



SCHOOL OF
ECONOMICS AND
MANAGEMENT

Google Searches and Stock Volatility

Evidence from the Danish market

Bachelor thesis in Economics
Department of Economics, Lund University

Author:
Nyström, Oliver Albin Wiegaard

Supervisor:
Anders Vilhelmsson

Course:
NEKH03

Date:
May 25, 2021

Abstract

In this thesis, we investigate the effect of online search activity approximated by Google searches on the volatility of the largest, most actively traded stocks listed on the Danish stock exchange. The interest in this relationship stems from an increasing number of retail investors entering the market in recent years. Driven by low-to-negative deposit rates, easily accessible online trading platforms with limited entry regulation, and most recently, the Covid-19 pandemic, we argue that online information seeking of “amateur investors” with little to no prior trading experience can be linked to the dynamics of return volatility. By conducting both statistical and regression analyses we investigate a dataset containing information on 27 Danish stocks for the years 2016-2021. Contrary to our expectations, we fail to establish a descriptive link between Google search activity and volatility at the market level, nor do we find predictive powers of Google searches on volatility patterns for our sample. Each result is somewhat controversial and lacks support in most of the established body of literature covering developed markets. We hypothesize that this can be explained by the methodology of which search data was obtained, leaving valuable knowledge for future investigators.

Acknowledgement

I would like to thank my supervisor, Professor Anders Vilhelmsson for his continued support in writing my thesis. The completion of this work is largely owed to his extensive knowledge and invaluable feedback, which has helped me immensely. I would also like to extend my gratitude towards my ever-supporting grandmother, Inger Nyström, for her invaluable input.

Table of Contents

1. Introduction	4
2. Literature review	5
3. Limitations	9
3.1 Robustness	9
3.1 Google SVI	9
4. Data	10
4.1 Selection of stocks	10
4.2 Stock performance data	11
4.3 Google Trends data	11
5. Methodology	14
5.1 Problem analysis	14
5.2 Statistical analysis	15
5.2.1 Hypothesis	15
5.2.2 Abnormal search volume index	15
5.2.3 Realized volatility	16
5.2.4 Student's t-test	16
5.3 Predictive regressions	18
5.3.1 Regression variables	19
5.3.2 The autoregressive model	20
6. Results	20
6.1 Descriptive analysis	22
6.1.1 Obtaining t-test scores	22
6.1.2 Results	23
6.2 Predictive analysis	25
6.2.1 Results	26
7. Conclusion	27
7.1 Further steps	28
Bibliography	29
Journal Articles	29
Working papers	32
Web pages	32
News articles	33
Appendix	34
Appendix A – result tables	34
Appendix B – Stata do files	36

1. Introduction

For years, researchers have investigated the driving forces of stock market volatility. A general agreement within the field lies in the belief that information is the most highly demanded asset in financial markets and must therefore be a key explanatory of volatility patterns. As a result of this, one area of the research focuses strictly on the link between information flow and volatility. A well-known hypothesis trying to explain this relationship states that measures of market activity, e.g. the volatility of returns, can be directly linked to the amount of information arriving in the market [(Clark, 1973), (Epps & Epps, 1976), (Tauchen & Pitts, 1983), (Darolles, et al., 2017)]. The Mixture of Distributions Hypothesis (MDH henceforth) explains the relationship between volatility and trading volume by introducing joint dependence of both trading volume and returns on the inflow of information in the market. Therefore, the MDH argues, the behavior of market volatility and trading volume reflects that of information flowing into the market [(Clark, 1973), (Epps & Epps, 1976), (Tauchen & Pitts, 1983), (Darolles, et al., 2017)].

However, historically researchers have struggled to find a viable direct source of such information flows. This has changed in recent years with digitalization forcing an increasing fraction of information seeking to be conducted online. The most popular online platform used to search for this type of information is Google. Since 2004, Google has made search data obtained through their search engine public, initially through Google Insights, which is known as Google Trends today. What sets Google search data apart from previous proxies of information demand is the fact that it quantifies the behavior of retail investors, which to a larger extent relies on information available online, e.g., news articles, when making investment decisions (Da, et al., 2011). Therefore, the introduction of Google search data as a direct approximation of information demand has recently been found to be a reliable source of data. This has especially become evident during the Covid-19 pandemic, with the number of retail investors increasing substantially [(Glenwood, 2021), (Winding, 2021)].

In this thesis, the relationship between information demand proxied by Google search data (SVI henceforth) and volatility of Danish stocks is investigated. We aim to address the following

research question: Does the increasing number of retail investors seeking information for prospective investments online imply that the online search activity for a given stock can be linked to the volatility of that stock? To explain this relationship better, we conduct both descriptive and predictive analyses. For the descriptive analysis, we perform a Student's t-test to investigate whether such a relationship can be explained with statistical significance or not. For the second analysis, we aim to explore the predictive power of SVI on volatility utilizing an autoregressive model. For each test, we analyze the sample of stocks both at an aggregate level to identify a general trend and individually. Furthermore, we classify all stocks in the sample as per the Global Industry Classification Standard (GICS henceforth), to investigate whether any sector specific trends can be explained.

The rest of the study is organized in the following way. Chapter 2 introduces a literature review covering past research related to our chosen topic. Chapter 3 discusses possible limitations associated with our findings. Chapter 4 introduces the reader to the data selection process. Here, we first touch on our approach to securing a stock sample well fitted for a study on information demand and the source of stock performance data. Additionally, we present the reader with a formal description of how SVI is constructed and discuss the chosen approach for creating the search queries chosen for collection. Chapter 5 states the research questions, and introduces the chosen methodology used in the experiments. The first part of the chapter focuses on the statistical testing used for the descriptive analysis while the second part presents the autoregressive model used for prediction. Chapter 6 takes the reader through the experimental procedure. We touch on the software used to conduct the analyses, discuss challenges associated with comparing our results to prior findings and outline the GICS classification. Furthermore, we present our experimental findings, with a comparison to evidence from past research where possible. Chapter 8 concludes the experimental findings of the thesis. The significance of the results is discussed along with our contribution to the state of the art. We conclude with an outline of possible improvements and further steps related to the study.

2. Literature review

In recent years, the relation between online search activity and asset performance has attracted a lot of attention. In the first of such studies, (Mondria, et al., 2010) uses the AOL click-through

series, a sample of online search queries performed by more than half a million anonymous users covering a three-month period, to investigate the link between home bias and attention allocation. The study finds empirical evidence of a bi-directional relationship. (Da, et al., 2011) first introduced the use of SVI data as a direct measure of investor intention. Analyzing a sample of Russell 3000 stocks, they find SVI to be correlated with but different from existing approximations of investor attention. Additionally, they find SVI to have significant predictive capabilities. They find that an increase in SVI predicts increasing returns in the subsequent two weeks. However, they conclude that these positive short term price effects are reversed within the following year. Furthermore, the investigation links SVI to large first-day returns while undermining long-run performance for IPO stocks. Similarly, (Joseph, et al., 2011) examines the ability of SVI, constructed with tickers as keywords, to predict abnormal stock returns and trading volume. The study, which investigates a sample of S&P 500 stocks, concludes that online search intensity is a reliable predictor of abnormal stock returns and trading volume. Furthermore, they find significant evidence of a positive link between SVI, the volatility of returns and the difficulty of which a stock can be arbitrated.

The mentioned findings align with the price pressure hypothesis formulated by (Barber & Odean, 2008). They note that, when individuals are buying stocks, they face a decision problem. There is a large selection of stocks to choose from which all have different levels of performance potential. When facing such a problem, the benefit of obtaining information to help the investor make the optimal buying decision is relatively high. However, investors do not face this problem when selling because they tend to make the selling decision based on past returns, which is typically available on the online trading platform at which the transactions are taking place. As a result of this, they hypothesize that attention determines what choices are available before preferences determine the actual choice. Therefore, an increase in the search intensity for a stock should be accompanied by an increase in buying pressure, resulting in a rise in the stock price. However, this has been challenged by (Bilj, et al., 2016). In their study of the link between SVI for company names and returns of S&P 500 stocks, they link increasing investor attention to negative returns.

The link between SVI and the U.S. stock market has also been investigated with a strict focus on volatility. (Vlastakis & Markellos, 2012) study information demand and supply using SVI for 30

of the largest stocks traded on the New York Stock Exchange and Nasdaq. They find market level information demand to be positively correlated with historical and implied measures of volatility and trading volume. They also find significant evidence that information demand increases in periods of higher returns. Furthermore, they find empirical evidence for the hypothesis that as investors grow increasingly risk averse, they demand more information.

The link between online search activity and performance of other types of assets has also been studied. (Vozlyublennaiia, 2014) investigates the link between Google search probability and performance of indexes in broad investment categories. They find that short term return of a given index can be linked to searches for that index. Specifically, after a period of rising attention, returns can either significantly increase or decrease, which depends on the investment horizon. These findings differ from previous literature, where increasing searches create a buying pressure on prices. Instead their results point towards retail investors creating either a buying [(Da, et al., 2011), (Joseph, et al., 2011)] or selling pressure (Bilj, et al., 2016), depending on the nature of the information discovered. Furthermore, they find that the return shock leads to a long-term increase in attention, which reduces the predictability of index returns from SVI. This leads to diminishing index volatility, ultimately improving market efficiency. SVI data has also been used to investigate the driving forces behind volatility in the Bitcoin market. (Byström & Krygier, 2018) investigate whether Bitcoin volatility can be explained by retail investor driven internet search volumes. They find a positive link between Bitcoin volatility and SVI on related keywords, especially for the word “Bitcoin”. They also find SVI to have strong predictive skills on Bitcoin volatility.

Though initial research has primarily been focused on the US stock market, other markets have recently attracted attention from researchers. (Gwilym, et al., 2016) examines how Chinese stocks are affected by investors’ speculative demand, which they construct from SVI on the keyword “concept stocks”. They find evidence of a positive link between SVI and returns. (Bank, et al., 2011) find that an increase in SVI is linked to higher trading activity, improved liquidity, and higher future short-run returns of German stocks. In another German study, (Fink & Johann, 2014) examine the link between short-run changes in SVI and liquidity, turnover, volatility and returns. They find that turnover and volatility of large stocks, stocks with a lower, cross-sectional SVI level, and stocks popular among retail investors significantly increase on days with high search

volumes. (Aouadi, et al., 2013) investigate the French market, finding a link between SVI and trading volume. They also find significant descriptive power of SVI on both stock market illiquidity and volatility. (Takeda & Wakao, 2014) investigate the relationship between SVI and trading behavior in the Japanese stock market, finding strong evidence for a positive correlation between SVI and trading volume, while finding a weaker relationship between SVI and stock prices. [(Swamy, et al., 2019), (Swamy & Dharani, 2019), (Aziz & Ansari, 2021)] examine the link between Google searches and Indian stocks, all finding a positive link between search activity and stock returns.

However, certain investigations focusing on smaller markets have explored the link between returns, volatility, and online search activity with mixed results. These findings are of a particular interest since we are conducting an investigation on a smaller market as well. (Kim, et al., 2019) investigates whether SVI can explain current and predict future returns, trading volume, and volatility of Norwegian stocks. They conclude that, while SVI does predict increased volatility and trading volume, it fails to find significant predictive power on future abnormal returns. Similarly, (Osarumwense, 2021) examines whether investors' online information demand can be used to describe or forecast the dynamics of returns, trading volumes and volatility for Nigerian stocks. Using both SVI and Wikipedia page views as a measure of information demand, they provide robust evidence that neither Google searches, nor Wikipedia clicks can explain or predict stock returns, trading volumes or volatility. On the contrary, (Bui & Nguyen, 2019) find a strong positive correlation between SVI and trading volume, stock liquidity and volatility for Vietnamese stocks. Similarly, (Tan & Taş, 2019), who investigate the Turkish stock market, find that firms attracting higher attention measured by SVI earn higher returns, with the price pressure being stronger on smaller companies. They also find the predictability of SVI on abnormal returns to persist for three weeks but note that this is followed by a price reversal in the following year.

The past research has strongly influenced our work. Our approach to standardizing SVI and approximating latent variance follows the methodology of (Da, et al., 2011) and (Andersen & Bollerslev, 1998) respectively, which will be described in chapter 5. Furthermore, the autoregressive model employed for the predictive analysis largely resembles the model used in (Kim, et al., 2019). The state-of-the-art shows that in most cases SVI has been a strong predictor

of stock performance. However, to the best of our knowledge, this is yet to be tested on the Danish market.

3. Limitations

The limitation of our work lies primarily within uncertainties associated with the use of Google data as a measure of online search activity and a lack of additional testing done to secure robustness of the results.

3.1 Robustness

For this study, a limitation lies in the fact that we do not evaluate our results in comparison with different model alternatives. This could have been done using different means of standardization in respects to SVI, e.g., following the methodology defined in (Bilj, et al., 2016), or by changing the number of weeks used to obtain the standardized SVI values. Furthermore, we could have used a different approach to modelling volatility, by conducting our analysis with SVI obtained using the Finance filter or narrowed the areas at which searches were included, as opposed to the default versions.

3.2 Google SVI

We have identified a number of uncertainties associated with the use of SVI as a proxy for general online search activity. As noted in (Fink & Johann, 2014), one problem that arises when using SVI as a measure of the demand for information is that the construction of the search volume index itself remains relatively unknown to users since the actual click rates are not published. They also note that the measures taken to convert actual search queries into SVI are not known to the user which makes it unclear whether SVI is what it claims to be.

Another challenge identified in past research is the lack of certainty connected to keywords used to query the index. (Da, et al., 2011) use stock tickers to construct their search queries but note problems associated with ‘noisy’ tickers with generic meaning. This results in stocks with ticker symbols such as ‘DNA’, ‘BABY’ and ‘ALL’ being omitted from their sample. Authors also discuss the possibility that a keyword’s SVI may differ depending on when it was downloaded.

This problem arises because Google calculates SVI from a random subset of historical search data. However, (Bilj, et al., 2016) challenge this approach, noting that searches for company names are more efficient in explaining stock returns than ticker symbols. This adds to the mentioned uncertainty of which queries to use. Furthermore, [(Vlastakis & Markellos, 2012), (Takeda & Wakao, 2014)] note that certain components of company names are irrelevant to investment purposes. Based on an assumption that these components are random, (Takeda & Wakao, 2014) create a list of abbreviation from company names such as ‘Co.’, ‘Ltd.’, ‘Inc.’ and ‘Holdings’ which they exclude from their search queries.

Additionally, we, along with past research, identify the lack of certainty associated with the fact that not all searches for the keywords included in this study are necessarily related to investors seeking information for trading causes (Vozlyublennaiia, 2014). Furthermore, we note that using Google SVI as a proxy for online search activity does not cover all search activity on a global scale. For instance, Baidu has the largest market share in China while Yandex is the largest search engine in Russia as per HubSpot (Forsey, 2019). However, as to the best of our knowledge, SVI is the best source of such approximation with Google having a global market share of 86.6% as per February 2021 (Johnson, 2021).

4. Data

4.1 Selection of stocks

In Denmark, financial instruments are generally traded through two different marketplaces. Nasdaq Copenhagen serves as the central marketplace, while Nasdaq First North Premier Growth Market serves as an alternative marketplace for smaller companies across the Nordic region.

For this study, we focus on stocks represented in the OMX C25 index from Nasdaq Copenhagen. The OMX C25 index consists of the 25 largest and most actively traded equities on the Danish main market. The index, which began in December 2016, is an expansion of the former OMX C20 CAP index. The whole period that this study covers is from March 2016 to March 2021. The OMX C25 index is the primary source of our selection of stocks covering stock data from January 2017 to March 2021. For the period from March 2016 to December 2016 we included stocks represented

in the OMX C20 CAP index. This index was chosen because of its close resemblance to the OMX C25 index.

The OMX C25 index is revised twice a year in June and December and the selection process has two primary steps: initially, the 35 listed shares with the highest free float market cap are selected for further screening. Of these 35 shares, the 25 shares with the highest trading volume during the past six months are selected and included in the index. Stocks with high trading volumes indicate a larger pool of interested investors who will likely want to search for information on these stocks. This makes the stocks from the C25 index suitable for an investigation such as the one we are conducting. The index weights of the 25 shares selected are based on the free float adjusted market cap to ensure that it is exclusively shares available for trading that are included (Nasdaq, 2020). The availability of the stocks further supports investor incentive to seek stock information. Within the specific 5-year period selected for this study, a total of 33 stocks were included in the two indexes, which formed our initial sample. The initial sample was then screened for price data through Yahoo! Finance and search data through Google Trends. Due to either a lack of historical price data or misleading keywords when collecting search data, 6 stocks were excluded. Thus, our final sample consists of 27 stocks.

4.2 Stock performance data

For each of the stocks included in the sample, performance data is obtained from Yahoo! Finance. Our data covers 261 weeks and is collected at a daily frequency. For each of those 261 weeks, we obtain historical data on daily opening price, high, low, closing price, adjusted closing price and trading volumes. The adjusted closing price adjusts for both dividends and splits, making it a suitable measure of company performance over time (Yahoo! Finance, 2021).

4.3 Google Trends data

Online search data is collected from Google Trends. Google Trends is a public tool which quantifies the frequency at which a specific search query is searched for on Google relative to the total search volume, over a defined range of time. Google quantifies this as a Search Volume Index

(SVI), which is calculated using the daily search volume for the query of interest. The SVI is calculated primarily through three steps. First, search interest is calculated as follows:

$$(4.1) \text{ Search_interest}_t = \frac{\text{Number of searches for specific query on day } t}{\text{Total number of Google search queries on day } t}$$

Then, within the date range defined by the user, each of the resulting search interests is divided by the highest interest point for the specific query. Lastly, search interest is indexed to relative values ranging from 0 to 100. By indexing the search volume using the aforementioned methodology, one controls for changes in general search activity over time (Google, 2021). The user can filter the tool exclusively to include searches for a specific geographical area. However, for this particular study, we chose to include search data on a global scale. Google Trends also allows the user to filter for specific categories. Amongst these is a finance filter. However, we selected the default filter, ‘All Categories’ since the finance filter returned too many null values. Lastly, Google Trends filters data according to what type of search was conducted: ‘Web Search’, ‘Image Search’, ‘News Search’, ‘Google Shopping’ or ‘YouTube Search’ (Google, 2021). We chose ‘Web Search’ as our search type.

SVI can be collected in both real time and non-real time samples. Real-time samples cover the last seven days while non-real time samples are separate samples that can be collected as far back as 2004 up to 36 hours before the search in question. As we are looking for search data covering an extensive period, non-real time data samples are chosen for collection. Non-real time data can be collected in different time spans ranging from the past hour till 2004 to present as well as custom-made ranges. Depending on the range of the dataset in question, the time frequency of the of the data varies. For example, the maximum time range for a set of daily SVI is 90 days. Due to this restriction, our dataset consists of weekly SVI covering a maximum allowed 5-year period between March 2016 and March 2021 (Google, 2021).

For the collection of SVI, we use a combination of ticker symbols and company names to complete our search queries. The former increases the likelihood that a substantial fraction of the search activity included is in direct relation to the chosen stocks (Da, et al., 2011). For example, an

unemployed sailor searching for vacant positions is less inclined to search for ‘MAERSK-B’ or ‘MAERSK-A’ than he is to search for ‘Maersk’. However, searching only for ticker symbols resulted in insufficient search volumes for several stocks, resulting in a lack of available SVI. Additionally, several ticker symbols either have generic meaning (see Topdanmark, Danske Bank) or are abbreviations of company names. To combat this problem, we decided to either add or exclusively use the latter for the queries. As noted in (Bilj, et al., 2016), searching for the name of firms explains stock returns better than ticker symbols. We rely on their results as justification for this inclusion. Furthermore, words that are either too general or leads to misleading results were excluded. For example, searches for ‘Nets’ and their ticker symbol ‘NETS’ returned relatively large search volumes in both the United States and Singapore. We found that the high U.S volumes could be related to search queries for the Brooklyn Nets, an NBA basketball team while the high Singaporean search volumes could be related to the Singapore-based Network for Electronic Transfers (NETS). As a result, we chose to omit Nets A/S from the sample. We also excluded the term ‘A/S’ based on the findings of (Takeda & Wakao, 2014). Table 6.1 presents the stocks included along with their ticker symbols and the search terms we used to obtain SVI. We note that company names and ticker symbols are listed as described by (Yahoo! Finance, 2021).

COMPANY NAME	Stock ticker	Google search query
<i>AMBU A/S</i>	AMBU-B	AMBU-B +AMBU
<i>Bavarian Nordic A/S</i>	BAVA	BAVA+Bavarian Nordic
<i>Carlsberg A/S</i>	CARL-B	CARL-B+ Carlsberg
<i>Chr. Hansen A/S</i>	CHR	CHR+Chr. Hansen
<i>Coloplast A/S</i>	COLO-B	COLO-B+Coloplast
<i>Danske Bank A/S</i>	DANSKE	Danske Bank
<i>Demant A/S</i>	DEMANT	DEMANT+Demant
<i>DSV Panalpina A/S</i>	DSV	DSV+DSV Panalpina
<i>FLSmidth & Co. A/S</i>	FLS	FLS+FLSmidth & Co.
<i>Genmab A/S</i>	GMAB	GMAB+Genmab
<i>GN Store Nord A/S</i>	GN	GN+GN Store Nord
<i>ISS A/S</i>	ISS	ISS+ISS
<i>Jyske Bank A/S</i>	JYSK	JYSK+Jyske Bank
<i>H. Lundbeck A/S</i>	LUN	LUN+H. Lundbeck
<i>A.P. Møller – Mærsk A/S</i>	MAERSK-A	MAERSK-A+AP Moller Maersk
<i>A.P. Møller – Mærsk A/S</i>	MAERSK-B	MAERSK-B+A.P. Moller Maersk
<i>Nordea Bank Danmark A/S</i>	NDA-DK	NDA-DK+Nordea Bank Danmark
<i>Novo Nordisk A/S</i>	NOVO-B	NOVO-B+Novo Nordisk
<i>Novozymes A/S</i>	NZYM-B	NZYM-B+Novozymes
<i>Pandora A/S</i>	PNDORA	PNDORA+Pandora
<i>Rockwool International A/S</i>	ROCK-B	ROCK-B+Rockwool International
<i>Royal Unibrew A/S</i>	RBREW	RBREW+Royal Unibrew
<i>SimCorp A/S</i>	SIM	SIM+SimCorp
<i>Sydbank A/S</i>	SYDB	SYDB+Sydbank
<i>Topdanmark Forsikring A/S</i>	TOP	Topdanmark Forsikring
<i>Tryg Forsikring A/S</i>	TRYG	TRYG+Tryg Forsikring
<i>Vestas Wind Systems A/S</i>	VWS	VWS+Vestas Wind Systems

Table 4.1: Company name, stock ticker and Google search query for each stock in the sample.

5. Methodology

5.1 Problem analysis

We aim to investigate the effect of Google search activity on the volatility of Danish stocks utilizing two different methodological approaches. The first experiment focuses strictly on the descriptive power of Google search volume on volatility. Specifically, we will assess correlations between abnormal SVI and realized volatility using statistical testing. The results will help us determine whether such correlations exist and, if so, in which way SVI affects volatility.

The second experiment aims to investigate the predictive power of Google search activity on volatility. For this aim, we perform predictive regressions using a lagged abnormal SVI variable to test whether past search volumes can be used to predict future realized stock volatility. The

results will help us determine whether and how well realized stock volatility can be projected from Google search activity.

5.2 Statistical analysis

5.2.1 Hypothesis

We aim to investigate the relationship between realized volatility among our selection of Danish stocks and their corresponding SVI. The research question we aim to address is: “Is there a statistically significant correlation between Google search volumes and volatilities of Danish stocks?”. Based on the vast results from previous studies, we hypothesize that SVI does affect volatility. Thus, our high-level research hypothesis is: “There is a statistically significant correlation between Google search volumes and volatility in Danish stocks.”, and the null hypothesis is: “There is no statistically significant correlation between Google search volumes and volatility in Danish stocks.”. The relationships will be investigated by testing the relationship between SVI and realized volatility both for the sample at an aggregate level and for each stock individually.

5.2.2 Abnormal search volume index

We use abnormal SVI for the investigation (ASVI henceforth). ASVI is calculated to standardize SVI data. A standardization is necessary because the value of raw SVI is set relative to the period of which data was collected, e.g., raw SVI of week one in March 2018 depends on whether we collect data from 2017-2020 or 2018-2021. By using ASVI instead of raw SVI, we ensure that our search data is standardized in relation to its past history. Abnormal SVI is calculated utilizing the methodology used by (Da, et al., 2011), taking the natural logarithm of SVI during the week minus the natural logarithm of median SVI during a previous chosen number of weeks. For this study, we choose to standardize over the previous 52 weeks:

$$(5.1) ASVI_w = \log(SVI_w) - \log [Median(SVI_{w-1}, \dots, SVI_{w-52})]$$

5.2.3 Realized volatility

We use the adjusted daily closing price to calculate realized variance as an approximated historical volatility measure. Realized variance measures the daily standard deviation of log returns of the stock over a predefined period. (Andersen & Bollerslev, 1998) showed that latent variance can be approximated by summing the squared returns of a higher frequency than the time period, during which the variance is measured. This implies that weekly realized volatility of week w can be approximated by summing the squared daily returns for all t days in week w . Furthermore, as noted in (Andersen, et al., 2001) “*Realized volatilities and correlations show strong temporal dependence and appear to be well described by long-memory processes*”. This makes it well fitting for the purpose of correlating it with Google search volumes.

First, daily returns are calculated. Let $r_t = \ln(p_t) - \ln(p_{t-1})$, where p is the asset’s adjusted closing price on day t , denote the natural logarithm of the continuously compounded t day return. Then, weekly realized volatilities are calculated by taking the sum of the squared daily returns for each week. For this study, we assume five trading days per week. However, due to holiday, some weeks had less than five trading days. For all non-trading weekdays, we use the adjusted closing price from the previous trading day as a proxy. Based on the realized volatility literature, we construct our m -week volatility measure as follows:

$$(5.2) \quad RV_w = \sum_{t=1}^{n=5} r_{t,w}^2$$

where RV_w is the realized volatility for week w , n is the amount of trading days in week w , and $r_{t,w}$ denotes the natural logarithm of the daily return for t days in week w .

5.2.4 Student’s t-test

We use a student’s t-test to investigate whether SVI affects volatility. Specifically, we aim to test whether the corresponding ASVI values for weeks with high volatility significantly differ from those of weeks with low volatility. The t-test allows us to pairwise compare the mean value between groups in a population. If the mean value between the two groups tested differ

significantly, the phenomenon in focus affects the population. That is, for each stock, if the corresponding mean ASVI between two groups of realized weekly volatility vary significantly, where one group contains ASVI for weeks with high volatility and the other for weeks with low volatility, then search activity influences volatility.

One problem with pairing ASVI and volatility is the fact during a regular week, stocks are traded from Monday through Friday, while search volumes are collected from Monday through Sunday. In order to avoid possible errors as a result of this, we introduce a lag of one week in the ASVI value when pairing with realized volatility. That is, each pair will look as follows:

$$(5.3) \text{ pair}_{i,w} = (RV_{i,w}, ASVI_{i,w-1})$$

where $RV_{i,w}$ is the realized volatility of stock i over week w and $ASVI_{i,w-1}$ is the abnormal search volume index value for stock i over the previous week. What this essentially implies is that we hypothesize that the search volumes collected in week w is correlated with realized volatility of week $w+1$.

We use a two-tailed Student's t-test assuming unequal variances. This version of the t-test allows us to evaluate differences between two independent samples, assuming that the respective means are different. The two-tailed t-test is chosen because the correlation between the ASVI and realized volatility can be either positive or negative. Our null hypothesis assumes no difference in the means between the two samples of ASVI in the population. We define a t-distribution of the population for when the null hypothesis is true and choose a significance level α of 0.05 at which we try to reject the null hypothesis. The t-score is calculated using the following formula:

$$(5.4) t_{score} = \frac{\mu_{high} - \mu_{low}}{\left(\frac{S_{high}^2}{N_{high}} + \frac{S_{low}^2}{N_{low}} \right)^{1/2}}$$

where μ denotes the mean ASVI of the two samples, s denotes the standard deviation and N denotes the sample sizes. The results are distributed as student's t with df degrees of freedom, where df is calculated using Satterwaithe's formula:

$$(5.5) \quad df = \frac{\left(\frac{s_{high}^2}{N_{high}} + \frac{s_{low}^2}{N_{low}}\right)^2}{\frac{\left(\frac{s_{high}^2}{N_{high}}\right)^2}{N_{high} - 1} + \frac{\left(\frac{s_{low}^2}{N_{low}}\right)^2}{N_{low} - 1}}$$

where s denotes the standard deviation and N denotes the sample size (Stata, 2021).

We then plot our obtained t -scores on the x -axis of a graph and calculate the probability $P(t \geq t\text{-score})$ or $P(t \leq t\text{-score})$, which depends on whether the t -value is positive or negative. This value, which is commonly referred to as the p -value, represents an area below the t -distribution curve. If the p -value $\leq \alpha$, we can reject the null hypothesis within our chosen confidence level. Specifically, if the p -value ≤ 0.05 , then we reject the null hypothesis, which implies that there is a statistically significant link between SVI and volatility. This link can either be positive or negative, which means that increasing search volumes for stock i in week w either implies higher or lower volatility for the stock in week $w+1$.

5.3 Predictive regressions

The aim is to address the following research question: "Can the volatility of a given stock be predicted from Google searches for that stock? If so, how accurately?". To answer the research question, we construct an autoregressive model with weekly realized volatility as the dependent variable using only one-week lagged variables as possible predictors. We test this relationship both for the entire sample at an aggregate level and for each stock individually.

5.3.1 Regression variables

To perform the regressions, several control variables are calculated to be included in the model. Realized volatility, which is the dependent variable, and ASVI were calculated as specified in 7.2.3 and 7.2.2.

To include a control variable measuring trading activity, we calculate average trading volume. We use average as opposed to total volume to account for the fact that some weeks have less than five trading days. Additionally, we take the natural logarithm of the average trading volume to normalize the data. The formula can be specified as follows:

$$(5.6) \text{ATV}_w = \log \left(\frac{\sum_n TV_t}{n} \right)$$

Where ATV_w is the average trading volume for week w , TV_t is the trading volume on day t , and n denotes the number of days in week w .

As a last control variable, we calculate returns for each stock. First, daily returns are calculated by taking the natural logarithm of the ratio between the adjusted closing price on day t and $t-1$:

$$(5.7) rt_t = \log \left(\frac{\text{adjusted closing price}_t}{\text{adjusted closing price}_{t-1}} \right)$$

Then, weekly returns are calculated by taking the sum of rt for all t in week w . For this calculation, each week is assumed to have five trading days. As previously mentioned, certain weeks have less than five trading days due to holiday. For these weeks, we apply the same methodology as described in section 7.2.3. Thus, our m -week return measure is constructed as follows:

$$(5.8) RT_w = \sum_{t=1}^{n=5} rt_{t,w}$$

Where rt_w denotes the returns for week w , n denotes the amount of trading days in week w and $rt_{t,w}$ denotes the daily returns on day t in week w .

5.3.2 The autoregressive model

For the predictive analysis, we largely utilize the methodology used in (Kim, et al., 2019). We test the weekly lead-lag relationship between realized volatility and Google search volumes using a first order autoregressive model with multiple explanatory variables (AR(1) henceforth). The AR(1) equation includes a constant, the variables one week lagged values, one week lagged values of all other variables included in the model and an error term. A general AR(1) model can be written as:

$$(5.9) Y_t = c_i + \beta_1 Y_{t-1} + \beta_2 X1_{t-1} + \beta_2 X2_{t-1} + \dots + \beta_n Xn_{t-1} + \epsilon_t$$

Where Y_t is the variable investigated at time t , β 's denote the regression coefficients, Y_{t-1} is the lagged value of the variable regressed, X 's are all other variables included, c is a constant and ϵ is the error term.

We regress realized volatility at week w against realized volatility, ASVI, ATV, and RT. This will allow us to evaluate the impact of ASVI relative to the impact of the control variables. We run the AR(1) on the entire sample as an aggregate and for each stock with four years of weekly data with the following model:

$$(5.10) RV_{w,i} = c_i + \beta_1 RV_{w-1,i} + \beta_2 ASVI_{w-1,i} + \beta_3 ATV_{w-1,i} + \beta_4 RT_{w-1,i} + \epsilon_{w,i}$$

Where RV denotes realized volatility at week w for firm i , β 's denotes the regression coefficients for the previous weeks value of realized volatility, ASVI, ATV and RT.

6. Results

The software application used to conduct the statistical analysis and run the regressions is Stata, a general-purpose statistical software owned by StataCorp.

We aim to compare our results to those found in past research. However, past research aiming at specific sectors has primarily been focused on the link between SVI and stock returns. To include them in our analysis, we put forward an assumption that stock returns and volatility are inversely correlated [(Campbell & Hentschel, 1992), (Bekaert & Wu, 2000), (Wu, 2001) and (Bae, et al., 2007)]. Thus, we will compare our results to studies showing correlations between SVI and returns and inverting the relationships shown in them.

The stocks were classified according to the Global Industry Classification Standard (GICS henceforth) in order for us to better comment on our results. GICS was originally developed by MSCI and S&P Dow Jones Indices in 1999 with the purpose of efficiently capturing the breadth, depth, and development of industry sectors. Companies are classified at the sub-industry level in accordance with its principal business activity, which is determined based on revenues, earnings, and market perception. The classification consists of 11 sectors which are divided into 24 industry groups. The industry groups are further divided into 69 industries, which finally consists of 158 sub-industries (MSCI, 2021). Of the 11 GICS sectors, eight are represented in our sample: eight Health Care stocks, six Industrials stocks, six Financials stocks, two Consumer Staples stocks, one Consumer Discretionary stock, one Information Technology stock and one Utility stock. The classification of each stock can be found in table 8.1.

TICKER	GICS SECTOR
AMBU-B	Health Care
BAVA	Health Care
CARL-B	Cons. Staples
CHR	Materials
COLO-B	Health Care
DANSKE	Financials
DEMANT	Health Care
DSV	Industrials
FLS	Industrials
GMAB	Health Care
GN	Health Care
ISS	Industrials
JYSK	Financials
LUN	Health Care
MAERSK-A	Industrials
MAERSK-B	Industrials
NDA-DK	Financials
NOVO-B	Health Care
NZYM-B	Materials
PNDORA	Cons. Discretionary
ROCK-B	Industrials
RBREW	Cons. Staples
SIM	IT
SYDB	Financials
TOP	Financials
TRYG	Financials
VWS	Utilities

Table 6.1: Sector of each stock as per the Global Industry Classification Standard.

6.1 Descriptive analysis

6.1.1 Obtaining t-test scores

Having downloaded SVI and historical price data as described in section 5, we first pre-processed the price data, then calculated realized volatilities and ASVI as described in section 5.2.2-5.2.3. Then, for each stock, ASVI was divided into two subsets: one containing ASVI for weeks with low volatility and one for weeks with high volatility. Specifically, ‘Low’ contained all ASVI values paired with a volatility strictly less than the median volatility and ‘High’ contained all ASVI values paired with a volatility greater than or equal to the median. These subsets served as two independent groups in the population. We then calculated the t-scores and the degrees of freedom for each stock in accordance with equations 5.4 and 5.5 through Stata (Stata, 2021). The corresponding p-values were also obtained through Stata. Based on the p-values and t-scores, we either rejected or retained the null hypothesis. The process was performed on the stocks, both at

an aggregate level as well as individually for each of the 27 stocks with a significance level of 0.05.

6.1.2 Results

The significant results of the t-test are summarized in table 6.2. On the aggregate level, the results failed to show significant descriptive power of ASVI on volatility for our sample of Danish stocks. These results differ from the majority of past research. For example, both (Kim, et al., 2019) and (Bui & Nguyen, 2019) conclude that an explanatory link between the current weeks SVI and stock price volatility in the subsequent week exists. Additionally, (Fink & Johann, 2014) and (Aouadi, et al., 2013) both find a significant descriptive link between SVI and volatility.

However, the results showed statistically significant correlation between ASVI and realized volatility for some of the stocks when tested individually. This implies, as according to our hypothesis, that search activity does affect volatility. Specifically, of the 27 stocks tested, three showed strong correlations with significance at the 99% confidence level. An additional three showed significant correlations at the 95% confidence interval. Of the six stocks with significant correlation, three showed a positive correlation, while three showed a negative correlation. A positive correlation implies that an increase in search volumes in week w increases volatility in week $w+1$, while a negative correlation implies that an increase in search volumes in week w decreases volatility in week $w+1$. Stocks representing four of the GICS sectors showed significant results. Each of those will be commented in detail below.

Health Care

Of the eight Health Care stocks tested, the experimental findings showed that Coloplast and Genmab had significant correlation at the 95% confidence level. The p-values range from 0.0097 for Coloplast to 0.0111 for Genmab, providing us with sufficient evidence for rejecting the null hypothesis. While Coloplast had a positive t-score, Genmab somewhat surprisingly had a negative t-score, indicating a negative correlation: increasing ASVI implies decreasing volatility. There was no sufficient evidence to show significant correlation between ASVI and volatility for: Bavarian Nordic, Demant, GN Store Nord, H. Lundbeck and Novo Nordisk. From our results, only the negative correlation found for Genmab falls in line with evidence from past research. (Smales,

2020) found a positive correlation between the SVI and excess return of Health Care stocks in their sample. A similar positive link was found in (Stejskalová, 2019), which would imply a negative correlation between the SVI and volatility, as according to asymmetric volatility literature.

Industrials

For the six Industrials stocks, statistical testing revealed that search volumes for DSV yielded significant correlation with volatility at the 95% confidence level. For ISS, the findings showed significant correlation at the 90% confidence level, though this will not be included in further commenting. For DSV, strong statistical evidence supports rejecting the null hypothesis. The t-score was positive, which implies a positive relationship between ASVI and volatility. There was no sufficient evidence to reject the null hypothesis for Maersk-A, Maersk-B, Rockwool, and FLSmidth & Co. To our knowledge, no past research has yet found any statistically significant link between searches for company names/tickers and stock performance for stocks representing the Industrials sector.

Financials

Based on experimental findings, we found that search volume significantly correlated with volatility for one out of six Financials stocks at the 95% confidence level: Tryg. We found strong statistical evidence for rejecting the null hypothesis for Tryg, with a p-value of 0.0233. The t-score indicated a negative correlation. This differs from the findings in (Smales, 2020) who found the SVI to have a negative correlation with returns of US Financials sector stocks. Assuming the aforementioned negative link between returns and volatility, the study pointed towards a positive link between SVI and volatility. There was no sufficient evidence to reject the null hypothesis for Danske Bank, Jyske Bank, Nordea, Sydbank, and Topdanmark. However, (Di, et al., 2021) explores whether investor sentiment, measured by SVI, has an impact on the stock performance of Islamic and conventional banks. They find SVI to significantly affect returns of both conventional and Islamic banks. This implies, as per the established body of literature, that volatility of bank stocks is affected by SVI.

Consumer Staples

An interesting result of the statistical testing was the fact that both Consumer Staples stocks yielded significant results. Consumer Staples includes stocks producing essential products used by consumers. The category includes stocks of companies producing foods and beverages, alcohol and tobacco. We found strong significance for a positive correlation between ASVI and volatility for Carlsberg, while we found a statistically significant negative correlation for Royal Unibrew. Evidence from past research falls in line with our mixed results. Our findings for Royal Unibrew find support in (Smales, 2020), who found a positive link between search activity and excess returns for Consumer Staples stocks, which per the asymmetric volatility literature implies negative correlation between SVI and volatility. Contrary to this, the findings of (Stejskalová, 2019), who found a negative link between searches for prices of Consumer Staples stocks and returns, support the results for Carlsberg.

<i>Ticker</i>	<i>t-score</i>	<i>p-value</i>	<i>df</i>	<i>GICS sector</i>
<i>CARL-B</i>	4.0914	0.0001***	206.665	Cons. Staples
<i>COLO-B</i>	2.6106	0.0097***	202.296	Health Care
<i>DSV</i>	4.3842	0.0000***	199.137	Industrials
<i>GMAB</i>	-2.5636	0.0111**	205.157	Health Care
<i>ISS</i>	1.8718	0.0627*	199.372	Industrials
<i>RBREW</i>	-2.0714	0.0396**	199.611	Cons. Staples
<i>TRYG</i>	-2.2856	0.0233**	203.642	Financials
<i>VWS</i>	-1.8038	0.0728*	200.596	Utilities

Table 6.2: Significant results of the *t*-test. The symbols *, **, *** denotes significance at the 10%, 5% and 1% levels, respectively.

6.2 Predictive analysis

To assess the significance of Google searches for predicting volatility, we perform predictive regressions in accordance with the AR(1) model defined in section 5.3.2. Following the preprocessing as described in section 5.3.1 we ran the regressions, first at an aggregate level to identify a general trend, and then individually for each stock.

A common problem with a regressive model such as ours is that it can be highly sensitive to outliers. To avoid misleading results, we use the logged versions of all variables when running the regressions. Additionally, we perform regressions with heteroskedasticity-consistent standard errors on our sample following the methodology of Huber and White (Stata, 2021). The more conservative standard errors are introduced based on the assumption that stock returns generally

are heteroskedastic (Schwert & Seguin, 1990). Thus, our results are presented with heteroskedastic robust standard errors.

6.2.1 Results

The significant results are summarized in table 6.3. On the aggregate level, the results failed to show any predictive power of ASVI on volatility for our sample. These results fall in line with evidence found in (Osarumwense, 2021), where ASVI failed to predict future volatility in the Nigerian stock market. On the contrary (Kim, et al., 2019) successfully found SVI to be a significant explanatory of volatility of Norwegian stocks in the following week. The same link is found in (Vlastakis & Markellos, 2012), where demand for information at the market level measured by SVI was found to be positively correlated with historical and implied measures of volatility.

Furthermore, the results mostly failed to find any predictive power on the stocks when tested individually. However, we did find evidence of ASVI as a significant predictor of volatility for some stocks. Specifically, Rockwool showed ASVI as a significant predictor at the 95% confidence level as the only stock. Additionally, we found weak evidence of ASVI as a predictor for Chr. Hansen, Jyske Bank and Tryg at the 90% confidence level. For Chr. Hansen and Tryg, increased search activity was associated with increasing volatility in the subsequent week. For Jyske Bank and Rockwool, increased search activity was associated with decreasing volatility in the subsequent week. Due to the overall lack of significant evidence, comments will not be made on a sector level.

VARIABLES	(5) CHR	(14) JYSK	(23) ROCK-B	(27) TRYG
VOLAPAST	0.82345*** (0.11281)	0.86226*** (0.10963)	0.77435*** (0.09151)	0.87831*** (0.08337)
ASVI	0.00041* (0.00022)	-0.00049* (0.00029)	-0.00112** (0.00048)	-0.00024* (0.00014)
ATV	0.00003 (0.00010)	0.00001 (0.00011)	-0.00007 (0.00012)	0.00005 (0.00005)
RT	-0.00064 (0.00090)	-0.00003 (0.00077)	-0.00148 (0.00112)	0.00004 (0.00060)
Constant	-0.00021 (0.00126)	0.00006 (0.00135)	0.00107 (0.00128)	-0.00062 (0.00062)
Observations	209	209	209	209
R-squared	0.66239	0.74389	0.62307	0.78688

*Table 6.3: Significant results of the multivariate regressions. Robust standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$*

7. Conclusion

In this paper, we introduce a novel approach to creating search queries when investigating the link between Google search volume and stock performance, as we use a combination of ticker symbols and company names to create the search queries. This approach, which relies on the vast evidence found in past research, is used to optimize the accuracy of using Google search data as an indicator of information demand [(Da, et al., 2011), (Vlastakis & Markellos, 2012), (Takeda & Wakao, 2014) and (Bilj, et al., 2016)].

The aim of the study is to explore whether the dynamics of stock volatility in the Danish stock market can be explained from online search activity. Specifically, we investigate whether SVI can explain and predict volatility. In a sample of the largest and most actively traded stocks listed on OMX Nasdaq Copenhagen, we fail to find significant evidence for neither the descriptive nor the predictive power of search volumes on volatility of Danish stocks. However, when testing the stocks on an individual basis, we do find significant evidence of both an existing descriptive and predictive relationship for a number of stocks. For the descriptive investigation, such relationship is most notably found for all stocks classified within the consumer staples sector which falls in line with evidence from past research [(Stejskalová, 2019) and (Smales, 2020)]. However, we do not manage to state whether such relationship is positive or negative, as we find significant evidence pointing towards both a positive and a negative descriptive link among the individual stocks tested. On the contrary, we fail to identify any sector specific trends when performing the predictive

analysis. To our knowledge, this is yet to be tested in other research, leaving room for further investigation.

7.1 Further steps

A shortcoming of this investigation is the fact that we did not test each sector on an aggregate level to investigate whether more general descriptive and predictive links between SVI and GICS sectors exist. Recent research has managed to establish such links at a descriptive level, though this, to the best of our knowledge, is yet to be tested on a broader CIGS level for a smaller market such as the one we are exploring [(Stejskalová, 2019), (Smales, 2020)] . This leaves room for future investigation, which should preferably be conducted on a larger sample of stocks than the one used in this study. Furthermore, we only analyzed search queries on the mix of tickers and firm names as opposed to also checking for search queries of each of the two separately. By doing this, we would have ensured that the lack of significant results obtained in this study was not a result of the construction of the search queries. Therefore, a similar investigation with a different approach to the construction of search queries might be a cause for future investigation.

Additionally, it is worth noting that we conduct our analysis over one given period. A future source of research lies within the need of conducting a similar investigation for different time periods to better explain whether and how the link between search activity and volatility has changed over time. An interesting implication could be to conduct the experiments for samples covering periods pre- and post-Covid 19, to see how the increasing appearance of retail investors following the pandemic has contributed to this relationship (Glenwood, 2021).

Lastly, an intriguing area of future research lies in the approximation of online search activity through other data sources than Google SVI. The established body of research has primarily used SVI for this approximation due to its availability and the strong market position of Google as a search engine. Though some researchers have investigated such approximation by the use of Wikipedia searches, e.g. (Osarumwense, 2021), and more recently Twitter activity [e.g. (Ranco, et al., 2015), (Bartov, et al., 2018) and (Rakowski, et al., 2021)] future researchers might seek out search data through different platforms such as Reddit.

Bibliography

Journal Articles

Andersen, T. G. & Bollerslev, T. (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts. *International Economic Review*, 39(4), pp. 885-905.

Andersen, T. G., Bollerslev, T., Diebold, F. X. & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1), pp. 43-76.

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

Aziz, T. & Ansari, V. A. (2021). How Does Google Search Affect the Stock Market? Evidence from Indian Companies. *Vision: The Journal of Business Perspective*. 25(2), pp. 224-232

Bae, J., Kim, C.-J. & Nelson, C. R. (2007). Why are stock returns and volatility negatively correlated?. *Journal of Empirical Finance*, 14(1), pp. 41-58.

Bank, M., Larch, M. & Peter, G. (2011). Google Search Volume and its Influence on Liquidity and Returns of German Stocks. *Financial Markets and Portfolio Management*, 25(3), pp. 239-264.

Barber, B. M. & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *Review of Financial Studies*, 21(2), pp. 785-818.

Bartov, E., Faurel, L. & Mohanram, P. S. (2018). Can Twitter Help Predict Firm-Level Earnings and Stock Returns?. *THE ACCOUNTING REVIEW*, 93(3), pp. 25-57.

Bekaert, G. & Wu, G. (2000). Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, Volume 13(1), pp. 1-42.

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

Bui, V. X. & Nguyen, H. T. (2019). Stock market activity and Google Trends: the case of a developing economy. *Journal of Economics and Development*, 21(2), pp. 191-212.

Campbell, J. Y. & Hentschel, L. (1992). No news is good news: an asymmetric model of changing volatility in stock returns. *Journal of Financial Economics*, 31(3), pp. 281-318.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1), pp. 135-155.

Darolles, S., Le Fol, G. & Mero, G. (2017). Mixture of distribution hypothesis: Analyzing daily liquidity frictions and information flows. *Journal of Econometrics*, 201(2), pp. 367-383.

Da, Z., Engleberg, J. & Gao, P. (2011). In Search of Attention. *The Journal of Finance*, 66(5), pp. 1461-1499.

Di, L., Shaiban, M. S. & Hasanov, A. S. (2021). The power of investor sentiment in explaining bank stock performance: Listed conventional vs. Islamic banks. *Pacific-Basin Finance Journal*, Volume 66. Available online: <https://www.sciencedirect.com/science/article/pii/S0927538X21000160>, [Accessed 15 May 2021]

Epps, T. W. & Epps, M. L., (1976). The Stochastic Dependence of Security Price Changes and Transaction Volumes: Implications for the Mixture-of-Distributions Hypothesis. *Econometrica*, 44(2), pp. 305-321.

Gwilym, O. a., Hassan, I., Wang, Q. & Xie, R. (2016). In Search of Concepts: The Effects of Speculative Demand on Stock Returns. *European Financial Management*, 22(3), pp. 427-449.

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), pp. 1116-1127.

Kim, N., Lučivjanská, K., Molnár, P. & Villa, R. (2019). Google searches and stock market activity: Evidence from Norway. *Finance Research Letters*, Volume 28, pp. 208-220.

Mondria, J., Wu, T. & Zhang, Y. (2010). The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics*, 82(1), pp. 85-95.

Osarumwense, O.-I. (2021). Investors' attention: does it impact the Nigerian stock market activities?. *Journal of Economics and Development*, 3(1). Available online: <https://www.emerald.com/insight/content/doi/10.1108/JED-02-2020-0015/full/html>, [Accessed 15 May 2021].

Rakowski, D., Shirley, S. E. & Stark, J. R. (2021). Twitter activity, investor attention, and the diffusion of information. *Financial Management*, 50(1), pp. 3-46.

Ranco, G. et al. (2015). The Effects of Twitter Sentiment on Stock Price Returns. *PLoS ONE*, 10(9). Available online: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0138441>, [Accessed 17 May, 2021]

Schwert, G. W. & Seguin, P. J. (1990). Heteroskedasticity in Stock Returns. *The Journal of Finance*, 45(4), pp. 1129-1155.

Smales, L. A. (2020). Investor attention and the response of US stock market sectors to the COVID-19 crisis. *Review of Behavioral Finance*, 13(1), pp. 20-39.

Stejskalová, J. (2019). Behavioural Attention to Financial Indicators: Evidence from Google Trends Data. *Finance a Uver: Czech Journal of Economics & Finance*, 69(5), pp. 440-462.

Swamy, V. & Dharani, M. (2019). Investor attention using the Google search volume index – impact on stock returns. *Review of Behavioral Finance*, 11(1). Available online: <https://www.emerald.com/insight/content/doi/10.1108/RBF-04-2018-0033/full/html>, [Accessed on 17 May, 2021]

Swamy, V., Dharani, M. & Takeda, F. (2019). Investor attention and Google Search Volume Index: Evidence from an emerging market using quantile regression analysis. *Research in International Business and Finance*, Volume 50, pp. 1-17.

Takeda, F. & Wakao, T. (2014). Google search intensity and its relationship with returns and trading volume of Japanese stocks. *Pacific-Basin Finance Journal*, Volume 27, pp. 1-18.

Tan , S. D. & Taş, O. (2019). Investor attention and stock returns: Evidence from Borsa Istanbul. *Borsa Istanbul Review*, 19(2), pp. 106-116.

Tauchen, G. E. & Pitts, M. (1983). The Price Variability-Volume Relationship on Speculative Markets. *Econometrica*, 51(2), pp. 485-505.

Vlastakis, N. & Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking & Finance*, 36(6), pp. 1808-1821.

Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking & Finance*, Volume 41, pp. 17-35.

Wu, G. (2001). The determinant of asymmetric volatility. *Review of Financial Studies*, Volume 14, pp. 837-859.

Working papers

Fink, C. & Johann, T. (2014). May I Have Your Attention, Please: The Market Microstructure of Investor Attention, working paper, Universität Mannheim, September 17, 2014.

Byström, H. & Krygier, D. (2018). What Drives Bitcoin Volatility?, working paper, no 2018:3, Lund University, 2018.

Web pages

Google Trends. (2021). Available online: <https://trends.google.com/trends/?geo=US>, [Accessed 17 April 2021].

Google Trends. (2021). Trends Help. Available online: https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052, [Accessed 16 April 2021].

Johnson, J. (2021). Statista. Worldwide desktop market share of leading search engines from January 2010 to February 2021. Available online: <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>, [Accessed 16 May 2021].

MSCI. (2021). The Global Industry Classification Standard. Available online: <https://www.msci.com/gics>, [Accessed 24 April 2021].

Nasdaq. (2020). Semi-annual review of the OMX Copenhagen C25 Index. Available online: <https://www.nasdaq.com/press-release/semi-annual-review-of-the-omx-copenhagen-c25-index-2020-06-10>, [Accessed 17 04 2021].

Stata. (2021). ttest. Available online: <https://www.stata.com/manuals/rttest.pdf>, [Accessed 21 April 2021].

Stata. (2021). rregress. Available at: <https://www.stata.com/manuals/rregress.pdf>, [Accessed 16 May 2021].

Yahoo! Finance. (2021). Historical data. Available online: <https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC>, [Accessed 19 April 2021].

News articles

Forsey, Caroline. (2019). The Top 7 Search Engines, Ranked by Popularity. HubSpot. Available online: <https://blog.hubspot.com/marketing/top-search-engines>, [Accessed 16 May 2021].

Glenwood, M. (2021). Everyone's a Day Trader Now. The Wall Street Journal. 25 July 2020, Available online: <https://www.wsj.com/articles/everyones-a-day-trader-now-11595649609>, [Accessed 17 May 2021].

Winding, Mikkel Holmgaard. (2021). Danskerne handler danske aktier som aldrig før. 20 February 2021, Available online: <https://finans.dk/privatokonomi/ECE12772369/danskerne-handler-danske-aktier-som-aldrig-foer/?ctxref=ext>, [Accessed 20 May 2021].

Appendix

Appendix A – result tables

Table 6.4: Results of the t-test. The symbols *, **, *** denotes significance at the 10%, 5% and 1% levels, respectively.

Ticker	t-score	p-value	df	GICS sector
<i>Aggregate</i>	-0.4135	0.6793	5627.22	N/A
<i>AMBU-B</i>	0.9919	0.3224	202.85	Health Care
<i>BAVA</i>	-1.4498	0.1487	192.963	Health Care
<i>CARL-B</i>	4.0914	0.0001***	206.665	Cons. Staples
<i>CHR</i>	0.7832	0.4346	173.472	Materials
<i>COLO-B</i>	2.6106	0.0097***	202.296	Health Care
<i>DANSKE</i>	0.0238	0.9811	203.375	Financials
<i>DEMANT</i>	1.3389	0.1822	192.305	Health Care
<i>DSV</i>	4.3842	0.0000***	199.137	Industrials
<i>FLS</i>	1.0552	0.2926	195.173	Industrials
<i>GMAB</i>	-2.5636	0.0111**	205.157	Health Care
<i>GN</i>	-0.2972	0.7666	206.998	Health Care
<i>ISS</i>	1.8718	0.0627*	199.372	Industrials
<i>JYSK</i>	0.9198	0.3588	205.283	Financials
<i>LUN</i>	1.0944	0.2750	206.879	Health Care
<i>MAERSK-A</i>	0.9797	0.3285	189.899	Industrials
<i>MAERSK-B</i>	0.8293	0.4079	200.534	Industrials
<i>NDA-DK</i>	1.4006	0.1629	200.767	Financials
<i>NOVO-B</i>	0.3867	0.6993	206.427	Health Care
<i>NZYM-B</i>	0.0286	0.9772	194.337	Materials
<i>PNDORA</i>	-0.3765	0.7070	199.277	Cons. Discretionary
<i>ROCK-B</i>	-0.5299	0.5968	206	Industrials
<i>RBREW</i>	-2.0714	0.0396**	199.611	Cons. Staples
<i>SIM</i>	-0.1739	0.8621	198.842	IT
<i>SYDB</i>	-1.5098	0.1326	204.303	Financials
<i>TOP</i>	-0.0757	0.9397	206.881	Financials
<i>TRYG</i>	-2.2856	0.0233**	203.642	Financials
<i>VWS</i>	-1.8038	0.0728*	200.596	Utilities

Table 6.5.1-6.5.5: Result of the AR(1) regressions. Robust standard errors are marked in parantheses. The symbols

*, **, *** denotes significance at the 10%, 5% and 1% levels, respectively.

VARIABLES	(1) AGG	(2) AMBU-B	(3) BAVA	(4) CARL-B	(5) CHR	(6) COLO-B
VOLAPAST	0.80890*** (0.04258)	0.82231*** (0.08196)	0.78609*** (0.05850)	0.83497*** (0.10268)	0.82345*** (0.11281)	0.85816*** (0.07687)
ASVI	-0.00004 (0.00003)	-0.00008 (0.00038)	-0.00007 (0.00049)	0.00008 (0.00019)	0.00041* (0.00022)	0.00002 (0.00016)
ATV	-0.00002 (0.00002)	-0.00010 (0.00010)	-0.00010 (0.00015)	0.00001 (0.00009)	0.00003 (0.00010)	-0.00010 (0.00007)
RT	0.00022 (0.00029)	-0.00126 (0.00084)	0.00126 (0.00100)	0.00046 (0.00041)	-0.00064 (0.00090)	-0.00128* (0.00071)
Constant	0.00048** (0.00023)	0.00171 (0.00138)	0.00179 (0.00186)	0.00001 (0.00114)	-0.00021 (0.00126)	0.00134 (0.00092)
Observations	5,643	209	209	209	209	209
R-squared	0.65643	0.68307	0.62287	0.69931	0.66239	0.74872
VARIABLES	(7) DANSKE	(8) DEMANT	(9) DSV	(10) FLS	(11) GMAB	(12) GN
VOLAPAST	0.87495*** (0.13660)	0.77047*** (0.10053)	0.84305*** (0.10191)	0.77416*** (0.13461)	0.82969*** (0.09167)	0.81305*** (0.12594)
ASVI	-0.00013 (0.00049)	-0.00017 (0.00014)	-0.00037 (0.00038)	-0.00007 (0.00049)	-0.00018 (0.00021)	-0.00004 (0.00004)
ATV	-0.00006 (0.00008)	0.00007 (0.00010)	-0.00011 (0.00011)	0.00035 (0.00021)	0.00024 (0.00031)	0.00020 (0.00013)
RT	0.00031 (0.00045)	-0.00061 (0.00081)	-0.00084 (0.00078)	-0.00034 (0.00154)	0.00318** (0.00138)	-0.00089 (0.00069)
Constant	0.00094 (0.00116)	-0.00072 (0.00125)	0.00157 (0.00151)	-0.00406 (0.00260)	-0.00265 (0.00371)	-0.00249 (0.00168)
Observations	209	209	209	209	209	209
R-squared	0.77635	0.59365	0.70087	0.60704	0.69421	0.67584
VARIABLES	(13) ISS	(14) JYSK	(15) LUN	(16) MAERSK-A	(17) MAERSK-B	(18) NDADK
VOLAPAST	0.86437*** (0.10822)	0.86226*** (0.10963)	0.74633*** (0.20097)	0.86203*** (0.12761)	0.84488*** (0.11761)	0.88803*** (0.10335)
ASVI	0.00003 (0.00005)	-0.00049* (0.00029)	-0.00001 (0.00026)	-0.00208 (0.00126)	-0.00160 (0.00137)	-0.00031 (0.00057)
ATV	-0.00004 (0.00004)	0.00001 (0.00011)	0.00003 (0.00020)	-0.00028 (0.00025)	-0.00028 (0.00026)	0.00006 (0.00007)
RT	0.00022 (0.00034)	-0.00003 (0.00077)	0.00492 (0.00369)	-0.00074 (0.00146)	-0.00022 (0.00138)	-0.00071 (0.00073)
Constant	0.00058 (0.00052)	0.00006 (0.00135)	0.00007 (0.00250)	0.00278 (0.00216)	0.00318 (0.00270)	-0.00065 (0.00088)
Observations	209	209	209	209	209	209
R-squared	0.75139	0.74389	0.55780	0.69421	0.68890	0.79148

VARIABLES	(19) NOVO-B	(20) NZYM-B	(21) PNDORA	(22) RBREW	(23) ROCK-B	(24) SIM
VOLAPAST	0.73960*** (0.17450)	0.76349*** (0.14273)	0.77604*** (0.08899)	0.76756*** (0.14290)	0.77435*** (0.09151)	0.83158*** (0.08255)
ASVI	0.00109 (0.00111)	0.00003 (0.00026)	0.00001 (0.00014)	0.00001 (0.00006)	-0.00112** (0.00048)	-0.00001 (0.00011)
ATV	-0.00067 (0.00051)	0.00010 (0.00027)	-0.00001 (0.00017)	-0.00012 (0.00010)	-0.00007 (0.00012)	-0.00002 (0.00012)
RT	0.00414 (0.00421)	0.00060 (0.00351)	-0.00029 (0.00099)	0.00003 (0.00127)	-0.00148 (0.00112)	-0.00092 (0.00067)
Constant	0.01035 (0.00752)	-0.00102 (0.00350)	0.00049 (0.00224)	0.00165 (0.00122)	0.00107 (0.00128)	0.00037 (0.00135)
Observations	209	209	209	209	209	209
R-squared	0.56010	0.58375	0.60049	0.58941	0.62307	0.69860

VARIABLES	(25) SYDB	(26) TOP	(27) TRYG	(28) VWS
VOLAPAST	0.86775*** (0.11579)	0.88927*** (0.11046)	0.87831*** (0.08337)	0.83700*** (0.10994)
ASVI	-0.00002 (0.00030)	-0.00002 (0.00009)	-0.00024* (0.00014)	-0.00052 (0.00036)
ATV	0.00011 (0.00007)	0.00006 (0.00006)	0.00005 (0.00005)	-0.00011 (0.00022)
RT	-0.00029 (0.00075)	0.00017 (0.00039)	0.00004 (0.00060)	-0.00052 (0.00117)
Constant	-0.00118 (0.00089)	-0.00068 (0.00067)	-0.00062 (0.00062)	0.00172 (0.00302)
Observations	209	209	209	209
R-squared	0.75186	0.78405	0.78688	0.71334

Appendix B – Stata do files

T-test:

```
//T-test command for aggregate test
```

```
clear all
```

```
import excel "/Users/olivernystrom/Desktop/Bachelor project - NEKH02/T-test2.xlsx",  
sheet("Sheet6") cellrange(A1:B2836) firstrow
```

```
tttest High == Low, unpaired unequal
```

```
// T-test commands for individual tests
```

```
clear all
```

```
import excel "/Users/olivernystrom/Desktop/Bachelor project - NEKH02/T-test2.xlsx",
sheet("Sheet4") cellrange(A3:BB109) firstrow
ttest AMBUB_high == AMBUB_low, unpaired unequal
ttest MAERSKA_high == MAERSKA_low, unpaired unequal
ttest MAERSKB_high == MAERSKB_low, unpaired unequal
ttest BAVA_high == BAVA_low, unpaired unequal
ttest CARLB_high == CARLB_low, unpaired unequal
ttest CHR_high == CHR_low, unpaired unequal
ttest COLOB_high == COLOB_low, unpaired unequal
ttest DANSKE_high == DANSKE_low, unpaired unequal
ttest DEMANT_high == DEMANT_low, unpaired unequal
ttest DSV_high == DSV_low, unpaired unequal
ttest FLS_high == FLS_low, unpaired unequal
ttest GMAB_high == GMAB_low, unpaired unequal
ttest GN_high == GN_low, unpaired unequal
ttest ISS_high == ISS_low, unpaired unequal
ttest LUN_high == LUN_low, unpaired unequal
ttest NOVOB_high == NOVOB_low, unpaired unequal
ttest NZYMB_high == NZYMB_low, unpaired unequal
ttest PNDORA_high == PNDORA_low, unpaired unequal
ttest ROCKB_high == ROCKB_low, unpaired unequal
ttest RBREW_high == RBREW_low, unpaired unequal
ttest SIM_high == SIM_low, unpaired unequal
ttest TRYG_high == TRYG_low, unpaired unequal
ttest VWS_high == VWS_low, unpaired unequal
ttest JYSKE_high == JYSKE_low, unpaired unequal
ttest NDADK_high == NDADK_low, unpaired unequal
ttest SYDB_high == SYDB_low, unpaired unequal
ttest TOP_high == TOP_low, unpaired unequal
```

Regressions:

```
// Install outreg2
```

```
clear all
```

```
ssc install outreg2
```

```
// Regression on aggregate level
```

```
clear all
```

```
import excel "/Users/olivernystrom/Desktop/Bachelor project - NEKH02/Autoregvals.xlsx",  
sheet("regagg") firstrow
```

```
reg VOLA VOLAPAST ASVI ATV RT, robust
```

```
outreg2 using AR10.doc, dec(5) replace ctitle(AGG)
```

```
// Regressions of individual stocks and creation of output table
```

```
clear all
```

```
import excel "/Users/olivernystrom/Desktop/Bachelor project - NEKH02/Autoregvals.xlsx",  
sheet("IND_1") firstrow
```

```
reg VOLA_AMBUB VOLAPAST ASVI ATV RT, robust
```

```
outreg2 using AR10.doc, dec(5) append ctitle(AMBU-B)
```

```
reg VOLA_BAVA VOLAPAST ASVI ATV RT, robust
```

```
outreg2 using AR10.doc, dec(5) append ctitle(BAVA)
```

```
reg VOLA_CARLB VOLAPAST ASVI ATV RT, robust
```

```
outreg2 using AR10.doc, dec(5) append ctitle(CARL-B)
```

```
reg VOLA_CHR VOLAPAST ASVI ATV RT, robust
```

```
outreg2 using AR10.doc, dec(5) append ctitle(CHR)
```

```
reg VOLA_COLOB VOLAPAST ASVI ATV RT, robust
```

```
outreg2 using AR10.doc, dec(5) append ctitle(COLO-B)
```

```
reg VOLA_DANSKE VOLAPAST ASVI ATV RT, robust
```

```
outreg2 using AR11.doc, dec(5) replace ctitle(DANSKE)
```

```
reg VOLA_DEMANT VOLAPAST ASVI ATV RT, robust
```

```
outreg2 using AR11.doc, dec(5) append ctitle(DEMANT)
```

```
reg VOLA_DSV VOLAPAST ASVI ATV RT, robust
```

outreg2 using AR11.doc, dec(5) append ctitle(DSV)
reg VOLA_FLS VOLAPAST ASVI ATV RT, robust
outreg2 using AR11.doc, dec(5) append ctitle(FLS)
reg VOLA_GMAB VOLAPAST ASVI ATV RT, robust
outreg2 using AR11.doc, dec(5) append ctitle(GMAB)
reg VOLA_GN VOLAPAST ASVI ATV RT, robust
outreg2 using AR11.doc, dec(5) append ctitle(GN)
reg VOLA_ISS VOLAPAST ASVI ATV RT, robust
outreg2 using AR12.doc, dec(5) replace ctitle(ISS)
reg VOLA_JYSK VOLAPAST ASVI ATV RT, robust
outreg2 using AR12.doc, dec(5) append ctitle(JYSK)
reg VOLA_LUN VOLAPAST ASVI ATV RT, robust
outreg2 using AR12.doc, dec(5) append ctitle(LUN)
reg VOLA_MAERSKA VOLAPAST ASVI ATV RT, robust
outreg2 using AR12.doc, dec(5) append ctitle(MAERSK-A)
reg VOLA_MAERSKB VOLAPAST ASVI ATV RT, robust
outreg2 using AR12.doc, dec(5) append ctitle(MAERSK-B)
reg VOLA_NDADK VOLAPAST ASVI ATV RT, robust
outreg2 using AR12.doc, dec(5) append ctitle(NDADK)
reg VOLA_NOVOB VOLAPAST ASVI ATV RT, robust
outreg2 using AR13.doc, dec(5) replace ctitle(NOVO-B)
reg VOLA_NZYMB VOLAPAST ASVI ATV RT, robust
outreg2 using AR13.doc, dec(5) append ctitle(NZYM-B)
reg VOLA_PNDORA VOLAPAST ASVI ATV RT, robust
outreg2 using AR13.doc, dec(5) append ctitle(PNDORA)
reg VOLA_RBREW VOLAPAST ASVI ATV RT, robust
outreg2 using AR13.doc, dec(5) append ctitle(RBREW)
reg VOLA_ROCKB VOLAPAST ASVI ATV RT, robust
outreg2 using AR13.doc, dec(5) append ctitle(ROCK-B)
reg VOLA_SIM VOLAPAST ASVI ATV RT, robust
outreg2 using AR13.doc, dec(5) append ctitle(SIM)

```
reg VOLA_SYDB VOLAPAST ASVI ATV RT, robust
outreg2 using AR14.doc, dec(5) replace ctitle(SYDB)
reg VOLA_TOP VOLAPAST ASVI ATV RT, robust
outreg2 using AR14.doc, dec(5) append ctitle(TOP)
reg VOLA_TRYG VOLAPAST ASVI ATV RT, robust
outreg2 using AR14.doc, dec(5) append ctitle(TRYG)
reg VOLA_VWS VOLAPAST ASVI ATV RT, robust
outreg2 using AR14.doc, dec(5) append ctitle(VWS)
```