV	0.326	0.417	0.204	0.555	0.272	0.236
UNH	**0.001*****	**0.000*****	**0.000*****	**0.003*****	0.213	**0.037****
V	**0.048****	**0.000*****	**0.000*****	**0.000*****	**0.003*****	**0.000*****
VZ	0.389	**0.006*****	**0.006*****	0.461	**0.013****	**0.005*****
WMT	**0.076***	**0.005*****	**0.002*****	0.637	**0.018****	**0.016****
XOM	**0.031****	**0.003*****	**0.000*****	**0.000*****	**0.060***	**0.000*****

5.2. Search Traffic

In this section, a new and direct measure of retail investor attention is proposed which uses Search Traffic (ST) as an alternative to SVI. ST has three important advantages over the SVI. First, it has an extensive flexibility because the search traffic can be obtained for any relevant website URL. In this study, the search traffic for the financial website, www.finance.yahoo.com, is used to gauge retail investor attention. Second, an attention proxy based on the search traffic overcomes the challenges of keyword selection since no keyword is needed to measure it. Third, it offers higher frequency for observations since it is available on a daily basis. Descriptive statistics for the ST model variables are documented in Table 5.11. Note that the volatility variable is not included in the ST models because of the inconvenient calculation of daily volatility. The box plot and histogram of index variables are plotted in Figure 5.4 and the correlation matrix of index observations is reported in Table 5.12.

Table 5.11: Descriptive statistics for ST model variables

This table reports the descriptive statistics of daily DJI returns, SST, VLM, and VXD. The sample period is from 18 July 2016 to 28 August 2020.

	Return	SST	VLM	VXD
Mean	0.0004	0.000	0.037	0.001
Median	0.001	-0.374	-0.017	-0.005
Standard Deviation	0.013	1.000	0.302	0.072
Kurtosis	26.232	0.045	3.044	3.974
Skewness	-1.237	0.998	1.271	0.682
Range	0.246	4.326	2.596	0.740
Minimum	-0.138	-1.090	-0.994	-0.408
Maximum	0.108	3.236	1.602	0.332
Jarque-Bera	29725	172	674	755
Observations	1038	1038	1038	1038

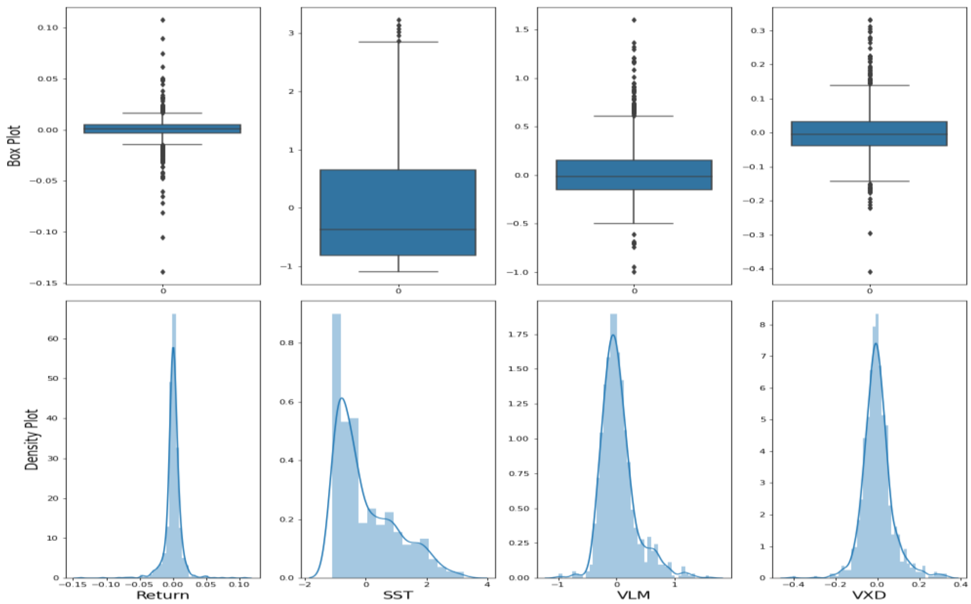


Figure 5.4: Box plot and histogram of index variables used in ST models. The figure shows box plot and histogram of each variable at daily frequency. The sample period is from 18 July 2016 to 28 August 2020.

In Figure 5.4, it can be observed that each variable has many outliers and especially the SST variable has a highly right-skewed distribution. These observations suggest that observations do not show characteristics of a normal distribution.

Table 5.12: Correlation matrix for ST model variables

This table shows correlations among the index variables at daily frequency. The variables that do not have a subscript represent the whole sample period. Lagged variables used in the models are also reported. $Return_{t-1}$ defines the 1-week lagged return.

	Return	SST	VLM	VXD
SST	-0.041			
VLM	-0.041	-0.051		
VXD	-0.566	0.006	0.126	
Return -1	-0.239	-0.006	-0.113	0.115

5.2.1. Impact of ST on Index Returns

The impact of ST on index returns is examined by estimating the following regression model. Results are reported in Table 5.13.

$$R_t = \beta_0 + \beta_1 SST_t + \beta_2 VLM_t + \beta_3 VXD_t$$

The effect of search traffic is negative and significant for the 0.10 quantile and positive and significant for the 0.90 quantile. In addition, it is insignificant in both the mean and median models. These results are similar to those of the SVI proxy model, showing that investor attention measured with either the ST or the SVI has a positive impact during bullish periods and negative influence during bearish periods. The major difference between these two proxies is that SVI has a larger impact under bullish conditions while ST has a larger effect under bearish conditions. Since these two proxies measure the same type of attention with different frequencies, it can be said that retail investor attention has a stronger effect during bearish conditions within a day but its effect is stronger during bullish conditions within a week. It is worth to note that trading volume has similar effects with retail attention across the quantiles, meaning that it has also asymmetric impact on index return, which is consistent with the Chen [5] study. Furthermore, similar to the SVI estimations, the DJI volatility index, VXD, has a negative and significant relationship with the index returns in all models.

Table 5.13: Impact of ST on DJI returns

The coefficient values are marked with significance levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Robust and bootstrapped standard errors are in parentheses for the conditional mean and quantile parameters, respectively. R^2 represents the pseudo R^2 for quantile regressions.

		Quantiles		
	Mean	0.10	0.50	0.90
Intercept	**0.001****	**-0.007*****	**0.0003***	**0.009*****
	(0.000)	(0.000)	(0.000)	(0.000)
SST	-0.001	**-0.004*****	-0.0002	**0.003*****
	(0.001)	(0.000)	(0.000)	(0.000)
VLM	-0.002	**-0.008*****	0.000	**0.007*****
	(0.002)	(0.001)	(0.001)	(0.001)
VXD	**-0.103*****	**-0.087*****	**-0.082*****	**-0.079*****
	(0.009)	(0.007)	(0.002)	(0.006)
R^2	0.323	0.368	0.233	0.221

5.2.2. Impact of Index Returns on ST

In order to moderate the strong correlation between index return and VXD, multiple regressions are estimated by increasing the lag of the return variable. The effect of return is significant and negative at the conditional median up to four days lagged return but its magnitude decreases as the lag increases. Estimation results of following regression where the dependent variable is the SST and the focus variable is the one day lagged index return are reported in Table 5.14.

$$SST_t = \beta_0 + \beta_1 R_{t-1} + \beta_2 VLM_t + \beta_3 VXD_t$$

The coefficient of the index return is significant only in the median model and it has a negative association with ST. In contrast, the index return had a positive and significant influence on SVI. In other words, retail investor attention increases as index returns increase at weekly frequency and increases as index returns decrease at daily frequency. This indicates that investors pay more attention to negative index returns on a daily basis while they pay more attention to positive returns at a weekly frequency, implying that retail investors respond to price reversals more quickly. These findings are consistent with Klemola, Nikkinen, and Peltomäki [19]. However, it should be noted that the models explain only a small fraction of the variation in attention at a daily frequency.

Table 5.14: Impact of DJI returns on ST

The coefficient values are marked with significance levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Robust and bootstrapped standard errors are in parentheses for the conditional mean and quantile parameters, respectively. R^2 represents the pseudo R^2 for quantile regressions.

		Quantiles		
	Mean	0.10	0.50	0.90
Intercept	0.008	**-0.998*** **	**-0.347*** **	**1.615*** **
	(0.031)	(0.017)	(0.036)	(0.081)
Return_{t-1}	-1.053	-0.507	**-8.822*** **	2.263
	(4.372)	(0.646)	(2.702)	(12.739)
VLM	-0.185	0.022	**-0.259*** **	0.225
	(0.119)	(0.058)	(0.119)	(0.282)
VXD	-0.211	0.068	0.166	0.094
	(0.471)	(0.271)	(0.495)	(1.270)
R^2	0.003	0.001	0.005	0.002

5.2.3. Visualization of the ST-DJI Models

Fitted lines for the focus explanatory variables are plotted and presented in Figure 5.5. In these plots, it is possible to observe the asymmetric effect of retail investor attention between the upper and lower quantiles. In addition, the linear association is considerably weak for the independent variable index return and dependent variable attention, which also validates that model explains only 0.5% variation in the ST.

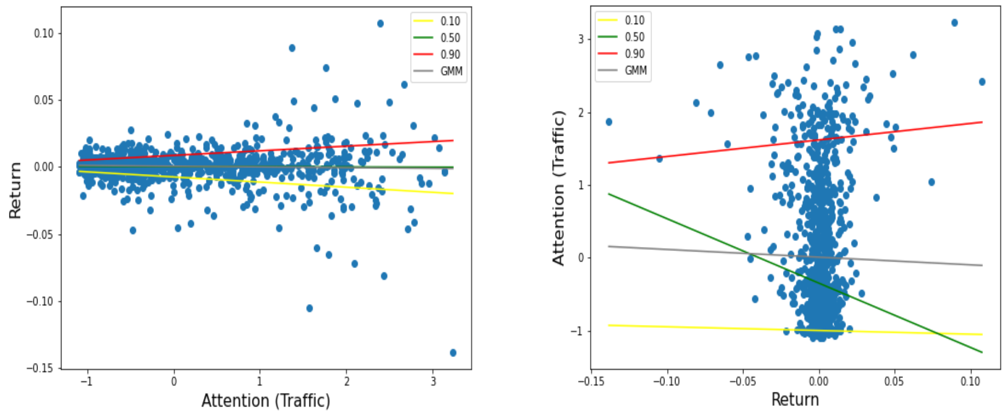


Figure 5.5: Scatter and fitted ST-DJI model plots. The figure shows the scatter plot and fitted lines for the estimated quantile regression models. Red lines represent 0.90, green lines represent 0.50, yellow lines represent 0.10 quantile regressions, and grey lines represent the GMM model fits as seen in the legends.

The coefficient changes for the focus variables are also plotted in Figure 5.6. Retail investor attention loses its effect rapidly until the 0.30 quantile and its effect changes steadily around the median. The growth of attention's effect accelerates above the 0.70 quantile. Hence, there is a clear asymmetric relation across the quantiles with greater shifts at tails for the response variable index return. On the other hand, the coefficient of index return is significant in a range of roughly 0.20 to 0.60 quantiles and the confidence envelope becomes wider dramatically above the median.

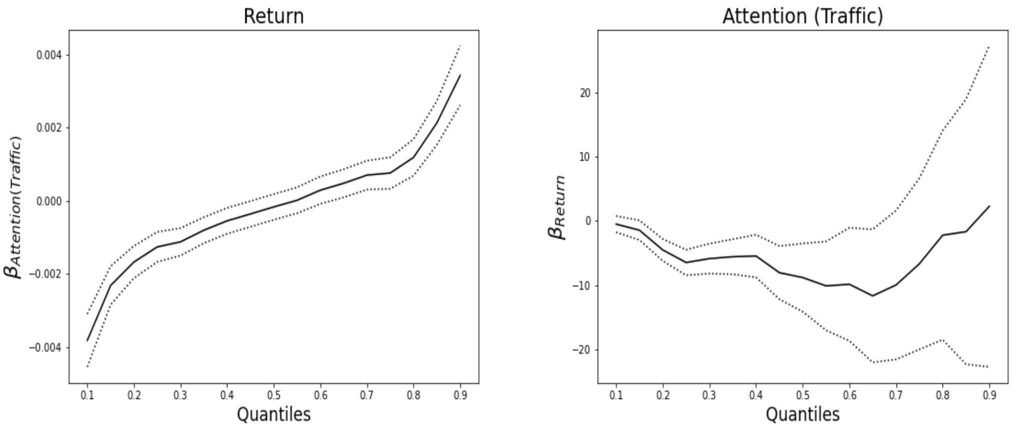


Figure 5.6: Coefficients across quantiles for ST-DJI models. The figure shows coefficient changes of ST-DJI models over the quantiles, ranging from 0.10 to 0.90 with their 5% confidence intervals. Titles of graphs indicate the dependent variable of the model.

5.2.4. Robustness for ST-DJI Models

Table 5.15 documents the robustness test results of index models when attention is measured with the search traffic. The effect of attention statistically differs across all the selected quantiles. Please note that search traffic has a negative effect in the 0.10 quantile and positive effect in the 0.90 quantile (Table 5.13). Test results in Table 5.15 confirm that this asymmetric effect is statistically significant. Finally, the effect of index returns on retail attention is significantly different only between the 0.10 quantile and the median.

Table 5.15: Slope equality tests for the ST-DJI models

The *p*-values of Wald tests are reported for slope equality test. *, **, and *** indicate that the null hypothesis is rejected at the 10%, 5%, and 1% significance levels, respectively. (1), (2) and (3) represent null hypotheses $\beta_1(\tau_{0.10}) = \beta_1(\tau_{0.50})$, $\beta_1(\tau_{0.50}) = \beta_1(\tau_{0.90})$ and $\beta_1(\tau_{0.10}) = \beta_1(\tau_{0.90})$, respectively.

	(1)	(2)	(3)
Dependent variable: Return Coefficient: SST	**0.000*** **	**0.000*** **	**0.000*** **
Dependent variable: SST Coefficient: Return	**0.000*** **	0.120	0.708

5.2.5. Impact of ST on Excess Returns of Constituent Stocks

The impact of ST on the excess returns of individual stocks is tested by estimating the following regression model. Results are reported in Table 5.16.

$$ER_{t,i} = \beta_{0,i} + \beta_{1,i}SST_{t,i} + \beta_{2,i}VLM_{t,i} + \beta_{3,i}VXD_t$$

At the 0.10 quantile, the coefficient of ST is significant and negative for 24 stocks (out of 27). On the contrary, at the 0.90 quantile, ST's coefficient is significant and positive for 24 stocks (out of 27). For the mean and median models, the relationship between retail attention and stock return is insignificant for the majority of the constituent stocks. These results are consistent with the corresponding model that uses SVI as an attention proxy. Consequently, ST has a robust relationship with the index return as well as the excess return of individual stocks and the relation is asymmetric across the quantiles for the majority of individual stocks.

Table 5.16: Impact of ST on excess returns of individual stocks

Coeff represents the coefficient of the SST variable in the conditional mean and different quantile models where the dependent variable is the individual stock's excess return. ***, **, * indicate levels of significance at 1%, 5%, and 10%, respectively. R^2 represents pseudo R^2 for quantile models.

			Quantiles					
	Mean		0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
AAPL	-0.001	0.028	**-0.001*****	0.113	0.000	0.015	**0.002*****	0.114
AXP	0.000	0.010	**-0.005*****	0.169	**-0.001*****	0.008	**0.005*****	0.098
BA	**-0.003****	0.029	**-0.012*****	0.292	**-0.002*****	0.017	**0.008*****	0.178
CAT	0.000	0.036	**-0.002*****	0.082	0.000	0.016	**0.001****	0.094
CSCO	0.000	0.013	**-0.002*****	0.067	0.000	0.013	**0.001*****	0.069

(*continued*)

Table 5.16–*Continued*

			Quantiles					
	Mean		0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
CVX	0.000	0.001	**-0.004*****	0.043	0.000	0.006	**0.003*****	0.074
DIS	0.000	0.010	**-0.004*****	0.112	0.000	0.010	**0.003*****	0.129
GS	-0.001	0.030	**-0.001*****	0.095	0.000	0.009	0.000	0.074
HD	0.000	0.007	0.000	0.049	**0.001***	0.006	**0.001****	0.027
IBM	0.000	0.006	**-0.002*****	0.091	0.000	0.002	**0.001*****	0.073
INTC	0.000	0.022	**-0.003*****	0.121	0.000	0.013	**0.003*****	0.102
JNJ	0.000	0.065	**-0.002*****	0.070	0.000	0.040	**0.003*****	0.137
JPM	-0.001	0.018	**-0.003*****	0.149	**-0.001*****	0.006	**0.003*****	0.122
KO	0.000	0.084	**-0.001****	0.068	0.000	0.040	**0.001*****	0.114
MCD	0.000	0.061	**-0.001****	0.117	0.000	0.035	**0.002*****	0.136
MMM	0.000	0.009	**-0.003*****	0.103	0.000	0.001	**0.003*****	0.083
MRK	0.000	0.031	**-0.002*****	0.064	**0.001***	0.020	**0.002*****	0.122
MSFT	0.000	0.041	**-0.004*****	0.113	**0.001***	0.023	**0.004*****	0.150
NKE	0.000	0.002	**0.001***	0.061	0.000	0.002	0.000	0.074
PFE	0.000	0.023	**-0.002*****	0.061	0.000	0.020	**0.002*****	0.089
PG	0.001	0.084	**-0.004*****	0.093	**0.001***	0.046	**0.004*****	0.160
TRV	0.000	0.009	**-0.002*****	0.049	0.000	0.022	0.001	0.049
UNH	0.000	0.004	**-0.003*****	0.076	0.000	0.003	**0.004*****	0.091
V	0.000	0.011	-0.001	0.043	0.000	0.003	**0.001*****	0.036
VZ	0.001	0.077	**-0.001****	0.097	0.000	0.030	**0.001*****	0.152
WMT	0.001	0.061	**-0.003*****	0.159	**0.001***	0.044	**0.004*****	0.219
XOM	0.000	0.002	**-0.004*****	0.089	**-0.001*****	0.005	**0.003*****	0.076

5.2.6. Impact of Excess Returns on ST

The relationship between ST and excess return of individual stocks is also tested in the reverse direction by estimating the following regression model. Results are reported in Table 5.17.

$$SST_t = \beta_0 + \beta_1 ER_{t-1} + \beta_2 VLM_t + \beta_3 VXD_t$$

As can be seen in Table 5.17, it is not possible to show any evidence that retail investor attention is significantly affected from the excess returns on a stock. This result is also consistent with the findings of the SVI model.

Table 5.17: Impact of excess returns on ST

Coeff represents the coefficient of the ER variable in the conditional mean and different quantile models where the dependent variable is SST. ***, **, * indicate levels of significance at 1%, 5%, and 10%, respectively. R^2 represents pseudo R^2 for quantile models.

			Quantiles					
	Mean		0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
AAPL	-3.421	0.004	-0.415	0.001	-2.623	0.001	**-10.388*****	0.015
AXP	-3.326	0.007	-1.036	0.000	-2.298	0.001	-4.334	0.024
BA	**-5.583***	0.067	**-0.846*****	0.003	**-3.738*****	0.010	-4.524	0.097
CAT	-1.358	0.001	0.501	0.000	0.233	0.001	0.249	0.003
CSCO	1.554	0.001	-0.250	0.002	1.220	0.002	5.749	0.009
CVX	-1.163	0.006	1.395	0.001	-2.969	0.002	-5.045	0.003
DIS	2.451	0.003	1.012	0.002	1.086	0.000	3.078	0.001
GS	-4.002	0.004	-1.698	0.002	-3.368	0.002	-3.881	0.005
HD	3.654	0.007	9.310	0.043	0.958	0.002	**8.653****	0.004
IBM	-1.206	0.008	-1.005	0.004	-1.112	0.001	-3.808	0.031
INTC	-0.414	0.005	1.109	0.000	-0.277	0.002	-0.399	0.018
JNJ	1.804	0.021	**9.031****	0.002	-0.828	0.001	0.776	0.070
JPM	-7.084	0.011	**-2.700****	0.002	**-3.993***	0.001	-7.904	0.021
KO	0.568	0.005	0.393	0.000	4.234	0.003	-1.350	0.012
MCD	1.071	0.009	3.546	0.008	**4.540***	0.005	1.311	0.009
MMM	1.373	0.003	0.091	0.000	-3.472	0.001	1.193	0.005
MRK	1.691	0.001	0.117	0.000	-0.801	0.002	-1.397	0.002
MSFT	3.167	0.016	0.016	0.000	0.972	0.001	4.638	0.020
NKE	0.958	0.006	2.505	0.005	-0.132	0.001	1.362	0.024
PFE	2.131	0.001	-0.738	0.008	-1.850	0.004	0.849	0.001
PG	3.827	0.019	0.089	0.001	3.601	0.001	2.020	0.007
TRV	-1.321	0.001	0.475	0.000	-3.433	0.001	-1.089	0.002
UNH	-0.999	0.004	0.059	0.000	0.528	0.000	-2.367	0.007
V	2.825	0.001	-1.534	0.000	4.109	0.002	11.570	0.003
VZ	3.583	0.002	**5.140***	0.003	2.190	0.001	-1.009	0.002
WMT	6.080	0.025	0.424	0.001	0.131	0.005	8.607	0.020
XOM	-1.710	0.044	-0.589	0.002	**-3.451***	0.005	-4.685	0.040

5.2.7. Robustness for ST-Stock Models

Table 5.18 reports the results of slope equality test for the model where the response variable is individual stocks’ excess returns. The coefficient estimates for retail attention are significantly different from each other across all tested quantiles for most of the constituent stocks. Hence, the test results are consistent with those of the SVI model. Please note that since excess returns do not have a significant effect on search traffic, tests of slope equality are not conducted for the model where attention is the dependent variable.

Table 5.18: Slope equality tests for the ST-Stock model

The *p*-values of Wald tests are reported for slope equality test. *, **, and *** indicate that the null hypothesis is rejected at the 10%, 5%, and 1% significance levels, respectively. (1), (2) and (3) represent null hypotheses $\beta_1(\tau_{0.10}) = \beta_1(\tau_{0.50})$, $\beta_1(\tau_{0.50}) = \beta_1(\tau_{0.90})$ and $\beta_1(\tau_{0.10}) = \beta_1(\tau_{0.90})$, respectively.

	(1)	(2)	(3)
AAPL	**0.027****	**0.001*****	**0.000*****
AXP	**0.000*****	**0.000*****	**0.000*****
BA	**0.000*****	**0.000*****	**0.000*****
CAT	**0.011****	**0.027****	**0.001*****
CSCO	**0.000*****	**0.000*****	**0.000*****
CVX	**0.000*****	**0.000*****	**0.000*****
DIS	**0.000*****	**0.000*****	**0.000*****
GS	**0.036****	0.333	**0.042****
HD	0.183	0.323	**0.087***
IBM	**0.000*****	**0.002*****	**0.000*****
INTC	**0.000*****	**0.000*****	**0.000*****
JNJ	**0.000*****	**0.000*****	**0.000*****
JPM	**0.000*****	**0.000*****	**0.000*****
KO	**0.006*****	**0.060***	**0.001*****
MCD	**0.004*****	**0.000*****	**0.000*****
MMM	**0.000*****	**0.000*****	**0.000*****
MRK	**0.003*****	**0.009*****	**0.000*****
MSFT	**0.000*****	**0.000*****	**0.000*****
NKE	0.684	0.952	0.715
PFE	**0.000*****	**0.000*****	**0.000*****
PG	**0.000*****	**0.000*****	**0.000*****
TRV	**0.006*****	0.190	**0.003*****
UNH	**0.004*****	**0.000*****	**0.000*****
V	**0.034****	**0.022****	**0.001*****
VZ	**0.006*****	**0.015****	**0.000*****
WMT	**0.000*****	**0.000*****	**0.000*****
XOM	**0.000*****	**0.000*****	**0.000*****

5.3. Click Per Search and Search Volume

In this section, two alternative attention proxies that are measured on a monthly basis are tested.. The first measure is the Search Volume (SV), which is equal to the average number of searches in a month on Google for a target keyword. The second is the Click Per Share (CPS) which is the ratio of clicks to search volume and it shows the average number of clicks in websites following the search for a target keyword. Since the SV measure is still based on the search activity on Google, the main variable of interest in this section is the CPS rather than the SV proxy. Table 5.19 shows the descriptive statistics of the index variables used in the CPS models. Figure 5.7 illustrates the box plot and histogram of index variables for visualizing their characteristics of distribution such as outliers and skewness. In addition, Table 5.20 reports the correlation matrix to display the correlation coefficients between variables.

Table 5.19: Descriptive statistics for CPS model variables

This table reports the descriptive statistics of monthly DJI returns, SCPS, SSV, VLM, Volatility and VXD. The sample period is from September 2015 to December 2019.

	Return	SCPS	SSV	VLM	Volatility	VXD
Mean	0.010	0.000	0.000	0.030	0.034	-0.013
Median	0.013	-0.372	-0.291	-0.021	0.028	-0.003
Standard Deviation	0.034	1.000	1.000	0.248	0.018	0.180
Kurtosis	1.636	-0.220	0.246	7.303	0.632	0.195
Skewness	-0.717	0.189	1.008	2.215	1.188	0.126
Range	0.184	4.731	3.735	1.430	0.072	0.838
Minimum	-0.102	-2.735	-1.072	-0.337	0.014	-0.360
Maximum	0.082	1.997	2.664	1.093	0.086	0.478
Jarque-Bera	8.264	0.503	8.322	132	11.978	0.139
Observations	52	52	52	52	52	52

In Table 5.20, CPS and SV are highly and negatively correlated, implying that they measure different types of attention and the interaction is strong. As seen from Figure 5.7, the variables do not have many outlier observations but they have highly skewed distributions.

Table 5.20: Correlation matrix for CPS model variables

This table shows correlations among index variables at monthly frequency.

	SCPS	SSV	Return	VXD	Volatility
SSV	-0.628				
Return	-0.015	-0.326			
VXD	-0.060	0.092	-0.685		
Volatility	-0.006	0.683	-0.545	0.253	
VLM	0.114	-0.011	-0.136	0.246	0.133

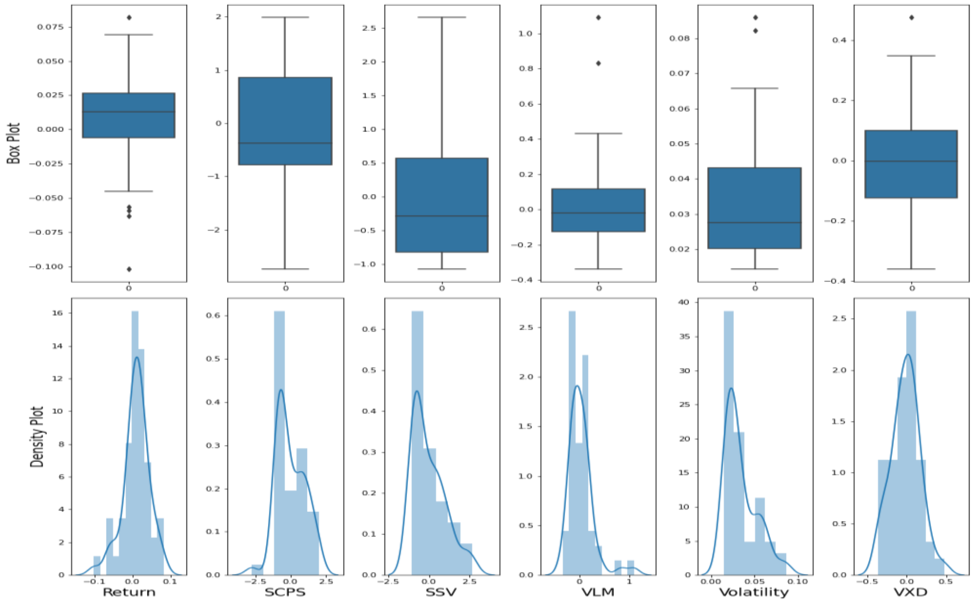


Figure 5.7: Box plot and histogram of index variables used in CPS models. The figure shows the box plot and histogram of each variable at a monthly frequency. The sample period is from September 2015 to December 2019.

5.3.1. Impact of CPS on Index Returns

SSV is not included in the model where the dependent variable is the index return since the two attention proxies are highly correlated. Thus, the following regression model is estimated to test the relationship between DJI returns and investor attention where the attention is proxied by the Click Per Share (CPS) measure.

$$R_t = \beta_0 + \beta_1 SCPS_t + \beta_2 VLM_t + \beta_3 Volatility_t + \beta_4 VXD_t$$

As seen in Table 5.21, regardless of the model, there is no evidence of a significant relationship between the SCPS variable and index returns. On the other hand, similar to the previously estimated SVI-DJI and ST-DJI models, investor sentiment, which is measured by the volatility index VXD, has a significant and negative impact on index returns in all models. Investor sentiment presents symmetric effects across quantiles, implying that index returns increase as VXD decreases regardless of the market's condition. The Pseudo R^2 statistics are considerably large with a tendency of a decrease as the quantile level increases.

Table 5.21: Impact of CPS on DJI returns

The coefficient values are marked with significance levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Robust and bootstrapped standard errors are in parentheses for the conditional mean and quantile parameters, respectively. R^2 represents the pseudo R^2 for quantile regressions.

		Quantiles		
	Mean	0.1	0.5	0.9
Intercept	**0.034*****	0.005	**0.027*****	**0.071*****
	(0.007)	(0.010)	(0.007)	(0.014)
SCPS	-0.002	0.002	-0.002	-0.006
	(0.003)	(0.005)	(0.003)	(0.008)
VLM	0.010	-0.007	0.010	**0.078*****
	(0.013)	(0.011)	(0.013)	(0.019)
Volatility	**-0.765*****	**-0.672****	**-0.606*****	**-0.944****
	(0.174)	(0.316)	(0.180)	(0.371)
VXD	**-0.116*****	**-0.139*****	**-0.113*****	**-0.143****
	(0.021)	(0.030)	(0.019)	(0.060)
R^2	0.593	0.580	0.398	0.331

5.3.2. Impact of CPS on Index Volatility

In Table 5.20, it is seen that the VXD and return variables are highly correlated. Therefore, for this part of the analysis, the VXD variable is dropped from the model in order to avoid multicollinearity. The following regression model is estimated to examine whether CPS has a significant impact on the DJI volatility. Results are presented in Table 5.22.

$$Volatility_t = \beta_0 + \beta_1 SCPS_t + \beta_2 R_t + \beta_3 VLM_t$$

Table 5.22: Impact of CPS on DJI volatility

The coefficient values are marked with significance levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Robust and bootstrapped standard errors are in parentheses for the conditional mean and quantile parameters, respectively. R^2 represents the pseudo R^2 for quantile regressions.

		Quantiles		
	Mean	0.1	0.5	0.9
Intercept	**0.036*****	**0.019*****	**0.031*****	**0.055*****
	(0.002)	(0.003)	(0.004)	(0.004)
SCPS	0.000	-0.002	0.000	-0.004
	(0.002)	(0.001)	(0.003)	(0.005)
Return	**-0.282*****	**-0.130*****	**-0.229****	-0.189
	(0.072)	(0.048)	(0.100)	(0.147)
VLM	0.005	0.002	-0.003	**0.049*****
	(0.008)	(0.006)	(0.014)	(0.008)
R^2	0.257	0.010	0.096	0.301

As can be seen in Table 5.22, SCPS is not significant in any of the models and R^2 statistics are quite small, implying that CPS and volatility are statistically not related at a monthly frequency and the models explain only a small proportion of the variation in volatility.

5.3.3. Impact of Index Returns on CPS

The question of whether index returns affect investor attention is studied by estimating the following model. Please note that the volatility variable is excluded from the model since volatility and SSV are highly correlated. Findings are reported in Table 5.23.

$$SCPS_t = \beta_0 + \beta_1 R_t + \beta_2 SSV_t + \beta_3 VLM_t$$

Return is significant only in the conditional mean model but there is no significant relationship between monthly index returns and Click Per Share when the model is estimated for the quantiles separately. In addition, SV has a significant and negative relationship with CPS in all four models. This finding implies that the more people search for a specific keyword, the less they click on websites per searched keyword.

Table 5.23: Impact of DJI returns on CPS

The coefficient values are marked with significance levels. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. Robust and bootstrapped standard errors are in parentheses for the conditional mean and quantile parameters. R^2 represents the pseudo R^2 for quantile regressions.

		Quantiles		
	Mean	0.1	0.5	0.9
Intercept	0.056	**-0.707*****	-0.050	**1.078*****
	(0.099)	(0.128)	(0.150)	(0.194)
Return	**-6.798***	-4.678	-6.210	-3.673
	(3.713)	(3.641)	(4.443)	(9.119)
SSV	**-0.704*****	**-0.518*****	**-0.585*****	**-0.846*****
	(0.091)	(0.078)	(0.151)	(0.296)
VLM	0.300	0.174	0.441	**1.726*****
	(0.210)	(0.430)	(0.584)	(0.307)
R^2	0.420	0.139	0.342	0.303

5.3.4. Visualization of CPS-DJI Models

Figure 5.8 plots the fitted lines for the conditional mean and 0.10, 0.50, and 0.90 quantile levels. In addition, Figure 5.9 presents the graphical views of the coefficients across quantiles which makes it possible to observe the location and shape shifts across the quantiles. The insignificant relationships between CPS and returns, and also CPS and volatility are once again observed in Figure 5.9 since the confidence envelope always involves the horizontal 0 line. For the model where CPS is the dependent variable, there is weak evidence that return has a negative effect on CPS between the 0.20 and 0.30 quantiles.

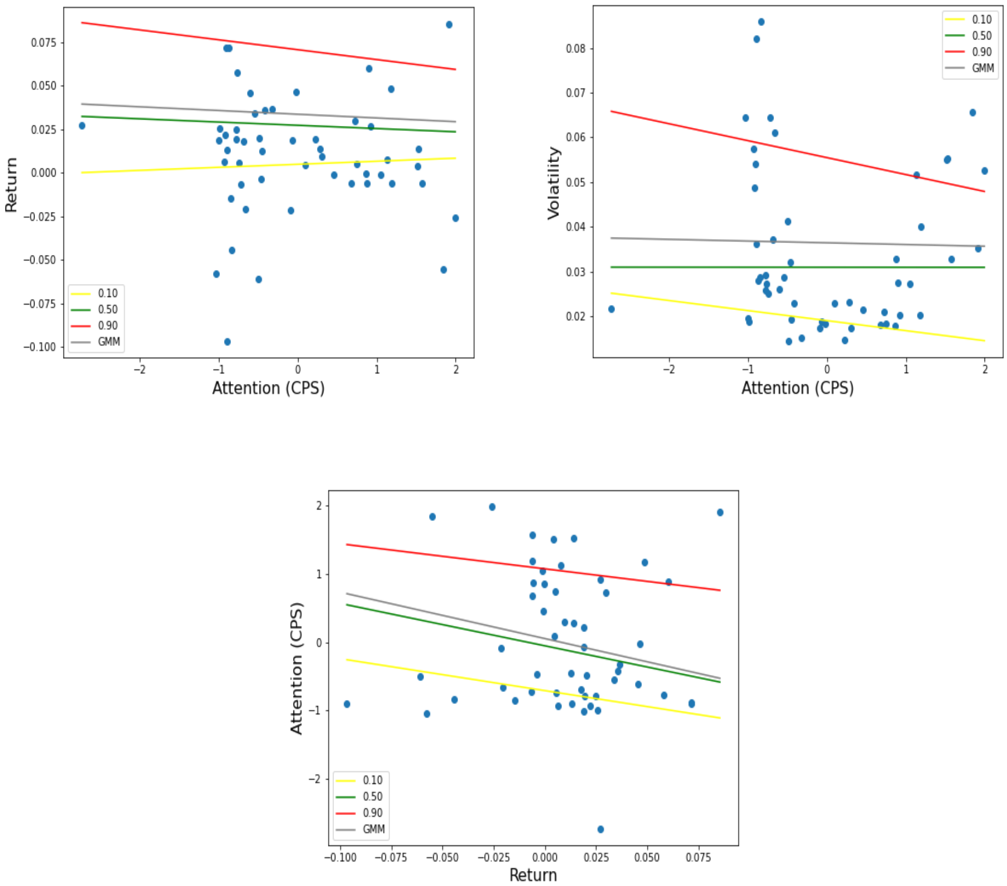


Figure 5.8: Scatter and fitted CPS-DJI model plots. The figure shows the scatter plot and fitted lines for the estimated quantile regression models. Red lines represent 0.90, green lines represent 0.50, yellow lines represent 0.10 quantile regressions, and grey lines represent the GMM model fits as seen in the legends.

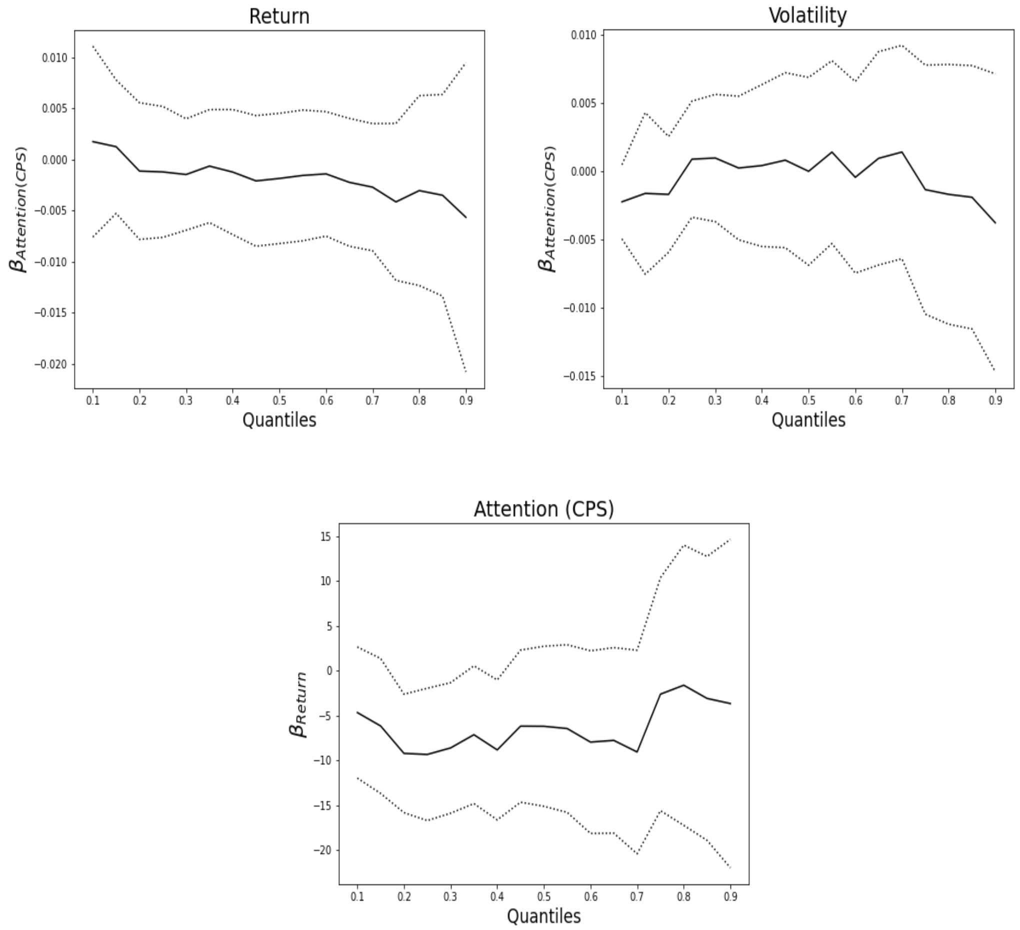


Figure 5.9: Coefficients across quantiles for CPS-DJI models. The figure shows coefficient changes of CPS-DJI models over the quantiles, ranging from 0.10 to 0.90 with their 5% confidence intervals. Titles of graphs indicate the dependent variable of the model.

5.3.5. Impact of CPS on Excess Returns of Constituent Stocks

In this part of the analysis, the relationship between retail investor attention, as proxied by the CPS, and returns is examined individually for the DJI's constituent stocks. The following regression model is estimated and results are reported in Table 5.24.

$$ER_{t,i} = \beta_{0,i} + \beta_{1,i}SCPS_{t,i} + \beta_{2,i}VLM_{t,i} + \beta_{3,i}Volatility_{t,i} + \beta_{4,i}VXD_t$$

Similar to the results for the index returns, there is no consistent evidence that there is a significant relationship between the CPS proxy of investor attention and individual stock returns. Also, the R^2 values are typically smaller relative to those of the DJI model so predicted lines fit better in market index model rather than the individual stock models.

Table 5.24: Impact of CPS on excess return of individual stocks

Coeff represents the coefficient of the SPCS variable in the conditional mean and different quantile models where the dependent variable is excess return. ***, **, * indicate levels of significance at 1%, 5%, and 10%, respectively. R^2 represents pseudo R^2 for quantile models.

			Quantiles					
	Mean		0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
AAPL	-0.006	0.093	**-0.038***	0.244	-0.009	0.065	0.003	0.025
AXP	-0.004	0.178	-0.001	0.336	0.004	0.065	-0.001	0.031
BA	0.004	0.039	-0.003	0.284	-0.006	0.034	-0.009	0.157
CAT	-0.008	0.213	-0.011	0.21	-0.005	0.17	-0.009	0.172
CSCO	-0.001	0.063	**-0.029***	0.351	-0.005	0.036	0.005	0.124
CVX	-0.003	0.036	-0.016	0.08	-0.004	0.019	0	0.182
DIS	-0.005	0.158	0.001	0.148	0.001	0.048	-0.011	0.272
GS	0.001	0.023	-0.007	0.143	0	0.056	0.011	0.049
HD	0.003	0.08	-0.001	0.264	0.002	0.006	0.011	0.075
IBM	-0.003	0.162	-0.009	0.351	0.001	0.025	0.004	0.246
INTC	-0.007	0.033	**0.016****	0.036	0	0.025	-0.008	0.053
JNJ	**0.008***	0.101	0.015	0.204	0.005	0.108	0.004	0.158
JPM	-0.003	0.172	0.001	0.188	-0.012	0.144	-0.001	0.032
KO	0.001	0.194	-0.005	0.245	0	0.113	-0.005	0.276
MCD	0.006	0.258	-0.003	0.237	0.009	0.079	0.001	0.215
MMM	**-0.015*****	0.237	-0.006	0.261	**-0.012***	0.141	-0.014	0.127
MRK	-0.006	0.067	0	0.249	-0.013	0.028	-0.014	0.189
MSFT	**0.010****	0.072	**0.027****	0.09	0.006	0.084	0.008	0.174
NKE	0.001	0.122	-0.006	0.058	-0.001	0.053	**0.014***	0.245
PFE	-0.01	0.169	0.002	0.146	-0.011	0.098	-0.009	0.153
PG	0.003	0.205	-0.012	0.202	0.001	0.155	0.005	0.424
TRV	-0.006	0.218	-0.008	0.253	-0.004	0.108	-0.006	0.072
UNH	0.001	0.039	**0.020*****	0.244	0.005	0.027	-0.004	0.156
V	-0.006	0.218	**-0.019****	0.305	-0.004	0.069	-0.01	0.148
VZ	-0.005	0.141	-0.005	0.142	0.004	0.051	-0.012	0.327
WMT	-0.005	0.302	-0.004	0.371	0	0.199	0.01	0.365
XOM	-0.001	0.06	0.011	0.109	-0.004	0.047	-0.007	0.077

5.3.6. Impact of CPS on Stock Return Volatilities

The following regression model is estimated to investigate whether CPS has an impact on the volatility of individual stock returns. Estimations are presented in Table 5.25.

$$Volatility_{t,i} = \beta_{0,i} + \beta_{1,i}SCPS_{t,i} + \beta_{2,i}ER_{t,i} + \beta_{3,i}VLM_{t,i} + \beta_{4,i}VXD_t$$

As seen in the table, it is not possible to present any statistical evidence that CPS has an impact on the volatility of individual stock returns since coefficient significance is very infrequent and rather sporadic across the models.

Table 5.25: Impact of CPS on volatility of individual stocks

Coeff represents the coefficient of the SPCS variable in the conditional mean and different quantile models where the dependent variable is volatility. ***, **, * indicate levels of significance at 1%, 5%, and 10%, respectively. R^2 represents pseudo R^2 for quantile models.

			Quantiles					
	Mean		0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
AAPL	0.004	0.261	0.002	0.208	0.005	0.059	0.005	0.268
AXP	**0.007*****	0.265	0.002	0.250	**0.009*****	0.089	**0.013***	0.079
BA	**-0.005***	0.315	-0.006	0.234	-0.004	0.102	-0.006	0.068
CAT	-0.003	0.196	-0.005	0.225	-0.003	0.163	-0.009	0.200
CSCO	-0.001	0.354	-0.001	0.211	0.001	0.048	-0.004	0.352
CVX	-0.001	0.128	-0.001	0.087	-0.002	0.021	**0.010****	0.207
DIS	**0.004****	0.315	**-0.004***	0.132	0.003	0.066	0.003	0.312
GS	0.001	0.282	-0.002	0.183	0.000	0.023	0.000	0.053
HD	-0.004	0.257	-0.004	0.253	-0.003	0.033	-0.008	0.116
IBM	-0.001	0.128	-0.001	0.440	-0.002	0.024	-0.008	0.192
INTC	-0.006	0.220	-0.003	0.064	-0.004	0.036	-0.008	0.061
JNJ	-0.001	0.378	-0.001	0.188	-0.003	0.074	-0.001	0.141
JPM	0.000	0.285	0.000	0.173	**0.006***	0.068	-0.002	0.053
KO	0.002	0.120	0.000	0.120	0.004	0.119	0.000	0.264
MCD	**0.005****	0.320	0.006	0.226	**0.006****	0.068	**0.007***	0.224
MMM	**-0.011*****	0.392	**-0.007*****	0.157	**-0.010*****	0.259	**-0.015****	0.221
MRK	0.003	0.255	0.002	0.120	0.001	0.025	0.002	0.196
MSFT	**0.007****	0.432	0.003	0.134	0.004	0.085	0.009	0.156
NKE	0.000	0.419	0.004	0.037	0.000	0.040	-0.004	0.209
PFE	0.003	0.457	0.004	0.056	**0.006****	0.109	0.002	0.183
PG	**-0.004***	0.261	-0.003	0.087	-0.003	0.173	-0.006	0.431
TRV	-0.001	0.368	0.001	0.128	-0.001	0.079	-0.004	0.031
UNH	**-0.006****	0.312	-0.001	0.226	**-0.010*****	0.033	-0.003	0.133
V	-0.001	0.302	0.000	0.272	0.001	0.072	-0.006	0.133
VZ	0.004	0.190	0.002	0.076	0.002	0.050	0.010	**0.286****
WMT	0.000	0.433	0.001	0.294	-0.001	0.175	-0.010	0.270
XOM	**0.006*****	0.324	0.006	0.128	**0.008****	0.098	0.001	0.144

5.3.7. Impact of Excess Stock Returns on CPS

The following regression model is estimated to test whether the returns generated by individual stocks affect the CPS as an attention proxy. Results are presented in Table 5.26.

$$SCPS_{t,i} = \beta_{0,i} + \beta_{1,i}ER_{t,i} + \beta_{2,i}SSV_{t,i} + \beta_{3,i}VLM_{t,i} + \beta_{4,i}VXD_t$$

The findings are similar to those of the previous models. There is no consistent statistical evidence that excess stock returns have a significant effect on the CPS. It should be remembered that CPS and SV have a strong negative correlation. Hence, CPS may have the potential to have a significant effect on individual stock returns or return volatilities but it is not possible to identify this relationship when data are examined on a monthly frequency.

Table 5.26: Impact of excess returns on CPS

Coeff represents the coefficient of the ER variable in the conditional mean and different quantile models where the dependent variable is SCPS. ***, **, * indicate levels of significance at 1%, 5%, and 10%, respectively. R^2 represents pseudo R^2 for quantile models.

			Quantiles					
	Mean		0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
AAPL	-2.143	0.067	-1.889	0.119	-1.371	0.061	-3.661	0.088
AXP	-3.127	0.269	1.112	0.145	-3.368	0.161	-7.238	0.227
BA	-0.796	0.176	2.126	0.223	-3.430	0.156	-0.568	0.127
CAT	-2.393	0.261	1.877	0.053	**-5.303***	0.176	-2.965	0.231
CSCO	0.044	0.181	-3.745	0.220	0.795	0.080	3.837	0.178
CVX	-2.676	0.126	-5.673	0.178	-4.416	0.085	1.937	0.066
DIS	-0.064	0.010	0.092	0.065	0.695	0.015	-7.066	0.035
GS	0.440	0.362	**5.621***	0.256	1.065	0.305	-0.742	0.337
HD	3.011	0.249	4.486	0.098	4.677	0.170	-1.901	0.182
IBM	-1.124	0.022	4.257	0.103	-1.970	0.023	-1.410	0.112
INTC	-2.890	0.185	**-7.851***	0.089	-3.420	0.182	2.013	0.135
JNJ	**5.253***	0.201	3.804	0.121	**9.639***	0.143	-1.725	0.115
JPM	-2.766	0.183	**-9.780*****	0.253	-0.793	0.090	-3.836	0.134
KO	0.119	0.011	-1.275	0.024	-1.385	0.056	7.400	0.084
MCD	3.450	0.131	-1.732	0.108	-0.544	0.050	**12.179*****	0.243
MMM	**-7.812****	0.147	-4.333	0.139	-5.255	0.073	**-10.581***	0.140
MRK	-2.055	0.086	-0.512	0.107	-0.755	0.079	-6.353	0.111
MSFT	5.721	0.106	**10.174****	0.113	1.351	0.069	5.279	0.095
NKE	-0.148	0.178	0.893	0.146	2.081	0.097	-1.634	0.144

(*continued*)

Table 5.26–*Continued*

	Mean		Quantiles					
			0.10		0.50		0.90	
	Coeff	R^2	Coeff	R^2	Coeff	R^2	Coeff	R^2
PFE	**-6.249***	0.111	-6.943	0.087	**-9.229****	0.091	-0.790	0.039
PG	0.926	0.030	-1.949	0.055	2.182	0.032	-3.604	0.078
TRV	-3.777	0.103	-11.131	0.144	1.048	0.068	2.853	0.147
UNH	-1.368	0.409	1.611	0.171	-1.853	0.351	4.402	0.236
V	-2.239	0.606	**-11.674*****	0.336	-3.500	0.439	1.552	0.385
VZ	-1.641	0.137	-3.367	0.133	-4.733	0.084	0.984	0.158
WMT	-2.376	0.081	-2.839	0.043	-0.226	0.027	-1.480	0.147
XOM	-0.245	0.052	-1.373	0.208	2.427	0.020	-6.240	0.044

CHAPTER 6

CONCLUSION

This study investigates the relationship between retail investor attention and stock market returns and volatilities by using a new approach, Search Engine Optimization (SEO). These relationships are analyzed for the returns and volatilities of the Dow Jones Index (DJI) as well as its constituent stocks individually. The main contribution of the study is its use of the SEO method to build the attention proxies used in the models. The SEO methodology is argued to generate superior attention measures in comparison to the traditionally used proxies in the literature since this method makes it possible to identify website traffic loads and search results in search engines by maximizing and optimizing quality and quantity of measures. Accordingly, the study proposes two new attention proxies by taking advantage of the functionality of SEO with various tools. A new and direct measure, Search Traffic, that refers to how much traffic a particular website gets from search results is proposed as the first alternative attention measure. With this proxy, it is possible to measure investor attention based on financial website URLs without having to specify any search keywords. As such, the Search Traffic measure provides a robust alternative to the traditionally used Google SVI, which may suffer from problems related to selecting the comprehensive set of relevant keywords. As a second alternative proxy, the Click Per Search (CPS) measure is proposed. CPS shows the average number of clicks in websites following the search for a keyword and may provide a more “active” attention measure since it identifies whether a keyword search is followed by a further search of information on a website. In addition to the newly proposed measure, the traditional Google SVI proxy is constructed by using the SEO methodology which helps to optimize the number of keywords that should be included in the calculation of SVI. Therefore, the Google SVI proxy used in this study is a broader version of the traditionally calculated measure that is used in the existing literature.

The effect of different investor attention proxies on stock returns and volatilities is estimated by employing the Generalized Method of Moments methodology for the conditional mean and the Quantile Regression methodology to observe the potentially changing nature of the relationship at different quantile levels of the dependent variable. Findings suggest that retail investor attention has significant and asymmetric effects for both the index return and the excess returns of individual stocks. Specifically, investor attention affects the index return as well as the individual stock returns negatively during periods of bearish market conditions and positively during periods of bullish market conditions.

The effect of investor attention also changes when attention and returns are measured with different frequencies. When attention and returns are calculated on a weekly basis, attention has a relatively larger effect on positive index returns compared to negative returns. Conversely, when attention and returns are calculated on a daily basis, attention has a relatively larger effect on negative index returns compared to positive returns. In other words, an increase in investor attention increases positive index returns in the longer term, but it decreases negative index returns in the shorter term.

Results also suggest that volatility is significantly influenced by retail investor attention for both the index as well as its individual constituent stocks. An increase in attention predicts higher volatilities across all quantiles and its growth of impact accelerates as the quantile level increases, confirming that attention, volatility and uncertainty are positively related.

It is also found that, when the relationship is analyzed from the reverse angle, negative index returns have more instant impact on attention compared to positive returns. In other words, investors pay more attention to the market when the overall market, but not individual stocks, is bearish, rather than bullish. Whereas, returns have less impacts on retail investors' attention compared to influence of retail attention on returns.

Estimation results provide no evidence of a significant relationship between Click Per Search and stock returns or volatilities. Interestingly, Click Per Search is negatively and strongly correlated with Search Volume. This implies that the more people search for a financial keyword, the less they click on websites per searched keyword. Thus, Click Per Search may have a potential to explain some other stock market parameter statistically even though it is not statistically related to return and volatility.

Overall, retail investor attention has a significant impact on both return and volatility for both index and individual stocks and it has considerable potential to forecast future index returns, excess returns, and volatilities.

REFERENCES

[1] Audrino, Francesco, Fabio Sigrist, and Daniele Ballinari, 2020, The impact of sentiment and attention measures on stock market volatility, International Journal of Forecasting 36, 334–357.

[2] Barber, Brad M., and Terrance Odean, 2008, All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, Review of Financial Studies 21, 785–818.

[3] Ben-Rephael, Azi, Zhi Da, and Ryan D. Israelsen, 2017, It depends on where you search: Institutional investor attention and underreaction to news, Review of Financial Studies 30, 3009–3047.

[4] Bijl, Laurens, Glenn Kringhaug, Peter Molnár, and Eirik Sandvik, 2016, Google searches and stock returns, International Review of Financial Analysis 45, 150–156.

[5] Chen, Shiu-Sheng, 2012, Revisiting the empirical linkages between stock returns and trading volume, Journal of Banking & Finance 36, 1781–1788.

[6] Chen, Tao, 2017, Investor attention and global stock returns, Journal of Behavioral Finance 18, 358–372.

[7] Chordia, Tarun, and Bhaskaran Swaminathan, 2000, Trading volume and cross-autocorrelations in stock returns, Journal of Finance 55, 913–935.

[8] Da, Zhi, Joseph Engelberg, and Pengjie Gao, 2011, In search of attention, Journal of Finance 66, 1461–1499.

[9] Fama, Eugene F., 1965, The behavior of stock-market prices, Journal of Business 38, 34–105.

[10] Fama, Eugene F., 1970, Efficient capital markets: A review of theory and empirical work, Journal of Finance 25, 383–417.

[11] Focke, Florens, Stefan Ruenzi, and Michael Ungeheuer, 2020, Advertising, attention, and financial markets, Review of Financial Studies 33, 4676–4720.

[12] Gervais, Simon, Ron Kaniel, and Dan H. Mingelgrin, 2001, The high-volume return premium, Journal of Finance 56, 877–919.

[13] Hautsch, Nikolaus, and Dieter Hess, 2007, Bayesian learning in financial markets: Testing for the relevance of information precision in price discovery, Journal of Financial & Quantitative Analysis 42, 189–208.

[14] Herwartz, Helmut, and Fang Xu, 2021, Structural transmissions among investor attention, stock market volatility and trading volumes, European Financial Management 28, 260–279.

[15] Hsieh, Shu-Fan, Chia-Ying Chan, and Ming-Chun Wang, 2020, Retail investor attention and herding behavior, Journal of Empirical Finance 59, 109–132.

[16] Huang, Melody Y., Randall R. Rojas, and Patrick D. Convery, 2020, Forecasting stock market movements using Google Trend searches, Empirical Economics 59, 2821–2839.

[17] Joseph, Kissan, M. Babajide Wintoki, and Zelin Zhang, 2011, Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from an online search, International Journal of Forecasting 27, 1116–1127.

[18] Kahneman, Daniel, 1973, *Attention and Effort* (Prentice-Hall, Englewood Cliffs, NJ).

[19] Klemola, Antti, Jussi Nikkinen, and Jarkko Peltomäki, 2016, Changes in investors' market attention and near-term stock market returns, Journal of Behavioral Finance 17, 18–30.

[20] Koenker, Roger, and Gilbert Bassett, 1978, Regression quantiles, Econometrica 46, 33–50.

[21] Li, Ruihai, Xuewu (Wesley) Wang, Zhipeng Yan, and Yan Zhao, 2019, Sophisticated investor attention and market reaction to earnings announcements: Evidence from the SEC's EDGAR log files, Journal of Behavioral Finance 20, 490–503.

[22] Li, Xian, James A. Hendler, and John L. Teall, 2016, Investor attention on the social web, Journal of Behavioral Finance 17, 45–59.

[23] Lin, Mei-Chen, and Yu-ZhenWu, 2019, Interest conflicts, stock recommendations and investor attention, International Review of Financial Analysis 65, 101054.

[24] Mbanga, Cedric, Ali F. Darrat, and Jung Chul Park, 2019, Investor sentiment and aggregate stock returns: The role of investor attention, Review of Quantitative Finance and Accounting 53, 397–428.

[25] Merton, Robert C., 1987, A simple model of capital market equilibrium with incomplete information, Journal of Finance 42, 483–510.

[26] Mittnik, Stefan, Nikolay Robinzonov, and Martin Spindler, 2015, Stock market volatility: Identifying major drivers and the nature of their impact, Journal of Banking & Finance 58, 1–14.

[27] Padungsaksawasdi, Chaiyuth, Sirimon Treepongkaruna, and Robert Brooks, 2019, Investor attention and stock market activities: New evidence from panel data, International Journal of Financial Studies 7, 30.

[28] Pastor, Lubos, and Pietro Veronesi, 2009, Learning in financial markets, Annual Review of Financial Economics 1, 361–381.

[29] Reyes, Tomas, 2019, Negativity bias in attention allocation: Retail investors' reaction to stock returns, International Review of Finance 19, 155–189.

[30] Subramaniam, Sowmya, and Madhumita Chakraborty, 2020, Investor attention and cryptocurrency returns: Evidence from quantile causality approach, Journal of Behavioral Finance 21, 103–115.

[31] Swamy, Vighneswara, and M. Dharani, 2020, Google search intensity and the investor attention effect: A quantile regression approach, Journal of Quantitative Economics 18, 403–423.

[32] Swamy, Vighneswara, M. Dharani, and Fumiko Takeda, 2019, Investor attention and Google Search Volume Index: Evidence from an emerging market using quantile regression analysis, Research in International Business and Finance 50, 1–17.

[33] Vozlyublennaia, Nadia, 2014, Investor attention, index performance, and return predictability, Journal of Banking & Finance 41, 17–35.

[34] Wang, Changyun, 2003, Investor sentiment, market timing and futures returns, Applied Financial Economics 13, 891–898.

[35] Wang, Yaw-Huei, Aneel Keswani, and Stephen J. Taylor, 2006, The relationships between sentiment, returns and volatility, International Journal of Forecasting 22, 109–123.

[36] Zhang, Wei, and Pengfei Wang, 2020, Investor attention and the pricing of cryptocurrency market, Evolutionary and Institutional Economics Review 17, 445–468.

APPENDIX

KEYWORDS FOR DOW JONES INDEX

This table reports the list of keywords related to "Dow Jones Index" used in the construction of CPS, SV, and SVI measures.

Keywords		
dow jones	djia chart	dow jones today chart
dow	dow live	dow.jones
djia	current dow jones	dow jones chart today
dow jones today	current dow	djia now
dow jones live	djia live	dow.jones today
dow jones index	dow industrial	dow stock market
dow jones industrial average	dow chart	dow average
djia today	dow now	todays dow
dowjones	the dow today	dow stock price
dow today	dow 30	dow jones close today
dow jones chart	live dow jones	dow today live
dow index	dow jones real time	djia stock price
dow jones industrial	dow jones live ticker	dow ticker
down jones	dow jones stock price	dow industrials
dow jones now	dow jones stock market	dow jones daily
dow jones average	dow jones historical data	dow close today
dowjones today	dow jones ticker	the dow
dow jones stock price today	dow jones industrial today	today's dow jones
dowjones index	dow jones average today	dow industrial average
dow jones live update	dow jones today chart live	stock market today dow
dow jones stock market today	how is the dow doing today	what is the dow doing today
dow jones industrial average for today	dow jones industrial average today	how is the dow jones doing today
current dow jones average		