Erasmus University Rotterdam

Erasmus School of Economics bachelor's thesis

Google Search Volume: Explaining and Predicting Returns for the European Stock Market

Student Name: Randrick Croes

Student ID number: 614444

Supervisor: Dr. Sjoerd van den Hauwe

Second assessor: Wei Hou MSc

Date: 1-07-2024

Abstract

This study examines the relationship between investor attention measured through the Google Search Volume Index (SVI) and stock returns of the European market. The SVI is measured by the search volume of company tickers on Google. The sample consists of 50 randomly chosen stocks from the STOXX 600 Index. The SVI is obtained monthly from the Google Trends database for the period between January 2004 and December 2023. A descriptive and predictive regression model is used to investigate the hypotheses. The results indicate that there is no significant impact of Google searches on stock returns for both models. Although the relationship is insignificant, it is worth mentioning that the abnormal SVI has a negative relationship with abnormal stock returns. This negative relation is in line with previous studies such as Da et al. (2011) who imply that internet searches cannot predict future stock returns. The findings of this thesis have important implications because it adds a new sample in a different setting, and it helps to fill in the gap for literature study on this topic for the European market as a whole. Limitations of this thesis includes a sample size of 50 stocks from the STOXX 600 Index. It could limit the generalizability of the findings. Future research could explore specific business-sectors in order to further investigate the attention-effect on stock returns.

Table of Contents

Chapter 1. Introduction

1.1 Research Problem and Motivation

How do investors react to the news? This has been a question that has been discussed for decades. Now with the ever rise of technological advancements and social media, news can be received instantaneously. A lot of the news in the world of stock market seems to circulate around the stock market of the United States. This is no surprise since the United States stock exchanges such as the New York Stock Exchange and Nasdaq dominates the stock world as the two biggest exchanges. They cover around 42% of the global market capitalization. But on the number three spot we have the Euronext (Neufeld, 2023). Although the United States has the biggest and leading stock market, the European market offers something different and has a lot of room for growth (Financial Times, 2024). This potential for growth of the European market obviously comes with a lot of attention. Attention in the form of news that could possibly alter investment decisions of investors and returns of certain stocks. The European market is an interesting context to study the effect of news on the stock market returns, because the European exchange is the third largest exchange in the world. Hence, it is an important market for further research. Therefore, the research question that this thesis aims to answer is: *How does attention-based news affect stock market returns for the European market?*

1.2 Related Literature and Contribution

Researchers, in particular the fields of Finance, have been interested in modeling and explaining this phenomenon. Early studies such as Barber and Odean (2008) show that attention towards certain stocks does indeed have an effect on buying and selling behavior of individual and institutional investors. They measure attention grabbing indirectly using three proxies which are, abnormal daily trading volume, extreme returns, and whether the firm appeared in that day’s news or not. Their evidence supports that attention is primarily related to abnormal trading volume. Da et al. (2011) expanded on the literature of Barber and Odean (2008) by providing a way to measure daily attention paid to stocks directly, since Barber and Odean (2008) could not manage a way to do this. Da et al. (2011) measure attention-grabbing events directly by using search frequency in Google (SVI). They suggest a positive predictive power of Google SVI on stock returns of companies in the Russel 3000 index. Bijl et al. (2016) findings however document the negative predictability of the Google SVI in the S&P 500 companies.

While these studies mostly focus on the United States, some researchers have tested the idea in other countries as well. Bank et al. (2011) for example studied the effect of search queries on the German market and they indicate that a rise in search queries leads to more trading activity. For the French stock market Aouadi et al. (2013) documents that investor attention is correlated with trading volume. Takeda and Wakao (2014) find that online search activity is strongly correlated with trading volume, but weakly correlated to stock returns in the Japanese stock market. In the Norwegian stock market however, the Google search activity does not predict stock returns (Kim et al., 2019). On the other hand, Google searches do explain and predict trading volume in the Norwegian stock market. Most of these studies were based on European countries. However, to this date the European stock market as a whole has not been studied through this lens.

A study that comes close to the concept of measuring a continent or multiple countries was conducted by Tantaopas et al. (2016). They study the relationships between investor attention and return, volatility, and trading volume from selected Asia Pacific equity market and find causality mostly from market variables and attention. As you can see existing literature is largely based on evidence on individual countries and focused on the U.S. markets. This paper tries to fill in that gap by focusing on the European market.

The main finding of this paper is, that attention measured by Google searches neither explains contemporary nor predicts future stock returns in the European market. The latter finding is in accordance with previous studies such as Da et al. (2011) who imply that internet searches cannot predict future stock returns. The result of this study provides insight that can help the decision-making process of investors and firms that invest in the European stock market.

1.3 Research Outline

The remainder of the thesis will be explained as follows. In Chapter 2, a thorough review of the literature will be performed on the past studies done on the topic of the effect of attention on the stock market. Based on past studies, the two hypotheses will be formulated. In Chapter 3, the data will be explained. After that, the research methodology will be presented in Chapter 4. This will include a detailed explanation of how the hypotheses will be tested as well as testing the necessary assumptions. The results of this analysis will be displayed and explained in Chapter 5, (where the summary statistics of the findings of this paper will be presented along with a discussion of the regression results). Lastly, Chapter 6 will consist of the conclusions that can be drawn from the evidence provided in this paper followed by the research implications, limitations, and suggestions for future research on this topic.

Chapter 2. Literature Review

This literature review will analyze the existing research done in the areas of attention effect and Google search volume as measure of attention. The existing literature will assist in the development of the two hypotheses that this thesis aims to test.

2.1 Attention and proxies for attention

Before studying how attention affects stock return it is important to understand what attention is and the proxies that have been used over the years to measure attention. One early definition of attention that is still relevant for today is provided by Kahneman (1973), who stated that attention is a scarce cognitive resource. This definition of attention disproves the assumption of the traditional asset pricing model, which assumes that all information is accounted for in the price. This assumption implicitly supposes that investors have all information about the market. This is inconsistent with Kahneman (1973), since investors are unable to consume all information due to time constraints.

In addition, Barber and Odean (2008), explains that when there are many alternatives, we are more likely to choose options that attract our attention, and we usually ignore the options that do not seem appealing to us. When we evaluate each option critically, attention can be beneficial. If not, attention may lead to suboptimal choices and results. In their paper, Barber and Odean (2008) introduces three direct proxies for measuring attention.

The first proxy they introduced is abnormal trading volume, since heavily traded stocks must be attracting investor's attention. The second proxy is extreme one-day returns since- whether good or bad- these are likely to occur simultaneously with attention-grabbing events. The third direct measure of attention the used is and whether a firm appeared in that day's news or not. Suggesting that when a firm is in the news it is more likely to attract attention to that stock. This buying pattern seems consistent with the media effect discussed by Fang and Peress

(2009), who explains that individuals' buying pressure temporarily pushes up the prices of attention-grabbing (in-the-news) stocks, but such pressure subsequently reverses.

Aboody et al. (2010) empirically tests whether limited attention can explain the phenomenon of stocks with the strongest prior 12-month returns experience a significant larger return during the five trading days before their announcement compared to the five trading days after the announcement. They find that during pre-announcement periods small and medium-sized traders evidence a significantly positive abnormal order imbalance, but large traders do not. After the earnings announcements the small and medium-sized traders' positive abnormal imbalances disappear. This indicates that the past returns of a particular stock can act as proxy for individual attention.

Another popular proxy for individual attention is the nearness to extreme prices suggested by Li and Yu (2012). In their study Li and Yu (2012), provides explanation for the predictive power for future returns of the nearness to extreme prices. This finding supports the study of Huddart et al. (2009), who report that extreme prices in a stock's past price path affect investor trading decisions in equity markets. Thus, capturing investors' attention.

These are only a few examples of a bunch of proxies used in prior literature to capture investors' attention. Here is a brief summary of other relevant proxies: Analyst coverage (Hirshleifer et al.2013); changes in advertising expenses (Lou, 2014); search traffic on the Electronic Data Gathering, Analysis and Retrieval (EDGAR) system (Lee et al. 2015) (Drake et al. 2015) (Drake et al. 2017); and Google search volume (Da et al. 2011) which will be further discussed in the following paragraph.

2.2 Google as a proxy for attention

With the rapid rise of technology and digitalization of the world, the internet has become one of the primary sources that people search for information. Of these online search engines Google has a cut above all, with a market share of 90.8% during May 2023 and May 2024 (Statcounter Global Stats, n.d.) of the reasons that Google is used for research purposes is that its data is public and accessible. The study of Ginsberg et al. (2009), introduced the new direct measure of attention besides the classical indirect measures, which is the Google's Search Volume Index (SVI). The SVI is a proxy of investor attention that portrays search volume of a selected keyword. The first benefit of the SVI is that it reflects a precise query from a particular

Google user. Second, the database is expanding everyday due to the large amount of internet usage. Finally, Da et al. (2011) finds that SVI captures attention more accurately compared to the traditional indirect proxies for attention (Barber and Odean, 2008).

2.3 Google as predictor of stock return

Several studies have aimed to investigate the relationship between attention effect and stock returns. The study of Da et al. (2011) measure attention-grabbing events directly by using search frequency in Google (SVI). They suggest a positive predictive power of Google SVI on stock returns of companies in the Russel 3000 Index. They found that an increase in SVI has a positive relationship with stock returns for the following two weeks, after the two weeks the prices bounce back to normal. These results are in line with those of Barber and Odean (2008), who found that an increase of investor attention leads to a temporary rise in price and eventual reversal. Joseph et al. (2011) finds similar results by using search volume on stock returns of companies in the S&P 500. In the short-term returns are positive and significant, but in the medium to long term horizon search volume seems to be negatively correlated with stock returns. Bank et al. (2011) also came to the same results when studying the relationship between Google search volume in the German market. They found that an increase in SVI is associated with temporarily higher future returns.

According to Adachi et al. (2017), they investigated the relationship between investor attention and stock price movements on the Japanese market. They found that a positive relationship exists between Google’s search volume and the stock price. Unlike the previously mentioned studies that reported a price reversal after a sudden increase in stock price, these prices did not revert to the old price therefore demonstrating the possibility of positive stock returns in the long term. Similar results are obtained in the paper of Takeda and Wakao (2014) also conducted in Japan.

On the other hand a study conducted by Bijl et al. (2016) between SVI and the companies listed on the S&P 500 Index, found that high Google search volumes are followed immediately by negative returns. One possible explanation they provide for predicting negative returns instead of positive return is that Google searches is now incorporated into the market faster and therefore weekly data allows them to detect only the subsequent negative returns. Bijl et al. (2016) are insinuating that because they are using newer data[1]. Their data is incorporated faster into the market in comparison to earlier studies.

Kim et al. (2019), investigate whether Google search activity can explain and predict activity in the Norwegian stock market. They focus on predicting three variables which are stock returns, trading volume, and volatility. They found that Google searches can forecast trading volume and volatility, but not returns. Similarly, the study of Tantaopas et al. (2016) found no causality of SVI for stock return in the Asia-Pacific stock market.

Nguyen et al. (2019) studies the relationship between Google search and stock returns for 5-emerging markets namely, Indonesia, Malaysia, Philippines, Thailand, and Vietnam. The authors found that an increase in Google searches increases the annual stock returns of the Malaysian stocks, and it reduces annual stock returns in the Philippines, Thailand, and Vietnam. For the Indonesian stock market however, no relationship between SVI and stock return was found.

2.4 Hypothesis Development

Upon reviewing the existing literature, it is worth noting that there have been mixed results on the relationship between Google search volume and stock returns. Some studies[2] conclude that there is no relationship between the two. Other studies[3] found a negative correlation between Google search volume and stock returns. Nonetheless, the positive relationship between a Google search volume and stock return seems to be the most prevalent[4].

As covered in the previous paragraph, most of the existing literature is based on evidence on individual countries. There is recent research however, conducted by Costa et al. (2024) that studies the connection between Google searches and the Euronext stock returns which consists of namely, Amsterdam, Brussels, Lisbon, and Paris. Their evidence suggests Google searches had

[1] Bijl et al. (2016) base their data on the period 2008-2013 compared to earlier papers i.e. Da et al. (2011) and Joseph et al. (2011) that use data from the period 2004-2008.
[2] The study of Tantaopas et al. (2016) and Kim et al. (2019) both came to the result that there is no correlation between SVI and stock returns.
[3] The study of Bijl et al. (2016) and Nguyen et al. (2019) both came to the outcome that there is a negative connection between SVI and stock returns.
[4] The study of Da et al. (2011), Joseph et al. (2011), Bank et al. (2011), Takeda and Wakao (2014), and Adachi et al., (2017) all came to the result that there is a positive relationship between SVI and stock returns.

no significant impact on Euronext stock Returns. Although this study covers multiple European countries it is not as comprehensive as one that examines the relationship between SVI and the European market represented by STOXX 600 Index. This thesis aims to fill this gap by studying the European market as a whole. Considering that many prior studies corroborate a positive relationship between a higher SVI and increase stock return. Therefore, I propose the following two hypotheses**:**

H1: There is a positive significant relation between the Google searches and the contemporary stock returns of the stocks listed on the STOXX 600 Index.

H2: There is a positive significant relation between the Google searches and the future stock returns of the stocks listed on the STOXX 600 Index.

Chapter 3. Data

3.1 Search Volume

In this research, I analyze the attention of investors by using data drawn out from Google Trends. The Google Trends search engine provides a way to access data and search term frequency that goes back to January 2004. For the purpose of this study the monthly Search Volume Index (SVI) was used for individual stocks. The index ranges between a score of 0 and 100 for each search term. A score of 0 means that there were not enough searches for this term and a value of 100 is the peak popularity of the term. The Search volume Index (SVI) represents a search interest relative to the total number of searches within a selected time frame. According to Takeda et al. (2014) the "relative" in search interest can be defined in two ways: time-series and cross-sectional. For time-series, the SVI for a particular month may differ over a given period because the SVI gives the value 100 when the number of searches is highest over the period specified in the search condition. For cross-sectional, the SVI does not increase when the number of searches of a particular keyword is less than that of another keyword. Additionally, Da et al. (2011), considers the possibility that SVI of the same keyword could be different at different point in times, since Google calculates SVI form a random subset of the actual historical search data[5]. I downloaded the data in June 2024 and used it in its original form.

3.2 Stock Identification

The next challenge concerns the identification of a stock in Google. To solve this problem the identification method of Da et al. (2011) is followed. In the study of Da et al. (2011) they identify a stock by using its ticker as the search query. This technique avoids including irrelevant components of the search volume index. On the other hand, one could argue why not just use the company names as search queries as a proxy for measuring attention. This method is called the "naïve approach". In the study of Bank et al. (2011) and Vlastakis et al. (2012) they use the "naïve approach" to capture the attention of Google users. They use company names as search queries to account for the search volume. Although I think this is another method to identify stocks, I would rather lean on the point of view of Da et al. (2011), who argue that searching only with company names could be problematic, due to two points.

The first reason would be that investors or Google users would use company name for reasons unrelated to investing. For example, in Appendix 1 the first stock mentioned is Adidas, one may search "Adidas" for online shopping purposes rather than to collect financial information about the company. This problem gets worse if the firm name has diverse meanings. To give an example, in Appendix 1 there is a company named Frontline from Norway. When a Google user searches for the word "Frontline" it could either indicate that they are interested in the financial information of the company, or it could mean that they are interested in a military terminology.

The second reason is investors may search for the same firm using different variations of its name. For example, ABN AMRO Bank is given the company name of "ABN AMRO Bank" in Eikon DataStream. Nonetheless, investors may search for the firm name in Google using different search queries, i.e. "ABN", "ABN AMRO", "ABN Bank" or "ABN AMRO Bank". To avoid and eliminate such random noise I opted to use only stock tickers. Additionally,

---

[5] Da et al. (2011) finds that the correlations of SVIs downloaded at different points of time are greater than 97%.

identifying a stock by its ticker also removes the problem of multiple reference names. Because a firm's ticker is always assigned uniquely.

With respect to other search conditions, I use the data on "all categories"[6] of "web search" conducted in "worldwide"[7] (location) for the period "January 2004 and December 2023." I downloaded the monthly data for 240 sample periods over these 20 years.

3.3 Stock Data and Sample selection

To capture the effect of attention-based news for the European market, I collected data for the STOXX 600 index. The STOXX 600 Index contains 600 largest European companies spanning over 17 countries. It captures more than 90% of the European investable market. In this paper, however not all 600 stocks will be observed, but only a sample of 50 stocks from the STOXX 600 Index. This is due to the fact that extracting data from Google trends is time consuming. To keep the validity of the study, a random sample of 50 stocks was chosen. For each stock in the sample, I manually draw the corresponding time series of Internet search activity for the period between 1st January 2004 to 31st December 2023. This period covers 20 years' worth of data, including some rough financial times, in particular the financial crisis of 2008 and more recently the Covid-19 pandemic.

Furthermore, the Eikon DataStream was used to extract data regarding the constituent of the STOXX 600 Index. As previously reported, only 50 stocks were used in this study. To determine which stocks these were going to be, a random sample was performed over the 600 stocks. The 50 stocks that were selected as the sample were drawn out using Excel. Appendix 1 presents the list of the 50 stocks remaining after the random sample, along with the corresponding stock tickers, country, and applied search queries.

[6] For category "all categories" was chosen instead of only the category "finance" to prevent bias or missing information.

[7] The location "worldwide" was chosen because there was not an option for Europe.

3.4 Stock Variables

This section will explain how the variables related to stock and investment attention are obtained or calculated. The stock variables 'Price', and 'Volatility' are all obtained from the Eikon DataStream. For the variable 'Volume' the historical data of Yahoo!Finance was used to extract the data. Next to these variables that I extracted from databases there were also three variables that need to be calculated using formulas. Firstly, the simple approach (Spaulding, 2002, Chapter 12) was applied to the variable 'Ret' which represents the rate of return for a particular stock:

$$Ret_{i,t} = (P_{i,t} - P_{i,t-1})/P_{i,t-1} \tag{1}$$

where:
$P_{i,t}$ = stock price of the current month
$P_{i,t-1}$ = stock price of the previous month
$Ret_{i,t}$ = rate of return

Secondly, the variable 'Abn Ret' which stands for abnormal return is calculated by following the characterized-adjusted returns method of Brown et al. (1985). Where the 'Abn Ret' is determined by the actual return minus the market index return. For this paper the market index is obviously the STOXX 600 Index. In order to determine the return for the market index the simple approach (Spaulding, 2002, Chapter 12) was applied to the variable 'RMI':

$$RMI_t = (RI_t - RI_{t-1})/RI_{t-1} \tag{2}$$

where:
RI_t = return index of the current month
RI_{t-1} = return index of the previous month
RMI_t = rate of return for the market index

The next step is to calculate the abnormal return. I define abnormal return for stock i traded on month t, Abn $Ret_{i,t}$ to be

$$Abn\ Ret_{i,t} = Ret_{i,t} - RMI_t \tag{3}$$

where $Ret_{i,t}$ the rate of return for stock i traded on month t as reported in the Eikon Datastream for the STOXX 600 Index sample of 50 stocks. The t ranges from 1 through 240[8] since the sample period is from January 2004 up until December 2023. $RMI_{i,t}$ is the rate of return for the market index.

The last variable that needs to be calculated is ‘Abn Volume’ which represents the abnormal trading volume of a particular stock. I will follow the method of Chordia et al. (2007) to standardize trading volume. The first step is to calculate the abnormal trading volume. Thus, I define abnormal trading volume for stock i on month t, Abn Trading $Volume_{it}$ to be

$$\textit{Abn Trading Volume}_{i,t} = \textit{Trading Volume}_{i,t} - \textit{Mean Volume}_{i} \tag{4}$$

where Trading $Volume_{it}$ is the quantity of shares for stock i traded on month t as reported in the historical data of Yahoo!Finance for the STOXX 600 Index sample of 50 stocks. Mean $Volume_{i}$ is the average trading volume across the whole sample period for a particular stock i. The next step is to calculate the Standardized Abnormal Volume using the following formula:

$$\textit{Abn Volume}_{i,t} = \textit{Abn Trading Volume}_{i,t} / \textit{Standard Deviation Volume}_{i} \tag{5}$$

where the Standard Deviation $Volume_{i}$ is the standard deviation for a stock i in the sample period. Table 1 explains all the main variables used in this paper.

Table 1: Variable Definitions

Variable	Definition
Variables from Google Trends	
SVI	Aggregate search frequency from Google trends based on stock ticker
ASVI	The log of SVI during the month minus the log of median SVI during the previous 2 months
Variables related to investment attention/ sentiment	
Price	Price of Stock

[8] The t ranges from 1 to 240 because the period covers 20 years which is 240 months.

Ret	Rate of Return as in (Spaulding, 2002, Chapter 12)
Abn Ret	Characterized-adjusted return as in Brown el al. (1985)
Volume	Trading volume in number of shares
Abn Volume	Standardized abnormal volume as in Chordia, Huh, and Subrahanyam (2007)

3.5 Stock Currencies

As you can imagine working with data containing multiple countries you would encounter different currencies. For the STOXX 600 Index sample of 50 stocks there were 8 different kinds of currencies. The majority of companies data were given in Euros, 58% to be exact. The other 42 percent was divided between these 7 currencies: Pound Sterling, Swiss Franc, Polish Zloty, Danish Krone, Slovak Koruna, and US Dollars. Table 2 provides a summary of the currencies along with their respective quantity of companies and percentages based on the sample. I converted the 7 currencies all to Euro's[9] using the flexible exchange rate found in Eikon DataStream for the sample period. I opted for the flexible exchange rate to better capture the price movements instead of opting for a fixed exchange rate. The reasoning behind only using one currency, in this case the Euros, is to mitigate the effect of monetary policy changes across countries with different currencies. Because according to Ioannidis and Kontonikas (2008), monetary policy changes affect contemporary and future stock returns.

[9] I converted all currencies to Euros since this study is focused on the European Market and most of the companies are valued in Euros (see table 2).

Table 2: Overview of the Currencies.

This table considers the currencies with their respective symbol, number of quantities, and fraction based on the STOXX 600 Index sample of 50 stocks.

Currencies	Symbol	Quantity of Companies	Fraction
Euro	€	29	58%
Pound Sterling	£	7	14%
Swiss Franc	₣	6	12%
Polish Zlotly	**zł**	3	6%
Danish Krone	kr.	2	4%
Slovak Koruna	Sk	2	4%
U.S. Dollars	$	1	2%

3.6 Descriptive Statistics

The sample contains data on 50 stocks listed on the STOXX 600 Index over the 20 years for a total of more than ten thousand observations for each variable. The total observations vary for each variable because when analyzing the data I found that 22 of the 50 stocks have incomplete data. See appendix 10 for an overview of the 22 stocks with their missing variables. To give an example of missing values, for a certain stock the data begins from June 2012 instead of the start of the sample period which is January 2004. To make sure that the missing values do not alter the conclusions of this paper, I performed a robustness check with the 28 stocks that do have complete data. The results of this robustness check are discussed in paragraph 5.3.

In table 3, the complete summary statistics can be found. These statistics provide for each variable a summary of the distribution, means, and standard deviation. The variables: abnormal SVI, abnormal return, and abnormal volume, are all calculated variables (see formula 3 (for abnormal return), 5 (for abnormal volume), 6 (for ASVI) given in proportions. The abnormal SVI has an average value of 0.4% meaning that the SVI of a current month t for a particular stock i is slightly higher than the median SVI of the prior 8 months.The abnormal return has a mean value of -2.5%. This means that on average a stock underperforms the market index, which is the STOXX 600 Index, by 2.5%. The abnormal trading volume has a mean value of 0.769, which means that the trading volume is 0.769 standard deviations above the average trading volume. A higher value (e.g. a maximum value of 94.038) indicates a period of uncommon higher trading activity. The volatility is directly retrieved from the Eikon DataStream and is also stated in proportions. In the sample, the average volatility is 31,9%. This is generally considered a high volatility.

With a sample size of more than 10 thousand observations, it can be concluded that the sample size is large enough to apply the Central Limit Theorem. This theorem states that when an adequate large sample size is used the sample would be approximately normally distributed. Therefore, the assumption that a regressions analysis must be a normal distributed sample is satisfied.

Table 3: Summary statistics for the 50 sample stocks.

	N	Minimum	Maximum	Mean	Standard Deviation
ASVI	10998	-0.352	0.745	0.004	0.070
Abn Return	10691	-0.712	1.855	-0.025	0.079
Abn Volume	10738	-0.832	94.038	0.769	4.513
Volatility	10726	0	1.208	0.319	0.159

Note: *This table displays the descriptive statistics for the 50 sample stocks from the STOXX 600 Index over the sample period of 20 years. The total observations vary for each variable because each variable has a different amount of missing data. All the variables are defined in Table 1 and are given as proportions.*

Next, I present the correlation between the variables in table 4. The correlation between the variables is quite low.

Table 4: Correlations matrix for the 50 sample stocks.

The table shows the correlations among variables of interest measured at monthly frequency for the sample 50 stocks from the STOXX 600 Index. The variables are defined in table 1. The sample period is from January 2004 to December 2023.

	ASVI	Abn Return	Volume	Volatility
ASVI	1.000	-0.001	0.002	-0.011
Abn Return	-0.001	1.000	-0.026	0.040
Abn Volume	0.002	-0.026	1.000	-0.063
Volatility	-0.011	0.040	-0.063	1.000

Chapter 4. Methodology

4.1 Research Design

Following the study of Da et al. (2011), I will focus on search volume data provided from Google Trends dating back to 1st January 2004 to 31st December 2023. Where this paper differentiates form the study of Da et al. (2011) is that I will be using monthly instead of weekly search volume index due to the larger sample period of 20 years compared to the 4 years period they used. The STOXX 600 Index will be used to measure the effect of attention on stock returns in the European market. As previously touched upon in chapter 3 of Data, an individual stock is identified in Google by using its ticker. I downloaded the SVI data in order to gather information on the 50 stocks in the sample. For SVI to be useful you have to standardize it first. The SVI is further processed into abnormal search volume indices (ASVI) using the formula from Da et al. (2011), which is defined as

$$ASVI_t = \log(SVI_t) - \log[Med(SVI_{t-1},..., SVI_{t-8})] \quad (6)$$

where $\log(SVI_t)$ is the logarithm of SVI during month t, and $\log[Med(SVI_{t-1},..., SVI_{t-8})]$ is the logarithm of the median value of SVI during the prior 8 months[10]. According to Da et al. (2011), the median captures over a longer window the "normal" level of attention in a way that is robust to recent spikes. A large positive ASVI certainly indicates a rise of investor attention, and it also makes it possible to compare search volume index via Google across stocks in the cross-section.

The methodology further involves panel data regression to investigate whether ASVI can explain or predict return. Following the study of Kim et al. (2019) two types of regressions were formed namely, descriptive regressions and predictive regressions. In the descriptive regressions, the ASVI is contemporary with the outcome variable. The descriptive regression is used to test the relationship between ASVI and contemporary stock returns, in order to answer the first hypothesis (H1). For the predictive regressions, the ASVI is lagged by one month to see if past ASVI can actually predict the outcome variables. The predictive regression is used to test the

[10] Da et al. (2011) measures SVI weekly and they use the prior 8 weeks to calculate median logarithmic value. Since this study measures monthly SVI, the formula has been adjusted to proportion.

relationship between ASVI and future stock returns, in order to answer the second hypothesis (H2). The regressions are presented in the next paragraph.

4.2 Regressions

The descriptive regression is performed on stock returns against ASVI and the other control variables presented in paragraph 3.4. According to Kim et al. (2019) this permits the impact of ASVI to be isolated to that of the control variables. In order to test the first hypothesis (H1), I define abnormal return for stock i on month t, Abn Ret_{it} to be

$$Abn\ Ret_{i,t} = \alpha_i + \beta_1\ Abn\ Ret_{i,t-1} + \beta_2\ ASVI_{i,t} + \beta_3\ Volatility_{i,t} + \beta_4\ Abn\ Volume_{i,t} + \varepsilon_{i,t} \quad (7)$$

where βs are the coefficients for the lagged abnormal return, ASVI, volatility, and the abnormal trading volume.

The predictive model is computed in a very similar way. According to Kim et al. (2019), the predictive models make use of only past information to predict future values. Therefore, only lagged variables are used as independent variables. Here is the predictive regression tested for the second hypothesis (H2):

$$Abn\ Ret_{i,t} = \alpha_i + \beta_1\ Abn\ Ret_{i,t-1} + \beta_2\ ASVI_{i,t-1} + \beta_3\ Volatility_{i,t-1} + \beta_4\ Abn\ Volume_{i,t-1} + \varepsilon_{i,t} \quad (8)$$

Where Abn Ret_{it} is the output variable abnormal return investigated for stock i on month t. Abn Ret_{it} is regressed on its lagged value, lagged ASVI, and the other two lagged control variables (i.e. the control variables for abnormal returns are abnormal volume and volatility).

4.3 Research Assumptions

When using regressions, the data must hold against a few assumptions for it to be statistically accurate. That is why it is necessary to test these three assumptions before conducting a regression.

The first assumption that needs to be tested is for perfect multicollinearity. Testing this assumption is important because it makes sure that the results of a regression analysis are reliable and valid. Multicollinearity is present when the independent variables of a regression have some level of correlation with each other. To analyze this phenomenon the Variance Inflation Factor (VIF) is used. Since the aim of this study is to research the relationship between ASVI and stock return this relationship will be used as an example. Appendix 2 presents the result of this analysis for the descriptive regression of abnormal return and Appendix 3 for the predictive regression of abnormal return. The regressions can be performed because the VIF values are close to 1, which is an acceptable level of multicollinearity.

The second assumption that needs to be tested is homoscedasticity. This signifies that the error terms of independent variables should have constant variances. In order to test if this holds true the Breusch-Pagan test is performed for both regressions of abnormal return. Appendix 4 and Appendix 5, display the result of this analysis. The p-value is smaller than 0.1, hence the hypothesis of constant variances is rejected, and heteroscedasticity is present.

The third assumption is that there should be no autocorrelation within the sample. Meaning, there can be no correlation between a variable and its lagged values. Given the fact that for regression analysis errors are uncorrelated it is crucial to analyze this. The Woolridge test for autocorrelation in panel dataset is used for this test. In Appendix 6 and Appendix 7 the result of this test is portrayed for the descriptive and predictive regressions. The p-value is lower than the threshold 0.5, hence the hypothesis of no autocorrelation is rejected, and there is autocorrelation present.

Since heteroscedasticity and autocorrelation were detected the Arellano method was applied to control for them (Arellano, 1987). Therefore, the results in the next section are presented with robust standard errors.

Chapter 5. Results

5.1 Regression Results

I use panel data regression models with fixed and random effects to determine the significance of Google searches for explaining and predicting abnormal returns. A Hausman test was performed to assess which effects to use. The outcome of the two Hausman tests are presented in appendix 8 and 9. The results support the fixed effect model for both regressions. Hence, only the results for panel data regression with fixed effects are displayed. As previously mentioned in chapter 4.3 there was heteroscedasticity and autocorrelation detected for the two regressions, but the Arellano method is used to control for them (Arellano, 1987). Therefore, the results in the table are displayed with robust standard errors.

The first model, displayed in column 1 of table 5, is the descriptive model regression with abnormal return as the dependent variable and its lagged value, abnormal SVI, abnormal trading volume, and volatility as independent variables. The result of table 5 shows that ASVI has a negative value that is not statistically significant. Meaning that there is no evidence to suggest that ASVI has an impact on contemporary abnormal returns. Abnormal trading volume is also statistically insignificant. The independent variable volatility is statistically significant at 1%. Meaning that when volatility rises with 1% the abnormal return is expected to increase by 4.29%. The one-month lagged abnormal return is also statistically significant at 5%. This implies that when the one-month lagged abnormal return rises by 1%, the abnormal return is expected to increase by 3.6%. H1 stated that there is a positive significant relationship between ASVI and the contemporary stock returns. Seeing that ASVI coefficient is negative and not statistically significant, no conclusion can be drawn. It is worth noting that the descriptive model has a relatively low R-Squared, with a value of 0.0061, meaning that 0.61% of the variation in the instances of abnormal return is caused by the independent variables in this model.

In column 2 the regression results for the predictive model are given. This model uses one-month lagged variables to predict future stock return. Just as the first model, the lagged variables for ASVI and abnormal trading volume are statistically insignificant. The one-month lagged volatility and one-month lagged abnormal return are statistically significant at a significance level of 1% and 5% respectively. H2 stated that there is a positive significant relationship between ASVI and the future stock returns, but since the one-month lagged ASVI is not statistically significant, no conclusion can be drawn. It is worth mentioning though that the coefficient for the one-month lagged ASVI is negative. Meaning that, past ASVI cannot predict long term returns, which is line with Da et al. (2011).

Table 5: Regression results for the 50 sample stocks.

	Dependent variable: Abnormal_Return	
Variables	**Descriptive Model**	**Predictive Model**
(Constant)	-0.0374***	-0.0398***
	(0.0030)	(0.0030)
ASVI	-0.0061	
	(0.0114)	
Abnormal_Volume	-0.0005	
	(0.0004)	
Volatility	0.0429***	
	(0.0099)	
L1_Abnormal_Return	0.0360**	0.0355**
	(0.0147)	(0.0145)
L1_ASVI		-0.0049
		(0.0109)
		-0.0004
L1_Abnormal_Volume		(0.0003)
L1_Volatility		0.0499***
		(0.0097)
R^2	0.0061	0.0050
F-statistics	6.97	5.27
N	10079	10040
N groups	50	50

Note: *This table displays the regression analysis of the relationship between abnormal SVI and abnormal returns. Column 1 displays the results of the descriptive model regression while column 2 displays the predictive model regression. The dependent variable abnormal return is described as unusually large profits or losses compared to a stocks normal or expected return. All the independent variables are denoted in proportions. Robust standard errors are given in parentheses. The significance is denoted as *p<0.10, **p<0.05, and ***p<0.01.*

5.2 Robustness Check

In order to enhance the reliability and generalizability of the results presented in this paper, a robustness check is performed. The empirical results of this paper will be tested on the 28 stocks in the data that do not have any missing values. This analysis will gather insight into whether the conclusions drawn from this study still hold when all the stocks with missing values are removed. Table 6 gives the summary statistics for the 28 stocks. There is not much difference between the summary statistics for the sample 50 stocks and the selected 28 stocks with no missing values. The only noticeable difference is for the variable abnormal trading volume and volatility, where the mean slightly increased.

Table 6: Summary statistics for the 28 stocks without missing values

	N	Minimum	Maximum	Mean	Standard Deviation
ASVI	6496	-0.352	0.745	0.0004	0.061
Abn Return	6692	-0.712	1.855	-0.025	0.086
Abn Volume	6720	-0.832	94.038	1.374	5.481
Volatility	6720	0	1.182	0.340	0.165

Note: *This table displays the descriptive statistics for 28 out of the 50 sample stocks from the STOXX 600 Index that have no missing values over the sample period of 20 years. The total observations may not seem to align, but this is not due to missing values. The lower number of total observations for ASVI*[11] *and Abnormal Return*[12]*are due to calculation reasons. All the variables are defined in Table 2 and are given as proportions.*

Table 7 displays the results that are obtained by re-running both the descriptive model and predictive model on the 28 stocks with no missing values. The robustness check indicates that the initial finding of a positive relationship between abnormal SVI and contemporary stock returns, which counters the results of this paper. However, the relationship between ASVI and contemporary stock return is also statistically insignificant just as this paper concludes. A possible explanation for this positive relation is provided by Da et al. (2011), who establish that ASVI predicts short term price increases up to two weeks following a price reversal for the long term. When studying whether search interest may have a significant influence on stock returns, the results prove to be insignificant, confirming the results that were acquired in the previous paragraph.

[11] In order to calculate ASVI, the previous 8 months are entered into the formula (see formula 6). As a result, the first 8 months of 2004 are empty. There are 224 missing data point due to the 28 stocks holding no value for 8 data points, which is exactly the difference between the number of observations for ASVI (6496) and the other two variables (6720)

[12] In order to calculate Abn Return, the previous month is entered into the formula (see formula 2 and 3). As a result, the first month of 2004 is empty. There are 28 missing data points due to each stock missing one data point, which accounts for the difference between the number of observations between Abn Return (6692) and the other two variables (6720).

Table 7: Regression results for the 28 stocks without missing values.

	Dependent variable: Abnormal_Return	
Variables	**Descriptive Model**	**Predictive Model**
(Constant)	-0.0406***	-0.0398***
	(0.0033)	(0.0034)
ASVI	0.0048	
	(0.0199)	
Abnormal_Volume	-0.0004	
	(0.0004)	
Volatility	0.0509***	
	(0.0101)	
L1_Abnormal_Return	0.0443**	0.0431**
	(0.0173)	(0.0172)
L1_ASVI		-0.0013
		(0.0146)
L1_Abnormal_Volume		-0.0003
		(0.0003)
L1_Volatility		0.0547***
		(0.0104)
R^2	0.0070	0.0074
F-statistics	7.88	7.99
N	6946	6468
N groups	28	28

Note: *This table displays the regression analysis of the relationship between abnormal SVI and abnormal returns. Column 1 displays the results of the descriptive model regression while column 2 displays the predictive model regression. The dependent variable abnormal return is described as unusually large profits or losses compared to a stocks normal or expected return. All the independent variables are denoted in proportions. Robust standard errors are given in parentheses. The significance is denoted as *p<0.10, **p<0.05, and ***p<0.01.*

Chapter 6. Conclusion and Discussion

6.1 Overview

In conclusion, this paper aimed to investigate the relationship between attention, measured by Google searches, and stock returns. This was done by conducting a descriptive and predictive regression using data from a random sample of 50 stocks listed on the STOXX 600 Index over a period of 20 years. The first hypothesis (H1) states that there exists a positive significant relation between Google searches and the contemporary stock returns. The second hypothesis (H2) states that there is a positive significant relation between Google searches and the future stock returns.

However, the main finding is that there is no significant relation between Google searches and stock returns. This applies to both hypotheses. The search volume does have a negative association with stock return, but the variable is not significant, so no conclusion can be drawn about the two hypotheses. This finding contradicts many papers who found that Google searches does have a significant positive effect on stock returns such as Da et al. (2011) and Joseph et al. (2011) for the U.S. market, Bank et al. (2011) for the German market, and Adachi et al. (2017) for the Japanese market.

The results of this paper do however resonate with the more recent studies of Kim et al. (2019) and Costa et al. (2024). Both papers show that the number of Google searches had no significant impact on stock returns. Kim et al. (2019) conducted their study on the Norwegian market and Costa et al. (2024) conducted their study on the Euronext exchange. A possible reason for the difference in findings between previous literature could be due to location-specific factors.

Thus, to answer the research question of *"How does attention-based news affect stock market returns for the European market?"*, the main findings point to an insignificant negative effect in the sample of 50 stocks listed on the STOXX 600 index. This means that no conclusion can be drawn whether attention influences stock returns in the European market. It is worth mentioning though that the coefficient for the one-month lagged ASVI is negative. Meaning that, past ASVI cannot predict long term returns, which is line with Da et al. (2011).

6.2 Research Implications

This paper contributes to the current literature by studying the relationship between Google searches and stock returns for the European market. Not until recently Costa et al. (2024) conducted a study focused on the European exchange, but prior to that the European continent has not been extensively studied as other continents or countries have been. Therefore, this paper adds a new sample in a different setting. This paper also reinforces the results obtained in the study of Costa et al. (2024), that Google searches have no significant impact on stock returns in the European market. Therefore, this study helps to fill the gap in the literature and helps to enhance the knowledge on this topic.

6.3 Limitations and Suggestions for Future Research

No study is without its limitations. As for this thesis a sample size of 50 stocks from the STOXX 600 Index was used. This may limit the generalizability of its findings, as only a fraction of the whole index was studied. Future research could expand to more stocks or perhaps even better the whole STOXX 600 Index, for a more diverse sample, to provide a deeper understanding of the relationship between attention measured by internet search volume and the STOXX 600 Index returns.

In addition, this study only focused on the European companies, thus limiting the conclusions that can be made for sector specific companies. By concentrating on specific business sectors such as finance, technology, health care, energy or manufacturing, you could get a better understanding of how attention affects the return of certain sectors in comparison to other sectors. By addressing these limitations and conducting studies with a larger sample, future research can enhance the current understanding of the relationship.

References

Aboody, D., Lehavy, R., & Trueman, B. (2010). Limited attention and the earnings announcement returns of past stock market winners. *Review of Accounting Studies*, 15, 317-344.

Adachi, Y., Masuda, M., & Takeda, F. (2017). Google search intensity and its relationship to the returns and liquidity of Japanese startup stocks. *Pacific-Basin Finance Journal*, 46, 243-257.

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

Arellano, M. (1987). Computing robust standard errors for within-groups estimators. *Oxford Bulletin of Economics & Statistics*, 49(4).

Bank, M., Larch, M., & Peter, G. (2011). Google search volume and its influence on liquidity and returns of German stocks. *Financial markets and portfolio management*, *25*, 239-264.

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *The review of financial studies*, *21*(2), 785-818.

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

Brown, S. J., & Warner, J. B. (1985). Using daily stock returns: The case of event studies. *Journal of financial economics*, 14(1), 3-31.

Chordia, Tarun, Sahn-Wook Huh, and Avanidhar Subrahmanyam, 2007, The cross-section of expected trading activity, *Review of Financial Studies 20,* 709–740.

Costa, L., Vieira, E., & Madaleno, M. (2024). Does Googling Impact Euronext Stock Returns?. *IBIMA Business Review*, 2024.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The journal of finance*, *66*(5), 1461-1499.

Drake, M. S., Jennings, J., Roulstone, D. T., & Thornock, J. R. (2017). The comovement of investor attention. *Management Science*, 63(9), 2847-2867.

Drake, M. S., Roulstone, D. T., & Thornock, J. R. (2015). The determinants and consequences of information acquisition via EDGAR. *Contemporary Accounting Research*, 32(3), 1128-1161.

Fang, L., & Peress, J. (2009). Media coverage and the cross-section of stock returns. *The journal of finance*, 64(5), 2023-2052.

Financial Times (2024). *The unfulfilled potential of Europe's stock markets*. Retrieved from https://www.ft.com/content/4f94ab66-25c6-45a1-96df-1fd6cd2ad509

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457(7232), 1012-1014.

Hirshleifer, D., Hsu, P. H., & Li, D. (2013). Innovative efficiency and stock returns. *Journal of financial economics*, 107(3), 632-654.

Huddart, S., Lang, M., & Yetman, M. H. (2009). Volume and price patterns around a stock's 52-week highs and lows: Theory and evidence. *Management Science*, 55(1), 16-31.

Ioannidis, C., & Kontonikas, A. (2008). The impact of monetary policy on stock prices. *Journal of policy modeling*, 30(1), 33-53.

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

Kahneman, D. (1973). Attention and effort (Vol. 1063, pp. 218-226). Englewood Cliffs, NJ: Prentice-Hall.

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

Lee, C. M., Ma, P., & Wang, C. C. (2015). Search-based peer firms: Aggregating investor perceptions through internet co-searches. *Journal of Financial Economics*, 116(2), 410-431.

Li, J., & Yu, J. (2012). Investor attention, psychological anchors, and stock return predictability. *Journal of financial economics*, 104(2), 401-419.

Lou, D. (2014). Attracting investor attention through advertising. *The Review of Financial Studies*, 27(6), 1797-1829.

Neufeld, D. (2023, October 18). *Mapped: The Largest Stock Exchanges in the World*. Visual Capitalist. Retrieved from https://www.visualcapitalist.com/largest-stock-exchanges-in-the-world-mapped/

Nguyen, C. P., Schinckus, C., & Nguyen, T. V. H. (2019). Google search and stock returns in emerging markets. *Borsa Istanbul Review*, 19(4), 288-296.

Search Engine Market Share Worldwide | Statcounter Global Stats. Retrieved June 5, 2024, StatCounter Global Stats. https://gs.statcounter.com/search-engine-market-share

Spaulding, D. (2002). *The rate-of-return formula can make a difference.* In Performance Measurement in Finance (pp. 329-341). Butterworth-Heinemann.

STOXX Ltd. (n.d.). STOXX® Europe 600 index. Retrieved May 10, 2024, from https://stoxx.com/index/sxxp/

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

Tantaopas, P., Padungsaksawasdi, C., & Treepongkaruna, S. (2016). Attention effect via internet search intensity in Asia-Pacific stock markets. *Pacific-Basin Finance Journal*, *38*, 107-124.

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

Appendix

Appendix 1: Overview of the 50 stocks drawn out of the random sample and the STOXX 600 index with their search queries.

Stock	Ticker	Country	Search Query
Adidas	ADS	Germany	"ads"
ABN AMRO Bank	ABN	Netherlands	"abn"
Acciona	ANA	Spain	"ana"
ADP	ADP	France	"adp"
Adyen	ADYEN	Netherlands	"adyen"
Alstom	ALO	France	"alo"
Aviva	AV	United Kingdom	"av"
Avolta Ag	AVOL	Switzerland	"avol"
Baloise Holding	BALN	Switzerland	"baln"
Bank of Ireland Group	BIRG	Ireland	"birg"
Cd Projekt	CDR	Poland	"cdr"
Croda International	CRDA	United Kingdom	"crda"
CTS Eventim	EVD	Germany	"evd"
Deutsche Bank	DBK	Germany	"dbk"
Diageo	DGE	United Kingdom	"dge"
EDP Energias De Portugal	EDP	Portugal	"edp"
Elia Group	ELI	Belgium	"eli"
Flutter Entertainment	FLTR	Ireland	"fltr"
Frontline	FRO	Norway	"fro"
ING Groep	INGA	Netherlands	"inga"
KBC Group	KBC	Belgium	"kbc"
Lanxess	LXS	Germany	"lxs"
Choc.Lindt & Spruengli	LISN	Switzerland	"lisn"
Legrand	LR	France	"lr"
LondonMetric Property	LMP	United Kingdom	"lmp"
LPP	LPP	Poland	"lpp"
Nestle	NESN	Switzerland	"nesn"
NKT	NKT	Denmark	"nkt"

OMV	OMV	Austria	"omv"
Pandora	PNDORA	Denmark	"pndora"
Bank Polska Kasa Opieki	PEO	Poland	"peo"
Rheinmetall	RHM	Germany	"rhm"
Glanbia	GL9	Ireland	"glb"
Sagax B	SAGAB	Sweden	"sagab"
Serco Group	SRP	United Kingdom	"srp"
Shell	SHEL	Netherlands	"shel"
Sofina	SOF	Belgium	"sof"
Soitec	SOI	France	"soi"
Symrise	SY1	Germany	"sy1"
Talanx AG	TLX	Germany	"tlx"
Thule Group	THULE	Sweden	"thule
TietoEVRY	TIETO	Finland	"tieto"
UBS Group	UBSG	Switzerland	"ubsg"
UCB	UCB	Belgium	"ucb"
Unicredit	UCG	Italy	"ucg"
Verallia	VRLA	France	"vrla"
Vivendi	VIV	France	"viv"
Weir Group	WEIR	United Kingdom	"weir"
Worldline	WLN	France	"wln"
Zurich Insurance Group	ZURN	Switerland	"zurn"

Appendix 2: Multicollinearity test (VIF) table for the descriptive abnormal return model.

Variable	VIF	1/VIF
Volatility	1.01	0.9948
Abnormal_Volume	1.00	0.9957
L1_Abnormal_Return	1.00	0.9986
ASVI	1.00	0.9997
Mean VIF	1.00	

Appendix 3: Multicollinearity test (VIF) table for the predictive abnormal return model.

Variable	VIF	1/VIF
L1_Volatility	1.01	0.9946
L1_Abnormal_Trading_Volume	1.00	0.9955
L1_Abnormal_Return	1.00	0.9980
L1_ASVI	1.00	0.9999
Mean VIF	1.00	

Appendix 4: Heteroscedasticity Breush-Pagan test for descriptive abnormal return model.

H_0: Constant variance Chi2(1)

= 23.89

Prob> Chi2 = 0.0000

Appendix 5: Heteroscedasticity Breush-Pagan test for predictive abnormal return model.

H_0: Constant variance Chi2(1)

= 8.01

Prob> Chi2 = 0.0047

Appendix 6: Autocorrelation Woolridge test for the descriptive abnormal return model.

H_0: No first-order autocorrelation

F(1,49) = 635.491

Prov>F = 0.0000

Appendix 7: Autocorrelation Woolridge test for the predictive abnormal return model.

H_0: No first-order autocorrelation

F(1,49) = 590.604

Prov>F = 0.0000

Appendix 8: Hausman test for the descriptive abnormal return model.

	Coefficients			
	(b)	(B)	(b-B)	Standard error
	fe	re	Difference	
L1_Abnormal_Return	0.3460	0.0454	-0.0094	0.0007
ASVI	-0.0061	-0.0008	-0.0053	0.0016
Abnormal_Volume	-0.0005	-0.0004	-0.0001	0.0001
Volatility	0.0429	0.0189	0.0239	0.0056

Test of H_0: Difference in coefficients not systematic.

Chi(4) = 231.94

Prob> Chi2 = 0.0000

Appendix 9: Hausman test for the predictive abnormal return model.

	Coefficients			
	(b)	(B)	(b-B)	Standard error
	fe	re	Difference	
L1_Abnormal_Return	0.0353	0.0455	-0.0100	0.0007
L1_ASVI	-0.0049	0.0010	-0.0060	0.0016
L1_Abnormal_Volume	-0.0004	-0.0003	0.0000	0.0001
L1_Volatility	0.0450	0.0224	0.0275	0.0056

Test of H_0: Difference in coefficients not systematic.

Chi(4) = 288.44

Prob> Chi2 = 0.0000

Appendix 10: Overview of the 22 stocks with their missing variables.

This table gives an overview of the 22 stocks in the sample of 50 stocks from the STOXX 600 Index that have large sets of missing values. The stocks are shown with their ticker for clarity. I only include the missing input variables, thus variables that are directly extracted from the databases. The variable SVI is extracted from Google Trends, Trading volume is obtained from Yahoo!Finance, and Price, and Volatility is acquired from the Eikon Datastream. The 'X' shows if a certain stock misses that variable. At the bottom of the table the total quantity of missing variable is stated for each variable.

Stock	Ticker	SVI	Price	Trading Volume	Volatility
ABN AMRO Bank	ABN		X	X	X
ADP	ADP		X	X	X
Adyen	ADYEN	X	X	X	X
Alstom	ALO			X	
Avolta Ag	AVOL	X	X	X	X
Baloise Holding	BALN	X			
Croda International	CRDA	X			
Elia Group	ELI		X	X	X
Frontline	FLTR	X			
Lanxess	LXS	X	X	X	X
Choc.Lindt & Spruengli	LISN	X			
Legrand	LR		X	X	X
LondonMetric Property	LMP		X	X	X
Pandora	PNDORA	X	X	X	X
Glanbia	GL9	X			
Sagax B	SAGAB	X	X	X	X
Symrise	SY1	X	X	X	X
Talanx AG	TLX		X	X	X
Thule Group	THULE		X	X	X
UBS Group	UBSG	X			
Verallia	VRLA	X	X	X	X
Worldine	WLN	X	X	X	X
Total:	22	14	15	16	15