Helene Wilhelmsen

Google Search Volume as an Attention Proxy in the Stock Market

A study of Nasdaq Copenhagen

Master's thesis spring 2020

Oslo Business School

Oslo Metropolitan University

Master's program in business administration

Acknowledgements

I would like to thank my supervisor Jacopo Bizzotto for guidance and constructive feedback. I will also like to thank Google for providing insights into user and search statistics that have been necessary for the execution of this master's thesis. At last, I want to thank my family and friends for supporting me throughout the writing of my thesis and my education. I want to dedicate this master's thesis to my late father, who sowed my interest in economics very early in life. He is undoubtedly the reason why I have developed such a great interest in economics.

Helene Wilhelmsen

Haugesund, April 21st, 2020

Abstract

This paper investigates whether Google search queries data correlates with or predicts stock market parameters on the Nasdaq Copenhagen (Copenhagen Stock Exchange), and whether excess return can be generated utilizing trading strategies based on Google search volume. I use abnormal returns, abnormal trading volume, and volatility as measures of market activity. Daily and weekly Google search data are collected. Google search data is obtained based on company ticker and company name. The analysis utilizes panel data regression to investigate if abnormal Google Search volume can describe or predict market parameters. The results find a small positive correlation between Google searches and abnormal stock returns. However, the findings from the panel data regressions find no evidence that abnormal Google search volume correlates with or predicts abnormal returns at a statistically significant level. However, the regression results show that both weekly and daily abnormal Google search volume correlates with abnormal trading volume and volatility. Trading strategies involving Google search volume show that there is economic value in utilizing abnormal Google search volume as a parameter in purchase decisions of financial products when transaction costs are not considered. The thesis also discusses the complex dynamics of the stock market and how this creates endogeneity challenges. The endogeneity problem is central to the discussion of the validity of the analysis results.

Sammendrag

Denne masteroppgaven har som formål å undersøke hvordan Google søkevolum korrelerer med og predikerer parametere på Nasdaq Copenhagen (Københavns Fondsbørs), og om en kan oppnå meravkastning ved bruk av søkevolum som en indikator ved aksjehandel. Jeg tar i bruk unormal avkastning, unormalt volum og volatilitet som mål på markedsaktivitet. Både daglig og ukentlig Google søkedata er studert. Google søkedata er innhentet basert på selskapets ticker og selskapets navn. I analysen blir panel data regresjon gjennomført for Google søkevolum på ukentlig og daglig basis. Resultatene viser at det er en liten positiv korrelasjon mellom unormalt Google søkevolum og unormal aksjeavkastning. Funnene fra regresjonsmodellene finner imidlertid få holdepunkter for at unormalt høyt Google-søkevolum korrelerer med eller predikerer unormal avkastning på et statistisk signifikant nivå. Regresjonsresultatene viser imidlertid at Google-søkevolum korrelerer med unormalt volum og volatilitet i både det ukentlige og daglige datasettet. Trading strategier som involverer Google søkevolum viser at en kan oppnå økonomisk gevinst ved å bruke Google-søkevolum som en parameter i kjøps- og salgsbeslutninger av finansielle produkter når transaksjonskostnader ikke er tatt med i beregningen. Denne oppgaven drøfter også den komplekse dynamikken til aksjemarkedet og hvordan dette skaper utfordringer med endogenitet. Endogenitetsproblemet står sentralt i drøftingen av validiteten til analysens resultater.

Table of Contents

List of Figures

List of Models

List of Tables

List of Abbreviations

AR	Abnormal returns
ASVI	Abnormal search volume index
ATV	Abnormal trading volume
HML	High minus low
OMXC20	OMX Copenhagen 20
SMA	Simple moving average
SMB	Small minus big
SVI	Search volume index
TV	Trading volume

1. Introduction

1.1 Motivation

Since the introduction of personal computers, access to the stock market has increased and information sharing has become faster. Personal computers enable individual investors to carry out their own stock research and purchases, making the stock markets more efficient. However, computers also make it more difficult for individual investors to generate abnormal returns due to pre-programmed algorithms implemented by institutional investors. The efficient market hypothesis states that new information about the value of traded securities is quickly priced into the stock price, and as a result we should rarely observe overpriced or underpriced securities. (Bodie, Kane and Marcus, 2014). Anomalies and insider information cause markets to be inefficient. Inefficient markets open the opportunity to generate profits from trading over- or underpriced stocks, because the stock prices may not reflect all information about the companies. Anomalies based on sentiment has been documented in varies of studies (Kamstra, Kramer and Levi, 2003; Joseph, Wintoki, and Zhang, 2011; Baker and Wurgler, 2007). Intangible assets[1] and new technologies are hard to valuate for investors, this can result in a collective mispricing of stocks. When stock evaluation is hard, investor attention and sentiment may play a large role in asset pricing. Because the price of a stock is the equilibrium price derived from the supply and demand at any given point in time, Internet search volume may reflect interest and public opinion about the attractiveness of a stock and give a clue about its future price prospects.

The motivation to study the correlation and the predictive power Google search volume has on stock market parameters is to explore how Abnormal Search Volume Index (ASVI), performs as an indicator for investor attention. ASVI is a relatively new proxy for investor attention and an increasing number of studies of ASVI are being conducted. I chose to study the effect of ASVI on the Nasdaq Copenhagen in order to add research on ASVI from a marked that has not been studied before.

[1] For a definition of *intangible asset* see the Cambridge dictionary https://dictionary.cambridge.org/dictionary/english/intangible-asset

1.2 Research Problem and Contribution

The aim of this paper is to study the effect of Google Search Volume on market factors of stocks on the Nasdaq Copenhagen in recent years using panel data regressions. The paper seeks to test the following hypotheses:

H1: *Google search volume correlates with abnormal stock returns, abnormal trading volume and volatility in stocks listed at Nasdaq Copenhagen.*

H2: *Google search volume predicts future abnormal stock returns, abnormal trading volume and volatility in stocks listed at Nasdaq Copenhagen.*

H3: *Daily data of Google search volume correlates with and predicts abnormal stock returns, abnormal trading volume and volatility differently than weekly Google search volume in stocks listed at Nasdaq Copenhagen.*

H4: *A trading strategy created based on AVSI generates abnormal returns*

This thesis will investigate if ASVI can be linked to the pricing of shares, trading and volatility on the Nasdaq Copenhagen. It will also investigate the difference between daily and weekly Google search attention. The thesis will also shed light on the implications of suggesting a causal connection between Google search volume and stock market data. Lastly, the thesis will go through trading strategies previously used to see if ASVI can be used to generate abnormal returns on the Nasdaq Copenhagen.

This thesis investigates the *correlation* between Google search volume and market parameters of the current period. It also investigated whether Google search volume can predict current market parameters by using Google search data the day and week before changes in the parameters of the market. Correlation is a measure of how two variables co-vary (Johannessen, Christoffersen and Tufte, 2011). From a regression model we can only observe correlation between variables, where we can observe patters and regularities, not *causality* (Hume, 2011).[2] Hume (2011) states that science needs to be cautious about making causal claims. One cannot be certain that all the requirements are met. This is because conclusions about causal direction is a subjective assessment, where one is looking for plausible mechanisms that can explain a phenomenon. Since a solid foundation is needed in order to draw causal conclusions, I would like to emphasize that this study looks at the correlation between Google search volume and market parameters and does not draw conclusions about causation.

The analysis of the weekly data is based on the methodology of Kim, Lučivjanská, Molnár and Villa (2019). The analysis will have three additions to the methodology of Kim et al. (2019): In addition to studying weekly Google search data, this thesis is studying daily Google search data. Second, this thesis will account for time fixed effects in the panel data regressions as done in Da, Engelberg and Gao (2011). Last, this thesis will also highlight and test the robustness of the regression assumptions, and how violations of these are handled. Google searches based on company ticker and company name will both be included in the analysis. A weak positive correlation with abnormal returns is detected for ASVI for company name and for ASVI for ticker symbol. Using a panel data regression with time- and entity fixed effects yield little evidence of a statistically significant relationship between ASVI and abnormal returns. However, in the dynamic regression, ASVI for company name have statistically significant predictive power for abnormal stock returns. The results of the analysis show that ASVI correlates with abnormal trading volume and volatility in the current period in the weekly data and in the daily data. Previous values of ASVI correlates with future values of the market parameters in the weekly data, while no such evidence is found for the daily data. Trading strategies utilizing ASVI is shown to yield abnormal returns when not accounting for transaction costs.

---

[2] Causality is the relationship between a cause and effect. One can say there is a causal relationship between two variables if one event occurs as a result of another event occurring (Tufte, 2018). Requirements for causal inference include: 1) Correlation or co-variation, 2) Causal direction, 3) Non-spurious context and 4) Empirical association (Tufte, 2018).

1.3 Limitations of the Study

Endogeneity

In regression analyzes, biased estimates of causal effects are often the result of explanatory variables that are endogenous, that is, correlated with the error term (Tufte, 2013). The explanatory variables are correlated with the error term when there is an omitted variable, simultaneous causality, autocorrelation, measurement error, or reciprocal causation. Stock markets are dynamic and continual change complex variables A major issue when studying stock markets is that they are complex and the equilibrium prices for any given moment are likely to be a result of simultaneously determining processes. That is, there is a two-way relationship between the dependent and the independent variables (Stock and Watson, 2014). For example, a higher trading volume might lead to higher search volume within a very short time as the market adapts quickly. This can cause problems in the analysis, as the beta coefficients from the ordinary least squares (OLS) regression may be biased and inconsistent. This could be a problem when examining hourly data. However, this thesis studies weekly and daily data. This analysis looks at search volume data one day before market parameter data, instead of the same day. The main issue when studying stock markets is the omitted variable bias. Many econometric papers are focused on the violation of the exogeneity assumption. Pooled OLS estimates have been used in most of the studies of the relationship between Google search volume and stock market parameters (Da et al., 2011; Drake et al., 2012; Kim et al., 2019; Tan and Taş, 2019). Because of the complex dynamics of stock markets, the OLS regressions are likely to suffer from omitted variable bias. Some studies also include time-and entity fixed effects. However, it is likely that the omitted variables are not constant. In which case, they are not accounted for in an OLS regression with entity- and time fixed effects. Based on the reasoning above, the strict exogeneity assumption is likely violated. When interpreting the results of the OLS analysis one should be vigilant. It is virtually impossible to circumvent the endogeneity problem. Statistical methods to reduce the endogeneity problem exists. However, these go beyond the scope of this master’s thesis.

The Shortcomings of Google Search Index

Google search volume is only provided as a relative measure of search queries on a scale from 0 to 100. The Google Trends site does not provide the search queries in absolute numbers. In addition, the Google site does not provide data for search queries done by relatively few people. This causes problems in the sample selection of this study, because only companies that have a certain number of search queries can be included in the study.

Investor Sentiment

Another implication of using Google search volume index as a measure of investor attention, is that the underlying cause of attention in search queries are not explicitly known. In the literature review the ambiguous findings of research papers of the causal direction between ASVI and abnormal returns will be highlighted. Search volume can rise based on both positive and negative predictions of how the stock prices are going to evolve. Based on this, Google search volume should be supplemented with other measures of investor sentiment such as: news articles, stock forums, subjects of conversations etc.

Sample Size

The sample in this study reflects 24,8 percent of the population, that is companies listed at Nasdaq Copenhagen. In other words, the sample most likely does not reflect the population. The central limit theorem states that when the sample size gets larger, the sampling distribution of the mean is approximately normally distributed (Mehmetoglu and Jakobsen, 2017). The sample size only matters if the population that is being studied is small. The sample size needed in percentage for a small population is greater than the sample size in percentage needed for a larger population (Mehmetoglu and Jakobsen, 2017). The Nasdaq Copenhagen holds a relatively small population of 137 companies as of January 2020. Sample size can cause problems to the normal distribution of the mean. The confidence intervals become larger because of greater standard errors, i.e. the estimates become less precise. The small size of the sample used in this study is mainly due to information shortages for small companies, and search word implications such as "noisy tickers" that will be explained further in Section 3.3. On the one hand, small sample size is not necessarily a threat to internal validity, that is whether statistical inference and conclusions are valid for the population and setting that is studied. On the other hand, small sample size is a threat to external validity and the precision of the estimates. External validity is the extent to which statistical inferences and conclusions from the study can be generalized to other populations and settings (Stock and Watson, 2014). This thesis will discuss sample size and definition in detail in Section 3.3.

1.4 Structure of the Paper

This Thesis is divided into 7 chapters. Chapter 1 has introduced the motivation, research problem, research contribution, the limitations of the study and the structure of the paper. Chapter 2 will provide a literature review. Chapter 3 seeks to familiarize the reader with the characteristics of the Danish stock market and the Google Trends dynamics. This chapter will also go through the sample definition and the variables included in the analysis. Chapter 4 will test the panel data regression assumptions and will go through the methodology. Chapter 5 will test the hypotheses of this thesis and provide the static and dynamic panel regression results. In Chapter 6 results regarding trading strategies will be presented. Chapter 7 will present the conclusion and reflections of the study.

2. Literature Review

This section outlines the early literature on investor attention and investor sentiment and their influence on stock prices and market efficiency. The chapter will then review the results of other studies utilizing the Google Search Volume Index as an attention proxy in American, European and emerging stock markets.

2.1 Investor Attention

The Efficient Market Hypothesis states that in strongly efficient capital markets there will be no excess return to gain by using technical or fundamental analysis (Bodie et al., 2014). This is because the stock prices reflect all relevant available information about the firms. However, anomalies have been found in the capital markets, which can yield excess returns by technical or fundamental analysis (Bodie et al., 2014). There are many anomalies in the stock markets that are discussed in previous literature, such as: mean reversion (Poterba and Summers, 1988), herding (Devenow and Welch, 1996), momentum, and calendar effects that are due to the psychology of the investors. Literature investigating the efficient market hypothesis has shown that information is priced into the stock prices prior to public announcements (Keown and Pinkerton, 1981; Busse and Green, 2002). If uninformed investors could detect that other investors are in possession of non-public information in advance through Google search volume, these uninformed investors could also act accordingly in advance of public announcement.

Studies have shown that investors have limited attention (Kahneman, 1973). Merton (1987) modeled the effects of investor attention on financial markets. His model predicts that the firm's security value increases with firm recognition, but that the expected return is decreasing with firm recognition. Intuitively this is the case because larger firms will have greater recognition and a larger investor base. To gain excess returns from mispricing of larger firms will be harder than for less recognized firms. This is because the larger investor base ensures more accurate pricing of the shares. Fang and Peress (2009) found that stocks with no media coverage earn higher returns than stocks with high media coverage, and that the results are more pronounced among small stocks. Barber and Odean (2008) confirm their hypothesis that individual investors are net buyers of attention-grabbing stocks. E.g. Stocks in the news, with high one-day returns and high abnormal trading volume. Seasholes and Wu (2007) find that when stocks hit upper price limit events, such as high returns, high volumes and news coverage, they attract investors' attention in the Shanghai market. Active individual investors may buy stocks that were not previously owned in the aftermath of attention-grabbing events. Such upper -limit events are followed by initial price increase and price mean reversion over the following week. DellaVigna and Pollet (2009) show that the response to earnings announcements on Fridays has a slower immediate response, and lower trading volume than on other weekdays. They suggest these findings to be caused by the fact that investor inattention is more likely on a Friday.

Baker and Wurgler (2007) studied the effect investor sentiment has in the stock market. They define investor sentiment broadly as "a belief about future cash flows and investment risks that is not justified by the facts at hand" (Baker and Wurgler, 2007, p.1). They build a sentiment index based on six proxies: trading volume, the dividend premium, the closed-end fund discount, the number of IPOs, first-day return on IPOs, and the equity share in new issues. The evidence from their research suggests that it is possible to measure investor sentiment. They also show that the average monthly return is higher for companies whose sentiment level was low the preceding month. Tetlock, Saar-Tsechansky and Macskassy (2008) studied how the language used in news articles can be used to predict individual firms' stock or accounting earrings. They found that the fraction of negative words in firm-specific news stories forecasts low firm earnings.

2.2 Internet Search Volume

Empirically there have been difficulties in measuring attention. That is because there is no exact measure for attention. To investigate the effects of investor attention, indirect proxies for attention has been implemented. One of the newest proxies for investor attention being internet search queries done through internet browsers such as Baidu (Shen, Zhang, Xiong and Zhang 2017), Yahoo (Mangold et al., 2005; Lawrence, Ryans, Sun and Laptev, 2016) and Google Search Volume (Da et al., 2011). Other proxies for investor attention from the internet include Wikipedia searches (Moat et al., 2013), Twitter (Bollen, Mao and Zeng, 2011; Bartov, Faurel and Mohanram, 2018), stock forums (Ackert, Jiang, Lee and Liu, 2016), amongst many others. A search volume proxy is a more direct measure of investor attention because a news article or media coverage does not guarantee attention unless an investor reads it (Da et al., 2011). Moat et al. (2013) find that Wikipedia data provide some insight into future trends in behavior for market actors. Evidence from their study shows that the number of page views of articles, that are related to financial topics or companies, increased before the stock market falls. Bollen et al. (2011) studied if the effect of public mood through Twitter feeds is correlated with the value of the Dow Jones Industrial Average. Evidence from the study suggests that public mood states are statistically significantly correlated with daily changes in the Dow Jones Industrial Average closing prices. Ackert et al. (2016) find that influential investors and users, who are popular in stock forums, are actively targeting large and liquid firms. They also prefer local investments in their messages. Predictions of influential investors on stock forums are more likely to indicate subsequent return compared to other investors (Ackert et al., 2016).

Google search volume has been used as a proxy for investor attention in more recent papers such as “In search of attention” by Da, Engelberg and Gao (2011), “Investor Information Demand: Evidence from Google Searches Around Earnings Announcements” by Drake, Roulstone and Thornock (2012) and “Google searches and stock returns” by Bijl, Kringhaug, Molnár and Sandvik (2016). Da et al. (2011) argue that Google search volume is an unambiguous attention measure because you are undoubtedly paying attention to a stock when conducting a search on that stock. Da et al. (2011) found that an increase in search volume index (SVI) predicts higher stock prices in the following two weeks, and an eventual price reversal within the year in a sample of Russel 3000 stocks. They also found that amongst the most searched stocks, the momentum effect is higher. Drake et al. (2012) found that Google search volume increased two weeks prior to earnings announcements and spikes markedly at the announcement for S&P 500 firms. They also found that preannouncement price and volume changes reflect more of the upcoming earnings news when the search volume prior to the announcement is higher. Preis, Reith and Stanley (2010) also studied S&P 500 firms and found a positive correlation between web searches and trading volume. A study by Tan and Taş (2019) found that high Google search volume results in higher returns and that the price pressure effect of Google searches is stronger among small stocks on the

Turkish stock exchange Borsa Istanbul. Tan and Taş (2019) argued that return premiums may be higher in emerging markets, such as the Turkish one, where information efficiency is lower. They also argued that higher search volume can be linked to higher abnormal stock returns because search activity on the web is more likely to be linked to the action of buying. Another study of an emerging market, the Bombay Stock Exchange, conducted by Swamy, Dharani and Takeda (2019) of S&P BSE found that higher quantiles of SVI predict positive and significant returns in the subsequent first and second week when using a quantile regression approach. Bank, Larch and Peter (2011) found that increases in search queries are associated with increased trading activity, liquidity and temporarily higher future returns in German stocks. Joseph et al. (2011) used internet search volume as a proxy for investor sentiment in their study of a sample of S&P 500 firms from 2005-2008. By allocating companies in quantiles based on their search volume, and re-sorting the quantiles based on new search volume every week, they find that search intensity predicts abnormal stock returns and trading volume. Like Baker and Wurgler (2007), Joseph et al. (2011) found that sensitivity to returns is positively related to the difficulty of a stock being arbitraged. A study conducted by Kim et al. (2019) of the Norwegian stock market, shows that Google searches are neither correlated, nor able to predict, abnormal returns. Nonetheless, they found that increased Google searches can predict increased trading volume and volatility. On the contrary, a study by Shen et al. (2017) found that increased search volume on the search engine Baidu leads to reduced stock prices. This is the same result as Moat et al. (2013) found for Wikipedia searches. Moat et al. (2013) argued that a possible explanation for this is the notion that investors are loss averse. That is, the fear of losing a euro is greater than the opportunity to gain a euro. Therefore, investors may execute more searches before taking on a trade when they view the investment to be of greater consequence. Similarly, Bijl et al. (2016) studied the S&P 500 in the time interval 2008-2013 and found that high values of Google search volume lead to negative returns. They show that, when transaction costs are not considered, buying stock with infrequent Google searches and selling stocks with frequent Google searches is profitable. However, this strategy is no longer profitable when taking transaction costs into account.

3. Data

This chapter reviews the characteristics of the Nasdaq Copenhagen in Section 3.1 and the dynamics of Google Trends in Section 3.2. This section also presents the variables that are included in the regression. Finally, Section 3.3 describes the sample characteristics and the sample selection process.

3.1 Stock Market Data

3.1.1 Nasdaq Copenhagen Characteristics

Copenhagen stock exchange has officially been named Nasdaq Copenhagen since 2014. Nasdaq Copenhagen is Denmark's stock exchange and has 137 listed companies as of January 2020. The domestic market capitalization of the exchange is 445 888 million USD (Sustainable Stock Exchanges Initiative, n.d.). Nasdaq Copenhagen is mostly dominated by the financial, industrial- and health care sectors. The main stock index for Nasdaq Copenhagen is the OMXC25. OMXC25 contains the 25 most traded and largest shares on Nasdaq Copenhagen. This index began trading on December 19th, 2016.[3] Because this index began trading during the time interval for this study, the index's predecessor, OMXC20, is used in this paper as a benchmark for market return. Similarly, this index consists of the 20 most actively traded shares on Nasdaq Copenhagen. The OMXC20 is a market-weighted price index and began trading in 1989.[4] Compared to the markets studied in the previous literature, Nasdaq Copenhagen has relatively few listed companies. Nevertheless, Nasdaq Copenhagen is a developed market as categorized by Kenneth R. French (n.d.). Shorting is also limited in the Danish market, as only 44 stocks can be shorted as of April 17th, 2020, through the online broker Nordnet.[5] The shares offered to be shorted change continuously and the margin requirements are typically high. This creates limits to arbitrage and abnormal

---

[3] Naqsdaq OMXC25 index information collected from https://indexes.nasdaqomx.com/Index/Overview/OMXC25

[4] Nasdaq OMXC20 index information collected from https://indexes.nasdaqomx.com/Index/Overview/OMXC20

[5] Nordnet list of stocks that can be shorted https://classic.nordnet.no/mux/page/blankninginl.html?valuta=DKK

returns utilizing trading strategies based on ASVI. This is discussed further in Chapter 5.

3.1.2 The Processing of Data and Market Variables

The stock market data from Nasdaq Copenhagen is obtained from Yahoo! Finance.[6] Both daily and weekly data are collected. The sample period for the weekly data is from 2016 to 2019. The daily data has a sample period from July 2019 to December 2019. Trading volume is standardized in order to determine the relative size of the company (Mehmetoglu and Jakobsen, 2017). This is because the trading volume may vary greatly based on the size of a company and the number of outstanding shares. A variable is standardized when the mean is subtracted from the values of the variable and divided by the standard deviation (Mehmetoglu and Jakobsen, 2017). In the weekly data, standardization is based on the previous year. Where standardization is needed in the weekly data, data from 2015 is used. The daily data is standardized based on the previous business month, which is 19 days. Where standardization is needed in the daily data, data from June 2019 is used. For the rest of this chapter, the explanation of how the variables are constructed will be based on the weekly data to keep the notation simple. The daily data is treated equally, but with daily data and standardization based on the previous 19 days.

Abnormal returns

Returns are calculated using natural logarithms in order to model relative effects and to reduce the effect of outliers. A skewed or pointy distribution can create problems for regression analysis (Mehmetoglu and Jakobsen, 2017). Transforming variables using natural logarithms makes the distribution more symmetrical and the data more interpretable (Mehmetoglu and Jakobsen, 2017). When calculating the returns, it is necessary to adjust for stock splits and dividends. Stock splits and dividends give

[6] Yahoo! Finance website is available at https://finance.yahoo.com/

signals to the market that may affect the stock price.

Yahoo's adjusted closing price is used in the calculation of returns, as it has already been adjusted for stock splits and dividends:

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

where R_t is the log returns, P_t is the adjusted stock price for week t, and P_{t-1} is the adjusted stock price from the previous week.

The expected returns are calculated using the Fama and French three-factor model (Fama and French, 1993).[7] The three factors used in calculating expected returns are market return, size factor (small-minus-big, SMB), and the value factor (high-minus-low, HML). The expected return is calculated using data from Kenneth French's online data library (Kenneth R. French - Data Library, 2020). The factor sensitivity betas are calculated using a 1-year rolling regression. The expected returns are then calculated as:

$$r_t = r_f + \beta_1(r_m - r_f) + \beta_2(SMB) + \beta_3(HML) + \varepsilon,$$

where r_f is the risk-free rate and the β's are the stock sensitivities to the Fama-French factors.

Abnormal stock returns are calculated as the difference between the actual returns and the expected returns derived from the Fama and French three-factor model:

$$AR_t = R_t - r_t.$$

[7] The factor data was constructed by Kenneth R. French and is available at https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Abnormal trading volume

Abnormal trading volume is standardized by subtracting the average trading volume in the previous year from this week's trading volume, and dividing by the standard deviation in the previous year:

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

where TV_t is this week's trading volume and σ_{TV} is the standard deviation of the trading volume in the previous year.

Volatility

Volatility is calculated using the opening-jump adjusted Garman-Klass (Garman and Klass, 1980) volatility estimator, as discussed in Molnár (2012). This estimator for volatility has the benefit of incorporating information about intraday variability in the estimation of volatility. To capture volatility for a given week, the daily variance is calculated:

$$Variance_{it} = \frac{1}{2} \cdot (h_t - l_t)^2 - (2log(2) - 1) \cdot c_t^2 + j_t^2,$$

where

$h_t = log(high_t) - log(open_t),$

$l_t = log(low_t) - log(open_t),$

$c_t = log(close_t) - log(open_t),$

$j_t = log(open_t) - log(close_{t-1}).$

$open_t$ is the price at the beginning of the day, i.e. $t = 0$. $close_t$ is the price at the end of the day, i.e. $t = 1$. $high_t$ is the highest price of the day. low_t is the lowest price of the day.

Following Kim et al. (2019), the volatility is then calculated on a weekly basis, as the square root of average daily variance:

$$Volatility_{it} = \sqrt{\frac{1}{|S_t|} \sum_{i \in S} Variance_{it}} \; .$$

3.2 Google Search Data

This section will show how the Google Search Volume Index values are constructed and how the Search volume Index values are standardized.

3.2.2 The Dynamics of Google Trends

The internet search volume data is obtained from the Google Trends site.[8] The Google Trends website provides a data sample of the relative search volume of specific search terms done through the Google browser. Sample data is used to display interest in a search term on a global, national or city-level. The search queries are normalized on a scale from 0 to 100 in order to compare search data (Google, 2020). Google Trends normalizes the data based on the time and location of a search query. The search data is normalized based on the following process as written on the Google Trends site (Google, 2020):

[8] Google Trends website is available at https://trends.google.com/trends/?geo=US

- Each data point is divided by the total searches of the geography and time range it represents to compare relative popularity. Otherwise, places with the most search volume would always be ranked highest.
- The resulting numbers are then scaled on a range of 0 to 100 based on a topic's proportion to all searches on all topics.
- Different regions that show the same search interest for a term don't always have the same total search volumes.

An upward sloping trend line means that the search term's relative popularity compared to other searches is increasing, not necessarily that the total number of searches on that term is increasing. (Google, 2020). Google Trends filter searches by excluding certain data such as: searches that are made by very few people for a given time period, duplicate searches from the same user and special characters.

Google Trends search data can be filtered through the following: (1) time, (2) geographical location, (3) category and (4) search channel. Google trends filter search queries based on time. Search activity can be filtered on an hourly, daily, weekly or monthly basis. However, the time format of search queries depends on the time-range set by the user. E.g. if you want to download search queries for a one-year time period, you are only able to obtain SVI data on a weekly basis. Daily data is available for time periods shorter than eight months. Hourly data is available for time periods shorter than 7 days. Google Trends filters information about search activity based on geographical location. Preis, Moat and Stanley (2013) fond that searches filtered through geographical location improve the explanatory power for queries in the geographic location in question. The geographical location in this study is thus set to Denmark. Setting the geographical location to the graphical location in question reduces the risk of contamination of search results as ticker or company name may have a different meaning in another language. Google Trends can also filter searches based on 25 categories.[9] The default filter is set to "all categories". Bijl et al. (2016)

[9] Google Trends categories filter include: (1) Arts & Entertainment, (2) Autos & Vehicles, (3) Beauty & Fitness, (4) Books & Literature, (5) Business & Industrial, (6) Computers & Electronics, (7) Finance, (8) Food & Drink,

find that using the finance filter does not outperform the unfiltered searches in terms of predicting stock returns. The last filter Google Trends provides is a filter for search channel. The search channel option filters the search queries according to the channel the search activity was done. The filters are web- (default), news-, image-, shopping- and YouTube search. In this study the search channel is set to the default, which is web-searches.

Figure 1 below illustrates the output data generated from the search term “Elgiganten” in Denmark for the last five years as of January 2020. “Elgiganten” is a Danish electronics store. Search volume peaks in the November and December months. The relatively high search volume in these months is likely the results of searches by consumers who are Christmas-shopping. This example illustrates how Google search volume is related to investors buying behavior.

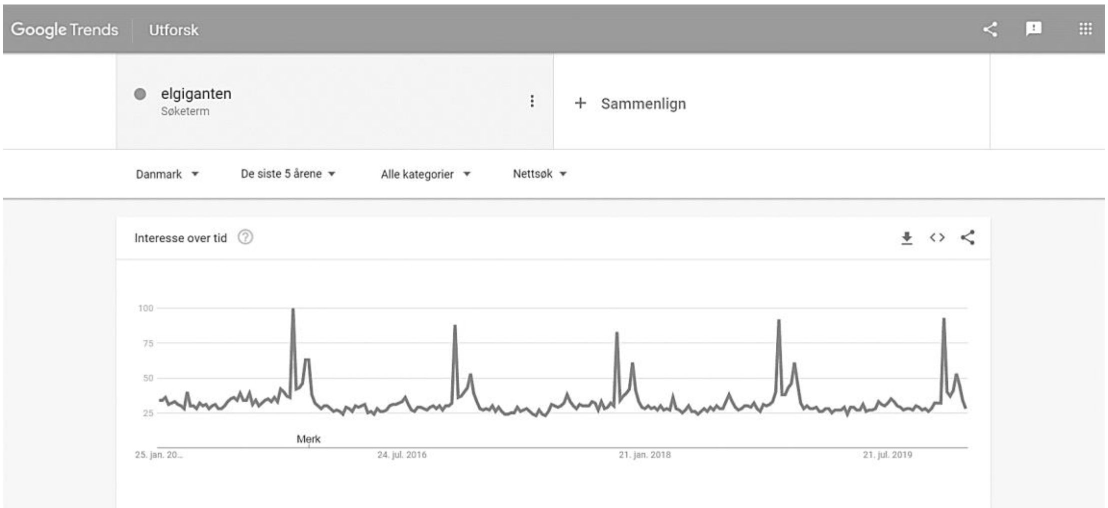


Figure 1 Example of Google Trends output from the search term "Elgiganten"

Web searches can be used as a proxy for investor attention. In order to use Search volume from Google Trends, which only captures search queries on the Google search

---

(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, (25) travel.

engine, I investigated search engines market share in Denmark through the webpage Statcounter. Table 1 is made from numbers provided by Statcounter. It shows that the Google search engine has dominated the search engine market in Denmark in the time period for this analysis. With a market share of over 95% in the last five years, Google searches are approximately equal to the general search behavior in Denmark.[10]

Table 1

Search Engine Market Share Percentage Denmark (2015-2020)

Year	Google	Bing	Yahoo!	Other
2015	95,32 %	2,19 %	1,87 %	0,62 %
2016	95,66 %	2,54 %	1,41 %	0,38 %
2017	95,54 %	2,73 %	1,21 %	0,52 %
2018	95,69 %	2,57 %	1,18 %	0,56 %
2019	97,17 %	1,37 %	0,83 %	0,63 %

Calculation based on numbers retrieved January 19th, 2020, from https://gs.statcounter.com/search-engine-market-share/all/denmark/#monthly-201501-202001

Google Search Volume data is obtained from the Google Trends site. In addition to real-time data, the engine provides historical data from 2004 up to 36 hours prior to search activity. Google Trends doesn't provide the total number of queries of a given search term. They provide a standardized scale from 0 to 100, where 100 represents the highest query volume during a considered time period and geographic region (Choi and Varian, 2012). Daily data is only available for time ranges up to eight months. Data series over five years are only available on a monthly basis. This analysis uses a four-year range for the weekly data in order to obtain weekly SVI and a six-month range in order to obtain daily SVI.[11]

[10] Source: Statcounter Retrieved January 19th, 2020, from https://gs.statcounter.com/search-engine-market-share/all/denmark/#monthly-201501-202001

[11] This obstacle cannot be circumvented by overlapping two or more time series because the normalization of SVI data depends on the time range set by the user.

Different studies have come to different conclusions as to whether ticker or company name gives the best search results. One the one hand, Bijl et al. (2016) find that using the company name results in a stronger relationship to stock market returns than using ticker symbols. On the other hand, Da et al. (2011) have two arguments for why it is better to use searches based on ticker symbols rather than the company name. The first argument is that searches on company names may be conducted for other reasons than investment. The second argument is that different investors may search with different variations of a company's name. This analysis will include both searches on company ticker and company name. Abnormal search volume index for company ticker is denoted *ASVI_ticker*. Abnormal search volume index based on company name is denoted *ASVI_name*.

3.2.2. Abnormal Search Volume Index

The search volume index values provided by Google trends is standardized to abnormal search volume index in order to make the SVI values more comparable. Following the formula from Bijl et al. (2016), average SVI from the past 52 weeks is subtracted from the raw weekly SVI, and their difference is divided by the standard deviation of the previous year:

$$ASVI_t = \frac{SVI\_n_t - \frac{1}{52}\sum_{i=1}^{52} SVI_{t-i}}{\sigma_{SVI,t}},$$

where SVI_n_t can either be *SVI_ticker* or *SVI_name* in week t, and $\sigma_{SVI,t}$ is the standard deviation of the SVI for the past 52 weeks.

The Nasdaq Copenhagen market reports data from when the stock exchange is open for trading, which is from Monday to Friday. Google Trends report their data from Sunday to Sunday. In order to match the two datasets, I will follow the procedure from Kim et al. (2019) for the weekly data. The Google SVI and trading volume is standardized. A standardized variable is a variable where the mean value of the variable is subtracted from each observation and then divided by the standard deviation of that variable (Mehmetoglu and Jakobsen, 2017). The weekly data is standardized based on the previous 52 weeks. The daily data will follow a similar approach but will be standardized based on the previous business month, which is 19 days. The weekly and daily data is standardized differently. This is because Google's search volume index presentation of data differs when searching for SVI at different time intervals. In the weekly panel data, the ASVI is given for every Sunday, while the stock market data refers to the day after, which is Monday. Due to differences in the weekly and daily series, a direct comparison of the results of the analysis, will not be appropriate. When interpreting the analysis results, the differences in the structure of the time series must be considered.

3.3 Sample Size and Definition

Nasdaq Copenhagen has 137 listed companies as of January 2020.The final sample size for this study consists of 34 companies. The sample is defined based on three selection criteria. The first selection criterion is that there is enough Google search data available. This means that companies that did not meet Google Trends lower threshold for search queries, and for which, therefore Google provide no data, is excluded from the sample. The second criterion in the selection process is to omit “noisy tickers” as done in the study by Da et al. (Da et al., 2011). “Noisy tickers” are ticker names that are ambiguous. Examples include names that are identical to the company name and may be used to search for other than investment purposes. Examples of tickers that were omitted are PARKEN, NORDIC and TOP, which have other meanings in Danish. Every ticker has been searched for manually. This is to ensure that the ticker is not ambiguous and search terms based on the company are the main search results. The third criterion is that only companies with both ticker and search names are included. This means companies who only have search results on either ticker or company name are excluded. The population of the study is all the 137 listed companies on Nasdaq Copenhagen. Out of the 137 listed companies, 67 had insufficient data information and 29 are categorized as a "noisy ticker" and therefore omitted from the sample. 7 companies are excluded because they were listed on the stock exchange during the sample period. The final sample consists of 34 companies. This sample size is smaller than for studies of larger stock markets such as Da et al. (2011). However, it resembles the sample sizes of Aouadi, Arouri and Teulon (2013) and Kim et al. (2019), where the sample sizes are 40 and 36 respectively. Table 2 shows the sample's distribution of sectors versus the distribution of sectors to the entire Nasdaq Copenhagen stock exchange. A list of all companies included in the study can be found in Appendix A. A list of the variables included in the analysis is available in Table 3.

Table 2

Industry Representation in Population Versus Final Sample

	Population		Sample	
Sector	Quantity	Proportion (%)	Quantity	Proportion (%)
Health Care	18	13 %	8	24 %
Financials	43	31 %	6	18 %
Consumer Goods	13	9 %	3	9 %
Oil and Gas	4	3 %	1	3 %
Industrials	36	26 %	9	26 %
Utilities	2	1 %	0	0 %
Technology	6	4 %	1	3 %
Consumer Services	14	10 %	5	15 %
Basic Materials	1	1 %	1	3 %
Total	137	100 %	34	100 %

Table 3

Variables Definition

Variable	Definition	
SVI	Search Volume Index	Scaled measure of weekly aggregate search frequency on a scale of 0 to 100.
ASVI	Abnormal Search Volume Index	Current week SVI minus the average SVI from the past 52 weeks (19 days for daily data) divided by the standard deviation of the SVI the past 52 weeks for the weekly data (19 days for daily data).
AR	Abnormal Return	The difference between actual returns and the expected return derived from the Fama and French three-factor model.
ATV	Abnormal Trading Volume	Current week trading volume minus the average trading volume from the past 52 weeks for the weekly data (19 days for daily data) divided by the standard deviation of the trading volume the past 52 weeks for the weekly data (19 days for daily data).
Volatility	Volatility	Jump-adjusted Garman-Klass volatility estimator.

4. Methodology

The outline of this chapter is as follows: Section 4.1 investigates the panel data regression assumptions. Section 4.2 presents the static regression model. Section 4.3 presents the dynamic regression model. All regressions are calculated using the statistical program Stata.[12] The Stata user-written commands can be found in Appendix B.

4.1 Regression Assumptions and Robustness Tests

In order to conduct a panel data regression with fixed effects, the following assumptions must hold (Stock and Watson, 2014):

I. $E(u_{it}|x_{i1}, x_{i2}, \dots, x_{iT}, \alpha_i,\ \lambda_t) = 0$. The conditional mean of the errors is zero.

II. $(x_{i1}, x_{i2}, \dots, x_{iT}, y_i)$ are i.i.d. (independent and identically distributed) hover the cross-section.

III. Large outliers are unlikely.

IV. There is no perfect multicollinearity

If these assumptions hold, then the OLS estimators are unbiased and consistent estimators that approximately have a normal distribution (Stock and Watson, 2014). Large amounts of missing data can cause a loss of efficiency in the models (Mehmetoglu and Jakobsen, 2017). The panel data in this study is strongly balanced, i.e. there are an equal number of time period observations per company.

[12] For information about the Stata software see the Stata website at https://www.stata.com/

The first regression assumption is the strict exogeneity assumption. The assumption states that any independent variable is uncorrelated with the error term. If this assumption doesn't hold, then we have an endogeneity problem. The most common violation is due to omitted variable bias.[13] An example of an omitted variable that can affect the stock prices is the management. Change in the management of a company, or change in the stock holdings of the management, might be a factor that affects stock prices indirectly through trading volume or volatility. If this variable is not included in the regression, then it is an omitted variable that causes a spurious relationship between trading volume and stock price.[14] This paper uses company- and time fixed effects. The omitted variables that are constant over time and constant for the companies can be accounted for. This reduces the omitted variable bias to some extent. However, it is likely that the omitted variables are not constant. In which case, they are not accounted for in an OLS regression with entity- and time fixed effects. As mentioned in the introduction, it is virtually impossible to circumvent the endogeneity problem. Gippel, Smith and Zhu (2015) state that textbook solutions also have implications and that natural experiments are the state-of-the-art solution to the endogeneity problem (Gippel et al., 2015). Chapter 6 will go through a trading strategy. The conditions for conducting a natural experiment are not optimal because Google does not present data for searches done by a few individuals. This leads to sample selection bias because only the searches of companies that exceed Googles threshold can be included. Despite this, it is possible to look at the outcome of the trading strategies in Section 6 as an informal experiment.

Intuitively, stock return, volatility, or trading volume for any given time are likely to be correlated with their past values. The Breusch-Pagan/ Cook-Weisberg test for heteroskedasticity (Breusch and Pagan, 1979; Cook and Weisberg, 1983) returns a significant value in all regression models (Results from the Breusch-Pagan/ Cook-Weisberg test are presented in Appendix C). This means there is a presence of heteroskedasticity.

---

[13] Omitted variable bias occurs if one of the dependent variables in the study returns a large and significant estimate on the dependent variable, which is due to variation in an omitted variable that is contained in the error term (Stock and Watson, 2014).

[14] E.g. an analysis can indicate that higher trading volume is the reason for higher stock prices. If the management buys more shares in the company this may be an indication of good prospects for the firm. Investors will in turn buy the stock and drive the price up. So, it is actually the changes in management stock holdings that drive the stock prices up, not the trading volume.

Heteroskedasticity leads to bias in the estimates of standard errors in the model (Mehmetoglu and Jakobsen, 2017). Cluster robust standard errors are used to account for this problem.

The Woolridge test for autocorrelation in panel data is conducted and the results can be read in Appendix D. The test rejects the null hypothesis of no autocorrelation in the regression models where the dependent variable is abnormal trading volume and volatility, in both the weekly and daily regression models. The serial correlation of the regression error causes the standard errors to be inconsistent. Consequently, cluster robust standard errors are applied to account for serial both autocorrelation and heteroscedasticity, as done in Da et al. (2011) and Drake et al. (2012). Cluster robust standard errors relax the assumption that errors are independent of each other and normally distributed. Coefficients are not affected by the use of robust standard errors, but the p-values will be reasonably accurate since the standard errors are changed (Mehmetoglu and Jakobsen, 2017).

The second regression assumption holds if a simple random sample is conducted. However, this study has data missing based on Google search volume. This is because Google does not provide data on search queries that are executed by a relatively low number of people as described in Section 3.3. On the one hand, data missing based on an independent variable is not necessarily a threat to the internal validity, that is whether statistical inferences are valid for the population and setting studied (Stock and Watson, 2014). On the other hand, this is a threat to the external validity of the study. The conclusions from the study can most likely not be generalized to other populations and settings (Stock and Watson, 2014).

The third regression assumption is that large outliers are unlikely. In order to detect influential observations in the model, Cook’s distance plots have been employed. The Cooks distance graphs in Appendix E and F show that some companies have a great influence on the regression models. In the daily dataset a special case is detected. The company Scandinavian Private Equity A/S (SPEAS) has had a very big influence on daily data. The company has had a tremendous decline in stock price, high trading volume, and volatility in the time period for the daily data. This largely affects the regression model and the statistics. Therefore, as recommended by Mehmetoglu and Jakobsen (2017), the daily regression is conducted again without the observations of this company. The regression model without the influential observations is used in the analysis, while the regressions with the influential observation are available in the appendices.

A VIF (Variation Inflation Factor) statistic is also conducted to see if multicollinearity is a problem in the dataset. An exact number for when multicollinearity is a problem is not set. A rule of thumb says that a variable has a high degree of collinearity with multiple variables if the VIF statistic is greater than five (Mehmetoglu and Jakobsen, 2017). The VIF statistics can be found in Appendix G. None of the VIF statistics exceed five, and we can conclude that there is probably not a large problem with multicollinearity in the regression models.

4.2 Static Model Regressions

This paper follows the methodology and the preparation of the variables that are done in Kim et al. (2019), with the three additions as discussed in the introduction. This thesis includes daily Google search data, time fixed effects, and investigation of the panel data regression assumptions.[15] Standard errors of coefficients may be underestimated when common time-specific factors are not accounted for, which in turn can result in coefficients incorrectly being statistically significant (Gow, Ormazabal and Taylor, 2010). Therefore, this study conducts panel data regression with company fixed effects as done in Kim et al. (2019), and with time fixed effects as done in Da et al. (2011). The data is sorted into panel data, following a single company over time. The advantage of panel data is that company- and time-specific effects can be held constant. This way we can overcome omitted variable bias, even if those variables are unobserved (Stock and Watson, 2014).

This study includes both static and dynamic regression. The interpretation of the two models will therefore differ. The static regression investigates if the market parameters are correlated with the dependent variable in the current period. The dynamic regression investigates whether the values of the market parameters in the previous period correlates with the future values of the dependent variable. This is of interest as one can then get insights into how ASVI correlates with market parameters at two different points in time.

[15] Entity fixed effects capture all characteristics that are constant over time for a given company and time fixed effects capture all characteristics that vary over time but are the same for all companies (Stock and Watson, 2014).

Model 1 investigates whether values of *ASVI*, *ATV* and *volatility* correlate with abnormal stock returns:

Model 1 Static Model for Abnormal Stock Returns

$$AR_{it} = \lambda_t + \propto_i + \beta_1 ASVI\_n_{it} + \beta_2 Volatility_{it} + \beta_3 ATV_{it} + u_{it},$$

where AR_{it} is abnormal stock returns for company i at time t, λ_t is time fixed effects, $\propto_i$ is entity fixed effects, $ASVI\_n_{it}$ can be either *ASVI_ticker* or *ASVI_name* and u_{it} is the error term.

Model 2 investigates whether values of *ASVI*, *AR* and *volatility* correlate with abnormal trading volume:

Model 2 Static Model for Abnormal Trading Volume

$$ATV_{it} = \lambda_t + \propto_i + \beta_1 ASVI\_n_{it} + \beta_2 Volatility_{it} + \beta_3 AR_{it} + u_{it},$$

where ATV_{it} is abnormal trading volume for company i at time t, λ_t is time fixed effects, $\propto_i$ is entity fixed effects, $ASVI\_n_{it}$ can be either *ASVI_ticker* or *ASVI_name* and u_{it} is the error term.

Model 3 investigates whether values of *ASVI*, *ATV* and *AR* correlate with volatility:

Model 3 Static Model for Volatility

$$Volatility_{it} = \lambda_t + \propto_i + \beta_1 ASVI\_n_{it} + \beta_2 ATV_{it} + \beta_3 AR_{it} + u_{it},$$

where $Volatility_{it}$ is the return volatility for company i at time t, λ_t is time fixed effects, $\propto_i$ is entity fixed effects, $ASVI\_n_{it}$ can be either *ASVI_ticker* or *ASVI_name* and u_{it} is the error term.

4.3 Dynamic Model Regressions

A dynamic model is also developed in order to investigate if previous values of ASVI can predict current values of stock returns, trading volume and volatility. In the dynamic models, past values (lags) of abnormal Google search volume, volatility, volume and stock returns are included, as they can correlate with future values of the dependent variable. In these regressions only lagged variables are included as explanatory variables as in Kim et al. (2019).

Model 4 investigates whether previous values of *ASVI*, *ATV* and *volatility* correlate with future values of abnormal stock returns:

Model 4 Dynamic Model for Abnormal Stock Returns

$$AR_{it} = \propto_i + \beta_1 AR_{i,t-1} + \beta_2 ASVI\_n_{i,t-1} + \beta_3 Volatility_{i,t-1} + \beta_4 ATV_{i,t-1} + u_{it}.$$

Model 5 investigates whether previous values of *ASVI*, *AR* and *volatility* correlate with future values of abnormal trading volume:

Model 5 Dynamic Model for Abnormal Trading Volume

$$ATV_{it} = \propto_i + \beta_1 ATV_{i,t-1} + \beta_2 ASVI\_n_{i,t-1} + \beta_3 Volatility_{i,t-1} + \beta_4 AR_{i,t-1} + u_{it}.$$

Model 6 investigates whether previous values of *ASVI*, *ATV* and *AR* correlate with future values of volatility:

Model 6 Dynamic Model for Volatility

$$Volatility_{it} = \propto_i + \beta_1 Volatility_{i,t-1} + \beta_2 ASVI\_n_{i,t-1} + \beta_3 ATV_{i,t-1} + \beta_4 AR_{i,t-1} + u_{it}.$$

5. Analysis Results

Panel data are employed with both fixed and random effects. Results from the Hausman test supports the use of entity fixed effects in all cases. [16] Entity fixed effects are therefore used in all the results presented below. This section will first present descriptive statistics in Section 5.1. Then the regression results for the static model for the weekly data in Section 5.2, followed by static model regression results for the daily data in Section 5.3. Lastly, Section 5.4 will present the dynamic model regression results for both the weekly and daily data.

5.1 Descriptive Statistics

Table 4 reports the summary statistics for the weekly data and Table 5 reports the summary statistics of the daily data. Table 6 and Table 7 provides the correlation matrix of the weekly and daily data respectively. The smaller sample size for the daily data is due to Google Trends download options for different time intervals as discussed in Section 3.2.

Table 4

Descriptive Statistics for all Variables – Weekly Data

Variable	N	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
AR	7,072	-.0001905	.0521923	-.8088319	.4513049	-1.016851	29.26772
ASVI_ticker	7,072	.0322341	1.082495	-3.815694	11.25487	1.436142	8.965146
ASVI_name	7,072	.0253463	1.121121	-3.912848	14.28353	2.023699	13.39268
ATV	7,072	.0410396	1.515649	-3.651698	53.87887	11.77373	312.0656
Volatility	7,072	.0211992	.0168105	0	.4081481	7.482244	127.9399

[16] The Hausman test runs a fixed effect model against a random effects model in order to check if the fixed effect model gives more consistent results (Mehmetoglu and Jakobsen, 2017).

Table 5

Descriptive Statistics for all Variables – Daily Data

Variable	N	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
AR	4,352	-.0004563	.02925	-.82224	.2152389	-6.56798	180.8399
ASVI_ticker	4,352	.1135501	1.090245	-3.296407	10.35055	2.267076	11.75883
ASVI_name	4,352	.1368864	1.009438	-2.798184	9.228454	2.120577	11.57386
ATV	4,352	.1889104	3.511629	-2.816477	188.2672	37.44277	1917.518
Volatility	4,352	.0008445	.0113029	0	.6722819	52.86147	2989.729

Comparing Table 6 and Table 7, we can see that ASVI has a small positive correlation with abnormal returns both in the daily and the weekly dataset. We can also see that *ASVI_ticker* and *ASVI_name* correlate less with abnormal returns, abnormal trading volume, and volatility in the daily data compared to the weekly data. ASVI is also positively correlated with abnormal trading volume, more so in the weekly dataset than in the daily. Differences in correlation between ASVI and the other variables in the weekly and daily data may be due to difference in sample size and time span. Some companies have an identical name and ticker symbol. Because of this, the correlation between *ASVI_name* and *ASVI_ticker* will be higher for these companies.

Table 6

Correlation Matrix for all Variables – Weekly Data

	AR	ASVI_ticker	ASVI_name	ATV	Volatility
AR	1.0000				
ASVI_ticker	0.0327	1.0000			
ASVI_name	0.0326	0.2357	1.0000		
ATV	0.0177	0.1456	0.1831	1.0000	
Volatility	-0.0263	0.0413	0.0625	0.1418	1.0000

Table 7

Correlation Matrix for all Variables – Daily Data

	AR	ASVI_ticker	ASVI_name	ATV	Volatility
AR	1.0000				
ASVI_ticker	0.0239	1.0000			
ASVI_name	0.0197	0.1120	1.0000		
ATV	0.0143	0.0460	0.0864	1.0000	
Volatility	-0.5020	0,0306	0,0046	0,0323	1.0000

5.2 Regression Results Static Model – Weekly Data

5.2.1 Abnormal Stock Return as Dependent Variable

In the first regression, with abnormal stock returns as the dependent variable, none of the coefficients are significant. Results of the regression can be found in Appendix H. ASVI for ticker and company name are insignificant in both the univariate and multivariate regression. The R^2 exhibits low values in all regressions with abnormal stock returns as the dependent variable. The independent variables included in the regression explain little variation of the movements in abnormal stock returns. These results are contrary to the findings of Da et al. (2011), but the results resemble findings by Kim et al. (2019) for the Norwegian stock market. Intuitively this makes sense because Google search frequency and trading volume are likely to be high whether the prospects of future earnings are good or bad. The lack of significance and correlation could be due to the lower sample size of these two studies or other market dynamics for the Scandinavian stock markets. To test the robustness of these results, two additional panel data regressions are conducted, where abnormal returns are defined differently. The second regression defines abnormal returns as the log of company return minus company beta multiplied by log market return: $log\ (R_i) - \beta(log(R_m))$ and in the last regression, abnormal returns is defined as the difference between the log company return and log market return: $log(R_i)\ -\ log(R_m)$. Where the market return is the return of the OMXC20 index. However, none of the definitions of abnormal returns correlates on a statistically significant level by ASVI for company ticker or ASVI for the company name. The last regression has larger R^2 values than the first two regressions. This might be because actual stock returns are used to calculate abnormal returns, rather than expected stock returns. The results from all three regressions can be found in Appendix I.

5.2.2 Abnormal Trading Volume as Dependent Variable

Results from the regression analyses where abnormal trading volume is the dependent variable are presented in Table 8. ASVI for ticker and company name are statistically significant at a 0,1% level in both the univariate and multivariate regression. This resembles the results of Barber and Odeon (2008), where they found that investors are net buyers of attention-grabbing stocks. Stocks with high search volume may in turn have higher trading volume. Volatility is also statistically significant in both the univariate and multivariate regressions. However, abnormal returns do not correlate with abnormal trading volume. The R^2 also exhibits low values. The results are like that of Kim et al. (2019), apart from a lower R^2. This is because Kim et al. (2019) have included previous lags of the dependent variable, that account for most of the variation in the dependent variable.

Table 8

Static Model Regression Results for Weekly Abnormal Trading Volume

	Dependent variable: *Abnormal trading volume*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ASVI_ticker	0.206^{***}				0.194^{***}		0.180^{***}
	(0.0396)				(0.0381)		(0.0385)
ASVI_name		0.252^{***}				0.237^{***}	
		(0.0428)				(0.0395)	
AR			0.595		0.545	0.505	0.707
			(1.096)		(1.129)	(1.075)	(1.207)
Volatility				15.64^{***}	14.88^{***}	14.37^{***}	15.10^{***}
				(2.860)	(2.779)	(2.869)	(2.852)
N	7072	7072	7072	7072	7072	7072	7072
Entity Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	NO	NO	NO	NO	NO	YES
R^2	0.022	0.035	0.000	0.024	0.043	0.055	0.099
adj. R^2	0.021	0.034	0.000	0.024	0.043	0.054	0.071
rho	0.00729	0.00794	0.00698	0.0117	0.0112	0.0118	0.0117

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.2.3 Volatility as Dependent Variable

Regression results where volatility is the dependent variable are presented in Table 9. *ASVI_ticker* and *ASVI_name* are significant at the 5% level in the univariate regressions. However, they become insignificant in the multivariate regression. This is the same result as obtained by Kim et al. (2019). ATV remains statistically significant in the multivariate regressions. When the regression allows for time fixed effects, the ASVI for company ticker is statistically significant at a 5% level again. The adjusted R^2 is 10,6% in model (7). Most of the variation in volatility can be attributed to time fixed effects. Once again, the R^2 for the regressions are lower than for Kim et al. (2019) because the first lag of the dependent variable is not included in this regression.

Table 9

Static Model Regression Results for Weekly Volatility

	Dependent variable: *Volatility*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ASVI_ticker	0.000783*				0.000490		0.000596*
	(0.000312)				(0.000247)		(0.000234)
ASVI_name		0.000976*				0.000625	
		(0.000362)				(0.000337)	
ATV			0.00152***		0.00147***	0.00144**	0.00141***
			(0.000403)		(0.000388)	(0.000403)	(0.000374)
AR				-0.00521	-0.00641	-0.00650	-0.00853
				(0.00769)	(0.00861)	(0.00868)	(0.00810)
N	7072	7072	7072	7072	7072	7072	7072
Entity Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	NO	NO	NO	NO	NO	YES
R^2	0.003	0.005	0.024	0.000	0.025	0.026	0.132
adj. R^2	0.003	0.005	0.024	0.000	0.025	0.026	0.106
rho	0.222	0.222	0.224	0.221	0.225	0.224	0.240

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.3 Regression Results Static Model – Daily Data

The regression results from the daily data differ in part from the weekly regression results. There are less statistically significant beta coefficients and the regressions exhibit lower values of R^2. The regression results without the exclusion of the company SPEAS, which exhibit extreme and influential values, are available in Appendix J.

5.3.1 Abnormal Stock Returns as Dependent Variable

Table 10 shows that the betas for *ASVI_ticker* and *ASVI_name* are not statistically significant in the univariate regressions. *ASVI_ ticker* becomes statistically significant in the multivariate regression (7), with both time and entity fixed effects. This is likely to be caused by a confounding variable that affects *ASVI_ticker* and that varies over time. When holding time effects constant, the coefficient for *ASVI_ticker* becomes statistically significant. Like Bijl et al. (2016), the correlation between ASVI and abnormal returns in the daily data is not statistically significant. The regressions also exhibit low values of R^2.

Table 10

Static model Regression Results for Daily Abnormal Returns

	Dependent variable: *Abnormal returns*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ASVI_ticker	0.00110				0.00108		0.00117^{*}
	(0.000615)				(0.000570)		(0.000541)
ASVI_name		0.000503				0.000457	
		(0.000556)				(0.000515)	
ATV			0.000125		0.000102	0.000104	0.0000895
			(0.000366)		(0.000312)	(0.000314)	(0.000298)
Volatility				0.187	0.0680	0.0921	0.0958
				(1.018)	(0.916)	(0.915)	(0.911)
N	4224	4224	4224	4224	4224	4224	4224
Entity Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	NO	NO	NO	NO	NO	YES
R^2	0.002	0.000	0.000	0.000	0.003	0.001	0.061
adj. R^2	0.002	0.000	0.000	0.000	0.002	0.000	0.031
rho	0.00635	0.00608	0.00618	0.00613	0.00637	0.00611	0.00657

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.3.2 Abnormal Trading Volume as Dependent Variable

Like the weekly regressions on abnormal returns, Table 11 shows that *ASVI_name* and *volatility* are statistically significant, both in the univariate and in the multivariate regression. Contrary to the weekly, the daily data for *ASVI_ticker* is not statistically significant, nor is abnormal returns. If we compare these results to the results from a regression without the exclusions of a largely influential observation, this regression has a larger R^2 and resemble the results of the weekly regression model, apart from *ASVI_ticker*, that is not statistically significant.

Table 11

Static Model Regression Results for Daily Abnormal Trading Volume

	Dependent variable: *Abnormal trading volume*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ASVI_ticker	0.154				0.0968		0.0829
	(0.0808)				(0.0588)		(0.0586)
ASVI_name		0.327^{***}				0.241^{**}	
		(0.0738)				(0.0817)	
AR			2.552		1.953	1.973	1.753
			(7.413)		(6.120)	(6.085)	(5.947)
Volatility				625.0^{**}	619.6^{**}	608.6^{*}	617.3^{**}
				(224.0)	(222.3)	(223.1)	(213.6)
N	4224	4224	4224	4224	4224	4224	4224
Entity Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	NO	NO	NO	NO	NO	YES
R^2	0.002	0.008	0.000	0.064	0.066	0.069	0.100
adj. R^2	0.002	0.008	0.000	0.064	0.065	0.069	0.071
rho	0.00667	0.00706	0.00664	0.0134	0.0132	0.0125	0.0132

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.3.3 Volatility as Dependent Variable

Table 12 below presents the regression results for daily volatility. The beta coefficient for *ASVI_name* is statistically significant at a 0,1% level in both the univariate and multivariate regression. *ASVI_ticker* is statistically significant at a 5% level. The R^2 value increases when abnormal trading volume is included as an independent variable. However, abnormal trading volume is not statistically significant contrary to the weekly regression results. In the regression with the highly influential observation in Appendix J, beta coefficients for both *ASVI_ticker* and *ASVI_name* are insignificant in all the regressions.

Table 12

Static Model Regression Results for Daily Volatility

	Dependent variable: *Volatility*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ASVI_ticker	0.0000888^{*}				0.0000729^{*}		0.0000787^{*}
	(0.0000343)				(0.0000278)		(0.0000305)
ASVI_name		0.000139^{***}				0.000106^{***}	
		(0.0000320)				(0.0000286)	
ATV			0.000103		0.000102	0.000100	0.000102
			(0.0000738)		(0.0000729)	(0.0000729)	(0.0000714)
AR				0.000631	0.000214	0.000289	0.000310
				(0.00341)	(0.00288)	(0.00286)	(0.00293)
N	4224	4224	4224	4224	4224	4224	4224
Entity Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	NO	NO	NO	NO	NO	YES
R^2	0.005	0.009	0.064	0.000	0.068	0.070	0.099
adj. R^2	0.004	0.009	0.064	0.000	0.067	0.069	0.071
rho	0.115	0.118	0.119	0.113	0.120	0.122	0.120

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.4 Regression Results from Dynamic Models

The dynamic models are arranged as follows: (1) to (4) use weekly data while (5) and (6) use daily data. The regression results from the dynamic regressions for weekly abnormal returns are shown in Table 13. The results show that the first lag of *ASVI_name* is statistically significant at a 5% level. This is contrary to the findings of Kim et al. (2019) in their dynamic regression of ASVI based on the company name.

Table 13

Dynamic Regression Results for Abnormal Returns

	Dependent variable: *Abnormal returns*							
	Weekly				Daily			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AR_{t-1}	-0.0217	-0.0218	-0.0212	-0.0213	-0.0556	-0.0549	-0.0764*	-0.0753*
	(0.0205)	(0.0205)	(0.0206)	(0.0206)	(0.0323)	(0.0320)	(0.0332)	(0.0331)
$ASVI\_ticker_{t-1}$	0.00106		0.00114		0.000475		0.000530	
	(0.000613)		(0.000623)		(0.000383)		(0.000386)	
$ASVI\_name_{t-1}$		0.00119*		0.00131*		-0.000272		-0.000193
		(0.000512)		(0.000561)		(0.000453)		(0.000476)
ATV_{t-1}			-0.000558	-0.000616			-0.000185	-0.000175
			(0.000515)	(0.000540)			(0.000252)	(0.000250)
$Volatility_{t-1}$			0.0334	0.0318			-0.103**	-0.102*
			(0.0446)	(0.0450)			(0.0378)	(0.0377)
N	7038	7038	7038	7038	3400	3400	3400	3400
R^2	0.001	0.001	0.001	0.001	0.003	0.003	0.005	0.004
adj. R^2	0.001	0.001	0.001	0.001	0.002	0.002	0.004	0.003
rho	0.00527	0.00531	0.00545	0.00547	0.0196	0.0196	0.0193	0.0194

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 14 present the dynamic regression results for the abnormal trading volume. In the weekly regression for abnormal trading volume, the first lag of *ASVI_ticker* and *ASVI_name* are statistically significant. For the daily regression however, beta coefficients for *ASVI_ticker* and *ASVI_name* are not statistically significant in any of the regressions.

Table 14

Dynamic Regression Results for Abnormal Trading Volume

	Dependent variable: *Abnormal trading volume*							
	Weekly				Daily			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
ATV_{t-1}	0.196***	0.187***		0.183***	0.0286	0.0293		0.0290
	(0.0414)	(0.0423)		(0.0417)	(0.0312)	(0.0300)		(0.0301)
$ASVI\_ticker_{t-1}$	0.0267		0.0619**		0.0430		0.0444	
	(0.0176)		(0.0199)		(0.0862)		(0.0797)	
$ASVI\_name_{t-1}$		0.0894*		0.0874*		-0.00838		-0.00979
		(0.0337)		(0.0329)		(0.0480)		(0.0485)
AR_{t-1}			0.653	0.511			2.693	2.733
			(0.527)	(0.423)			(1.603)	(1.739)
$Volatility_{t-1}$			5.237***	2.145			4.971*	4.757
			(1.296)	(1.178)			(2.171)	(2.462)
N	7038	7038	7038	7038	3400	3400	3400	3400
R^2	0.040	0.044	0.005	0.044	0.001	0.001	0.000	0.001
adj. R^2	0.040	0.043	0.005	0.044	0.000	0.000	-0.000	-0.000
rho	0.00461	0.00491	0.00700	0.00491	0.00725	0.00724	0.00774	0.00733

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The dynamic regression results for volatility are presented in Table 15. These results show that the first lag of *ASVI_ticker* and *ASVI_name* are statistically significant for the weekly data, but not in the daily data. The first lag of volatility, abnormal trading volume, and abnormal returns are also statistically significant in all the weekly dynamic regressions. The R^2 is also much larger for the weekly data compared to the daily data.

Table 15

Dynamic Regression Results for Volatility

	Dependent variable: *Volatility*							
	Weekly				Daily			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Volatility_{t-1}$	0.358^{***}	0.356^{***}	0.313^{***}	0.312^{***}	-0.0132^{***}	-0.0132^{***}	-0.00337	-0.00336
	(0.0682)	(0.0700)	(0.0640)	(0.0651)	(0.00178)	(0.00179)	(0.00843)	(0.00842)
$ASVI\_ticker_{t-1}$	0.00142^{**}		0.000962^{*}		-0.0000209		-0.0000479	
	(0.000405)		(0.000359)		(0.0000238)		(0.0000437)	
$ASVI\_name_{t-1}$		0.00161^{**}		0.00104^{*}		-0.0000567		-0.000105
		(0.000524)		(0.000413)		(0.0000600)		(0.000101)
ATV_{t-1}			0.00279^{***}	0.00275^{***}			0.000125	0.000127
			(0.000646)	(0.000614)			(0.000139)	(0.000141)
AR_{t-1}			-0.0501^{*}	-0.0501^{*}			0.00935	0.00935
			(0.0230)	(0.0230)			(0.00791)	(0.00791)
N	7038	7038	7038	7038	3400	3400	3400	3400
R^2	0.143	0.147	0.249	0.250	0.000	0.000	0.002	0.002
adj. R^2	0.143	0.147	0.248	0.249	-0.000	-0.000	0.001	0.001
rho	0.120	0.121	0.147	0.147	0.0171	0.0171	0.0173	0.0173

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6. Trading Strategy

In this section, the possibilities of generating abnormal returns through different trading strategies based on Google search volume are studied. First, a trading strategy inspired by Joseph et al. (2011) is studied. Secondly, a Tan and Taş (2019) inspired strategy is examined and lastly, a trading strategy inspired by Bollinger bands (Bollinger, 2001) is examined. All graphs present the abnormal returns in decimals calculated using log returns and Fama and French three-factors (Fama and French, 1993). ASVI is based on company ticker in all cases. The difference in cumulative returns is due to differences in the time interval for the two datasets. The weekly data span over four years, whereas the daily data span over six months.

6.1 Joseph et al. (2011)

By following the approach of Joseph et al. (2011), companies are divided into five quantiles (Q) based on previous week ASVI. Q1 contains firms with the lowest ASVI, Q5 contains firms with the highest values of ASVI. Without taking transaction costs and the ability to arbitrage into consideration, Figure 2 shows that high ASVI yields higher abnormal returns and low ASVI yields negative abnormal returns. These findings are in line with the findings of other studies (Tan and Taş, 2019; Bank et al., 2019; Bijl et al., 2016). However, the abnormal returns from quantile five are lower than the abnormal returns from quantile four. This can be related to the reasoning of Moat et al. (2013): stocks with high search volumes are considered speculative and of high risk. The expected returns will be higher, resulting in lower abnormal returns.

Weekly ASVI Daily ASVI

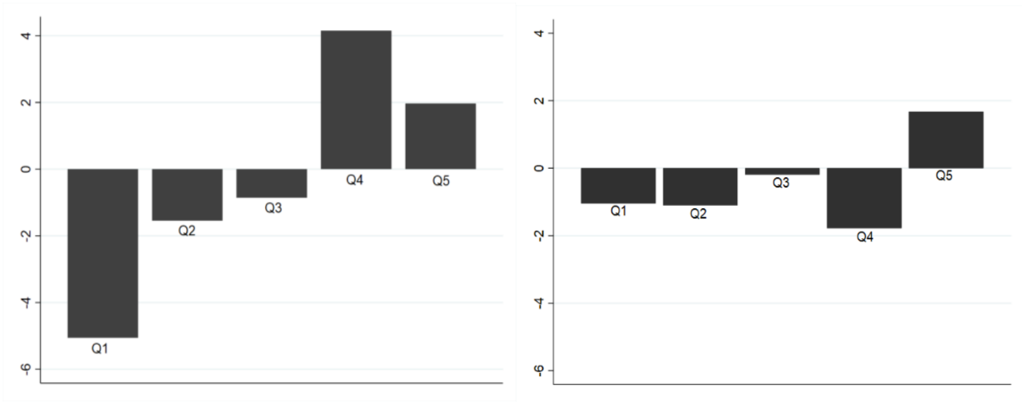


Figure 2 Cumulative Abnormal Returns from Joseph et al. (2019) Trading Strategy

6.2 Tan and Taş (2019)

A trading strategy suggested by Tan and Taş (2019) yields the highest cumulative abnormal returns of the different long position's strategies in this study. They suggested: going long in the high attention stocks that are above the 70th percentile and short the low attention stocks that are below the 30th percentile. Figure 3 shows cumulative abnormal returns when going long in the stocks, which is why the cumulative abnormal returns in the 30th percentile stocks are negative in the graph. However, shorting the lower 30th percentile yields positive abnormal returns.

Weekly ASVI Daily ASVI

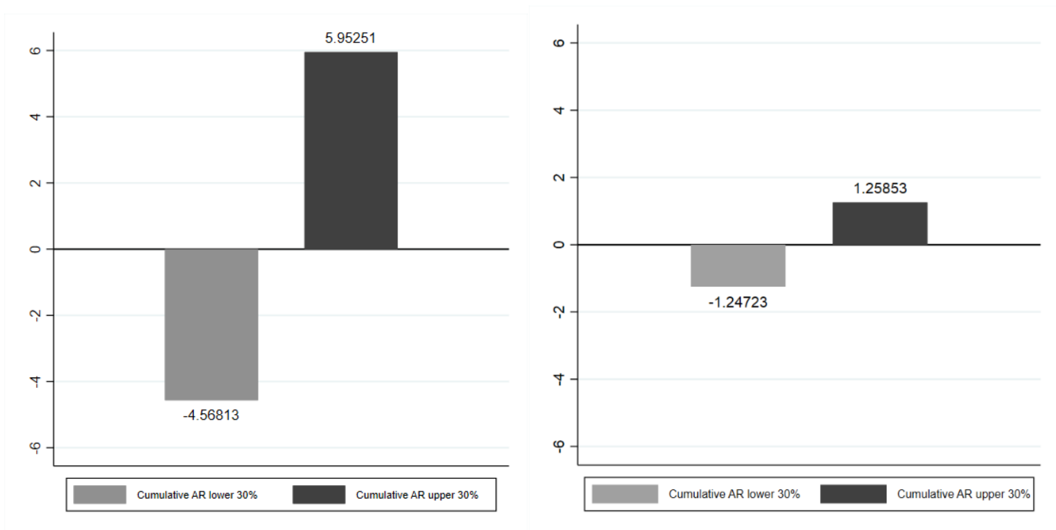


Figure 3 Cumulative Abnormal Returns from Tan and Taş (2019) Trading Strategy

6.3 Bollinger Bands Using ASVI

The following section investigates how the moving averages of ASVI perform as an instrument to detect extreme values of ASVI. The strategy is based on a Bollinger bands (Bollinger, 2001) trading strategy. Bollinger bands is a simple technical analysis tool developed by John Bollinger (Bollinger, 2001). The method consists of three bands around the stock price: a moving average, a lower band and an upper band. A simple moving average (SMA) is the average of the stock price for the last x days chosen by the investor. The upper and lower band are one standard deviation from the SMA. The bands expand with increased volatility and contract when volatility is low (Bollinger, 2001). It is used to define the trend in a stock (Grøtte, 2006). In a positive trend the stock price will most often be in the upper part of the band. That is, between the SMA and the upper band. The opposite applies to a negative trend (Grøtte, 2006). SMA is often the supporting line in a bull trend and the resistance point in a bear trend. Say you use a 20-day SMA. If the stock price exceeds the 20-day SMA, it can be used as an indication that the stock is in a rising trend, the opposite applies if the stock price is below the 20-day SMA. This analysis tool is very useful for options traders, as the value of an option increase with increased volatility (Brealey, Myers and Allen, 2017). Utilizing the Bollinger bands method in trading based on ASVI, results in the cumulative abnormal returns that is presented in Figure 4.

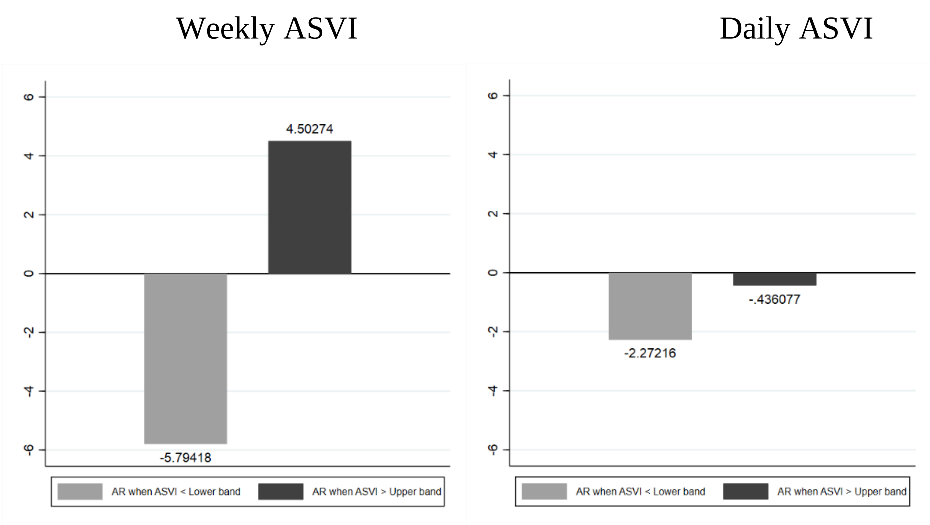


Figure 4 Cumulative Abnormal Returns from Bollinger Bands (2001) Trading Strategy

A simplified version of the Bollinger bands strategy would be to expect the mean of ASVI to be 0. If we set a threshold for buy/ short: buy whenever ASVI exceeds 1 and sell when ASVI is -1. This strategy yields the cumulative abnormal returns presented in Figure 5.

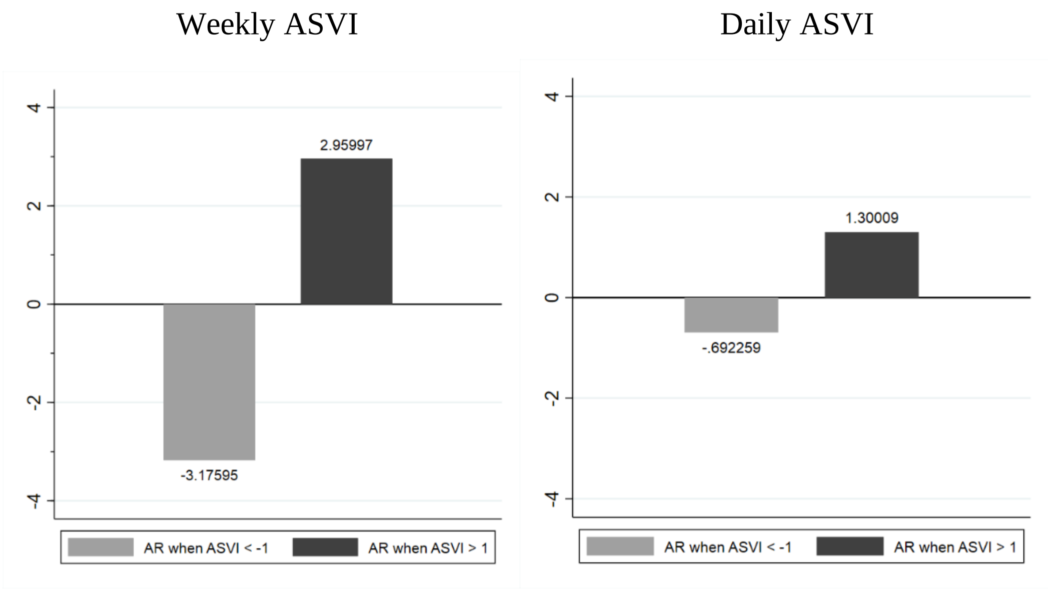


Figure 5 Cumulative Abnormal Returns Based on ASVI Threshold for Weekly and Daily Data

Based on the numbers and graphical analysis presented in this chapter, abnormal returns could be achieved by following simple trading strategies based on ASVI for the sample of stocks included in this study. However, the daily data are once again less clear. This is because of the short time period used for the daily data. The OMXC20 is the Nasdaq Copenhagen leading share index and consists of the 20 most actively traded firms on the exchange.[17] The index yields a cumulative return of 11,33% in the four-year time period 2016-2019. Holding shares in companies with high ASVI for company ticker previous week can generate abnormal returns that exceed the return of simply holding the OMXC20 index as graphed in this chapter. From Chapter 3 we know that Nasdaq Copenhagen is a relatively small stock exchange and the stocks offered for shorting are limited and vary. This puts constraints on arbitrage and trading

[17] OMXC20 information collected from https://indexes.nasdaqomx.com/Index/Overview/OMXC20

strategies explained in this section. Liquidity in stocks may also vary, which in turn also put restrictions on entering long and short positions. Brokerage fees and other costs related to the rebalancing of portfolios are not considered and will weaken the net profitability of these trading strategies considerably. Descriptive statistics of the trading strategies can be found in Appendix K and Appendix L.

7. Conclusion

This thesis has examined the relationship between investor attention and the stock market activity on the Nasdaq Copenhagen. Investor attention is hard to measure, and Internet search volume has been used as a proxy for investor attention in recent research. Da et al. (2011) proposed to use the Google search volume index as a new and direct measure for investor attention. This thesis contributes to the literature by studying Google search volume on the Nasdaq Copenhagen, which is the international stock exchange for Danish securities. Previous literature on American stock markets shows that high values of the Google search volume index correlates with or predict future abnormal returns (Da et al., 2011). A study by Kim et al. (2019) does not find the same results for the Norwegian market. However, their findings indicate that ASVI is more related to the future-, rather than the current trading activity. This analysis includes weekly data over a four-year period from 2016 to 2019, and daily data over a six-month period, from July 2019 to December 2019. 34 companies listed at Nasdaq Copenhagen are included in the study. The low sample size limits the external validity of the findings. Per contra, the sample size does not compromise internal validity, and the insights this study provides for the stocks in question. Google search volume index for company ticker and company name have been used as proxies for investor attention. The analysis uses static and dynamic panel data regressions with company- and time fixed effects. When investigating the first hypothesis, a weak positive correlation between abnormal stock returns and ASVI is detected. In a panel data regression for both weekly and daily ASVI, this study found little evidence of a link between abnormal stock returns and ASVI at a statistically significant level. However, there is evidence in support of the second hypothesis. ASVI for company ticker and ASVI for company name correlates with trading activity, such as trading volume and volatility. When comparing the weekly and daily data, I also find evidence in support of the third hypothesis. In the dynamic regressions, previous values of ASVI correlates with future values of the market parameters being studied in the weekly data, while no such evidence is found for the daily data. The differences between the daily and the weekly data may be due to a large difference in sample size, differences in time span, and because the daily data may be exposed to more noise when standardized into the Search Volume Index by Google. Finally, I also find evidence in support of the fourth hypothesis. Utilizing simple trading strategies inspired by Joseph et al. (2011), Tan and Taş (2019) and Bollinger (2001), can generate abnormal returns when liquidity issues, transaction- and rebalancing costs are not considered.

The first suggestion for further research is to try to find other ways to mitigate the endogeneity problem. Lack of statistically significant betas is likely to be due to shortcomings of utilizing panel data regression to examine causal effects in the stock market. As mentioned in Section 4.1, there is a strict exogeneity assumption. This assumption holds if the independent variables in the regression are uncorrelated with the error term. This assumption is not likely to hold when studying stock markets. Because of the complexity of financial markets, regressions are likely to suffer from omitted variable bias, simultaneous causality, and other problems of endogeneity. One way to mitigate the endogeneity problem is through a natural experiment as described as a state-of-the-art solution by Gippel et al. (2015). Another suggestion is to use the common correlated effects estimator of Pesaran (2015), that allows for cross-sectional dependence, heterogenous slopes, fixed effects, and endogenous regressors.

The second suggestion for further research is to create an index for attention which can differentiate the positive and negative attention. Google search volume is just one of many factors that may reflect an investors' attention to a stock. Investors' buying behavior is complex and composed of many factors at the individual, micro- and macro levels. It is a high probability of spuriousness in any model of investor attention. A principal component analysis will be able to determine if any common dimensions measure the same, so we can merge them into an index and thereby reduce data.[18] That is to collect a set of indicators that explain the phenomenon (investor attention). By making a scale or an index, it is possible to measure a latent variable,[19] such as investor attention. E.g. an index that measures investor attention is created in a study by Baker and Wurgler (2007). ASVI only captures investor attention, mixing the good and the bad. ASVI does not differentiate whether investors have a positive or negative opinion about the company's financial development. The investor sentiment during stock trading may be crucial for the following returns. If the sentiment inherent in the search queries can be quantified through an index, this could enhance the model and the interpretation of ASVI.

[18] Principal component analysis is based on correlations between a set of observed variables. The analysis suppose observed variables are affected by one or more, general underlying variables. The analysis investigates whether the observed variables can be used as indicators for the more general factor (Tufte, 2018).
[19] A latent variable is an unmeasured or not-directly observed variable (Mehmetoglu and Jakobsen, 2017).

The third suggestion for further research is to study ASVI in relation to earnings announcements, mergers and acquisitions and initial public offerings information. Search volume may be an indicator of knowledge spread by informed insiders. Evidence of insider trading activity has been documented, among others, by Keown and Pinkerton (1981) in takeover attempts and Rendleman Jr., Jones and Latané (1982) in earnings announcements. If investors with inside information conduct Google searches, the Google SVI could contain information that something is going to happen in a company. If ASVI can be an indicator of such changes in structure, investors can use this information in stock trading to generate abnormal returns.

The last suggestion for further research is to investigate trading strategies utilizing options and futures based on search volume activity. This thesis investigates the possibilities of utilizing ASVI in stock market trading. Volatility and trading volume are factors that are to a greater extent linked to return opportunities in the options- and futures market. If internet search queries can predict future volatility or trading volume, this can create investment opportunities in the options- and futures market that can generate abnormal returns. If we assume investors who are willing to buy or perceive a stock as high risk will investigate the company through search queries before making an investment decision in the stock market, we can detect early signs of higher volatility that can be utilized in an options- or futures trading strategy.

References

Ackert, L. F., Jiang, L., Lee, H. S., & Liu, J. (2016). Influential investors in online stock forums. *International Review of Financial Analysis*, *45*, 39-46. Doi: https://doi.org/10.1016/j.irfa.2016.02.001

Aouadi, A., Arouri, M., & Teulon, F. (2013) Investor attention and stock martket activity: Evidence from France. *Economic Modelling*, *35*, 674-681. Doi: https://doi-org.ezproxy.hioa.no/10.1016/j.econmod.2013.08.034

Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*, *21*(2), 129-151.

Bank, M., Larch, M., & Peter, G. (2011). Google searches and its influence on liquidity and returns of German stocks. *Financial Markets and Portfolio Management*, *25*, 239.

Barber, B. M., & Odean, T. (2008). All That Glitters: The Effect of Attention and News on the Buying Behavior of Individual and Institutional Investors. *The Review of Financial Studies*, *21*(2), 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), 25-57. Doi: https://doi.org/10.2308/accr-51865

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

Bodie, Z., Kane, A., & Marcus, A. J. (2014). *Investments* – 10th Global Edition. New York: McGraw-Hill Education.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*, 1-8. Doi: https://doi.org/10.1016/j.jocs.2010.12.007

Bollinger, J. A. (2001). *Bollinger on Bollinger Bands.* New York: McGraw-Hill Professional.

Brealey, R. A., Myers, S. C., & Allen, F. (2017). *Principles of Corporate Finance* 12th ed. New York: McGraw-Hill Education.

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroskedasticity and random coefficient variation. *Econometrica*, *47*, 1287-1294.

Busse, J. A., & Green, T. C. (2002). “Market Efficiency in Real Time”. *Journal of Financial Economics*, *65*, 422.

Choi, H., & Varian, H., (2012). Predicting the present with Google Trends. *Economic Record*, *88*(s1), 2-9.

Cook, R. D., & Weisberg, S. (1983). Diagnostics for heteroskedasticity in regression. *Biometrika*, *70*, 1-10.

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

DellaVigna, S., & Pollett, J. M. (2009). Investor Inattention and Friday Earnings Announcements. *The Journal of Finance*, *64,* 709-749. Doi: https://doi.org/10.1111/j.15406261.2009.01447.x

Devenow, A., & Welch, I. (1996). Rational herding in financial economics. *European Economic Review*, *40*, 3-5, 603-615. Doi: https://doi.org/10.1016/0014-2921(95)00073-9

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2012). Investor Information Demand: Evidence from Google Searches Around Earnings Announcements. *Journal of Accounting Research*, *50*(4), 1001-1040.

Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*(1), 3-56. https://doi.org/10.1016/0304405X(93)90023-5

Fang, L., & Peress, J. (2009). Media Coverage and the Cross-section of Stock Returns. *The Journal of Finance*, *64*(5), 2023-2052. Doi: https://doi-org.ezproxy.hioa.no/10.1111/j.1540-6261.2009.01493.x

French, K.R. (n.d.). *Description of Fama/French 3 Factors for Developed Markets.* Retrieved from http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_3developed.html

French, K. R. (2020). Developed market factors and returns. Retrieved 18th of January, 2020 from https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Garman, M. B., & Klass M. J. (1980). On the Estimation of Security Price Volatilities from Historical Data. *The Journal of Business*, *53*(1), 67-78.

Gippel, J., Smith, T., & Zhu, Y. (2015). Endogeneity in Accounting and Finance Research: Natural Experiments as a State-of-the-Art Solution. *Abacus*, *52*(2), 143-168. Doi: 10.1111/abac.12048

Google, (2020). *FAQ about Google Trends data*. Retrieved January 19th, 2020 from https://support.google.com/trends/answer/4365533?hl=en

Gow, I. D., Ormazabal, G., & Taylor, D. J. (2010). Correcting for cross-sectional and time-series dependence in accounting research. *The Accounting Review*, *85*(2), 483- 512.

Grøtte, O. (2006). *Aksjekjøp og daytrading: Metode, psykologi, risiko og strategier* [*Stock buying and daytrading: Method, psychology, risk and strategies*]. Norway: Hegnar Media AS.

Hume, D. (2011) [1740]. *A Treatise of Human Nature* (D.F. Norton and M.J. Norton, eds). Oxford: Oxford University Press.

Johannessen, A., Christoffersen, L., & Tufte, P.A. (2011). *Forskningsmetode for økonomisk-administrative fag* (3. utg.). Oslo: Abstrakt forlag

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*, 1116-1127.

Kahneman, D. (1973). *Attention and Effort*. Prentice-Hall, Englewood Cliffs, NJ.

Kamstra, M. J., Kramer, L. A., & Levi, M. D. (2003). Winter blues: A SAD stock market cycle. *The American Economic Review*, *93*(1), 324-343.

Keown, A., & Pinkerton, J. (1981). Merger Announcements and Insider Trading Activity. *Journal of Finance*, 36. Doi: https://doi.org/10.1111/j.1540-6261.1981.tb04888.x

Kim, N., Lučivjanská, K., Molnár, P., & Villa, R., (2019). Google Searches and Stock Market Activity: Evidence from Norway. *Finance Research Letters*, *28*, 08-220. https://doi.org/10.1016/j.frl.2018.05.003

Lawrence, A., Ryans, J., Sun, E., & Laptev, N. (2016). Yahoo Finance Search and Earnings Announcements. Available at SSRN: Available at https://ssrn.com/abstract=2804353. Doi: http://dx.doi.org/10.2139/ssrn.2804353

Mangold, B., Dooley, M., Flake, G. W., Hoffman, H., Kasturi, T., Pennock, D. M., & Dornfest, R. (2005). The Tech Buzz Game [stock market prediction]. *Computer*, *38*(7), 94-97. Doi: 10.1109/MC.2005.243

Mehmetoglu, M., & Jakobsen, T. G. (2017). *Applied statistics using Stata: A Guide for the Social Sciences.* Great Britain: SAGE Publications Ltd.

Merton, R. C. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *The Journal of Finance*, *42*(3).

Moat, H. S., Curme, C., Avakian, A., Kenett, D. Y., Stanley, H. E., & Preis, T. (2013). Quantifying *Wikipedia* Usage Patterns Before Stock Market Moves. *Scientific Reports*, *3*, 1801. Doi: https://doi.org/10.1038/srep01801

Molnár, P. (2012) Properties of range-based volatility estimators. *International Review of Financial Analysis*, *23*, 20-29. Retrieved January 28, 2020 from https://doi.org/10.1016/j.irfa.2011.06.012

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometric Reviews*, *34*(6-10), 1089-1117.

Poterba, J. M., & Summers, L. H. (1988). Mean reversion I stock prices: Evidence and Implications. *Journal of Finance*, *22*(1), 27-59. Doi: https://doi.org/10.1016/0304405X(88)90021-9

Preis, T., Moat, H. S., & Stanley, H. E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*, *3*(1). Doi: https://doi.org/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 of London A: Mathematical, Physical and Engineering Sciences*, *368*(1933), 5707-5719. Doi: http://doi.org/10.1098/rsta.2010.0284

Rendleman Jr, R. J., Jones, C. P., & Latané, H. A. (1982). Empirical anomalies based on unexpected earnings and the importance of risk adjustments. *Journal of Financial Economics*, *10*, 3, 269-287. Doi: https://doi.org/10.1016/0304-405X(82)90003-4

Seasholes, M. S., & Wu, G. (2007). Predictable behavior, profits, and attention. *Journal of Empirical Finance*, *14*(5),590-610. Doi: https://doi.org/10.1016/j.jempfin.2007.03.002

Shen, D., Zhang, Y., Xiong, X., & Zhang, W. (2017) Baidu index and predictability of Chinese stock returns. *Financial Innovation*, *3*, 4. Doi: https://doi.org/10.1186/s40854-017-0053-1

Stock, J. H., & Watson, M. W. (2014). *Introduction to econometrics, Update, Global Edition* (3rd ed.). Boston: Pearson Education Limited.

Sustainable Stock Exchanges Initiative. (n.d.). *Nasdaq Copenhagen*. Retrieved April 7th 2020 from https://sseinitiative.org/stock-exchange/nasdaq-copenhagen/#

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*, *50*, 1-17. Doi: 10.1016/j.ribaf.2019.04.010

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

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More Than Words: Quantifying Language to Measure Firms' Fundamentals. *The Journal of Finance*, *63*(3), 1437-1467. Doi: https://doi-org.ezproxy.hioa.no/10.1111/j.1540-6261.2008.01362.x

Tufte, P. A. (2013). Å studere sosiale årsakssammenhenger [To study social causal relationships]. *Sosiologi i dag*, *43*(3), 91-109.

Tufte, P. A. (2018, November 11). Forklaring og kausalitet [Explanation and causality]. [Slides]. Retrieved from https://oslomet.instructure.com/courses/3444/files/487912?module_item_id=64671

Appendices

Appendix A Final Sample and Search Terms

Company name	Search term used - name	Ticker symbol	Sector
SP Group	SP Group	SPG	Basic Materials
Dantax	Dantax	DANT	Consumer Goods
Scandinavian Brake Systems	Scandinavian Brake Systems	SBS	Consumer Goods
United Int. Enterprises	United International Enterprises	UIE	Consumer Goods
AaB	Aalborg Boldspilklub	AAB	Consumer Services
Brøndby IF	Brøndby IF	BIF	Consumer Services
Matas	Matas A/S	MATAS	Consumer Services
SAS	SAS Danmark	SAS DKK	Consumer Services
Silkeborg IF Invest	Silkeborg IF	SIF	Consumer Services
Cemat	Cemat A/S	CEMAT	Financials
Kreditbanken	Kreditbanken	KRE	Financials
Newcap Holding	Newcap Holding	NEWCAP	Financials
Scandinavian Private Equity	Scandinavian Private Equity	SPEAS	Financials
Sydbank	Sydbank	SYDB	Financials
Tryg	Tryg	TRYG	Financials
Bavarian Nordic	Bavarian Nordic	BAVA	Health Care
Chr. Hansen Holding	Chr. Hansen	CHR	Health Care
Demant	Demant	DEMANT	Health Care
Genmab	Genmab	GMAB	Health Care
GN Store Nord	GN Store Nord	GN	Health Care
Novo Nordisk B	Novo Nordisk	NOVO B	Health Care
Onxeo	Onxeo	ONXEO	Health Care
Zealand Pharma	Zealand Pharma	ZEAL	Health Care
A.P. Møller - Mærsk A	A.P. Møller - Mærsk	MAERSK A	Industrials
DLH	Dalhoff Larsen & Horneman	DLH	Industrials
DFDS	DFDS	DFDS	Industrials
FLSmidth & Co.	FLSmidth	FLS	Industrials
G4S plc	G4S	G4S	Industrials
ISS	ISS	ISS A/S	Industrials
NKT	NKT A/S	NKT	Industrials
Schouw & Co.	Schouw	SCHO	Industrials
SKAKO	SKAKO A/S	SKAKO	Industrials
Vestas Wind Systems	Vestas Wind Systems	VWS	Oil & Gas
CBRAIN	cBrain	CBRAIN	Technology

Appendix B Stata commands

The following are user-written Stata commands applied to tests and regressions used in this thesis. The syntax descriptions are given by Stata and options for the commands can be found in the Stata manual.

hettest	Breusch-Pagan/ Cook-Weisberg test for heteroskedasticity.
pwcorr	Display correlation matrix.
summarize	Calculates and displays a variety of univariate summary statistics
vif	Calculates the variance inflation factor (VIFs) for the independent variables specified in a linear regression model.
xtreg	Fits regression models to panel data.
xtserial	Woolridge test for autocorrelation in panel data.

Appendix C Breusch-Pagan / Cook-Weisberg Test for Heteroskedasticity

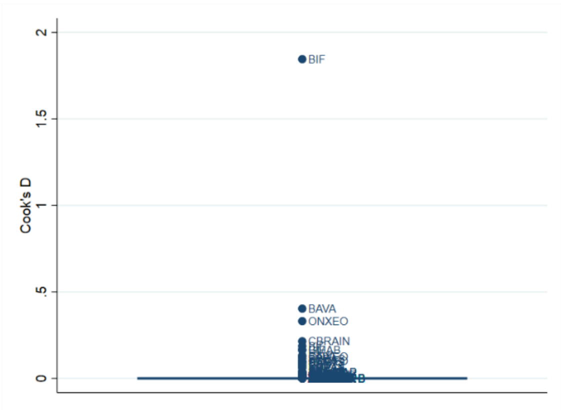
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity		
Regression model	Weekly	Daily
$AR = \beta_1 ASVI\_ticker + \beta_2 Volatility + \beta_3 ATV$	Ho: Constant variance Variables: fitted values of AR chi2(1) = 296.27 Prob > chi2 = 0.0000	Ho: Constant variance Variables: fitted values of AR chi2(1) = 414.18 Prob > chi2 = 0.0000
$AR = \beta_1 ASVI\_name + \beta_2 Volatility + \beta_3 ATV$	Ho: Constant variance Variables: fitted values of AR chi2(1) = 438.05 Prob > chi2 = 0.0000	Ho: Constant variance Variables: fitted values of AR chi2(1) = 405.65 Prob > chi2 = 0.0000
$ATV = \beta_1 ASVI\_ticker + \beta_2 Volatility + \beta_3 AR$	Ho: Constant variance Variables: fitted values of ATV chi2(1) = 2121.61 Prob > chi2 = 0.0000	Ho: Constant variance Variables: fitted values of ATV chi2(1) = 174.46 Prob > chi2 = 0.0000
$ATV = \beta_1 ASVI\_name + \beta_2 Volatility + \beta_3 AR$	Ho: Constant variance Variables: fitted values of ATV chi2(1) = 1398.84 Prob > chi2 = 0.0000	Ho: Constant variance Variables: fitted values of ATV chi2(1) = 3702.57 Prob > chi2 = 0.0000
$Volatility = \beta_1 ASVI\_ticker + \beta_2 AR + \beta_3 ATV$	Ho: Constant variance Variables: fitted values of Volatility chi2(1) = 3897.65 Prob > chi2 = 0.0000	Ho: Constant variance Variables: fitted values of Volatility chi2(1) = 815826.74 Prob > chi2 = 0.0000
$Volatility = \beta_1 ASVI\_name + \beta_2 AR + \beta_3 ATV$	Ho: Constant variance Variables: fitted values of Volatility chi2(1) = 4168.68 Prob > chi2 = 0.0000	Ho: Constant variance Variables: fitted values of Volatility chi2(1) = 816267.50 Prob > chi2 = 0.0000

Appendix D Wooldridge Test for Autocorrelation in Panel

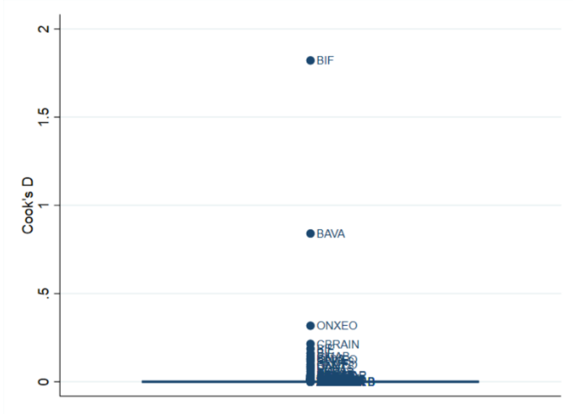
Wooldridge test for autocorrelation in panel		
Regression model	Weekly	Daily
AR $= \beta_1 ASVI\_ticker + \beta_2 Volatility + \beta_3 ATV$	xtserial AR ASVI_ticker ATV Volatility H0: no first order autocorrelation F(1, 33) = 1.180 Prob > F = 0.2853	xtserial AR ASVI_ticker ATV Volatility H0: no first order autocorrelation F(1, 33) = 0.791 Prob > F = 0.3801
AR $= \beta_1 ASVI\_name + \beta_2 Volatility + \beta_3 ATV$	xtserial AR ASVI_name ATV Volatility H0: no first order autocorrelation F(1, 33) = 1.164 Prob > F = 0.2885	xtserial AR ASVI_name ATV Volatility H0: no first order autocorrelation F(1, 33) = 0.725 Prob > F = 0.4005
$ATV = \beta_1 ASVI\_ticker + \beta_2 Volatility + \beta_3 AR$	xtserial ATV ASVI_ticker AR Volatility H0: no first order autocorrelation F(1, 33) = 21.065 Prob > F = 0.0001	xtserial ATV ASVI_ticker AR Volatility H0: no first order autocorrelation F(1, 33) = 8.178 Prob > F = 0.0073
ATV $= \beta_1 ASVI\_name + \beta_2 Volatility + \beta_3 AR$	xtserial ATV ASVI_name AR Volatility H0: no first order autocorrelation F(1, 33) = 22.593 Prob > F = 0.0000	xtserial ATV ASVI_name AR Volatility H0: no first order autocorrelation F(1, 33) = 7.894 Prob > F = 0.0083
$Volatility = \beta_1 ASVI\_ticker + \beta_2 AR + \beta_3 ATV$	xtserial Volatility ASVI_ticker ATV AR H0: no first order autocorrelation F(1, 33) = 18.463 Prob > F = 0.0001	xtserial Volatility ASVI_ticker ATV AR H0: no first order autocorrelation F(1, 33) = 0.044 Prob > F = 0.8355
$Volatility = \beta_1 ASVI\_name + \beta_2 AR + \beta_3 ATV$	xtserial Volatility ASVI_name ATV AR H0: no first order autocorrelation F(1, 33) = 18.254 Prob > F = 0.0002	xtserial Volatility ASVI_name ATV AR H0: no first order autocorrelation F(1, 33) = 0.061 Prob > F = 0.8058

Appendix E Cook's Distance Plots – Weekly Data

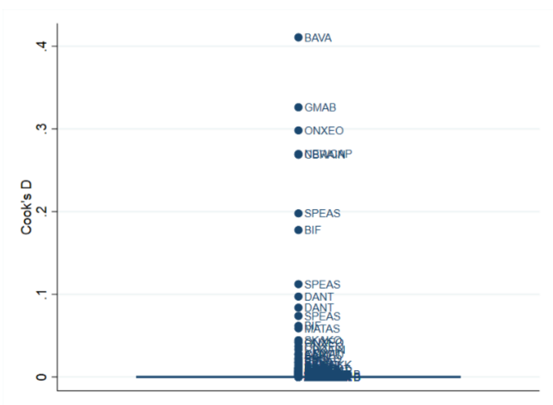
$\text{AR} = \beta_1 ASVI\_ticker + \beta_2 Volatility + \beta_3 ATV$



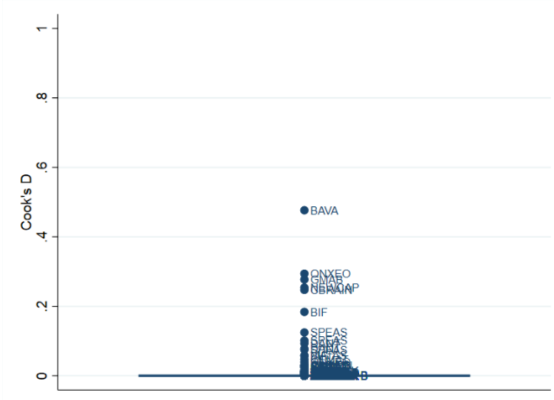
$\text{AR} = \beta_1 ASVI\_name + \beta_2 Volatility + \beta_3 ATV$



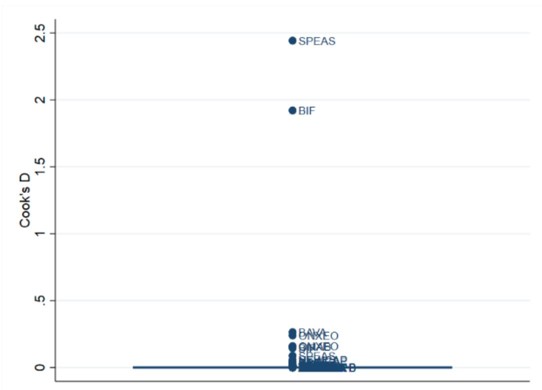
$ATV = \beta_1 ASVI\_ticker + \beta_2 Volatility + \beta_3 AR$



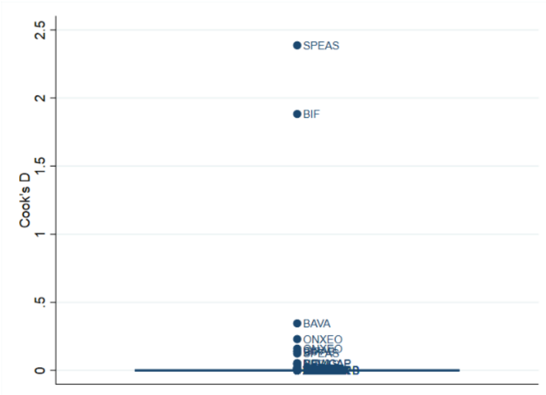
$\text{ATV} = \beta_1 ASVI\_name + \beta_2 Volatility + \beta_3 AR$



$Volatility = \beta_1 ASVI\_ticker + \beta_2 AR + \beta_3 ATV$

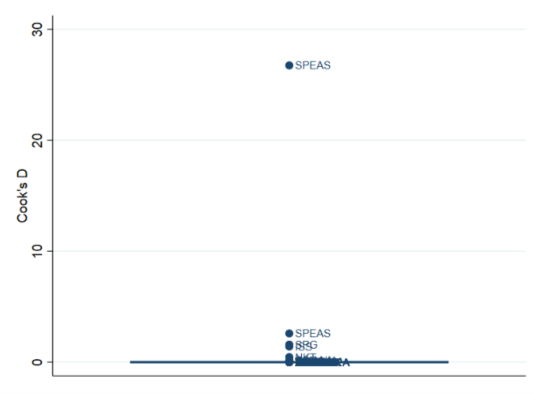


$Volatility = \beta_1 ASVI\_name + \beta_2 AR + \beta_3 ATV$

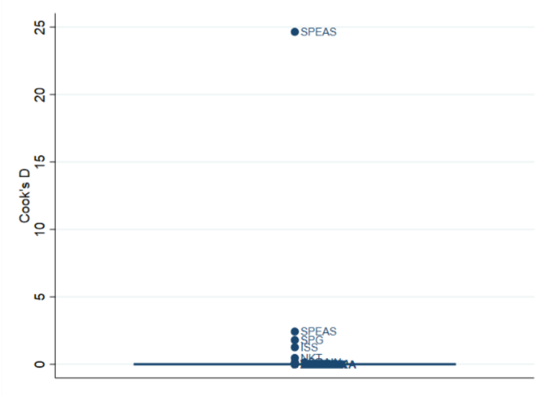


Appendix F Cook's Distance Plots – Daily Data

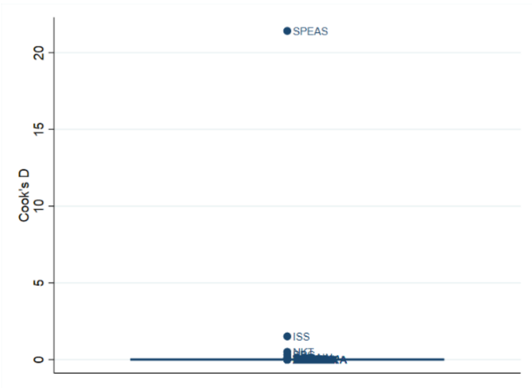
$\text{AR} = \beta_1 ASVI\_ticker + \beta_2 Volatility + \beta_3 ATV$



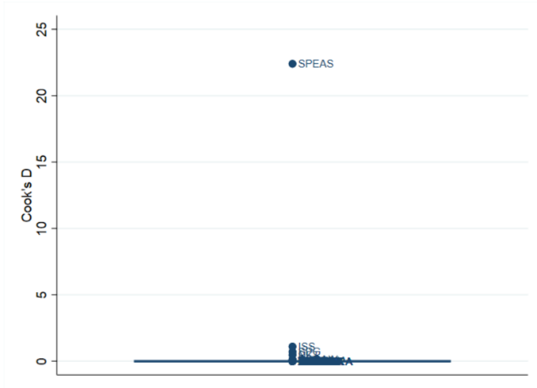
$\text{AR} = \beta_1 ASVI\_name + \beta_2 Volatility + \beta_3 ATV$



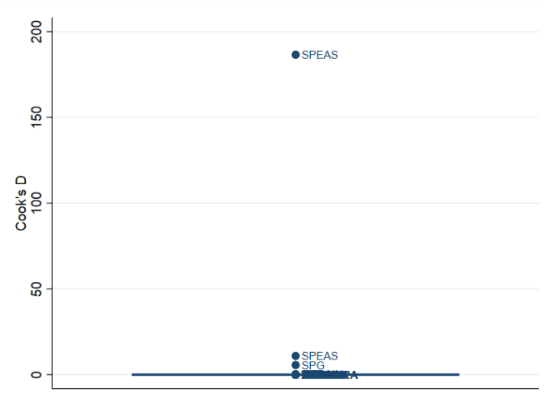
$ATV = \beta_1 ASVI\_ticker + \beta_2 Volatility + \beta_3 AR$



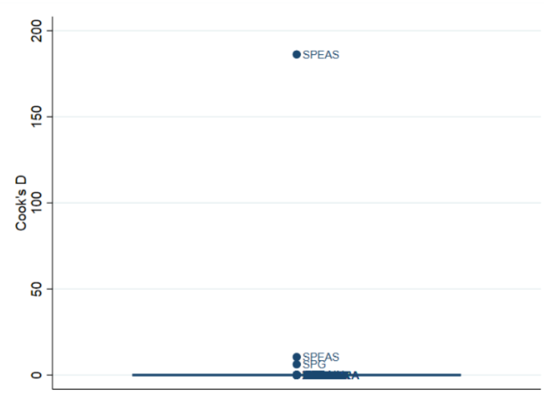
$\text{ATV} = \beta_1 ASVI\_name + \beta_2 Volatility + \beta_3 AR$



$Volatility = \beta_1 ASVI\_ticker + \beta_2 AR + \beta_3 ATV$



$Volatility = \beta_1 ASVI\_name + \beta_2 AR + \beta_3 ATV$



Appendix G VIF Statistics

		Weekly		Daily	
Regression Model	**Variable**	**VIF**	**1/VIF**	**VIF**	**1/VIF**
AR $= \beta_1 ASVI\_ticker$	ATV	1,04	0,960334	1,00	0,996926
$+ \beta_2 Volatility$	ASVI_ticker	1,02	0,97837	1,00	0,997033
$+ \beta_3 ATV$	Volatility	1,02	0,979458	1,00	0,998112
	Mean VIF	1,03		1,00	
AR $= \beta_1 ASVI\_name$	ATV	1,05	0,949407	1,01	0,991512
$+ \beta_2 Volatility$	ASVI_name	1,04	0,965098	1,01	0,992524
$+ \beta_3 ATV$	Volatility	1,02	0,978509	1,00	0,998955
	Mean VIF	1,04		1,01	
$ATV = \beta_1 ASVI\_ticker$	ASVI_ticker	1,00	0,997156	1,34	0,746183
$+ \beta_2 Volatility$	Volatility	1,00	0,997528	1,34	0,746453
$+ \beta_3 AR$	AR	1,00	0,998165	1,00	0,997007
	Mean VIF	1,00		1,23	
ATV $= \beta_1 ASVI\_name$	ASVI_name	1,01	0,994915	1,34	0,747511
$+ \beta_2 Volatility$	Volatility	1,00	0,995282	1,34	0,747784
$+ \beta_3 AR$	AR	1,00	0,998129	1,00	0,999331
	Mean VIF	1,00		1,23	
$Volatility = \beta_1 ASVI\_ticker$	ASVI_ticker	1,02	0,977899	1,00	0,997339
$+ \beta_2 AR$	ATV	1,02	0,978635	1,00	0,997706
$+ \beta_3 ATV$	AR	1,00	0,998761	1,00	0,999255
	Mean VIF	1,02		1,00	
$Volatility = \beta_1 ASVI\_name$	ASVI_name	1,04	0,965602	1,01	0,992187
$+ \beta_2 AR$	ATV	1,03	0,966325	1,01	0,99237
$+ \beta_3 ATV$	AR	1,00	0,998794	1,00	0,999453
	Mean VIF	1,02		1,01	

Appendix H Static Model Regression Results for Weekly Abnormal Return

	Dependent variable: *Abnormal return*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ASVI_ticker	0.00155				0.00147		0.00159
	(0.000865)				(0.000807)		(0.000792)
ASVI_name		0.00152				0.00144	
		(0.00106)				(0.000905)	
ATV			0.000707		0.000675	0.000633	0.000825
			(0.00135)		(0.00145)	(0.00140)	(0.00147)
Volatility				-0.0638	-0.0804	-0.0816	-0.106
				(0.0894)	(0.103)	(0.104)	(0.0951)
N	7072	7072	7072	7072	7072	7072	7072
Entity Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	NO	NO	NO	NO	NO	YES
R^2	0.001	0.001	0.000	0.000	0.002	0.002	0.115
adj. R^2	0.001	0.001	0.000	0.000	0.001	0.001	0.088
rho	0.00502	0.00506	0.00518	0.00473	0.00474	0.00477	0.00515

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix I Static Model Regression Results for Weekly Abnormal Return

	AR1		AR2		AR3	
	(1)	(2)	(3)	(4)	(5)	(6)
ASVI_ticker	0.00147	0.00159	0.00128	0.00152	0.000834	0.00156
	(0.000807)	(0.000792)	(0.000889)	(0.000854)	(0.000928)	(0.000852)
ATV	0.000675	0.000825	0.000716	0.000880	0.000853	0.000883
	(0.00145)	(0.00147)	(0.00148)	(0.00149)	(0.00151)	(0.00151)
Volatility	-0.0804	-0.106	-0.0832	-0.101	-0.0477	-0.102
	(0.103)	(0.0951)	(0.118)	(0.113)	(0.119)	(0.115)
N	7072	7072	7072	7072	7072	7072
Entity Fixed Effects	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	YES	NO	YES	NO	YES
R^2	0.002	0.115	0.002	0.108	0.001	0.219
adj. R^2	0.001	0.088	0.001	0.081	0.001	0.195
rho	0.00474	0.00515	0.00517	0.00556	0.00466	0.00547

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	AR1		AR2		AR3	
	(1)	(2)	(3)	(4)	(5)	(6)
ASVI_name	0.00144	0.00166	0.00143	0.00167	0.000942	0.00178
	(0.000905)	(0.000970)	(0.000944)	(0.000981)	(0.000952)	(0.000976)
ATV	0.000633	0.000778	0.000656	0.000827	0.000812	0.000820
	(0.00140)	(0.00144)	(0.00142)	(0.00145)	(0.00146)	(0.00147)
Volatility	-0.0816	-0.109	-0.0849	-0.103	-0.0489	-0.105
	(0.104)	(0.0957)	(0.118)	(0.113)	(0.119)	(0.114)
N	7072	7072	7072	7072	7072	7072
Entity Fixed Effects	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	YES	NO	YES	NO	YES
R^2	0.002	0.115	0.002	0.108	0.001	0.220
adj. R^2	0.001	0.088	0.002	0.081	0.001	0.196
rho	0.00477	0.00517	0.00518	0.00558	0.00467	0.00550

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Where

AR1 = $\log\left(\frac{P_t}{P_{t-1}}\right) - (r_f + \beta_1(r_m - r_f) + \beta_2(SMB) + \beta_3(HML) + \varepsilon)$.

AR2 = $log\ (R_i) - \beta(log(R_m))$.

AR3 = $log(R_i) - log(R_m)$.

Appendix J Static Model Regression Results Daily Data with Influential Variable

	Dependent variable*: Abnormal return*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ASVI_ticker	0.000635				0.00106		0.00117*
	(0.000769)				(0.000592)		(0.000565)
ASVI_name		0.000509				0.000562	
		(0.000536)				(0.000518)	
ATV			0.000129		0.000246	0.000247	0.000235
			(0.000362)		(0.000361)	(0.000365)	(0.000347)
Volatility				-1.289***	-1.295***	-1.292***	-1.296***
				(0.0294)	(0.0248)	(0.0255)	(0.0249)
N	4352	4352	4352	4352	4352	4352	4352
Entity Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	NO	NO	NO	NO	NO	YES
R^2	0.001	0.000	0.000	0.248	0.251	0.250	0.293
adj. R^2	0.000	0.000	0.000	0.248	0.250	0.249	0.272
rho	0.0145	0.0144	0.0145	0.00953	0.00975	0.00958	0.0100

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	Dependent variable: *Abnormal trading volume*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ASVI_ticker	0.152				0.143		0.134
	(0.0798)				(0.0819)		(0.0815)
ASVI_name		0.313***				0.309***	
		(0.0732)				(0.0729)	
AR			1.879		4.757	4.740	4.630
			(5.244)		(6.525)	(6.452)	(6.441)
Volatility				9.937	15.59	15.80	15.78
				(10.40)	(12.52)	(12.26)	(12.39)
N	4352	4352	4352	4352	4352	4352	4352
Entity Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	NO	NO	NO	NO	NO	YES
R^2	0.002	0.008	0.000	0.001	0.004	0.010	0.040
adj. R^2	0.002	0.008	0.000	0.001	0.004	0.009	0.011
rho	0.00661	0.00696	0.00657	0.00650	0.00663	0.00698	0.00667

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

	Dependent variable: *Volatility*						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
ASVI_ticker	0.000357				0.000461		0.000495
	(0.000275)				(0.000291)		(0.000304)
ASVI_name		0.000101*				0.000160	
		(0.0000489)				(0.000106)	
ATV			0.000102		0.000121	0.000123	0.000122
			(0.0000725)		(0.0000937)	(0.0000975)	(0.0000928)
AR				-0.193	-0.193	-0.193	-0.198
				(0.141)	(0.141)	(0.141)	(0.141)
N	4352	4352	4352	4352	4352	4352	4352
Entity Fixed Effects	YES	YES	YES	YES	YES	YES	YES
Time Fixed Effects	NO	NO	NO	NO	NO	NO	YES
R^2	0.001	0.000	0.001	0.248	0.252	0.250	0.279
adj. R^2	0.001	-0.000	0.001	0.248	0.251	0.250	0.257
rho	0.0138	0.0136	0.0135	0.00856	0.00898	0.00868	0.00898

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix K Descriptive Statistics Trading Strategies - Weekly Data

Descriptive statistics trading strategies -Weekly data							
Strategy	**Variable**	**Obs**	**Sum**	**Mean**	**Std, Dev,**	**Min**	**Max**
Joseph et al. (2011)	Quantile1	1456	-5,059598	-0,003475	0,041352	-0,192292	0,168862
	Quantile2	1456	-1,545816	-0,001062	0,050085	-0,808832	0,38329
	Quantile3	1248	-0,859181	-0,000688	0,049602	-0,238322	0,451305
	Quantile4	1456	4,151718	0,002852	0,047686	-0,279304	0,36892
	Quantile5	1456	1,965571	0,00135	0,067924	-0,696909	0,387712
Tan and Taş (2019)	30th percentile	2080	-4,568134	-0,002196	0,041335	-0,192292	0,28057
	70th percentile	2080	5,95251	0,002862	0,062304	-0,696909	0,387712
Bollinger (2001)	Under lower band	4031	-5,794185	-0,001437	0,048723	-0,808832	0,451305
Moving average	Over upper band	3037	4,502737	0,001483	0,056373	-0,696909	0,387712
Bollinger (2001)	Under -1	834	-3,175946	-0,003808	0,042931	-0,192292	0,168862
Average = 0	Over +1	1123	2,959973	0,002636	0,071231	-0,696909	0,387712

Appendix L Descriptive Statistics Trading Strategies - Daily Data

Descriptive statistics trading strategies -Daily data							
Strategy	**Variable**	**Obs**	**Sum**	**Mean**	**Std. Dev.**	**Min**	**Max**
Joseph et al. (2011)	Quantile1	889	-1,046275	-0,001177	0,024161	-0,22468	0,093527
	Quantile2	889	-1,100276	-0,001238	0,02282	-0,15794	0,115389
	Quantile3	762	-0,190191	-0,00025	0,025109	-0,134083	0,128421
	Quantile4	889	-1,777998	-0,002	0,036862	-0,82224	0,167292
	Quantile5	889	1,67663	0,001886	0,034088	-0,559361	0,215239
Tan and Taş (2019)	30th percentile	1280	-1,247227	-0,000974	0,023815	-0,22468	0,104833
	70th percentile	1280	1,25853	0,000983	0,031528	-0,559361	0,215238
Bollinger (2001)	Under lower band	3449	-2,272159	-0,000659	0,026323	-0,559361	0,215239
Moving average	Over upper band	788	-0,436077	-0,000553	0,040455	-0,82224	0,169198
Bollinger (2001)	Under -1	172	-0,692259	-0,004025	0,030061	-0,22468	0,07193
Average = 0	Over +1	775	1,300094	0,001678	0,034685	-0,559361	0,215239