GOOGLE TRENDS AND STOCKS

RETAIL INVESTING'S EFFECT ON TRADING PERFORMANCE OF SWEDISH LARGE CAP STOCKS

OSCAR EHINGER

SAMIER THOMAS ZEGERS MUSA

Bachelor Thesis

Stockholm School of Economics

2022

Google Trends and Stocks: Retail Investing's Effect on Trading Performance

Abstract:

We elaborate on the measure of retail investor attention by assessing search frequency in Google Trends. Search Volume Index (SVI) is explored as a proxy for attention that stocks get from retail investors, and we discuss its explanatory power on stock metrics. Our findings conclude a slight but distinct positive relationship between Google searches and stock metrics in terms of price changes, trading volume changes, and volatility changes on a 10-day rolling basis in Stockholm-listed Large Cap stocks. More particularly, SVI increases in company name searches show the strongest relationship in predicting the following week's stock movements in the tested parameters. Furthermore, the paper concludes that there are indeed industry differences in these effects within the tested stocks, where we see more evident results for Communication, Tech, Material, and Medical sectors for one-week lagged regressions. Our main addition to the existing research is the exploration of effects in Swedish markets as well as the contribution of more recent data. It also explores retail investor attention measures by discovering the dynamics in search method differences. When Google searches for the stock increases, it explains the variance better for all metrics. In contrast, when searches for the company increase, that is not necessarily the stock, it more accurately predicts the coefficient in the parameters tested.

Keywords:

Big Data, Google Trends, Retail Investing, Online Attention, Stock Trading

Authors:

Oscar Ehinger (24643)
Samier Thomas Zegers Musa (24637)

Tutor:

Riccardo Sabbatucci, Assistant Professor, Department of Finance

Examiner:

Adrien d'Avernas, Assistant Professor, Department of Finance

Bachelor Thesis
Bachelor Program in Business & Economics
Stockholm School of Economics

1. Introduction **4**
1.1 Background 4
1.2 Related Literature and Literature Reviews 5
2. Methodology and Specific Research Topic **7**
2.1 Investor Forums and Google Searches 7
2.2 Data Collection 8
2.2.1 Google Data 8
2.2.1.1 Topics vs. Terms 9
2.2.1.2 Groups of Data vs. Single Object Data 9
2.2.2 Company Stock Data 10
3. Data Management **12**
3.1 Datasets 12
3.2 Regression Model 13
3.3 Important Data Points: 13
3.3.1 P-values 13
3.3.2 Coefficient 14
3.3.3 R-squared (R2) 14
3.3.4 T-statistic 14
4. Results **15**
4.1 Panel A. ΔSVIstock Effects on Parameters. Large Cap 15
4.2 Panel B. ΔSVIcompany effects on parameters. Large Cap 15
4.3 Panel C. ΔSVIstock Effects on Parameters per Industry, One Week lagged 16
4.4 Panel D. ΔSVIcompany Effects on Parameters per Industry, One Week Lagged 16
4.5 Robustness Tests 17
5. Analysis **18**
5.1 The Effect of Different Lags on Result Significance 18
5.2 Robustness Tests 18
5.3 Entire Large Cap (Consolidated, industry-agnostic) 19
5.3.1. One-Week Lagged ΔSVI: Stock Performance Measures Comparison 19
5.3.2 ΔSVIcompany and ΔVIX Effects 19
5.3.3 ΔVolumeSVIstock 19
5.3.4 ΔVolumeSVIcompany 19
5.3.5 ΔCloseSVIcompany 20
5.3.6 ΔVol10SVIstock 20
5.3.7 ΔVol10SVIcompany 20
5.3.8 ΔVol30SVIstock 20
5.3.9 ΔVol30SVIcompany 20
5.3.10 Summary of Stock Performance measures When Compared to SVI Changes 21

5.4 Industry-divided Tests 21
5.4.1 ΔSVIstock and ΔVIX Effects. One Week Lagged 21
5.4.2 ΔSVIcompany and ΔVIX Effects, One Week Lagged 22
5.5 ΔSVI Effects on Parameters. Industry-divided, One Week Lagged 22
5.5.1 ΔVolumeSVIstock 23
5.5.2 ΔVolumeSVIcompany 23
5.5.3 ΔCloseSVIcompany 24
5.5.4 ΔVol10SVIstock 24
5.5.5 ΔVol10SVIcompany 24
5.5.6 ΔVol30SVIstock 24
5.5.7 ΔVol30SVIcompany 25
5.5.8 Summary of Industry-divided Analysis of Effects of ΔSVI 25

6. Discussion **26**
6.1 Interpretation of the Results 26
6.1.1 ΔSVIcompany vs ΔSVIcompany as a Regressor 26
6.1.2 R2: Better Explained by ΔSVIstock 26
6.1.2.1 R2 & ΔSVIstock: a Case Study on Swedbank 26
6.1.3 Coefficient: Better explained by ΔSVIcompany 27
6.1.4 Industry Differences 27
6.2 Potential for Further Research 28
6.3 Result Actionability 29

7. Concluding Remarks **30**

8. Appendix **31**
8.1 Robustness Tests 31
8.2 Supplements 33
8.3 References 36

1. Introduction

1.1 Background

The recent years have shed more and more light on the role of the retail investor in the economy as a whole. Something that has accelerated this discussion is the emergence of so-called "meme-stocks", where online forums such as Reddit, more specifically r/wallstreetbets, played central roles.[1] The effect that these online discussions could have on stock behavior has started becoming more and more of a reality. Perhaps most famously, Melvin Capital (a New York-based investment management firm) was badly bruised and down -46% on H1 2021 due to the Gamestop short-squeeze caused by retail investors that had coordinated through the r/wallstreetbets Reddit thread.[2]

The trend of investor forums has not only been an American occurrence but is also visible in Sweden. Both Twitter and Facebook profiles have gotten attention for a strong vocal presence regarding certain stocks. Many companies and people have been subject to investigations by Finansinspektionen (The Swedish Financial Supervisory Authority) for market manipulation.[3] One of the largest established Swedish investor forums, Placera, suddenly closed down during the latter parts of 2021. The reason is that the entire platform is to be re-imagined and will require the identification of its members for them to continue writing and reading posts.[4]

This paper aims to take a closer look at the ways in which internet attention regarding a certain company is correlated to the behavior of its stock on a publicly listed market. We believe that work in this area can be largely beneficial since there is not a lot of research regarding the relatively new phenomenon, which means we can gain valuable insight into how the trading patterns of retail investors have market-moving effects. Traditionally, the only actors that could make a significant difference in the markets were extremely wealthy individuals or institutional investors. It now appears that tables have turned to some extent. This makes the phenomenon in general rather new, and creates a dynamic that the public markets have not co-existed with during the majority of their life.

[1] Ponczek, Sarah, 2021. WallStreetBets Gains More Than a Million New Members Overnight, *Bloomberg,* 28 Jan
[2] Nagarajan, Shalini, 2021. The Hedge Fund Badly Bruised by Betting Against GameStop is Still Struggling After Ending the First Half with a 46% Loss, Report Says, *Business Insider,* 9 Jul
[3] Lindvall, Julia, 2021. Omfattande Aktiefusk i Chattrum Utreds av Ekobrottsmyndigheten, *SVT Nyheter,* 22 Dec
[4] TT, 2022. Då Öppnar Placera-forumet Igen, *Dagens Industri*, 21 Jan

1.2 Related Literature and Literature Reviews

While a relatively thinly covered topic, internet search and its correlation to stock movements have been examined before. Primarily, a 2011 paper published in the Journal of Finance, titled "In Search of Attention" and authored by Zhi Da, Joseph Engelberg, and Pengjie Gao stands out as the most relevant example in recent literature.[5]

The paper aims to determine the connection between a company's share movement and the attention a certain company or its stock gets. The proxy used to capture attention is found on Google Trends, called the Google Search Volume Index, which will forwardly be referred to as *SVI*. The reason for examining attention is the underlying assumption in the asset pricing models, which dictates that information is instantaneously incorporated into asset prices as soon as it arrives. For this to be true, investors must be paying attention, as attention is a prerequisite for demand. Hence, this paper will touch on the sentiment of retail investors and how attention is related to this, as discussed by Baker and Wurgler (2007).[6] Previous papers have also discussed using Google Trends as a tool to measure economic activities in real-time, notably Choi and Varian (2009).[7]

The "In Search of Attention"-methodology starts with looking at different Russell 3000 stock tickers on Google Trends, after which data is benchmarked against existing attention measures such as weekly returns, turnover, and news. After this, they aim to determine whose attention the SVI is displaying. Through the examination of SEC retail order execution, they establish a strong direct link between SVI changes and retail investor trading. This link is also stronger in market centers that attract less sophisticated investors. Lastly, they test the theory that individual investors are net buyers of stocks circulated with high attention, which would imply that a higher SVI should lead to higher buying pressure, which translated to positive price development. The reasoning behind this theory, first introduced by Barber and Odean (2008)[8], is that buying investors have substantial options to choose from, while selling investors can only sell the stocks currently in their portfolios. The framework explains that stock prices should see positive development in the short-term followed by reversals in the long term. This pattern should be more clearly distinguishable in stocks where individual investor attention has the largest impact.

When it comes to price pressure, the paper concludes that there is a link between an increase in SVI and higher stock prices during the first two weeks of the SVI spike. Further, it establishes a high likelihood of a price reversal within the year.

The paper, as a whole, was one of the first of its time to make use of internet search volumes in a financial economics context. It proves the importance and impact of search-proxied interest measures and how they affect large-scale financial performance.

The main difference between the above-discussed paper and this one is the geographical limitation. Furthermore, the societal role of retail investors between the US and Sweden is

[5] Da, Zhi, Joseph Engelberg, Pengjie Gao, 2011. In Search of Attention., *The Journal of Finance* 66, 1461-1499
[6] Baker, Malcolm, and Jeffrey Wurgler, 2007. Investor sentiment in the stock market, *Journal of Economic Perspectives* 21, 129-151
[7] Choi, Hyunyoung, and Hal Varian, 2009. Predicting the present with Google Trends, Working paper, Google Inc.
[8] Barber, Brad M., and Terrance Odean, 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies* 21, 785–818

an interesting dynamic to measure, which will be discussed in this paper. There might be a difference in cultural factors regarding retail investing; this paper could give more insight into how these cultures differ between the two countries. For instance, retail investing in Sweden has for a very long time been characterized by a long-term strategy where large and well-established Swedish stocks or ETFs have been widely popular. An example of this is parents placing funds in said assets in conjunction with the birth of their children later to transfer control of the assets to their kids once they grow older.

Another dimension to this paper that the previous paper has not mentioned is the industry effect of the SVI. This paper discusses the most affected sectors of retail investor sentiment and intends to bone out the possible underlying reasons. It also provides a more recent time period and explores the dynamics of the current market that we are in. Moreover, it discusses the dynamics of differences in search methods on Google.

2. Methodology and Specific Research Topic

2.1 Investor Forums and Google Searches

Typically, one would expect smaller firms to be affected to a greater extent by retail investors due to the lower trading volume and fewer owners of the stock. To exemplify, this phenomenon has been proved in various ways through market participants who aim to manipulate this sort of stocks for their own gain.[9] This paper aims to examine a similar trend on a somewhat larger scale where the connection between online presence and stock movement is not as prevalent. The focus will be put on industries and companies that have an established presence in retail consumer portfolios and focus on the situation in Sweden. In this paper, we will examine large Swedish companies instead of smaller ones.

To capture all retail investors is a tedious task and would require gathering data from all existing stock forums 5 years back. According to our research and attempts, it is a method prone to many errors and other intricacies, and might even be out of scope for this paper. To better encompass the behavioral pattern of more retail investors, Google searches will be examined, as many papers previously have explored. Naturally, Google searches will not fully represent how popular or trendy a stock currently is; however it can be used as a good proxy since it captures the most data points according to our research that is also viable within the scope of this paper. Targeting a specific forum could have potentially been a solid option, but it would also elevate the risk of biases and retro-fitting data as not all stocks are discussed in all forums. Many other papers discuss what captures investor attention and what metrics predict expected trading activity in stocks. Chordia, Huh, and Subrahmanyam (2006)[10] explore firm metrics and conclude that stock visibility is based on many factors, such as firm size, age, price or other financial metrics. Tetlock (2007)[11] concludes in his paper "Giving Content to Investor Sentiment: The Role of Media in the Stock Market" that unusually high or low pessimism predicts trading volume.

This paper will focus not on the underlying factors of the Google search increases, but rather the effects of the changes in Google searches. The Swedish stocks that have been selected have been filtered by size and by industry, where the entire OMX Large Cap has been examined. If a stock has too few searches or data points, there is a risk that the data loses its explanatory value, which is why some stocks have been removed from the sample. Another criterion is that the firm must have traded on a stock exchange for at least five years for statistical reasons and comparability possibilities.

Our method is fairly straightforward in terms of execution and offers a simple but clear-cut way of analyzing the research question: Does Google Search Index data predictably affect price data, volume, and volatility in Large Cap Stockholm? The collected data points are regressed against the different variables in establishing existing statistically significant relationships. The variables and data points that are tested against each other in our base scenarios will be primarily:

1. Google Trend Searches effect on Trading Volume

[9] Österberg, Tobias, 2022. Aktieforum Rensas upp efter Abrupt Stängning, *SVD Näringsliv*, 21 Jan
[10] Chordia, Tarun, Sahn-Wook Huh, and Avanidhar Subrahmanyam, 2007. The Cross-section of Expected Trading Activity, *Review of Financial Studies* 20, 709-740.
[11] Tetlock, Paul C., 2007, Giving content to investor sentiment: The Role of Media in the Stock Market, *Journal of Finance* 62, 1139-1168.

2. Google Trend Searches effect on Closing Price
3. Google Trend Searches effect on 10-Day Volatility
4. Google Trend Searches effect on 30-Day Volatility

Further, these variants will be analyzed from a time-frame perspective in order to determine the statistically most significant length of lag needed to establish the strongest results. Lastly, tests will be carried out when data is divided by industry to distinguish in which industries these relationships hold the strongest.

Our hypothesis will be in line with our reference paper's - that Google SVI will predictably affect movements in trading volume, closing price, and volatility in Large Cap Stockholm, and most notably will affect sectors where many retail investors are present, e.g. Tech based on empirical observations. Volatility increases are consistent with the findings of Foucault, Sraer, and Thesmar (2009) in their paper "Individual Investors and Volatility" which focuses on retail trading activity effect on the volatility of stock returns.[12]

2.2 Data Collection

The data needed for the empirical parts of this paper can be divided into two parts; Google Data and Company Stock Data.

2.2.1 Google Data

All Google Web Search Data used in this paper comes from Google's own online tool named "Google Trends". The tool's purpose is to allow for monitoring changes in web-searching behavior regarding certain topics or terms over time. How the data works is that search volume is presented as a Search Volume Index (SVI) where the peak value (=100) is assigned to the day that had the highest traffic in searches for the selected topic. Data is presented and downloadable on a weekly basis per Sunday every week. The data is rounded, implying that no decimals are presented.

The indexed data is transformed into weekly changes, or deltas (Δ), to make the regressions relevant and comparable. So for example, if SVI data for week 1 is 25 and for week 2 it is 75, the weekly change for week 2 will be presented as 75/25-1 = 200%. This will be the data included in the dataset for testing.

The manipulated data will be used in the tests as regressors and further referred to as:

1. $\Delta SVI_{company}$: The weekly change in Google SVI is represented by the search for only the company name, for example, "Atlas Copco", "Handelsbanken", or "Castellum".
2. ΔSVI_{stock} : The weekly change in Google SVI, the term searched for is only the company name followed by the Swedish word for stock "aktie" to isolate the Google search. For example: "Atlas Copco aktie", "Handelsbanken aktie", or "Castellum aktie".
3. ΔVIX: The weekly change in VIX is the volatility index based on the S&P500 index. This regressor will be used in the multivariate regressions as a test for robustness.

[12] Foucault, Thierry, David Sraer, and David J. Thesmar, 2011. Individual Investors and Volatility, *The Journal of Finance* 66, 1369-1406

2.2.1.1 Topics vs. Terms

Google has different types of categorizations for subjects to analyze with Google Trends. The entered term can either be analyzed at face value or as a search term. Output data will reflect search activity regarding the exact written term.

Instead of just looking at terms, Google also allows you to monitor trend data for topics. This output shows how developing an (unspecified) selection of terms looks. For instance, a search for the *term* "Microsoft" would show data for every time "Microsoft" has been searched, which most likely occurs in connection to their products more than pure searches for the company. Looking at "Microsoft", *topic*: Company, instead looks at searches regarding the company itself (which includes a combination of different*terms*). However, Google's function for this feature is limited and typically only works for the largest American blue-chip stocks. Not all companies in our sample have support for this feature, so we have selected other comparable types of data to avoid confusion, or "noise", in search data. For example, when searching for terms such as "Handelsbanken", "Swedbank", or "Nordea", searchers are not necessarily interested in the stock but rather in the services offered by the bank. This is why we have included the company name, followed by the Swedish translation for stock, "aktie". Data for company ticker has proved not to be widely searched in Sweden, which is why it will not be included in the analysis, e.g. the search volume has been too infrequent to be able to show any results on Google's side. By measuring both types of searches, we hope to find differences in dynamics in the results.

The total ownership of Swedish stocks was estimated to be distributed as follows: Swedish entities (45.5%), International entities (42.1%), and Swedish retail investors accounting for 12.1%.[13] We have delimited our search range in Google Trends to Sweden only as we expect most distribution among retail investors in these types of stocks to come from domestic markets. We explored the possibility of including international statistics and adding "stock" after each company name as well as searching for company-specific tickers. However, there were only occasionally successful searches, with most of them resulting in not having sufficient data to present, or with many weeks where the indexed number had been 0, which is why we have excluded international demand. This also confirms our notion that most of retail demand for our chosen stocks comes from domestic markets. We expect professional investors to not search for the company the same way as a retail investor would, e.g. Google, as they have more sophisticated software available such as Bloomberg Terminal, Refinitiv, Thomson Reuters, etc.

2.2.1.2 Groups of Data vs. Single Object Data

When inputting a term (or topic) to Google Trends, there is a possibility to compare up to 5 different terms simultaneously. The mechanics of the data that is output remains the same regardless of the number of terms analyzed; the term and date with the highest number of web searches will be indexed as 100, and all other data points are relative to this peak. All datasets from Google Trends will have one point with the value 100, and this holds true whether the amount of examined terms is 1 or 5. The data is presented on a weekly basis, and as such, we will present stock data on a weekly basis as well because there is no way of receiving data with tighter intervals than this from Google Trends.

[13] Statistiska Centralbyrån, 2021. Aktieägarstatistik.

As this report will examine more than 5 search terms, no crossover comparisons in relative terms will be made, but rather relative popularity to the stock itself. This is sensible from a statistical perspective as one would expect the number of search terms to be relative to the weekly trading volume, meaning that the absolute number of searches is not *as* important as the relative number of searches.

To clarify with an example: Suppose that we have two stocks, one that is extremely well-known, such as Apple, and one that is considerably less known internationally, such as Ratos. When searching for these two firms with a geographical limitation to Sweden, the data in *Figure A* is the data that is extracted. Compared to the data in *Figure B*, which only includes Ratos, it is apparent that our selected method is the most viable. Note that the data in *Figure A* and *Figure B* for Ratos is identical. *Figure A* is benchmarked against Apple's maximum number of searches, whereas *Figure B* is benchmarked against Ratos' maximum number of searches. Thus, grouping different terms or topics together can easily result in a misleading representation of development in a company's own Google searches as this becomes dependent on the subjective choice of companies included in the same group.

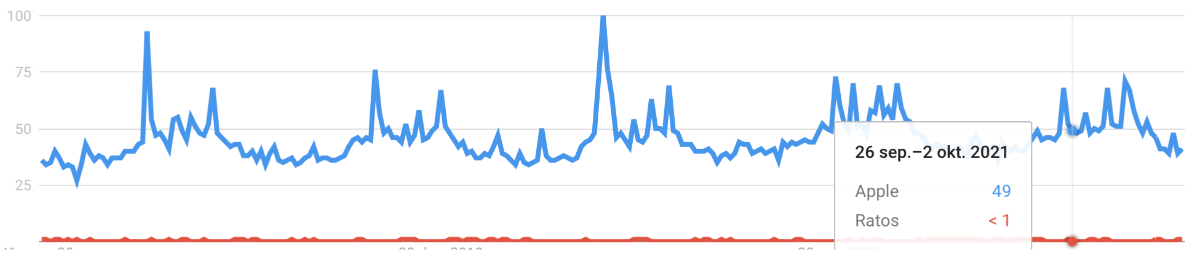


Figure A. Apple vs Ratos searches in Google Trends. Apple: Blue, Ratos: Red. Note that Ratos value is <1

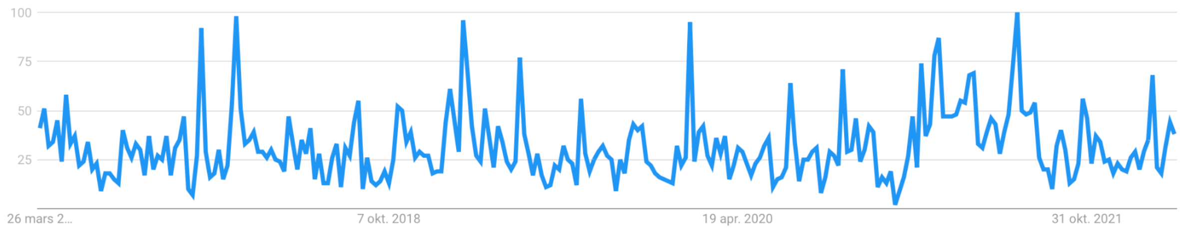


Figure B. Ratos searches in Google Trends in relation to itself

2.2.2 Company Stock Data

All company stock data is gathered from *Bloomberg* through its excel add-in and consists of the following data points collected from an L5Y period of all Large Cap Stockholm stocks with weekly intervals (per each Friday):

1. Closing Price
2. Volume Traded
3. 10-day volatility
4. 30-day volatility

To match our criteria, the firm has to 1. Be well-known and widely searched in Sweden according to Google Trends, and 2. Trade on Large Cap Stockholm. Usually, these two criteria coincide, but some companies have been discarded. Furthermore, stocks have been split into different categories, based on sector or industry, split by Avanza's (Swedish bank) segmentation. If a company has multiple stocks listed on the exchange (A-class, B-class, C-class etc.), the stock with the highest liquidity has been selected.

The data will be standardized in the same way as the Google SVI data. So, weekly percentage changes will be calculated for each data point. The data for week 2 will imply the change from week 1 to week 2. This will be done for all data points for Closing Price, Volume Traded, 10-Day Volatility and 30-Day Volatility.

The manipulated data will be used in the tests and further referred to as:

1. ΔVolume : The weekly change in volume traded is measured from Friday - Friday. The trading volume is the amount of stocks that have been traded.
2. ΔClose : The weekly change in closing price is measured from Friday - Friday. The closing price is the last price paid for any given stock.
3. ΔVol10 : The weekly change in the 10-day average backwards-looking volatility in stock price measured from Friday - Friday.
4. ΔVol30 : The weekly change in the 30-day average backwards-looking volatility in stock price measured from Friday - Friday.

3. Data Management

3.1 Datasets

Prior to testing, the data had to be manipulated to ensure that all inputs followed consistent formatting. Since the Google Trends data can only be downloaded weekly, this is the form the remaining data have been manipulated into using =INDEX(MATCH) functions in excel to align dates. Note that the data from Google SVI is per each Sunday, whereas the closing price and volume for each stock are per each Friday. This implies that the most logical way to perform the tests is with lagged data. Testing data with no lag will mean that the SVI from Saturday and Sunday are included, although the closing price is on Friday that same week. It further implies that in order to match the data points for the tests, the dates have to be aligned. This is done by simply matching the last data point for each week, which means that there is a slight discrepancy in the data points tested for each week. Ideally, albeit not possible, the SVI would have been downloaded on a weekly basis each Friday. On the other hand, as the stock markets are closed on weekends, one might argue that not a lot of searches are made on Saturdays and Sundays.

The same procedure is repeated for each tested data, with the difference being the amount of lag used. Regressions have been generated for data lagged from 1-4 weeks as well as with no lag. This is simply done by pulling the regressor one week forward, that is $\Delta SVI_{company}$ and ΔSVI_{stock}. The regressor's data from week 0 is matched with the endogenous variable's data for weeks 1, 2, 3, or 4, depending on the amount of lag.

As the one-week lag appeared to show the strongest levels of significance, coefficients, and R^2, these are the tests that have been further explored and analyzed at an industry level. The industries are defined per the internet bank Avanza's segmentation, which are:

1. Communication
2. Tech
3. Material
4. Industrial
5. Consumer
6. Medical
7. Financial
8. Real Estate

As many data points for the SVI included "0" (rounded to 0), the percentage change in levels could be deemed as a poor way of manipulating the data as there would be a significant amount of 0's involved in the manipulated data. To counteract this, each value in the original data containing 0 has been upwards-adjusted to 0.5, after which the tests showed stronger significance and better explanatory parameters.

With all data properly formatted and in place, the *Statsmodels* and *Linearmodels* Python modules were used to run the statistical tests. *Statsmodels* and *Linearmodels* support specifying models using formulas similar to R and support *pandas DataFrames* with OLS-methodology. Hence, it provides the simplicity of Python language and the *Pandas DataFrames* while still keeping *R*'s strength as a statistical programming language.

Both changes in closing price (ΔClose), trading volume (ΔVolume), and volatility on a 10-and 30-day moving average basis (ΔVol10, ΔVol30) will be examined by use of change in Company SVI ($\Delta SVI_{company}$) and Stock SVI (ΔSVI_{stock}) as regressors. To run the tests, the variables are tested and denoted in the following manner to examine SVI's impact on volume, closing price and volatility:

Table A	ΔClose	ΔVolume	Δ10-day Volatility	Δ30-day Volatility
$\Delta \mathbf{SVI_{stock}}$	$\Delta Close_{SVIstock}$	$\Delta Volume_{SVIstock}$	$\Delta Vol10_{SVIstock}$	$\Delta Vol30_{SVIstock}$
$\Delta \mathbf{SVI_{company}}$	$\Delta Close_{SVIcompany}$	$\Delta Volume_{SVIcompany}$	$\Delta Vol10_{SVIcompany}$	$\Delta Vol30_{SVIcompany}$

Table A - Matrix of the tests that will be conducted. $\Delta SVI_{company}$ and $\Delta SVI_{company}$ will serve as regressors.

With this way of structuring the data, operations will make it easy to analyze the connections between SVI and closing price, volume traded and volatility. These tests are first run on a no-lag basis, then weekly lagged levels as well as industry-divided for one week lag. To exemplify, $\Delta Close_{SVIstock}$ is a test where ΔSVI_{stock} is used as a regressor to determine the effect on the *y*-variable ΔClose. Based on empirical observations in the data, a normal distribution will be assumed for the tests.

3.2 Regression Model

The type of regression analyzed in this paper follows the Ordinary Least Squares (OLS) methodology for Panel Regressions. OLS looks at generating a coefficient that creates a graph that minimizes the sum of the squares of the variance, which is the difference between a given data point and the plotted line for the same independent variable.

To measure the effect that $\Delta SVI_{company}$ and ΔSVI_{stock} have on share metrics, ΔClose, ΔVolume, ΔVol10, and ΔVol30 are set as dependent variables. The tested variables are standardized per previous explanation and will be used as weekly % changes, or deltas. We have primarily assessed a univariate panel regression that has been robustness-tested with a multivariate panel regression, including ΔVIX.

3.3 Important Data Points:

Although the regression model has a wide array of output data, some metrics, or data points, are more important than others when it comes to outlining potential co-movements between SVI and share performance. These will be introduced and discussed below.

3.3.1 P-values

Before analyzing other output metrics derived from the regression model, the p-value should be examined as it determines whether the result is statistically significant. The definition of the p-value is that it acts as the probability of generated results being equally or more extreme than the observed results. This means that a low p-value indicates a low likelihood of results being more extreme than observed, which means that the observed result is more statistically significant. The method is based on a system of hypotheses built up of a null hypothesis supported by an alternative hypothesis. Using p-values, we can determine how likely our null hypothesis is false, which acts as a validator for the alternative hypothesis. A p-value of *x%* implies that we can reject our null hypothesis with a confidence level of *1-x%*. Thus, the alternative hypothesis has a *1-x%* statistical confidence level of being correct.

While it is dependent on context as well as the type of test, a p-value that is lower than 5% usually indicates an observed result which is significant. This value can be seen as a gatekeeper, or an indicator, used to determine whether the rest of the generated results carry any statistical significance. Thus, data containing a high p-value makes it ill-fitting to serve as base for further analytical conclusions. Therefore, looking at p-values becomes vital before trying to draw conclusions from other data regarding a certain company. The significance levels used in this paper are 5%, 1%, and 0.1%, which represent *,**, and *** respectively.

3.3.2 Coefficient

The parameter, or coefficient, describes the linear relation or slope of the regression between the independent (SVI) and dependent variables (ΔClose, ΔVolume, ΔVol10, and ΔVol30). An *x* shift in SVI will predict a specific *y* shift in closing price and trading volume, showing the nominal relationship of the two tested variables. By simply plugging in the value of X into the regression model, we will be able to predict Y. Since the values used in the tests are standardised, a 1-point increase in *x* will result in a-

$$b = coefficient * (standard\ deviation\ for\ Y)$$

-absolute increase in Y-value. The coefficient will represent b in

$$Y = a + bX.$$

All standard deviation data for specific metrics is found in Panel B1 in the appendix.

3.3.3 R-squared (R^2)

R^2 looks at the relationship between the independent (SVI) variable and dependent variable (ΔClose, ΔVolume, ΔVol10 and ΔVol30) and aims at describing the amount of variance in the dependent variable that the independent variable can explain. The value ranges between 0-1 where a value of 1 means that 100% of the dependent variable changes in value can be described by the changes in the value of the underlying independent variable (SVI).

There is no certain limit at which an R^2 value starts to prove explanatory power. Instead, the measure is more subjective and impacted by circumstantial factors. This means that there is a need for constant re-evaluation when looking at R^2 numbers, and connecting it back to the question examined in this specific case. Further, this infers that it is impossible to objectively state what a sufficient, or insufficient, R^2 values are. In the case of this paper, all R^2 values will

3.3.4 T-statistic

The t-statistic assumes a normal distribution and is used as a hypothesis testing tool to test whether we should support or reject the null-hypothesis. Values derived from the t-test are used to test the significance of the regressors. It will be used in conjunction with the p-values to inform what the likelihood is that the results would have happened. As such, it will be an important part of the results and is why we include it as a basis to decide which data points are significant enough to base further analysis on.

4. Results

Results represent significance levels of 5%, 1%, 0.1% (*,**,***)
Time frame for the tests are April 2017 - April 2022
All eligible companies from Large Cap Stockholm have been included (see 2.2.2)
Values in [brackets] represent the T-statistic

4.1 Panel A. ΔSVI_{stock} Effects on Parameters. Large Cap

Panel A	Δ Volume		Δ Close		Δ Vol10		Δ Vol30	
	Coefficient	R2	Coefficient	R2	Coefficient	R2	Coefficient	R2
No Lag	0.0045*** [4.2231***]	0.0023***	0.0001*** [4.1245***]	0.0009***	0.001*** [5.7145***]	0.001***	0.0001 [1.7047]	0.0001
One Week Lag	0.0059*** [9.0649***]	0.0039***	0.0001*** [4.036***]	0.0009***	0.0027*** [9.2279***]	0.0077***	0.0007*** [7.3621***]	0.0054***
Two Week Lag	0.0041*** [4.4732***]	0.0019***	0.0001** [2.9255**]	0.0004**	0.0015*** [6.1173***]	0.0024***	0.0003*** [4.9024***]	0.0011***
Three Week Lag	[0.003***] [5.9515***]	0.001***	0* [2.2957*]	0.0003*	0.0008*** [4.0046***]	0.0007***	0.0002*** [3.4142***]	0.0003***
Four Week Lag	[0.0045***] [5.4056***]	0.0023***	0.0001*** [4.0266***]	0.0006***	0.0013*** [6.3702***]	0.0017***	0.0003*** [5.2441***]	0.0006***

4.2 Panel B. $\Delta SVI_{company}$ effects on parameters. Large Cap

Panel B	Δ Volume		Δ Close		Δ Vol10		Δ Vol30	
	Coefficient	R2	Coefficient	R2	Coefficient	R2	Coefficient	R2
No Lag	0.0087 [1.6519]	0.0007	0 [0.1894]	0	0 [-0.0638]	0.00E+00	-0.0003* [-2.4573*]	0.0001*
One Week Lag	0.0169*** [5.2749***]	0.0027***	0.0003*** [4.2428***]	0.0008***	0.0052*** [5.5062***]	0.0023***	0.0009*** [4.1414***]	0.0007***
Two Week Lag	0.0115** [2.6358**]	0.0012**	0.0001* [2.0736*]	0.0001*	0.0028*** [5.021***]	0.0007***	0.0006*** [3.5726***]	0.0003***
Three Week Lag	[0.0079**] [2.6834**]	0.0006**	0.0001 [1.5514]	0.0001	0.0008* [2.0096*]	0.0001*	0.0004* [2.3242*]	0.0001*
Four Week Lag	[0.0122**] [2.6204**]	0.0014**	0 [0.5632]	0	0.0006 [0.9867]	0	0.0001 [0.508]	0

4.3 Panel C. ΔSVI_{stock} Effects on Parameters per Industry, One Week lagged

Panel C	Δ Volume		Δ Close		Δ Vol10		Δ Vol30	
	Coefficient	R2	Coefficient	R2	Coefficient	R2	Coefficient	R2
Communication n=4	0.0416*** [13.409***]	0.0082***	0 [-0.0023]	0	0.0233*** [6.4444***]	0.0224***	0.0069*** [11.679***]	0.022***
Consumer n=10	0.0047*** [3.5578***]	0.0073***	0 [0.9338]	0.0001	0.0025** [2.8067**]	0.0064**	0.0005* [2.3039*]	0.0024*
Tech n=8	0.0076** [3.1158**]	0.0072**	0.0002*** [5.4385***]	0.0052***	0.0041** [3.1608**]	0.0128**	0.0009* [2.5345*]	0.0059*
Real Estate n=12	0.0058*** [3.2942***]	0.0025***	0 [0.7292]	0.0002	0.0017*** [4.3702***]	0.007***	0.0004*** [3.972***]	0.0053***
Industrial n=27	0.0062*** [6.1868***]	0.004***	0.0001*** [3.6261***]	0.0015***	0.0027*** [5.7929***]	0.0089***	0.0009*** [4.659***]	0.009***
Financial n=14	0.0046*** [5.1176***]	0.0072***	0.0002 [1.7979]	0.0031	0.0029*** [3.3491***]	0.0054***	0.0007** [2.6367**]	0.0028**
Material n=7	0.0064** [2.9901**]	0.029**	-0.0001 [-1.1454]	0.0004	0.0046*** [3.7804***]	0.0226***	0.0014** [2.6828**]	0.0198**
Medical n=7	0.0055** [3.1122**]	0.0033**	0.0001* [2.1662*]	0.0016*	0.003** [2.8057**]	0.005**	0.001 [1.5645]	0.0037

4.4 Panel D. $\Delta SVI_{company}$ Effects on Parameters per Industry, One Week Lagged

Panel D	Δ Volume		Δ Close		Δ Vol10		Δ Vol30	
	Coefficient	R2	Coefficient	R2	Coefficient	R2	Coefficient	R2
Communication n=4	0.1031 [1.8208]	0.0086	0.0013*** [3.3623***]	0.0014***	0.0372** [2.9954**]	0.0098**	0.0066 [1.4229]	0.0034
Consumer n=10	0.0132*** [4.9732***]	0.0055***	0 [-1.2998]	0	0.0033 [1.7791]	0.001	0.0007 [1.2187]	0.0005
Tech n=8	0.0817 [1.8654]	0.0082	0.0006 [1.0187]	0.0003	0.0326*** [3.4378***]	0.0082***	0.0045* [2.1231*]	0.0016*
Real Estate n=12	0.0155** [2.9536**]	0.002**	0.0003*** [3.4761***]	0.0025***	0.0023*** [3.5968***]	0.0016***	0.0002 [1.4804]	0.0002
Industrial n=27	0.0167*** [4.0199***]	0.0013***	0.0003* [2.4476*]	0.0006*	0.0066*** [5.9881***]	0.0024***	0.0014*** [4.0968***]	0.0011***
Financial n=14	0.0259* [2.0503*]	0.0045*	0.0007 [1.5259]	0.0007	0.0118 [1.3231]	0.0018	0.0038 [1.3485]	0.0017
Material n=7	0.0073*** [4.2402***]	0.0066***	0.0003*** [16.9519***]	0.0014***	0.0034*** [4.2208***]	0.0021***	0.0004 [1.5067]	0.0003
Medical n=7	0.0245* [2.0974*]	0.0132*	0.0004** [2.6343**]	0.0024**	0.0094*** [5.1499***]	0.0102***	0.0015* [2.3746*]	0.0017*

4.5 Robustness Tests

To test the robustness of the regressions, we have run a multivariate panel regression where the lagged delta of the volatility index VIX has been included as an additional regressor. The results that we found had the most significance for our analysis have been selected for robustness checking. "*n*" represents the number of companies included in the tests. All of these have 263 observations each reflecting the weeks between April 2017 - April 2022

5. Analysis

5.1 The Effect of Different Lags on Result Significance

When examining the link between Google search indices and differences in various stock performance measures, the time aspect becomes a vital dimension to keep track of. This is grounded in a basic assumption regarding a search being made, which at a later stage affects the particular stock in question. However, pinpointing exactly which amount of lag would be the appropriate level to use when aiming to extrapolate potential correlations between SVI movements and stock performance behavior is complex. When looking at the different results generated, t-stat values are used as the basis for decisions regarding which amount of lag is most reliable to base further analysis upon. With this in mind, a One-Week Lag becomes a clear outlier in terms of displaying the highest t-stat values, which in turn means that the significance of later generated metrics are better fit as a basis for analysis. With an overall p-value of less than 0.1%, the tests performed on data where the SVI is compared to stock performance with a lag of one week (Row 2, Panel A) will serve as the basis for further analysis and discussion. This lag window is also mentioned and analyzed in the previously mentioned paper examining similar topics, "In Search of Attention".

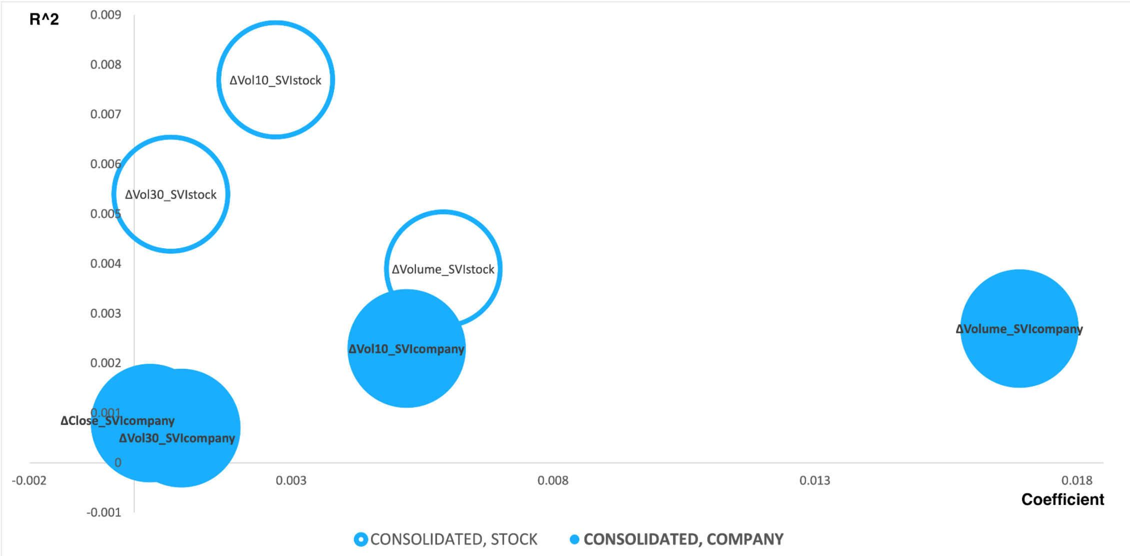
Looking at the exact methodology and the more intricate details of the lag, it is important to remember that the raw pricing data is measured per Friday, while the SVI data is measured per Sunday. However, all data points are measured on weekly bases, and the adjustment for lag is effectively pulling the SVI "forward" as many weeks as the test requires. This means that one week of lag means that the ΔSVI data is moved so that we examine how ΔSVI of any given date affects the stock performance metric *x* weeks in the future.

Apart from the one-week lag showing the highest t-stat values and, therefore, the lowest p-values, we can also see in Panel A that the R^2 values of the one-week lagged row are the highest. This further helps with establishing that a lag of one week seems to be the best time period to base further analysis upon.

5.2 Robustness Tests

A robustness test was done by adding regressor ΔVIX, which is the volatility index of S&P500 and a constant. The test is deemed robust if the regressor keeps its significance and does not change the coefficient notably. The tests that have shown to be robust and kept their core significance and coefficient will be elaborated on below. All robustness test values are available in Panel A1-A4.

5.3 Entire Large Cap (Consolidated, industry-agnostic)



Graph A: Visualization of stock performance measures: ΔSVI_{stock}vs $\Delta SVI_{company}$ as a regressor. Data significant at p≤5%

5.3.1. One-Week Lagged ΔSVI: Stock Performance Measures Comparison

When looking at stock SVI with a lag of one week, one of the metrics returned results that cannot be deemed robust. Therefore, $\Delta Close_{SVIstock}$ data on a consolidated level will not be analyzed further on a stand-alone basis, as to ensure the highest possible level of certainty in made conclusions.

5.3.2 $\Delta SVI_{company}$ and ΔVIX Effects

Focusing on $\Delta SVI_{company}$ instead resulted in all metrics being robust (see Panel A2). None of the data presented in these tables will be excluded from the base for analysis.

5.3.3 $\Delta Volume_{SVIstock}$

When primarily looking at how ΔSVI_{stock} data can serve as explanatory proxies for development in traded volumes, the coefficient of 0.0059*** tells us that there is a slight positive link between ΔSVI_{stock} and ΔVolume. As visualized in Graph A, $Volume_{SVIstock}$ is the ΔSVI_{stock} metric that returned the highest coefficient value. However, these numbers cannot be deemed to determine strong interconnectivity between the development metrics when it comes to co-movements compared to $\Delta SVI_{company}$ data. The R^2 value of 0.0039*** indicates that some ΔVolume variance can be described by ΔSVI_{stock}. However, this link is stronger in other examined metrics further developed below.

5.3.4 $\Delta Volume_{SVIcompany}$

Instead, when looking at the $SVI_{company}$ coefficient relationship, the value becomes considerably higher. With a value of 0.0169***, every point increase in $SVI_{company}$ means that we can predict an increase of ~1.7% * (volume std. deviations, see Panel B1 in appendix) in traded volume the following week. This value being a clear outlier is made evident by the visualization in Graph A. Despite this, the lower R^2 value of 0.0027*** points at less volume variance being described by changes in $SVI_{company}$ compared to the R^2 value of 0.0039*** that SVI_{stock} yields (see 5.3.1).

5.3.5 $\Delta Close_{SVIcompany}$

While the relationship between closing price and $SVI_{company}$ continues to lack strong evidence after tests, the values move in different directions when examining $\Delta SVI_{company}$ instead of ΔSVI_{stock}, as R^2 decreases from 0.0008*** to 0.0009*** while coefficient increases from 0.0001*** to 0.0003***. However, the overall close-price explanatory value in SVI can still be determined to be negligible at best, noted in Graph A by being positioned relatively close to origo.

5.3.6 $\Delta Vol10_{SVIstock}$

When examining volatility, a 10-day period in comparison to Google searches for stocks (ΔSVI_{stock}) shows higher explanatory power than closing price across all metrics. When looking at R^2 values in particular, the 0.0077*** R^2 is the highest of any examined metrics, which is visible in Graph A, meaning that a lot of the ΔVol10 variance can be described by ΔSVI_{stock}. Instead looking at coefficients yielded a 0.0027*** value, which is second-highest compared to other metrics in similar data groups.

5.3.7 $\Delta Vol10_{SVIcompany}$

Looking at Google searches for the company ($\Delta SVI_{company}$) instead of the stock, we can see that R^2 drops significantly from 0.0077*** to 0.0023***. This means that the explanatory power for variance in ΔVol10 more than triples when examining ΔSVI_{stock} as compared to simply looking at $\Delta SVI_{company}$. However, the higher coefficient that $\Delta SVI_{company}$ yields (0.0052*** vs 0.0027***) implies that ΔVol10 has almost twice the movement in the same direction when looking at $\Delta SVI_{company}$ as opposed to ΔSVI_{stock}.

5.3.8 $\Delta Vol30_{SVIstock}$

Increasing the examined volatility time period to 30 days, there are visible differences in a visible pattern regardless of looking at ΔSVI_{stock} or $\Delta SVI_{company}$. Initially, coefficients seem to drop significantly. In the case of company searches, coefficient values go down from 0.0027*** to 0.0007*** when moving from 10-day to 30-day volatility changes. Similarly, R^2 values go down from 0.0077*** to 0.0054*** when increasing the examined time period. This could hint to potential effects on stock performance that can be predicted by SVI, being temporary and not permanent. This is something that the article mentioned in the literature review strengthens, while that relationship was visible for impact on stock closing price in their study.

5.3.9 $\Delta Vol30_{SVIcompany}$

Looking at $\Delta SVI_{company}$, we can see that R^2 drops significantly from 0.0054*** to 0.0007***. This means that the explanatory power for variance in 30-day volatility almost increases 8-fold when moving from $\Delta SVI_{company}$ to ΔSVI_{stock}. However, the higher coefficient that $\Delta SVI_{company}$ yields (0.0007*** vs 0.0009***) implies that 10-day volatility is marginally more likely to move in the same direction as $\Delta SVI_{company}$.

5.3.10 Summary of Stock Performance measures When Compared to SVI Changes

When looking at the Graph A, which summarizes different ΔSVI's performances when it comes to stock performance predictability, a clear distinction between $\Delta SVI_{company}$ and ΔSVI_{stock} can be made. When it comes to R^2 values, all of the ΔSVI_{stock} numbers have higher R^2 values and thus higher explanatory power of stock performance variance measures than $SVI_{company}$ does. Within R^2 values, 10-day volatility (ΔVol10) numbers always end up in the top half of the highest values, regardless of analyzing Google searches for the company or for the stock.

Looking at coefficients, which give insight into which degree two metrics develop in the same direction, $\Delta SVI_{company}$ seems to generate higher coefficient values. This means that company name Google searches and stock performance metrics are probable to have a higher degree of correlative movements in the same direction than stock metrics do with ΔSVI_{stock}. When looking at the specific metric that generates the highest coefficients, the two clearly strongest metrics are 10-day volatility changes (ΔVol10) and volume changes (ΔVolume). While the degree of co-movements differs, all examined metrics have positive coefficient values that point to a positive relationship between Google searches and changes in closing price, volume, and volatility.

Conclusively, key takeaways that can be distinguished when looking at data on a consolidated, industry-agnostic, level show us that ΔSVI_{stock} can act as a relatively strong proxy for describing the variance in stock performance data (R^2). The stock performance metric that fits this behavior best, regardless of which type of ΔSVI is being examined, seems to be the changes in 10-day volatility.

Further, $\Delta SVI_{company}$ is more fit as proxies for describing the development of company stock metrics across the board. Within this, the strongest analyzed correlation seems to be evident in the relationship between ΔSVI_{stock} and the volume of said company's stock that is being traded. While this relationship differs in strength for different metrics, all examined relationships are positive, which means that stock metrics and Google searches tend to move in the same absolute direction.

5.4 Industry-divided Tests

5.4.1 ΔSVI_{stock} and ΔVIX Effects. One Week Lagged

Robustness tests for stock SVI on an industry level showed that Consumer, Tech, Real Estate, and Medical sectors returned non-robust data when analyzing ΔVolume data.

When it comes to ΔClose data, Consumer, Real Estate, Industrial, Financial, and Medical industries will be removed from the analysis.

Lastly, looking at Volatility change data, Consumer, Tech, and Medical data did not generate robust tests for neither 10-day nor 30-day windows. Additionally, the Material industry failed to generate robust data when looking at a 30-day period. All of these metrics will be excluded from the analysis.

5.4.2 $\Delta SVI_{company}$ and ΔVIX Effects, One Week Lagged

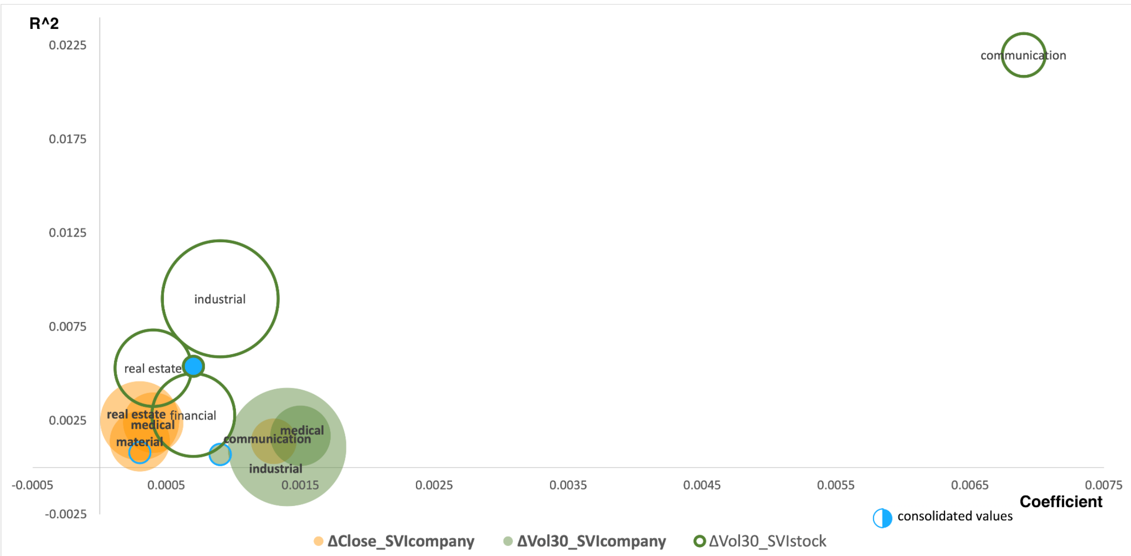
When looking at $SVI_{company}$, robustness for different metrics differed somewhat in comparison for stock SVI. For ΔVolume data, Communication, Tech, Real Estate, and Financial industries failed to generate robust data.

Looking at the change in closing price, Tech, Industrial, and Financial industries did not yield robust enough results to base analysis on.

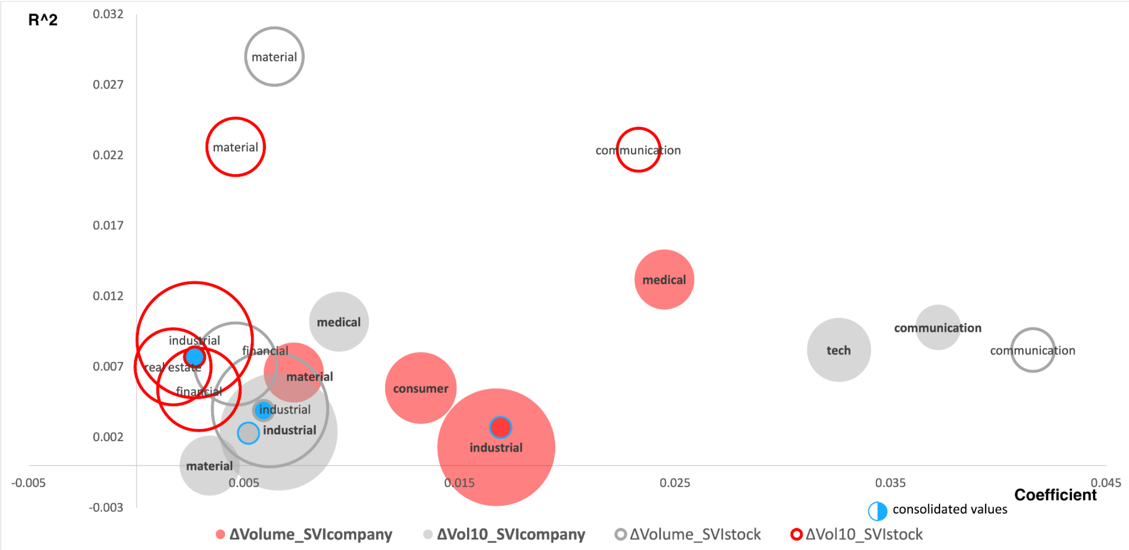
Volatility change data robustness requirements were not met by data in the Consumer, Real Estate, or Financial industries when looking at Volatility regardless of the length of the examined period. Additional metrics that fell short of statistical robustness for the 30-day Volatility change window were Communication, Tech, and Material industries. This means that for 30-day Volatility change data for $SVI_{company}$, the only statistically significant data points that are fit to base analysis on are the Industrial and Medical sectors.

5.5 ΔSVI Effects on Parameters. Industry-divided, One Week Lagged

When looking at metrics for different industries, it is important to keep both sample sizes and t-stat values in mind. These can be found in detail in chart Panel C. The below analysis is based on the face value of numbers presented in the above-mentioned chart.



Graph B: Visualization of ΔVol30 and ΔClose effects per industry. Size: n. Data significance at p≤5%



Graph C: Visualization of ΔVol10 and ΔVolume effects per industry. Size: n. Data significance at p≤5%

5.5.1 $\Delta\text{Volume}_{\text{SVIstock}}$

The tests between ΔSVI_{stock} and their potential explanatory power when it comes to predicting shares traded volume resulted in clear differences in values between different industries. With an R^2 value of 0.029**, the Material segment showed the highest explanatory power of traded volume variance, visable in the top left of Graph C. At the same time, the industry where this relationship was the weakest was Industrial, with an R^2 value of 0.004***. Important to keep in mind that the panel regression of all industries consolidated resulted in an R^2 value of 0.0039***. Splitting data up into industries resulted in 5/8 R^2 values that were ≥0.0039***. Looking at coefficients instead, the 0.0416*** coefficient value of the Communication industry shows that there is a much higher degree of correlation in this industry than in the consolidated data where the coefficient value was 0.0059***. On the other side, we can find a lower coefficient value of 0.0046*** in the Financial industry, pointing toward the correlation between ΔSVI_{stock} and ΔVolume being substantially weaker in the Financial industry.

5.5.2 $\Delta\text{Volume}_{\text{SVIcompany}}$

Analyzing the links between $\Delta SVI_{company}$ instead of company stocks and traded volumes instead puts the Medical industry at the top when it comes to R^2 with a value of 0.0132*, visible inte centre of Graph C. At the same time, the lowest visible R^2 value is seen in the 0.0013*** R^2 value of the Industrial sector. Both these values are substantially lower than the R^2 values of ΔSVI_{stock}, indicating that the overall explanatory power of ΔSVI when it comes to variance in ΔVolume is higher when looking at $\Delta SVI_{company}$ as compared to ΔSVI_{stock}. This is also something that is consistent with findings yielded from consolidated tests of all companies without industry segmentation. Focusing on coefficients instead shows us that the Medical industry has the highest value of 0.0245* which means that a 1-point increase in ΔSVI seems to predict a substantial increase of 0.0245*(std. deviations, see Panel B1 in appendix)) in traded volumes. While this is higher than the consolidated value of 0.0169***, it is important to note that they have somewhat different levels of statistical significance. However, this seems to indicate that the Medical industry sees a higher-than-average correlation between $\Delta SVI_{company}$ and traded volume of their shares. On the other side of the spectrum, the lowest coefficient values with statistical significance are derived from data gathered from the Material industry, with a coefficient value of 0.0073***. This makes the Medical industry the sector in which $\Delta SVI_{company}$ searches and traded volumes have the weakest linkage in development, albeit still having a positive relationship.

5.5.3 $\Delta Close_{SVIcompany}$

Moving from $\Delta SVI_{company}$ to ΔSVI_{stock}, the amount of significant data increases. Highest R^2 value can be seen in the Real Estate industry, and is at 0.0025**. The lowest R^2 value of statistical significance that can de be observed in the Communication and Material sectors and is 0.0014***. Both R^2 values are lower than the R^2 values for $\Delta SVI_{company}$ (0.0052*** & 0.0015***, see 5.3.1.3). This would indicate that there is larger explanatory power for ΔClose variance using ΔSVI_{stock} as a regressor instead of $\Delta SVI_{company}$ When it comes to coefficients, these are higher in general when the Google searches are for company names. The highest coefficient can be found in the Communication industry (0.0013***), and the lowest is shared by the Real Estate and Material sectors (0.0003***). While $\Delta SVI_{company}$ has less explanatory power for ΔClose variance, it has stronger positively correlated movement as coefficients are higher.

5.5.4 $\Delta Vol10_{SVIstock}$

Analyzing the ΔVol10 of company stocks and how SVIstock can help predict the volatility development shows us that the Material sector is the industry in which ΔSVI_{stock} has the largest explanatory value for ΔVol10 variance with an R^2 value of 0.0226***. The industry with the weakest explanatory value is the Financial sector with a substantially lower R^2 value of 0.0054***. The difference between sectors is rather substantial, and the large spread resulted in a 0.0077*** R^2 value in the consolidated analysis (see 5.3.6). They range from 0.0017*** in Real Estate to 0.0233*** in Communication when it comes to coefficients. Again, the consolidated data is much closer to the lower bound, with its 0.0027*** coefficient (see 5.3.6). This means the relationship between ΔSVI_{stock} and ΔVol10 is always positive, but this relationship differs very much in strength depending on which industry is being examined.

5.5.5 $\Delta Vol10_{SVIcompany}$

Staying within ΔVol10 but instead examining the predictability made possible by examining $\Delta SVI_{company}$, the R^2 values are considerably lower than when looking at ΔSVI_{stock}. Both the highest value of 0.0102*** (Medical) and the lowest value of 0.0021***(Material) are lower than the above-mentioned numbers for ΔVol10 R^2 in conjunction with SVI for stock searches. This would imply that the movement of ΔSVI_{stock} is a better predictor of the movements in ΔVol10 than $\Delta SVI_{company}$. When it comes to coefficients, the industry with the highest coefficient is Communication with a value of 0.0372**. The lowest value seems to be found in the Tech sector where the value is 0.0023***. These values are higher than the values visible for stock searches, indicating that ΔSVI for company searches has a higher correlation with ΔVol10 than ΔSVI_{stock}.

5.5.6 $\Delta Vol30_{SVIstock}$

Lastly, the ΔVol30 window gives somewhat different results when compared to the 10-day window. Both the highest (0.022***, Communication sector) and lowest (0.0028**, Financial sector) R^2 values are lower than the values seen in the ΔVol10 and ΔSVI_{stock} tests. This implies that the explanatory value of ΔSVI decreases as the volatility window increases in length. When examining coefficients instead, the highest value can be found in the

Communication industry with a value of 0.0069***, while the lowest value of 0.0004*** can be found in the Real Estate sector. Again, these are both lower than in the case of examining ΔVol10. This emphasizes the stance that effects on volatility lose explanatory power when the ex6mined time period grows in length.

5.5.7 $\Delta Vol30_{SVIcompany}$
As a result of low statistical significance, with only 3 metrics returning a ≤5% p-value, the explanatory value of $\Delta SVI_{company}$ on ΔVol30 can not be determined with satisfactory certainty.

5.5.8 Summary of Industry-divided Analysis of Effects of ΔSVI
Summarizing the use of Google SVI for stock performance in an industry-segmented environment, one can quickly divide the different metrics into two groups. This distinction can be made with R^2 and coefficient values as basis, since these values seem to be substantially higher for ΔVolume and ΔVol10 than they are for ΔClose and ΔVol30. This was also visible when conducting tests on a consolidated level, where the data was not divided per industry the company is active in.

Initially, taking a look at the somewhat weaker metrics (ΔClose & ΔVol30) shows that a large section of the data is centered around rather low R^2- and coefficient values. When trying to determine which industries might portray stronger explanatory values in ΔSVI, we can see that the Communication industry always seems to display stronger coefficients. Further, Industrial also seems to perform somewhat stronger when it comes to ΔSVI explanatory power for ΔClose and ΔVol30. Apart from that, we can clearly see that a majority of the different industries seem to generate results in line with consolidated values.

Looking at the data that generated results that indicated potentially stronger correlations with stock performance did so within mainly two metrics - ΔVolume and ΔVol10. Again, the main industry that stands out is Communication. While showing relatively high values for both R^2 and coefficients, the data is mainly an outlier on the x-axis, meaning coefficients are high in relation to other industries. Apart from Communication, other industries that generated clearly higher values than consolidated data did were the Material & Medical industries.

6. Discussion

6.1 Interpretation of the Results

6.1.1 $\Delta SVI_{company}$ vs $\Delta SVI_{company}$ as a Regressor

In the results, it is clear that the regressions including the regressor ΔSVI_{stock} generate more significant and explanatory results when it comes to R^2, while $\Delta SVI_{company}$ data as the regressor generates higher coefficient values. This is emphasized even more when analyzing the highly significant data points generated in tests between different industries against change in ΔVol10 and ΔVolume. Given the higher coefficient values, $\Delta SVI_{company}$ seems to be more fit as a predictive indicator for value of the parameters one week in the future. One general assumption that could lie behind the difference in metric value between the two data sets could be a completely different use case of the two types of Google searches. If company Google searches where to act as a medium for retail investors with a more long-term approach to investing then this could explain why the coefficient values returned higher, as the relationship between higher Google company searches and higher coefficients allow for more predictability at a reliable level. Continuing with this hypothetical example, retail investors that search on Google for company stock would have a more reactive and short term approach to buying or selling a particular stock, hence why that regressor's R^2 values are relatively higher across the board.

6.1.2 R^2: Better Explained by ΔSVI_{stock}

The phenomenon that all R^2 values are higher for ΔSVI_{stock} than for $\Delta SVI_{company}$ could be explained by the fact that when a stock receives attention, it becomes more widely searched. That does not necessarily imply a positive or negative relationship coefficient-wise. It rather explains that increased searches for the company is more predictive for the variance rather than direction of the movement in ΔVolume, ΔVol10, and ΔVol30 the following week. When a particular stock becomes the center of attention, it could be due to either very positive or very negative news. Potentially, it could be so that non-investors who search for the company stock on Google are not interested in buying or selling the stock - just interested in the news surrounding the stock price. If this is the case, it would explain why the coefficient is lower but the R^2 can explain variance more - there are still larger volumes traded but not necessarily due to the retail investors. Rather, it could be that the institutional or more sophisticated investors are trading the stock during events of high-impact news but that the Google searches coincide with this.

6.1.2.1 R^2 & ΔSVI_{stock}: a Case Study on Swedbank

Taking a look at Swedbank's SVI_{stock} , a clear peak during the last week of March 2019 is visible. When comparing this to $SVI_{company}$ data, we can see that the peaks do not align. However, the SVI_{stock} peak aligns with a sharp drop in share price during the end of March 2019. Upon further investigation, news articles released during the last week of March 2019 confirm that Swedbank, at that moment, were experiencing difficulties which resulted in large-scale media attention. This given example shows a relationship that could aid in understanding the integral mechanics, which leads to different metric values. If the relationship between stock Google searches and medially covered events holds, it could also explain why closing price data for SVI_{stock} failed to generate statistically significant results. As events that generate a lot of media attention usually lead to very steep share price increases or decreases, the data set as a whole results in a struggle to find correlation with movements in a certain direction. Theoretically, if half of all the events were negative and half of all the events were positive, then the measured link between stock SVI and closing price data would come back inconclusive in terms of acting as a proxy for closing price predictions. While the movements in share prices might be unpredictable, other metrics that do move in more predictable directions during medially noticed events are volume and volatility. This is due to the nature of the metrics, which can increase regardless of development in share price. For reference, please see Supplement C1-C4 in the Appendix.

6.1.3 Coefficient: Better explained by $\Delta SVI_{company}$

The phenomenon that all coefficients are higher for $\Delta SVI_{company}$ than for ΔSVI_{stock} can be explained by retail investors that are interested in getting to know the company, and their fundamentals, rather than purchasing the stock short term for a quick sell-off and are thus often holders of the stock. This is why a more positive relationship in the coefficient is noted. However, R^2 does not follow as strongly as it does for SVI_{stock}. It could also be that the investor that searches for the company is a long-term holder and thus regularly Googles the company to get updated on news. The short-term investor may be interested solely in the stock and is only interested during volatile times, hence it may not show as predictive coefficients, as well as being interested in movements in the stock, regardless of direction and hence rather explain the variance better than direction of movement. If Google searches for a company could act as a proxy for interest for a particular company, it would further explain the reason why the positive correlation seems to be relatively strong.

6.1.4 Industry Differences

For ΔVol10 and ΔVolume particularly, we see the strongest R^2 values and coefficients in the Communication, Tech, Material and Medical sector. The other sectors are closer centered to origo and do not experience as much deviation as the above-mentioned sectors have. One potential explanation for the Communication and Tech sector to be seemingly the most affected could be due to the recent hype in similar stocks, or the Tech and Communication sectors as a whole. Many firms in the Medical sector appear to be in roughly the same category and this may also be explained by Covid-19. As many Medical firms have gotten attention during the pandemic, it could mean that retail investors have opened up their eyes for this sector. It would also be consistent with the idea that these stocks only started to interest retail investors in the latter part of the sample, i.e. during Covid 2019-2022. This could explain why the numbers for the Medical sector show slightly lower explanatory values of the Communication and Tech sector, but still better than mean.

The Material sector's relatively strong R^2 values for $\Delta Volume_{SVIstock}$ and $\Delta Vol10_{SVIstock}$ compared to other sectors was slightly harder to understand as $\Delta Volume_{SVIcompany}$ and $\Delta Vol10_{SVIcompany}$ did not show nearly the same relative strength. The figures appear to contradict the notion that no retail investor would Google search for the company name without being interested in buying their stock, hence there would be no need for searching for the stock specifically. They would not be interested in Googling the stock for any other reason than buying the stock as the Material sector is characterized by a B2B business model. The results remain somewhat of a mystery but could potentially be explained by a smaller sample size, or that high-impact news specific to the Material sector have coincided with a stronger interest from the general public in seeing the stock price effects of the headlines.

For ΔClose and ΔVol30, we see rather weak results and predictable effects in virtually all industries in comparison to ΔVol10 and ΔVolume. The significance was overall lower, which left us with fewer data points to examine. However, in relative terms, the data appeared to coincide somewhat neatly. Many of the industries that were relatively strong in ΔVol10 and ΔVolume, were also relatively strong in ΔClose and ΔVol30. The most notable data point was Communication that was an outlier in the tests - also somewhat consistent with the previously discussed results. Most of the results for ΔClose and ΔVol30 we hence explain on a consolidated basis as there were not as many sector differences.

6.2 Potential for Further Research

It appears that stocks are affected by Google searches but not to the same extent that they have been in previous papers and in the United States. This could be due to a potentially larger retail market in the United States than in Sweden or due to the different time periods measured.

Something else that could explain this result is the size of the analyzed companies. In the reference paper, "In Search of Attention", data from all of the companies listed on the Russell 3000 index are included. This means that the relative amount of small companies in the dataset is higher, as all of the analyzed Swedish companies are listed as Large Cap corporations. If the assumption regarding smaller companies being more susceptible to ΔSVI holds, this becomes an important dynamic to keep track of. While the Russell 3000 index might be weighted, results in papers such as this one are not. This means that the results of a small company are just as important as the results of a larger company. Under the assumption that smaller cap stocks are more widely affected by retail investors, a higher relative share of small companies in a dataset leads to a higher representation of highly-correlative data. This in turn leads to overall results being more skewed towards a visible link between ΔSVI and stock performance metrics. This might be an explanation to the discrepancy between our results, and the one's in "In Search of Attention" (

To further develop this paper and explore potential areas of use, there is a wide range of research that still can be done. Much of this is related to the potential opportunities for creating viable trading strategies for hedge funds, looking at differences between other stock exchanges and potentially on a global level. Surely, there are considerable areas of retail investing's effects on stock markets that are yet to be explored. Hopefully, this paper will serve as an inspiration to further deep-dive into the intricacies of stock market effects from the internet era - a subject that is constantly evolving. As an example, one could explore stock investor forums further and deep dive into how specific forums such as Reddit (r/wallstreetbets) can affect stocks and if certain sites have stronger effects than others.

6.3 Result Actionability

Looking at potential practical use-cases of the results generated and conclusions drawn in this paper, it would be interesting to see if they could act as a base for developing a trading strategy. Based on which type of methodology that would be used for a trading strategy, different metrics discussed in this paper are of different value.

Starting off by examining which SVI metric to look at in order to be able to predict company stock performance would lead to $\Delta SVI_{company}$ becoming the most fitting option. This is due to the fact that no matter which metric to look at, $\Delta SVI_{company}$ always returned higher coefficients than ΔSVI_{stock} data. Since all of these coefficients returned positive and a week's lag was used to analyze data, a theoretical trading strategy could be to purchase securities that increase in value when volume, volatility or closing price increases. While this trading strategy would be the easiest to execute within would be $\Delta Close_{SVIcompany}$, taking long positions in stocks where $\Delta SVI_{company}$ increases and short positions in companies where $\Delta SVI_{company}$ decreases, it would be the least favorable since the coefficient within $\Delta Close$ was the lowest. Instead, the most favorable metric to pursue would be volume, since the coefficient generated is very high in relation to other metrics. However, this most likely requires the creation of a bespoke OTC product that would be relatively expensive which quickly would eat into the margins of the trade. Aiming to find a middle-ground that could potentially be both profitable and feasible to carry out in practice would lead attention toward the volatility, and taking straddle positions as $\Delta SVI_{company}$ increases as well as looking at taking short straddle positions as $\Delta SVI_{company}$ decreases. Although this strategy seems theoretically feasible, a pragmatic approach would most likely end up in most profits being rendered non-existent due to the fees required to take these positions.

Looking at which industry this type of trade would be the most beneficial in it leads attention to the Communication industry, as it returned higher coefficients than other industries regardless of examined metric. This would mean that the above-mentioned trades would be best fitted for the following Swedish Large Cap companies: Telia, Tele2, Sinch, and Millicom.

7. Concluding Remarks

Based on the analysis and discussion, we conclude that there is a small but distinct positive relationship between changes in Google searches and changing trading volume, closing price, and rolling 10-day rolling volatility, particularly on a one-week lagged basis. Albeit we see explanatory variables in other tests, the one-week lagged regressions are the ones with the highest explanatory power and strongest significance across the board after testing for robustness. When it comes to explanatory power of stock metric variance (R^2), Google searches for company stock (ΔSVI_{stock}) return the highest values, while Google searches for company names ($\Delta SVI_{company}$) yield the highest predictability for future development of stock metrics (coefficient).

Also on an industry-basis, the only time frame that was deemed worthy of further discussion was the one-week lagged regressions. This is since other time periods barely showed any significance or explanatory values in terms of R^2 and coefficients, which is why they were left out of the analysis and discussion. Overall, the Communication sector was proven to show the strongest relationship with regard to coefficient, followed by the Tech sector and the Material sector. When it instead comes to R^2 values, the Material and Medical industries displayed the highest explanatory values of stock performance metric variance when looking at ΔSVI.

Overall, this paper further establishes that there is an evident link between the ways in which companies and stocks are Googled and the behavior of said stocks on a publicly traded market. Driven by the increased prevalence of retail investors, more and more tools available to less sophisticated investors can be proven to have market-moving impacts on many of the largest listed Swedish corporations. Despite explanatory values being relatively low, rigorous testing has returned robust replicable results of high statistical significance. These results can further the discussion of how impactful different parts of society can be when it comes to topics that are usually assumed to be left to large institutional actors.

8. Appendix

8.1 Robustness Tests

Panel A1. ΔSVI$_{stock}$ and ΔVIX effects. Large Cap Consolidated

Panel A1	Δ Volume			Δ Close			Δ Vol10			Δ Vol30		
	Coefficient Δ SVIstock	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIstock	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIstock	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIstock	Coefficient Δ VIX	R2 Robust
No Lag	0.0012 [1.4737]	0.3348*** [5.7470]***	0.0017***	0.0000 [0.6502]	-0.0959*** [-44.500]***	0.1430***	-0.0004* [-2.1757]*	0.3259*** [15.443]***	0.0139***	-0.0002** [-2.9465]**	0.0918*** [14.647]***	0.0104***
One Week Lag	0.0026*** [-4.4612]***	-0.2045** [-2.9016]**	0.0013***	**0.0000** **[1.3033]**	-0.0094*** [-5.5027]***	0.0015	0.0015*** [5.3379***]	0.229*** [14.101***]	0.0087***	0.0006*** [5.7307]***	0.0774*** [12.243]***	0.01***
Two Week Lag	0.0007 [1.0730]	0.1962*** [4.3200]***	0.0006***	0 [-0.0263]	-0.0221*** [-14.536]***	0.0076***	0.0001 [0.4378]	0.0202 [1.1079]	0	0.0001* [1.9899]*	0.1109*** [17.354]***	0.0149***
Three Week Lag	-0.0005 [-1.0406]	-0.1141* [-2.2632]*	0.0002*	0 [-0.3924]	-0.0081*** [-4.5868]***	0.0010***	-0.0006** [-3.0023]**	-0.0132 [-0.9601]	0.0004**	0 [-0.6783]	0.0603*** [12.808]***	0.0044***
Four Week Lag	0.0011 [1.3495]	-0.0697 [-1.4803]	0.0002*	0 [0.8455]	0.0098*** [7.3746]***	0.0015***	-0.0001 [-0.5935]	-0.0049 [-0.3134]	0	0 [0.9367]	0.0645*** [14.177]***	0.0051***

Panel A2. ΔSVI$_{company}$ and VIX Effects. Large Cap Consolidated

Panel A2	Δ Volume			Δ Close			Δ Vol10			Δ Vol30		
	Coefficient Δ SVIcompany	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIcompany	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIcompany	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIcompany	Coefficient Δ VIX	R2 Robust
No Lag	0.0037 [0.7194]	0.3335*** [5.477]***	0.0017***	-0.0001 [-1.1045]	-0.0959*** [-44.518]***	0.1431***	-0.0021*** [-3.5617]***	0.3264*** [15.439]***	0.0141***	-0.0007*** [-4.3814]***	0.0919*** [14.628]***	0.0105***
One Week Lag	0.0120*** [4.2727]***	-0.2074*** [-2.9861]***	0.0020***	0.0002** [3.1583]**	-0.0095*** [-5.5242]***	0.0018***	0.0031*** [3.8442]***	0.2274*** [14.080]***	0.0075***	0.0006** [2.8395]**	0.0767*** [12.228]***	0.0073***
Two Week Lag	0.0065 [1.6465]	0.1954*** [4.3016]***	0.0009***	0 [0.4575]	-0.0221*** [-14.613]***	0.0076***	0.0007 [1.4621]	0.02 [1.1025]	0.0001	0.0002 [1.5813]	0.1108*** [17.317]***	0.0148***
Three Week Lag	0.0029 [1.0817]	-0.1135* [-2.3450]*	0.0003*	0 [0.4787]	-0.0081*** [-4.5858]***	0.0010***	-0.0013** [-2.8603]**	-0.0124 [-0.9032]	0.0002*	0.0001 [0.5142]	0.0604*** [12.826]***	0.0044***
Four Week Lag	0.0072 [1.7249]	-0.071 [-1.5273]	0.0006	-0.0001 [-0.9077]	0.0098*** [7.3192]***	0.0015***	-0.0015** [-2.5849]**	-0.0048 [-0.3042]	0.0002*	-0.0002 [-1.6399]	0.0644*** [14.1449]***	0.0041**

Panel A3. ΔSVI_{stock} and ΔVIX Effects per Industry, One Week Lagged

Panel A3	Δ Volume			Δ Close			Δ Vol10			Δ Vol30		
	Coefficient Δ SVIstock	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIstock	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIstock	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIstock	Coefficient Δ VIX	R2 Robust
Communication n=4	0.0309*** [6.0373]***	-0.3352* [-1.9762]*	0.0061	0.0309*** [6.0373]***	-0.3352* [-1.9762]*	0.0061	0.0190*** [7.1741]***	0.2314*** [4.0985]**	0.0203***	0.0065*** [9.2089]***	0.0777*** [7.1945]***	0.0255***
Consumer n=10	**0.0021** **[1.6365]**	-0.1743 [-1.7943]	0.0025	**0** **[-1.1489]**	-0.0138* [-2.1306]*	0.0027*	**0.0012** **[1.3857]**	0.0973* [2.5244]*	0.0022**	**0.0003** **[1.3474]**	0.0523** [3.1728]**	0.0034**
Tech n=8	**0.0026** **[1.4745]**	-0.2827** [-3.2485]**	0.0024***	0.0001* [2.5536]*	-0.0181*** [-4.4455]***	0.0065***	**0.0025** **[1.9030]**	0.2805*** [4.7860]***	0.0142***	**0.0006** **[1.9246]**	0.06*** [6.9026]***	0.0075***
Real Estate n=12	**-0.0001** **[-0.1516]**	-0.5014 [-1.5833]	0.0013	**0** **[-0.8289]**	0.0026 [-0.7216]	0.0004	0.0008* [2.0358]*	0.1812*** [6.7813]***	0.0067***	0.0003** [2.8135]**	0.0726*** [9.8115]***	0.0126***
Industrial n=27	0.0030** [2.6720]**	-0.1909 [-1.1482]	0.0013***	**0.0001** **[1.8524]**	-0.0141*** [-4.9671]***	0.0034***	0.0014** [3.1948]**	0.2454*** [8.4859]***	0.0104***	0.0007*** [3.6803]***	0.0771*** [9.3638]***	0.0134***
Financial n=14	0.0028*** [4.5726]***	-0.1059* [-2.0733]*	0.0035***	**0.0002** **[1.5012]**	-0.0047 [-0.9713]	0.0024	0.0016** [2.6144]**	0.2474*** [6.5287]***	0.0099***	0.0120* [2.1036]*	0.0850*** [4.9186]***	0.0098***
Material n=7	0.0051** [2.5843]**	-0.0186 [-0.5449]	0.0178*	-0.0001** [-3.0582]**	-0.0031 [-0.9331]	0.0017**	0.0037*** [3.3552]***	0.2100*** [8.4450]***	0.0207***	**0.0013** **[2.5037]**	0.558*** [4.6416]***	0.0205***
Medical n=7	**0.0013** **[0.5049]**	-0.0324 [-0.1973]	0.0002	**0.0001** **[0.7572]**	-0.0125* [-2.1485]*	0.0022	**0.001** **[1.1607]**	0.3608*** [4.2485]***	0.0118***	**0.0006** **[1.1398]**	0.1475** [3.0370]**	0.0140***

Panel A4. $\Delta SVI_{company}$ and VIX Effects per Industry, One Week Lagged

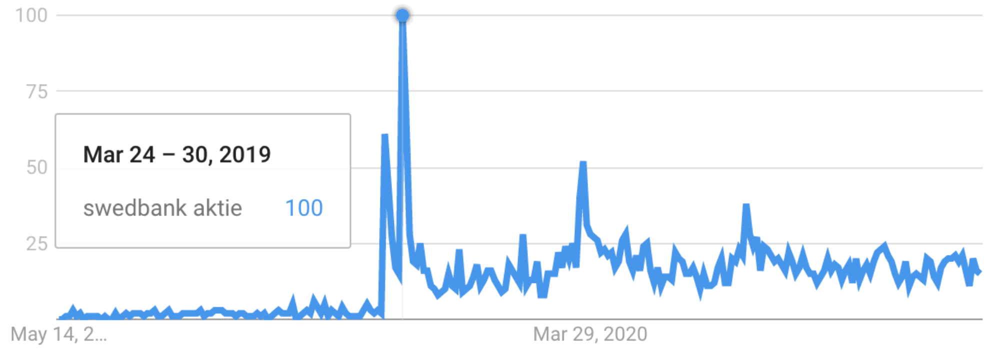
Panel A4	Δ Volume			Δ Close			Δ Vol10			Δ Vol30		
	Coefficient Δ SVIcompany	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIcompany	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIcompany	Coefficient Δ VIX	R2 Robust	Coefficient Δ SVIcompany	Coefficient Δ VIX	R2 Robust
Communication	0.0806 [1.8240]	-0.3352* [-2.1173]*	0.0069	0.001*** [8.8215]***	0.0083* [2.4531]*	0.0018***	0.027** [2.8007]**	0.2212*** [4.0158]***	0.0105***	0.0053 [1.2203]	0.0733*** [6.8993]	0.0087***
Consumer	0.0097*** [3.5442]***	-0.1815 [-1.8818]	0.0042***	-0.0001*** [-3.9874]***	-0.0137* [-2.1078]*	0.0026***	0.0012 [0.7732]	0.0947* [2.4776]*	0.0011*	0.0004 [0.6905]	0.0516** [3.1355]**	0.0028**
Tech	0.0623 [1.7038]	-0.2879*** [-3.3325]***	0.0065***	0.0002 [0.4930]	-0.0181*** [-4.4601]***	0.0048***	0.0255** [3.1527]**	0.2781*** [4.7877]***	0.0148***	0.0035 [1.7479]	0.0597*** [6.9021]***	0.0054***
Real Estate	0.007 [1.3846]	-0.493 [-1.5727]	0.0017	0.0002* [2.4278]*	-0.0022 [-0.6298]	0.0014*	0.001 [1.7120]	0.1795*** [6.6813]	0.0056***	0.0001 [0.4212]	0.0715*** [9.7504]***	0.0098***
Industrial	0.0109*** [3.4504]***	-0.1912 [-1.1589]	0.0010***	0.0002 [1.7488]	-0.0141*** [-4.9538]***	0.0033***	0.0043*** [5.2437]***	0.2453*** [8.6055]***	0.0091***	0.0011** [2.9270]**	0.0770*** [9.4774]***	0.0087***
Financial	0.02 [1.6891]	-0.1113* [-2.1950]*	0.0036***	0.0005 [1.2626]	-0.005 [-1.0325]	0.0009	0.0069 [0.8443]	0.2448*** [6.5556]***	0.0089***	0.003 [1.0997]	0.0841*** [4.9121]***	0.0095***
Material	0.0061*** [3.8670]***	-0.0392 [-1.2130]	0.0047***	0.0002*** [18.879]***	-0.0031 [-0.9200]	0.0011***	0.0021** [2.8956]**	0.1983*** [8.8841]***	0.0076***	0.0002 [0.6252]	0.0525*** [4.4923]***	0.0043***
Medical	0.0188* [2.0120]*	-0.298 [-0.1776]	0.0078	0.0003* [2.2597]*	-0.0125* [-2.1497]*	0.0032**	0.0067*** [5.4982]***	0.3619*** [4.3354]***	0.0164***	0.0010* [1.9912]*	0.1478** [3.0379]**	0.0133**

8.2 Supplements

Table B1: Standard Deviation, σ of Samples

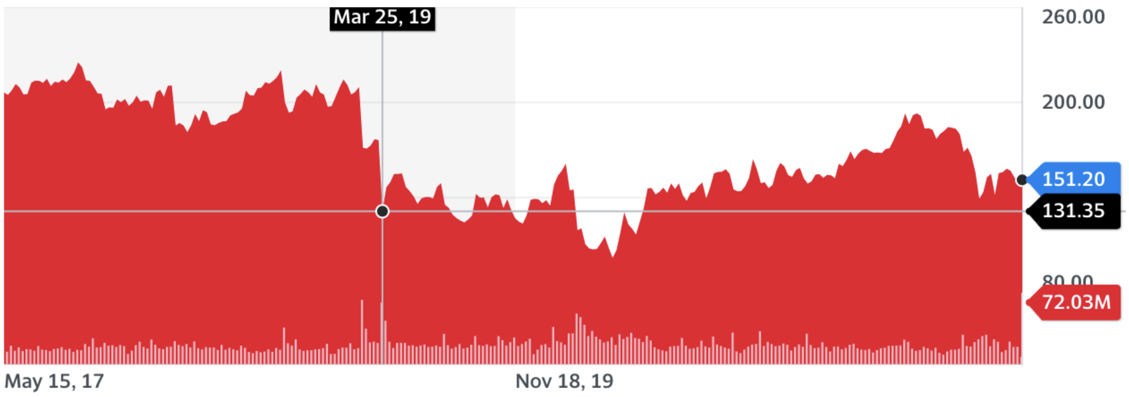
One Week Lag	$\Delta SVI_{company}$	$\Delta SVI_{company}$	ΔVolume	ΔClose	ΔVol10	ΔVol30
Consolidated	3.765	17.172	1.726	0.048	0.510	0.160
Communication	1.470	3.543	1.642	0.051	0.551	0.168
Consumer	5.659	17.880	0.999	0.051	0.582	0.188
Tech	1.494	14.404	1.329	0.049	0.533	0.168
Real Estate	7.709	21.612	2.638	0.042	0.458	0.135
Industrial	3.765	17.172	1.726	0.048	0.510	0.160
Financial	1.840	12.804	0.700	0.044	0.504	0.172
Material	6.219	14.676	0.557	0.044	0.457	0.151
Medical	6.888	15.016	1.465	0.053	0.639	0.247

Supplement C1: Swedbank SVI$_{stock}$: Google Searches for "Swedbank Aktie".



Supplement C2: Swedbank SVI$_{company}$: Google Searches for "Swedbank".

Supplement C3: Swedbank Stock Price and Trading Volume



Supplement C4: Swedbank Headline from News Article Posted During the End of March 2019

BANKS MARCH 28, 2019 / 11:35 AM / UPDATED 3 YEARS AGO

Swedbank dumps CEO as money laundering claims spook investors

By Esha Vaish, Johan Ahlander 5 MIN READ

STOCKHOLM (Reuters) - Swedbank dismissed Birgitte Bonnesen as chief executive on Thursday only an hour before its annual meeting after disgruntled shareholders rounded on her handling of a rapidly-growing money laundering scandal.

8.3 References

BARBER, BRAD M., AND TERRANCE ODEAN, 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors, *Review of Financial Studies* 21, 785–818

BAKER, MALCOLM, AND JEFFREY WURGLER, 2007. Investor sentiment in the stock market, *Journal of Economic Perspectives* 21, 129-151

CHOI, HYUNYOUNG, AND HAL VARIAN, 2009. Predicting the present with Google Trends, Working paper, Google Inc.

CHORDIA, TARUN, SAHN-WOOK HUH, AND AVANIDHAR SUBRAHMANYAM, 2007. The Cross-section of Expected Trading Activity, *Review of Financial Studies* 20, 709–740.

DA, ZHI, JOSEPH ENGELBERG, PENGJIE GAO, 2011. In Search of Attention., *The Journal of Finance* 66, 1461-1499

FOUCAULT, THIERRY, DAVID SRAER, AND DAVID J. THESMAR, 2011. Individual Investors and Volatility, *The Journal of Finance* 66, 1369-1406

LINDVALL, JULIA, 2021. Omfattande Aktiefusk i Chattrum Utreds av Ekobrottsmyndigheten, *SVT Nyheter*, 22 Dec

NAGARAJAN, SHALINI, 2021. The Hedge Fund Badly Bruised by Betting Against GameStop is Still Struggling After Ending the First Half with a 46% Loss, Report Says, *Business Insider,* 9 Jul

PONCZEK, SARAH, 2021. WallStreetBets Gains More Than a Million New Members Overnight, *Bloomberg,* 28 Jan

STATISTISKA CENTRALBYRÅN (SCB), 2021. Aktieägarstatistik. https://www.scb.se/hitta-statistik/statistik-efter-amne/finansmarknad/aktieagarstatistik/aktieagarstatistik/

TETLOCK, PAUL C., 2007. Giving content to investor sentiment: The Role of Media in the Stock Market, *The Journal of Finance* 62, 1139– 1168

TT, 2022. Då Öppnar Placera-forumet Igen, *Dagens Industri*, 21 Jan

ÖSTERBERG, TOBIAS, 2022. Aktieforum Rensas upp efter Abrupt Stängning, *SVD Näringsliv*, 21 Jan