Google and Financial markets: Can Google trends describe and predict the dynamics of Norwegian stock market?

By

Neri Kim

And

Roviel Villa

2017

UiS Business School

Universitetet
i Stavanger

DET SAMFUNNSVITENSKAPELIGE FAKULTET,
HANDELSHØGSKOLEN VED UIS

MASTEROPPGAVE

STUDIEPROGRAM:	OPPGAVEN ER SKREVET INNEN FØLGENDE SPESIALISERINGSRETNING:
Master i økonomi og administrasjon	**Anvendt finans**
	ER OPPGAVEN KONFIDENSIELL? (**NB**! Bruk rødt skjema ved konfidensiell oppgave)

TITTEL:

Google og det finansielle markedet: Kan Google trends forklare og forutse det norske aksjemarkedet?

ENGELSK TITTEL:

Google and Financial markets: Can Google trends describe and predict the dynamics of the Norwegian stock market?

FORFATTER(E)		VEILEDER:
Kandidatnummer:	Navn:	**Peter Molnár**
1085 …………………	**Neri Kim** …………………………………….	
1072 …………………	**Roviel Villa** …………………………………….	

TABLE OF CONTENTS

LIST OF EQUATIONS, FIGURES, TABLES, & MODELS

EQUATIONS

FIGURES

TABLES

MODELS

LIST OF ABBREVIATIONS

AR	– Abnormal returns
ASVI	– Abnormal search volume index
ATV	– Abnormal trading volume
bt	– business term
GT	– Google Trends
OSE	– Oslo stock exchange
OBX	– Oslo Børs index
st	– search term
SVI	– Search volume index
TV	– Trading volume

ACKNOWLEDGEMENTS

We would like to thank our thesis advisor, Peter Molnár, for his guidance and valuable contribution to our study. His extensive knowledge and previous work on the topic of Google search volume index and the stock market served as building block for our research. We also thank Gennady Zavyalov for giving us guidance and valuable input in using the statistical software R. We're also thankful for Katarina Lucivjanska for the feedback and additional comments she gave on our first draft. Our academic institution, the University of Stavanger (UiS) has also been helpful in giving us access to resource materials. Last but not the least, we thank our family and friends for their support. Their presence and understanding helped us to focus on the work at hand and see it through.

ABSTRACT

We investigate whether Google search volume index (SVI) can explain and predict trading activity at the Norwegian stock market (OSE). Our sample focuses on the companies listed on OSE's tradeable index OBX. We use abnormal returns, trading volume, and volatility as measures of market activity. SVI were classified as a) search term, which Google uses to keep track of word-specific search queries; and b) business term, which Google uses to keep track of all search queries done in any language and classifies these queries together under one topic. The regression models we developed were two-fold: (1) a descriptive model that tests whether a relationship exists between each of the three indicators and SVI; and (2) a predictive model that tests the predictive power of SVI towards the three indicators. Our results show that both SVIs neither exhibit a significant relationship nor a predictive power on abnormal returns. However, both SVIs show a significant positive relationship and a predictive power on trading volume. Lastly, search activity only exhibits a predictive power with volatility. Therefore, Google searches can tell more about future trading activity than current trading activity.

1. INTRODUCTION

Google's search engine is by far the most popular and highly utilized information gathering platform in the world. Close to 90% of searches are handled by the search giant worldwide and many business rely on being ranked highly in the platform's organic search results to attract attention from potential customers (Harford, 2017). Google also keeps track of statistics for various search queries done on their search engine and these are publicly available through their product Google Trends (henceforth GT). Information offered by the platform has garnered attention from the research community and was used to either identify trends or predict dynamics of, including among others, the stock market. Google's search volume index (henceforth SVI) was previously shown to be a significant proxy for investor attention (Da, Engelberg, & Gao, 2011) and investor sentiment (Joseph, Wintoki, & Zhang, 2011). SVI was also used to forecast stock returns over different time horizons and showed that its predictive power increased in recent years (Bijl, Kringhaug, Molnár, & Sandvik, 2016). Trading strategies based on information from SVI were also developed (Preis, Moat, & Stanley, 2013; Bijl et al., 2016), both indicating the potential for a profitable trading strategy. Challet & Ayed (2014) challenged methodology of Preis et al. (2013) and showed that random finance-related keywords were not better indicators of exploitable predictive information compared to other random keywords. At the same time, the former found out that keywords applied to suitable assets led to robust profitable strategies, which confirms the latter's intuition. Bijl et al. (2016) also noted that their trading strategy was only profitable if transaction costs were not considered.

Previous research was built around the observed relationship and predictive power of SVI on US stock market characteristics. This led us to wonder whether their earlier findings are replicable and applicable to a different and relatively smaller market. Da et al. (2011) suggests that SVI may be more pronounced in smaller markets. An ideal small market in this scenario must have high internet penetration, high internet activity, and high utility of Google's localized search engine. Norway meets these parameters, with the country's internet penetration rate reported at 96.8% (World Bank, 2015), daily internet access at 89% (SSB, 2017), and Google's localized search engine ranked as the top site visited with users averaging 6:20 minutes daily with 7.47 daily unique page views (Alexa, 2017). As researchers based in Norway, we also have the local insight and observation that can help contextualize our results. Thus, we decided to focus our study on SVI and the Norwegian stock market (henceforth OSE).

As far as we know, no previous research has been done to test whether findings on SVI and the stock market can be observed at a different geographical and economical setting, especially in Norway. We thus set-up this paper as a first look towards previously observed phenomenon in a different context and setting. Working with a smaller market bring with it challenges in data availability both from the market and GT side. To overcome this limitation, we thus focus our attention towards constituents of the market index for OSE, the Oslo Børs Total Return Index (henceforth OBX). Companies listed in the OBX index are the most liquid companies, making both their market data and Google SVI relatively easier to obtain.

Our general research question would thus be: can the SVI explain the dynamics of the Norwegian stock market? Tackling this question would require narrowing the scope further and work around measurable and tangible components. First would be to decompose both the SVI and the OSE constituents into values that can be standardized and regressed against each other. Da et al. (2011) and Bijl et al. (2016) gives two methods to standardize SVI into comparable values across different securities. On the contrary OSE, and financial markets in general, is dynamic that various ways can be used to measure its movement. Da et al. (2011) and Bijl et al. (2016) explored the effect of SVI on stock returns and found that SVI can predict stock market movements with subsequent reversal. Thus, we will use OSE returns to measure the market's movements. In addition, Preis, Reith, and Stanley (2010) found strong evidence of SVI's predictive power on trading volume, and thus we also use this as another measure of market activity. Lastly, volatility of the market is also a good indicator of market activity, especially with stock price dispersion (Molnár, 2012). In a later study, Molnár along with Kim and Paulsson (2017) found that a mutual relationship between Google searches and volatility exists and that SVI can improve volatility and correlation forecasting when they studied Google search activity and foreign currency exchange. Building on these discussions, we thus include volatility as our third measure of the OSE's dynamics.

Identifying all these measures then lead us to specific ways to answer our main research question. We compare the relation between SVI and returns, trading volume and volatility, in two ways. First, we study the contemporary relationship between SVI and these three measures or market activity. Second, we investigate whether SVI can even predict these variables. We find that even though GT do not predict of explain stock returns, the opposite is true for volatility and trading volume. Interestingly, Google trends have stronger relation to future volatility and volume than to current volatility and volume.

The remaining sections are organized as follows. Section 2 provides an overview of the GT platform, its development and how it was studied by previous research. We then discuss the data we gathered and the methodology in section 3. Section 4 then presents our findings and a discussion of the results. We then conclude this paper and give recommendations for further research in section 5.

2. GOOGLE TRENDS

Google Trends is a real-time daily index of the volume of queries users enter into Google (Choi & Varian, 2011). The platform also gives access to what it calls “non-real time data,” which pertains to historical data from 2004 up to 36 hours prior to search activity (Where Trends data come from - Trends Help, n.d.). Non-real time data can be viewable and downloadable in different time ranges: past hour, past 4 hours, past day, past 7 days, past 30 days, past 90 days, past 12 months, past 5 years, 2004 to present, and custom time range. However, time frequency in the dataset varies according to the time range set by the user: hourly data for the past hour up to the past day; daily data for the past week up to 90 days; weekly data from the past year up to 5 years; and monthly data for a time range beyond 5 years.

2.1 Previous research on Google SVI

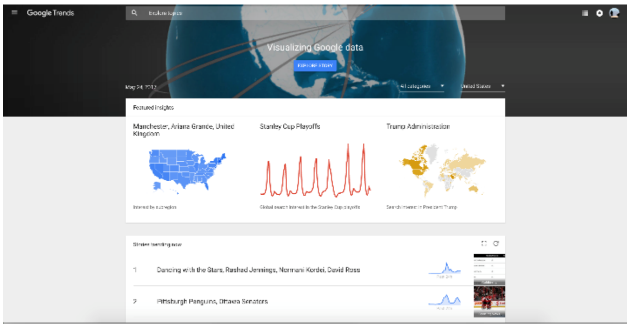
The traditional view on financial markets assume its efficiency and that all relevant information is incorporated in the existing share price (Fama, 1998). However, recent technological advancements that brought the digital age led to a shift from traditional industry to a digital information-based economy (Castells, 1999). This economic shift has provoked further investigation from researchers arguing against the market’s efficiency, with direct measures for investor attention (Da et al., 2011) and investor sentiment (Joseph et al., 2011) observed by tracking search activity in Google as suggested by Choi & Varian (2011). Da et al. (2011) proved that an increase in SVI predicted higher stock prices in the next 2 weeks and subsequent price reversal within the year using data from Russel 3000 stocks from 2004 to 2008. They also showed that SVI captures investor attention in a more timely fashion compared to other investor attention measures and that this attention is mostly from retail investors. Joseph et al. (2011) triangulates the former’s findings on SVI’s predictive power, this time on abnormal returns and trading volume over a weekly horizon using a sample of S&P 500 firms from 2005 to 2008. They examined the ability of online ticker searches and concluded that online ticker search serves as a valid proxy for investor sentiment.

Recently, Bijl et al. (2016) noted that a few attempts have been made to forecast financial markets based on GT data, which gave off mixed results. Preis et al. (2010) investigated the correlation between returns and search volume for company names, but found no significant correlation. They, however, found strong evidence that Google search data could be used to predict trading volume. Preis et al. (2013) investigated whether general search terms related to finance could be used to predict market movements. They found that a strategy where a market portfolio is bought, or sold, based on the Google search volumes for certain keywords could outperform the market index by 310% over the 7-year period they investigated. Their findings met staunch criticism for the paper's subtle biases and overfitted model, most notably Challet & Ayed (2014) who showed that random non-finance-related keywords replicated the same result. They however confirmed Preis et al.(2013)'s intuition when using applying keywords to suitable assets. Bijl et al. (2016) found out that search query data based on company names can be used to predict weekly stock returns for individual firms, with their results showing that high search volume predicted low future returns. However, the relationship between returns and search volume was weak yet robust and statistically significant. A trading strategy based on their findings also yielded weak profitability due to transaction costs.

2.2 Development

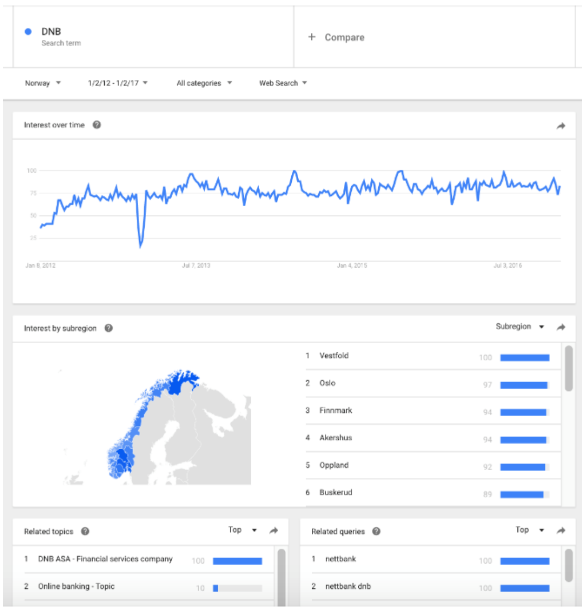
Research mentioned above examined GT back when it was still under development by Google. Since its original inception in 2004, GT had undergone changes in its interface and how it presents data (presented in Figure 1). Thus, features available by the time we conducted this study such as categories and channels (described further in section 3.3) may not have existed prior. In September 27, 2012, Google's Trends and Insights for Search platforms were merged together (Matias, 2012), paving the way to the current version of Google Trends. Google claims to update their GT information daily and Hot Trends hourly to reflect real time data (Tamir, 2015). They have also demonstrated in the past that they respond to demands for fresh updates from the online community (Weinberg, 2007). With a growing interest in utilizing GT data for different fields and advent of the data revolution, it can be expected that Google will update the platform to meet the demand. This then could lead to development of future features that does not exist during the time this paper is being written. Consequently, data that we might have not included in this paper due to the 90% or more occurrence of 0 may turn out to be significant in the future as Google's data infrastructure and algorithm improves.

Figure 1. Google Trends development



Panel A. Google Trends homepage, accessed on May 24, 2017

Panel B. Google Insights for Search, utilized by Choi & Varian (2011) to forecast near-term values of economic indicators.



Panel C. Google Trends' Explore feature for the search term "DNB" with geographic location set to Norway and time range set to custom, accessed on May 24, 2017.

3. DATA & METHODOLOGY

Our data was obtained primarily from Yahoo! Finance, Google Trends, and Bernt Arne Ødegaard's online data library. The sample period is from January 2nd, 2012 to January 2nd, 2017. However, data from 2011 were also obtained and used, because we standardize some of the variables with respect to their past values. Yahoo! Finance was used to collect daily open, close, high, low, adjusted close price, and trading volume for the companies listed in the Oslo Børs Total Return Index (henceforth OBX). Google Trends was used to obtain raw SVI, with two sets of five-year continuous data obtained as discussed further in section 3.3. For abnormal returns, weekly actual returns were calculated from Yahoo! Finance's daily adjusted closed price while weekly excess returns were calculated using daily Fama-French factors and daily risk-free rate for OSE obtained from Bernt Arne Ødegaard's asset pricing data library, presented in section 3.4. Lastly, weekly trading volume and volatility were calculated using daily data from Yahoo! Finance, presented further in sections 3.5 and 3.6.

Since the SVI is reported weekly, monthly, or not at all, for search words with low search volume, we are unable to include all companies from the OBX. In addition, we only include companies that were in the index from 2012 to 2017 and where we have complete stock data. Our final sample therefore includes 28 companies for the time period we focus on.

3.1 Time horizon

To capture the most recent joint dynamics (or relationship) between Google search volumes and the OSE market, we chose to collect data from the past five years (January 2nd 2012 to January 2nd 2017). Five years is a considerable period to capture any recent movement and test Google's descriptive and predicting power based on the most recent historical data, as was done by Bijl et al. (2016). We also consider developments that happened within the GT platform reported by Matias (2012) and Tamir (2015) and opine that the data we obtain from our specified period is sufficient.

Preis et al. (2010) showed that there is a correlation between GT data of company names and transaction volumes of the corresponding stocks on a weekly time scale. Bijl et al. (2016) found out that a significant and negative relationship exists between weekly abnormal search volumes and subsequent stock returns. Due to these earlier findings, we opted to look at our data using a weekly time horizon. This time horizon makes our task more straightforward as GT data is reported weekly for the time period set. It also gives us a reference point which aids us in assuring our dataset's consistency and minimal presence of noise.

Since we are looking at the impact of Google SVI, we had to match its weekly reporting to the daily reporting for the Norwegian stock market. Google reports their trend data from Sunday to Sunday, while the Norwegian stock market data reports data from Monday to Friday. To match the two datasets, we must frame them such that the SVI falls in between the gaps of the trading week, based on insight that search activity from previous week are reflected in subsequent weeks (Preis, Reith, & Stanley, 2010; Bijl et al., 2016). For this reason, we grouped our daily OSE data to Monday to Monday and we describe this procedure in more detail under section 3.4.

3.2 Norwegian stock market data

Oslo Børs is Norway's central marketplace for listing and trading financial instruments (About Oslo Børs, n.d.). It has five different marketplaces: Oslo Børs, Oslo Axess, Merkur Market, Nordic ABM and Oslo Connect. These five marketplaces offer listing and trading in equities, equity certificates, ETPs, fixed income products and derivatives products.

Oslo Børs has a tradeable index called the OBX, which consists of the 25 most traded securities based on a six-month turnover rating (OBX Total Return Index, n.d.). Within our specific five-year period, the semi-annual turnover occurred 10 times, of which 40 different companies in total has been listed in the OBX. Four companies were delisted in that span, which preliminarily narrowed our list down to 36.

As mentioned above, we obtained historical data for the companies that belong to OBX primarily from Yahoo! Finance. However, unfavorable news has been recently circling around Yahoo! by the time we conducted our study and consequently made us suspect the dataset's reliability. The platform also issues a warning of data unreliability and delayed reporting when accessing historical data for companies that trade in non-US stock markets. Therefore, we also obtained historical data using Thomson Reuter's Eikon to confirm our dataset's accuracy. Eikon is a paid financial information platform made accessible to us through our academic institution, the University of Stavanger. We compared data obtained from these two sources and found only slight differences, most likely due to a rounding error. Eikon's advantage was that their dataset tracked Norwegian public holidays, which explained the existence of zero trading values for reported trading volume from Yahoo! Finance's dataset. Eikon generally reported higher trading volumes compared to Yahoo! Finance, but when the raw trading volume is plotted together, their movement was uniformly the same. A detailed overview of the comparison between the two sources is included in Appendix 2.

Ultimately, Eikon doesn't report the adjusted closed price for the OSE, which Yahoo! Finance keeps track of. For this reason, we chose Yahoo's data as our primary dataset for the OSE, since the adjusted closed price considers the dividends paid out by companies that were included in the OBX, making further calculations more consistent and reliable. However, Yahoo's dataset had missing data for some companies in our list, further narrowing the 36 companies that we preliminarily had down to our final sample size of 28 companies.

3.3 Raw and Abnormal Google Search Volume Index

Bijl et al. (2016) found evidence that company name search activity has a stronger relationship to stock market returns than ticker searches. Based on this insight, we prioritized using words closest to the company's name. We drop words that are generally used in business names, such as "limited" and "ltd.", "group", and "international"; and words that were too general such as "seafood" and "petroleum". Companies with one-word names were the easiest to collect significant raw Google SVI because it provides data that have less than 5% occurrence of 0 values. Some one-word companies however didn't return any raw SVI when using their name as the search word. We went on to use other related words (i.e. company ticker) for these companies. For companies with names that contain more than one word, we checked and compared the words separately and chose the word that had lesser occurrences of 0 values. We tested using the complete name of the company as the search word, but it often led to data that consisted more than a half of 0 values, thus we opted to drop them. GT currently differentiates search words into two: (1) search terms, which show matches for all terms in the language the query was done; or (2) topics, which are a group of terms that share the same concept in any language (Compare Trends search term – Trends Help, n.d.). Thus, we ended up with two raw SVIs, search term (henceforth *st*) and business term for topics (henceforth *bt*). A list of the *st*, *bt*, and companies used in this study can be found in Appendix 1.

GT uses a standardized scale of 0 to 100, where 100 represents the highest query volume during a specific time period and geographic region (Choi & Varian, 2011). The difference in scaling consequently leads to little deviations in value (due to rounding) when downloaded at different time periods. However, Bijl et al. (2016) noted that these differences in value are small and correlation between them are close to 1. From this information, we then obtain two sets of 5-year continuous data. Five years is the maximum period for GT to generate weekly data and our standardization requires SVI from previous year. The first set consists of data from January 2nd, 2011 to January 2nd, 2016, while the second set consists of data from January 2nd, 2012 to January 2nd, 2017.

GT gives search volume information according to a specific geographic location. Previous research done by Preis et al. (2013) indicates that data filtered according to geographic location can better explain movements in the specific geographic location. In their research, they focused on the Dow Jones Industrial Average, one of the indexes based on the US stock market. Following their example, we filtered our data geographic location to Norway.

GT also filters information through the following categories: (1) Arts & Entertainment; (2) Autos & Vehicles; (3) Beauty & Fitness; (4) Books & Literature; (5) Business & Industrial; (6) Computer & Electronics; (7) Finance; (8) Food & Drink; (9) Games; (10) Health; (11) Hobbies & Leisure; (12) Home & Garden; (13) Internet & Telecom; (14) Jobs & Education; (15) Law & Government; (16) News; (17) Online Communities; (18) People & Society; (19) Pets & Animals; (20) Real Estate; (21) Reference; (22) Science; (23) Shopping; (24) Sports; and (25) Travel. The default filter is set to "All Categories". We checked our *st* and *bt* using the finance filter, but it yielded a dataset that mostly contained 0 values, thus we opted to drop them. This currently confirms Bijl et al. (2016) where they found out that the finance filter does not provide improvement over the unfiltered searches in terms of predicting stock returns.

Lastly, GT also filter their data according to which channel a search activity was done: Web, YouTube, News, Photos, and Google Shopping. Out of these channels, web searches yielded more variations in raw SVI for most of the companies, thus we only focused on obtaining web searches for our *st* and *bt*.

Raw SVI_{st} and raw SVI_{bt} are then used to compute abnormal SVI (henceforth ASVI). Standardization that we implement makes these raw SVIs comparable across companies in our dataset (Figure 2). We compute for $ASVI_t$ using two methods discussed by Bijl et al. (2016) and Da et al. (2011).

The first method, denoted as $ASVI_t^B$ in Equation 1, follows the formula used by Bijl et al. (2016), where the average of the past 52 weeks is subtracted from the weekly raw SVI and dividing their difference from the standard deviation of the previous year:

Equation 1. ASVI (Bijl et al.)

$$ASVI_t^B = \frac{SVI_t - \frac{1}{52}\sum_{i=0}^{52} SVI_{t-7*i}}{\sigma_{SVI}}$$

where SVI_t can either be SVI_{st} or SVI_{bt}, and σ_{SVI} is the standard deviation of the SVI for the past 52 weeks.

The second method, denoted as $ASVI_t^D$, follows the formula used by Da et al. (2011), where the log of the weekly raw SVI is subtracted from log of the median SVI in the past 52 weeks:

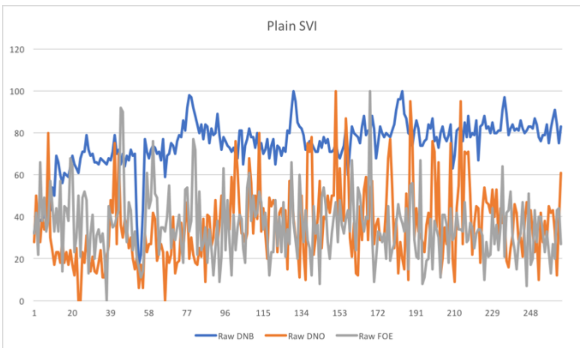
Equation 2. ASVI (Da)

$$ASVI_t^D = \log(SVI_t) - \log[Med(SVI_{t-1}, \dots, SVI_{t-52})]$$

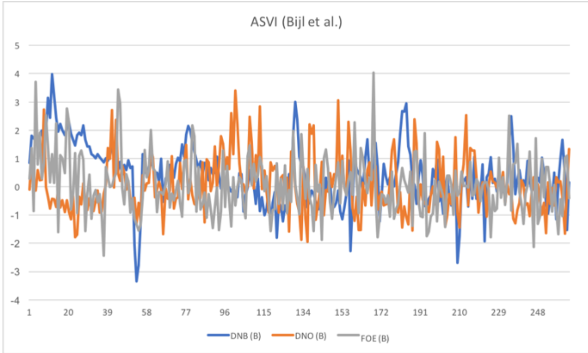
where SVI_t can either be SVI_{st} or SVI_{bt}.

We also study these standardizations over8 and 26 week time horizons.

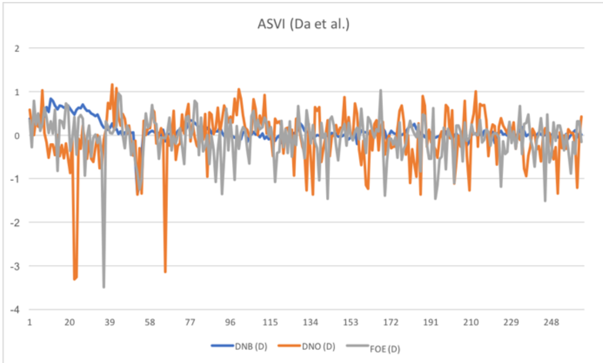
Figure 2. SVI vs ASVI



Panel A: Raw Google SVI for companies with ticker DNB, DNO, and FOE.



Panel B: ASVI for companies with ticker DNB, DNO, and FOE, computed using the formula used by Bijl et al. (2016)



Panel C: ASVI for companies with ticker DNB, DNO, and FOE, computed using the formula used by Da et al. (2011)

Figure 2. Comparison of search volumes for three companies before and after standardization.

3.4 Raw and Abnormal Weekly Stock Returns

Since stock price is reported daily, we created our own algorithm to identify the weekly price for a company. The easiest and most straightforward method is to use the daily price according to our identified start and end for the week, Monday to Monday. This in turn gave us a weekly dataset containing 261 weeks, each week represented by the stock price reported on the corresponding Monday of a certain week *t*. Some Mondays in our dataset still had missing data due to Norway-specific holidays and non-trading days. For these instances, we then used the closest previously reported stock price from previous trading day.

Yahoo's adjusted closed price already includes information for dividends paid out by the stock, thus we can go ahead and compute for the raw log return based on our weekly dataset:

Equation 3. Weekly log return

$$R_t = \log\left(\frac{p_t}{p_{t-1}}\right)$$

where R_t is the raw log return and p_t is the reported stock price for week t and p_{t-1} is the reported stock from the previous week.

For our regression model, we adjust the nominal return R_t with the factors in the Fama and French asset pricing model to compute abnormal return. We acquired our asset pricing data at OSE from Norwegian Financial Data (Ødegaard, n.d.). Based on previous research by Ødegaard (2017), we find the following pricing factors relevant to our computation: HML, SMB (Fama and French, 1998), PR1YR (Carhart, 1997) and LIQ (Ødegaard & Næs, 2009). We calculate the pricing factors beta coefficients from a 1 year rolling regression:

Equation 4. Fama-French returns

$$R_t = R_{f,t} + \beta_{mkt,t} \cdot \left(R_{mkt,t} - R_{f,t}\right) + \beta_{smb,t} \cdot R_{smb,t} + \beta_{hml,t} \cdot R_{hml,t} + \beta_{pr1yr,t} \cdot R_{pr1yr,t} + \beta_{liq,t} \cdot R_{liq,t} + \varepsilon_t$$

Where R_f is the risk-free rate and β are pricing factor loadings. The abnormal return AR_t is calculated as the difference between the actual return and the expected return:

Equation 5. Abnormal returns

$$AR_t = R_t - (R_{f,t} + \hat{\beta}_{mkt,t} \cdot (R_{mkt,t} - R_{f,t}) + \hat{\beta}_{smb,t} \cdot R_{smb,t} + \hat{\beta}_{hml,t} \cdot R_{hml,t} + \hat{\beta}_{pr1yr,t} \cdot R_{pr1yr,t} + \hat{\beta}_{liq,t} \cdot R_{liq,t})$$

The pricing factors and the risk-free rate were not available in weekly data. We thus converted daily to weekly data by compounding the returns from Monday to Monday. To adjust for seasonality (i.e. holidays) we took returns from the nearest previous trading day. Similarly, we compounded the risk-free rate *Rf = log (1+rf)*.

3.5 Weekly Trading Volume

Converting daily trading volume TV_i to weekly trading volume TV_t requires capturing the overall movement of the trading volume for our specified time period (Monday to Monday). This can be done by calculating the average trading volume, which yields $\overline{TV_t}$ shown in Equation 6.

Equation 6. Weekly trading volume

$$TV_t = \frac{1}{|S|} \sum_{i \in S} TV_i$$

where |S| is a number of trading days in a given week

We then calculate the abnormal trading volume (henceforth ATV). Based on the same formula used by Bijl et al. (2016), ATV is scaled by subtracting the mean of the past 52 weeks from the weekly trading volume and dividing their standard deviation of the previous year:

Equation 7. Standardizing trading volume

$$ATV_t = \frac{TV_t - \frac{1}{52}\sum_{i=0}^{52} TV_{t-1}}{\sigma_{TV}}$$

where σ_{TV} is the standard deviation of the volume for the past 52 weeks.

3.6 Weekly Open-High-Low-Close (OHLC) Volatility

Volatility is a popular measure to evaluate how the stock return vary over time. Prior studies examine the effect of volatility on future stock returns and find indeed a positive relationship (French et al., 1987; Banerjee, Doran, & Peterson 2007; Bollerslev & Zhou, 2009). We therefore find it necessary to include volatility as a control variable in our regression model explaining returns and volume, and also as a measure for the market activity. We measure volatility by using the jump-adjusted Garman-Klass volatility estimator discussed by Molnár (2012). The calculation uses open, high, low, close and adjusted close prices during a trading day to calculate the variance for that day, as shown in Equation 8.

Equation 8. Daily jump-adjusted variance

$$variance_t = \frac{1}{2} \cdot (h_t - l_t)^2 - (2\log(2) - 1) \cdot {c_t}^2 + {jadj_t}^2$$

where the notations are described further in Equation 9:

Equation 9. OHLC adjustment

$$\begin{aligned} c_t &= \log(close_t) - \log(open_t) \\ l_t &= \log(low_t) - \log(open_t) \\ h_t &= \log(high_t) - \log(open_t) \\ jadj_t &= j_t * \frac{radj_t}{r_t} \end{aligned}$$

and the jump adjustment $jadj_t$ is defined in Equation 10 as:

Equation 10. Jump adjustment

$$\begin{aligned} j_t &= \log(open_t) - \log(close_{t-1}) \\ r_t &= \log(close_t) - \log(close_{t-1}) \\ radj_t &= \log(adjclose_t) - \log(adjclose_{t-1}) \end{aligned}$$

After calculating for daily variance, we then get the square root of the average for the trading week (Monday to Monday), as summarized in Equation 11.

Equation 11. Weekly volatility

$$Volatility_t = \sqrt{\frac{1}{|S|}\sum_{i \in S} variance_i}$$

3.7 Statistics

Table 1. Descriptive statistics for all variables

Descriptive Statistics

	N	Mean	St.dev.	Min	Max	Skew	Kurtosis
$ASVI_{st}$	7308	0.0439	1.027	-4.422	6.317	0.572	1.619
$ASVI_{bt}$	7308	-0.0360	1.000	-5.109	6.702	0.800	2.932
Return	7308	-0.0005	0.092	-2.150	3.779	10.024	492.349
Volume	7308	0.0564	1.175	-2.507	6.967	1.481	3.194
Volatility	7308	0.0273	0.050	0.003	2.019	24.091	757.261

In Table 1 we present the summary statistics for the variables we generated from our dataset. The $ASVI_{st}$ and $ASVI_{bt}$ are calculated using Bijl et al. (2016) formula with 52-week time horizon discussed in section 3.2. After comparing the results with the different ASVI standardization methods and time horizons, we chose to use Bijl et al. 52-week as the benchmark because the standard deviation of this variable is close to 1, while Da et al. was close to 0.5. The reason why we chose the standardization of Bijl et al. (2016) is twofold. Firstly, this standardization already produces standardized variable with standard deviation very close to one, and this standardization produces most significant results. Return are the abnormal returns we calculated using pricing factors presented in section 3.4. Volume was standardized according to our discussion in section 3.5. Volatility was calculated using the weekly Garman Klass jump adjusted estimator we discussed in section 3.6.

Before we run the regression, we also tested for the correlation between the different variables:

Table 2. Correlation matrix for all variables

Correlation Matrix

	Return	Volatility	Volume	$ASVI_{st}$	$ASVI_{bt}$
Return	1	0.19	0	0.01	-0.02
Volatility	0.19	1	0.17	0.02	0.02
Volume	0	0.17	1	0.09	0.06
$ASVI_{st}$	0.01	0.02	0.09	1	0.50
$ASVI_{bt}$	-0.02	0.02	0.06	0.50	1

As presented in Table 2, the correlation across our different variables are close to 0, making them uncorrelated. Notably, the correlation of $ASVI_{st}$ to $ASVI_{bt}$ is 0.5, which means that both variables share some connection as search terms are matched to the exact word used in a search activity while business terms are the general topics consisting of different search terms and grouped according to Google's algorithm.

3.8 Model Specification

The data results we present later in section 4 was conducted using the statistical software R. We conduct panel data regressions with fixed and random effects and thereafter run the Hausman test to see which of the two should be applied. In the subsections below we present the models we develop to answer our specific research questions.

3.8.1 Descriptive model regression of stock returns

We measure the sensitivity of stock returns by regressing AR against the control variables. We include the control variables presented in section 3.5. and 3.6. It allows us to isolate the impact of the $ASVI$ to that of the control variables. This leads to the following regression model:

Model 1. Descriptive model for AR

$$AR_t = \alpha + \beta_1 AR_{t-1} + \beta_2 ASVI_t + \beta_3 Volatility_t + \beta_4 ATV_t + \varepsilon_t$$

Where AR_t is the abnormal return at time *t*. β are the regression coefficients for lagged abnormal return, Google search volume index, volatility, and trading volume.

3.8.2 Descriptive regression model of Trading Volume

We follow the Da et al. (2011)'s research on whether ASVI can be used as a proxy to capture investor's attention. We use weekly trading volume as a dependent variable to see if changes in search interest explain changes in trading volume. We also include lagged trading volume and volatility as control variables. In this way, we can isolate the impact of ASVI. This results in our second regression model:

Model 2. Descriptive model for ATV

$$ATV_t = \alpha + \beta_1 ATV_{t-1} + \beta_2 ASVI_t + \beta_3 Volatility_t + \beta_4 AR_t + \varepsilon_t$$

Where ATV_t is the abnormal return at time *t*. β are the regression coefficients for lagged trading volume, Google search volume index, volatility, and abnormal returns.

3.8.3 Descriptive regression model of Volatility

The last measure we want to investigate is whether there is an explanatory relationship between ASVI and volatility. We develop our third model:

Model 3. Descriptive model for volatility

$$Volatility_t = \alpha + \beta_1 Volatility_{t-1} + \beta_2 ASVI_t + \beta_3 ATV_t + \beta_4 AR_t + \varepsilon_t$$

Where $Volatility_t$ is the volatility at time *t*. β are the regression coefficients for lagged volatility, Google search volume index, trading volume, and abnormal returns.

3.8.4 Predictive regression models of stock returns, trading volume and volatility

We also developed predictive models based on the descriptive models we described above. Since we are looking also at the predictive power of ASVI on stock returns, trading volume, and volatility, we used lagged variables. Our control variables are included to isolate the effect of ASVI.

Model 4. Predictive model for AR

$$AR_t = \alpha + \beta_1 AR_{t-1} + \beta_2 ASVI_{t-1} + \beta_3 Volatility_{t-1} + \beta_4 ATV_{t-1} + \varepsilon_{i,t}$$

Model 5. Predictive model for ATV

$$ATV_t = \alpha + \beta_1 ATV_{t-1} + \beta_2 ASVI_{t-1} + \beta_3 Volatility_{t-1} + \beta_4 AR_{t-1} + \varepsilon_{i,t}$$

Model 6. Predictive model for volatility

$$Volatility_t = \alpha + \beta_1 Volatility_{t-1} + \beta_2 ASVI_{t-1} + \beta_3 ATV_{t-1} + \beta_4 AR_{t-1} + \varepsilon_t$$

4. RESULTS

The regression models were tested with fixed and random effects. The Hausman test supported the fixed-effect model when we compared the two results therefore we will present results with fixed effects. Breusch-Godfrey and Breusch-Pagan test were conducted to check for autocorrelation and heteroscedasticity. In general, we detected autocorrelation and heteroscedasticity in our dataset and consequently used the Arellano method to control for them (Arellano, 1987). Hence, the tables are presented with robust standard errors.

We organize this section by presenting and discussing the results for ASVI search terms on different dependent variables in section 4.1. Regression results and discussion for ASVI business terms follows in section 4.2 Lastly, we also compared regression results using ASVI standardized by using the two methods and in different time horizons in section 4.3.

4.1 Regression results for ASVI search terms

4.1.1 Returns as dependent variable

Table 3. Regression results on returns when ASVI is calculated from search term

	Dependent variable:								
	Return								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Return_{t-1}$	-0.029							-0.046	-0.022
	(0.047)							(0.058)	(0.036)
$ASVI_t$		0.001						0.0005	
		(0.0003)						(0.001)	
$ASVI_{t-1}$			0.002						0.002
			(0.001)						(0.001)
$Volatility_t$				0.383				0.382	
				(0.412)				(0.426)	
$Volatility_{t-1}$					-0.072				-0.069
					(0.088)				(0.072)
$Volume_t$						0.0003		-0.002	
						(0.001)		(0.004)	
$Volume_{t-1}$							0.001		0.001
							(0.001)		(0.0009)
Observations	7,280	7,308	7,280	7,308	7,280	7,308	7,280	7,280	7,280
R^2	0.001	0.00003	0.0005	0.041	0.002	0.00002	0.0001	0.042	0.003
Adjusted R^2	-0.003	-0.004	-0.003	0.038	-0.002	-0.004	-0.004	0.038	-0.002

*p **p ***p<0.01

Note: Values in columns 1-9 stands for regression outputs to the variables in the corresponding rows. Column 1 is a single regression for dependent variable against independent variable in the first row. Columns 2-9 are multiple regressions of the dependent variable against the independent variables in the corresponding rows. Robust standard errors are reported in parentheses. Lagged variables only have 7280 observations compared to the non-lagged variables that have 7308. When you regress with both lagged and non-lagged variables, R (the software program) will then match to 7280 observations for data balancing.

Table 3 summarizes the results of our regression for both the descriptive and predictive models for returns. ASVI are insignificant when tested both in single regression and multiple regression and regressions exhibit very low values of R^2. Therefore, search volume can neither describe the dynamics of OBX's returns nor predict its movement. These results are contrary to previous findings by Da et al. (2011) and Bijl et al. (2016) where they found out that search volume can predict returns for up to 2 weeks with subsequent reversal for the US stock market. This could be because our dataset was limited to 28 companies. Perhaps expanding the dataset to include more securities such as small cap companies could deliver the same results as the two studies have previously shown.

4.1.2 Volume as dependent variable

Table 4. Regression results on volume when ASVI is calculated from search term

	Dependent variable:								
	Volume								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Volume_{t-1}$	0.577^{***}	0.575^{***}	0.567^{***}	0.569^{***}	0.580^{***}	0.577^{***}	0.577^{***}	0.566^{***}	0.569^{***}
	(0.010)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
$ASVI_t$		0.075^{***}						0.072^{***}	
		(0.014)						(0.015)	
$ASVI_{t-1}$			0.129^{***}						0.129^{***}
			(0.017)						(0.017)
$Volatility_t$				2.332^{**}				2.385^{*}	
				(0.878)				(0.965)	
$Volatility_{t-1}$					-0.331				-0.339
					(0.377)				(0.391)
$Return_t$						0.008		-0.251	
						(0.279)		(0.123)	
$Return_{t-1}$							-0.117		-0.088
							(0.097)		(0.119)
Observations	7,280	7,280	7,280	7,280	7,280	7,280	7,280	7,280	7,280
R^2	0.332	0.336	0.345	0.342	0.332	0.332	0.332	0.346	0.345
Adjusted R^2	0.330	0.334	0.342	0.339	0.330	0.329	0.330	0.343	0.342

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: Values in columns 1-9 stands for regression outputs to the variables in the corresponding rows. Column 1 is a single regression for dependent variable against independent variable in the first row. Columns 2-9 are multiple regressions of the dependent variable against the independent variables in the corresponding rows. Robust standard errors are reported in parentheses. Lagged variables only have 7280 observations compared to the non-lagged variables that have 7308. When you regress with both lagged and non-lagged variables, R (the software program) will then match to 7280 observations for data balancing.

Table 4 shows that ASVI are significant when tested together with control variables at a 99% confidence level. Volume from the previous week along with ASVI of the current week can explain the variance of the current week's trading volume at 33.6%. Furthermore, volume and ASVI from the previous week can predict the variance of the current week's trading volume. These results are still consistent for the multiple regression that includes lagged and non-lagged volatility and returns. Therefore, search volume can both describe and predict trading volume. This observed relationship signifies that investor's sentiment and attention for companies trading in the OBX are captured by Google Search Volume.

4.1.3 Volatility as dependent variable

Table 5. Regression results for volatility when ASVI is calculated from search term

	Dependent variable:								
	Volatility								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Volatility_{t-1}$	0.346^{***}	0.345^{***}	0.344^{***}	0.337^{***}	0.340^{***}	0.354^{***}	0.343^{***}	0.345^{***}	0.336^{***}
	(0.091)	(0.091)	(0.091)	(0.093)	(0.094)	(0.088)	(0.084)	(0.091)	(0.088)
$ASVI_t$		0.001^{*}						0.001	
		(0.0001)						(0.0001)	
$ASVI_{t-1}$			0.003^{**}						0.003^{**}
			(0.0002)						(0.0002)
$Volume_t$				0.005^{***}				0.005^{***}	
				(0.001)				(0.001)	
$Volume_{t-1}$					0.002				0.001
					(0.001)				(0.001)
$Return_t$						0.114		0.113	
						(0.108)		(0.109)	
$Return_{t-1}$							0.008		0.008
							(0.053)		(0.052)
Observations	7,280	7,280	7,280	7,280	7,280	7,280	7,280	7,280	7,280
R^2	0.120	0.120	0.123	0.135	0.121	0.165	0.120	0.180	0.125
Adjusted R^2	0.116	0.117	0.120	0.132	0.118	0.161	0.116	0.176	0.121

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: Values in columns 1-9 stands for regression outputs to the variables in the corresponding rows. Column 1 is a single regression for dependent variable against independent variable in the first row. Columns 2-9 are multiple regressions of the dependent variable against the independent variables in the corresponding rows. Robust standard errors are reported in parentheses. Lagged variables only have 7280 observations compared to the non-lagged variables that have 7308. When you regress with both lagged and non-lagged variables, R (the software program) will then match to 7280 observations for data balancing.

Table 5 shows that ASVI when regressed only with lagged Volatility as the other control variable can explain 12% of the variance (R^2) in the volatility at the 90% confidence level. But when regressed along with lagged and non-lagged control variables trading volume and returns, it becomes insignificant. On the contrary, lagged ASVI remains significant at 95% confidence level when it is either regressed only with lagged Volatility or regressed with the other control variables incorporated in the model. Therefore, an explanatory relationship between current week's search activity with stock price volatility does not seem to exist. However, previous week's search activity can predict the subsequent week's volatility.

We also explored whether replacing weekly with monthly ASVI, volume and volatility would improve our model's findings, but the results were just replicated with lower R^2 values. Tables summarizing the results from the monthly variables can be found in Appendix 3.

We thus sum up our section on ASVI search terms with the conclusion that Google searches can tell us even more about future trading activity for the Norwegian stock market (represented by volatility and volume) than they tell us about current trading activity. The next section presents our findings and discussion for ASVI business terms.

4.2 ASVI business terms

We also tested the effects of business towards returns, trading volume and volatility as shown respectively in Table 6, Table 7, and Table 8. Results for bt were similar to st, however comparing the coefficients and R^2 values shows that bt is relatively weaker. Thus, in the next section where we compare the standardization methods, we will only focus on st.

Table 6. Regression results for returns when ASVI is calculated from business term

	Dependent variable:								
	Return								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Return_{t-1}$	-0.029							-0.046	-0.022
	(0.047)							(0.058)	(0.036)
$ASVI_t$		-0.001						-0.002	
		(0.001)						(0.001)	
$ASVI_{t-1}$			-0.000						-0.000
			(0.008)						(0.000)
$Volatility_t$				0.383				0.382	
				(0.412)				(0.426)	
$Volatility_{t-1}$					-0.072				-0.069
					(0.088)				(0.071)
$Volume_t$						0.000		-0.002	
						(0.001)		(0.004)	
$Volume_{t-1}$							0.001		0.001
							(0.001)		(0.009)
Observations	7,280	7,308	7,280	7,308	7,280	7,308	7,280	7,280	7,280
R^2	0.001	0.000	0.000	0.041	0.002	0.000	0.000	0.042	0.002
Adjusted R^2	-0.003	-0.004	-0.004	0.038	-0.002	-0.004	-0.004	0.038	-0.002

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: Values in columns 1-9 stands for regression outputs to the variables in the corresponding rows. Columns 1-7 are single regressions for dependent variable against independent variables in the corresponding rows. Columns 8-9 are multiple regressions of the dependent variable against the independent variables in the corresponding rows. Robust standard errors are reported in parentheses. Lagged variables only have 7280 observations compared to the non-lagged variables that have 7308. When you regress with both lagged and non-lagged variables, R (the software program) will then match to 7280 observations for data balancing.

Table 7. Regression results for volume when ASVI is calculated from business term

	Dependent variable:								
	Volume								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Volume_{t-1}$	0.577^{***}	0.576^{***}	0.571^{***}	0.569^{***}	0.580^{***}	0.577^{***}	0.577^{***}	0.568^{***}	0.573^{***}
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
$ASVI_t$		0.052^{**}						0.050^{**}	
		(0.017)						(0.017)	
$ASVI_{t-1}$			0.126^{***}						0.126^{***}
			(0.018)						(0.018)
$Volatility_t$				2.332^{**}				2.400^{*}	
				(0.878)				(0.967)	
$Volatility_{t-1}$					-0.331				-0.342
					(0.377)				(0.382)
$Return_t$						0.008		-0.237^{*}	
						(0.279)		(0.531)	
$Return_{t-1}$							-0.117		-0.058
							(0.097)		(0.117)
Observations	7,280	7,280	7,280	7,280	7,280	7,280	7,280	7,280	7,280
R^2	0.332	0.334	0.344	0.342	0.332	0.332	0.332	0.344	0.344
Adjusted R^2	0.330	0.331	0.341	0.339	0.330	0.329	0.330	0.341	0.341

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Note: Values in columns 1-9 stands for regression outputs to the variables in the corresponding rows. Column 1 is a single regression for dependent variable against independent variable in the first row. Columns 2-9 are multiple regressions of the dependent variable against the independent variables in the corresponding rows. Robust standard errors are reported in parentheses. Lagged variables only have 7280 observations compared to the non-lagged variables that have 7308. When you regress with both lagged and non-lagged variables, R (the software program) will then match to 7280 observations for data balancing.

Table 8. Regression results for volatility when ASVI is calculated from business terms

	Dependent variable:								
	Volatility								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$Volatility_{t-1}$	0.346***	0.346***	0.345***	0.337***	0.340***	0.354***	0.343***	0.345***	0.336***
	(0.091)	(0.091)	(0.091)	(0.093)	(0.094)	(0.088)	(0.084)	(0.091)	(0.088)
$ASVI_t$		0.001*						0.001	
		(0.0001)						(0.0001)	
$ASVI_{t-1}$			0.002*						0.002**
			(0.0002)						(0.0002)
$Volume_t$				0.005***				0.005***	
				(0.001)				(0.001)	
$Volume_{t-1}$					0.002				0.002
					(0.001)				(0.001)
$Return_t$						0.114		0.113	
						(0.108)		(0.109)	
$Return_{t-1}$							0.008		0.009
							(0.053)		(0.052)
Observations	7,280	7,280	7,280	7,280	7,280	7,280	7,280	7,280	7,280
R^2	0.120	0.120	0.121	0.135	0.121	0.165	0.120	0.180	0.123
Adjusted R^2	0.116	0.117	0.118	0.132	0.118	0.161	0.116	0.176	0.119

*p<0.1; **p<0.05; ***p<0.01

Note: Values in columns 1-9 stands for regression outputs to the variables in the corresponding rows. Column 1 is a single regression for dependent variable against independent variable in the first row. Columns 2-9 are multiple regressions of the dependent variable against the independent variables in the corresponding rows. Robust standard errors are reported in parentheses. Lagged variables only have 7280 observations compared to the non-lagged variables that have 7308. When you regress with both lagged and non-lagged variables, R (the software program) will then match to 7280 observations for data balancing.

4.3 ASVI in Different Time Horizons and Standardization Methods

This section presents the regression results for different ASVI values. As we mentioned before in section 3.3., we used two methods two methods to standardize ASVI: Bijl et al. (2016) and Da et al. (2011). We also used three different time horizons for our calculations, 8-week, 26-week and 52-week.

Table 9. Correlation matrix for ASVIs

Correlation Matrix

	ASVI 8-week (B)	ASVI 26-week (B)	ASVI 52-week (B)	ASVI 8-week (D)	ASVI 26-week (D)	ASVI 52-week (D)
ASVI 8-week (B)	1	0.86	0.80	0.70	0.66	0.64
ASVI 26-week (B)	0.86	1	0.95	0.67	0.74	0.73
ASVI 52-week (B)	0.80	0.95	1	0.66	0.74	0.76
ASVI 8-week (D)	0.70	0.67	0.66	1	0.94	0.92
ASVI 26-week (D)	0.66	0.74	0.74	0.94	1	0.98
ASVI 52-week (D)	0.64	0.73	0.76	0.92	0.98	1

Note: ASVI computed using Bijl et al. formula is denoted as (B). ASVI computed using Da et al. formula denoted as (D).

We expected the correlation within ASVIs to be high as they are calculated using the same method and is from the same time series (presented in Table 9). Correlation across ASVIs are closer to 1 as well since the source for computation is the same but relatively smaller compared to within ASVIs since the computation methods are different. Due to these observations, comparing the correlation for the different ASVIs is not enough. We also should assess the magnitude of change, to find out which relationship is greater among our ASVIs. Bijl et al.'s coefficients are stronger and easier to interpret because their standard deviation is close to 1.

The tables (Table 10, Table 11, and Table 12) are presented such that we only look at the descriptive and predictive power of the ASVI's on returns, trading volume and volatility with and without control variables. Hence, the table only consists of coefficients and R^2 values. The stars denote the significance level.

Table 10. ASVIs and returns

	Dependent Variable: Return											
	Da et al. (2011) - 8 weeks				Da et al. (2011) - 26 weeks				Da et al. (2011) - 52 weeks			
$Return_{t-1}$			-0.046	-0.021			-0.046	-0.021			-0.046	-0.021
$ASVI_t$	-0.004		-0.002		-0.003		-0.001		-0.003		-0.001	
$ASVI_{t-1}$		0.004		0.004		0.005		0.005		0.005		0.004
$Volatility_t$			0.382				0.382				0.382	
$Volatility_{t-1}$				-0.068				-0.068				-0.069
$Volume_t$			-0.002				-0.002				-0.002	
$Volume_{t-1}$				0.001				0.001				0.001
R^2	0.000	0.001	0.042	0.003	0.000	0.001	0.042	0.003	0.000	0.001	0.042	0.003
	Bijl et al. (2016) - 8 weeks				Bijl et al. (2016) - 26 weeks				Bijl et al. (2016) - 52 weeks			
$Return_{t-1}$			-0.046	-0.021			-0.046	-0.022			-0.046	-0.022
$ASVI_t$	-0.001		-0.001		0.000		0.000		0.001		0.000	
$ASVI_{t-1}$		0.002		0.002		0.002		0.002		0.002		0.002
$Volatility_t$			0.382				0.382				0.382	
$Volatility_{t-1}$				-0.069				-0.069				-0.069
$Volume_t$			-0.002				-0.002				-0.002	
$Volume_{t-1}$				0.001				0.001				0.001
R^2	0.000	0.000	0.042	0.003	0.000	0.001	0.042	0.003	0.000	0.000	0.042	0.003

Table 11. ASVIs and trading volume

	Dependent Variable: Volume											
	Da et al. (2011) - 8 weeks				Da et al. (2011) - 26 weeks				Da et al. (2011) - 52 weeks			
$Volume_{t-1}$	0.576***	0.570***	0.567***	0.571***	0.573***	0.567***	0.564***	0.568***	0.572***	0.566***	0.564***	0.568***
$ASVI_t$	0.151***		0.150***		0.161***		0.159***		0.162***		0.160***	
$ASVI_{t-1}$		0.245***		0.244***		0.234***		0.233***		0.230***		0.230***
$Volatility_t$			2.420*				2.412*				2.403*	
$Volatility_{t-1}$				-0.278				-0.287				-0.298
$Return_t$			-0.240				-0.241				-0.240	
$Return_{t-1}$				-0.059				-0.063				-0.062
R^2	0.336	0.342	0.346	0.342	0.337	0.342	0.347	0.342	0.337	0.342	0.347	0.342

	Bijl et al. (2016) - 8 weeks				Bijl et al. (2016) - 26 weeks				Bijl et al. (2016) - 52 weeks			
$Volume_{t-1}$	0.578***	0.571***	0.570***	0.573***	0.575***	0.567***	0.567***	0.569***	0.575***	0.567***	0.566***	0.569***
$ASVI_t$	0.069***		0.067***		0.077**		0.074***		0.075***		0.072***	
$ASVI_{t-1}$		0.148***		0.148***		0.131***		0.132***		0.129***		0.129***
$Volatility_t$			2.402*				2.385*				2.385*	
$Volatility_{t-1}$				-0.320				-0.337				-0.339
$Return_t$			-0.243				-0.249				-0.251	
$Return_{t-1}$				-0.071				-0.083				-0.088
R^2	0.336	0.348	0.345	0.348	0.337	0.346	0.347	0.346	0.336	0.345	0.346	0.345

Table 12. ASVIs and volatility

	Dependent Variable: Volatility											
	Da et al. (2011) - 8 weeks				Da et al. (2011) - 26 weeks				Da et al. (2011) - 52 weeks			
$Volatility_{t-1}$	0.346***	0.346***	0.345***	0.337***	0.346***	0.345***	0.345***	0.337***	0.345***	0.345***	0.345***	0.337***
$ASVI_t$	0.000		-0.001		0.001		-0.001		0.001		-0.0004	
$ASVI_{t-1}$		0.005***		0.005***		0.005***		0.005***		0.005***		0.005***
$Volume_t$			0.005***				0.005***				0.005***	
$Volume_{t-1}$				0.001				0.001				0.001
$Return_t$			0.113				0.113				0.113	
$Return_{t-1}$				0.009				0.009				0.009
R^2	0.120	0.122	0.179	0.123	0.120	0.122	0.179	0.124	0.120	0.122	0.179	0.124

	Bijl et al. (2016) - 8 weeks				Bijl et al. (2016) - 26 weeks				Bijl et al. (2016) - 52 weeks			
$Volatility_{t-1}$	0.346***	0.345***	0.345***	0.337***	0.345***	0.344***	0.345***	0.336***	0.345***	0.344***	0.345***	0.336***
$ASVI_t$	0.001		0.000		0.001		0.000		0.001*		0.001	
$ASVI_{t-1}$		0.003**		0.003**		0.003**		0.003**		0.003**		0.003**
$Volume_t$			0.005***				0.005***				0.005***	
$Volume_{t-1}$				0.002				0.001				0.001
$Return_t$			0.113				0.113				0.113	
$Return_{t-1}$				0.009				0.008				0.008
R^2	0.120	0.122	0.179	0.124	0.120	0.123	0.180	0.124	0.120	0.123	0.180	0.125

5. CONCLUSION

The aim of this paper is to investigate whether Google search activity can explain and predict the activity in the Norwegian stock market. Google classifies their data as a) search term, which keeps track of word-specific search queries; and b) business term, which keeps track of all search queries done in any language and are classified together under one topic through Google's algorithm. Our approach was to first check if ASVI can explain the three market measures: stock returns, trading volume, and volatility using their respective descriptive models. We then tested the predictive power of ASVI on these three measures using their respective predictive models. Both search terms and business terms behave in similar manner. Comparing their impact however made search term results more meaningful to look at.

Our results show that ASVI does not have a significant relationship nor does it have a predictive effect on abnormal returns. This is different from previous findings (Da et al., 2011; Bijl et al., 2016) and could be attributed to our small sample size. On the other hand, ASVI and trading volume have a significant relationship, and both ASVIs can describe and predict the movement of trading volume. This signifies that investors in the OSE uses information from Google along with other information channels in making investment decisions, which confirms the intuition from Da et al (2011). Lastly, ASVI does not have contemporary relation with the volatility, but it can predict future volatility. Altogether, we conclude that Google searches are not only related to, but can also predict trading activity measured by volatility and traded volume. Surprisingly, in this case, the predictive power of Google searches is even stronger than their contemporary explanatory power.

The findings we have for this paper can be researched further by expanding the sample of the dataset to include small cap companies. Lastly, due to time constraints, we were not able to develop a trading strategy based on information from SVI. We think this can be something worth considering.

REFERENCES

Alexa. (n.d.). Top Sites in Norway. Retrieved June 05, 2017, from http://www.alexa.com/topsites/countries/NO

Arellano, M. (1987) Computing robust standard errors for within-group estimators, *Oxford Bulletin of Economics and Statistics*, **49(4)**, pp. 431--434.

Banerjee, P. S., Doran, J. S., & Peterson, D. R. (2007). Implied volatility and future portfolio returns. Journal of Banking & Finance, 31(10), 3183-3199. doi:10.1016/j.jbankfin.2006.12.007

Bijl, L., Kringhaug, G., Molnár, P., & Sandvik, E. (2016). Google searches and stock returns. *International Review of Financial Analysis,45*, 150-156. doi:10.1016/j.irfa.2016.03.015

Bollerslev, T., & Zhou, H. (2009). Expected Stock Returns and Variance Risk Premia. SSRN Electronic Journal. doi:10.2139/ssrn.1315328

Castells, M. (1999). The information age: economy, society and culture. Oxford: Blackwell.

Challet, D., & Ayed, A. B. (2014). Predicting Financial Markets with Google Trends and Not so Random Keywords. *SSRN Electronic Journal*. doi:10.2139/ssrn.2310621

Compare Trends search terms - Trends Help. (n.d.). Retrieved May 28, 2017, from http://support.google.com/trends/answer/4359550

Da, Z., Engelberg, J., & Gao, P. (2011). In Search of Attention. *The Journal of Finance,66*(5), 1461-1499. doi:10.1111/j.1540-6261.2011.01679.x

Fama, E. F. (1998). Market Efficiency, Long-Term Returns, and Behavioral Finance. *SSRN Electronic Journal*. doi:10.2139/ssrn.15108

Harford, T. (2017, March 27). Just google it: The student project that changed the world. Retrieved June 04, 2017, from http://www.bbc.com/news/business-39129619

Joseph, K., Wintoki, M. B., & Zhang, Z. (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. *International Journal of Forecasting,27*(4), 1116-1127. doi:10.1016/j.ijforecast.2010.11.001

Kim, Y., Molnár, P., & Poulsson, A. E. (March 17, 2017). *Google Searches and Foreign Exchange Volatility*. Norwegian University of Science and Technology and University of Stavanger.

Kim, Y., Molnár, P., & Poulsson, A. E. (May 6, 2017). *Google Searches and Foreign Exchange Markets*. Norwegian University of Science and Technology and University of Stavanger.

Matias, Y. (2012, September 27). Insights into what the world is searching for -- the new Google Trends. Retrieved May 24, 2017, from http://search.googleblog.com/2012/09/insights-into-what-world-is-searching.html

Statistisk sentralbyrå (SSB). (n.d.). Norwegian Media Barometer. Retrieved June 05, 2017, from http://www.ssb.no/en/kultur-og-fritid/statistikker/medie

Molnár, P. (2012). Properties of range-based volatility estimators. *International Review of Financial Analysis,23*, 20-29. doi:10.1016/j.irfa.2011.06.012

Naes, R., & Ødegaard, B. A. (2009). Liquidity and Asset Pricing: Evidence on the Role of Investor Holding Period. SSRN Electronic Journal. doi:10.2139/ssrn.1095226

Oslo Børs. (n.d.). About Oslo Børs. Retrieved May 15, 2017, from http://www.oslobors.no/ob_eng/Oslo-Boers/About-Oslo-Boers

Oslo Børs. (n.d.). OBX Total Return Index. Retrieved May 15, 2017, from http://oslobors.no/ob_eng/markedsaktivitet/#/details/OBX.OSE/overview

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports,3*(1). doi:10.1038/srep01684

Preis, T., Reith, D., & Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,368*(1933), 5707-5719. doi:10.1098/rsta.2010.0284

Tamir, N. (2015, June 17). A new window into our world with real-time trends. Retrieved May 24, 2017, from http://googleblog.blogspot.no/2015/06/a-new-window-into-our-world-with-real.html?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A%2Bblogspot%2FMKuf%2B%28Official%2BGoogle%2BBlog%29

Torres-Reyna, O. (2007, December). *Panel Data Analysis Fixed and Random Effects Using Stata (v.4.2)* [PDF]. Princeton University Press.

Varian, H. R., & Choi, H. (2011). Predicting the Present with Google Trends. *SSRN Electronic Journal*. doi:10.2139/ssrn.1659302

Weinberg, N. (2007, July 30). Success! Google Trends Updated. Retrieved May 24, 2017, from http://google.blognewschannel.com/archives/2007/07/30/success-google-trends-updated/

Where Trends data comes from - Trends Help. (n.d.). Retrieved May 22, 2017, from http://support.google.com/trends/answer/4355213?hl=en&ref_topic=6248052

World Bank. (n.d.). Internet users as percentage of population. Retrieved June 05, 2017, from http://www.google.no/publicdata/explore?ds=d5bncppjof8f9_&met_y=it_net_user_p2&idim=country%3ANOR%3ASWE%3ANLD&hl=en&dl=en#!ctype=l&strail=false&bcs=d&nselm=h&met_y=it_net_user_p2&scale_y=lin&ind_y=false&rdim=region&idim=country:NOR:SWE:NLD&ifdim=region&hl=en_US&dl=en&ind=false

Ødegaard, B. (n.d.). Asset pricing data at OSE. Retrieved May 30, 2017, from http://finance.bi.no/~bernt/financial_data/ose_asset_pricing_data/index.html

APPENDICES

APPENDIX 1: Companies and Google search words used

Arranged according to the following pattern: st, bt - category, **company name** (remark, if any):

DNB, DNB ASA - Financial services company, **DNB**; dno, DNO ASA - Company, **DNO**; fred olsen, Fred. Olsen & Co. - Holding company, **Fred. Olsen Energy**; gjensidige, Gjensidige - Insurance company, **Gjensidige Forsikring**; GOGL, Tor Olav Trøim - Businessman, **Golden Ocean Group**; Marine Harvest, Marine Harvest - Company, **Marine Harvest** (insufficient data); Norsk Hydro, Norsk Hydro - Company, **Norsk Hydro**; norwegian, Norwegian Air Shuttle - Airline, **Norwegian Air Shuttle**; opera, Opera - Web browser, **Opera Software**; Orkla, Orkla Group - Conglomerate company, **Orkla**; pgs, Petroleum Geo-Services - Petroleum industry company, **Petroleum Geo-Services**; prosafe, Prosafe - Company, **Prosafe**; rec, Renewable Energy Corporation - Company, **REC Silicon**; Schibsted, Schibsted - Media company, **Schibsted A**; Seadrill, Seadrill - Oil industry company, **Seadrill** (insufficient data); Statoil, Statoil - Oil industry company; **Statoil**; Storebrand, Storebrand - Insurance company, **Storebrand**; subsea, Subsea 7 - Engineering company, **Subsea 7**; Telenor, Telenor - Telecommunications company, **Telenor**; TGS, TGS-Nopec Geophysical Company - Company, **TGS-NOPEC**; Yara, Yara International - Chemicals company, **Yara International**; **Aker BP** (insufficient data); BAKKA, Faroe Islands - Island country, **P/F Bakkafrost**; fro, Frontline Ltd. - Company, **Frontline Ltd.**; Grieg, Grieg Seafood - Company, **Grieg Seafood**; Lerøy, Ole Rasmus Møgster - Topic, **Lerøy Seafood**; salmar, SalMar - Company, **SalMar**; Songa, Songa Offshore - Topic, **Songa Offshore**; emgs, Electromagnetic Geoservices - Company, **ElectroMagnetic GeoServices**; polarcus, Polarcus - Company, **Polarcus**; semiconductor, Nordic Semiconductor - Semiconductor company, **Nordic Semiconductor**; **Schibsted B** (insufficient data)

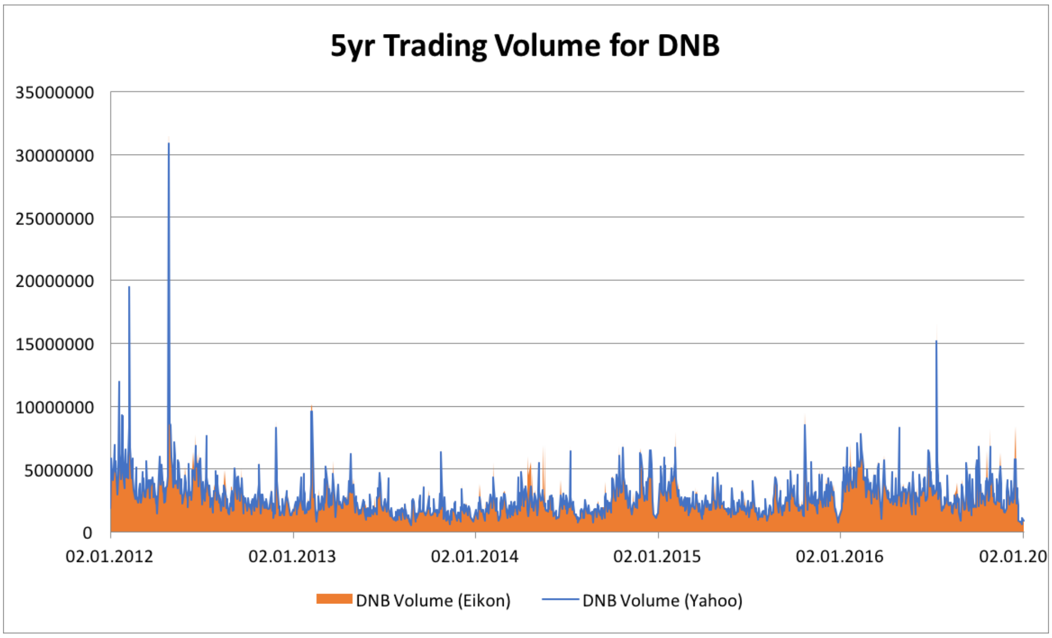
APPENDIX 2: Yahoo! Finance and Eikon Data

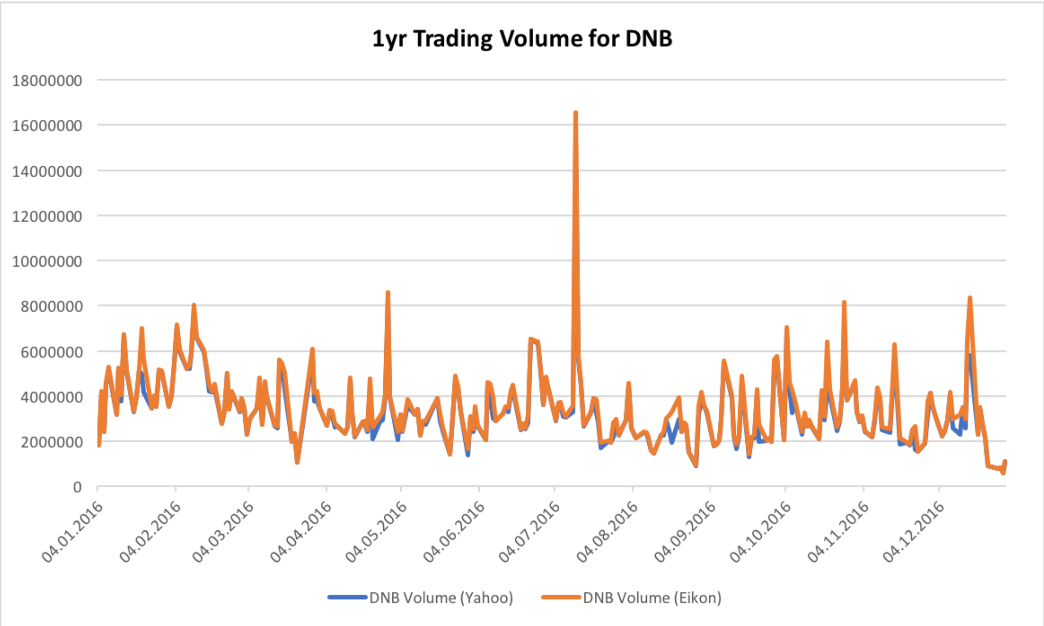
As we mentioned in section 3.2, Yahoo! had reported external issues with their information and security system at the time that we conducted our study. Their finance platform also warned us of possible data unreliability and delayed reporting for financial markets outside the United States of America. We thus obtained data from Thomson Reuter's Eikon to verify Yahoo! Finance data. We did a side-by-side comparison of figures from the two sources, checking for differences between the open, high, low, close price, and trading volume for each 28 companies in our dataset.

We found out that the reported open, high, low, and close price for both sources were mostly the same for each company, with some occurrence of differences attributed to rounding errors. For the sake of brevity, the table presented below shows the side-by-side comparison of data from DNB for the most recent month in our dataset.

	OPEN		HIGH		LOW		CLOSE		VOLUME	
DNB data	**YF**	**Eikon**	**YF**	**Eikon**	**YF**	**Eikon**	**YF**	**Eikon**	**YF**	**Eikon**
02/12/2016	127.0	127.0	128.4	128.4	125.7	125.7	126.0	126.0	3,173,800	3,174,514
05/12/2016	125.2	125.2	128.4	128.4	125.2	125.2	128.0	128.0	2,212,600	2,213,428
06/12/2016	127.7	127.7	129.9	129.9	127.3	127.3	129.9	129.9	2,449,100	2,452,480
07/12/2016	130.0	130.0	132.0	132.0	129.9	129.9	132.0	132.0	2,950,500	2,951,747
08/12/2016	132.4	132.4	135.4	135.4	132.2	132.2	135.4	135.4	4,044,600	4,173,788
09/12/2016	135.4	135.4	136.8	136.8	104.1	134.2	135.8	135.8	2,572,500	2,961,611
12/12/2016	136.5	136.5	138.2	138.2	135.8	135.8	136.9	136.9	2,307,000	3,193,738
13/12/2016	134.2	134.2	136.0	136.0	133.8	133.8	135.3	135.3	3,514,500	3,515,016
14/12/2016	134.8	134.8	135.0	135.0	133.6	133.6	134.0	134.0	2,588,800	2,990,666
15/12/2016	133.7	133.7	134.9	134.9	130.0	130.0	130.0	130.0	5,810,900	6,436,077
16/12/2016	130.3	130.3	131.7	131.7	129.4	129.4	131.2	131.2	5,798,900	8,350,437
19/12/2016	131.7	131.7	131.7	131.7	128.7	128.7	128.9	128.9	2,298,700	2,300,670
20/12/2016	129.0	129.0	129.7	129.7	127.8	127.8	128.7	128.7	3,508,300	3,516,405
21/12/2016	128.7	128.7	129.9	129.9	127.3	127.3	127.4	127.4	2,732,800	2,733,836
22/12/2016	127.5	127.5	129.2	129.2	126.9	126.9	129.1	129.1	2,140,100	2,140,206
23/12/2016	129.8	129.8	130.2	130.2	128.5	128.5	129.2	129.2	909,400	910,441
27/12/2016	129.4	129.4	129.7	129.7	128.7	128.7	129.0	129.0	784,400	787,640
28/12/2016	128.9	128.9	129.9	129.9	128.4	128.4	128.9	128.9	825,200	825,209
29/12/2016	129.0	129.0	129.2	129.2	128.3	128.3	129.0	129.0	607,000	609,656
30/12/2016	128.6	128.6	128.6	128.6	127.5	127.5	128.4	128.4	1,118,600	1,120,053
02/01/2017	128.0	128.0	129.9	129.9	127.6	127.6	129.6	129.6	869,700	876,229

Differences in trading from both sources looks greater compared to the other prices, but this is due mainly to scaling reasons. When scaled such that 1 million becomes 1.0, then the difference would seem less significant. Another way to intuitively view that the figures are close and almost similar is to illustrate the two figures in a side-by-side graph, as shown in the figures below.





APPENDIX 3: Other tables

The two tables below relate to section 4 and summarize the results for our models when we replace weekly with monthly variables. Results for returns were excluded because of insignificant values.

	Dependent variable:								
	Volatility								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$MonthlyVolatility_{t-1}$	0.416***	0.413***	0.414***	0.397***	0.403***	0.435***	0.412***	0.414***	0.398***
	(0.039)	(0.038)	(0.038)	(0.036)	(0.037)	(0.027)	(0.047)	(0.022)	(0.046)
$MonthlyASVI_t$		0.003**						0.002*	
		(0.000)						(0.000)	
$MonthlyASVI_{t-1}$			0.002*						0.002*
			(0.000)						(0.001)
$MonthlyVolume_t$				0.004***				0.004***	
				(0.001)				(0.001)	
$MonthlyVolume_{t-1}$					0.003*				0.003*
					(0.001)				(0.001)
$Return_t$						0.115		0.114	
						(0.102)		(0.102)	
$Return_{t-1}$							0.039		0.042
							(0.039)		(0.041)
Observations	7,280	7,280	7,252	7,280	7,252	7,280	7,280	7,280	7,252
R^2	0.073	0.074	0.074	0.079	0.075	0.118	0.078	0.125	0.082
Adjusted R^2	0.069	0.070	0.070	0.076	0.072	0.115	0.074	0.121	0.078

*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:								
	Volume								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$MonthlyVolume_{t-1}$	0.669^{***}	0.659^{***}	0.663^{***}	0.663^{***}	0.672^{***}	0.669^{***}	0.669^{***}	0.653^{***}	0.667^{***}
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
$MonthlyASVI_t$		0.115^{***}						0.113^{***}	
		(0.029)						(0.029)	
$MonthlyASVI_{t-1}$			0.053^{*}						0.054^{*}
			(0.024)						(0.024)
$MonthlyVolatility_t$				1.033				0.914	
				(0.707)				(0.193)	
$MonthlyVolatility_{t-1}$					-0.586				-0.626
					(0.586)				(0.556)
$Return_t$						0.025		0.004	
						(0.338)		(0.990)	
$Return_{t-1}$							-0.092		-0.091
							(0.206)		(0.186)
Observations	7,252	7,252	7,252	7,252	7,252	7,252	7,252	7,252	7,252
R^2	0.271	0.275	0.272	0.272	0.271	0.271	0.271	0.276	0.272
Adjusted R^2	0.268	0.273	0.269	0.269	0.268	0.268	0.268	0.273	0.269

$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$