Journal of
Risk and Financial Management

MDPI

Article

A Collection of Wisdom in Predicting Sector Returns: The Use of Google Search Volume Index

Hsiu-lang Chen [1,*] and Jolana Stejskalova [2,†]

1 Department of Finance, University of Illinois Chicago, 601 South Morgan Street, Chicago, IL 60607, USA
2 Department of Finance, Mendel University in Brno, 613 00 Brno, Czech Republic; jolana.stejskalova@pef.mendelu.cz
* Correspondence: hsiulang@uic.edu; Tel.: +1-312-355-1024
† For helpful comments, we thank participants at the conference, World Finance & Banking Symposium in Budapest, Hungary, 17–18 December 2021; ITISE 2022 (8th International conference on Time Series and Forecasting), Gran Canaria, Spain, 27–30 June 2022; the 12th Biennial Conference of the Czech Economic Society in Prague, Czech Republic, 25–26 November 2022; and World Finance Conference in in Kristiansand, Norway, 2–4 August 2023.

Abstract: This study investigates whether the aggregate investor information demand for all stocks in a sector demonstrated in the Google search volume index (SVI) can predict the sector's performance. The evidence shows that a sector's abnormal SVI can predict the sector's performance next month, even after controlling for the sector's contemporaneous standardized unexpected earnings and lagged returns on both the market and the sector. Also found is a partial reversal in the sector's long-run performance that is not completely consistent with the price pressure hypothesis. This indicates that some fundamental information about a sector can be captured by the sector's abnormal SVI on a timely basis.

Keywords: Google search volume index; S&P 500 sector indices; standardized unexpected earnings (SUEs)

JEL Classification: G10; G23

check for updates

Citation: Chen, Hsiu-lang, and Jolana Stejskalova. 2024. A Collection of Wisdom in Predicting Sector Returns: The Use of Google Search Volume Index. *Journal of Risk and Financial Management* 17: 452. https://doi.org/10.3390/jrfm17100452

Academic Editor: Marius Sikveland

Received: 19 August 2024
Revised: 30 September 2024
Accepted: 2 October 2024
Published: 5 October 2024

1. Introduction

Asset managers are in the business of acquiring information that they use to manage a portfolio of assets. Sector or industry information is one of the important inputs. Kacperczyk et al. (2005) present evidence that fund managers can create value by concentrating on industries in which they have an information advantage.[1] How to effectively gain information and insight about sectors or industries is an important but challenging task for investors. Our research examines the role of internet search to gain information about sector prospects.

There is a huge amount of information available today that demonstrates the need to allocate our attention efficiently.[2] Search reveals attention, and internet users commonly search the internet for information. Fallows (2005) estimates that about 84% of U.S. adults use a search engine to help them find information. In fact, search is one of the most popular internet activities (only sending and receiving e-mail ranks higher).

In a theoretical model, Moscarini and Smith (2002) prove that information demand is a decreasing function of the "informational content" of a signal, where informational content is a measure of how well a signal can help the receiver to differentiate states of the economy. When the informational content of a signal is minimal, there is uncertainty with respect to the state of the economy, so the demand for information grows. The implication of the Moscarini and Smith (2002) model is demonstrated by Drake et al. (2012), who find that Google search volume ramps up around the earnings announcements of firms that have wider spreads and higher idiosyncratic risk. In an everyday context, Google search volume is quite likely to serve as a proxy for general investor attention. The Google search engine continues to be the favorite of internet users.[3]

Search volume is one objective way to reveal and quantify the interests of investors. Investors may start to pay attention to a stock and search it out in Google ahead of a prescheduled news event (e.g., an earnings announcement or a release of macroeconomic statistics), or simply follow an observation from a conversation in their economic networks or social networks. A surge in Google search for a stock could be considered a signal that could include a little piece of information about the stock, with huge noise.

As noise dominates information in a search, an increase in Google search is hardly determinative information about a stock, particularly if the stock is covered by enough analysts that information is fully priced into the stock. If we can aggregate signals across Google searches for stocks in the same industry or sector, however, and significantly reduce noise, we might be able to confirm that an increase in Google search is informative about the industry or sector.

Da et al. (2011) directly measure investor attention using the search volume index (SVI) constructed by Google and conclude that an increase in abnormal SVI (ASVI) predicts higher stock prices in the next two weeks and an eventual price reversal within the year.[4] Drake et al. (2012) study investor information demand from Google searches around times of earnings announcement. They argue that the ASVI data reflect investor effort to obtain firm-specific information. Their evidence shows that the market reaction to earnings news is partially preempted when investor search activities in the predisclosure period are abnormally high.

This finding has two implications. First, it suggests that investor search via the internet offers timely price discovery of earnings information. Second, their evidence is inconsistent with the idea that the ASVI measure captures solely the behavior of less sophisticated retail investors (or noise traders).

ASVI does more than confirm investor attention to information. It also reveals a piece of information that the investor might possess. The information arrives with uncertainty, and the investor searches for some confirmation of his or her interpretation of the information. Investors can use internet searches to acquire information already in the public domain that has not yet been fully impounded into prices. Even if a surge in ASVI is triggered only by a news event, it still conveys useful information about the amount of attention that the news event ultimately generates among investors.

Increased attention paid to genuine news may indeed increase the rate at which information is incorporated into prices. When the aggregate search is abnormal, this likely indicates that the information is confirmed. Da et al. (2011) show that abnormal search volume index (ASVI) on a stock predicts a big move in its stock prices. As in Barber and Odean (2008), Da et al. (2011) argue that individual investors are net buyers of attention-grabbing stocks, so an increase in investor attention results in temporary positive price pressure.

Increases in investor attention could also simply reflect positive fundamental information about a firm that is captured by ASVI more quickly than usual. We know different economic agents have access to different information. All information recipients want to validate information they have, and their searches reveal that possessed information for validation.

A 2006 survey on internet use by the Pew Internet & American Life Project shows that while 70% of internet users have used the internet to look up the meaning of a scientific concept or term, 65% have used the internet to learn more about a science story or discovery first heard of offline.[5]

In the field of public health, several studies have shown that web search logs can provide one of the most timely, broad-reaching influenza monitoring systems available today. Polgreen et al. (2008) document that the frequency of internet searches provides information regarding infectious disease activity. Ginsberg et al. (2009) find that search data for 45 terms related to influenza predicted flu outbreaks one to two weeks before Centers for Disease Control and Prevention (CDC) reports.

So, how do we harness the collective intelligence of millions of users in Google web search logs in order to provide a timely and effective predictor of a subject of interest? While a large body of research in literature examines the impact of investor information demand via Google searches on an *individual stock* price, we believe ours is the first study to attempt to investigate such an impact on *sector* performance. Sector or industry information is a key input in a mutual fund manager's investment decision. Thus, whether sector performance is predictable is a first-order question for asset managers. We show that the aggregate investor information demand revealed via Google searches on all stocks in a sector—a collection of wisdom—can predict sector performance.

Not surprisingly, the prospect for a sector is full of uncertainty. Firms within a sector with different resources, business structures, and market positions might react differently to the same economic shock. The information about effects of the shock on the sector could be indeterminate. And even though firms disclose earnings quarterly, firms in the same sector likely have different fiscal year-ends. All this further challenges investors' ability to extract information about sector prospects from a firm's routine financial disclosures.

Internet search data, however, could offer the potential to describe interest in a variety of economic activities in real time. A surge of sector ASVI might reflect positive fundamental information about the sector that is captured on a timely basis. For example, initiation of a new credit policy on the use of renewable energy could trigger positive responses from firms in the green energy sector. Such an initiative would initially move the ASVI higher in the green energy sector as people start to search stocks in the sector, which "predicts" an eventual price increase as this positive news is gradually incorporated into the prices of the majority of stocks in the sector.

In information economics, a study of situations in which different economic agents have access to different information, Milgrom (1981) models the role of monotonicity in information exchange such that more favorable news leads to higher stock prices.[6] Could a surge in a sector's ASVIs be a proxy for a more favorable signal about prospects for the entire sector? We examine empirically whether the aggregate ASVI at a sector level can predict sector performance. Can a collection of wisdom predict sector performance?

Firms disclose earnings quarterly, and financial analysts commonly make forecasts on a firm's earnings routinely. Quarterly standardized unexpected earnings (SUEs) for a firm represent information that has not yet been priced in the firm's stock but that could be associated with the firm's business prospect. Drake et al. (2012) in a study of investor information demand from Google searches find that abnormal Google search increases about two weeks prior to an earnings announcement, spikes markedly at the announcement, and continues at high levels for some time after the announcement. Our work links ASVI to SUEs at an industry level. We show that the sector SUEs, which are the value-weighted SUEs of firms in the same sector, are contemporaneously associated with the sector returns. Even with the sector SUEs as a control, we show that a lagged sector ASVI can still predict the sector performance.

This research makes several contributions to the literature. First, we demonstrate that investor information demand could play a role in determining a sector's performance. Sector performance can be predicted by the aggregate abnormal search volume index across all firms within the sector, even after taking into consideration lagged returns on both the market and the sector. Second, the one-month-ahead predictability in sector returns does not completely reverse in a year. That is, the initial price increase associated with ASVI reflects at least some fundamental information about firms. Finally, with the sector SUEs as a further control, a lagged sector ASVI can still predict the sector's performance.

This paper is organized as follows. We describe the data in Section 2. At the level of individual stocks and sectors, we show the quartile distribution of ASVIs across the largest 3000 U.S. stocks and the S&P 500 stocks. In Section 3, we investigate the predictability of a sector's ASVI for sector performance the next month, and examine whether the positive predictability is due to the price pressure. In Section 4, we then explore the link between the sector SUEs and sector ASVI. We investigate whether the positive predictability of the sector's ASVI for sector performance could reflect fundamental information about the sector on a more timely basis. We conclude in Section 5.

2. Data

Our main data sources are the Center for Research in Security Prices (CRSP) return files, the Institutional Brokers Estimate System (IBES) Unadjusted Detail Historical Estimates and Unadjusted Actuals Database, and the Compustat database. The sample period is from January 2004 through December 2018. We consider the largest 3000 common stocks (*shrcd* 10 and 11) from the CRSP only. Each year we define the largest 3000 U.S. stocks with stock prices over USD 1 according to a stock's market capitalization at the beginning of January.

The S&P 500 is widely regarded as the best single gauge of large-cap U.S. equities. The index includes 500 leading companies in 11 market sectors based on the Global Industry Classification Standard (GICS): Energy, Materials, Industrials, Consumer Discretionary, Consumer Staples, Health Care, Financials, Information Technology, Communication Services, Utilities, and Real Estate. Over USD 11.2 trillion is indexed or benchmarked to the index, and we focus on the S&P 500 sectors.

We retrieve annual index constituents and monthly returns of S&P 500 sector indices directly from Compustat. To quantify a sector's prospect as comprehensively as possible, we also apply S&P 500 sector classification to the largest 3000 U.S. stocks according to a stock's GICS code. Monthly returns on an S&P sector based on the largest 3000 stocks (S&P Sector_3000) are value-weighted monthly returns of all stocks in the sector at the beginning of each year, using a stock's market capitalization at the beginning of each month as a weight.

To analyze the relation between investor attention and sector returns, we use data about search volume from Google Trends, which provides information on search term frequency dating back to January 2004. The application provides the search intensity for a keyword or a group of keywords. It generates a time-series index from 0 to 100 in a selected time frequency. The measured value is obtained based on the order of 100 million searches per day. As in Da et al. (2011) and Drake et al. (2012), we use a stock's ticker to represent investor search for a stock.[7]

Following Da et al. (2011), we define the abnormal search volume index (ASVI) as follows:

$$ASVI_{n,t} = log(SVI_t) - log[Med(SVI_{t-1}, \ldots, SVI_{t-n})] \quad (1)$$

where $logSVI_t$ represents the logarithm of search volume index (SVI) during month t, and $log[Med(SVI_{t-1}, \ldots, SVI_{t-n})]$ is the logarithm of the median value of SVI during the prior n months. For example, the $ASVI_3$ for April is constructed as the *log(SVI)* for April minus the log of the median over the previous three months (January, February, and March). We use $ASVI_1$, $ASVI_2$,..., $ASVI_{12}$ with a one-month lag as an independent variable to check how the impact of ASVIs on sector performance decays across months.

ASVIs not only reflect information demand by investors, but also confirm what information the investors might have. We use an equally weighted average of individual ASVIs (*EW_ASVI*) across stocks within a sector to quantify information about the sector's prospect from an investor search perspective. When the abnormal search volume index for a sector is value-weighted and not equal-weighted, search volume intensity on a big firm in the sector would be assumed to carry more weight in the sector's prospects. Such an assumption implies that whoever searches a bigger firm is smarter, so their search volume describes more about the sector's prospects next month.

This assumption is totally unsound. Further, blue chip stocks are the object of scrutiny of a large number of financial analysts and money managers. Their stock prices most likely reflect all information. Any surge in ASVIs of these stocks could simply be noise. A value-weighted average of ASVIs would amplify noise.

Equation (2) posits the decomposition of a signal φ into its information content θ about the firm in Sector I and errors ε:

$$\varphi_{i \in I} = \theta_i^I(N_i) + \varepsilon_i \quad (2)$$

$$\|\varepsilon_i\| \gg \left\|\theta_i^I\right\| \; \forall i$$

$$\frac{d\theta_i^I}{dN_i} < 0$$

In this context, Firm i's ASVI is considered as a signal φ about the firm. Given that a firm's ASVI reflects changes in attention from average investors, it is reasonable to assume that errors unconditionally dominate information. Therefore, it is unlikely that a firm's ASVI can be individually used to predict its stock price. When N_i, the number of financial analysts who cover Firm i, increases, information about the firm is increasingly incorporated in the stock price, and thus the information content $\theta_i^I(N_i)$ in the signal φ_i is diminished.

Firms in the same sector compete with one another to a certain degree. When investors overestimate the signals about information on a firm's rivals, they more likely underestimate their signals about information on the firm's allies. As we aggregate ASVIs across firms in a sector, the errors in their signals might be minimized because of this offset, making the aggregated signals become informative about the sector.

Table 1 presents the cross-sectional quantile distribution of ASVI at the individual stock level in Panel As and at the sector level in Panel Bs. One would expect ASVIs based on individual stocks to fluctuate more and thus be noisier than those based on sectors. Notably, while the cross-sectional average of ASVIs based on individual stocks does not differ much from that based on sectors, its standard deviation shrinks dramatically at the sector level. Furthermore, a sector's ASVI based on the largest 3000 stocks has a lower standard deviation than the one based on the S&P 500 stocks. This highlights that the ASVIs of S&P 500 stocks contain the most noise because information about these firms has been priced. Thus, a sector's ASVI based on the universe of the largest 3000 stocks could be a better choice to quantify information about the sector's prospects.

Table 1. Summary statistics for ASVI3 and ASVI6. ASVI3 and ASVI6 are calculated for each stock in each month from January 2004 through December 2018. Panel A1 presents the quantile distribution for the universe of the largest 3000 U.S. stocks as well as the mean value and standard deviation. The largest 3000 stocks are defined each year based on a stock's market capitalization at the beginning of the year. These statistics are presented in a format of percentages for December of 2004, 2009, 2014, 2018, with the average across all months in Row "All". Panel A2 presents statistics for S&P 500 stocks. Equally weighted averages of ASVIs (*EW_ASVI*) for each sector in each month are presented at a sector level for the universe of the largest 3000 U.S. stocks in Panel B1 and S&P 500 stocks in Panel B2.

	Year	Quartiles			Mean	SD
		25%	Median	75%		
Panel A1. Individual stock ASVIs based on the largest 3000 stocks						
ASVI3	2004	−15.42	−3.23	9.89	−2.59	33.78
	2009	−16.99	−5.13	5.72	−6.66	30.85
	2014	−16.71	−5.51	2.74	−7.11	28.45
	2018	−17.48	−6.19	2.67	−7.81	28.74
	All	−9.96	−0.21	9.58	−0.32	29.90
ASVI6	2004	−15.47	−3.85	8.96	−3.84	29.25
	2009	−18.23	−6.06	5.41	−8.07	30.06
	2014	−15.96	−5.04	4.36	−6.47	28.47
	2018	−18.23	−6.52	1.74	−8.99	28.90
	All	−9.96	−0.30	9.46	−0.55	28.67

Table 1. *Cont.*

	Year	Quartiles 25%	Median	75%	Mean	SD
Panel A2. Individual stock ASVIs based on the S&P 500 stocks						
ASVI3	2004	−11.99	−2.44	6.90	−2.74	26.85
	2009	−12.17	−3.85	4.47	−4.28	18.44
	2014	−11.69	−4.40	1.87	−4.74	17.19
	2018	−14.55	−6.54	0.00	−7.59	16.73
	All	−6.84	−0.07	6.87	0.26	19.90
ASVI6	2004	−13.98	−3.63	6.65	−4.30	25.55
	2009	−14.55	−4.70	4.50	−5.17	21.01
	2014	−11.94	−4.37	2.90	−4.28	18.22
	2018	−14.90	−6.54	0.00	−8.20	19.45
	All	−7.06	−0.13	7.12	0.37	20.11
Panel B1. Sector ASVIs based on the largest 3000 stocks						
ASVI3	2004	−4.07	−2.90	−1.65	−3.28	3.44
	2009	−9.43	−5.01	−2.72	−5.08	3.30
	2014	−9.72	−6.39	−4.43	−6.05	3.27
	2018	−10.57	−7.85	−4.98	−7.51	3.09
	All	−2.07	−0.29	1.16	−0.08	2.55
ASVI6	2004	−4.56	−3.86	−3.18	−5.35	6.03
	2009	−10.80	−6.26	−4.01	−6.56	3.24
	2014	−7.94	−5.96	−3.47	−5.18	3.62
	2018	−12.14	−9.00	−5.88	−8.50	3.37
	All	−2.16	−0.40	1.11	−0.14	2.64
Panel B2. Sector ASVIs based on the S&P 500 stocks						
ASVI3	2004	−5.61	−2.87	−0.65	−2.20	3.70
	2009	−7.34	−4.47	−3.04	−4.53	3.34
	2014	−7.55	−4.93	−3.57	−5.91	3.78
	2018	−9.30	−7.93	−6.07	−7.69	2.63
	All	−2.50	−0.06	2.09	0.33	4.19
ASVI6	2004	−9.43	−5.40	−1.75	−5.64	6.59
	2009	−9.84	−4.54	−2.75	−5.26	4.89
	2014	−6.39	−4.99	−4.02	−5.20	4.05
	2018	−11.23	−7.60	−5.75	−7.70	3.16
	All	−2.36	0.10	2.27	0.49	4.37

3. Is a Sector's ASVI Informative about the Sector's Prospects?

We adopt a pooled-OLS panel regression:

$$Y_{i,t} = \beta_1 EW\_ASVI_{i,t-1} + \beta_2 Y_{i,t-1} + \beta_3 R_{t-1}^{S\&P500} + \mu_i + \varepsilon_{i,t} \quad (3)$$

where $Y_{i,t}$ is value-weighted sector i's returns based on either the largest 3000 U.S. stocks, $R_{i,t}^{S\&P\ Sector\_3000}$, or the S&P 500 stocks, $R_{i,t}^{S\&P500\ Sector}$, in month t. The return on the S&P 500 Index, a proxy for the market, with a one-month lag is included. We scale all independent variables by their standard deviation, so the estimated coefficients are directly informative about the economic significance of the effects. The sector-fixed effect (μ_i) is included, and all standard errors are adjusted for error correlations clustered by both sector and month according to Petersen (2009).

Table 2 clearly shows that a sector's abnormal search volume index (ASVI), when it is defined as the search volume index (SVI) in excess of the median of SVIs over the prior two to six months, can predict the sector's performance next month even after controlling for returns on the market and the sector. When the ASVI is defined as the search volume index in excess of the median of SVIs over a period beyond the prior seven months, however, it does not have any predictability (results not tabulated). The predictability of ASVI for the S&P sector's performance is maintained whether S&P sectors are constructed based on the universe of the largest 3000 stocks (S&P Sector_3000) or on S&P 500 stocks (S&P 500 Sector Index).

Table 2. Prediction of abnormal search volume on S&P sector returns next month. The sample includes the largest 3000 U.S. common stocks at the beginning of each year. Each stock is assigned to one of eleven S&P sectors according to the stock's GICS code. We calculate the monthly value-weighted returns of the eleven sectors over a year following the sector formation, using as a weight the market capitalization of a stock at the beginning of each month. Following Da et al. (2011), we construct the stock's ASVI. We calculate the equally weighted monthly ASVIs for each S&P sector. A panel regression on prediction of S&P sector returns is performed. The dependent variable is an S&P sector's return (in percentage) in month t calculated for the largest 3000 stocks in Panel A and the S&P 500 Sector Index return in Panel B. The explanatory variables include the ASVI on the S&P sector, return on the S&P sector, and return on S&P 500 Index in month t - 1, all scaled by their standard deviation. All variables are in percentages. The regression coefficients with p-values in parentheses are reported. All standard errors are adjusted for error correlations clustered by both sector and month according to Petersen (2009). ***, **, and * indicate significance of coefficients at the 1%, 5%, and 10% level, respectively. We examine Granger causality in F-tests on the null hypothesis that lagged sector ASVIs do not Granger-cause the S&P sector returns. The sample period is from January 2004 through December 2018.

Independent Variables	Panel A: Dependent Variable: $R_t^{S\&P\ Sector\_3000}$							
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
$ASVI1_{t-1}$	0.274							
	(0.173)							
$ASVI2_{t-1}$		0.409 *						
		(0.054)						
$ASVI3_{t-1}$			0.483 **				0.479 **	0.482 **
			(0.031)				(0.032)	(0.031)
$ASVI4_{t-1}$				0.407 **				
				(0.049)				
$ASVI5_{t-1}$					0.355 *			
					(0.076)			
$ASVI6_{t-1}$						0.354 *		
						(0.073)		
$R_{t-1}^{S\&P\ Sector\_3000}$	−0.022	−0.039	−0.040	−0.031	−0.029	−0.027	0.374	
	(0.890)	(0.815)	(0.807)	(0.851)	(0.863)	(0.876)	(0.342)	
$R_{t-1}^{S\&P\ 500}$	0.528	0.537	0.520	0.511	0.513	0.520		0.488
	(0.133)	(0.126)	(0.143)	(0.154)	(0.153)	(0.147)		(0.257)
Sector-Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2 (%)	0.78	1.16	1.40	1.12	0.95	0.98	1.05	1.45
Sector-Months	1958	1947	1936	1925	1914	1903	1936	1936
Granger Causality Test								
F-statistic	6.170	13.427	18.871	13.095	9.874	9.691	18.470	18.833
p-value	0.0131	0.0003	0.0000	0.0003	0.0017	0.0019	0.0000	0.0000
Independent Variables	**Panel B: Dependent Variable: $R_t^{S\&P\ 500\ Sector\ Index}$**							
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
$ASVI1_{t-1}$	0.245							
	(0.214)							
$ASVI2_{t-1}$		0.370 *						
		(0.070)						
$ASVI3_{t-1}$			0.459 **				0.458 **	0.460 **
			(0.035)				(0.034)	(0.035)
$ASVI4_{t-1}$				0.368 *				
				(0.064)				

Table 2. *Cont.*

Independent Variables	Panel B: Dependent Variable: $R_t^{S\&P\ 500\ Sector\ Index}$							
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**
$ASVI5_{t-1}$					0.322 *			
					(0.093)			
$ASVI6_{t-1}$						0.308		
						(0.109)		
$R_{t-1}^{S\&P\ 500\ Sector}$	0.065	0.057	0.052	0.064	0.070	0.070	0.291	
	(0.779)	(0.808)	(0.823)	(0.787)	(0.769)	(0.775)	(0.433)	
$R_{t-1}^{S\&P\ 500}$	0.327	0.331	0.317	0.306	0.303	0.311		0.356
	(0.295)	(0.289)	(0.313)	(0.336)	(0.342)	(0.331)		(0.365)
Sector-Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2 (%)	0.25	0.57	0.87	0.54	0.40	0.39	0.74	0.93
Sector-Months	1806	1796	1786	1776	1766	1756	1786	1786
Granger Causality Test								
F-statistic	4.717	10.492	16.253	10.245	7.745	7.014	16.137	16.359
p-value	0.0300	0.0012	0.0001	0.0014	0.0054	0.0082	0.0001	0.0001

Table 2 also shows that ASVI3, the SVI in excess of the median of SVIs over the prior three months, has the greatest economic significance of the six different ASVIs in predicting the S&P sector's performance. A one-standard-deviation increase in ASVI3 predicts a rise of 0.483% (0.459%) in next-month S&P sector performance defined by the universe of the largest 3000 stocks in Panel A (the S&P 500 stocks in Panel B). A low adjusted R-squared is not uncommon in a simple predictive regression of the broad market or general sector performance. In studying whether the returns of industry portfolios predict the stock market, Hong et al. (2007) show an adjusted R-squared of 0.9% in their predictive regression when one-month-lagged returns of the market and the metal industry are the only two independent variables.

In a robustness check, we examine if ASVIs do Granger-cause the S&P sector returns in F-tests. Granger (1969) put forth a framework for assessing whether one series is predictive of another. A time series X is said to Granger-cause Y if it can be shown, usually through F-tests on lagged values of X (and with lagged values of Y also included), that those X values provide statistically significant information about future values of Y. According to F-statistics in Table 2, the null hypothesis that the lagged sector ASVI is not Granger causal of the S&P sector's performance is rejected at a significance level of less than 3% in every model.

A Barber and Odean (2008) model predicts that attention shocks lead to net buying by retail traders. Because retail traders are uninformed on average, this should lead to temporarily higher returns. While the positive coefficient on ASVIs is consistent with such a price pressure hypothesis, it could also indicate that ASVIs capture positive fundamental information about the firm on a more timely basis.

For example, suppose a firm announces a technology innovation to which consumers react positively. Such a positive reaction immediately translates into a higher SVI as people start to search out the company stock. As this positive news gradually gets incorporated into the stock price, a later price increase is predicted.

We first attempt to validate why the equally weighted average of individual ASVIs across stocks within the sector should quantify information about a sector's prospects. In Table 3, we consider stocks in five groups: the S&P 500 stocks, the largest 1000 stocks, the second largest 1000 stocks (1001st–2000th), the smallest 1000 stocks (2001st–3000th), and all the largest 3000 stocks for a sector's ASVI calculation, either equally weighted (EW) or value-weighted (VW). We run the regression in Equation (3), but allow the ASVI variable to be either EW or VW. We report only the coefficients of a sector's ASVI 3 and ASVI 6 in the last columns.

Table 3. Sector prediction of abnormal search volume based on portfolios by size. We report the sector prediction coefficients of ASVI for portfolios of S&P 500 stocks, the largest 1000 U.S. stocks, the second largest 1000 stocks, the smallest 1000 stocks, and the largest 3000 U.S. stocks. Both equally weighted (EW) and value-weighted (VW) ASVI3 and ASVI6 are calculated for each portfolio. We calculate monthly value-weighted ASVIs, using as a weight the market capitalization of a stock at the beginning of each month. For each firm each month, we also collect analysts' current-fiscal-year annual earnings per share forecasts reported in the IBES Summary History file, and calculate the number of IBES analysts (N_Analysts) and the analyst dispersion (AD). N_Analysts is the number of analysts who provide fiscal-year-one earnings estimates. AD is the ratio of the standard deviation of analysts' forecasts to the absolute value of the mean forecast. We report the average for each portfolio. A panel regression on prediction of S&P sector returns is performed as in Table 2. The dependent variable is an S&P sector's return (in percentage) in month t, calculated for the largest 3000 stocks in Panel A and for the S&P 500 Sector Index return in Panel B. The explanatory variables include the sector ASVIs constructed for the portfolios with different size, return on the S&P sector, and return on S&P 500 Index in month t - 1, all scaled by their standard deviation. All variables in regressions are in percentages. Only the coefficients of sector ASVIs with p-values in parentheses are reported. Adjusted R^2s are in brackets. All standard errors are adjusted for error correlations clustered by both sector and month according to Petersen (2009). ***, **, and * indicate significance of coefficients at the 1%, 5%, and 10% level, respectively. The sample period is from January 2004 through December 2018.

Panel A: $R_{i,t}^{S\&P\ Sector\_3000} = \beta_1 ASVI_{i,t-1} + \beta_2 R_{i,t-1}^{S\&P\ Sector\_3000} + \beta_3 R_{t-1}^{S\&P\ 500} + \mu_i + \epsilon_{i,t}$

ASVIs are constructed for	N_Analysts	AD	Coefficient of ASVI3		Coefficient of ASVI6	
			EW	VW	EW	VW
S&P 500	17.1	0.095	0.220	−0.032	0.086	−0.106
			(0.142)	(0.850)	(0.544)	(0.518)
			[0.64]	[0.45]	[0.50]	[0.52]
Largest 1000	11.0	0.173	0.211	0.148	0.139	0.034
			(0.230)	(0.266)	(0.346)	(0.795)
			[0.63]	[0.53]	[0.55]	[0.48]
Second Largest 1000 (1001st–2000th)	6.4	0.227	0.459 ***	0.392 **	0.389 **	0.294 *
			(0.006)	(0.024)	(0.040)	(0.100)
			[1.31]	[1.07]	[1.08]	[0.82]
Smallest 1000 (2001st–3000th)	3.4	0.310	0.280 *	0.299 *	0.285 **	0.250 *
			(0.065)	(0.060)	(0.030)	(0.062)
			[0.49]	[0.81]	[0.51]	[0.72]
Largest 3000	8.9	0.205	0.483 **	0.053	0.354 *	−0.062
			(0.031)	(0.772)	(0.073)	(0.725)
			[1.40]	[0.45]	[0.98]	[0.49]

Panel B: $R_{i,t}^{S\&P\ 500\ Sector\ Index} = \beta_1 ASVI_{i,t-1} + \beta_2 R_{i,t-1}^{S\&P\ 500\ Sector\ Index} + \beta_3 R_{t-1}^{S\&P\ 500} + \mu_i + \epsilon_{i,t}$

ASVIs are constructed for	N_Analysts	AD	Coefficient of ASVI3		Coefficient of ASVI6	
			EW	VW	EW	VW
S&P 500	17.1	0.095	0.152	−0.094	−0.003	−0.194
			(0.374)	(0.603)	(0.982)	(0.204)
			[0.07]	[0.01]	[0.00]	[0.15]
Largest 1000	11.0	0.173	0.242	0.101	0.172	−0.049
			(0.141)	(0.431)	(0.221)	(0.738)
			[0.22]	[0.01]	[0.11]	[0.00]
Second Largest 1000 (1001st–2000th)	6.4	0.227	0.426 **	0.363 **	0.269 *	0.209
			(0.020)	(0.039)	(0.097)	(0.182)
			[0.75]	[0.54]	[0.29]	[0.17]
Smallest 1000 (2001st–3000th)	3.4	0.310	0.267 *	0.295 *	0.271 *	0.260 *
			(0.093)	(0.090)	(0.053)	(0.078)
			[0.28]	[0.34]	[0.30]	[0.27]
Largest 3000	8.9	0.205	0.459 **	−0.038	0.308	−0.164
			(0.035)	(0.829)	(0.109)	(0.290)
			[0.87]	[0.00]	[0.39]	[0.10]

We take the average of number of financial analysts (*N_Analyst*) and analyst dispersion (*AD*) as proxies for the degree of information that has been priced. *N_Analysts* is the number of IBES analysts who provide fiscal-year-one earnings estimates. *AD* is defined as the ratio of the standard deviation of analysts' current-fiscal-year annual earnings per share forecasts to the absolute value of the mean forecast, as reported in the IBES Summary History file. Diether et al. (2002) argue that dispersion in analysts' earnings forecasts reveals different opinions on the stock valuation. We expect that larger firms are scrutinized by more analysts and will show less analyst dispersion. Thus, information is most likely priced in their stocks, and any surge in ASVIs will be nothing except noise. Table 3 clearly confirms these conjectures.

S&P 500 firms are on average covered by about 17 analysts' coverage and enjoy less analyst dispersion, where the standard deviation of analysts' earnings forecasts is just about 0.095 of the mean forecast. Therefore, any change in a stock's ASVI could just be noise, and the collection of individual ASVIs is still subject to noise. The noise is magnified in the value-weighted calculation for a sector's ASVI3, where the R^2 of the regression is lower in the value-weighted than the equally weighted calculation.

The second largest 1000 firms in the largest 3000 pool have about 6.4 analysts' coverage, and the *AD* is about 0.227. For the smallest 1000, analyst coverage is 3.4, with an *AD* of 0.310. Therefore, any change in the ASVIs of these stocks could reflect both information and noise. The aggregate individual ASVIs in these stocks can describe their sector's prospects and significantly predict sector performance, in terms of equally weighted or value-weighted calculations. The result holds for the S&P 500 Sector Index returns analyzed in Panel B of Table 3, except for the coefficient of ASVI6 in the value-weighted calculation for the second largest 1000 stocks.

In short, using equally weighted ASVIs in a sector as a proxy for the sector's prospect can guard against the problem that noise is magnified in the value-weighted ASVIs.

Da et al. (2011) find that ASVI predicts both the initial price increase and subsequent long-run price reversal. They conclude a long-run price reversal is more consistent with the price pressure hypothesis than the information hypothesis. Both Barber and Odean (2008) and Da et al. (2011), of course, investigate attention predictability in terms of individual stocks and present evidence supporting the price pressure hypothesis.

The main distinguishing feature between the price pressure hypothesis and the information-based alternative is the prediction of long-run returns. If an initial price increase is due to temporary price pressure, we would expect it to reverse in the long run. If, however, the initial price increase partially reflects fundamental information about the firm, then long-run reversal would not be expected to be severe. For our research, we ask whether the aggregate ASVI at a sector level can possibly be informative about the sector's prospect. We examine long-run returns for Quarters 2 through 4 in Table 4.

Panel A of Table 4 clearly shows that the initial price increase is reversed about 76% in one year. The insignificant negative coefficient of −0.367 on ASVI3 in predicting S&P Sector_3000 performance from quarters 2 through 4 is about 76% the size of the initial price pressure in the first month. Da et al. (2011) find that the incremental predictive power of ASVI persists in the first two weeks and is reversed from weeks 5 through 52. Their reversal rate is about 86%.

In Panel B, a negative coefficient of −0.658 on ASVI3 in predicting S&P 500 Sector Index performance from quarters 2 through 4 exceeds the degree of 0.459 in the first month. Since S&P 500 companies are highly covered and held by many institutional investors, we know fundamental information is most likely fully priced in these stocks. Thus, a high ASVI indicates attention-grabbing purchases. It predicts both the initial price increase and subsequent long-run price reversal.

Table 4. Prediction of abnormal search volume on long-run S&P sector returns. The sample includes the largest 3000 U.S. common stocks at the beginning of each year. Each stock is assigned to one of eleven S&P sectors according to the stock's GICS code. We calculate the value-weighted monthly returns of these eleven sectors over a year following the sector formation, using as a weight the market capitalization of a stock at the beginning of each month. We calculate the equally weighted monthly ASVI3 for each S&P sector. A panel regression on prediction of S&P sector returns is performed. The dependent variable is an S&P sector's future return (in percentage) during the first 3 months and during quarters 2 to 4. Monthly sector returns are compounded for quarterly returns calculated for the largest 3000 stocks in Panel A and for the S&P 500 Sector Index return in Panel B. The explanatory variables include the ASVI on the S&P sector, return on the S&P sector, and return on S&P 500 Index in Month 0, all scaled by their standard deviation. All variables are in percentages. The regression coefficients with *p*-values in parentheses are reported. All standard errors are adjusted for error correlations clustered by both sector and month according to Petersen (2009). ***, **, and * indicate significance of coefficients at the 1%, 5%, and 10% level, respectively. The sample period is from January 2004 through December 2018.

Independent Variables	**Panel A: Dependent Variable: $R_t^{S\&P\ Sector\_3000}$**						
	Quarter 1			**Quarter 2**	**Quarter 3**	**Quarter 4**	**Quarter 2–4**
	Month 1	**Month 2**	**Month 3**				
$ASVI3_0$	0.483 **	0.203	−0.131	−0.360	−0.044	−0.051	−0.367
	(0.031)	(0.337)	(0.532)	(0.383)	(0.907)	(0.889)	(0.680)
$R_0^{S\&P\ Sector\_3000}$	−0.040	0.170	0.095	−0.495 *	0.180	0.048	−0.330
	(0.807)	(0.514)	(0.631)	(0.061)	(0.784)	(0.856)	(0.726)
$R_0^{S\&P\ 500}$	0.520	−0.300	0.193	0.242	−0.122	−0.101	−0.264
	(0.143)	(0.470)	(0.595)	(0.669)	(0.879)	(0.837)	(0.834)
Sector-Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2 (%)	1.40	0.00	0.00	0.00	0.00	0.00	0.10
Sector-Months	1936	1925	1914	1881	1848	1815	1815
Independent Variables	**Panel B: Dependent Variable: $R_t^{S\&P\ 500\ Sector\ Index}$**						
	Quarter 1			**Quarter 2**	**Quarter 3**	**Quarter 4**	**Quarter 2–4**
	Month 1	**Month 2**	**Month 3**				
$ASVI3_0$	0.459 **	0.158	−0.137	−0.490	−0.059	−0.177	−0.658
	(0.035)	(0.480)	(0.528)	(0.236)	(0.869)	(0.588)	(0.412)
$R_0^{S\&P\ Sector\_3000}$	0.052	−0.026	−0.099	−0.405 ***	0.048	−0.004	−0.388
	(0.823)	(0.892)	(0.464)	(0.000)	(0.923)	(0.989)	(0.540)
$R_0^{S\&P\ 500}$	0.317	−0.137	0.441	0.177	−0.027	−0.038	−0.166
	(0.313)	(0.681)	(0.117)	(0.705)	(0.967)	(0.933)	(0.876)
Sector-Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2 (%)	0.87	0.00	0.02	0.02	0.00	0.00	0.40
Sector-Months	1786	1775	1764	1731	1698	1665	1665

4. Link between Sector SUEs and a Sector ASVI

What might make the sector ASVI informative about sector prospects? The predictability of ASVI could be simply the result of data mining. Drake et al. (2012) find that abnormal Google search increases about two weeks prior to an earnings announcement and spikes markedly at the announcement. We argue that a sector's ASVI could capture fundamental information associated with the sector's standardized unexpected earnings (SUEs) on a timely basis.

Earnings surprises represent a specific type of information that investors are reasonably expected to collect.[8] Firms in the same sector are likely exposed to common demand shocks and similar technologies. Unlike firm-specific information, at which only a few investors may be looking, sector earnings surprise information is likely to affect a collection of firms and thus paint an outlook of the sector's prospects.[9]

We construct four quarterly SUEs for a firm in the largest 3000 U.S. stocks. Following Livnat and Mendenhall (2006), we first construct SUEs using three methods.[10] The first method (SUE1) is given by

$$SUE_{j,t} = (X_{j,t} - X_{j,t-4})/P_{j,t} \tag{4}$$

where $X_{j,t}$ is primary Earnings Per Share (EPS) before extraordinary items for firm j in quarter t, and $P_{j,t}$ is the price per share for firm j at the end of quarter t from Compustat. $X_{j,t}$ and $P_{j,t}$ are unadjusted for stock splits, but $X_{j,t-4}$ is adjusted for any stock splits and stock dividends during the period $[t-4, t]$.

Method 2 (SUE2) excludes "special items" (Compustat variable SPIQ) from the Compustat. Specifically, to estimate *SUE* from the Compustat data after exclusion of special items, we subtract from the primary EPS the amount of special items times 65%, divided by the number of shares used to calculate primary earnings per share.

Method 3 (SUE3) replaces the forecast ($X_{j,t-4}$) with a measure of analysts' expectations. Considering only the most recent forecast for each analyst, the measure of analysts' expectations is the median of forecasts reported to IBES in the 90 days prior to the earnings announcement. SUE3 is based solely on IBES median estimates/actuals in Unadjusted Detail History Estimates and Actuals data, and does not use Compustat data. Following the method used in Mendenhall (2004), we deflate a stock's earning surprise by the dispersion of analysts' forecasts on the stock in SUE4 instead of deflating it by the stock price in SUE3.

We calculate the value-weighted monthly SUEs of eleven S&P sectors over a year following the sector formation, using as a weight the market capitalization of a stock at the beginning of each month. Table 5 provides summary statistics for S&P sector SUEs. In Panel A, when S&P sectors are based on the universe of the largest 3000 stocks, on average, SUE3 results in the most negative mean and median, while SUE4 results in higher volatility in estimations. This indicates that analysts, on average, are more optimistic in forecasting S&P sector earnings. In the smallest 500 stocks within the universe of the largest 3000 stocks in Panel C, it seems that analysts have more difficulty in making the correct call on sector earnings forecasts. Fewer financial analysts follow the bottom 500 firms, so these firms' ASVIs more likely convey more information.

To address differences in earnings surprises across months and industries as well as outliers, we transform the SUEs into percentile ranks. Following the method used in Chen (2018), we assign a stock a percentile rank score between 0 (the lowest) and 1 (the highest) every month according to its SUEs and then subtract 0.5 from the SUEs rank score in order to assign a score of 0 to the median observation. When firms do not have a quarterly earnings announcement in a given month, we exclude them from the ranking for that month. We then construct the scaled SUEs for each S&P sector monthly and still use the market capitalization of a firm at the beginning of each month as a weight. Constructing sector-wide earnings surprises based on reports by a few firms in some months will be less informative. Still, monthly earnings surprise signals better represent *information flow*, from which street investors can gradually glean sector prospects, than quarterly earnings surprise signals do.

Consistent with the literature on post-earnings-announcement drift, the results in Table 6 show that a sector's SUEs, regardless of different constructions of SUEs, are contemporaneously associated with the sector's prospects. A one-standard-deviation increase in SUEs is significantly associated with a contemporaneous rise in sector performance, ranging from 0.157% to 0.210%. Among the four SUEs, SUE2 is the most important while SUE3 is the second most important measure associated with contemporaneous sector performance.

Table 5. Summary statistics of sector earnings surprises. The sample includes the largest 3000 U.S. common stocks at the beginning of each year. We calculate a stock's quarterly standardized earnings surprises (SUEs) using 4 methods discussed in the text. Each stock is assigned to one of eleven S&P sectors according to the stock's GICS code. We calculate the value-weighted monthly SUEs of these eleven sectors over a year following the sector formation, using as a weight the market capitalization of a stock at the beginning of each month. Panel A presents cross-sectional quartile distribution for the eleven sectors along with the mean value and standard deviation. Panel B presents the same statistics for sector SUEs based on the S&P 500 stocks while Panel C presents the statistics for the smallest 500 stocks in the pool of the largest 3000 stocks. The table reports the statistics for December of 2004, 2009, 2014, 2018 and across all months. Numbers are in percentages. The sample period is from January 2004 through December 2018.

Variables	Year	Quartiles			Mean	SD
		25%	Median	75%		
Panel A: Summary statistics for sector SUEs based on largest 3000 stocks						
SUE1	2004	−0.44	0.12	1.91	0.47	16.10
	2009	−0.95	0.00	0.17	0.45	5.08
	2014	−0.27	0.16	1.63	−2.87	10.90
	2018	−0.54	0.05	1.62	−0.00	1.56
	All	−0.09	0.01	0.13	−0.00	4.66
SUE2	2004	−0.76	0.16	2.48	1.02	16.10
	2009	−0.30	0.13	5.83	3.33	8.02
	2014	−0.30	0.03	1.58	−3.12	10.90
	2018	−0.85	0.02	0.29	−0.62	2.34
	All	−0.08	0.01	0.13	−0.08	4.59
SUE3	2004	−0.80	−0.33	0.30	−3.84	14.30
	2009	−3.07	−0.44	0.17	−0.00	12.70
	2014	−4.11	−0.34	0.11	−5.13	15.00
	2018	−34.60	−2.05	−0.31	−13.80	17.40
	All	−0.15	−0.00	0.08	−0.56	7.38
SUE4	2004	−2.39	−0.42	0.38	−4.40	10.90
	2009	−2.45	−0.21	0.92	1.81	17.90
	2014	−3.68	−0.48	0.36	−5.43	13.90
	2018	−34.60	−1.07	0.24	−12.30	23.60
	All	−0.10	0.04	0.18	−0.50	7.77
Panel B: Summary statistics for sector SUEs based on S&P 500 stocks						
SUE1	2004	−0.86	0.45	4.35	4.23	12.40
	2009	−4.45	−0.19	−0.01	−7.12	13.90
	2014	−0.35	0.43	3.36	1.50	3.18
	2018	−0.28	0.03	0.42	0.19	2.98
	All	−0.37	0.04	0.51	0.02	6.97
SUE2	2004	−1.70	0.64	3.34	4.82	12.30
	2009	0.10	0.35	0.46	5.19	14.10
	2014	−0.44	0.43	3.38	1.47	3.30
	2018	−0.04	0.00	0.21	−0.28	3.55
	All	−0.36	0.06	0.52	0.23	6.76
SUE3	2004	−7.65	0.09	4.92	1.81	13.90
	2009	−6.13	−1.66	−0.10	−5.09	7.78
	2014	−2.47	−0.73	−0.34	−1.41	1.89
	2018	−1.36	−1.28	−1.17	−2.70	3.85
	All	−0.36	0.02	0.41	0.23	7.36
SUE4	2004	−8.56	−0.78	1.32	−2.97	6.20
	2009	−6.36	−0.10	1.00	−4.51	8.43
	2014	−1.70	1.15	1.57	−0.06	2.99
	2018	−3.51	−2.07	0.46	−3.85	7.03
	All	−0.23	0.27	0.88	0.40	7.10

Table 5. *Cont.*

Variables	Year	Quartiles			Mean	SD
		25%	Median	75%		
Panel C: Summary statistics for sector SUEs based on smallest 500 stocks						
SUE1	2004	−12.70	−1.61	−0.73	−10.70	16.90
	2009	−3.81	1.54	4.29	2.34	10.40
	2014	−5.79	0.86	2.39	0.10	9.08
	2018	−13.80	2.54	12.20	−1.44	17.40
	All	−1.24	0.11	1.62	0.21	12.60
SUE2	2004	−12.40	−0.73	−0.52	−11.20	19.30
	2009	−3.41	0.14	3.59	1.99	10.50
	2014	−6.21	1.70	4.80	1.12	7.34
	2018	−17.90	−1.49	10.40	−3.74	17.10
	All	−1.08	0.11	1.66	0.38	12.70
SUE3	2004	−9.90	−4.51	−1.84	3.45	25.20
	2009	−40.20	−1.42	2.74	−13.30	22.10
	2014	6.09	13.80	19.20	10.70	18.30
	2018	−12.30	4.43	6.93	−0.32	10.50
	All	−2.41	0.05	2.14	−0.40	16.80
SUE4	2004	−11.90	−6.64	−2.06	−1.52	16.70
	2009	−38.80	−2.70	0.19	−14.30	21.30
	2014	−5.00	3.81	12.60	6.56	20.80
	2018	−12.60	−4.03	6.96	−3.22	9.80
	All	−3.40	−0.17	1.19	−1.46	15.70

Table 6. S&P sector SUEs and S&P sector returns. The sample includes the largest 3000 U.S. common stocks at the beginning of each year. Each stock is assigned to one of eleven S&P sectors according to the stock's GICS code. We calculate a stock's quarterly standardized earnings surprises (SUEs) using 4 methods discussed in the text. To address the outliers and differences in the earnings surprises across months and sectors, we transform the SUEs into percentile ranks according to Chen (2018). We then construct scaled SUEs for each S&P sector monthly, using the market capitalization of a firm at the beginning of each month as a weight. A panel regression on explaining S&P sector returns is performed. The dependent variable is an S&P sector's return (in percentage) calculated for the largest 3000 stocks in month t. The explanatory variables include the S&P sector SUEs in month t as well as returns on both S&P sector and S&P 500 Index in month t - 1, all scaled by their standard deviation. All variables are in percentages. The regression coefficients with p-values in parentheses are reported. All standard errors are adjusted for error correlations clustered by both sector and month according to Petersen (2009). ***, **, and * indicate significance of coefficients at the 1%, 5%, and 10% level, respectively. The sample period is from January 2004 through December 2018.

Independent Variables	Dependent Variable: $R_t^{S\&P\ Sector\_3000}$					
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
$SUE1_t$	0.175 ***					
	(0.000)					
$SUE2_t$		0.210 ***				
		(0.000)				
$SUE3_t$			0.187 ***		0.200 ***	0.187 ***
			(0.001)		(0.000)	(0.001)
$SUE4_t$				0.157 ***		
				(0.000)		
$R_{t-1}^{S\&P\ Sector\_3000}$	−0.047	−0.043	−0.020	−0.022	0.408	
	(0.783)	(0.806)	(0.906)	(0.896)	(0.315)	
$R_{t-1}^{S\&P\ 500}$	0.539	0.537	0.537	0.538		0.521
	(0.146)	(0.147)	(0.145)	(0.145)		(0.242)
Sector-Fixed	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2 (%)	0.60	0.68	0.71	0.67	0.33	0.76
Sector-Months	1949	1951	1917	1917	1917	1917

When a surge in a stock's ASVI comes at the time of a firm's quarterly earnings disclosure, ASVI relays investor attention to earnings-related information. ASVI also reveals the information that the investor has. The information arrives under uncertainty, and the investor then searches for a confirmation of his or her interpretation of the information. Even if a surge in ASVI is triggered completely by a news event, ASVI carries further information about the extent of the attention. Increased attention paid to genuine news may increase the rate at which information is incorporated into prices. When the aggregate search is abnormal, this indicates either positive or negative confirmation.

To test whether a sector ASVI can still predict the sector's performance after controlling for sector earnings surprise, we examine a pooled-OLS panel regression:

$$Y_{i,t} = \beta_1 SUE_{i,t} + \beta_2 EW\_ASVI_{i,t-1} + \beta_3 Y_{i,t-1} + \beta_4 R_{t-1}^{S\&P500} + \mu_i + \varepsilon_{i,t} \tag{5}$$

where $Y_{i,t}$ is value-weighted sector i's returns based on either the universe of the largest 3000 U.S. stocks, $R_{i,t}^{S\&P\ Sector\_3000}$ or S&P 500 stocks, $R_{i,t}^{S\&P500\ Sector}$, in month t.

Table 7 shows that sector ASVI still provides significant predictability of S&P sector performance even after controlling for the sector's SUEs and others. The significant and positive coefficient on ASVIs still reflects positive fundamental information about the firm that is captured on a timely basis. Inclusion of the sector's ASVI not only enhances the predictability of contemporaneous SUEs for S&P sector performance, but also doubles the R^2, the explanatory power of the model. Furthermore, the economic significance of the lagged ASVI on sector performance is greater than that of the contemporaneous SUEs. A one-standard-deviation increase in SUE3 is significantly associated with a contemporaneous rise of 0.194% in sector performance while a one-standard-deviation increase in ASVI3 significantly predicts the sector returns increasing by 0.453% next month.[11]

Table 7. Prediction of abnormal search volume on S&P sector returns with SUEs as a control. The sample includes the largest 3000 U.S. common stocks at the beginning of each year. Each stock is assigned to one of eleven S&P sectors according to the stock's GICS code. We calculate the value-weighted monthly returns of these eleven sectors over a year following the sector formation, using as a weight the market capitalization of a stock at the beginning of each month. For each S&P sector, we construct its ASVI and SUEs. The detailed construction and variable definition are described in Tables 4 and 6. A panel regression on prediction of S&P sector returns is performed. The dependent variable is an S&P sector's return (in percentage) calculated for the largest 3000 stocks in month t. The explanatory variables include the S&P sector SUEs in month t as well as the S&P sector ASVI3, return on the S&P sector, and return on S&P 500 Index in month t - 1, all scaled by their standard deviation. All variables are in percentages. The regression coefficients with p-values in parentheses are reported. All standard errors are adjusted for error correlations clustered by both sector and month according to Petersen (2009). ***, **, and * indicate significance of coefficients at the 1%, 5%, and 10% level, respectively. We examine Granger causality in F-tests on the null hypothesis that lagged sector ASVI3 does not Granger-cause the S&P sector returns. The sample period is from January 2004 through December 2018.

Independent Variables	Dependent Variable: $R_t^{S\&P\ Sector\_3000}$					
	Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**
$SUE1_t$	0.185 ***					
	(0.000)					
$SUE2_t$		0.216 ***				
		(0.000)				
$SUE3_t$			0.194 ***		0.207 ***	0.195 ***
			(0.001)		(0.000)	(0.001)
$SUE4_t$				0.160 ***		
				(0.001)		
$ASVI3_{t-1}$	0.475 **	0.472 **	0.453 **	0.452 **	0.449 **	0.452 **
	(0.041)	(0.042)	(0.047)	(0.048)	(0.049)	(0.048)
$R_{t-1}^{S\&P\ Sector\_3000}$	−0.062	−0.057	−0.035	−0.037	0.389	
	(0.724)	(0.748)	(0.840)	(0.829)	(0.349)	

Table 7. *Cont.*

Independent Variables	Dependent Variable: $R_t^{S\&P\ Sector\_3000}$					
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
$R_{t-1}^{S\&P\ 500}$	0.533	0.530	0.530	0.531		0.502
	(0.156)	(0.158)	(0.156)	(0.156)		(0.267)
Sector-Fixed	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2 (%)	1.44	1.51	1.47	1.42	1.10	1.52
Sector-Months	1916	1918	1885	1885	1885	1885
Granger Causality Test						
F-statistic	17.981	17.804	16.096	15.994	15.777	16.070
p-value	0.0000	0.0000	0.0001	0.0001	0.0001	0.0001

In a robustness check, we examine if ASVIs can still Granger-cause the S&P sector returns after taking contemporaneous SUEs into consideration. According to the F-statistics shown in Table 7, we conclude that the lagged sector ASVI3 is indeed Granger causal of the S&P sector's performance.

When we include sector SUEs in the regression, Table 8 clearly shows that the initial price increase associated with SUE3 reflects information about sector prospects and does not entail long-run reversal. The initial price increase led by ASVI3 is reversed by about 86% in one year, however. A negative coefficient of −0.388 on ASVI3 in predicting S&P Sector_3000 performance from quarters 2 to 4 is about 86% of the initial price pressure revealed in the first month. With a control of contemporaneous sector SUEs, ASVIs might overstate the information content about a sector even though ASVIs can capture information about sector prospects on a more timely basis.

Table 8. Prediction of abnormal search volume on long-run S&P sector returns with SUEs as a control. The sample includes the largest 3000 U.S. common stocks at the beginning of each year. Each stock is assigned to one of eleven S&P sectors according to the stock's GICS code. We calculate the value-weighted monthly returns of these eleven sectors over a year following the sector formation, using as a weight the market capitalization of a stock at the beginning of each month. For each S&P sector, we construct its ASVI and SUEs. The detailed construction and variable definition are described in Tables 4 and 6. A panel regression on prediction of S&P sector returns is performed. The dependent variable is an S&P sector's future return (in percentages) during the first 3 months and during quarters 2 to 4. Monthly sector returns are compounded for quarterly returns. The explanatory variables include the S&P sector SUEs in Month 1 as well as the S&P sector ASVI3, return on the S&P sector, and return on S&P 500 Index in Month 0, all scaled by their standard deviation. All variables are in percentages. The regression coefficients with p-values in parentheses are reported. All standard errors are adjusted for error correlations clustered by both sector and month according to Petersen (2009). ***, **, and * indicate significance of coefficients at the 1%, 5%, and 10% level, respectively. The sample period is from January 2004 through December 2018.

Independent Variables	Dependent Variable: $R_t^{S\&P\ Sector\_3000}$						
	Quarter 1			Quarter 2	Quarter 3	Quarter 4	Quarter 2–4
	Month 1	Month 2	Month 3				
$SUE3_1$	0.194 ***	−0.038	−0.021	−0.119	−0.028	0.037	0.002
	(0.001)	(0.742)	(0.849)	(0.727)	(0.894)	(0.913)	(1.000)
$ASVI3_0$	0.453 **	0.148	−0.104	−0.348	−0.044	−0.071	−0.388
	(0.047)	(0.494)	(0.629)	(0.414)	(0.909)	(0.852)	(0.667)
$R_0^{S\&P\ Sector\_3000}$	−0.035	0.150	0.153	−0.513 *	0.146	0.073	−0.368
	(0.840)	(0.572)	(0.455)	(0.100)	(0.830)	(0.795)	(0.711)
$R_0^{S\&P\ 500}$	0.530	−0.248	0.136	0.266	−0.080	−0.123	−0.216
	(0.156)	(0.555)	(0.714)	(0.654)	(0.923)	(0.814)	(0.868)
Sector-Fixed	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R^2 (%)	1.47	0.00	0.00	0.00	0.00	0.00	0.07
Sector-Months	1885	1876	1865	1834	1804	1771	1771

5. Conclusions

Search is a revealed attention measure: if a person searches for a stock, the person is definitely paying attention to it. We know information acquisition is costly. Yet, the exact cost of collecting any piece of information depends on timing, location, a person's private information set, and so on. Other uncertainties include the idiosyncratic characteristics and complexities of the information signal and of the asset itself.

The Google search engine continues to be the favorite of internet users. Search volume index (SVI) is an objective way to reveal and quantify the interests of investors. Investors could start to pay attention to a stock and use Google to search it well ahead of a prescheduled news event (e.g., an earnings announcement or a release of macroeconomic statistics). Even a surge in SVI completely triggered by a news event carries further information about the amount of attention ultimately generated among investors. Increased attention paid to genuine news may increase the rate at which information is incorporated into prices.

Our work shows a sector's abnormal search volume index (ASVI), defined as the search volume index (SVI) in excess of the median of SVIs over the prior two to seven months, can predict the sector's performance next month, even after controlling for returns on the market and the sector. The initial price increase due to a rise in ASVI is reversed about 76% in one year. In other words, parts of fundamental information about the sector can be captured by the sector ASVI on a timely basis.

When a surge of a stock's ASVI occurs at the time of a firm's quarterly earning disclosures, ASVI conveys the degree of investor attention to information as well as a piece of information that the investor possesses. With controls for a sector's contemporaneous earnings surprises and other factors, the sector's ASVI still exhibits significant predictability for S&P sector performance. The positive coefficient on ASVIs could at least partially reflect positive fundamental information about the sector captured on a timely basis.

Author Contributions: Conceptualization, H.-l.C.; methodology, H.-l.C.; software, H.-l.C. and J.S.; validation, H.-l.C. and J.S.; formal analysis, H.-l.C. and J.S.; investigation, H.-l.C. and J.S.; resources, H.-l.C. and J.S.; data curation, H.-l.C. and J.S.; writing—original draft preparation, H.-l.C.; writing—review and editing, H.-l.C. and J.S.; visualization, H.-l.C. and J.S.; supervision, H.-l.C.; project administration, H.-l.C. and J.S.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from Compustat, CRSP, IBES, and Google, and are available from the authors with the permission of Compustat, CRSP, IBES, and Google.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1. According to Van Nieuwerburgh and Veldkamp (2010), information acquisition may be viewed as a technology to "resolve uncertainty earlier". That is, information lets an investor learn sooner what he would otherwise observe only after asset payoffs are realized. In this case, under-diversification arises optimally because of increasing returns to scale in learning.
2. When attention is a precious cognitive resource and investors have limited attention, information may not be incorporated instantaneously into asset prices upon arrival. Economists have recognized that investors do not have unlimited processing capacity but are better characterized as only boundedly rational. Several authors propose theoretical models in which limited attention could affect asset pricing (See Merton 1987; Hirshleifer and Teoh 2003; Sims 2003; Peng and Xiong 2006).
3. As of January 2021, Google accounted for 62% of the U.S. desktop search (93% of mobile search) queries (see https://www.statista.com/statistics/265796/us-search-engines-ranked-by-number-of-core-searches/; accessed on 15 May 2021). Google makes the Search Volume Index of search terms public via the product Google Trends (http://www.google.com/trends). Trends data can provide a powerful lens on what Google users are curious about and how people around the world react to important events. Weekly SVI for a search term is the number of searches for that term scaled by its time-series average.
4. Joseph et al. (2011) and Vlastakis and Markellos (2012) further confirm the conclusion of price reversal.
5. According to a 2006 survey about the internet use by the Pew Internet & American Life Project, the internet is a research tool for 87% of online users. That translates to 128 million adults in a U.S. population of about 298 million in 2006. See https:

//www.pewresearch.org/internet/2006/11/20/the-internet-as-a-resource-for-news-and-information-about-science/ (accessed on 15 January 2020).

6 In the case of EntreMed (ENMD), a biotechnology company, Huberman and Regev (2001) document that a no-new-news New York Times article caused the stock price to more than double and on a permanent basis. They question to what extent stock prices reflect fundamental information or just publicity.

7 Using a firm's name in a search may not indicate that the user in interested in fundamental-related information about the firm. For example, one may search "Apple" for food or shopping iPhones. Several tickers may have a generic meaning and have nothing to do with attention paid to the stocks with these ticker symbols. For example, "BIG", "BOX", "CAT", "SKY", "WIN". An inclusion of these "noisy" tickers guard against us finding results supporting ASVI's predictability.

8 The predictability of stock returns after earnings announcements has attracted substantial attention since the late 1960s. Bernard and Thomas (1989, 1990) and Bartov (1992), among others, document that the post-earnings-announcement drift represents the market's failure to fully reflect the attributes of the stochastic process underlying earnings. Bhushan (1994) concludes that transaction costs preclude professionals from taking positions that would eliminate the drift. Mendenhall (2004) further highlights that arbitrage risk impedes arbitragers who attempt to profit from the drift.

9 Chen (2018) presents new evidence that industrywide earnings surprises diffuse gradually across the supply chain at both industry and individual-firm levels.

10 We use "esurprises.sas" provided by Wharton Research Data Services (WRDS) to construct SUEs defined by Livnat and Mendenhall (2006). We adopt their data selection criteria. The detailed description of SUE construction can be found in Chen (2018) and the WRDS web page, https://wrds-www.wharton.upenn.edu/pages/support/sample-programs/ibes/calculate-quarterly-standardized-earnings-surprises-sue/. (accessed on 15 September 2020).

11 When sector performance is calculated based on the S&P 500 stocks, the results in Table 7 still hold for ASVI2 to ASVI4. However, the explanatory power of contemporaneous SUEs disappears. The results are not tabulated to save space but available upon request.

References

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

Bartov, Eli. 1992. Patterns in unexpected earnings as an explanation for post-announcement drift. *The Accounting Review* 67: 610–22.

Bernard, Victor L., and Jacob K. Thomas. 1989. Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research (Supplement)* 27: 1–36. [CrossRef]

Bernard, Victor L., and Jacob K. Thomas. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13: 305–40. [CrossRef]

Bhushan, Ravi. 1994. An informational efficiency perspective on the post-earnings-announcement drift. *Journal of Accounting and Economics* 18: 45–65. [CrossRef]

Chen, Hsiu-lang. 2018. Information Diffusion of upstream and downstream industry-wide earnings surprises and its implications. *Review of Quantitative Finance and Accounting* 51: 751–84. [CrossRef]

Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2011. In search of attention. *Journal of Finance* 66: 1461–99. [CrossRef]

Diether, Karl B., Christopher J. Malloy, and Anna Scherbina. 2002. Differences of opinion and the cross section of stock returns. *Journal of Finance* 57: 2113–41. [CrossRef]

Drake, Michael S., Darren T. Roulstone, and Jacob R. Thornock. 2012. Investor information demand: Evidence from Google searches around earnings Announcements. *Journal of Accounting Research* 50: 1001–40. [CrossRef]

Fallows, Deborah. 2005. *Search Engine Users*. Washington, DC: Pew Internet & American Life Project.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant. 2009. Detecting influenza epidemics using search engine query data. *Nature* 457: 1012–14. [CrossRef]

Granger, C. W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424–38. [CrossRef]

Hirshleifer, David, and Siew Hong Teoh. 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36: 337–86. [CrossRef]

Hong, Harrison, Walter Torous, and Rossen Valkanov. 2007. Do industries lead stock markets? *Journal of Financial Economics* 83: 367–96. [CrossRef]

Huberman, Gur, and Tomer Regev. 2001. Contagious speculation and a cure for cancer: A nonevent that made stock prices soar. *Journal of Finance* 56: 387–96. [CrossRef]

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

Kacperczyk, Marcin, Clemens Sialm, and Lu Zheng. 2005. On the industry concentration of actively managed equity mutual funds. *Journal of Finance* 60: 1983–2011. [CrossRef]

Livnat, Joshua, and Richard R. Mendenhall. 2006. Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research* 44: 177–205. [CrossRef]

Mendenhall, Richard R. 2004. Arbitrage risk and post-earnings-announcement drift. *Journal of Business* 77: 875–94. [CrossRef]

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

Milgrom, Paul. 1981. Good news and bad news: Representation theorems and applications. *Bell Journal of Economics* 12: 380–91. [CrossRef]

Moscarini, Giuseppe, and Lones Smith. 2002. The law of large demand for information. *Econometrica* 70: 2351–66. [CrossRef]

Peng, Lin, and Wei Xiong. 2006. Investor attention, overconfidence and category learning. *Journal of Financial Economics* 80: 563–602. [CrossRef]

Petersen, Mitchell A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22: 435–80. [CrossRef]

Polgreen, Philip M., Yiling Chen, David M. Pennock, and Forrest D. Nelson. 2008. Using internet searches for influenza surveillance. *Clinical Infectious Diseases* 47: 1443–48. [CrossRef]

Sims, Christopher A. 2003. Implications of rational inattention. *Journal of Monetary Economics* 50: 665–90. [CrossRef]

Van Nieuwerburgh, Stijn, and Laura Veldkamp. 2010. Information acquisition and portfolio under-diversification. *Review of Economic Studies* 77: 779–805. [CrossRef]

Vlastakis, Nikolaos, and Raphael N. Markellos. 2012. Information demand and stock market volatility. *Journal of Banking & Finance* 36: 1808–21.