Turun yliopisto
University of Turku

IS FORECASTING NON-PROFESSIONAL INVESTORS TRADING BEHAVIOR POSSIBLE WITH GOOGLE TRENDS?

Trading volumes based on Google's Search Volume Index as a proxy for investor attention

Master´s Thesis
In Accounting and finance

Author:
Arto Lerkki 418986

Supervisors:
Prof. Markus Granlund
Dr. Jan Pfister

8.5.2015
Turku

Turun kauppakorkeakoulu • Turku School of Economics

CONTENTS

1 INTRODUCTION 7

1.1 Background 7
1.2 Theoretical framework 10
1.3 The purpose of the study 12
1.4 Methodology in comparison to previous studies 14
1.4.1 Contribution 14
1.4.2 Methods 16
1.5 Validity, reliability and limitations with generalization 16
1.6 Structure of the study 17

2 THEORETICAL FRAMEWORK AND LITERATURE 19

2.1 Attention and the efficient markets 19
2.1.1 Limited attention and overconfidence 20
2.1.2 Measuring attention with proxies 21
2.1.3 Neoclassical approach and Efficient Market Hypothesis (EMH) 22
2.1.4 The efficiency of the Finnish securities markets 23
2.2 Behavioral finance 24
2.2.1 Behavioral finance and market sentiment 24
2.2.2 Noise trading 28
2.2.3 Limiting factors to arbitrage trading 31
2.2.4 Psychological factors influencing markets 34
2.2.5 Heuristics and manners 38
2.2.6 Risk management in behavioral finance 42
2.2.7 Sociological factors influencing markets 44
2.3 Literature specific to Google Trends 50
2.3.1 Non-finance specific literature on Search Volume index and Google Trends 50
2.3.2 Finance specific literature on Search Volume index 52
2.4 Development of the hypotheses 56
2.4.1 Hypotheses in this study 56
2.4.2 Comparison to previous studies 59

3 METHODOLOGY, METHODS AND DATA 62

3.1 Regression methods and the testing of study hypotheses 62
3.1.1 The use of multivariate models instead of structural 63
3.1.2 The application of VAR models with SVI and Trading Volume 65
3.2 Data and descriptive statistics 67

3.3 Generating the time series ... 70
3.3.1 Abnormal and expected returns ... 72
3.3.2 Abnormal trading volume ... 74
3.3.3 Abnormal Search Volume index ASVI ... 75
3.4 Data for OLS and multivariate analysis ... 78
3.5 Granger causality ... 82

4 EMPIRICAL RESULTS ... 85

4.1 Regressions and testing ... 85
4.1.1 Ordinary least squares ... 85
4.1.2 Multivariate and ARMA-GARCH ... 87
4.2 Granger causality ... 93
4.3 Results of the hypothesis testing ... 94
4.4 Fama-French three-factor model and robustness checks ... 99

5 CONCLUSIONS ... 102

5.1 The purpose of the study and conclusions ... 102
5.2 Future studies and limitations ... 104

REFERENCES ... 106

APPENDIX 1 ALL LISTED COMPANIES ... 119

APPENDIX 2 SUMMARY OF OLS REGRESSIONS ... 122

APPENDIX 3 ATRIA ... 128

APPENDIX 4 FORTUM ... 129

APPENDIX 5 KEMIRA ... 130

APPENDIX 6 KESKISUOMALAINEN ... 131

APPENDIX 7 KONE ... 132

APPENDIX 8 GRANGER SVI_GLO ... 133

APPENDIX 9 GRANGER SVI_FIN ... 135

APPENDIX 10 OMXH ALL SHARE CONSTITUENTS FOR THIS STUDY ... 138

LIST OF FIGURES

Figure 1 Google market share compared to other search engines from 2010 to 2014 (Alexa 2014)........ 9
Figure 2 The keyword “vippi” indicating weekly fluctuation of SVI (Google 2014) 10
Figure 3 Tikkurila abnormal daily returns for year 2013........ 73
Figure 4 Abnormal trading volume for Tikkurila for year 2013........ 75
Figure 5 Abnormal Search Volume Index for Tikkurila for year 2013........ 77
Figure 6 Normal distribution testing, skewness and kurtosis........ 90

LIST OF TABLES

Table 1 A summary of studies analyzing Search Volume Index (DEG 2010; Wuoristo 2012; Dimpfi & Jank 2012)........ 55
Table 2 List of the hypotheses in this study as compared to previous studies........ 59
Table 3 Example: FTSE AllShare Composition (Wuoristo 2012, 24)........ 68
Table 4 Variables used in the study........ 71
Table 5 An example of a noisy ticker search for the Keller resulting in omitting from study (Wuoristo 2012, 30)........ 81
Table 6 Summary of OLS regressions........ 86
Table 7 Atria multivariate regression results........ 89
Table 8 Q-statistic probabilities for ARMA terms........ 90
Table 9 Fortum ARMA........ 91
Table 10 Kemira ARMA........ 91
Table 11 Keskisuomalainen ARMA........ 92
Table 12 Kone ARMA........ 93
Table 13 Conclusions of the hypotheses........ 99

1 INTRODUCTION

1.1 Background

The introduction of Google Trends in 2006 has made it possible to perform research with a new approach. The general ability to measure the trends of queries and the demand for information contains large quantities of new data to be analyzed. Previously that data has not been easily in a quantifiable form. According to Alexa (2015) Google has achieved large market share in global internet search queries and it provides a unique tool in almost real time to measure the global search query amounts the different topics receive from public queries.

The search query studies have begun around 2010 in modern finance. One of the first remarkable research papers was Da, Engelberg, Gao (2011) (from now on referenced as DEG). The first version was published in 2009 but was later revised with new data. This particular study will replicate such previous studies and tests done by DEG, but with different time period, different target group and different search query key words. The methods and methodology will have common approaches, making this study in reality comparable to previous studies in the same context.

"Google is a company that specializes in digital data facilitation; for general common issues there is the basic Google service, for academia there is Google Scholar and for finance related information there is Google Finance "(Wuoristo 2012, 1). During the past few years Google searches have become the industry standard how to navigate web. In Europe, UK and USA Google is the most common way compared to other search engines to search the internet for various topics. During the last five years Google has also become the most visited site measured by internet traffic (Alexa 2015).

Google trends provides the aggregated time-series data. It is generally referenced as "Search Volume Index". The commonly used abbreviation for it is SVI. This information is publicly available to all Google account holders from a single web page (Google Trends). The qualification for the data to appear on Google trends is that there is enough search queries made each week. The data is also given out in weekly frequency. "The weekly SVI value is the number of searches for that term scaled by the time-series average" (Wuoristo 2012, 2). Further filters can be added to sort the time-series data. Few different filters can, for example, be:

- Country from which the queries originate
- Product categories, in example Apple (computers) and apple (food)
- Start of the time-series period
- End of the time-series period

- Total number of weeks in the sample

The data is generally available dating back to early 2004, and ends in the previous week when the query for the CSV Google Trends data is made. For example, executing a query for "Nokia" share during week 10 2014, would therefore result in time-series from 2004 to query date week 9. Therefore, it can be perceived Google Trends has one full week of buffer until a new series is released for public. With longer series this is not relevant or statistically significant for this study, but in case of shorter periods, the one week duration can have a greater weight. The problem with too short a duration for time series can be easily avoided by observing a longer time frame or extrapolating the missing values.

Because Google has a high penetration rate in most markets and it can provide reliable results in terms of measuring attention trough different query words (Google 2014). Google's market share has not been always so high. The data further away from current situation or time period is not always reliable. The search queries relational proportions might have been divided differently between specific search engines. For example, retail investors have made their queries on a certain search engine or platform. In such case they may have been preferring an alternative engine over Google. Academic research does not have many studies or statistics over this. Therefore, this study does not take the possible effect into account at all. Such competitors are, for example, Altavista which no longer holds any significant part of queries compared to Google or Microsoft's Bing.

Over the few last year's Google's popularity as the number one search engine has remained fairly steady as seen from the figure below (Alexa 2014). Other search engines are slowly gaining popularity, but at the same time the size for total market is growing rapidly and is estimated to experience double digit growth for years to come (Alexa 2014). This further strengthened the hold of Google as the most popular and used engine according to surveys and query amount histories published by Google. The very near past situation (more precisely years 2010-2014) for Google can be seen from the following figure.

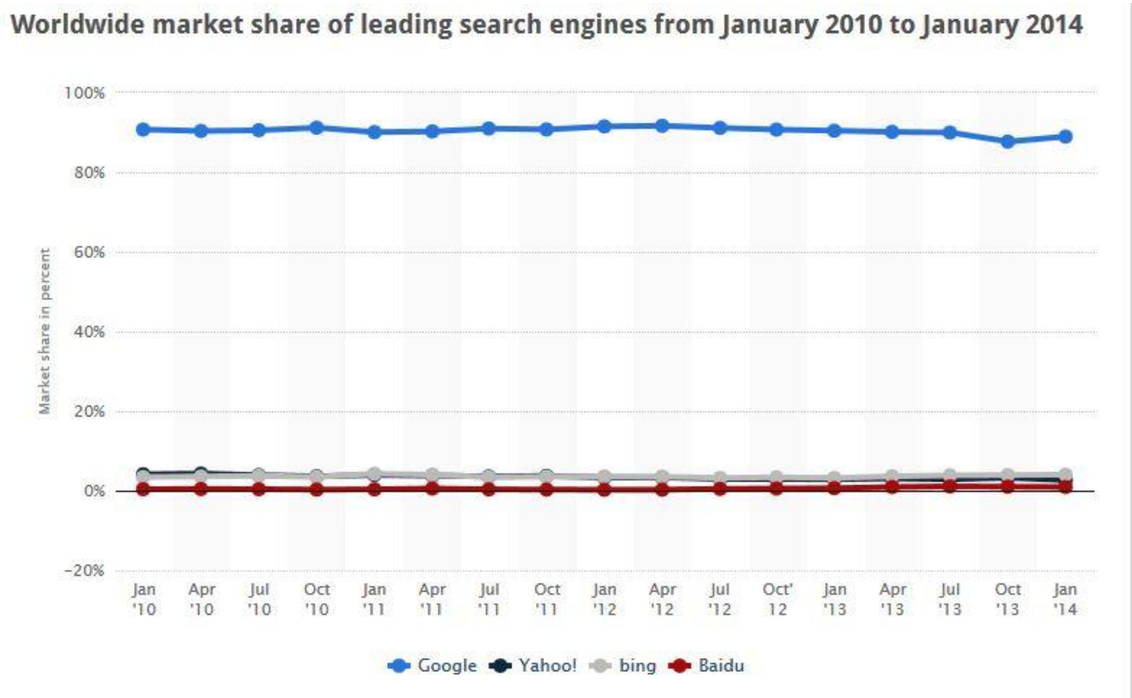

Figure 1 Google market share compared to other search engines from 2010 to 2014 (Alexa 2014)

In the figure one it can be seen Google takes a major part of the internet queries. Previous finance studies have taken the distribution of queries (between different search engines) as an insignificant factor when modeling behavioral phenomenon. This study presumes the differences are also insignificant for validity and reliability. The keyword *vippi* is used as an example because the attention it receives comes almost solely from consumers. It can be seen as a reliable proxy for attention as it also has very little if any noise in it. The word always infers with borrowing money and does not have any other meanings in Finnish language.

The graph depicts an example of SVI. The keyword is the Finnish word “vippi” which means a very short term loan. Commonly the duration of the loan varies from 7-14 days accompanied with a high interest rate. The following figure states the attention such a keyword receives:

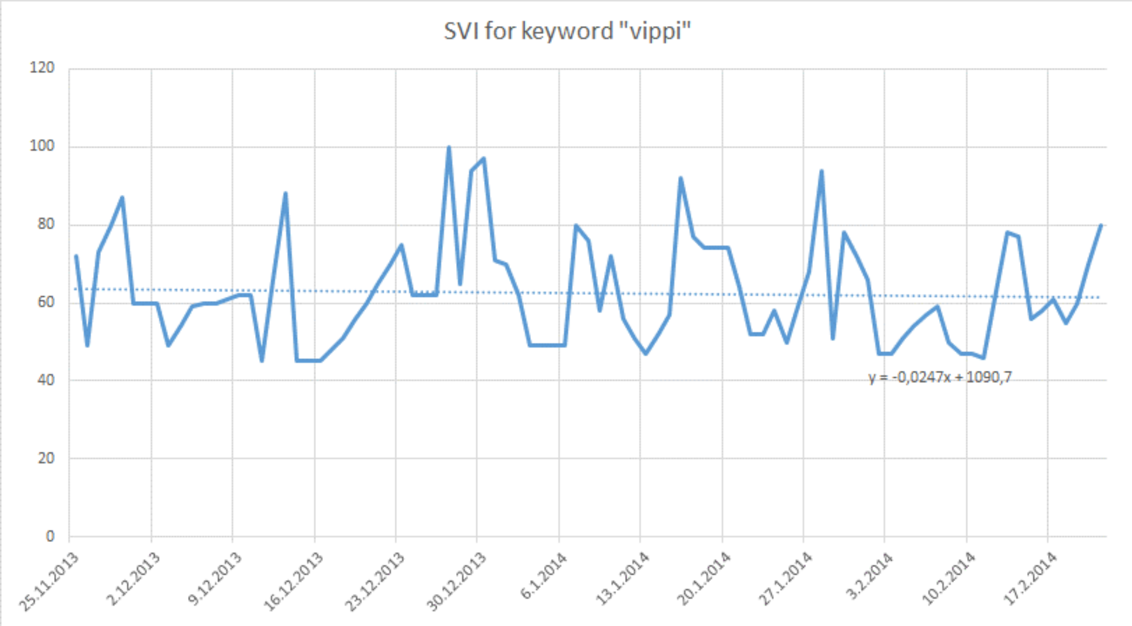

Figure 2 The keyword "vippi" indicating weekly fluctuation of SVI (Google 2014)

There are clear and definite peaks from Thursday to Friday therefore indicating the short-term financing needs of individuals seems to be mainly for weekends. Also there is a significant spike before New Year's Eve as can be seen from the graph near the end of the 2013. Therefore, it's safe to conclude SVI captures the attention well in this case. There is clear pattern in the fluctuation of SVI for the keyword. The actual level of SVI may vary, but the time-series are stationary. This is most likely because of the relative nature of the time-series since Google Trends uses normalization. The actual formula for this specific action is unknown.

Google Trends Search Volume Index (SVI) has been shown to have predictive powers in forecasting influenza epidemics (Ginbsberg et al, 2009). The predictive powers have not been commonly tested in neither corporate finance or in finance research in general besides few geographically isolated studies (mainly in the UK and in the US). Nevertheless generally its predictive powers have not been tested to full extent, and with the on-going development of the tool (Google Trends and Google Insight) it is very likely the forecasting powers will significantly increase in future and new fields of study will emerge. Also since the release of Google Trends in 2006 the amount of data gathered is significantly greater than in the studies previously conducted in 2009.

1.2 Theoretical framework

Classic asset pricing models and efficient market hypothesis claim all information to be instantaneously available for all investors. Therefore, it will be reflected in the share price immediately, and there are no possibilities of arbitrages or gaining excess returns (or at least they would be extremely short term). Kahneman (1973) argues attention being a scarce cognitive resource and therefore investors are forced with different alternatives for their focus of attention. Thus generally investors are presumed to have a predefined set of choices or alternatives for their individual investment decisions. This framework is then used to make the said investment decisions without further analysis.

Based on previous studies investor attention is presumed to be a significant factor in defining trading volumes for assets (Barber and Odean 2008). The easiest and most practical approach to measure attention is using different proxies. Few examples of these (indirect) proxies according to previous studies are:

- References in newspapers or journals (Barber and Odean 2008).
- Attention in online discussion forums or chat rooms (mouth to mouth advertising) (Barber and Odean 2008).
- Advertising expenses (Chemmanur and Yan 2009).
- Price limits (Seasholes and Wu 2007).
- Trading volume (Barber and Odean 2008).
- Excess positive returns (defined by CAPM/Beta and Fama-French three-factor model).
- Excess negative returns (defined by CAPM/Beta Fama-French and three-factor model).

Most of the studies indicate that those shares, stocks or other investment assets grasp the general public's attention are most commonly based on individual preferences. This indicates that they are not acquired on grounds of classic financial theory. "In 2011, the first study to use a *direct* measure of investor attention was published (DEG 2009). The paper uses Google's Search Volume Index (SVI) as a proxy for investor attention based on the assumption that searching on the internet for a company reflects acute and direct interest." (Wuoristo 2012, 4) The study claims there is a linear dependency between Search Volume Index and stock trading volume's with shares listed in the UK.

DEG (2011) show in which way and how much attention the SVI captures of retail investors. The presumption is that institutional investors turn towards more sophisticated tools, for examples, the Reuters or the Bloomberg terminals (DEG 2011; Wuoristo 2012; Mondria and Wu 2011). This means the investors that are outside the use of Bloomberg or Reuters terminals would be nearly always using Google as their main means to acquiring information. This is in line with academic journals and is a likely result of Google gaining near monopoly market penetration.

The study of Mondria and WU (2011) indicates that information asymmetry is also a decisive factor in the following way:

- Local attention high, non-local low -> SVI can be efficiently used to predict future share prices
- Local attention high, non-local high -> no statistically significant predictive powers
- Local attention low, non-local high -> no statistically significant predictive powers

Therefore, SVI predicts only (or at least better) when there is information asymmetry. The predictive powers are solely defined by ordinary least squares (OLS) regressions in all the mentioned previous studies. Any more complex or more thorough analysis methods have not been used before. First example of these methods is examining the changes in variance with the regression. Another example is using Granger causality to determine if another variable Granger causes the other. In these cases there could be predictability and possibilities of arbitraging. The lack of use of these methods previously can be of several reasons. Firstly because OLS is a very fast method of acquiring reliable results from vast sample sizes. Secondly it is unclear if OLS is leaving something out. In reality this means more complex methods such as VAR may not provide any information that a simple OLS cannot produce. The studies generally do not give results that could not be contradicted. In many cases similar studies conducted with different types of stocks give the opposite results. This finding is very different from what is stated in DEG (2011) study. Nevertheless, some of this can be explained because DEG (2011) used weekly data – as will this study also – but Mondria and Wu (2011) used monthly. In many cases the autocorrelation or predictive powers can be seen only in shorter frequency. Both the studies mentioned use OLS as a method of regression, and do not imply the use of VAR or other more complex methods. Also the causality between SVI and share price as also with trading volume is not clear, but the statistically significant regression still holds true.

1.3 The purpose of the study

The objective is to analyze if Google Trends can predict stock trading volume with Finnish OMXH listed shares. This is done by using search volume index as a proxy of non-professional investor attention. Furthermore the study is conducted by using the company names as the queries. This can be further studied for possibilities of predicting trades or prices using Google Trends search results as the proxy on publicly listed companies in Finland based on search query results using weekly data from Google Trends and index price data from Data Stream. The financial data on individual companies and indices is available from Data Stream but also from alternative sources as

Yahoo Finance. Therefore, the thesis examines non-professional investor behavior with using Google search volume index as a proxy for attention. The following three individual research questions are all defining the main question more clearly. The first question is if search volume index is a viable proxy to be used for studying. The second question is a follow-up to the first. So if search volume index is actually a viable proxy for investor attention and quantifies the common attention, does it correlate with trading volumes at a statistically significant level? So if investor attention rises do investors buy more shares and does it affect the liquidity of stocks and how much? The third question maps if search volume index can be used to predict trading volumes in the future. If there are higher than normal trading volumes, are they connected to returns? The questions can be stated as follows:

1. Does Search Volume Index capture the attention of non-professional investors in the Finnish market?
2. Does Search Volume Index correlate with the stock market trading volume (in a statistically significant level)?
3. Can Search Volume Index be used to predict and forecast an individual company's stock turnover?

The motivation of this study is to test the initial findings of Da, Engelberg and Gao (2011) and Wuoristo (2012) and replicate certain parts of their studies with different geographical area, target group and stock market index. Both of the studies conclude differences in the target groups. They also clearly and specifically indicate they expect different results to arise from different areas. By conducting the study in a similar manner with the same sources of data, the study can be easily verified and replicated, and is also comparable. Also, since this study is made at a later date, Google Trends is offering longer time series and it will also probably dilute the effects of the 2008-2009 financial crisis better and thus offer a better long term approximation of the predicting powers of SVI. This study will solely focus on keywords that are company names, while the previous studies generally have been studying queries done with ticker symbols. Examples of this are, for example, searching the time-series for Apple computers by just using the ticker APPL or NASDAQ:APPL.

In general, this would indicate this study is one of the first conducted with either Nordic or Finnish material. Only exceptions would be the tickers overlapping the 3 different market places because of multiple listings (i.e. Helsinki, UK, and US). The amount of these companies is very limited. The assumption is that the general public will indicate strong home bias –effect. Therefore, a high value in SVI (keywords with share ticker or company name) will probably mean higher general public interest in the stock and thus have higher trading volumes. This in return results as higher demand for the stock, and depending on the volume of trades would influence the market price of the share. If there is clear indication of predictability, such information could be used efficiently to arbitrage and gain excess returns (and maybe enhance the phenomena of rising trading volumes based on raising search volume index values). It is possible if there is evidence of clear predictability. Then, it could be presumed the investors want to utilize such an inefficiency by creating more demand, increasing prices and causing more volatility.

The studies conducted by Da, Engelberg and Gao (2011) and Wuoristo (2012) focus on a general mean level key figures such as average turnover, average returns and median prices for a number of different stocks. There is a definite lack of studies that solely focus on individual stocks and shares. This is mainly based on few separate reasons. Firstly, the sheer amount of data collection is extremely time consuming especially if the vast majority of the companies need to be omitted for various reasons. Until Google or any other major source provides easy and fast means of acquiring data, conducting studies will be difficult. Secondly, the sample may end being too small to reliably predict future values for variables. In many cases for industry sectors and indices the sample sizes are vast and ample. Although, this is not the case for all sectors. For example, the OMXH: Helsinki stock index only has few companies that solely work with consumers instead of business to business trade. If the remaining sample size for studies would range from only 2 to 3 different distinct companies, it would be difficult to make generalizations based on the sample's findings (De long et al. 2008).

1.4 Methodology in comparison to previous studies

1.4.1 Contribution

There are not many studies in finance conducted with the use of *search volume index* yet. The most common reason is that the actual tool is rather new and therefore the time-series are only few years in length. This forces many of the studies to be focused on short term events instead of longer term predictability and analysis. Because of the massive amount of data gathering that is needed the analysis is often limited to simple ordinary least squares regressions. Nevertheless, as time passes, the series will be more ample and will cover different events in the general economy, for example, long bull markets and longer periods of recession. As far as the author knows, this study is the first study to focus on solely Finnish securities and SVI. The main focus and idea is still to replicate certain limited parts of the previous study by DEG (2011). Although, this study will have very different focus on certain aspects:

- Size of market (the US vs Finland)
- Type of the markets (US companies have ownership from several different types of investors whereas in Finnish companies are mainly owned by institutional investors),
- Nordic markets are more focused on certain industry fields where as US markets have various types of companies representing a larger field of economy,
- Company sizes vary from Finnish small caps (20 M market capitalization) to over 200 B USD in US markets).
- Time horizon from the beginning of the year 2013.
- The market conditions under which the study is conducted. For example, studies from 2008 to 2009 will have different economic background than studies conducted after 2010.
- Both VAR and OLS models are being used in this study while previous studies commonly use only ordinary least squares methods.
- Granger-causality testing for estimating future values.
- More in depth focus on what influences the SVI from the aspect of behavioral finance instead of typical classical Bayesian estimation.

Based on previous studies, it seems that retail investors act differently in different markets (Trueman 1988). In this study the presumption is the investors react differently in the US markets than they act in the Finnish and the Nordic markets. However, it is not clear at all if this has an effect on the results of the study based on any theory. The first example for the differences is the home bias in different countries. Some countries may be more eager to invest to local markets instead of global (Tay 2009). This varies obviously from country to country. The second major difference between markets is the population's attitude towards investing related activities in general (Sornette 2003). In the US, people generally invest and act actively in markets following news and reading about companies to invest in. In Finland, people keep their distance to investing or have market exposure at best with mutual funds or through pension funds in general. This allows the Finnish investors to be exceedingly more passive and the American and people from the United Kingdom to be more active and with greater participation in investing activities. It is unclear how well SVI can capture the non-professional investor attention. It could be questioned if the non-professional investors use Google as a source of information when they are interested in investing in companies. Also, if non-professional investors use Google for seeking information, is it consistent with all companies or unique to specific type of companies? For example, are small caps receiving more queries than large cap companies? Is SVI capturing the attention of the public in the first place?

Therefore, the major sources of contribution in this study are:

1. Understanding the Search Volume index in Finnish markets with non-professional investors.
2. Deepening the understanding on non-professional investor behavior in general.
3. Learning the uses of SVI as a proxy of attention.

1.4.2 Methods

The main method is statistical inference. Another data gathering method is to gather information from Google Trends by hand individually one company at a time. The process starts by going into Google Trends landing page and searching the keywords one at a time. It is unclear if Google allows some sort of data mining or web crawling to be used in gathering information, but it is not available for public when this study was conducted. It is also noted that the previous studies conducted had no data mining tools available either, so it can be concluded the amount of companies is limited because of work and time limitations. Samples of hundreds of companies would require massive amount of manual data extraction, which is prone to errors. Nevertheless, when the data is gathered one at a time by hand, the sample size has to be limited because the actual process is extremely time consuming. Google Trends provides CSV-files[1] which in turn have to be processed to be in a format that can be input to EViews[2] or similar statistical analysis application. From this data the different tests will be conducted.

1.5 Validity, reliability and limitations with generalization

The previous studies have also verified most of the keywords and tickers are not "noisy" (DEG 2011; Wuoristo 2012, 6; Mondria and Wu 2011; Ginbsberg et al. 2009). There does not appear to be this problem with the Finnish data either. There are quite a lot of problems with European and US data on the other hand, because many of the brand names can have multiple different meanings. As Wuoristo (2012) indicates in his study: "Another limitation is that many companies have noisy tickers, such as British American Tobacco's ticker BATS, which cannot be used to measure SVI since it would also pick up on searches for the nocturnal animals." For Finnish companies a similar problem exists only to some extent. For example, if you search for the keyword Fortum, you might be interested in finding information about electricity contracts or prices. Nevertheless, this

[1] CSV-File is a Comma-Separated-Values file with raw numerical data in a standard table form defined by RFC 4180 format.
[2] EViews is a Windows based statistical program specifically designed for economic analysis.

also is a signal of interest or attention towards the company so it is not grounds for omitting the data from the research. Therefore, such coincidences will be left within the research data. In many cases the Nordic companies have very specific and straight forward names and are not easily confused with anything else. For example, a query for the steel company "Rautaruukki" would result only in results related to the steel manufacturing company or the products it manufactures. Both of the mentioned options would indicate attention towards the company and thus are valid and acceptable choices for the study and are also presumed to be reflected in the underlying search volume index values for the company (Mondria and Wu 2011; Ginbsberg et al. 2009).

1.6 Structure of the study

The structure of the thesis is based on five main chapters. The first chapter is the introduction. The second is the theory, the third being the methods, methodology and data and the fourth chapter introduces the results. The fifth and final chapter is about the conclusions that can be made.

Chapter two introduces the theoretical part of the study. In general, it introduces the relevant theories that are needed to exercise this study and to replicate the tests. The chapter is used as theoretical basis for the study. The chapter 2.1 consists of the introduction of the attention based theory and different aspects how it can be measured and what kind of proxies are viable for further study. It also introduces the notable theory called efficient market hypothesis, which is one of the corner stones of current financial theories and explains why this study also indirectly tests the efficient market hypothesis. Chapter 2.2 examines behavioral finance as an alternative to efficient market presumptions. In practice many of the efficient market assumptions must be discarded and a more behavioral approach is chosen as a preference. The chapter then carries on with noise trading to behavioral heuristics and ends with risk management in finance and some aspects related to sociology as an option for approach. Chapter 2.3 includes the relevant studies and theories to Google trends as a tool and as a source of data for this study. For academic studies Google as a proxy is a new tool with only less than a decade of history. The studies are therefore limited and new information surfaces often. Chapter 2.4 introduces the hypotheses of this study by explaining them one at a time. The hypotheses are closely tied to the previous studies but also justified from a theory aspect.

The third chapter contains the methods, methodology and the data used in this study. Chapter 3.1 introduces the regression methods which are used in previous studies such as the ordinary least squares regressions. It also compares the different approaches this particular study has such as multivariate models as a different approach to the same problems.

Chapter 3.2 carries with the introduction of the sample data and descriptive statistics. Some requirements for the data are introduced in regard to the sample size and the types of companies that are viable for further studies. It carries on with the generation of the data series. The sources and also alternative sources for the data are given and listed. The major variables in the study are abnormal stock returns, abnormal trading volumes and abnormal search volume. The abnormality of the values means their deviation from the expected values based on different theories. Later on the chapter the selected data for ordinary least squares and multivariate analysis is shown. The reasons for the selected companies are also carefully determined and evaluated. The chapter ends with the introduction of Granger causality and the presumptions that are required for Granger causality's application with the null hypotheses.

Chapter four consists of the empirical results. The chapter starts with the given regressions and their testing. This is done to both ordinary least squares and to multivariate (ARMA-GARCH). Chapter 4.3 contains comparison of the hypotheses. This means each of the previous study hypotheses and the hypothesis unique to this study are gone through one at a time either accepting or rejecting the hypotheses. There are also notes if the results differ greatly from previous studies. After the results there are robustness tests which mean the same types of regressions are done with different ways to calculate the expected returns. This is to verify that the results hold if the calculated returns differ based on the type of theory used. In this study the expected results are first calculated using CAPM-model, and during the robustness tests using Fama-French three-factor model. If both expected return values produce similar type of results they can be considered to be more reliable.

The fifth chapter introduces the conclusions of the study and reviews possible problems with generalizing the results within a larger context.

2 THEORETICAL FRAMEWORK AND LITERATURE

This chapter reviews the relevant literatures for this study. There are four subdivisions. First two cover the literature specific to attention theory and how it compares the differences between behavioral economics and efficient market hypothesis. Generally this study will not go to deep analysis with the efficient market hypothesis. As a theory it is well known and commonly used within finance. What is not so clear is the relation to Search Volume Index and to forecasting in general. The first chapter also briefly explains efficient market hypothesis when trying to predict market movements based on proxies or historical events.

The latter parts of the chapter are about the theory behind Google Trends and SVI in general. The SVI subchapters are divided into financial and non-financial literature reviews. The chapter ends with the development of the hypotheses.

2.1 Attention and the efficient markets

The attention based theories can be divided into two different approaches. Firstly, by analyzing them through different proxies and efficient market hypothesis. Secondly, by taking the behavioral approach and also using different proxies. The attention itself is very difficult to measure without using proxies, which at best only reflect the attention but are unlikely to ever be exactly accurate. Attention "theory" is very different from the traditional efficient market hypothesis, which assumes a lot of different factors (Merton 1987). In the past decade finance has had an increasing amount of studies focusing on behavioral economics instead of taking the traditional approach. There are few different reasons for that. The biggest driving factor is the assumption that information is immediately at hand for all investors. In reality it is easy to prove this is not the case. Because the traditional finance model requires such approach, information asymmetry accounts for many different approaches for pricing. For traditional models it can be said that: "These models assume that investors have undivided attention to all assets and their corresponding information streams. However that is understandably not the case since attention is a scarce Cognitive resource" (Kahneman 1973). This fact forces the investors or an individual to focus on only few things at a time. With the flood of information the focus and attention of an individual easily becomes very scarce (Kahneman 1973). The lack of infinite information required by the market model and frictions like costs are being explained by Merton (1987). The paper sheds light to the deficiencies of actual markets and focuses on capital market equilibriums under lack of information with transaction costs and asymmetrical information in general. The limitations of attention have been studied both by economists and psychologists.

2.1.1 Limited attention and overconfidence

Behavioral and psychological approach to finance have brought many interesting results to attention. The behavioral approach has studied the actors of transactions. For example, Corwin and Coughenour (2008) study the attention among specialists in finance. The specific study focuses on turnover of their portfolios with NYSE as the data. Their main findings are:

- Most active stocks have more allocation during periods of higher activity/attention.
- During times of increased attention and activity the assets which have less attention, become also less liquid.
- The effects of running out of liquidity lead to price improvements and greatly increased transaction costs.

A significant paper for this study is the study by Peng and Xiong (2006) which shows actor interaction while experiencing limited or asymmetrical information. Firstly, their most important findings are that when actors are unable to acquire the relevant information or they have to make decisions with adequate background information they tend to start focusing on category-level information. Category-level information in this case means a more general approach to investing than careful thorough individual asset analysis. An example of category-level news is the economic conditions deteriorating greatly. Another is, for example, rising unemployment levels within a nation. Secondly, they also conclude that while experiencing lack of information, overconfidence is prevalent. Overconfidence often leads to making decisions based on general sentiment among other investors instead of company related data or information. In many real world cases there is very little if any financial analysis involved when sentiment dictates buying decisions. Since many if not all investors have fairly limited attention it is interesting and valid to ask what are the specific driving factors behind their interests and motives. What are the defining factors in grabbing investor’s attention? Wuoristo (2012) refers to the study by Odean (1999) which is about the problems of picking up individual stocks from extremely large amount of different alternatives. The solution according to the paper is the investors limit their common searches to assets that have gained their attention in the very near past. In reality this means only days or weeks of timespan. Therefore it can be concluded attention is the framework which offers the options of which to choose the investments and make the decisions. The actual investment decisions nevertheless can be of different ends. Barber and Odean (2008) offer different strategies such as contrarian or momentum investing. “The availability of heuristic is one of the most common explanations for attention allocation for uninformed investors (Wuoristo 2012, 15).”

The price limits rely on disposition effect. It means when an asset, stock, bond or a ticker reaches recent highs or lows in price, it acquires vast attention from general investing public. It also is seen to gather more “relative attention” to what the turnover of the asset would indicate. “Advertising expenses work as a proxy by assuming that the more money a company puts into advertising the more familiar it is with investors and the more familiar it is results in higher attention, since investors are more likely to follow firms they know compared to unfamiliar ones.” (Wuoristo 2012, 15) The advertising expenses spent approach is commonly used in marketing, and many companies try to optimize the gain from advertising. Nevertheless, studies are conflicted how advertising expenses influence the stock price in the long term. Few studies indicate some advertising clearly raises the value of the company compared to a situation where no money is spent on advertising. After some marginal value, the value gained from advertising stops and excess money spent has no effect on value. Although the topic is very controversial and between different industries different results have arisen. Naturally the commercial marketing industry pushes on studies heavily that claim investing has clear returns for shareholders (Shiller 2009).

2.1.2 Measuring attention with proxies

Proxies are the easiest way to measure investor attention but they act indirectly. In reality, proxies are the only way to measure something that does not have a quantitative value that can be observed by the researcher. This is a common dilemma for many studies. Many of the commonly used proxies in finance literature are extreme returns, media attention and trading volume (Barber and Odean 2008; Wuoristo 2012, 15). The common approach is to consider the information push. If an individual sees or hears from a single topic often, he is more prone to pay greater attention to that specific topic. This can also trigger interest in the observer and he or she becomes interested in acquiring further information. Wuoristo (2012) mentions a situation where a company exhibits overly positive or on the other hand negative returns or return periods. Therefore, it becomes exceedingly likely the individual investors will pay more attention to that compared to a more stable day in the terms of price fluctuation. Two unique proxies are also mentioned in many studies. The first being price limits mentioned by Seasholes and Wu (2007) and the second advertising expenses spent (Chemmanur and Yan, 2009). Both of the proxies have been found to be trustworthy in modeling and capturing the attention changes significantly and reliably.

One of the most important things about investor attention in this study is how it affects the individual’s behavior. Barber and Odean (2008) have clearly confirmed individual investors tend to be net buyers of stocks that have high levels of attention. This directly indicates the more attention a share gets the more net buyers it will have. The initial thinking behind this hypothesis is when an average investor is choosing a stock for his portfolio, he makes the choice based on items he is familiar with or has heard of (Barber and Odean 2008). Nevertheless when they choose to sell something out of their portfolio, the options are naturally limited to the shares they own. This approach does not take short selling into account. This one-sided approach generally means retail investors as a group lean towards being net buyers of stocks or assets that experience high attention. Like in previous studies, this study also tests this theorem with the hypotheses that are introduced later.

2.1.3 Neoclassical approach and Efficient Market Hypothesis (EMH)

The most central theory within finance is the efficient market hypothesis (EMH). It has been tested and studied frequently, and it has remained in a significant focus on financial studies. Many of the published research papers either directly or indirectly test the EMH theory or test how efficient the markets actually are. A very common topic for studies is to try to prove markets to be inefficient or biased in several different ways. Several years after the development of EMH a lot new theories have emerged, mainly pointing out facts the EMH cannot justify.

Based on the viewpoint EMH has few different meanings. In economics it often refers to the process of price forming. For example, how the changes in demand and supply are reflected in a specific utility's market price. In finance the most common approach is to analyze what kind of information is included in the price of an asset. In short, how well the price of a security reflects the information in the markets. According to the original theory, there are three (or four) distinct levels of market efficiency. Level zero usually indicates a clear statement: markets are not efficient at all, and the prices do not reflect the information available and this rule is very strict. This is not mentioned often though, but the theory is represented with the three levels commonly. The levels are as follows (Elton et al. 2011, 396):

- Weak-form efficiency in which all existing information contained in historical prices are fully reflected (without delay) in current security prices.
- Semi-strong-form efficiency in which all public information is immediately and thoroughly reflected in current security prices.
- Strong-form efficiency in which all public and private information is (without delay) fully reflected in current prices.

There are many debates on what level the markets operate in a real world situation. The strong-form especially is often seen only as a theoretical approach, as no markets are able to reach the requirements for it. Although it is also highly debatable what it actually means being fully reflected in the prices of securities. Jensen, one of the most notable researches of the topic, claims in his study (1986) that, for example, transaction costs to be added into the equations. For example, in a real world situation the strong-form in its most theoretic form cannot exist. But if one loosens the requirements for the strong-form, so transaction costs are omitted (or not taken into account), the strong market efficiency can be achieved in markets with very high liquidity. It requires the price changes to be differentially small and extremely frequent (Jensen 1986).

One of the other problems with the theory is forecasting future returns. Fama (1970, 384) points out the “joint hypothesis problem”. He concludes that the market efficiency tests are always tests on the whole market’s equilibrium and also the same time tests on the expected return models. So without understanding the expected return models, it is extremely difficult to judge on the market equilibrium. This study generally uses CAPM to calculate market returns, but for the sake of the robustness tests also applies Fama-French 3factor models. In final chapters of this thesis the market efficiency will be indirectly tested. As a principle there should be no possibilities of arbitrage or reliable forecasting possibilities, at least not for extended periods of time according to CAPM (Jensen 1986). This thesis will not introduce Capital Asset Pricing Model any further, since it is one of the most notable and well known theories within finance at the moment. This study utilizes mainly CAPM expected returns.

2.1.4 The efficiency of the Finnish securities markets

The Helsinki Stock Exchange was founded in the year 1912. There had been some activity before the given date but it was rather unregulated and the required organizations for day to day functions were missing (Nyberg & Vaihekoski 2010). After the initial start, for over half a century the Finnish stock market was extremely tightly regulated (as in many countries in proximity to Soviet Union). This period ended in the 1980’s, and that point the markets were moderately underdeveloped as in many eastern countries, opposite to the western markets.

The process of deliberation of the securities markets started in the 1980’s and resulted in an boom, bust cycle forcing Finland first into a recession and then to the depression of the 1990’s. Foreign investments and ownership were heavily restricted in the end of the 1980’s and until the beginning of 1990’s. After these periods the Finnish legislation was unified with the European Union standards. “The Finnish Stock market was harmonized with other Nordic exchanges to create OMX (Ripatti 2010, 12; Vaihekoski 1997; Pörssisäätiö 2007).

The Finnish stock market efficiency has been studied directly and indirectly. Studies had begun with Korhonen (1997) and he found the markets at that point to be weak-form efficient. Nevertheless, the Finnish markets have differentiated vastly from other western markets (Korhonen 1997). In most western markets investing has been a sought after activity for non-professional investors for decades. In Finland most of the companies were government owned. The original studies by Korhonen (1997) were conducted with only 18 companies. In his thesis Ripatti (2010) states as follows: ”Prices did fluctuate in a random walk way and no leads or lags were found that could be exploitable.” Yet there are studies of exactly the opposite findings. Virtanen and Yli-Olli (1987) had found many different anomalies within the Finnish market and questioned the market efficiency and the randomness of the price forming process.

2.2 Behavioral finance

Behavioral finance plays a significant role in finance research. According to the efficient market hypothesis there should be no long term possibilities of excess gains. All assets should revert to their estimated future earnings in the long term. Nevertheless, there are several studies proving the other way around. This chapter is divided into seven subchapters. At first the general market sentiment theory will be glanced through. The second subchapter will be about noise trading the possibilities of making money based on sentiments and other investor reactions. The third chapter is about the limiting factors to trading and how arbitraging works according to literature. The fourth, fifth and sixth subchapters introduce psychological aspects and risk management from other than efficient market hypothesis approach. The seventh chapter introduces sociological aspects instead of the psychological factors that have been prevalent in the first 6 chapters.

2.2.1 Behavioral finance and market sentiment

The grand idea behind behavioral theories is that the individual investors do not act in a rational way. Therefore, most of the decisions they decide to make are influenced heavily by biases and also by irrelevant information. As such it can be concluded the actors are being bombarded with several mental and psychological factors (Shiller 2003).

“Behavioral finance provides a broader social science perspective by incorporating sociology and psychology in finance” (Nofsinger 2008, 4; Shiller 2003, 83). Generally the time period of odd behavior vary a lot. Between different individuals there can be a great deal of correlation in decisions. This can be persistent and constant with long periods of time. In reality this means market prices can deviate from their innate fundamentals for extremely long time, and also because there are no limits to the possibilities of arbitrage that deters the informed investors from trying to eliminate the mispricing in markets. Therefore, generally betting or taking opposite positions to the said individuals acting irrationally (in example noise traders) is not generally lucrative. The mispricing in market can stay for a long time and even strengthen over time (Schmeling 2009; Neal & Wheatley 1998). Some traders are more interested in mispricing than others. Generally many non-professional investors act as the securities prices are "given". Depending on the type of the investor, some are more prone to investor sentiment changes than others. Noise traders in specific aggregate the mood of the market more than others (Shleifer 2008). This is not a problem that exists solely among non-professional investors. Opposite to non-professional investors are the professional investors, commonly referenced as institutional investors. Often they are considered to be smarter, more aware of prevalent information and commonly to be smarter traders. Nevertheless these professional investors are still found to be herding in and out of stocks and use momentum-related trading strategies, which are directly against the efficient market hypothesis (Shiller 2003). Since professional investors usually have higher amount of resources, they may end up strengthening the market movements and over exaggerating. So when the market sentiment is overly positive, they end up buying a lot of securities. When the markets are down, or it is a bear-market, they tend to sell in quantities therefore causing and strengthening the market movements out of their balance situations. Trying to make money out of the movements tends to cause bigger movements in markets. Nevertheless the situation may also be the other way around. In some markets professional investors may be aware of the underlying psychological biases, and therefore they try to benefit from them excessively. So it can be used as a clear arbitrage point, that the prices of securities should reflect their underlying value in the long term. According to Schmeling (2007) there are professional institutional investors who do this type of arbitrage regularly.

Some other researches have studied arbitraging further. For example, Baker and Wurgler (2007) state that investor sentiment thoroughly means the propensity to speculate by the marginal investors. In these kind of situations the general market sentiment is the driving force behind asset demand with speculative securities. This demand causes ripples and cross-sectional correlation with other assets. Thus increasing market prices for all assets, despite there are possibilities of arbitrage. The individual characteristic with securities is, it is difficult to determine their true value (Shiller 2003). As an example of difficult to assess companies are the ones that possess lots of intangible assets, have great growth potential or are debt burdened. Each of the factors may affect the value of the security remarkably. Also as another example smaller companies are generally more affected by market sentiment than their larger counterparts. And on a more macro level idea, securities or stocks in general can be viewed from optimistic or pessimistic prospects. For example, there are periods in history where stocks are generally valued with extremely low price/earnings values (P/E<7) within SP500. And then there are times such as market bubbles where the P/E-ratios are generally closer to 30 or even higher. It is unknown if institutional investors are as prone as non-professionals to employ momentum-based investment strategies without noticing. Nevertheless professionals do take part in strong bull markets pushing stock prices over their fundamental values (Schmeling 2007). Often it can be noted, that harder the security to value, the easier it is for noise traders to cause price shifts. Stocks that are more of speculative investments than long term holdings, are generally harder to use with arbitration and have higher valuations when sentiment is higher. Therefore, one should be able to find a contrarian-type of relation between market sentiment and the expected future earnings of securities.

Baker and Wurgler (2006) have found the demand for stocks varies with time and macroeconomic conditions. During better times the demand for stocks is significantly higher than during times of recession or depression. It can be said that “the sentiment fluctuations can affect the demand for speculative or stable and profitable stocks when, for example, there are flights to quality…” (Baker & Wurgler 2007). The irrational exuberance is quite common in the past. The phenomena has occurred with very many different types of assets. The recipe for these is similar in most if not all the cases. The sentiment starts building up slowly at first but then accelerates rapidly. This is called *feedback modeling*. A stock starts to go up and thus creates success stories and happy investors in media. Public attention increases, the news and media repeatedly report of individuals getting rich, word-of-mouth enthusiasm grabs non-professional investors and therefore bringing more and more investors, creating excessive demand. The circle gets repeated and prices tend to keep rising. The longer the process continues the higher the prices get, and more investors want to pile in and have a share of the infinite success of others. In these bubble situations there are only winners in the market. Nevertheless, the boom-bust cycle commonly ends in the bust especially in situations where the prices are way bound rational valuations. The following situation has been tested by two psychologists. Andreassen and Krauss ran tests on the perceptions how people conceived different market situations. In these tests non-professional investors where shown historical prices of different stocks. Then they were asked to start trading in a simulated but fictive securities market. The findings where the people tried to find obvious trends and patterns in the historical prices and trade according to these. In reality it means they tried to extrapolate the price curves and expect the same inclines or declines happen as have happened in the past (Shiller 2003, 91-94). This is clear evidence of the general overall atmosphere being a decisive factor for many investors when they are valuating future expected returns.

It is commonly known noise traders influence market prices, and even to some extent the 2008 financial crisis was thought to be because of excessive trading (Shiller 2003). According to studies this is controversial though. Even though there are several papers concerning the issue none of those can clearly prove decisively which is the correct amount of effect on noise trading to overall market status or equilibrium. Delong, Shleifer, Summers and Waldman (1990) were the first ones to start studying the subject in general. The key finding of their studies was noise trades influencing markets. The second important finding was that they create a new risk called *noise trader risk*. This risk was unheard of before and their studies indicated it is something that should be priced into asset prices overall. The mentioned risk in reality means that the trader's personal sentiment should be taken into consideration. If they are overly happy or delighted, they tend to create asset demand. If they are feeling negatively over the markets they will tend to sell assets thus further lowering prices in bear markets. Both of the feelings are generally unpredictable for other traders. According to the studies these feelings may persevere for extended periods, and on short term the asset prices may not revert back to their long term means (Portniaguina 2006). Such behavior will enhance both bull and bear markets.

When comparing the effect of noise traders on different types of assets, also different types of results arise. Generally according to Baker and Wurgler (2006) the earlier studies indicate the noise trader interest lies within speculative assets. In reality it means assets that they are not planning on holding for more than a very limited period. Because of this De Long et al. also claim that noise traders are actually the ones causing many of the experienced extreme market anomalies. Some noise trader's do not rely solely on financial analysis based on the market prices or estimated future earnings. Instead they choose to rely on pure volatility or volatility derivatives. Because of such behavior the stocks with extreme high volatilities will face even more difficulty in pricing them correctly. Controversially according to Baker and Wurgler (2006) the high volatility stocks earned significantly better during low sentiment market conditions, and on the other hand, experienced very low returns on high sentiment market conditions. The high volatility stocks in their study consisted of high earnings growth securities. This type of results have been later verified, for example, by Lemon and Portniaguina (2006), but their studies consisted of mainly value stock instead of growth stocks. This on the other hand in same cases contradicts the earlier findings because there are same type of behavior but it represents itself in a different way. As further findings for studies of this sort are the ones conducted by Finter, Niessen & Ruenzi (2010). They found that stocks more sensitive to sentiment changes are higher in short term returns, but these returns are reverted in a three month period. With time periods longer than three months the sentiment insensitive stocks are deemed to outperform the sensitive stocks.

As it seems noise trading is a valid and viable strategy in investing in the short term (Finter, Niessen & Ruenzi 2010). According to efficient market hypothesis this should not be the case, not for a very long time at least. If noise trading would be a long term viable option, there should be possibilities to arbitrage that, and with time, revert the noise trading gains to zero.

2.2.2 Noise trading

In many cases investors are divided to subcategories. One approach is by simply categorizing them to professionals and non-professionals. The professional investors are also often referred to as institutional investors. There are other options in categorizing investors in modern securities markets. Needless to say each of the different ways to categorize the investors have their benefits and downfalls. One way to categorize them is:

- rational arbitrageurs,
- smart money investors,
- noise traders,
- and the passive investors that buy and hold.

Behind each of these categories there is a different type of group of investors. For example, smart money investors are deemed to be often professional, possibly institutional investors. They often represent a larger fund or holding or are wealthier private investors that are both interested and have the interests to gather explicit information about investing and assets (Kelly 1997). On the other hand, noise traders are seen to be small-time investors with generally smaller portfolios, also experiencing lower income. Noise traders usually do not want to behave as rational investors, but are prone to taking more risks to possibly reap greater rewards. In reality this means holding possibly only one or two different stocks (Kelly 1997). This means very little diversification. Some of the noise traders justify their trades based on “gut feeling” or just personal beliefs of gaining profits. The literature refers to pseudo-signals, which in many cases mean common investment advice from friends, brokers, possibly technical analysis or any other probable information that is not deemed essential or relevant by financial theories. Also this information is commonly already included in the current market prices (Shleifer & Summers 1990). The changes in stock prices - especially if the prices go up a lot – brings in a lot of new investors to stock markets. Many of these aforementioned investors are not interested in fundamental pricing of stocks or intrinsic values, but more the historical price trend that they conclude will keep on going on in the same direction (Nofsinger 2008). It is extremely common these type of investors buy when the prices are going up, and sell when the prices are going down thus going with the flow (Tokic 2007). Also another type of trend appears. The similar type of noise traders tend to focus on small cap companies that have a bigger amount of outstanding shares. The more the outstanding shares with small caps, the lower the individual share price usually. This leads to many non-professional investors believing they can acquire greater gains because of low individual share price. Also according to Shleifer (2000) they are usually very optimistic about the performance of their portfolio and they tend to systematically underestimate the risks. Nofsinger (2008, 89) claims overoptimistic nonprofessional investors commonly conduct a lot less analysis on securities and also systematically decide to disregard any possible negative news or information about their own portfolio. Often it is that the analysis that are not based on pure financials tend to be misjudged. It is found that the pseudo-signal based strategies also are correlated with each other. The reason behind this is the strategies are originally formed using the same biases affecting the process of judgment (Nofsinger 2008). This can be further vilified by observing how test subjects in psychological tests and exams tend to make the same mistakes over and over again, clearly indicating a pattern. These patterns often lead to demand shifts and therefore will influence the prices of different securities (Shleifer & Summers 1990).

Often noise traders are seen as a problem for the markets, but in reality this may not be the case. Instead of detracting the efficiency of the markets, they are providing more liquidity and according to Black (1986, 531-532) the noise makes the financial markets more possible, but also the same time makes them imperfect. He has argued further that the noise trading may make the market seem imperfect but it is in the end very crucial to overall financial markets and also providing great liquidity for those in need of it. Since many different people have their own beliefs on market movement, a singled out individual noise trader may have his own piece of information. This person generally understand other similar type of noise traders have their own pieces of information and thus is not encouraged to go into trading based on his own beliefs. But as the noise increases it becomes feasible for many more investors to start trading with their own information (Nofsinger 2008). This means it also becomes more feasible to start gathering information and conducting further research and thus deepening the understanding of market movements.

Studies by De long et al. (1990) are inferring that a new risk is born with noise trading. This risk is deemed to be called *noise trading risk* and should be valued into the prices of securities. The risk in reality means more volatility of assets and most noise trader stocks also experience stronger reversion to their mean prices over the long term. “Noise traders can also garner larger returns than rational investors by bearing more of the risk that these assets themselves create" (Ripatti 2010). For the typical arbitrageurs noise traders present a possibility. Many of the assets that are strongly connected with noise traders may be experiencing chronic undervaluation. Thus for the arbitrageurs and long term investors they may represent a possibility of a cheap stock. However, noise traders can also create rather a strong price pressure in some cases. These cases usually are in bubble market situations where traders keep pushing the prices higher and higher even the valuations may be wrong and easily to be noted to be too high. This in turn of course lowers the estimated future returns of the asset and makes the single stock purchases worse investments. However, market sentiment is not only affecting the non-professional investors but also the professionals.

For the smart money investors market sentiments can also cause problems. Noise might be perceived but they do not always make corrections or rebalancing of portfolios because of it. Also, the noise in markets may present an opportunity for the fund managers to attract new investors. The fund managers may think it is acceptable to do noisy investment decisions based on market movements. This is because the general public that constitutes mainly of nonprofessional investors has a strong positive feeling and trust in the fund managers. They are believed to have superior knowledge about the stock market. It may be seem when the fund managers starts trading he or she has superior knowledge and a personal interest to get better results. Therefore, it may seem to the outsiders that the more trades the fund manager does the better he is and the more information he has. He may start trading without any meaningful information (Trueman 1988).

Investment banking often offers financial advisory services as their line of business (Ripatti 2010). These include the possible Initial Public Offerings (also known as IPO's), splits and mergers, various different services related to trading stocks and also investment research services. Since the bread and butter of trading business for banks is the sheer amount of transactions (and turnover), they try to maximize the trades widely and thoroughly (Ripatti 2010). This in return results to giving out significant amount of positive reviews and suggestions. The more the people buy shares, the more volume gets created and the more profit the bank acquires. Especially since the general public are commonly either unfamiliar with short selling or simply dislike it. Another option is for the bank to create extremely biased opinion and review about a certain company, if the said company is a customer to them. This may strengthen their relationship and ensure the business together continues (Michaely & Womack 1999).

Also personal rewarding may cause some biases. The schemes how banks decide to reward their employees may be directly tied to the profit the bank makes. This may prove to be an incentive for biased behavior. It is often more safe to give instructions to buy securities than it is to sell them, because over the long term prices do get higher (Marttila 2001, 146-148). Therefore, suggesting to sell off shares is more because declines happen more rarely and also because deviating for the general consensus could end up costing the analyst his or her job (Womack 1996, 165).

Womack (1996) studied the analyst recommendations further and found out that instructions to buy securities often led to increases in share prices. This is in accordance to financial theories. The price increases were stronger on small cap firms with lower liquidity. This finding also confirms findings of other researchers on the same topic. The situation leaves the investment banks with vast amount of power to influence share prices and also the incentives to abuse these powers on their own benefit.

2.2.3 Limiting factors to arbitrage trading

Arbitraging in markets does have very distinct limits. If this was not the case, in theory it could be possible to eliminate all noise trading. Commonly investors that arbitrage are more risk averse. Their aim is to benefit from market imperfections and at least in theory a perfect arbitrage has no risk involved. This in reality is not the case, because of transaction costs and other limiting factors (De Lang et al. 1990). Because of these facts arbitrageurs commonly do not go against noise traders strategies. For example, the noise trader's beliefs on the market direction may not revert back to their long term means in any sensible time. For example, the market can be valuing companies falsely for extended periods of time. In some cases the extreme over or undervaluation may even get stronger with time, instead of reverting to their intrinsic values. Therefore, if an arbitrageur has to start liquidating his or her position before the arbitrage is fully finished or the stock prices recovered, there may be possibilities for a significant losses (De Lang et al. 1990).

As noted before, the rational arbitrageurs mainly consist of professional investors, for example, managers from hedge funds. According to De Lang et al. (1990) their investment time span for arbitraging may be short. This time span is based on the requirements of the customers of the funds. They can come and ask their share of the fund to be liquidated almost immediately, or they may be in dire need of quick cash and have parts of the fund sold. In these cases the manager is forced to act on these premises. Also commonly the customers are not fully aware of the managers long term plans for the fund, and even the shorter term plans may be cloudy. They are not familiar of the arbitrage positions that need to be kept till the arbitrage is dissolved in a way previously planned. Liquidating too early may give unwarranted and unwanted results (De Lang et al. 1990). The time span problem is very difficult to indicate for the investors. For example, investors can get only monthly reports, while an arbitrage position can take out few months to solve. Therefore, the net returns are not realized on the monthly reports and investors may jump to easy conclusions on the monthly reports (Shleifer & Vishny 1997).

Many nonprofessional investors who invest in hedge funds or generally in funds may act in an unforeseen, unpredictable way thus creating problems for the fund's executive manager's ability to foresee the amount of owners in their particular funds. This can only be solved by not taking long term opposite market positions because the rarely if ever are entirely risk free (Gemmill & Thomas 2002). The funds needed to create the arbitraging positions usually require initial cash that can be acquired by borrowing and therefore creating a steady flowing cost that needs to be covered. As these costs and fees keep accumulating over time, it can begin to represent a new risk to the portfolio (De Lang et al. 1990). However, the lengthier the arbitrage, the bigger the risk the cost of capital composes. Purely because of this the fund managers can only create short term arbitraging (Sheifer & Summers 1990). Even with shorter periods there are very few riskless arbitrage positions, and with added transaction costs, there are close to none. In some cases the arbitrageur can see one stock being overvalued and another being clearly undervalued. He can short sell the undervalued to buy the overvalued, but considering these positions may revert to their long term means during different times, there are plenty of risks to be carried. Many different market events may turn the arbitrage sour. For example, sudden positive short term news about the overvalued company may push the price further up. Or on the other hand, the fundamentally undervalued company may experience a sudden selloff, thus causing a loss on the arbitrage if realized (Shleifer & Summers 1990; Malkiel 2007, 241).

When considering the risks of selling securities short, the following list can be observed:

- bureaucratic risks
- psychological risks
- social risks

This chapter will go through each of these risks. Short selling in general can reap benefits, but also if things go awry create an almost never-ending infinite loss because of rising securities prices (Malkiel 2007). Although short selling is not possible in the first place with some countries since it may have been banned permanently by law or regulated heavily to an extent where it does not happen. Also in many countries it is someway limited or restricted, and during some periods in time it can be banned altogether. Short selling is commonly more difficult task than simply buying assets. For example, many political parties object to this idea, thus creating a political anti-short selling atmosphere. In many financial crisis the short sellers have received the blame for the situation. There is a widespread antipathy towards short sellers (Shiller 2005, 182-183).

The psychological risks on the other hand rely on two unique factors. Firstly the owner of the share might want it back so the position must be closed. Secondly the possibility of growing loss because of rising share price. According to behavioral finance individuals tend to feel more agitated and scared of losses, than they are happy or enthusiastic about gains. Usually trying to avoid the unwanted feelings of loss is countered by making the short sales position even bigger, thus increasing the risks. Although on the opposite side individual investors generally tend to avoid taking large short selling positions because of the innate risks it holds. Shiller (2003, 97-102) also notes that there are times in history of finance where there have been no stocks available for short selling which means no one is willing to lend securities. The demand for the short selling securities originates from few different places. Most notable of these are the hedge funds. Needless to say their objectives are not to return the prices to their long term means, but more or less to bend the market in direction and take the maximum benefits out of that specific direction (Malkiel 2007). Brunnermeier and Nagel (2004) studied the funds after the IT-bubble and found out hedge funds in general never acted in a manner that would return the security prices to their long term mean values. They instead rode the bubble and enforced it as much as they could, and as such many of the hedge funds were extremely heavily leaning towards portfolios of internet stocks even at times when the price per earnings ratios were seen in the hundreds. In a situation where extremely strong growth is needed for the foreseeable future or the stock should be considered as overvalued, or will end up with very low earnings and thus low earnings potential for investors as such. Brunnermeier and Nagel (2004) also found out the outcome of the price race could have been foreseen. However, Malkiel (2007, 241) concludes that buying a security worth of mere 15 dollars can be bought with 30 dollars if the price of the stock can be seen to go to 60, therefore clearly indicating he just follows the market flow and rides the bubbles.

In many cases long term arbitraging possibilities remain rather thin. Many if not all arbitraging positions are taken advantage of in a rather fast manner when they are spotted in open market situations. But there are also few known situations where long term arbitraging positions have remained in the market and have not been entirely closed. De Jong, Rosenthal and Dijk (2008) conducted an extensive study about the topic. They calculated intrinsic values for companies – which they thought would be fair values for long term investments – and compared there values to current market prices. In this way they tried to find out possible mispricing situations or generally any possibilities of arbitrage. They found the most obvious possibilities of arbitrage to remain within stocks that are publicly listed in more than one market. These are commonly referred as dual listed companies. One example of such a case of dual listing is the Shell and Royal Dutch. Second example would be Unilever NV and Unilever PLC. The underlying companies are same subjects naturally and therefore the quoted prices should be the same in different markets (Malkiel 2007). Naturally the currency should be taken into account but the underlying intrinsic value remains the same. De Jong et al. (2008) found out there are significant pricing differences which do carry consistently over time. Their main conclusion was that mispricing exceeded twelve percent on average and the absolute maximum deviations exceeded fifty percent. The second finding was about the standard deviations these companies experienced. All in all, the standard deviations were almost double compared to companies that were not dual listed. The values of the companies were closely tied to the underlying market conditions. In example a stock listed in one market often followed the local market fluctuations, therefore having high correlation with the dominant index instead of following the stock price on other markets or following the underlying valuation (Malkiel 2007).

The most common possibilities of arbitraging rely on more speculative investments than long term buy and hold –type of stocks. The more speculative the nature of the stock the more difficult it is generally to price and the higher the possibilities of external factors influencing the valuation. In these cases psychological factors step in and begin influencing investor decision making. For example, such factors can be the representativeness or overconfidence biases.

2.2.4 Psychological factors influencing markets

The decision making of individuals is often plagued with psychological factors. This applies to both professional investors but also to non-professionals. In both cases the results of such behavior are often suboptimal compared to a more analytical approach. Under uncertainty people's decisions rely often on systematic yet misinformed decisions. For example, the changes in the sentiment investors are facing have been found to have profound implications to asset returns. "The Trading decisions of individuals have been recognized to be highly correlated and one possible explanation for this is the same shared psychological biases among people" (Barder, Odean & Zhu 2009).

Most biases originate from two separate factors. The first is trusting other people too much and the second is overestimating yourself and trusting your own skills exceedingly much. These both factors often lead into over estimating the forecasting possibilities about the future. Malkiel (2005) claims people tend to trust their own personal judgment heavily and therefore they are considered to be overconfident. One example of overconfidence in a common situation is the individual's driving capability or skill. Over 60% of people consider themselves to be far superior drivers compared to the other drivers. This is an obvious bias and it can be verified to some extent from traffic accidents. Overconfidence is often not deterred even after the accidents. The victims of the bias often feel a period of disappointment but in the end return back to their old habits, without trying to alter their behavior even despite the obvious misconceptions. This often proves the biases have extremely strong roots within the human psychology and external factors rarely change them unless the investor at hand chooses to reflect their actions thoroughly.

Another example of being overconfident – and also the same time extremely over optimistic – is when people are asked to predict future stock prices. The tests with Kahneman and Riepe (1998) concluded that when people were given a chance to ask where the stock prices are going with 1% possibility of statistical error, they concluded the prices of securities only could fluctuate less than twenty percent in either direction. While in reality this is an extremely narrow possibility the fluctuations are so little, and also considering in many cases they can easily exceed 50% in reality. It can be concluded overconfidence and over optimism often lead into too rosy future expectations, while the real life is truly more volatile in its nature.

Overconfidence generally means three different aspects as listed here (Barder, Odean & Zhu 2009):

- Over estimating understanding or knowledge about a specific topic at hand.
- Underestimating and belittling risks about securities and market conditions.
- Overestimating their possibilities of taking control of complex situations, in which both results and variables are reliably unknown.

Being in control is a major driver for overconfidence (Nofsinger 2008, 10). There is a specific effect that is named after overconfidence called *magical thinking*. In these cases the victims genuinely feel they are extremely lucky. They feel the next thing will go in their favor if they act in a certain illogical fashion, even in many if not in all cases simply acting logically would yield better results. As an example of such illogical thinking is betting on a flip of a coin in a certain way. Studies show people are more eager to put large wagers on the result of a coin flip when the coin has still not been flipped, than when it is in the air. Another example of illogical thinking might be relation to lottery numbers and already purchased tickets. For example, people tend to believe their picked numbers are some way luckier than an average randomly selected set of numbers. This is also displayed how they demand compensation if they were offered a chance to exchange their current lottery ticket into another one of same financial value, in which the numbers are picked by someone else or purely out of random sequence (Nofsinger 2008).

Shiller (2005, 152-153) has also studied the effect called hot hands. It is based on the idea that previous successful actions on the market are always signs of great astonishing personal skill backed with knowledge and profound intuition. This often leads into excessive trading in reality. Often overconfidence is tightly knit to the beliefs on individual investor's information level. He or she somehow feels the knowledge at hand is somehow superior in some specific way (Nofsinger 2008). In many cases the practical behavior of individual investors is not rational in its nature. For example, an investor may have a view about a security. If the empirical evidence rejects this view, the facts remain ignored and the confidence of the investor receives only a mild bump. If the empirical evidence on the other hand verifies the original view, the investor begins to feel he has special knowledge and his trading prospects activity increases (Barder, Odean & Zhu 2009). Most bad news are ignored or understood as “bad luck”. The reaction to positive news leads into momentum in purchases. After that noise traders pile up on the stocks with positive news despite their risk levels. And when the sentiment starts rising further the noise traders end up buying more and more of the same assets and holding remarkably larger portions of their portfolio in those shares than the so called rational investors (Wang 2001).

In situations like these it can be concluded that repeated success strengthens the investor’s feeling of overconfidence. This type of overconfidence also creates extremely high volumes of trading and also at the same time grants more courage to take on bigger risks generally. Among non-professional investors there are possibilities of creating a vicious cycle. The short term success in trading leads into further trading and even greater risks, and if these get verified repeatedly, it may be the course towards market bubbles. In case of a bubble many participants in the market need to be excited and joining in for the momentum. As unlikely as it seems, these events have happened quite regularly in modern markets (Shiller 2005). During booms or bull markets investor feel the success of their investments come from their own know-how instead of the market conditions and they generally experience overconfidence. During the bull markets trading volumes are expected to go higher. “Overconfidence has been seen as a fundamental factor promoting the high volume of trade observed in speculative markets.” (Shiller 2005, 152-154)

Often being overconfident is combined with ignorance and not understanding the risks. It is common to believe this is a fairly popular trait among non-professional unsophisticated investors, but Dittrich, Guth & Maciejovsky (2005) conclude also professional investors act in a similar fashion. It is easiest to observe when any task difficulty is either moderate, difficult or even extremely difficult. In such occasions overconfidence becomes prevalent. As a symptom of overconfidence excessive trading often occurs (Shiller 2005). Instead of analytical careful observations and analytical probability assessments, overconfident individuals trust more their own private ideas instead of public signals. These private signals and beliefs in turn can turn into market direction at least for some periods of time (Wang 2001). Statistically based on sex men are more eager to turn overconfident. This applies to both to real life activities and to securities trading and investing. Overconfidence has not been a successful trait in the past when observing returns on the strategy. In many cases it leads to more risks than anticipated, more trading and lower returns (Shiller 2005).

According to Allen & Evans (2005) the root and source of overconfidence often comes from the individual’s capabilities of processing complex information. Having strong beliefs and biases distort the decision making process and causes the individual to lean towards easier solutions which are often not correct. Nofsinger (2008) also went further with his research and concluded overconfidence to be heavily tied to false beliefs or illusions about the future. Since investors cognitive capabilities are limited – as is their time span to analyze information also – they jump into conclusions and instead of finding better information individuals compensate the uncertainties by simply acquired more knowledge. This new vast amount of knowledge may be irrelevant to the task at hand.

Based on different types of securities, stocks experience the most overconfidence. A new aspiring growth company may be seen as an opportunity to get rich quick and easy. This type of behavior pushes the stock price ever the higher and with the stock price going up, new investors pile up and the original owners feel the emotions of success, and also start to feel very confident about their investing skills and understanding of markets. As long as the bull market carries the more overconfident the so called winners get. The overconfidence bias is also affected by another bias called hindsight. When the bull markets keep on going many investors start to feel they were foreseeing this and start to judge the past actions differently (Malkiel 2007). In reality the hindsight bias thinks he did better in the past than he actually did. The hindsight bias can be seen as a mechanics to deceive yourself. As time passes the individual’s judgment on the past changes. Past actions present themselves differently and the reasoning becomes clouded. Based on sex females are prone to be more careful and risk averse. Observations by Barber and Odean (2001) indicate females tend to trade significantly less. In example the average turnover of portfolios for female investors was approximately 50%, while the same time men had nearly 80% turnover y-o-y. In Barber and Odean’s study (2001) they found men also gained less returns because of transaction costs. Therefore, it can be concluded the excessive trading does not pay off for the average investor. Although how non-professional investors reflect their losses and gains do differ based on sex. Men we deemed to be more certain of their actions while women questioned the history and their decisions more. Despite this men decide to do more trades (Malkiel 2007).

As a description for overconfident investor literature gives few examples. Grinblatt & Keloharju (2009) found out men are more action oriented gamblers and seek experiences of thrill and excitement, while women do not show such behavior within non-professionals or professional investors. Behavioral finance generally uses the term *sensation seeking*. This occurs most often with stocks but also with bonds to some extent. Bonds are generally less volatile so they grant lesser thrills and sensations. The higher the volatility the higher the excitement, as trading with changing values is exciting per se. The sensation oriented trading strategies are in reality text-book examples of non-profitable noise trading (Malkiel 2007). Often they yield very low or negative results but generate thrill and excitement for the traders. The focus of such behavior often is directed to more volatile and speculative markets. The same markets are also the easiest to manipulate for price changes and bubbles do occur there more often (Malkiel 2007).

2.2.5 Heuristics and manners

The world is exceedingly complex place and the human brain seems not to be capable of processing all the information. Individuals are offered an abundance of information and data, and they rarely have the time nor the interest to carefully analyze all the important aspects to reach proper educated decisions. The human brain as an organ is inclined to make decisions and situations easier than they really are. Thus, heuristics are created. To quickly analyze and process major flows of information and after that taking a shortcut to valid solutions. But in some cases these shortcuts may be misinformed and instead lead to systematic failures and biased yet repetitive mistakes (Bazerman 2006, 13). Kahneman and Tversky (1974) were the original researchers to introduce the concept of biases. The theory they created led into the development of behavioral finance and heavily attacked against the Bayesian estimation process, in which operators were believed to be able to make near-perfect estimations on the outcome of the future. The heuristics that Kahneman and Tversky have used in their studies originally were as the following list will indicate.

- The representativeness bias,
- the availability bias,
- and the anchoring bias.

Each of the aforementioned biases can be tested from statistical data to some extent, although the representativeness bias often represents itself in few different ways. The problems with the biases is how to measure them. This can be commonly done with using a proxy, but there are always some problems or possibilities of misinterpretation with such an approach. For example, if the proxy measures the correct variables or if it is plagued with noise.

The first mentioned bias, the representativeness bias, means mixing up two different subjects that are very alike each other. One example of this could be that a person smiles. Smiling is often believed to reflect happiness, but there can be many other reasons for smiling too. Still often smiling indeed represents the feeling of happiness. This idea relies heavily on stereotypes and stereotypical thinking even at the cost of disregarding the surrounding factors. In these cases statistical probabilities can be used or using average means of the population. According to Bazerman (2006, 22-23) the largest donator to representativeness bias is the lack of intuitive skills on sample sizes. We tend to understand intuition with basic information, but when we are asked to intuitively interpret anything quantitative about sample size the task is remarkably more difficult to perceive. The sample size systematically ends up being ignored on decision making. An example given by Kahneman & Tversky (1974) is as follows: A person is quiet and shy by nature, he likes staying home and reading books in quiet. He is very meticulous and enjoys keeping his items in order, and shows very little or no interest at all in meeting new people and socializing. After the description, test subjects are asked to pick a profession for person in question. Options are farmer, a librarian, an airline pilot and a firefighter. The test subjects repeatedly chose librarian since it fits the stereotypical description of the person, even farmers accounted for ¾ of the entire population in the area. This applies for almost all people unless they explicitly choose to work against the representativeness bias. It can be generalized the sample size is of great relevance when judging using representativeness bias. Unfortunately the heuristics and intuition work against the analytical approach. Even it is very easy and fast to make decisions by using the bias, it systematically causes problems. Also because the samples taken from the population may not be statistically significant, and also because individuals tend to make generalizations on subjects when there is not sufficient data to decide anything at all.

With investment decisions the situation is very similar than described by the Kahneman & Tversky studies. Barber and Odean (2009) later studied the bias among investors. They concluded all investors tend to overweight the information and value given by past securities returns. It distantly reminds of momentum investing strategy, where one expects the future gains or losses to continue happening at the same pace. Instead of expecting returns that are most likely, many investors overweight the returns that are most representative calculated from previous returns. This is against the efficient market hypothesis directly. The representativeness bias is extremely strong when buying securities, but not very prevalent when selling (Barber et al. 2009). Another example of representativeness is the classic coin flip situation. The test subjects are given three different possibilities how flipping the coin five times may end. In the options H represents heads and T tails. The subjects understand the possibilities are fifty-fifty by nature.

- Option 1: H-T-H-T-H
- Option 2: H-H-H-H-H
- Option 3: T-T-T-T-T

In almost all cases test subjects choose option 1 and consider it to be more likely than options two or three. Needless to say all the options have similar probabilities. These misconceptions of probabilities can be devastating for individual investor portfolios. The test and the problem is known as the gamblers fallacy. This type of behavior is fairly common among non-professional investors but portfolio manager can experience from the same fallacy. In some cases a good company does not sum up as a good investment, because the purchasing price needs to be taken into account. For example, a company may have launched a once in a lifetime new astonishing product which boosts earnings incredibly high on short term basis. This may not be a repetitive situation and it may fade out quickly. Nevertheless individual investors experiencing the bias are prone to start discounting the future earnings and cash flows based on the short term historical values, even it is clear the short term earnings may only represent future earnings and are not guaranteed to happen.

The same type of behavior can happen for prolonged periods of time. For example, some long bull markets keep on going even the earnings of companies have already gone down. Investors are discounting historical values into the future without questioning their durability. Malkiel (2007) also states individual investors are trying to find hot mutual funds that have high past earnings. A company with high past earnings can easily represent itself as one where management has “hot hands” and are lucky in picking up profitable projects to participate in.

If an investors chooses willingly or unwillingly to take part in valuation of companies with the use of representativeness between different assets, some errors are prone to happen. According to finance literature the effect is especially strong during bull markets and combined with the rising price level of assets can lead to financial bubbles (Malkiel 2007, 227-231).

The second heuristic or a shortcut to be analyzed is the *availability*. It is similar to representativeness but consists of frequency of events. For example, events that occur more often are more easily in one’s thoughts and are also easier to recollect from memories. For example, when people estimate their chances of getting fatal cancer they at first try to remember if there are any patients in their own blood line. After this they move into relatives and possibly friends with similar perks. Another very similar example to the previous stated is originally by Tversky and Kahmeman (1974) who found similar behavior among middle aged men who had fears of heart attacks. They started to recall similar cases from past from their friends and neighbors and random encounters that they remembered. Generally they concluded that events that are easy to remember are also often considered to be more important of nature than the ones that did not leave a permanent memory mark. The incidents that are easier to remember also often leave an emotional marker. Therefore, the general idea behind availability is that situations that are closer to present time and that are easier to recall are commonly considered also to be more probable. Hirshleifer (2001) also concluded situations that have emotional ties to be easier to imagine and are considered more likely to happen. The roots for the phenomena often lie within personal experiences. May it be experience, past memories or simply emotionally arousing imagination of the individuals, the mind can create alternative versions of reality and thus use these as replacements for decision making (Wärneryd 2001). Often the reason for strong memories is the media. If an individual sees something on television or reads about it from a paper, a certain type of mental image is created. This image may not reflect the truth. The individuals have very limited cognitive resources and therefore can be affected rather easily with external messages (Barber & Odean 2008).

With stock markets the availability heuristic is commonly used (Harvey 2006). With limited cognitive resources it is easier to pay attention to attention grabbing news or broadcasts, thus often leading to investments in well-known companies that advertise a lot and also receive positive news coverage on grand scale. This is believed to lead to overvaluation and bubble behavior with certain assets. As prices go up new investors pour into the same companies and want to participate in the bull market in which everyone is winning all the time. The major factors for luring in new investors to these market situations are high trading volume periods, extreme past returns and also common news that seem enticing.

The last of the heuristics introduced is the *anchoring* bias. This means individual investors use a certain point in time to anchor their decisions to. For example, they buy a stock at a random period in time. Based on that date and price they reflect the return of their investment. The problems that arise from this approach are commonly related to the information. In many situations individual investors only pick information that are related to the initial anchor point instead of estimating the investment in a larger scale. Also it forces people to ignore information that is not consistent with the anchor point. “People often put too much emphasis on first impressions and fail to adjust their opinions appropriately at a later date.” (Bazerman 2006) One of the first proofs of anchoring bias’s existence was made by Tversky and Kahneman (1974) where random participants in the test were asked for an estimate about the number of African nations in the United Nations. The test began with the spin of a wheel. This wheel produced randomly numbers between zero and hundred. After the test subjects had seen the result of the wheel they were asked to answer the question. In this test the result was significantly dictated by the wheel and not by previous information or knowledge about the topic. People were also asked if they recognized the wheel had nothing to do with the number of nations, and they concluded but still answered in accordance to the wheels random numbers. It can be concluded that many people have the tendency to forecast based on the very recent events by simply extrapolating the same results to continue happening. One specific feature to extrapolating time series is the single last value of the time series. This point often is not revised with time, instead the initial anchoring point works as a fertile ground for future decisions. All the news and changes that might affect the company’s future earnings are reflected to this specific point (Harvey 2007). If anchoring is compared to noise trading that was covered in previous chapters there are many similar perks to be taken note of. Like individual non-professional investors also professionals like hedge fund managers have been found to be suffering from anchoring bias. According to studies by Marshden, Veeraravaghan & Ye (2008) analysts are the most likely victims to fall to anchoring.

On general global market level different biases play a role in generating bull and bear markets, and have known to be driving forces in specific market crashes. Often it is seen that over exaggeration is because of influence of biases (Harvey 2007).

2.2.6 Risk management in behavioral finance

Professional and also non-professional investors mainly use two different approaches to analyze risks. The modern classic way of approaching the concepts of risk and risk awareness relies often on statistical analysis, logical deduction and Bayesian estimation of probabilities. The second way is the intuitive approach. There are few typical views to assess this approach (Tversky and Kahneman 1974):

- Relies on images, expressions, views,
- intuition,
- association to feelings,
- and a strong emphasis on trusting one’s gut.

If an individual has strong feelings or emotions about an aspect or activity he or she believes the risks to be lower if the action seems favorable. On the other case if the emotions are of discomfort the risks are perceived to be higher. Often making decisions based on gut feeling or intuition is considered unwise and leading to bad outcomes, even recent studies show analytical approach combined with intuition can be rather productive way of deciding (Ackert, Church & Deaves 2003). How often is intuition then used? According to Schwarz (2010) it depends on the popularity of the task. If a task is a routine one that is commonly used in everyday life the need for intuition is lesser. On the other hand if the task is rather unknown and forces the individual to leave his or her comfort zone, the incentive to use intuition and emotions in decision making is greater. In some cases although it is possible to use intuition and emotions even they are fairly common tasks simply because it is faster to reassess changes in patterns and routines (Schwarz 2010).

In recent studies there have been two separate research branches studying emotional influences. The first is how different expressions and images influence how an individual person as an investor sees securities and how those feelings influence the decisions they decide to go through with. Lucey & Dowling (2005) found out investors tend to considered popular companies that are in the media as good investments. Simply comparing different companies in media one can find out which ones are those that draw the most public attention. Public attention according to the studies translates directly into demand for the company’s stock. This demand in return will drive up the market prices and possible causes a bull run. However, the study does not indicate specific what the probabilities of overvaluation are. One key concept in the studies conducted by Lucey & Dowling (2005) is the meaning of bounded rationality.

Different studies have taken a closer look about emotions, changes in weather and their correlation to securities market excess returns. Hirsleifer and Shumway (2003) concluded there is strong statistically significant correlation between the US weather and the market return. More specifically this study states when it shines and the weather is pleasant the correlation is positive to markets. Rainy days did not have such a significant correlation so the emotions have more emphasis on positive external factors. The researchers also conclude it is exceedingly difficult to create a working arbitrage strategy based on the weather conditions because transaction fees will become too high for it to be efficient. Another example of external emotional factors influencing securities markets are positive sports results. Edmans, Garcia and Norli (2006) claim the trading days after sports team wins are positive while if the investor’s team loses there is no negative correlation. This effect was found to be statistically significant at 99% level in soccer enthusiastic countries such as Brazil, Germany and southern America in general. The effect was existent but not particularly strong in other countries. The tests did not take into account for the presence of other sports activities than soccer. These results lead to the fact the general public react more eagerly to positive news (Hirsleifer and Shumway 2003). The waiting of a positive outcomes from a sports game creates positive charge into the minds of the investors and as such pushes risk taking forward and causes individuals to buy more stocks. The strength of the positive charge vastly depends on the imagination of the test subject. For example, individuals with more vivid imagination are more eager to create positive outcome scenarios and generally get more excited about these ideas. This type of behavior leads into distorted analysis on risks and rewards and often is combined with extrapolating positive realized results or returns from near history. This effect is believed to be one of the driving factors behind investor sentiment (Paterson 2002).

Many of the researches within the field agree emotions, feelings and insights play a significant role in assessing risk and in the decision making process in general (Lucey and Dowling 2005; Paterson 2002; Mehra and Sah 2002). The two extremes of feelings in here are extreme fear and extreme greed. The studies show that valuation factors vary over time based on how greedy or fearful investors are. This can be measured with different proxies, the most common probably being the VIX volatility index. Also one can calculate risk preferences from implied option volatility. As a conclusion for the chapter it can be said that emotions and feelings can play a huge part in market booms and busts, but that they also influence everyday trading situations. According to several studies uneducated and inexperienced investors are more prone in falling victims of trading based on emotions. Often the very thing called investor sentiment is simply the aggregated emotions of the investors, but especially in extreme situations like end of bull or bear markets the emotions of individuals influence the market the most (Paterson 2002).

2.2.7 Sociological factors influencing markets

Often sociological factors influence the decision making. For example, one can receive tips from friends who think the company they have invested in is the greatest thing out there. Many people took part in the US housing bubble, simply because individuals do not want to be left out in situations where everyone is making money. This was recognized one of the driving factors behind the US housing market during the years from 2000 to 2006. Quoting the famous US investors Warren Buffett: “When your neighbors were making big bucks it didn’t make a lot of sense for the common people to stay out. They wanted to be a part of it.” Thus often the combination of flock behavior and easy to reach information and news from internet will drive the boom cycles higher.

2.2.7.1 Flock behavior

Flocking or herding is the phenomena when individuals mimic the actions of other people without thinking themselves (Sornette 2003, 94). The effect varies greatly based on the market the researchers are observing. In some cases listening to the public opinion or following the crowd leads into sentiment extremes as deep bear markets or speculative all time high bubbles (Caparrelli, D’Arcangelis & Cassuto 2004). The effect’s strength depends how well information is spread in the markets. For example, markets with very transparent dissemination of information bubbles and bear markets are rarer. On the other situations there are markets where information is scarce and only a handful of individual investors have access to it. Herding occurs and investors follow the few that have better insight and have access to more quality information. This is known to lead into bubbles (Zhou & Lai 2009, 42).

With flocking there is a distinct difference between what is private and public information. When the public information is scarce the private information becomes very valuable that only a limited amount of people have access. This is commonly part of the bubble creation process. Hwang and Salmon (2004) found out that when private information is suppressed in markets, herding or flock behavior becomes very common and leads to situations where the market prices of assets no longer reflect their underlying value nor the information at market. The effect is called *informational cascade* according to them. In reality this means the individual investors do not act according to their private information but instead replace this falsely with public opinion. The individual investors simply mimick the behavior of few other investors. Sornette (2003) also states this situation holds for the general public if it holds for individual investors.

Further studies by Sornette (2003) have shown it is very common for investments to be done in open market situations. This means both transactions – selling and buying – are known by other people operating also with similar assets. These open markets or networks include fellow professional and non-professional investors, friends, family and random members of other communities. These networks influence the decision making significantly, but there are no further studies how the effect applies for different groups of these networks. According to Sornette (2003) the herding effect is because of two separate factors. First ones are the networks investors have and the second one is media. Under uncertainty individual investors tend to use networks more than the media, but under normal market conditions media also plays an significant part. Nofsinger (2008) claims that in US fifty percent of the investment decisions for non-professional investors are done because they received a tip from someone in their networks. This is an evidence that if a major part of the investor population feels it is the time to buy or sell stocks generally, this can and will influence the market prices and push the sentiment into one direction. Wärneryd (2001, 205) also concludes that in modern times internet has become a vessel for disseminating information and is a very relevant part of networks. The ability to chat online, share views on forums or bulletin boards and use messaging services extends the ability to influence other people's opinion significantly. Such behavior will influence the market prices (Tay 2009, 207).

When comparing which group is the most prone being under the influence of herding, the academic results are conflicting. One might conclude that moving in herds is common for the non-professional uneducated investors who are ignorant of their surroundings. Academic literature seems to differ with the hypothesis. Herding or flocking is known to occur with different types of investors and surprisingly often. Grinblatt, Titman & Wermers (1995) have studied the same effect among professional investors such as pension and hedge fund managers. During strong bull markets even the professional investment officers fall subject of herd type of behavior following the footsteps of others in the same direction. Such behavior appears to be surprising considering the background of many investors yet they repeat the known mistakes again and again. There reasons for this are complex according to literature. The first reason possibly is the investors generally tend to use same type of forecasting models and expect certain type of behavior from the assets. When the traders or investors start using the same type of models the returns for their portfolios become synchronized. This causes all of the portfolios to start following their own benchmark index and when they are comparing results, everyone appears to have done equally good. This helps them keep their jobs and thus not failing in generating subpar results, even the absolutely values for the returns might be questionable. For example, if there was a general conception that the stock of Google will go up since the company seems to have good prospects and a working business model. One trader or investment advisor begs to differ and tells his or her clients the prospects of Google seem dim and they should sell their stocks. In reality, the price of the stock would go higher and the investment advisor who recommended to sell is seen as a bad unqualified advisor. He or she probably loses the job and status quo remains in the market, and the ones advising according to the public opinion get to keep their jobs. According to literature there seems to be very few incentives to actually do good instead of remaining in the comfort of public opinion (Grinblatt, Titman & Wermers 1995).

So is flock type of behavior present always even with professional investors? There are several studies on the issue and, for example, Zhou and Lhai (2009) found that the investing environment is more critical in herding to form than the type of the investor. In extreme market situations such as the ending periods of bull and bear markets, flocking behavior is almost always recognized. The findings of these studies commonly are that speculative stocks and small caps receive larger flocks of people pouring money into them. This is also prevalent in securities that have very high price to book ratios. For example, because of possible high growth prospects or speculative valuation in general.

In general, it can be said that individual investors tend to act as a flock and invest in companies other people are investing especially during bull markets. Everyone seems to be making money and no one wants to be left behind. The information that provides the momentum for this effect comes from the public media or internet forums commonly.

2.2.7.2 Media and internet

Chat forums in internet have become a common everyday way of disseminating information. Individuals use Google as a very important source for information. This new source of information has increased the amount of messages we receive every day. Google itself provides advertisements mixed in the search query results, as well as the landing pages are littered with different types of electronic advertisement. These messages obviously alter our perception of reality. Going back in time about two decades, the main source of information about investments most likely was only news as in papers as partly as in television. These both channels have commonly been plagued by sensationalism: selling topics that are quick and easy to read and have profound implications to emotions such as fear, love, and curiosity. These emotions in particular have extreme effects on our behavior. The more feelings the news arouse the higher the revenue of the news companies (Read 2009). This leads into very sensational news broadcasting which may either entice or scare investors depending on the market conditions.

One of the most common topics for today's news is financial or someway economy related topics. For example, changes in unemployment rate, changes in currency or politics related to these topics. They often catch the common eye and receive both clicks in internet pages and also purchases of papers. Stock markets in general offer very mixed feelings for the common public. Some people consider it as a way of investing while others can relate to more gambling related approach. During bubbles its common for people to be shown on TV who have made significant amount of money and thus creating an atmosphere where everyone investing in stock markets are believed to be making money. During recessions the channels are filled with disappointed investors who lost their life savings. Both of the time periods are very dramatic and offer a lot of revenue for media companies to bathe in. According to the studies by Shiller (2005) financial and sports related topics count for roughly half of the topics in national news in US. Both of the areas are related to fast moving or changing atmosphere, exciting events and surprising results. In many cases investing in the media is treated in a similar way to gambling instead of careful financial analysis. According to Shiller (2005) the proportion how interested the common public are in investing does vary over time. In some periods the participation rate is much higher, but after market crashes and in the end of boom-bust cycles the public steers clear of stocks. Closer to peaks of bull markets the investors flock back into stocks because the commonplace is to make money easily. Because the public demands scandals and drama to remain entertained, the magazines and television broadcasters meet this demand. This is a requirement for the media to survive in competition (Shiller 2005, 60-90).

In the late 1980's investing was not too common in the States and even rarer in the European zone. Nevertheless, after this the investing behavior has steadily been rising over the decades, peaking in the last IT-bubble. The bubble made many people extraordinary rich but for many others was the ruin of their personal finances. What was clearly a bubble afterwards was thought to be a great opportunity for common investors at that current time. Afterwards it has become clear it was a bubble and also that simple herding and flock type of behavior was exceedingly common in the beginning of the 2000's. For media influence, the higher the prices of the stocks the more news to be had, the more happy faces in television and the more revenue generated by media and advertisement sales. Therefore, it is obvious the media companies have an incentive to push individuals towards herding. According to Shiller (2005) the media is actually an active participant in markets, creating as speculative environment as possible to create more speculative price movements. These movements are then interesting for the general public.

In many cases it seems to be easier to sell negative news and scandals then information where nothing is out of the ordinary. Because of investor's fears this usually leads into price dips and excess volatility. This creates more reporting for the news companies.

Tetlock (2007) found out small companies are more prone to be influenced by news and react easier to positive and negative news. Pessimism and optimist both vary based on the time and effect the securities prices significantly in a way that is not justified with their underlying values. The other findings of the studies are that small companies are commonly owned by the less informed investors and react more strongly to news and market optimism or pessimism.

2.2.7.3 Extreme market conditions

There are several notable historical events with securities markets. Black Monday is one, where the index named Down Jones crashed over 15% in few hours, and for the day over 20%. The event took place in 1987. This was mainly blamed on automatic trading with computers. In the 19th century there was a bubble with railroads both in the old and new lands. The possibility to haul goods with very low costs were believed to create investors notable profits. They did, but at a significantly later time than presumed. 1929 had a crash where the combination of industrialization, mass production, investment banking and mass marketing were believed to create near infinite possibilities for investors. They also did, but at a significantly later time and thus the 1929 was rather bad for the investors.

As general statistics the stock markets do not seem to act like they were normally distributed in the long term. Stock market returns are close to normal distributed on short term but long term far from it. There are ample amount of examples where the returns deviate massively to both directions. The US markets generally are considered to be the most efficient markets, and as such the other markets deviate even more. Markets in smaller underdeveloped countries are more prone to fluctuations. These are commonly caused by macroeconomic factors but also because of lack of buyers for securities. Illiquid markets are easier to manipulate and experience higher volatility (Shiller 2002). There are few studies how much the liquidity of the markets influence the formation of prices. Sornette (2003, 2-8) concluded the same result as many other researchers. If the markets experience illiquid behavior the price changes are significantly stronger. Currently many of the US markets experience over 80% of the trading done with computers, while there are no accurate figures for the Nordic or Finnish data, its presumed to be somewhere around 40% from daily trading.

The strongest bull markets where prices keep on going up despite underlying values or macroeconomic conditions have been experienced in the past two decades. The biggest bubble most likely was the notorious IT-bubble around year 2000. This great bull market lead to at first extreme positive returns on assets, and after the turn of the millennia, to extreme negative returns. Neither of the returns can be justified with financial theories nor normal distribution expectations. According to Ferguson (2009) the returns seen at those times should not exceed 20% ever and even the 10% returns should be extremely rare. With bull and bear markets it is very common for individual investors to change their investment horizon. Generally in case of downturns fear grips the minds of investors and the horizon become shorter. It is considered that the shorter the investment horizon the less the risks experienced. Then the average holding periods for stocks go down during bear markets and trading creates more volume. This effect is strengthened if investors are in danger of losing money that they would require elsewhere, which is the case with non-professional investors often. Also based on the market atmosphere investors seem to weight in different things. During bull runs the focus lies within the positive news and many of the bad events are disregarded. On the case of bear markets, the negative news remain in close focus while the positive signals are left unnoticed. The markets are at extremities controlled by the feelings of fear and greed (Marttila 2001). Psychologists, sociologist and behavioral finance in general have repeatedly tried to tackle why investors are so eager to react to fear and greed. The result for this seems to be that individuals are controlled and ran by the biases and trends in the markets, while flocking and herding is quite common especially in the extreme market situations. The other extreme market maker besides feelings is the access to excessive liquidity. This has been evident in many of the market crashes during the past century. 1929 crashes before the great wars were at least partly influenced because of massive amount of nearly free liquidity in the market. Even the crash did not remove the liquidity, even it did become scarce (Ferguson 2009, 100-105). Although the difference was with the IT-bubble compared to the 1929 crisis in the required collateral. In 1929 only five to ten percent of the stock's current price was needed for collateral thus creating extreme leverages. In the latter bubbles the collateral from banks and institutions have been significantly higher.

So what is the decisive factor a bull run turns into a bitter crash? A crash can be caused by internal or external factors to market. Initial impulse may come from either but it is often followed by sentiment change among investors which turns rapidly to panic and flock-like behavior. Sornette (2003) has conducted several studies on the topic and concludes just before the crashes markets seem to be fairly stable with good macroeconomic situation. After that the situation changes to worse by surprise and catches most of the investors without preparation. This leads into panic, herding and mass selloffs of securities causing lower prices and thus feeding the cycle. The economic cycles ending in bear markets have been a common situation for hundreds of years, although with the market activity increasing over the last century they have occurred more often (Sornette 2003). The Finnish stock markets have experienced some bubbles same like other Nordic markets. The success of Nokia was the initial trigger for some US trusts to find their way to Finnish markets. This quite likely acted as a driver for the IT-bubble atleast. The markets were steadily gaining momentum prior to year 2000 (Marttila 2001, 75-85). The last global financial crisis starting between 2007 and 2008 had several factors contributing to the crash. Some of the reasons yet remain the same from the previous times. The massive amount of liquidity and debt on markets is a driving factor according to several studies (Shiller 2009). Although in the latest crisis there was also new financial products that were to change risk management. Nevertheless, the bear markets started in a very similar fashion than in 1929 and in many other cases. At first there was a major selloff that was followed by panic and mass fear combined with herd-type behavior (Shiller 2009).

2.3 Literature specific to Google Trends

2.3.1 Non-finance specific literature on Search Volume index and Google Trends

Google Trends as a product has been released in 2006. After this it has been developed actively and is considered an important tool in search data analysis (DEG 2011). Google Trends started providing CSV exporting around year 2008. After that point a massive amount of statistical data became available to general public. "Google makes public the Search Volume Index (SVI) of search terms via its product Google Trends. Weekly SVI for a search term is the number of searches for that term scaled by its time-series average (DEG, 2009, 2)." The general literature about SVI is rather ample even it has a very short history dating back only 6 years during the writing of this study.

However, the use of internet to predict real world phenomena is a common event also in the past. However, it is easy to understand the rise of Google Trends and similar products from the competitors of Google. One of the most famous studies is the Johnson et al. (2004) which indicated flu influenzas can be predicted with internet web browsing history and patterns. "The results were moderately strong and no clear connection could be established at the time, however it paved road for future studies (Wuoristo, 2012, 8). At this time Google did not provide an easy way of gathering data so the researches had to get hold of the web browsing history with other means. The study started a trend of health and influenza related scientific studies. One of these is the Cooper et al. (2005) that studies cancer and internet search relations. Many other studies were also conducted, but many of them did not have statistically significant results. The first study with Google trends to show clear and distinct predictive power of internet searches was the Ginsberg et al. (2009) which showed flu influenza can be predicted two weeks before Center for Disease control and Prevention (CDC) reports. This was a significant finding since it proved many of the previously used methods had become obsolete because of SVI. The Ginsberg et al. (2009) has become the cornerstone for many of internet search related studies and is considered basis for academic research according to DEG (2009). The influenza trends can be now found directly from Google Trends. It is uncertain if Google has further plans to incorporate other premade searches into their Trends product.

In studies conducted by Googles Chief Economist Hal Varian, he suggests SVI can describe interest in many different economic activities in real time or with very little delay. A study claiming such is by Choi and Varian (2009). The study indicates SVI has predictive powers in house sales and tourism trends. The lags of predictive powers differ though. The presumptions in the study are straightforward. An increase in SVI is a clear indication of plans to travel to a location. The aforementioned study observes the different queries for the term “Hong Kong” from specific nine different locations and compares the given indices to Hong Kong Tourism Board’s monthly visitor statistics. The HKTB includes the location of which the travelers come from. In that study they find very strong correlation between the two given parameters, only omitting Japan. Several correlation studies have been done within movie industry. Goel et al. (2010) found a strong correlation between box office revenue and search volumes. This is to date used as a common metrics to analyze cash flows for premier movies. Similar studies have been done in music and media industry in general. The correlations vary usually quite significantly, but in most cases are still statistically significant at 95% (5% chance of being entirely something else) confidence level. Predicting box office revenue for premier weekends for movies has been by far the strongest in prediction power, while sales for music CD records its fairly low if none existent in many cases. This may however be explained with different sales channels and with differences in how the product is delivered. For example, music can be consumed using channels such as Spotify or iTunes, whereas movies are usually distributed only in movie theaters at first. The channel undeniable makes finding correlations harder as observed by Goel et al. (2010) and also Lui et al. (2011). Therefore, SVI is not a tool that can be used in every single situation but fairs strongly with single channel sales.

SVI has also been studied among voters by Lui et Al. (2011). Their findings were fairly mixed. Conclusion was that there is no real valid correlation with SVI and election results. Nevertheless it is obvious search queries are made to find out information about candidates and the events that evolve around them. Therefore, even elections have in reality only one channel to show the voter activity (vote for candidate, or do not vote at all), the study indicates no causality or predictive powers of SVI. Although the higher the SVI, the higher the participation rate among voters. So they conclude the more attention the elections get, the more information the retail voters want to acquire.

SVI is being used fairly regularly in every day economics to forecast consumer trends but also more common variables such as inflation, unemployment rate and loans demand among few things. Central banks use such tools as advisory data for decision making. Originally this was mentioned by Ettredge (2005) but many later researches have mentioned such variables. In US and UK the SVI for unemployment benefits correlates with economic cycles, and it has modest predictive powers. Such studies are also conducted by Baker and Fradkin (2011) and specifically among unemployment by Askitas and Zimmerman (2010).

Inflation has also been a very central topic among correlation research. Guzman (2011) had implications with inflation and global SVI. His main findings were: ”Google Inflation Search Index (GISI) has the lowest forecast error of all the inflation expectation indicators tested.” Journal articles by Goeffrey (2012) show extreme significant predictive powers of SVI and volatility in foreign currency markets. Such research have severely questioned the efficient market hypothesis since most studies indicate statistical significant predictive powers of SVI with 99 percent confidence levels.

2.3.2 Finance specific literature on Search Volume index

A common research topic is how SVI occurs with price pressure of stocks. An example of this is initial public offering –situations (IPO). In several occasions a direct relation can be seen. “A natural context where such price pressure may occur is during a stock’s initial public offering (IPO). Since trading-based attention measures are not available prior to the IPO, SVI offers a unique opportunity to empirically study the impact of retail investor attention on the IPO returns. Ritter and Welch (2002) and Ljungqvist, Nanda and Singh (2006) argue that over-enthusiasm among retail investors may explain high first-day returns and low longer running returns for initial public offer stocks (Loughran and Ritter, 1995 and 2002)” (DEG 2009, 4). Many of the following studies find the similar outcomes. The higher the attention higher the SVI peaks. Nevertheless in some cases the initial spikes in SVI are being reversed within either a few weeks period, or in a one year period. “We also document significant long-run return reversals among IPO stocks that experience large increases in search pre-IPO and large first day returns post-IPO. These patterns are confirmed using cross-sectional regressions” (DEG 2009, 4).

"In summary, we find that SVI is related to but different from alternative proxies of attention proposed in the literature, highlighting the distinct feature of SVI in capturing the demand for attention or active attention on a real-time basis” (DEG 2009, 4). To summarize the use of SVI in finance it is in most cases about quantifying attention to companies, stocks or assets in general (DEG 2011; Modria and Wu 2011; Modria et al., 2010). The mentioned studies mainly observe the changes in attention towards different assets. The sources of attention vary depending on the study but central key terms would include how many times an event is mentioned in media, possible higher or lower returns than expected or general investment sentiment, for example, with Russell 3000 stocks during time period of 2004 to 2008 (Modria et al. 2010; Wuoristo 2012). The most important findings of the previously mentioned studies are:

- SVI is a leading indicator when compared to any other proxy,
- SVI captures the attention immediately without any lag terms,
- There can be no extreme returns among stocks without pre-existing investor attention.

This leads to an event study –related approach. Many excess attention peaks evolve around earnings announcements or just after profit warnings. One of the important assumptions with SVI is that a market center such as a local stock exchange or bank services in general are used mainly by retail investors. Therefore, the “more informed” investors such as professional investors or fund managers, are more prone to use exchanges like NYSE or Bloomberg terminals. Professional investors are believed to have an easy access to Bloomberg, while retail investors have easier access to Google and rely on the information from the search engine. The study also finds that SVI is fairly strongly correlated with a Barber and Odean’s (2008) price pressure hypothesis (Modria et al., 2010; Wuoristo 2012). One STDEV increase in given abnormal search volume index (ASVI) will lead to a positive price fluctuation among Russell 3000 stocks. However, this does not apply to the sample thoroughly. “The positive price pressure is only present in the smaller half of the stock sample and is stronger in retail investor driven Dash-5 trading volume than total trading volume. Price reversal is evident after the third week, and the positive change is completely reversed in under one year” (Wuoristo 2012). According to the studies conducted by DEG (2011) abnormal search volume index (ASVI) is the only quantifiable attention measurement that has predictive powers and can foretell the price reversal also. The later part of the study focuses on product oriented correlations and with IPO price changes. The main results are that if there are significant changes in SVI within the IPO week, the price usually has an upward trend for two to three weeks prior to the IPO event. The price spikes during the IPO week but is shortly reverted to pre-IPO levels in the coming few weeks. Interestingly SVI statistically predicts first-day IPO (excess) returns, and also that high ASVI IPO’s underperform lower ASVI IPO’s because there is significantly lower price pressure (DEG 2011).

There are also more studies on the topic. The aforementioned Mondria et al., (2010) and Mondria and Wu (2012) base their hypotheses mainly from the perspectives of Barber and Odean (2008). The price pressure hypotheses are used as grounds for study. SVI is used as a direct attention proxy for stocks from S&P 500. These studies also divide the searches by using Google Trends location filters by choosing different SVI’s from queries originating from different locations. In those studies they also divide the locations by states within US. “In the first study they evaluate the effect of home bias by analyzing search queries and show strong support for the anomaly, since local investors disproportionately search for local companies. They further expand their analysis to show that when local attention rises without a similar increase in non-local attention it indicates that some internal local news has entered the market” (Wuoristo 2012, 11). This local versus non-local effect is seen to be strongest in distant, remote areas. The reasoning behind this is distance. It is presumed in the study that information spreads slower in remote areas. The predictability or estimation of future expected results is also heavily based on local attention according to the study (Mondria and Wu, 2011). On the other hand, DEG (2011) does not find these results. In their studies, SVI is consistent with both local and non-local queries. Apparently there are no other studies that have tried to replicate these previous studies exactly, so it is unknown if the situation with opposing results is prevalent. Nevertheless, both findings are statistically significant, although they are opposing. One possible reason for the results is the data frequency. DEG (2009) and later in 2011 uses weekly data. Mondria and Wu (2011) use monthly data. It is possible the differences are because of frequency, but this has not been verified or replicated after their initial studies.

The latest more known research according to Wuoristo (2012) is written by Vlastikis and Markellos (2012). Unlike the previous studies which evolve around investor attention, they used “information demand” instead. The basic idea of the study is the investors crave for more information when their risk aversion increases. In reality that means the more they want to take risk, the more information they want to acquire. The Vlastikis and Markellos (2012) study uses only 30 of the largest market cap companies from New York Stock Exchange. They compare information supply – which is the amount of news in this study in Reuters – compared to SVI. Their findings are not similar to previous studies. According to them information demand is mainly driven by historical volatility of the company stock, and more importantly the trading volumes for the asset. Supply on the other hand is extremely periodic and systematic. This study is also the first study to show that using expected risk premium for S&P 500 as a proxy for time-varying risk aversion means that information demand increases with the level of risk aversion.

According to Wuoristo (2012) the following table will depict the relevant studies with their names, dates, data range and frequency, motivation of the study and which is used for SVI proxy.

Table 1 A summary of studies analyzing Search Volume Index (DEG 2010; Wuoristo 2012; Dimpfi & Jank 2012)

Researchers	Name of study	Journal	Year	Data	Range and frequency	Motivation	SVI proxy	Results
Da, Engelberg and Gao	The sum of All FEARS	Working paper	2010	Russell 3000	2004-2008, Weekly	Measuring market sentiment via search volume.	Queries related to household concerns (bankruptcy, recession, credit card debt)	Increase in SVI proxy is associated with low returns today and predict high returns tomorrow. They also predict excess volatility and daily mutual fund flow.
	Internet Search and Momentum	Working paper	2010	Russell 3000	2004-2008, Weekly	The momentum effect strength in relation to search volume.	Company ticker signs	Companies with high SVI have stronger momentum effect.
	In search of Fundamentals	Working paper	2010	Russell 3000	2004-2008, Weekly	Revenue surprises and earnings announcement surprises in relation to search volume.	The name of a firms most popular product	Product search volume has strong predictability for returns around earnings announcements.
	In search of Attention	Journal of Finance	2011	Russell 3000	2004-2008, Weekly	Measuring investor attention via search volume.	Company ticker signs, product name and company name	Ticker search volume predicts future share price movements and price reversal.
Mondria, Wu and Zhang	The determinants of international investment and attention allocation	Journal of International Economics	2010	S&P 500	2004-2009, Monthly	Measuring home bias via search volume.	Local attention versus non-local attention on company ticker	Local companies are searched significantly more locally than non-locally.
Mondria and Wu	Asymmetric Attention and Stock Returns	Chicago Meetings Papers	2011	S&P 500	2004-2009, Monthly	The effect of asymmetric information, as measured by search volume, on share price.	Local attention versus non-local attention on company ticker	Local asymmetric attention predicts stock increase.
	Familiarity and Surprises in International Financial Markets	Working paper	2012	S&P 500	2006-2010, Monthly	Measuring how attention to local stocks and foreign stocks differ using search volume.	Foreign equity exchange index ticker	US Investor attention is different from foreign stocks. Attention peaks during foreign market downturns.
Vlastakis and Markellos	Information Demand and Stock Market Volatility	Journal of Banking and Finance	2012	30 largest stocks in NYSE	2004-2009, Weekly	Measure information demand via search volume for company and compare it to information supply as measured by Reuters news articles.	Company name	Information demand does not correspond with the supply.
Dimpfl and Jank	Can Internet search queries help predict stock market volatility	Working Paper	2012	Dow Jones index	2006-2011, Daily	Measure investor confidence via search volume.	Dow Jones index name	Search queries Granger cause index volatility.

After 2012 there has not been many new studies that would have different results than the original studies had. Many of the studies are replications of the original versions with same results. The purpose for these studies has been to replica the original findings and test for market efficiency under different market conditions. Nevertheless, some of the studies have been replicated by other researchers but mainly with the original data sources and longer time-series. Most studies have a distinct focus on generating market-level results instead of focusing on forecasting values for single individual companies.

2.4 Development of the hypotheses

This chapter covers the hypotheses. The first subchapter describes the basic hypotheses that will be studied and tested later. The second subchapter compares the hypotheses of this study to the ones previously tested by DEG (2011) and Wuoristo (2012).

2.4.1 Hypotheses in this study

This chapter introduces the hypotheses concerning SVI and company shares. The hypotheses share the same OLS principles as in previous studies conducted by DEG (2011) and Wuoristo (2012), with two entirely new hypotheses that are H5 and H6. Also hypotheses H1 to H4 are tested with multivariate methods where there are initially positive correlation with OLS. Another difference is that instead of using ticker queries, SVI includes the company name as proxy. All of the hypotheses also differ in data compared to DEG (2011) and Wuoristo (2012). The study hypotheses are as follows:

- Hypothesis 1 (H1): Search volume index captures the general attention in the Finnish market.

The hypothesis simply means if individuals use Google search to find information about companies. For example, if a company name would receive the value of zero with Search Volume Index, it would mean individual investors do not use Google for finding information about the company. Also if there are no fluctuation with the SVI for the time-series it can be questioned if it is working as intended. The SVI is presumed to be fluctuation based on news, interim reports, profit warnings and macroeconomic factors.

If SVI and a stock’s trading volume have an apparent link it can be quantified by means of statistical analysis. Previous studies indicate there are some differences between US and UK retail investors, but there is no prevailing information or data about the Finnish investors in this sense. The previous studies assume that increase in SVI results in higher demand for the stock.

“Based on the price pressure hypothesis of retail investors presented by Barber and Odean (2008), and previously shown to be accurate in the US market by DEG (2011), this paper states the hypothesis that with an increase in company ticker SVI, the company share price is more likely to rise than decrease (Wuoristo, 2012).” However, it is utterly unknown if there is a complete reverse effect during any price period. Previous studies such as DEG (2011) found such phenomena to exist in one year period. There are also studies such as Wuoristo (2012) who find different time windows for this effect.

Because there are no other known studies in using SVI with Finnish queries it is unclear if the variable SVI_FIN has positive values. The variable measures the SVI from local sources. For global queries the Google searches are more studied and as a priori information the values are known to be positive. If the peaks occur with local queries and news, interim reports or profit warnings the SVI captures the attention. Also based on DEG (2011) during the SVI peaks there can also be abnormal return peaks.

- Hypothesis 2 (H2): Search volume index correlates with the stock market trading volume but the trading volumes revert back to their mean values with some lag.

This hypothesis examines the correlation between the variables and the length of the effect period in a more precise manner. The short-run effect in share price has previously been stronger near initial public offering (IPO) situations like described in DEG (2011). However what is the long term effect is unclear and should be studied case by case. The hypothesis will be studied at first with OLS. The companies that experience correlation either from Finnish or global sources will be taken into further examination for multivariate models.

- Hypothesis 3 (H3): Search volume index can be used to predict and forecast an individual company's stock turnover or abnormal returns.

The third hypothesis is as it has been in the previous studies. "The third hypothesis presented in this paper is that the effect of price pressure due to individual buying activity should be more present in smaller stocks" (Wuoristo, 2012, 19). The idea behind this hypothesis is that smaller market cap stocks usually have lesser liquidity and therefore the prices are more easily affected. The smaller the stock in market cap the greater the price impact. This has been observed initially in DEG (2011) but also verified by other later studies.

- Hypothesis 4 (H4): Companies with high consumer public exposure are more affected by retail investor attention than companies that are not visible to consumers.

The fourth hypothesis is similar to the initial study by DEG (2011). It examines the behavior of individual non-professional investors which are also called retail investors. Chemmanur and Yan (2009) found retail investors are more prone to buy into stocks they are familiar with. Testing of this hypothesis is easiest done among consumer goods related companies. In Finnish market such companies would likely by Kesko, Stockmann, and telecompanies such as Elisa and Sonera but also utility companies such as Fortum. The exact companies will be listed in data gathering. The initial proposal is that companies with higher consumer affiliation will be more influenced by retail investor attention measured by the SVI.

- Hypothesis 5 (H5): The SVI results from Finland are better predictors for Finnish based company share price than the SVI results gathered from global queries.

In previous studies such as DEG (2011) the fifth hypothesis has been with the way in which data gathering has happened for SVI. It is given the data will have “noise” in it because the nature of the queries. However, the Nordic companies company names are in almost all cases simple to understand and very difficult to confuse to any other words with a similar meaning. At first DEG (2011) and then Wuoristo (2011) have used the method of narrowing down the markets to reduce the effects of noise. More importantly the effect of home bias also supports the hypothesis. The presumption is that local investors are more inclined to invest in local stocks which are familiar to them and they have experience of. There are studies like Tesar and Werner (1995) which compare the home bias effect in different markets. Their findings suggest that the UK has the least effect of home bias compared to the other bigger markets such as Germany, Canada, US and Japan.

- Hypothesis 6 (H6): According to Granger causality, SVI has forecasting power to trading volume with selected companies.

The sixth hypothesis requires a VAR-model to be created for few selected companies. After that these companies will be tested if there is Granger causality. In practice there will be 5 lags from VAR-model and Granger causality will test if they are jointly zero. If they are not there are predictive powers of SVI to trading volume.

The following table will list the assorted hypotheses and the most important relevant studies that have been conducted with those topics previously. With every hypothesis there are no known studies with the data selected for this particular study with Finnish companies. The previous studies have been conducted in a different environment with companies that are listed in either United States or United Kingdom markets. Many of the hypotheses are alterations from previous studies and the correct counterparts are referenced in the table.

Table 2 List of the hypotheses in this study as compared to previous studies

(H#)	Stating	Previous studies
1	Search volume index captures the general attention in the Finnish market.	No previous studies. Attention capturing by DEG (2011)
2	Search volume index correlates with the stock market trading volume but the trading volumes revert back to their mean values with some lag.	DEG (2011)
3	Search volume index can be used to predict and forecast and individual company's stock turnover or abnormal returns.	No previous studies. Mean level tests by DEG (2011)
4	Companies with high consumer public exposure are more affected by retail investor attention than companies that are not visible to consumers.	Wuoristo (2012)
5	The SVI results from Finland are better predictors for Finnish based company share price than the SVI results gathered from global queries.	No previous studies
6	According to Granger causality, SVI has forecasting power to trading volume with selected companies.	No previous studies

This study replicates the first three hypotheses of the DEG (2011) study, one hypothesis from the Wuoristo (2012) study and introduces two new that are geographically oriented. The hypothesis 2 and 5 actually contains two types of different regressions. The search volume from local sources and the search volume from global sources. Both are trying to explain the local trading volume within OMHX.

2.4.2 Comparison to previous studies

This study will replicate the testing of hypotheses originally by DEG papers from years 2009 and 2011. Even the data is significantly different from previous markets it should not cause problems with the testing of the hypotheses. Because of the structure of the indices (more specific the OMX compared to DJ/NASDAQ/FTSE) the data is assumed to be more reflecting the efficient market hypothesis in other indices. The OMXH consist of more from industrial originated companies compared to their US or European counterparts. This will allow the study to prevent general noise related issues. One example from the previous studies is from Wuoristo (2012) with the company Keller (Ticker symbol KLR). When searching for KLR in Google, the query results are mainly about a Kawasaki KLR motorbike. This obviously means the ticker and the company must be omitted from the research data. There does not seem to be such an issue with Nordic companies. The tickers and company names relate closely to the company itself with only few exceptions. Nokia phones queries result in phone models manufactured by Nokia.

The original study conducted by DEG (2009) does take note of the noise, but is moderately relaxed in the results. In general, the situation can be problematized in a following fashion: "A search engine user may search for a stock in Google using either its ticker or company name. Identifying search frequencies by company name may be problematic for three reasons. First, investors may search the company name for reasons unrelated to investing. For example, one may search "Best Buy" in order to do online shopping rather than to collect financial information about the firm" (DEG 2009, 6). The following aspects may be listed as problems:

- Search queries unrelated to investing (Google, Best Buy, Apple, Amazon etc.)
- Google Trends does not allow non-alphabetical terms (3M, 7-Eleven etc.)
- Different investors may search the same firm using several variations of its name. For example, American Airlines (AMR corp) with "AMR Corp", "AMR", "AA", or "American Airlines".
- Investors may use more random words for landing on the company's webpage, and after that navigate to the investor section. Thus not being shown on query results even if there would be demand for investor related information from the specific individual.

One related topic is not listed in the previous studies. A retail investor could search for a global ticker or search word while actually buying the stock from a global market place. The problem is more theoretical, and not presumed to be common. However, for example, Nokia is listed in few different market places. A retail investor in Finland or Sweden could search for "Nokia share" and still end up buying an American ADR from the US markets. Based on different markets part of stock price fluctuation could not be explained with SVI if the searches originate from different geographical locations. SVI locally would see an increase in value, while the demand for the share locally would see no change.

Another viable problem that might cause distortion in the results is the keywords of queries. For an investor curious about Nokia investment related information, he could Google search a Nokia phone model directly, land on the Nokia website and navigate to investor relations and investor information. This would not be shown in this study's SVI results at all, though there is underlying demand for investor information. This study considers this of a lesser problem because Nokia is omitted from the study.

3 METHODOLOGY, METHODS AND DATA

3.1 Regression methods and the testing of study hypotheses

The first hypothesis (H1) is tested as a relation between the trading volume and the acquired SVI values. The method of analysis is a simple ordinary least squares (OLS) regression where the trading volume (variable TV) is used as the depended variable and the SVI, stock returns (variable R) and the logarithm of market cap (variable MC) as independent variables.

$$TV_i = \beta_0 + \beta_{1i} SV_i + \beta_{2i} R_i + \beta_{3i} MC_i + \varepsilon_i, \tag{01}$$

The hypotheses H1, H2, H3, H4 and H5 are tested with regression. The hypothesis H6 is tested with Granger causality and Wald test. The OLS regression method is the similar method used originally in DEG (2011) and later in Wuoristo (2012). The method is fairly simple as it provides estimates on the marginal effect of the given explanatory variable. This is the reason given in DEG (2011) why it is a proper way to analyze and observe the predictive powers of ASVI-variable on different stock prices. The Fama-Macbeth method is commonly used for very large quantities of panel data that is gathered over time. The method also provides standard errors which in turn are corrected for cross-sectional correlation in the data.

The regression is conducted in two specific different stages. First stage is analyzing the effect of ASVI in every single asset separately. The results of first stage are that we have a table of data indicating how different variables influence the returns of individual companies. The second stage of regression is trying to understand the premium that is rewarded for each exposure. The second stage is a different kind of regression. The study uses different time horizons as previously in DEG (2011). The time horizons are the first four weeks, and the fifth horizon is the rest of that year (weeks 5-52 in specific). The variables for the second stage are: dependent variable is the future (t+1) abnormal returns. The abnormal returns are given in basis points in this study.

$$AR_{it} = \gamma_{ot} + \sum_{k=1}^{K} \gamma_{kt} X_{kit} + \varepsilon_{it} \qquad t = 1,2,\dots,T, \tag{02}$$

The variables are as follows for all the given formulas:

- AR is the abnormal return on stock I in the given week t,
- X's are the potential explanatory variables in cross-sectional expected returns,
- SVI is the search volume index,
- ASVI is the abnormal search volume index,
- TV is the trading volume,
- ABTV is the abnormal trading volume,
- MC is the market cap of a given company.

The same logic will follow every step of the study. For example, AR in one place also indicates AR in other formula. The letter A in front of a variable indicates the variable's abnormality. For example, SVI is Search Volume Index, while ASVI is abnormal Search Volume Index. This is consistent with every formula.

3.1.1 The use of multivariate models instead of structural

When studying time series jointly it is possible to find the correlation and the dynamic relation each of the series have. For example, two unique time series can be studied jointly with ordinary least squares methods or with VAR-models. In studies the individual time series in analysis are usually referred to as unique *components*. Each of the component has a correlation which the other components in the analysis as stated by Tsay (2005). In financial and economic studies the models used are inherited from long history of statistical inference. In many cases they have proved to be reliable source of information through proxies. In many cases there are no recognizable economic theories to back up the study results, but based on the statistical findings a new theory or hypotheses can be put together. Often in history the studies in finance have been built with structural equation models and in these cases it is common to examine the methods of intervention analysis and transfer function analysis. Both of the methods assume the time series to be independent, and the theory tests if independent variables have an effect on the time series. The models are often very cumbersome and indeed require a prior information or economic theories or equilibrium models to be usable. With any time series an independent researcher may find correlation, but it does not necessarily infer causation unless there is solid theory to back up the study results. Sims (1980) introduced vector autoregressive approach to the multivariate time series methods. This allows the researches to have less a prior information about the study at hand. The biggest shortcomings of the commonly used OLS models can be overcome with using multivariate models. As an example multivariate models allow the dependent variable to affect the values or the time path an independent variable. This is not common with OLS methods.

The multivariate (VAR) models can be represented in primitive, reduced or structural forms (Enders 2004). Starting from the simpler versions there is the standard structural form of the multivariate model. It stands as follows:

$$\begin{aligned} y_t &= b_{10} - b_{12}z_t + y_{11}y_{t-1} + y_{12}z_{t-1} + \epsilon_{yt} \\ z_t &= b_{20} - b_{21}y_t + y_{21}y_{t-1} + y_{22}z_{t-1} + \epsilon_{zt} \end{aligned} \tag{03}$$

The model is used in a simple way. There are some time shocks happening with z_t that have an impact to y_t in specific. If the variable b has any other value than zero then the error term have simultaneous effects ($\epsilon_{yt}\ and\ \epsilon_{zt}$). In this formula the error terms generally represent shocks that have an effect on y_t and z_t. According to Enders (2004) the next stage of the formula is creating the reduced form of the model based on the structural form. The structural form itself is commonly and often presented with matrices as follows (Enders 2004):

$$\begin{bmatrix} 1 & b_{12} \\ b_{21} & 1 \end{bmatrix} \begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} b_{10} \\ b_{20} \end{bmatrix} + \begin{bmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} \tag{04}$$

From this model the reduced form can be created. Since this study does not evolve around creating the model but only uses it, only the end result is there presented as it is.

$$\begin{aligned} y_t &= a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + \epsilon_{1t} \\ z_t &= a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + \epsilon_{2t} \end{aligned} \tag{05}$$

This model has practical uses in this study. The reduced form of the multivariate (VAR) model is a method to study the trading volume dependencies between several time series. This model also doesn't require the a priori information about the process. This is particularly useful when using another time series as proxies to represent. Nevertheless Enders (2004) states VAR models can often acquire too many parameters and thus reducing its possibilities of forecasting future values. This is not a problem though if there study's only goal is to understand the relation between SVI and trading volume. Although it is obvious the forecasting possibilities are greatly hindered because of the method then. In some cases the VAR model cannot be estimated easily. For these situations there is a method called Cholesky decomposition method in which the structural form can be achieved by applying certain restrictions to the estimated reduced form of the equation. This unfortunately leads to lost information of the time series as was the case with differentiation of the time series (Enders 2004).

The other problem caused by time series is if they are stationary or not. SVI is a normalized figure calculated by Google so therefore it is always stationary by nature (index from 0 to 100). This study uses logarithms of the trading volumes so thus the series become close to normal distributed. The literature by Sims (1980) mentions that when using differentiated time series information may be lost. For series that obviously are not stationary vector moving averages can be used (VMA). For Vector Moving Averages (VMA) the reduced form stands as follows:

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^i \begin{bmatrix} e_{1t-1} \\ e_{2t-1} \end{bmatrix} \tag{06}$$

In cases where both of the time series (y and z) are transformed into average values the VMA can be altered to the following general form:

$$\begin{bmatrix} y_t \\ z_t \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \bar{z} \end{bmatrix} + \sum_{i=0}^{\infty} \begin{bmatrix} \phi_{11}(i) & \phi_{12}(i) \\ \phi_{21}(i) & \phi_{22}(i) \end{bmatrix} \begin{bmatrix} \varepsilon_{yt-1} \\ \varepsilon_{zt-1} \end{bmatrix} \tag{07}$$

In which the multipliers $\phi_{11}(11, 21, 12, 22)$ are multipliers called the impulse response function. In this case the ε_{zt} presents the total cumulated effect to this date. According to Enders (2004) it is the sum that affects to the dependent variable y or z when time goes or changes from time zero to time i. This can be used to observe the strength of the impulse at various times and examine how strong the impulses are. Since this study uses mostly proxies to examine changes in trading volume there are no reasons to go to more exact models. The base values for SVI at best contain large uncertainty and thus creating more precise estimates of future values is not justified.

3.1.2 The application of VAR models with SVI and Trading Volume

VAR analysis can be a better method of studying time series than common OLS regression methods. However, VAR models have their negative aspects in the presumptions. For example, Enders (2004) brings up the case with Choleski decomposition. Often there are no economic theories or interpretations that are backing up the results. For example, there is no proven causal effect between SVI and trading volumes even there is a theory that suggest attention causes trading. It is not mathematically obvious the theory holds true, but this is the case often when using proxies (Enders 2004). At best it can be an accurate estimate. When using the different models often the analysis becomes mechanical and straight forward but the actual interpretation of the results is solely based on the individual researcher, and thus the reliability of the results are questionable. This study uses the reduced form of the models since there are no known causal relation between Search Volume Index and Trading Volume, other than the previous statistical inferences. There is no obvious causality between the two that can be verified prior to the statistical analysis. The attention theory presumes there is relation but the actual theory does not quantify the strength of the relation. Therefore, statistical methods must be used to acquire values. In any case there is no distinct causality between SVI and trading volume although previous studies have found correlation that holds true for different markets. According to Enders (2004) if the reduced model of the VAR models is built with lagged variables, and even if there is no lagged relation between SVI and TV, it still remains possible SVI and trading volume (TV) have an contemporaneous relationship.

In the past the studies have focused on OLS because the nature of the research field. There are very few studies in finance that study search volume index. Thus when studying an entirely new approach OLS provides reliable results. One can compare the differences between the structural equations and VAR models. For example, the structural equations commonly do require assumptions about the nature of the variables. For this study the hypothesis is formulated so that it claims there is a distinct relation between SVI and TV. Generally no such assumption is required with VAR models. They are free from assumptions. For estimation purposes the structural equations use OLS and maximum likelihood –methods. VAR models focus on testing of the lag lengths. When testing the hypothesis of studies, the structural equations tend to test the individual coefficients at some specific level, while VAR models test for either exogeneity or endogeneity of the variables. Both structural and VAR models allow forecasting possibilities, however there might be differences in the reliability of the values. The last lackluster problems for both models are as follows: structural models generally require assumptions about the nature of the variables and specific relevant theories to back up the results. VAR models lack the possibility of contemporaneous effect with the variables. In this study the contemporaneous effect means that the SVI and TV values are moving to the same direction at the same time. Although since this study uses both of the methods, the differences can be noticed and possibly examined further and because there are very few prior studies to the subject and even scarcer amount of studies with both methods, the results must be examined carefully. Unfortunately, since there is very little previous research data all the results in this study with VAR models are to be interpreted with care.

3.2 Data and descriptive statistics

This chapter contains some analysis and descriptions about retrieved stock market data and the SVI series. This specific study uses data from the Nordic markets. The index studied is the OMX Helsinki All Share Index (Ticker OMXH:HEX). The time series start from the beginning of 2013 and end in the last day of 2013. The method where the time series do start from the start of the year are commonly used by Shiller (2005). Similar method is also commonly used by Fama but to study event anomalies. The other option would be to begin the series from more randomized time. This would cause the possible anomalies to act differently. For example, if January is commonly a period of bull market (that is more positive returns than negative), with time series starting from midyear only one January anomaly period is taken into the data. Since the time series start from 1.1.2013 the researcher loses relevant information regarding the event that takes place annually. For example ,the January effect that means a bull market month in most years can actually start from the last month of the previous year. A yearly time series has the possibility of not taking such occasions into count. On the other hand, the next years January effect would be uncovered also, but would only experience the beginning of the next effect. The sample is over 251 trading days, but the regression is run with 250 trading day changes. This was a countermeasure if the sample would not have enough non-zero trading days during the period. In cases where the non-zero trading days exceeded 5 for the entire year, the company was entirely omitted from the study based on lack of liquidity.

The OMX Helsinki All Share Index represents estimated over 120 companies within the Finnish industry from healthcare to paper mills. The previous studies like DEG (2011) have chosen a similar type of All Share Index, mainly because it covers a vast amount of different company and industry types and the also data is more reliable with larger samples. Nevertheless, in previous studies the studied data has been from either the US or the UK markets solely. The composition of those markets is different both in cap size of the companies, but also in the industry distribution. The participation rate in markets for general public varies a lot from country to country, but unfortunately there is very little reliable data of it. The investing environment is also different. For example, the FTSE (UK markets) have very large number of companies representing many fields of industry as the following table depicts.

Table 3 Example: FTSE AllShare Composition (Wuoristo 2012, 24)

Industry	No of Companies	Weight in size (%)
Oil and Gas	26	16,99
Basic Materials	36	10,1
Industrials	109	9,06
Consumer Goods	35	13,7
Health Care	13	7,27
Consumer Services	85	9,47
Telecommunications	8	6,08
Utilities	7	3,99
Financials	257	21,88
Technology	25	1,47
Totals	**601**	**100**

As seen in the table for comparison to this study, the previous study's indices cover a very large number of different companies. There are several different types of industries represented. Financials and Oil & Gas do represent over 35% of the index, it still covers also consumer products, basic materials and telecommunications. For this study the indices chosen (OMX All Share) represent a smaller number of different industries. Not because the chosen index would only represent a small number of industries, but more or less because geographically Finland has large quantities of pulp and paper (forestry in general), basic industry and steel companies. While FTSE covers vast majority of oil and gas, the OMX has equally large share of industrial companies. Neither the industrial companies nor the Oil and Gas do not experience high consumer attention. Nor these types of industries have been part of the recent hypes and stock bubbles.

The OMX Helsinki All-Share Index includes all the shares listed on the Helsinki Stock Exchange at current time. The aim of the index is to reflect the current status and changes in the market. The HEX Index is therefore broken down using the ICB Classification as of February 1, 2012. The index was developed with a base level of 1,000 as of December 28, 1990. Outside of tech industry the index has remained fairly stable with the companies. The list of companies presented here is exactly the same list that is used to retrieve the search volume index data from Google. If a company's name is, for example, Apetit, then the exact search key for Google SVI is "Apetit". To make it clear and easy to verify the results, no additional words are added to queries. It is presumed if an individual person wants to acquire information about a company he searches with the company name. This is the presumption in the previous studies also except for those that directly accept only the ticker name as the search indicator. The reason for this approach with the company that’s name is quite obvious and it is indeed the most likely option and therefore is the closest proxy to attention. The problems with the method are when the same word is confused with an alternative meaning. For example, Nokia would suggest an individual Nokia phone model. Because of this the company is omitted from the study. Another example is Apetit which also produces fish servings, filet fish and such of the likes, but all of the queries originating from these subdivisions are omitted by default. The companies that are not omitted are listed in the appendix 10.

Each of the companies listed can be found from Google Finance, Yahoo Finance or from DataStream with the aforementioned ticker names or simply by searching for the company name. The company names are given with exactly the same letters they are listed in the index so there is no confusion or possibility of a confusion between two individual company names that would be close to each other. The individual time series used for these are acquired from the Finnish markets, therefore the HEX company ticker names. An individual company can be listed in many different markets, and therefore it is vital to indicate the correct ticker name. The prices for the companies are similar in different markets but there are differences how the prices change within countries and markets. In reality it means that there could be different correlations for SVI and trading volume with different market places.

Ticker symbols are listed mainly for volume, market cap, and price searches from DataStream. The same data can be acquired from other sources too. Some examples of this could be Yahoo Finance, Kauppalehti online or CNN Money. Many of the companies did not experience enough queries from local Finnish locations to qualify into the final sample. These are automatically omitted from further study.

The companies listed here are the ones with sufficient SVI data for further studies. The selecting of the time series data is straight forward with the three steps as stated here:

- Select companies with sufficient search volume index values for the company name. A series of zeroes is omitted because their absolutely values may be positive but after Google’ normalization insufficient to create regression time series. The company needs 245 trading days with positive search queries (SVI) for 2013.
- Select companies with daily changes in trading volume. For example, if the time-series is with static value it would be omitted entirely from the study. The volume has to be dynamic.
- Select companies with nearly everyday trading volume (max 5 days per year without trades). This is considered as sufficient liquidity for reliable price information.

If a company did not have enough SVI values it is automatically omitted without further studies or investigation why the lack of search volume. The SVI value time series must be constant throughout from 2013 to 2014. Secondly the companies that are selected have every day trading and are therefore considered liquid enough for the prices to reflect the current market situation and to price in any possible both external and internal risks to the business. The third step is to include additional companies that have had less or equal than 5 days without or with extremely low trading volumes throughout the 250 trading day time window. In some cases a company may have had only one or two trades for a day and can be considered a low liquidity small cap. In almost all cases the trading is fluent and total amount of trades is at minimum few hundred individual trades per day.

The results of data collection sum up to following figures. Companies that have both weekly global and weekly Finnish local data, add up to 51 companies out of 142. For monthly global and monthly weekly data, there are 13 individual companies. Rest of the companies have only either of the search origins in weekly form, therefore resulting in omitting from the study. In many cases the situation is there is enough or adequate search volume globally, but not locally. This is the first set of qualifications for the data. Both series have to be at minimum on weekly level to be able to make sufficient time series regressions. All the DEG studies have been done with weekly time series. There would be possibilities to run monthly regressions if one would start the time series earlier, since the beginning of available SVI data. But the previous studies have not been done in this manner and as this study tries to replicate them, one week is also chosen to be the time period.

3.3 Generating the time series

The previous studies used three different derived values from the data. They are listed here with the source of raw data:

- Abnormal stock returns (DataStream)
- Abnormal trading volumes (DataStream)
- Abnormal search volume (Google Trends)

All the values are abnormal in their nature. The abnormality is defined by a deviation from the expected values. It can be concluded that: abnormal stock returns are the returns which are derived from comparing realized returns to expected returns. Expected returns are calculated in a similar fashion as in previous studies. The method is to calculate from the benchmark index using the company specific beta. With this study the betas are calculated from daily data, then multiplied it with a company specific risk measure. This is almost exact replication to DEG (2011) and Wuoristo (2012) but with different data. This also can be used as verification the results are comparable and acquired the same way. For the beta two years of regression is used. In previous studies the beta is recalculated every two years to ensure validity. During the examination period the beta does not experience variation so therefore both methods as recalculating the beta or using a rolling beta do not affect the results. This 1 year beta approach is the chosen method in previous studies to take note of the changing market environment. There are some issues with it especially considering the time window. It is obvious the exact choosing of the time window can influence the results. Although the same method is used for every asset and the start date is the same. This will reduce the effect of the single events during the time-series.

Table 4 Variables used in the study

Variables	
AR_{it}	Abnormal returns.
R_{it}	Returns.
$E(R_{it})$	Expected returns.
R_{mt}	Returns from market porffolio
B_{it}	Beta
$_i$	Company i
$_t$	Week t

The reason for the weekly variable usage is because of the data limitations. One week is currently the shortest data period available in Google Trends, therefore the same variables are used for DataStream. Although DataStream is able to provide day to day data, which is downloadable to excel file. The weekly SVI values can be evened out to approximate daily values. For example, by using mean average values for the change. In some cases this is not absolutely correct but it is a common approach used with proxies. Also the changes in values are normal distributed and do not experience skewness or kurtosis with daily series. This provides more reliable statistical inference and removes many disadvantages that linear returns experience often. The method itself is well known and commonly used within finance studies.

This study consists of 54 selected companies that have the required SVI and trading data available. The omitted companies are entirely omitted because of search volume index related issues. The list for the companies selected for this study can be found from appendix 10.

Out of the 124 companies that are listed in the appendix 10 only 51 have the required prerequisites. Because of the high frequency data and the use of four different time series, abnormal returns, abnormal trading volume, search volume indices from global and local sources the total sample size for the study is 50128 individual observations. Less than 0.5 percent of the observations are zero. This is considerable larger sample size than in any of the previous studies conducted with search volume index and stock trading volumes.

3.3.1 Abnormal and expected returns

The exact formulas for abnormal returns and expected returns based on capital asset pricing model *CAPM* are as follows:

$$AR_{it} = R_{it} - E(R_{it}),$$

$$E(R_{it}) = R_{mt}\beta_{it}, \qquad (08)$$

The first variable is the abnormal returns while the second being the expected returns. The variables used in the equations are explained in the table. The same explanations apply for the entire study and they are named the same way as they have been in the previous studies to avoid confusion. Abnormal returns are calculated for daily values by means of iteration. The calculation is simple. From the daily realized returns the expected returns are calculated. The result is the abnormal positive or negative returns. The expected returns are calculated with CAPM using the calculated industry betas. This study uses one year beta as a static method. This means the beta is reliant on the market conditions for the year 2013. However, in some other studies by Fama (1998) for market returns rolling betas or five year betas are used. The choosing of the beta depends on the study at hand and all of the different alternatives have different results. Since the time window for SVI is one year it is natural to use the beta for the same time period, but there are very few factors that would prevent using any other possible beta. For realized returns the logarithm values are used as in the study by DEG (2009), the exact formula is taking the logarithm from (current value divided by previous value). The logarithm values can be summed as it is to create the return time series.

Tikkurila Oyj can be used as an example of the abnormal series. There is no specific reason to choose the company Tikkurila specifically, but it is used as a general example to indicate how the figures can be viewed. The changes are given from -1 to 1. Since the values are calculated from logarithm values they can be added as such. So two days of 0.02 returns together would equal a 4 percent increase to the original base value. Because of the logarithmic values the abnormal time series is stationary and does not have an increasing or declining trend. The graph for Tikkurila abnormal daily returns can be calculated when compared to the expected values of CAPM. The graph is stationary because of logarithm values with base values and therefore can be used for statistical methods.

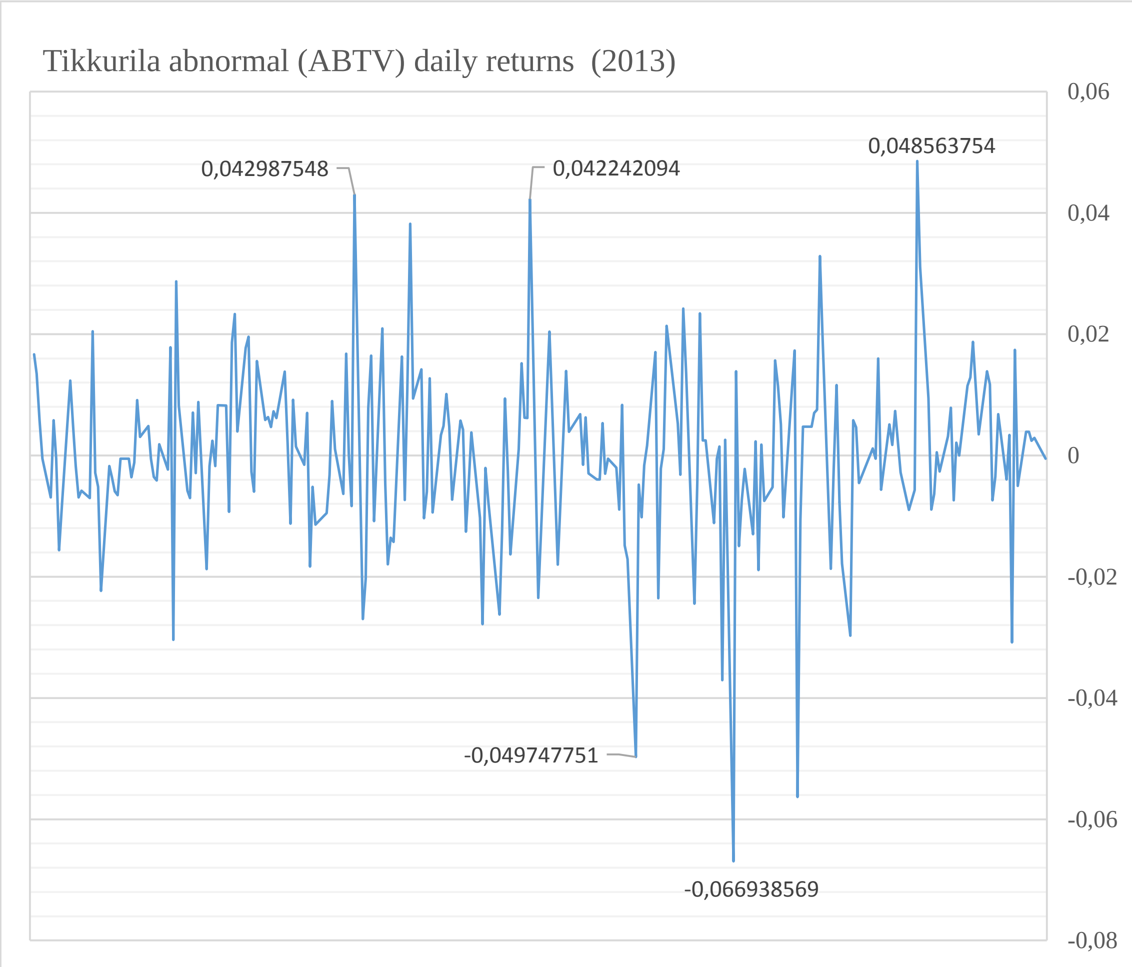

Figure 3 Tikkurila abnormal daily returns for year 2013

The figure 3 clearly indicates peaks in the abnormal returns. The series is stationary around 0 returns as presumed since logarithm is used to calculate returns. The result is to be expected and clearly indicates peaks around events. The frequency of the peaks varies and is 20 unique abnormal returns that exceed plus or minus 2% daily return. The abnormal values have to be verified for the chosen companies to make sure the time series is not just zeroes. In this case the regression with search volume index would be rather pointless. In all the cases all Finnish companies experience abnormal positive or negative returns during year 2013 when CAPM expected returns are used. There is another way of calculating abnormal returns also. This is the Fama-French three-factor model. The actual model is introduced in chapter 4.3 because it is used for robustness tests. The values from the Fama-French three-factor model do not differ from the values of the CAPM in this study. This is not the same for different indexes and for every different index the expected returns should be recalculated. The exact formula for the Fama-French model is as follows:

$$r = R_f + \beta(K - R_f) + bL * SMB + bS * HML + \alpha \tag{09}$$

The Fama-French gives a single value for each asset and it is not recalculated over time in this study. For year 2013 there is only one beta for each company according to the three-factor model.

3.3.2 Abnormal trading volume

The variables are calculated in different ways. Weekly abnormal volume is calculated from the specific changes in turnover of the share (variable T). This is calculated by relation of trading volume (variable TV) to outstanding shares for the company (variable OS). “This is used to normalize data since stocks have different amounts of shares and naturally a company with many shares should have more trade volume than a company with fewer shares (Wuoristo, 2012, 36).” Some companies in the data might have too low trading volumes to reliably generate regression, this can be the case especially with very small cap companies. The method for calculating abnormal in a similar fashion to DEG (2009) this study uses the same method. ASVI is calculated by comparing the current turnover of the stock to the median of the past eight weeks. This is done to find abnormal high or low turnovers and examine the peaks in the data.

$$T_{it} = \left(\frac{TV_{it}}{OS_{it}}\right),$$

$$AV_{it} = \log T_{it} - \log Med(T_{it} - 1, \ldots, T_{it} - 8)\_ , \tag{10}$$

The formula is very straightforward. Outstanding shares are updated from database every time there are changes in the free shares. This is the shares that are available for public or private trading. The shares owned by the company that cannot be bought or sold are not included in the calculations. Often these amounts are very small and do not affect the results in any significant way. The logarithm from the median for trades with 8 lags is used as method in DEG (2009) originally for SVI studies. As a general method it is used commonly to acquire homoscedastic error terms, or to prevent heteroskedastic error terms. As an example for the abnormal trading volume the company Tikkurila can be used again. See figure 1.

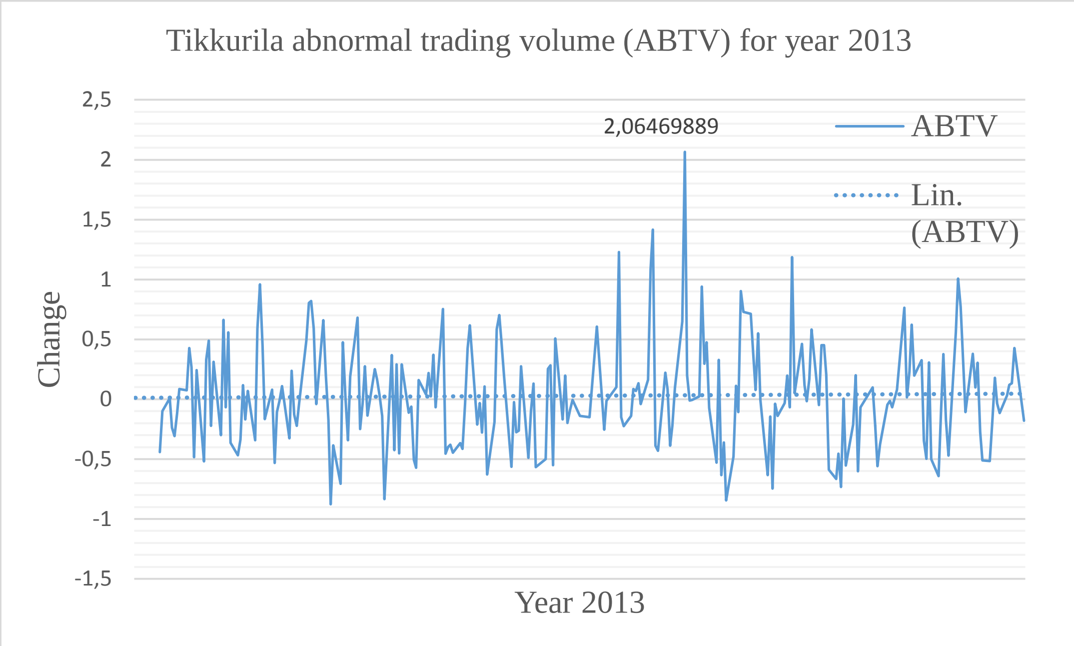

Figure 4 Abnormal trading volume for Tikkurila for year 2013

The common graph for many of the companies in the study is very similar to the one presented here. Because the values are also abnormal in nature and not absolute values but changes the distribution is close to normal distribution and the time series are stationary. There is a very minor linear trend within the ABTV but this is concluded to be insignificant. The similar trend can be recognized within most of the companies within OMX AllShare index. This is because the trading volumes generally have been rising from 2010 to 2013. Removing the trend is possible with statistical methods but does not change the results in any significant or even recognizable way. The positive trend with ABTV is less than 0.01 percent per year and is disregarded for further analysis. This is important for further regressions. There are obvious peaks in the data for abnormal trading volume for all the companies which is also to be expected. In many cases it implies there are market events to which investors react. The abnormal trading volume as such can be used to recognize the peaks that differ from the expected values defined by CAPM. The last of the calculated abnormal values is the abnormal search volume index.

3.3.3 Abnormal Search Volume index ASVI

The many variables are calculated in different ways. Weekly abnormal volume is calculated from the specific changes in turnover of the share (variable T). This is calculated by relation of trading volume (variable TV) to outstanding shares for the company (variable OS). “This is used to normalize data since stocks have different amounts of shares and naturally a company with many shares should have more trade volume than a company with fewer shares (Wuoristo 2012, 36).” Some companies in the data might have too low trading volumes to reliably generate regression, this can be the case especially with very small cap companies. The method for calculating abnormal values is a similar fashion to DEG (2009). This study uses the exact same method.

It is fairly vital to recognize the abnormal peaks in attention in regard to search volume index data. To accomplish this we will use the abnormal search volume variable called ASVI as in the previous studies. Visually observing a figure can also reveal if a series has significant amount of spikes. The abnormal value is calculated by using LOG of SVI during the current week minus the log median SVI during the past eight week period given. This method is exactly the same as used originally in DEG. The ASVI variable is very important for this study. The peaks are compared to the specific company’s share price and trading volume seen in markets. Using logarithms makes the time series stationary and therefore easier to use for statistical methods. Some information may be lost with the approach but it is commonly used method. The logarithmic time series are both stationary and do not experience kurtosis. The exact formula for this is as follows:

$$ASVI_{it} \ = \ log\, SVI_{it} \ - \ log\, Med(SVI_{it} - 1, \ldots, SVI_{it} - 8)_{-}\,, \tag{11}$$

In those specific situations where there is no available data for share price, trading volumes or search volume from Google the actual current week will be omitted from the data. If there are more than 5 of these weeks, the company is entirely omitted from the study. As such ASVI can be calculated to be originating from either global queries or from local to finish markets. The time series are naturally not identical but in many cases experience strong correlation and are contemporaneous. This practically means if an article of news is published, both global and local investors react to these news the same time. This is to be expected as most news are distributed through electronic sources such as television or internet web pages and therefore all investors receive the information the same time. Such behavior is common for financial markets. They are connected and the information is generally available to all investors almost the same time. Thus the investors react to the incoming news in a synchronous manner. For example, the Tikkurila stock abnormal search volume indices can be depicted in the following manner:

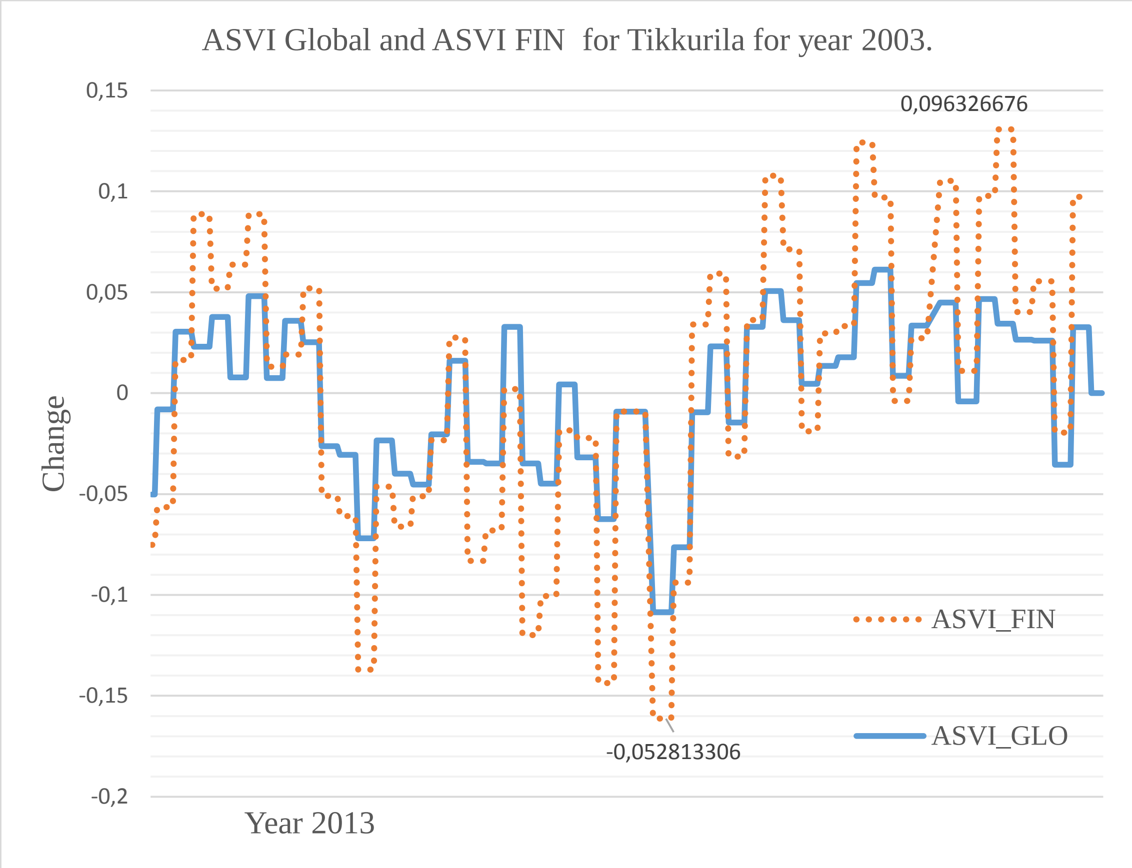

Figure 5 Abnormal Search Volume Index for Tikkurila for year 2013

There are systematic spikes in abnormal search volume index that originates from local Finnish sources. The overall peaks are happening simultaneously with both the global and the local searches. This is according to the attention theory based on that investors react to news items and search more information regarding the individual news topics. It also refers to herding or crowding effects as the risen levels of attention maintain in the markets for some time. It can be concluded there are distinct periods of optimism and distinct periods of pessimism. In both of the situations the search volume index levels remain abnormal. In pessimistic periods the levels are abnormal low and in more optimistic periods the values are abnormally high. What is prominent to local queries is the fact ASVI is stronger originating from local compared to global. The peaks are occurring simultaneously but FIN is significantly stronger. As the SVI values vary from 0 to 100 it can be concluded that the Finnish stock market is significantly more important for local people than for global investors. In all cases of Finnish stocks the ASVI values are close to zero change for the entire period. This can be interpreted in different ways. First, the attention the shares experience is rather low, the general public are not extremely interested in investing activities or searching information through Google. The second interpretation is that there are no interesting or ground breaking news during the period. The SVI the shares experience is solely reaction to the news the companies publish.

3.4 Data for OLS and multivariate analysis

This chapter contains the characteristics definition and sample size analysis. The OMX Helsinki All Share Index has currently 132 companies. These companies are all listed in 2014. The survivorship bias will be present in the index. To rule out survivorship bias the panel data should consist from data gathered over few decades. Because the fact SVI did not exist that far away in time the study will be affected by survivorship bias. This is unavoidable at this point of time and the previous studies considering SVI also experience the same phenomena. Nevertheless the OMX index has been fairly stable over the years, mainly because of lack of listings and also because the lack of corporate buyouts. There has not been ample delisting situations nor that many initial public offerings. This is a characteristic of Finnish markets. This has not been the situation with the previous studies such as Wuoristo (2012) or DEG (2011), where the UK and the US markets have had very significant changes. Wuoristo (2012) states the UK market has had 601 in FTSE All Share Index with 57 that had to be added to the initial sample. The DEG (2011) study has a far greater number of companies, and significantly more active capital market atmosphere in the US.

The previous studies have had four different unique processes in how to process and select the sample size. The basis for those processes is the replication of the previous studies as closely as possible to verify the comparability between the unique studies generally made during different time periods. As a difference to this study, Wuoristo (2012) and DEG (2011) have had their chosen indices (namely FTSE and US markets) have investment trusts as a part of the data originally. These investment trusts have been omitted from the data because of the belief, that investment trusts as such are not so likely to attract retail investor attention. The majority of retail investor attention is therefore presumed to be directed to equities in general. Investment trusts are traded in London Exchange. The initial string of thought with the Wuoristo (2012) study follows as such: "Investment trust funds are reallocated based on the fund manager's disposition, and are therefore fundamentally different types of securities." This current study does not include investment trusts in the data, nor are they part of the OMX All Share Index.

The second chosen process is how noise is being dealt within the data. In any case, no matter how carefully the data is chosen there will be always some amount of noise which may possibly distort the results. Noise generally is caused by company names (or tickers in Wuoristo (2012) study) that have different alternative meanings. As an example from the UK markets within FTSE index, there is a companies called American Tobacco and

Glencore. Their aforementioned company tickers are named BATS and GLEN. Therefore Wuoristo (2012) concludes these companies cannot be used as such for proxies of investor attention. They are more likely to capture something else than the retail investor attention for the equities.

This study replicates the steps of the Wuoristo and (2012) and DEG (2011) in regards to data cleaning. In reality this means clearing the obviously noisy tickers out of the data. For OMX there are few borderline cases. Second step of the process is checking the SVI data. For SVI data Google Trends informs the end-user with two separate values for a single search. First is the "Top searches" which is the most common words associated with the word. The scaling for the top searches is from 0…100. In this case the highest given value does not mean it is associated always with the given word, but that it is relatively the most common word to be associated with. Unfortunately and deliberately Google provides no absolute values for the searches. The second provided value is the "Rising searches" which tries to show the different unique terms that are experiencing a rise in popularity during the given time period in comparison with the preceding period. There is also a case of "breakout" which means a single term has experienced over or equal to 5000% growth during the chosen time period.

In an approach where each individual company data is manually reviewed, it is possible to avoid many of the noisiest company names. Compared to the situation where such an approach would not be taken, the quality of data is significantly better. As an example of noisy ticker from UK markets Wuoristo (2012, 30) introduces the company named Keller. The ticker for Keller is KLR. The aforementioned ticker seems to experience seasonality and significant peaks. Such anomalies are usually considered to be indications of noise. The reason for noise in this example case is: KLR is also a model for the Kawasaki motorcycles. Because it's obvious a lot of the searches are actually for the motorcycle instead of the equity, the ticker has to be omitted from the data. Therefore the previous studies take the approach that if a company name or ticker has a clear and undeniable alternative that is the probable cause for most of the queries, it should always be omitted from the study. If this on the other hand is not the situation, the results are accepted as is. The most optimal situation naturally is, where the search word would be associated with words such as info, news, share, price, share price, investor and with other related words. In these cases it is almost obvious the query is related to the individual trying to acquire information about the company. Nevertheless Wuoristo (2012, 30) concludes that: "It must be noted that there will be a certain amount of noise in a search volume based study, but by monitoring the results the risk can be minimized." This is the situation with the OMX data also. It is possible the data contains significant amount of noise in many cases, but unfortunately with company name related searches it is impossible to avoid, and as the mere ticker name searches do not provide enough SVI data, other approaches are not possible.

The Wuoristo (2012, 30) example for the Keller company data shows both the Top searches and Rising searches paragraphs very clearly. The ticker (and the company name for that matter) are with high noise that cannot be easily cleared from the data. This results in the company name and ticker to be omitted from study as such. This example is one of the most clear in any data. There are several non-related rising searches in or above the breakout limit (5000%) which indicates without doubt the searches are not related to the equity but to the motorcycle. Similar cases as this always cause the company at hand to be omitted from the study. The clearest example of a company resulting in omitting is Nokia. Nokia search queries consist almost solely of mobile hand set related search keywords. With closer observation there are very few if any investing related topics in the top hundred query keywords. This can be avoided by choosing the queries to be originating from investing related pages only such as Google Finance, Yahoo Finance or CNN Money for example. After these modifications the query words are majorly investment activity related such as queries for the ticker, a graph of the stock's returns or queries related to the upcoming interim reports. This procedure is not used in this study at all, simply because to make the SVI values commonly generalized this should be done to every single company. Investing related queries as relative search volume index terms coming from direct Google queries are significantly lower portion of the total amount of searches, than the searches originating from purely financing related services. For example, if an individual finds Nokia's personal page from Yahoo finance and compares to Google search volume index values, almost 100 percent of the queries originating from Yahoo are investing related. This time-series would not be equivalent to those originating from common Google queries. Therefore, any company with the majority of the search queries being entirely unrelated to investing related activities must be omitted from the study. In previous studies this has resulted in over 80% of the companies to be omitted from the study. With OMXH the case is not so severe since 52 out of 124 indeed get accepted into the final sample. This relative value difference does not originate from the used methods but more because both FTSE and the US markets consist of more consumer oriented companies that experience direct product search queries by consumers.

Table 5 An example of a noisy ticker search for the Keller resulting in omitting from study (Wuoristo 2012, 30)

Top searches for klr	
klr 650	100
kawasaki klr	90
kawasaki klr 650	50
klr 250	45
klr 600	25
klr for sale	20
kawasaki klr 250	20
kawasaki klr 600	10
Rising searches for klr	
kawasaki klr 250	Breakout
kawasaki klr 600	Breakout
klr 250	Breakout
klr 600	Breakout
klr for sale	Breakout
klr 650	180 %
kawasaki klr 650	40 %

It is very likely – although not clearly stated - none of the given searches in the SVI are related to the company the original study wanted to study. Unfortunately in many cases the situation is not as clear as this, but there might be underlying noise that is not evident even when looking at the top searches and rising searches paragraphs. Still in all the cases the approach where each of the SVI queries is handpicked and studied individually and verified for noise, the SVI data will contain less noise. There are cases where the query does not necessarily mean investor attention at the exact moment the query is done through Google, but general attention towards a company. This can transform into investor attention with or without lag in line with the theory that investors tend to invest in companies they are familiar with. The underlying presumption is if an individual is searching for information, he or she wants to know something about the query word that is typed into Google search.

The third step to data analysis in this study is omitting the companies which experience the lack of data. There was previously two different dominant reasons for this. Firstly, if the company is listed in very near history, there is lack of price history for the share. In some cases the company may be result of divestment or fusion. In these cases the company is omitted. The second reason for omitting is the lack of sufficient SVI data. This can be caused by two different things. Firstly, if the company is very young, and the situation is exactly the same as with very little price history. Secondly, if the company has not experienced investor attention at all, or only by such amount that Google Trends is not able to generate SVI data for it. Both, the lack of price history or lack of SVI data, result in omitting the company.

The fourth and also same time the final step into sample gathering process is to verify that the time period (frequency) is chosen on one week level. This is dictated because the previous studies (DEG 2011) and (Wuoristo 2012, 31) have made it in a similar fashion. Both studies found price pressure and price reversal being strong in one or two week periods.

Wuoristo (2012, 32) study had consumer-visible companies to be popular and common in the sample. The consumer-oriented companies tend in both the US and UK markets to generally get more retail investor attention which is what the theory also predicts and indicates. With previous studies the global SVI has carried more random noise than local SVI. This “makes singular attention peaks more difficult to attain (Wuoristo 2012, 33).”

After the careful selection process, only a handful of companies remain in the final data sample. The list that is in figure 10 has the companies that meet the previously given criteria for Search Volume Index. The raw data must be acceptable on SVI and on DataStream, meaning it has to be reliable from google and it has to have enough active trading days to be accepted into further study. For example, Valmet corporation is recently listed into OMXH (it is split from Metso corporation), and therefore lacks the daily trading volume to be reliable source of analysis data. In previous studies as DEG (2011) and Wuoristo (2012) they have not disqualified companies that have excessive amount of zero trading days. This study differs from those. The main reason for omitting the companies in this is the following: if companies with half a year of trading days would be accepted, the SVI and trading values could be measuring any random phenomena. One of these could be the initial changes that occur because of IPO. This study is not capable of measuring such events, and therefore accepting them into the data would be questionable. Considering all the requirements, only 42 companies are accepted into the final sample. The most common reason for omitting is the lack of SVI or the reliability of the SVI is not high enough. The second reason for omitting is the lack of active trading days. In practice meaning there are no trades at all for several days in the time series. In few cases there are stocks which have 2-5 days per year where there is no trade. These shares have been kept in the final sample.

3.5 Granger causality

One of the aim of the study is to study the possible predictive capabilities of search volume index towards a company’s stock’s trading volume. A probable outcome is SVI cannot explain the trading volume or vice versa. If either was able to predict the other statistically significantly and consistently the efficient markets hypothesis should be rejected. Based on the attention the share's experience there should be a relation between SVI and TV. The theory explains there should be predictive powers that are statistically significant and therefore can be replicated for profits. If either have predictive powers over the other is not automatically transformed to causality. If SVI predicts trading volume the natural outcome is not that SVI is causing the changes in the trading volume. There is a testing method called Granger causality that helps the researches understand the relation. When using more complex VAR-models more variables can be taken into account when studying the causality. The method to test the causality is based on the hypothesis of Granger. Enders (2004) suggest the test is run in a way that lags of one variable are entering into the equation of some other variable. For Granger causality this means that if all the coefficients in a VAR-model equal to zero there is no Granger causality. If the coefficients do not equal to zero there is Granger causality. This can be explain also as a formula for typical F-testing.

$$\begin{aligned} &If\ a_{12}\ \neq 0\ Search\ Volume\ Index\ Granger\ causes\ Trading\ Volume \\ &If\ a_{21}\ \neq 0\ Trading\ Volume\ Granger\ causes\ Search\ Volume\ Index \end{aligned} \tag{12}$$

If the coefficients remain zero the variables do not contribute to the forecasting powers of the other variables. The same test has to be run for all the valid variables. Commonly the test does not conclude contemporaneous effect at all, instead just refers to the past values of the variables. By testing for contemporaneous effect the test would also gain the ability to study further if the variables are endogenous or exogenous. The tests begin with the simple equation. To able to test if search volume index Granger causes the trading volume the test is run in a following way:

$$\begin{aligned} TV_t = a_{10} + a_{11}\text{Lenght}(SVI\_FIN_t) + a_{12}\text{Lenght}(SVI\_GLO_{\text{t}}) \\ + a_{13}Lengh(ABTV_t) + \varepsilon_{1t} \end{aligned} \tag{13}$$

The trading volume is explained with SVI_FIN and SVI_GLO and also by the abnormal trading volume itself. The length before the variables means the specific lag length for the actual variable used. The different VAR-models are identified differently and different number of lag lengths is used based on the model. The VAR-Model used here is heteroscedastic and therefore either White's robust method or Newey's and West's (1987) standard calculation for standard errors can be used. The interpretation for the results can be seen as if the impact on the values is not reversed within the calculation period the search volume index contains relevant and statistically significant information about trading volumes. The previous information therefore is not calculated in the share prices and thus trading volumes should change. The test is run only to the selected few companies that have statistically significant variables with OLS. OLS methods alone would not grant predictability simply because the error terms are heteroskedastic. The Granger causality can be calculated for each of the companies in the study separately but it also could be calculated for the mean values for the entire index. The mean value calculation is not performed in this study simply for the fact that only a handful of companies had any statistically significant variables with OLS regression. In previous studies such as DEG (2011) or Wuoristo (2012) there were mean level linear dependencies. This indicates efficient market hypothesis is valid and accepted with the Finnish indices although the requirements for the strong market efficiency are not fulfilled. The efficient market hypothesis applies to over 90% of the companies in the index which is a greater percent than compared to UK or US markets in the previous studies by DEG (2009) and Wuoristo (2012). Both of the previous studies had found linear dependencies even those dependencies diminished one or two weeks after they had been initially found in the market time-series. The remaining few companies in this study that have statistical significant variables with low p-values could be used for arbitrage or profiting at the time, but because of the weak nature of the dependencies the deviances from the efficient markets is minor at best. According to EMH the market should profit out the arbitraging positions or they should be very short lived to begin with.

The following companies had statistically significant variables. The following list is for ASVI_GLO significant at 95% level: Atria, Fortum, Keskisuomalainen and Kone Oyj. For ASVI_FIN the only options Kemira and Kone Oyj. The Granger causality testing is performed with a method as simple as possible. The estimation is performed as it was with ordinary least squares or VAR-models. The dependent variable is ABTV or Abnormal returns with lags as variables. The coefficient testing is performed with Wald test with lags. The idea is $L\text{enght}_1 = L\text{enght}_2 = L\text{enght}_3 = L\text{enght}_4 = L\text{enght}_5 = 0$. The testing is therefore quite simple. The error terms are HAC for heteroscedastic qualities but, for example, White’s robust heteroscedastic adjusted errors are also viable. The results are very similar.

4 EMPIRICAL RESULTS

This chapter is divided into two subchapters. The first subchapter collects the regression results and testing values from ordinary least square methods to multivariate and the final part is Granger causality. The second subchapter provides information to Fama-French robustness testing. In the robustness checks the regressions are rerun by using Fama-French three-factor model betas instead of the ones from capital asset pricing model.

4.1 Regressions and testing

4.1.1 Ordinary least squares

The complete list of the companies in the study can be found from the appendix 1 while the companies chosen for further analysis can be found from appendix 10. The main focus of this study is in the companies that have statistically significant variables with 95% confidence level. The aim of the study was to find out the relation between search volume index and trading volume of a stock. The results are presented in this chapter with a list of companies that have statistically significant companies while omitting all those companies that have no statistical significance. Total 17 companies from the sample have variables that are statistically significant with the given regression model. In few cases the variable C is of statistical significance but there is no interpretation for it. According to Enders (2004) if the variable C has statistical significance but theory suggest no sensible interpretation for it the factor can be reported but disregarded from further studies. Surprisingly only 4 companies in the OMXH had linear dependency between abnormal trading volume and search volume index that originates either from local Finnish sources or from global queries. In all cases the R^2 values are below 2%, so even with the few significant variables the model only explains a fraction of the fluctuation. This result is not in line with any of the previous studies conducted with SVI and trading volumes. Both the DEG (2009) and Wuoristo (2012) found SVI to have predictive powers at least for short term and also linear dependency between local SVI and abnormal trading volume. As a priori information both of the UK and US markets should be presumed to be closer to strong-form market efficiency, while the OMXH is found to be weak-form efficient at best in most studies. However, this study does not directly test the efficient market hypothesis. If there was to be a clear and decisive statistical significance between SVI and trading volume or prices consistently in the data, then the efficient market hypothesis could be rejected or noted that the markets are not efficient at all. The appendix 10 holds the list for all the companies in the study. The summary of the ordinary least square regressions for the selected companies in the final sample can be seen from the following table.

Table 6 Summary of OLS regressions

Summary of OLS regressions					
Sample size 42*252*4=42336 individual observations					
Dependent variable: ABTV (abnormal Trading Volume)					
regr. ABTV = B0 + B_1*ASVI_FIN+B_2*ASVI_GLO+B_3*ABRETURNS+e					
Sample period year 2013. Companies with variables at 95 % significance level chosen					
Company	**Variable**	**Coefficient**	**Std. Error**	**t-Stat.**	**Prob.**
aktia	C	1.798185	0.281252	6.393501	0.0000
atria	ASVI_FIN	0.671978	0.241391	2.783780	0.0058
basware	ABRETURNS	5.085843	1.720657	2.955755	0.0034
digia	ABRETURNS	6.777225	3.345574	2.025729	0.0439
f-secure	ABRETURNS	3.389221	1.112710	3.045916	0.0026
finnlines	C	0.843695	0.144461	5.840300	0.0000
fortum	ASVI_GLO	0.783738	0.286140	2.739001	0.0066
honkarakenne	C	0.811511	0.113780	7.132257	0.0000
kemira	ASVI_FIN	0.450251	0.160747	2.800984	0.0055
kemira	ABRETURNS	-3.437501	1.096988	-3.133581	0.0019
keskisuom.	C	0.239236	0.100293	2.385357	0.0179
keskisuom.	ASVI_GLO	-7.293762	3.444725	-2.117371	0.0353
kone	ASVI_FIN	0.870357	0.277238	3.139387	0.0019
kone	ASVI_GLO	-2.019694	0.531953	-3.796753	0.0002
kone	ABRETURNS	-3.153284	0.758561	-4.156927	0.0000
martela	C	0.354358	0.083690	4.234187	0.0000
nurminen	C	1.049403	0.197308	5.318602	0.0000
outokumpu	ABRETURNS	-0.625354	0.302348	-2.068326	0.0397
outotec	ABRETURNS	-2.677220	0.654307	-4.091688	0.0001

ponsse	ABRETURNS	5.900366	1.887932	3.125306	0.0020
rautaruukki	ABRETURNS	3.830386	0.460842	8.311704	0.0000
SRV	ABRETURNS	5.236674	1.737124	3.014565	0.0029
stockmann	ABRETURNS	4.769340	1.137114	4.194248	0.0000
viking line	C	0.650049	0.118930	5.465797	0.0000

There are 24 statistically significant variables in this study. The total number of variables is 496 for the regressions. The statistical significant variables therefore represent 4.84 percent of all the variables. This can be considered as a minor value considering the attention theory predicts retail investors to be involved in investing related activities based on how they show interest in finding information about companies. 11 of the significant variables are abnormal returns that are positively correlated with volume. This is the basis for momentum investing strategies. In reality this means the higher the stock price goes the higher the trading volume it experiences in those cases where the correlation is positive. This also an evidence of flock-type of behavior or herding. In most cases the abnormal return variable is one of the highest values in the regression. It can be concluded that abnormal returns are better predictors of future abnormal trading volume than search volume indices. Retail investors buy stocks that go up in price and sell stocks that are going down. Based on this notion they are acting with the market changes and even enforcing them. Seven variables are the constant C. This also indicates there is abnormal trading involved compared to the expected return models. This study does not go into great detail about momentum investing, but the behavior can be found and verified from the results stated above.

The exact R^2 values are given in the appendices 2 to 7 for OLS regressions. In almost all cases the values are below 0.05 which means the chosen variables explain only less than five percent of the total fluctuation in trading volume. For further studies the companies Atria, Fortum, Kemira, Keskisuomalainen and Kone will be used for multivariate models because they have linear regression at 95 percent significance level with SVI variables either from local or from global sources. The implication is to use multivariate models with both autoregressive and moving average variables to find a model that explains the fluctuation best for these five selected companies.

4.1.2 Multivariate and ARMA-GARCH

The multivariate and ARMA-GARCH models produce different results for the selected companies that had statistically relevant variables. The model chosen is

ARMA(2,2). This means two lags of autoregressive and two lags of moving averaging. This decision is based on the previous studies how investors react to changes and also because of the residuals in the model. In practice the variables in the regression are RESID-1, RESID-2 and GARCH-1 and GARCH-2 factors. The model chosen is verified on residuals from the analysis. In some cases it proves that only one lag would be needed but the model is still ran with 2 lags. For model verification purposes the different lag lengths were observed. If there is statistically significant autocorrelation with the residuals more lags are added. The best model was iterated this way to be the ARMA(2,2). With this model the factors are significant at 99 percent level at times and in all cases the p-values are lower than they are with OLS tests. In some cases the first lag is insignificant but the second is significant. These results reflect the previous studies. Both DEG (2009) and Wuoristo (2012) found the 2 weeks lag also the key part of the results at very high significance levels (from 99 to 99.9 percent significance). It is also noted that ARMA models are not designed to model the contemporaneous effect of two variables moving in the same direction.

The companies will be reviewed one at a time, starting from Atria. ARMA(2,2) was used for studying the company for the period year 2013. The amount of trading day changes equals the amount of possible trading in year 2013 through direct market operations. Search volume index originating from global sources is statistically significant at 99 percent level. The regression result can be seen from the table below.

Table 7 Atria multivariate regression results

Dependent Variable: TV	Atria			
Method: ML - ARCH (Marquardt) - Normal distribution				
Convergence after 32 iterations				
Presample variance: backcast (parameter = 0.7)				
Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.018865	0.018092	-1.042738	0.2971
SVI_FIN	0.696616	0.244146	2.853279	0.0043
SVI_GLO	-2.140435	0.515857	-4.149283	0.0000
AR(2)	-0.196540	0.001152	-170.5638	0.0000
MA(2)	0.354740	0.046813	7.577804	0.0000
Variance Equation				
C	0.002223	0.000813	2.733019	0.0063
RESID(-1)^2	0.175081	0.083293	2.101988	0.0356
RESID(-2)^2	-0.225736	0.084519	-2.670840	0.0076
GARCH(-1)	0.966116	0.157304	6.141721	0.0000
GARCH(-2)	0.062700	0.158853	0.394701	0.6931
R-squared	0.021359	Mean dependent var		-0.003071
Adjusted R-squared	0.004486	S.D. dependent var	0.310332	

Using 95 percent significance the GARCH term with two lags is not significant. The results for the company are equivalent to the results with OLS in the sense that at least one of the variables is statistically significant. With OLS method only the global search volume index was significant stating the queries from outside of Finland are linear dependent with trading volume. For ARMA-GARCH both sources of queries are relevant with global queries being significant at 99.9 percent level. Findings for this company are similar that were the findings for DEG (2009) studies with company mean values. Abnormal volumes correlate with more abnormal volumes. This is the basis for momentum strategy. To make sure there are no relevant and significant lags left out from the model the residuals must be observed and examined. The q-statistics are as seen from the following table.

Table 8 Q-statistic probabilities for ARMA terms

Q-statistic probabilities adjusted for 2 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.115	0.115	3.2009	
		2	-0.034	-0.048	3.4748	
		3	-0.007	0.003	3.4852	0.062
		4	0.023	0.022	3.6167	0.164
		5	-0.073	-0.079	4.9001	0.179

The lags beyond the first two are not statistically significant. The third lag would be significant at 90 percent level but not at 95 percent. Thus it will be omitted from the model. The same two lag length is valid for this company as it was in the previous studies at weekly level. Thus it can be concluded that the relevant changes between search volume index and trading volume happen one to two weeks after the initial change. This driver for change can be either exo- or endogenous. Such events can be market rumors, news or interim reports or profit warnings.

For the normal distribution testing there are different ways. In this case the Jarque-bera is also checked. The Jarque-Bera measures the goodness-of-fit. It tests if the sample data has excess skewness and kurtosis. If the sample is close to normal distribution the Jarque-Bera is close to 0. The distribution of the residuals is as follows:

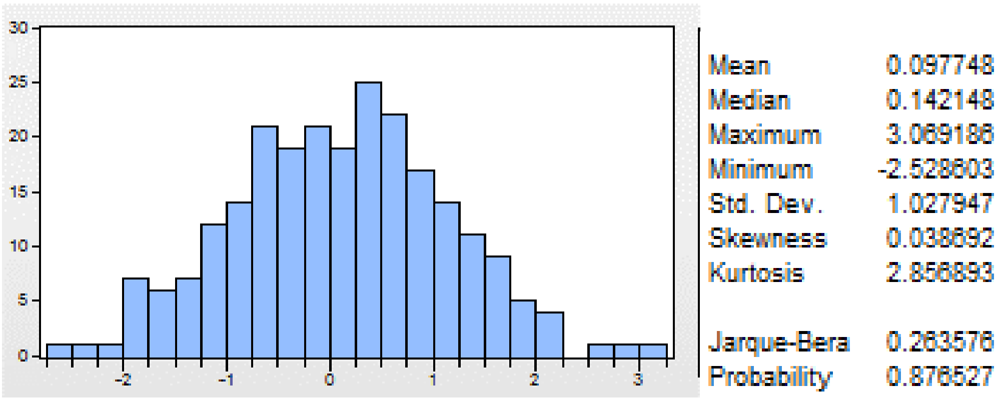

Figure 6 Normal distribution testing, skewness and kurtosis

The testing for normal distribution, skewness and kurtosis can be seen from the figure. The distributions have minor skewness and the default value for kurtosis is presumed to be 3. This is typical for return distributions since and it is very common. There are very few outliers and the distribution has only one single peak. The models can be accepted based on the fact the residuals are very close to normal distributed, have only one single peak, and kurtosis value is close to 3. There are tails but they are not considered to be problematic in this study.

The second company for the multivariate analysis is Fortum. Only the factors are noted here, the entire regression with the given lags can be found from appendix. The OLS results for Fortum were SVI_GLO being the only statistically significant variable with the factor of 0.78. So if there is a 1 unit change with global SVI value for Fortum the abnormal trading volume rises 0.78. The change is positive and there is no reversing of the effect according to OLS. For multivariate the results are as follows:

Table 9 Fortum ARMA

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.028377	0.008284	-3.425567	0.0006
SVI_FIN	-0.273801	0.154363	-1.773749	0.0761
SVI_GLO	0.829859	0.223463	3.713623	0.0002

The ARMA(2,2) finds SVI_GLO also to be the relevant variable, thus indicating the non-professional investor attention is more meaningful originating from global sources outside Finland. The entire regression can be found from appendix 4.The SVI_FIN is also significant at 90 percent level but not at 95 percent. Interestingly the SVI_FIN value is negative, but the absolute value is very minor. If Finnish investors use Google trends to find information about Fortum it causes abnormal trading volumes to sink. However, the factor is very minor and therefore the result should be questioned or disregarded entirely. The R^2 for Fortum's ARMA model is 0.0117 and R^2 adjusted only 0.0033 which can directly be translated the search volume index changes have extremely little effect on the trading volumes. The efficient market hypothesis is valid for the company, there are no weak-form efficiency deficiencies that could be used for any arbitraging positions and the markets experience strong efficiency in this regard.

The third company is Kemira. The OLS regression was that SVI_FIN factor value was 0.45 and significant at 99 percent level. For ARMA model the results are as follows:

Table 10 Kemira ARMA

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.014482	0.017756	-0.815615	0.4147
SVI_FIN	0.340225	0.165106	2.060638	0.0393
SVI_GLO	0.144046	0.303848	0.474073	0.6354

The results are similar in the table. The entire regression can be found from appendix 5. ASVI_FIN stands at 0.34 with 95 percent significance level. So if queries originating from Finland increase 1 the abnormal trading volume gains 0.34. The global queries are clearly irrelevant for trading volume. So one unit change in Finnish SVI creates a 0.34 change in abnormal trading volume. The more individual investors seek information the more demand for the share it creates. The entire regression can be found from appendix with the lags also. The R^2 stands at 0.048 so the model explains roughly 4.8 percent of the total fluctuation in trading volume. The lags prove to be statistically insignificant.

Fourth company is Keskisuomalainen with only global queries as the relevant variable for explaining abnormal trading volumes in this model. The OLS value was -0.729. This can be interpreted as global queries being very important for the trading volume. As a company that operates and is owned mainly by Finnish investors the results are surprising. The following table shows the values for the ARMA-GARCH.

Table 11 Keskisuomalainen ARMA

Variable	**Coefficient**	**Std. Error**	**z-Statistic**	**Prob.**
C	0.169215	0.068568	2.467843	0.0136
SVI_FIN	5.265797	3.405495	1.546265	0.1220
SVI_GLO	-8.105870	2.959078	-2.739323	0.0062

The results can be found in the appendix 6 also. The entire regression can be found from appendix 6 which also show the lags prove to be insignificant for this company. For efficient market hypothesis the Keskisuomalainen share does not indicate any arbitrary positions based on Google queries. Either the stock does not experience enough investor's attention or the purchasing of the shares is not influenced with attention.

The fifth and last company is Kone Oyj. The entire regression can also be found in the appendix. Kone was the only company in OLS studies to have both the Finnish and global queries significant at 95 percent level. This result is also surprising because Kone stock is one of the most liquid stocks in the local markets. It experiences also very high investor attention according to Google trends. Abnormal search volume index from Finland was 0.87 and abnormal search volume index from global sources stands at -2.01. The variables having different directions is conflicting information. Reason for this can be that global investors use other sources of information more than Google, but this study cannot reliable answer the question. The ARMA-GARCH results are as the following table indicates. The entire regression with the variables and lags can be found from appendix 7 for Kone Oyj.

Table 12 Kone ARMA

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.011535	0.011854	0.973086	0.3305
SVI_FIN	0.808568	0.281566	2.871685	0.0041
SVI_GLO	-2.070548	0.542033	-3.819968	0.0001

If global queries increase by change of 1 unit then the abnormal trading volume goes down by approximately 2 units. This is in direct conflict with attention based behavior which states increased attention causes more purchases than sells. Increased attention results in less trading according to this regression. Instead for Finnish sources the results are the opposite. Finnish non-professional investors that are searching for information about the company end up with higher trading volumes. This would indicate flock or herding behavior and a behavioral bias very similar to what has been found in the previous studies. The R^2 stands at 0.068 so the model explains 6.8% of the abnormal trading volume fluctuation. Approximately 93% of the fluctuation is explained with other variables that are not included in the current regression model. Kone oyj has the strongest linear dependency between SVI and trading volume in this study and the absolute amounts state that SVI does not explain any meaningful amount of the trading volume, precisely less than 90 percent. Other reasons and variables are behind excess trading than information demand of non-professional investors. This finding is similar to the findings of previous studies. Compared to the DEG (2009) this studies R^2 values are lower so for the chosen companies the regression model explains less while other factors that are left outside the regression contribute a more significate proportion of the total fluctuations.

After the multivariate models the results can be examined further if there are possibilities of finding causality. For this specific use there is the Granger causality. Granger causality is not causality per se but studies the lags.

4.2 Granger causality

Granger causality can be used to assess the possibilities if another variable causes the depended variable as explained in chapter 3.5. To understand the behavior of Finnish non-professional investors this test can be used to see if SVI causes the ABTV. For this purpose only three companies are chosen because of their positive OLS variables. The companies are Atria, Kemira and Kone. Each of them had ASVI_FIN significant at 95 percent level. The test is conducted as in 3.7 but the values for statistically non-significant variables is set to 0. If there is Granger causality with abnormal search volume index then the abnormal returns are also tested. The abnormal returns testing could be performed for all companies that have positive results. Mechanically the testing can be done for global SVI values that are negative too but the interpretation is problematic since it is against the theory behind the expectations. The theory claimed that the more people have attention with the assets the more eager they are to acquiring them, while the reasons for selling are usually not related to individual companies but to macroeconomic signals. In both cases the increased attention to search volumes is presumed to be positively correlated with trading volume – the higher the query amounts the higher the trading volume. The testing will be done only for few selected companies. The following list is for SVI_GLO significant at 95% level: Atria, Fortum, Keskisuomalainen and Kone Oyj. For SVI_FIN the only options remaining are Kemira and Kone Oyj. By adjusting the significance level of the VAR and OLS models from 95 percent to 90 percent or less would increase the amount of acceptable companies but in many studies the 95 percent level is the most commonly used level.

The results for the SVI_GLO Granger Causality testing can be seen from appendix 8. For all companies the results are clear, the null hypothesis is accepted in all occasions. The null hypothesis states the search volume index does not have forecasting power to trading volumes. The five lags chosen are clearly jointly zero with over 99.99 percent confidence level. The global search query index values cannot be used to forecast the trading volumes happening in the OMXH. Granger causality is not the same as the generally understood causality between variables but it can be said there is very little evidence search volumes would have an causal relation with trading volumes even the theory suggest this. Statistical inference does not find Granger causality. This is not tested with the previous studies such as DEG (2011) or Wuoristo (2012) so it is difficult to find comparison to verify the test results from other sources.

The results for the SVI_FIN Granger Causality testing can be seen from appendix 9. The results are very similar to the results with the SVI_GLO as they also indicate the null hypothesis should be accepted. The null hypothesis states there is no forecasting power with SVI to trading volume. In comparison to previous studies such as DEG (2011) or Wuoristo (2012) this study provides new information even though the evidence points out SVI is unviable for forecasting. However, the results are clear for this study.

4.3 Results of the hypothesis testing

This study replicated some parts of the DEG (2011) study but also methods from Wuoristo (2012). This chapter will go through the comparison and answer as many of the hypotheses as possible in those cases where the sample or the methods allow it. In some hypotheses this study proves to be inconclusive because of sample size or the types of the companies. There are results for all the regressions but in some cases the sample size is too small to accept the results reliably. The comparison process will carry on one hypothesis at a time either accepting, rejecting or stating it cannot be answered with the current study.

> Hypothesis 1. Search volume index captures the general attention in the Finnish market.

The previous studies such as DEG (2011) and Wuoristo (2012) found SVI to be statistically significant for the mean values of share volumes and SVI in general even the effects were reversed in one or two weeks after. This study finds SVI and trading volumes move together and experience correlation. Peaks in global SVI or local SVI both happen close to peaks with trading volume. Therefore, it can be concluded SVI can capture the attention of the public investors and non-professional investors indeed use Google to search investing information on companies. This is a result of either Google Finance as a service or news bulletins giving incentives for investors to do queries about the companies they follow. In both cases SVI and attention are tied together close to interim reports and financial statements in general. The absolute query amounts are unknown and the financial queries in general experience low interest from investors. This is likely because of non-professional investor base being significantly smaller than it is in the counterparty countries from previous studies such as US or UK markets. This study cannot answer to the absolute values of the query numbers or how well the SVI captures the attention. Hypothesis 1 is rejected conditionally.

> Hypothesis 2. Search volume index correlates with the stock market trading volume but the trading volumes revert back to their mean values with some lag.

This study found only 17 companies out of 124 to have statistically significant variables such as global SVI, local SVI, variable C which is a constant for individual company and abnormal returns in relation to trading volume. In those selected companies the F-tests proved to be statistically insignificant and resulted in accepting the null hypothesis: SVI cannot explain abnormal trading volume. The result is surprising since markets such as UK FTSE and the US markets are considered to be closer to efficient markets according to efficient market hypothesis. Yet those markets have strong indication SVI can be used to forecast trading volumes. In Finnish markets SVI proved to be less of a usable tool. This can be because of two different factors. Firstly the Finnish markets are more efficient than US or UK markets and arbitrage trading in here happens faster. The possible excess returns gained from market are taken advantage of faster in the Finnish markets. The second possible option is that Google finance or search volume index in general plays a lesser role as a tool for searching information. The non-professional investors acquire their news from other sources. The answer to the hypothesis remains as: SVI has a positive correlation with trading volume on few selected companies (6 out of 124). The trading volumes do not revert back to their means, but instead act according to moving averages. That means higher trading volume periods are followed by higher trading volumes. On the other hand lower liquidity periods are followed by lower liquidity periods. Hypothesis 2 is rejected conditionally.

> Hypothesis 3. Search volume index can be used to predict and forecast and individual company's stock turnover or abnormal returns.

In previous studies SVI could be used for predicting short term abnormal returns on both US and UK markets, but it was reversed under two weeks. This study finds no such predictability. Since only 10 companies out of 124 had significant variables at 95% significance level using the mean values to generalize the market are meaningless and unusable for reliable forecasting purposes. According to this study SVI either from global or local sources cannot be used to forecast future trading volumes or abnormal returns with OMX AllShare index. Considering the index contains the companies used in many of the smaller indices such as OMXH25 it can be concluded the SVI is not a possibility to arbitrage in the Finnish markets. The possibility to forecast remains only for a handful of companies and this study cannot answer if there is a causality or Granger causality between search volumes and abnormal returns. Also the liquidity of companies for the Finnish markets is questionable at times, therefore resulting in questioning if there are enough free shares for an investor to fully arbitrage the positions if they should occur in some specific situations. The global and local SVI's are positively correlated with every company in the sample. Based on the regressions the SVI is not a usable tool for Finnish markets either because of strong market efficiency or the lack of Google queries performed by investors. The year 2013 has not been special for Google queries in other locations. Neither, the Google Finance or the common queries have experienced very low periods of query amounts. Instead the number of queries have been steadily growing for nearly a decade without any strong decreases in any field. The relational proportion of the investor queries has diminished over time, but not because investing related queries would be more rare but because Google is receiving ample amount of other queries. So it can be concluded that the relational values have been declining (investing related queries divided by total queries) because the number of individual non-investing related queries have arisen. Google is used for more information queries now than it was back in 2005. Hypothesis 3 is rejected.

> Hypothesis 4. Companies with high consumer public exposure are more affected by retail investor attention than companies that are not visible to consumers.

The hypothesis cannot be directly answered based on the OMXH. The size of the index presents problems that cannot be overcome. The companies in the index that experience high consumer vicinity and exposure are omitted from the study based on either queries that are not related to the company or lack of trading volume. One example of such companies would be Nokia. Unfortunately Nokia as a query word results in results that have nothing to do with investing related activities. The majority of the search results imply to the mobile phones unit's products such as handsets. The phones are no way related to the investing activity. Therefore using such search volume index results in creating the regressions would result in obvious false results or spurious regression. The possible correlations would not be relevant because there is no foreseeable connection between buying consumer electronics and overwhelmingly investing in the stock just based on this. Based on the final sample there is no direct approach to accept or reject the hypothesis. Hypothesis 4 is not accepted or rejected. This study finds the hypothesis inconclusive because of lack of data.

> Hypothesis 5. The SVI results from Finland are better predictors for Finnish based company share price than the SVI results gathered from globally.

Based on this study neither source of search volume index are viable predictors for either trading volume or prices as SVI_FIN and SVI_GLO prove to be indecisive. The markets function in an efficient way in which search volume index presents very little possibilities to forecast future returns. The reason for this according to previous studies can be because of the following reasons:

- The forecasting possibilities are quickly taken advantage of and vanish in intraday trading and therefore cannot be verified from either daily or weekly frequency data.
- SVI does not capture the public's attention to investing activities.
- Search volume index, trading volume and asset returns are not statistically significantly correlated because of efficient markets.

This study does not have the required information to accept or reject any of the possibilities listed above. At the moment in 2014 Google does not provide intraday SVI data and therefore further conclusions about the possibilities of SVI forecasting remain unknown. The values acquired for intraday data are through iteration and linear regression and therefore can be questionable at times. The average changes remain correct but the individual values from forecast models are of course only iterations. The intraday variance is different than the variance calculated from the aggregated data. All though this dilemma has been persistent in the previous studies also and has been left mainly ignored. The data for intraday SVI is available to Google internal use only. It can be perceived it is extremely difficult for the general public to take advantage of the possible trends within stock market attention. Hypothesis 5 is rejected. Global queries are better predictors of future values than local queries.

> Hypothesis 6. According to Granger causality there is forecasting power with SVI regards to trading volume with selected companies.

With any of the selected companies there is no Granger causality. In all tests for both global and local SVI the five lags are jointly zero, and thus the null hypothesis is accepted as is. SVI does not cause trading volume and trading volume does not cause search volume index. For all the companies tested the search volume index does not have any predicting power on either trading volume or abnormal trading volume. This hypothesis has been accepted in other studies such as DEG (2011). In the mentioned study the search volume index did have forecasting power based on ordinary least squares regressions but the trading volumes reverted back to their means in one or two weeks. In this study there is no predicting powers even though there are occasional correlation in one or two week frequency with search volume index and trading volume. Hypothesis 6 must therefore be rejected.

There relation between ABTV and abnormal returns was not directly studied, but the OLS regressions had the returns as a variable. The results for Granger causality for SVI and returns are similar to SVI and trading volume. The lags are jointly zero and therefore based on the theory behind Granger causality there is no forecasting power with any of the selected companies (Enders 2007). This hypothesis has not been verified with the previous studies and therefore there are very little grounds for comparison. Enders (2007) also points out Granger causality itself is not a causality based on pure theory so even the possible positive findings should be taken into further studies. However, based on efficient market hypothesis the markets function in in this study an efficient way. Hypothesis 6 must also be rejected. As the previous studies did not test Granger causality nor any positive results of any Granger causality tests there are no comparable data to different markets at current time.

The summary of the hypotheses is according to the following table.

Table 13 Conclusions of the hypotheses

In this table you will find the hypotheses and the conclusions.		
(H#)	**Stating**	**Conclusion**
1	Search volume index captures the general attention in the Finnish market	Rejected conditionally.
2	Search volume index correlates with the stock market trading volume (one or more variables statistically significant at 95% confidence level).	Rejected conditionally (6 out of 124 companies have statistical correlation with trading volume).
3	Search volume index can be used to predict and forecast and individual company's stock turnover or abnormal returns	Rejected.
4	Companies with high consumer recognizability are more affected by retail investor attention than companies that are not visible to consumers.	Unable to conclude. Lack of data.
5	The SVI results from Finland are better predictors for Finnish based company share price than the SVI results gathered from globally.	Rejected. The global queries are better predictors.
6	According to Granger causality, there is forecasting power with SVI regards to trading volume with selected companies.	Rejected. The lags are jointly zero. No predicting power.

The table summarizes the hypotheses in the study. None of the hypotheses are accepted. This study did not specifically test for mean level values like in the DEG (2011) studies, but considering the low mount of statistically significant variables in this study all of the hypotheses can be rejected also on mean-value level.

4.4 Fama-French three-factor model and robustness checks

For robustness tests purposes the expected returns can be calculated by using Fama-French 3-factor model. Currently for the original OLS regressions the expected returns were calculated using CAPM-model as indicated previously. This was the chosen method for the previous studies such as DEG and as this study is a replication of the selected few hypotheses the same method must be applied here also. The Fama-French 3-factor model is considered a better approach for expected returns by some researchers. It is known to have better forecasting properties because of the lack of information for CAPM. CAPM varies depending on for what the values are calculated for. The Fama-French 3-factor model is reported to be more accurate for asset return calculations. For example, the commonly used traditional capital asset pricing model is bound to using a single variable in describing the aggregated returns and considers the markets as a very large but practically immeasurable. There are different versions to CAPM that do take different factors to note. The Fama-French three-factor model divides the markets into different sections as there are three different betas. The exact formula for the model is as follows:

$$r = R_{\mathrm{f}} + \beta(K - R_{\mathrm{f}}) + \mathrm{b}L * \mathrm{SMB} + bS * HML + \alpha \qquad (14)$$

In which R_{f} is the risk free return of the assets and K is the return of the market portfolio. The variable bL stands for small market capitalization minus big market capitalization and HML for high book-to-market-ratio minus low book-to-market ratio. In these analysis the market returns for small caps are measured generally long term and there are differences between small and large cap company returns. The values can be calculated daily or with any chosen frequency but the results are always close to the normal beta. If an investor chooses to create a diverse portfolio where there are different portions of small cap and large cap companies combined with a different ratio for so called growth stocks compared to blue chips the beta may deviate from the standard beta significantly. In the case of this study the beta is almost identical because the portfolio that is being examined is exactly the index itself. However, the values for the Fama-French three-factor models are similar to the values from the capital asset pricing model. The robustness tests are therefore verified and the results are reliable after this observation. In the cases of the previous studies such as DEG (2011) and Wuoristo (2012) the robustness tests differ more from the standard values because the sample is different. Nevertheless the robustness tests are ran and they provide same statistics. Generally the Fama-French three-factor model provides more time-varying betas for markets that have larger number of growth stocks or either are more prone to using innovation with accounting. There are studies such as Foye, Mramor and Pahor (2013) that add factors to the classic three-factor model because of accounting manipulation and target seeking behavior. This is also considered to be more specific for Indian and Asian markets. The European and Nordic markets have experienced significantly less account manipulation compared to the eastern and southern counterparts. This study does not presume the OMXH All Share index is under pressure from account manipulation or deliberate manufacturing of false accounting information and therefore the classic Fama-French three-factor model is used as it is presented in the given formula, without any additional variables or modifications. The single biggest problem with reliable robustness testing is the Fama-French three-factor model beta. Because the sample size is limited since many companies had to be omitted from the study the betas do not differ significantly. The reason for omitting the small companies commonly was because they lacked sufficient trading or search volume index data. The companies that could deviate the Fama-French three-factor model beta greatly from the CAPM beta are commonly omitted. This could be avoided by still calculating the three factor model for the index without omitting the companies that lack the required information. In this case the beta would not reflect the expected returns of the sample at all and would render the results unacceptable and also incomparable. This results in the Fama-French three-factor model robustness testing to produce same results as the CAPM models, and therefore the robustness testing itself may be compromised and not reliable.

5 CONCLUSIONS

5.1 The purpose of the study and conclusions

The purpose of this study was to analyze if Google Trends can be used to predict stock trading volumes with Finnish OMXH listed shares. This was done by using Google Trend's Search Volume Index (SVI) as a proxy for investor attention. In practice if individual investors are interested in acquiring information about companies, they go to Google search landing page and type the company name to find information. This type of behavior in turn is used as a proxy for general attention towards investing in companies. The search and trading volumes are available in quantifiable form. That data is then used to create regressions and to study the results.

The conclusions for this study are different compared to the studies that were replicated, but the results in this study represent no surprises if reflected to the underlying theories. In previous studies there were clear indication of trading volume predictability and correlation. This correlation occurred between abnormal trading volume (ABTV) and abnormal search volume index. This study finds the efficient market hypothesis theory applies for the Finnish market. There are indications of behavioral biases and psychological factors influencing trading in Finnish markets. This finding has been common for previous studies. The findings are tied to the theories introduced in this study in chapter 2.2. There were three different biases commonly tied to investing activities. The representativeness, the availability and the anchoring. This study only finds anchoring bias within search volume index based on that the absolute values indicate both longer term bull and bear periods. This type of results are also found previously by Zhou (2009). High abnormal trading volumes are followed by high abnormal trading volumes, and also if the trading volumes are lower than expected, they are followed with lower than expected values. For trading volumes in almost all cases there is no reverse-to-mean effect.

According to the OLS regressions the OMXH index has significantly lower amount of companies that have abnormal trading volume correlated with abnormal search volume index compared to either the US markets or the UK stocks. This result is surprising when compared to the findings of DEG (2011) and Wuoristo (2012). In most other finance studies the US and the UK markets are considered to be more efficient based on financial theory and efficient markets (Wuoristo 2012). The reasons for higher efficiency expectations are because of higher liquidity and transparent company policies (Fama 1991). According to classical finance theory those markets have less possibilities of arbitrage than the markets in peripheral areas such as the Nordic countries where arbitrages are considered to be risky and more inefficient thus causing mispricing and behavioral bias influences (Shleifer & Vishny 1997). The reason for such results are either because the Finnish markets are closer to strong-form efficient than the counterparts or because Google's search volume index does not experience active enough use to be a reliable proxy for investor attention for the chosen market. As there are no public absolute values for the query amounts this study is unable to provide reliable results. The search volume originating from global queries resulted to be better predictor of trading volume than the search volume from local sources. Previous studies such as DEG (2011) and Wuoristo (2012) found SVI from both local and global sources to be a reliable proxy for investor attention. This study found global sources are more reliable than the local queries, and the vast majority of both sources to be inaccurate in explaining trading volume behavior. This can be seen as an indirect test for market efficiency. For Finnish stocks the markets operate in an efficient way and SVI cannot be used as a tool to arbitrage at least not with the used frequencies such as a weekly or daily data. As stated in the beginning of this study the major contributions this study can have are:

1. Does Search Volume Index capture the general attention in the Finnish market?

As seen from results or data SVI does fluctuate over time and there are positive search volume index values for all of the companies within the OMX index. The positive values indicate the public uses Google as a tool to acquire information about the companies. This provides the ground to run analysis and create regressions. In the study many of the companies had to be omitted because of several different reasons. These reasons are not enough searches or that the searches did not represent financial related queries. In comparable studies Google proved to be a more reliable tool with US or UK markets than with Finnish markets. This study cannot answer what are the reasons for having less investor related queries from Finland then there are from global sources for other markets. This is not because of low Google penetration since it is reported by different traffic observation web pages to be one of the most popular sites in Finland.

2. Does Search Volume Index correlate with the stock market trading volume (in a statistically significant level)?

The SVI can be used as proxy for only a few companies in the Finnish market. The results occur with lower absolute values than in the comparable studies such as DEG (2011) and Wuoristo (2012) where there were findings of reliable correlation with mean values.

3. Can Search Volume Index be used to predict and forecast an individual company's stock turnover?

Forecasting trading volumes in some cases provide possibilities for arbitraging. According to the efficient markets these situations should not prevail for extended periods of time. According to the study by Mondria and WU (2011) information asymmetry is also a decisive factor. The different combinations are indicated here:

- Local attention high, non-local low -> SVI can be efficiently used to predict future share prices
- Local attention high, non-local high -> no statistically significant predictive powers
- Local attention low, non-local high -> no statistically significant predictive powers

This study finds exactly the same results as Mondria and Wu (2011). The SVI focus comes from non-local sources and therefore the findings indicate there is no significant predictive powers with SVI. The predictive powers represent themselves according to Mondria and Wu only when there is information asymmetry and only when the local attention is higher than the global. In this study the global attention is higher compared to the local attention. The results are therefore according to the theory for predictability. SVI cannot be used to predict the abnormal turnover unlike in the studies by DEG (2009) or Wuoristo (2012). Interestingly, this study finds abnormal returns to be correlated with abnormal trading volume. This can be seen as the basis for momentum investing within Finnish markets. Abnormally high returns calculated by capital asset pricing models expected returns are correlated with abnormally high trading volume calculated by moving averages. This is a clear contradiction to the efficient market hypothesis.

5.2 Future studies and limitations

The possibilities for Google Trends are immense. Using the search volume index as a proxy for investor attention is only one option among the many other possibilities. Almost all of the studies that are conducted have been made with data from either the United States or from the UK. This leaves many possibilities to conduct studies with more local data. As it is seem from this study there are differences between the different countries.

Another more complex way to study the topic would be to start using combinations of words. For example, using set of query strings such as *name+ticker*. Another option would be to use the strings with omitting a keyword from them. For example, an option could be *name+ticker-buy*. The word *buy* could be replaced with the words purchase, acquire and with many others. This could reduce the noise within the data significantly and would in turn increase the sample size. In many cases the queries result in individuals trying to find information about the company's products instead of wanting to invest in the company. From the financial study perspective this is the reason many of the companies were omitted from the data sample. This study as the also the previous studies by DEG (2011) and Wuoristo (2012) had to omit a large portion of the companies that were initially in the index. In this study over 50 percent of the companies had to be omitted, and the study by Wuoristo (2012) over 80 percent of the companies were omitted.

This study also suffered from survivorship bias. This means the sample consists of companies that are currently listed in the chosen index. The companies that have failed in the past are not represented at all. To overcome this the time-period must be chosen differently. But with longer periods of time there can be situations where a company does not experience enough trading. This can result in omitting the company from the study. Another option is to change the frequency of the data. Instead of using interpolated daily data or weekly data, a month or even a quarter's length could be utilized. These results are not comparable because of different variance values naturally, but the new regressions may produce different results than cannot be observed in different frequencies. The limitations of the study are clear. Based on the available frequencies there can be a lot of correlation or forecasting possibilities that remain unfound. For example, SVI could correlate with trading volumes in intraday data because of automated trading. With OMXH the automated machine trading has grown both in size and in relative portion to other trading activities every year. The machines react to trends far faster than individuals who search information manually from Google queries. Unfortunately Google does not provide intraday data for public use.

REFERENCES

About Lex <http://www.ft.com/intl/cms/s/2/7b4c9706-7e1d-11df-8478-00144feabdc0.html#axzz3920NIjRS> Retrieved 12.1.2015

Ackert, L. – Church, K. – Deaves, R. (2003) Emotion and financial markets. *Federal Reserve Bank of Atlanta Economic Review*, Q2/2003, 33-41.

Akerlof, G. – Shiller, R. (2009) How animal spirits destabilize economies. *Mckinsey Quarterly*, No. 3, 126-135.

Allen, W. – Evans, D. (2005) Bidding and overconfidence in experimental financial markets. Journal of Behavioral Finance , Vol. 6, No. 3, 108-120

Askitas, N. and Zimmermann, K. F., 2010. *Google econometrics and unemployment forecasting*, Applied Economics Quarterly, 2009, 55 (2), 107-120

Antweiler, W. and Frank, Z.F., 2004. Is All that Talk Just Noise? The information content of Internet stock message boards, *Journal of Finance* 59(3), 1259-1294.

Baker, M. – Stein, J. (2004) Market liquidity as a sentiment indicator. Journal of Financial Markets, Vol. 7, 271-299.

Baker, M. –Wurgler, J. (2002) Market timing and capital structure. *The Journal of Finance*, Vol. 57, No. 1, 1-32.

Baker, M. – Wurgler, J. (2004) A catering theory of dividends. *The Journal of Finance*,Vol. 59, No. 3, 1125-1165.

Baker, M. – Wurgler, J. (2006) Investor sentiment and cross-section of stock returns. *Journal of Finance*, Vol. 61 (4), 1645–1680.

Baker, M. – Wurgler, J. (2007) Investor sentiment in stock market. *Journal of economic perspectives*, Vol. 21 (2), 129–151.

Banerjee, P. S. – Doran, J. S. – Peterson, D. R. (2007) Implied volatility and future portfolio returns. *Journal of Banking & Finance*, Vol. 31 (10), 3183–3199.

Bandopadhyaya, A. (2006) Measures of Investor Sentiment: Who wins the horse race? *Financial services forum*, University of Massachusetts.

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*, Vol. 21 (2), 785–817.

Barber B and Odean T (2000), “Trading is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors”, *Journal of Finance*, Vol. 55, No. 2, pp. 773-806.

Barber B and Odean T (2001), “Boys Will be Boys: Gender, Overconfidence and Common Stock Investment”, *Quarterly Journal of Economics,* Vol. 116, No. 1, pp. 261-292.

Baumeister, R. – Bratslavsky, F.E. – Finkenauer, C. – Vohs, D. (2001) Bad is stronger than good. *Review of general psychology*, Vol. 5 (4), 323-370.

Bhattacharya, U. – Galpin, N. – Ray, R. – Yo, X. (2009) The role of the media in the Internet IPO bubble. *Journal of financial and quantitative analysis,* Vol. 44 (3), 657–682.

Berglund, T. – Wahlroos, B. – Örnmark, A. (1983) The Weak-form Efficiency of the Finnish and Scandinavian Stock Exchanges. *Scandinavian Journal of Economics*, Vol. 85, No. 4, 521-530.

Bergman, N. – Roychowdhury, S. (2008) Investor sentiment, expectations and corporate disclosure. *Journal of Accounting Research* , Vol. 46, No. 5, 1057-1083.

Biais, B. – Weber, M.(2008) Hindsight bias, risk perception and investment performance. To appear in: *Management Science*.

Bhandari G and Deaves R (2006), “The Demographics of Overconfidence”, *Journal of Behavioral* Finance, Vol. 7, No. 1, pp. 5-11.

Black, F. (1986) Noise. *Journal of Finance,* Vol. 41 (3), 529–543.

Brown, G. W. – Cliff, M. T. (2004) Investor sentiment and the near-term stock market. *Journal of empirical finance,* Vol. 11 (1), 1–27.

Booth, G. – Kallunki, J. – Martikainen, T. (1999) Earnings news and the behavior of large and small traders in the Finnish stock market. *Applied Economic Letters*, Vol. 6, 771-774.

Brooks, C. (2002) Introductory econometrics for finance. *Cambridge University Press*, Great Britain.

Canbas, S. (2009) Investor sentiment and stock returns: Evidence from Turkey. Emerging Markets Finance and Trade, Vol. 45, No. 4, 36-52.

Caparrelli, F. – D'Arcangelis, A. – Cassuto, A. (2004) Herding in the Italian stock market: A case of behavioral finance. The Journal of Behavioral Finance, Vol. 5, No. 4, 222-230.

Chan, K. – Hameed, A. – Tong, W. (2000) Profitability of momentum strategies in the international equity markets. Journal of Financial and Quantitative Analysis, Vol. 35, No. 2, 153-17

Chaarlas, L.J. (2012), Behavioral finance A boon to investors. *Journal of Finance, Accounting and Management*, Vol. 3 (1), 32–44.

Cen, L. – Hilary, G. – Wei, K. C. J. (2013) The Role of anchoring bias in equity markets Evidence from analysts' earnings forecast and stock returns *Journal of Financial and quantitative analysis,* Vol. 48 (1), 47–76.

Chughtai, S. (2010) *Semi strong Market Efficiency: An empirical study of Karachi stock exchange*. Interdisciplinary Journal of Contemporary Researching Business, Vol. 2 (2), 179–205.

Chordia, T., Huh, S-W. and Subrahmanyam, A, 2007.The cross-section of expected trading activity, *Review of Financial Studies* 20 (3), 709–740.

Cooper, C., Mallon, K., Leadbetter, S., Pollack, L. and Peipins, L. 2005.Cancer Internet search activity on a major search engine:United States 2001-2003. *J Med Internet res* 7(3) , e36.

Corwin, A.S. and Coughenour, F.J., 2008. Limited Attention and the Allocation of Effort in Securities. Trading. *Journal of Finance* 63(6), 3031-3067.

Da, Z., Engelberg, J. and Gao, P., (a) 2010. *Internet Search and Momentum,* Working paper, University of Notre Dame and University of North Carolina at Chapel Hill.

Da, Z., Engelberg, J. and Gao, P., (b) 2010. *The sum of all FEARS: Investor sentiment and Asset Price,* Working paper, University of Notre Dame and University of North Carolina at Chapel Hill.

Daniel, K. – Hirshleifer, D. – Subrahmanyam, A. (1998) Investor Psychology and Security Market Under-and Over-Reactions. Journal of Finance, Vol. 53, No. 6, 1839-1885.

Davis, N. (2003) The Triumph of Contrarian Investing. McGraw-Hill, USA

DeLong, J. – Shleifer, A. – Summers, L. – Waldman, R. (1990a) Noise trader risk in financial markets. *The Journal of Political Economy*, Vol. 98 (4), 703 – 738.

DeLong, J. – Shleifer, A. – Summers, L. – Waldman, R. (1990b) Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance,* Vol. 45 (2), 374–397.

Dimpfl, T. and Jank, S., 2012. *Can Internet search behavior help to predict stock market volatility?* Paris December 2012 Finance Meeting EUROFIDAI-AFFI Paper. Available at <http://ssrn.com/abstract=1941680> or <http://dx.doi.org/10.2139/ssrn.1941680> Retrieved 9.6.2014

Dittrich, D. – Güth, W – Maciejovsky, B. (2005) Overconfidence in investment decisions: An experimental approach. *The European Journal of Finance,* Vol. 11, No. 6, 471-491.

Dougherty, C. (2002) Introduction to Econometrics. *Oxford University Press,* New York.

Drakos, K. (2009) Terrorism activity, investor sentiment and stock returns. *Review of financial economics*, Vol. 19, No. 3, 128-135

Dzielinksi, M., 2012.Measuring economic uncertainty and its impact on the stock market, *Finance Research Letters* **9**(3), 167-175.

Enders, W. (2004) *Applied econometric time series*. 2nd ed. John Wiley & Sons, inc. Hamilton printing, United States.

Eviews (2009) Eviews 7 user guide.

Engle, R. F.– Granger, C. W. J. (1987) Co-integration and error correction: Representation, estimation and testing. *Econometrica,* Vol. 55 (2), 251–276.

Ettredge, M., Gerdes, J. and Karuga. G., 2005. Using web-based search data to predict macroeconomic statistics. *Communications of the ACM* 48(11), 87-92.

Fama, E. (1965) Random Walks in Stock Market Prices. *Financial analyst journal*, Vol. 21 (5), 55–59.

Fama, E. (1970) Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*, Vol. 25 (2), 383–417.

Fama, E. (1991) Effiecent capital markets II. *Journal of Finance*, Vol. 46 (5), 1575–1617.

Fang, L. – Peress, J. (2009) Media coverage and the cross-section of stock returns. *Journal of Finance*, Vol. 64 (5), 2020–2052.

Fama, E.F. and Macbeth, J. D., 1973. Risk return and equilibrium: Empirical tests, *Journal of Political Economy* 81(3), 607–636.

Ferguson, N. (2008) *Rahan nousu*. Hakapaino, Helsinki.

Geman, H. (2002) Pure jump levy processes for asset price modeling. *Journal of Banking and Finance*, Vol. 26 (7), 1297–1316.

Gentzkow, M – Shapiro J. (2006) Media bias and reputation. *Journal of political economy*, Vol. 114 (2), 280–316.

Gentzkow, M. – Shapiro, J. (2010) What drives media slant? Evidence from U:S daily newspapers. *Econometrica,* Vol. 78 (1) 35–71.

Gibbons, M. R – Hess, P. (1981) Day of the week effects and asset returns. *Journal of Business,* Vol. 54 (4), 579–596.

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

Guzman, G,. 2011. Internet search behavior as an economic forecasting tool. *The Journal of Economic and Social Measurement* 36(3), 119-167.

Gutter M S, Fox J and Montalto C P (1999), "Racial Differences in Investor Decision Making", *Financial Services* Review, Vol. 8, No. 3, pp. 149-162.

Gutter M S and Fontes A (2006), "Racial Differences in Risky Asset Ownership: A Two-Stage Model of the Investment Decision-Making Process", Financial Counseling and Planning, Vol. 17, No. 2, pp. 64-78.

Goel, S., Hofman, J.M., Lahaie, S, Pennock, D,M. and Watts, D,J. 2010. Predicting consumer behavior with web search, *Proceedings of the National Academy of Sciences*, **7**(41), 17486-17490 .

Grinblatt, M. –Keloharju, M. (2001) What makes investors trade? *The Journal of Finance*, Vol. 56, No. 2, 589-616.

Grinblatt, M. – Keloharju, M. (2009) Sensation Seeking, overconfidence and trading activity. *The Journal of Finance*, Vol. 64, No. 2, 549-578.

Grinblatt, M. – Titman, S. – Wermers, R. (1995) Momentum investment strategies, portfolio performance and herding: a study of mutual fund behavior. *The American Economic Review,* Vol. 85, No. 5, 1088-1105.

Granger, C. W. J. (1969) Investigating causal relations by econometric models and cross-spectral methods. *Econometrica,* Vol. 37 (3), 427–438.

Granger, C. W. J. (1981) Some properties of time series data and their use in econometric model specification. *Journal of Econometrics*, Vol. 16 (1), 121–130.

Google Trends <www.google.com/trends> Retrieved 15.8.2014

Google Trends <www.google.org/flutrends/> Retrieved 20.8.2014

Hair, J. – Anderson, R. – Tatham, R. – Black, W. C. (1998) 5th ed. *Multivariate Data Analysis*. Prentice-Hall Upper Saddle River, New Jersey.

Hamilton, D. (2011) Forecasting with structural models and VARs: Relative advantages and the client connection. *Foresight, The international journal of applied forecasting*, Vol. 1 (23), 37–42

Haliassos M and Bertaut C C (1995), "Why Do So Few Hold Stocks?", *The Economic Journal*, Vol. 105, No. 432, pp. 1110-1129.

Hirshleifer, D – Lim, S. S. – Teoh S. H (2009) Driven to distraction: Extraneous events and under reaction to earnings news. *Journal of Finance*, Vol. 64 (5), 2289–2323.

Hirshleifer, D. – Teoh, S. H. (2003) Limited attention, information disclosure, and financial reporting. *Journal of accounting and Economics*, Vol. 36 (1), 337–386.

Hong H, Kubik J D and Stein J C (2004), "Social Interaction and Stock Market Participation", *The Journal of Finance*, Vol. 59, No. 1, pp. 137-163.

Johanssen, S. (1991) Estimation and hypothesis testing of cointegration and vectors in Gaussian vector autoregressive models. *Econometrica*, Vol. 59 (6), 1551–1580.

Johnson, H., 2004. Analysis of Web access logs for surveillance of influenza. *Stud. Health Technol. Inform.*107(2), 1202–1206.

Jokivuori, P. – Hietala, R. (2007) *Määrällisiä tarinoita*: Monimuuttujamenetelmien käyttö ja tulkinta.. WSOY, Helsinki.

Joutz, F. L. – Maddala, G. S. – Trost, R. P. (1995) An Integrated Bayesian vector autoregression and error correction model for forecasting electricity consumption and prices. *Journal of Forecasting*, Vol. 14 (3), 287–310.

Jurafsky, D. – Martin, J. H. (2009) *Speech and Language Processing*. Prentice Hall, Upper Saddle River, New Jersey.

Kahneman, D. – Tversky, A. (1979) Prospect theory: An analysis of decision under risk. Econometrica, Vol. 47 (2), 263–292.

Kahneman, D.,1 973, *Attention and efforts*. Prentice-Hall, Englewood Cliffs, NJ.

Kahneman, D. and Tversky, A., 1982. The psychology of preferences. *Scientific American* 246(2), 160–173.

Keim, D. B. – Madhavanb, A. (1995) Anatomy of the trading process Empirical evidence on the behavior of institutional traders. *Journal of Financial Economics*, Vol. 37 (3), 371– 398.

Keynes, J.M. (1936) *The General Theory of Employment, Interest, and Money*. Harcourt, Brace, New York.

Klibanoff, P. – Lamont, O. – Wizman, T. A. (1998) Investor Reaction to Salient News in Closed-End Country Funds. *Journal of Finance*, Vol. 53 (2), 673–699.

Langer E and Roth J (1975), "Heads I Win, Tails it's Chance: The Illusion of Control as a Function of the Sequence of Outcomes in a Purely Chance Task", Journal ofPersonality and Social Psychology, Vol. 32, No. 6, pp. 951-955.

Leinweber, D. – Sisk, J. (2010) Relating news analytics to stock returns. http://www.machinereadablenews.com/images/dl/Leinweber-Sisk-CARISMA-Jan26-BestPDF%20%282%29.pdf Retrieved 21.6.2014

Li, F. (2006) Do stock markets investors understand the risk sentiment of corporate annual reports. *SSRN Working paper*

Lui, C., Panagiotis, T. and Mustafaraj, E., 2011. On the predictability of the U.S. elections through search volume activity, *Department of Computer Science*, Wellesley College.

Long, J. B. – Shleifer, A. – Summers, L. H. Waldmann R. J. (1990) Noise trader risk in financial markets. *Journal of political economy*, Vol. 98 (4), 703–738.

Loughran, T – Mcdonald B. (2011) When is a liability not a liability? Textual analysis, Dictionaries and 10-Ks. *Journal of Finance*, Vol. 66 (1), 35–65.

Malkiel, B. (2003) The efficient market hypothesis and its critics. *The Journal of Economic Perspectives*, Vol. 17, No. 1, 59-82.

Malkiel, B. (2005) Reflections on the efficient market hypothesis: 30 years later. *The Financial Review Vol.* 40, No. 1, 1-9.

Malkiel, B. (2007) *Sattumankauppaa Wall Streetillä*. Talentum, Helsinki.

Marsden, A. – Veeraraghavan, M. – Ye, M. (2008) Heuristics of representativeness, anchoring and adjustment, and leniency: Impact on earnings' forecasts by Australian analysts. Quarterly Journal of Finance and Accounting, Vol.47, No. 83-102.

Marttila, J. (2001) Järki ja tunteet osakemarkkinoilla. Edita Oyj, Helsinki.

Mehra, R. – Sah, R. (2002) Mood fluctuations, projection bias and volatility of equity Prices. Journal of Economic Dynamics and Control, Vol. 26, No. 5, 869-887.

Metsämuuronen, J. (2008) *Monimuuttujamenetelmien perusteet.* Gummerus, Jyväskylä.

Mitra, L. – Mitra, G. (2010) Applications of news analytics in finance: a Review <http://optirisk-systems.com/papers/Opt0014.pdf> Retrieved 20.6.2014

Miller D T and Ross M (1975), "Self-serving Bias in Attribution of Causality: Fact or Fiction?", Psychological Bulletin, Vol. 82, No. 2, pp. 213-225.

Moilanen, K. – Pulman, S. – Zhang, Y. (2010) Packed Feelings and Ordered Sentiments: Sentiment Parsing with Quasi-compositional Polarity Sequencing and Compression. 19th *European conference on artificial intelligence Lisbon, Portugal*. http://www.sutd.edu.sg/cmsresource/faculty/yuezhang/wassa10.pdf> Retrieved 10.6.2014

Mondria, J. and Wu, T., 2012. *Familiarity and Surprises in International Financial Markets: Bad news travels like wildfire, good news travels slow*, Working paper. University of Chapel Hill and University of Santa Cruz.

Mondria, J. and Wu, T., 2011. *Asymmetric Attention and Stock Returns*, Society for Economic Dynamics, Meeting Papers 134.

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

Merton, R.C., 1987. A simple model of capital market equilibrium with incomplete information, *Journal of Finance* 42(3), 483–510.

Newey, W. K. – West, K. D. (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation. *Econometrica*, Vol. 55 (3), 703–708.

Schadéwitz, H. – Blevins, D. (1998) A preliminary analysis of the semi-strong efficiency of the Helsinki stock exchange. *International Journal of Management*, Vol. 15, No. 2.

Schmeling, M. (2009) Investor sentiment and stock returns: Some international Evidence. Journal of empirical finance No. 16, 394-408.

Schwarz, N. (2010) Feelings-as-information theory. To appear in Handbook of theories of social psychology, eds. Van Lange, P. – Kruglanski, A. – Higgins,T., Sage.

Sherman, S. – Cialdini, R – Schwartzman, D. – Reynolds, K. (1985) Imagining can heighten or lower the perceived likelihood of contracting a disease: The mediating effect of ease of imaginery. *Personality and social psychology bulletin*, Vol.11, 118-127.

Sheung, H. – Lu, Y. – Wei, Y. (2010) Causalities between sentiment indicators and stock market returns under different market scenarios. *The International Journal of Business and Finance Research*, Vol. 4, No. 1, 159-171.

Shiller, R. (2003) From efficient markets theory to behavioral finance. *The Journal of Economic Perspectives,* Vol. 17, No. 1, 83-104.

Shiller, R. (2005) Irrational exuberance. Princeton University Press, New Jersey. Shiller, R. (2009) The Subprime Solution. *Princeton University Press*, New Jersey.

Shleifer, A. (2000) *Inefficient markets: an introduction to behavioral finance*. Oxford university press, Oxford.

Shleifer, A. – Summers, L. H. (1990) The noise trader approach to finance. *Journal of economic perspectives,* Vol. 4 (2), 19–33.

Shleifer, A. – Vishny, R. W. (1997) The limits of arbitrage. *Journal of Finance,* Vol. 52 (1), 35–55.

Schmeling, M. (2009) Investor sentiment and stock returns: Some international evidence. *Journal of empirical finance,* Vol. 16 (3), 39 –408.

Sims, A. (1980) Macroeconomics and reality. *Econometrica,* Vol. 48 (1), 1–47.

Solomon, D. H. (2012) Selective publicity and stock prices. *Journal of Finance* Vol. 67 (2), 599–537.

Scott, J. – Stumpp, M. – Xu, P. (2003) Overconfidence bias in international stock prices *The journal of portfolio management,* Vol. 29 (2), 80–89.

Seasholes, M,S. and Guojun W., 2007. Predictable behavior, profits, and attention, Journal of Empirical Finance 14(5), 590–610.

Solomon, D. H. (2012) Selective publicity and stock prices. *Journal of Finance* Vol. 67 (2), 599–537.

Odean, T., 1999. Do investors trade too much? *American Economic Review* 89(5), 1279–1298.

Park, H. – Sohn, W. (2013) Behavioral Finance: A survey of literature and recent development. *Seoul Journal of Business,* Vol. 19 (1), 3–42.

Pavabutr P (2002), *"*Investor Behavior and Asset Prices*"*, Sangvien Conference.

Peress, J. (2011) The Media and the Diffusion of Information in Financial Markets: Evidence from Newspaper Strikes. *INSEAD Working paper*

Pesaran, H. H. – Shin, Y. (1998) Generalized impulse response analysis in linear multivariate models. *Economic letters*, Vol. 58 (1), 17–29.

Pompian M M (2006), *Behavioral Finance and Wealth Management; How to Build Optimal Portfolios that Account for Investor Bias*, pp. 104-111, John Wiley & Sons, Inc.

Riffe, D. – Lacy, S. – Fico, F. (2005) *Analyzing media messages: Using quantitative content analysis in research.* Lawrence Erlbaum associates, Mahwah, New Jersey.

Tabachnick, B. – Fidell, L. (2007) *Using multivariate statistics.* Pearson, Boston.

Tetlock, P. C. (2007) Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *Journal of Finance*, Vol. 62 (3), 1139–1168.

Tetlock, P. C. – Saar-tsechansky, M. – Macskassy S. (2008) More Than Words: Quantifying Language to Measure Firms' fundamentals. *Journal of Finance*, Vol. 63 (3), 1437–1467.

Tesar, L.L., and Werner, I., 1995. Home bias and high turnover, Journal of International Money & Finance 14(4), 467-492.

Tsay, R. S. (2005) *Analysis of financial time series*. 2nd ed. John Wiley & Sons, Inc., Hoboken, New Jersey.

Tversky A (1990), "The Psychology of Risk", Conference Proceedings of the Quantifying the Market Risk Premium Phenomenon for Investment Decision Making Conference, Conducted by AIMR, Association for InvestmentManagement and Research, Charlottesville, VA, Cited in Schalkwyk C (n.d.), "Creating Wealth in a Developing Economy by Incorporating Investor Psychology in Portfolio Construction.

Tversky A and Kahneman D (1981), "The Framing of Decisions and the Psychology of Choice", Science, Vol. 211, No. 1, pp. 453-458.

Vaihekoski, M. (1997) The Finnish stock market : Recent trends and important events. Liiketaloudellinen aikakauskirja, Vol. 4, 526-543.

Vaihekoski, M. (2004) Rahoitusalan sovellukset ja Excel. WSOY, Helsinki.

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

Virtanen, I. – Yli-Olli, P. (1987) Forecasting stock market prices in a thin security Market. Omega International *Journal of Management*, Vol. 15, No. 2, 145-155.

Yatiwella, W. B. (2011) Calendar anomalies: evidence from Colombo stock exchange. *Economics, Management and financial markets*, Vol. 6 (2), 84–105.

Wuoristo, L. (2012) Can you google the future? A study on the predictive power of Google Trends on company shares in the UK. *Aalto University*, Department of Finance.

Wood A S (1997), "Behavioral Risk: Anecdotes and Disturbing Evidence", The Journal of Investing, Vol. 8, No. 1, pp. 8-11.

Xiao J J (1992), "Effects of Selected Family Characteristics in Interrelated Components of Household Asset Portfolios", Ph.D. Dissertation, Oregon State University, Corvallis, OR. *Counseling and Planning*, Vol. 6, No. 1, pp. 99-106.

Xiao J J and Olson G I (1993), "Mental Accounting and Saving Behavior", *Home Economics Research Journal*, Vol. 22, No. 1, pp. 92-109

Yu, W. W – Lui, E. C. K. – Wang, J.W. (2010) The predictive power of the implied volatility of options trader OTC and exchanges. *Journal of Banking & Finance*, Vol. 34 (1), 1–11.

Zhang, – Y. Difang, W. – Fu L. (2013) Impact of media on stock returns: an in-depth empirical study in China. *Chinese Management Studies*, Vol. 7 (4), 586–603.

Zhou, R. – Lai, R. (2009) Herding and information based trading. *The Journal of Empirical Finance*, Vol. 16, 388-393.

APPENDIX 1 ALL LISTED COMPANIES

OMX Helsinki All Share Index Constituents

Company name	Ticker symbol
Affecto PLC	AFE1V:HEX
Ahlstrom Oyj	AHL1V:HEX
Aktia Bank PLC	AKTRV:HEX
Alma Media Oyj	ALN1V:HEX
Amer Sports Oyj	AMEAS:HEX
Apetit Oyj	APETI:HEX
Aspo Oyj	ASU1V:HEX
Atria Oyj	ATRAV:HEX
Basware Oyj	BAS1V:HEX
Biohit Oyj	BIOBV:HEX
Citycon Oyj	CTY1S:HEX
Componenta Oyj	CTH1V:HEX
Comptel Oyj	CTL1V:HEX
Cramo Oyj	CRA1V:HEX
Digia Oyj	DIG1V:HEX
Dovre Group Oyj	DOV1V:HEX
Efore PLC	EFO1V:HEX
Elisa Oyj	ELI1V:HEX
Etteplan Oyj	ETT1V:HEX
F-Secure Oyj	FSC1V:HEX
Finnair	FIA1S:HEX
Finnlines Oyj	FLG1S:HEX
Fiskars Oyj	FIS1V:HEX
Fortum	FUM1V:HEX
Honkarakenne Oyj	HONBS:HEX
Huhtamaki Oyj	HUH1V:HEX
Incap Oyj	ICP1V:HEX
Ixonos Oyj	XNS1V:HEX

Kemira Oyj	KRA1V:HEX
Keskisuomalainen Oyj	KSLAV:HEX
Kesko Oyj	KESAV:HEX
Kesla Oyj	KELAS:HEX
Kone Oyj	KNEBV:HEX
Konecranes Abp	KCR1V:HEX
Lemminkainen Oyj	LEM1S:HEX
M-real Oyj	MRLAV:HEX
Marimekko Oyj	MMO1V:HEX
Martela Oyj	MARAS:HEX
Metso Oyj	MEO1V:HEX
Neo Industrial Oyj	NEO1V:HEX
Neste Oil Corporation	NES1V:HEX
NOKIA	NOK1V:HEX
Nokian Tyres PLC	NRE1V:HEX
Nordea Bank AB	NDA1V:HEX
Nurminen Logistics Oyj	NLG1V:HEX
Olvi Oyj	OLVAS:HEX
Oriola KD Oyj	OKDAV:HEX
Outokumpu	OUT1V:HEX
Outotec Oyj	OTE1V:HEX
Ponsse Oyj	PON1V:HEX
Poyry Oyj	POY1V:HEX
Ramirent Oyj	RMR1V:HEX
Rapala VMC Corp	RAP1V:HEX
Rautaruukki Oyj	RTRKS:HEX
Raute Oyj	RUTAV:HEX
Revenio Group Oyj	REG1V:HEX
Sampo Oyj	SAMAS:HEX
Soprano Oyj	SOPRA:HEX
Sponda Oyj	SDA1V:HEX
SRV Yhtiot Oyj	SRV1V:HEX
Stockmann Oyj	STCAS:HEX

Stora Enso	STERV:HEX
Takoma Oyj	TAM1V:HEX
Talvivaaran Kaivososakeyhtio Oyj	TLV1V:HEX
Technopolis Oyj	TPS1V:HEX
Tecnotree Oyj	TEM1V:HEX
Teleste Oyj	TLT1V:HEX
TeliaSonera	TLS1V:HEX
Tieto Oyj	TIE1V:HEX
Tikkurila Oyj	TIK1V:HEX
Trainers' House Oyj	TRH1V:HEX
Tulikivi Oyj	TULAV:HEX
Turvatiimi Oyj	TUT1V:HEX
UPM Kymmene Oyj	UPM1V:HEX
Uponor Oyj	UNR1V:HEX
Vacon Oyj	VAC1V:HEX
Vaisala Oyj	VAIAS:HEX
Valmet Corp	VALMT:HEX
Viking Line Abp	VIK1V:HEX
Wartsila Oyj	WRT1V:HEX
Yit Oyj	YTY1V:HEX
Yleiselektroniikka Oyj	YLEPS:HEX
Total number of companies	(124)

APPENDIX 2 SUMMARY OF OLS REGRESSIONS

Summary of OLS regressions							
	Dependent variable: ABTV (abnormal Trading Volume)						
regr. ABTV = B0 + B_{1*}ASVI_FIN+B_{2*}ASVI_GLO+B_{3*}ABRETURNS+e							
	Company	Variable	Coefficient	Std. Error	t-Statistic	Prob.	R-squared
1	aktia	C	1.798185	0.281252	6.393501	0.0000	0.016002
		ASVI_FIN	20.19162	11.96650	1.687345	0.0929	
		ASVI_GLO	-16.61221	11.42394	-1.454157	0.1472	
		ABRETURNS	-15.33881	15.36756	-0.998129	0.3192	
2	apetit	C	0.040927	0.027491	1.488728	0.1379	0.013225
		ASVI_FIN	0.133857	0.109420	1.223328	0.2224	
		ASVI_GLO	-0.207724	0.849436	-0.244544	0.8070	
		ABRETURNS	-3.031556	2.294215	-1.321392	0.1877	
3	atria	C	-0.002924	0.020073	-0.145650	0.8843	0.034978
		ASVI_FIN	0.671978	0.241391	2.783780	0.0058	
		ASVI_GLO	-0.759360	0.625660	-1.213695	0.2261	
		ABRETURNS	-1.367055	1.388824	-0.984326	0.3260	
4	basware	C	0.017197	0.038228	0.449844	0.6532	0.038182
		ASVI_FIN	0.180325	0.361642	0.498629	0.6185	
		ASVI_GLO	0.026422	0.451209	0.058557	0.9534	
		ABRETURNS	5.085843	1.720657	2.955755	0.0034	
5	cramo	C	0.009063	0.021523	0.421094	0.6741	0.009435
		ASVI_FIN	0.038336	0.199597	0.192069	0.8479	
		ASVI_GLO	-0.038196	0.265426	-0.143904	0.8857	
		ABRETURNS	1.536265	1.048900	1.464644	0.1444	
6	digia	C	0.075475	0.052926	1.426064	0.1552	0.021780
		ASVI_FIN	0.207873	0.200127	1.038702	0.3000	
		ASVI_GLO	0.056706	0.533130	0.106365	0.9154	
		ABRETURNS	6.777225	3.345574	2.025729	0.0439	

7	elisa	C	0.001430	0.011698	0.122272	0.9028	0.015151
		ASVI_FIN	-0.689054	0.597046	-1.154104	0.2496	
		ASVI_GLO	0.357341	0.324990	1.099546	0.2727	
		ABRETURNS	1.153099	0.982233	1.173957	0.2416	
8	f-secure	C	0.046163	0.028399	1.625529	0.1054	0.050237
		ASVI_FIN	-0.715093	0.418923	-1.706978	0.0891	
		ASVI_GLO	0.768244	0.572727	1.341380	0.1811	
		ABRETURNS	3.389221	1.112710	3.045916	0.0026	
9	finnlines	C	0.843695	0.144461	5.840300	0.0000	0.030146
		ASVI_FIN	-0.871542	1.871568	-0.465675	0.6419	
		ASVI_GLO	-4.292896	2.309901	-1.858476	0.0644	
		ABRETURNS	-1.320157	6.806816	-0.193946	0.8464	
10	fiskars	C	0.018089	0.026632	0.679212	0.4977	0.015448
		ASVI_FIN	-0.068418	0.304262	-0.224864	0.8223	
		ASVI_GLO	0.419543	0.471014	0.890724	0.3740	
		ABRETURNS	4.023129	2.349741	1.712159	0.0882	
11	fortum	C	0.002438	0.011214	0.217396	0.8281	0.046962
		ASVI_FIN	-0.305794	0.185467	-1.648778	0.1005	
		ASVI_GLO	0.783738	0.286140	2.739001	0.0066	
		ABRETURNS	-1.000498	0.954023	-1.048714	0.2954	
12	honkarakenne	C	0.811511	0.113780	7.132257	0.0000	0.029611
		ASVI_FIN	0.123509	1.424762	0.086688	0.9310	
		ASVI_GLO	-2.576838	1.557597	-1.654367	0.0994	
		ABRETURNS	4.328182	5.684565	0.761392	0.4472	
13	kemira	C	-0.018244	0.016472	-1.107576	0.2692	0.091782
		ASVI_FIN	0.450251	0.160747	2.800984	0.0055	
		ASVI_GLO	0.226140	0.234343	0.964995	0.3355	
		ABRETURNS	-3.437501	1.096988	-3.133581	0.0019	
14	keskisuomalainen	C	0.239236	0.100293	2.385357	0.0179	0.027682

		ASVI_FIN	5.184984	3.266613	1.587266	0.1138	
		ASVI_GLO	-7.293762	3.444725	-2.117371	0.0353	
		ABRETURNS	6.897241	5.241048	1.316004	0.1895	
15	kesko	C	0.010504	0.013661	0.768892	0.4427	0.006310
		ASVI_FIN	0.126180	0.278044	0.453815	0.6504	
		ASVI_GLO	-0.208201	0.265146	-0.785231	0.4331	
		ABRETURNS	0.711735	0.875169	0.813255	0.4169	
16	kone	C	0.016035	0.010837	1.479633	0.1403	0.134683
		ASVI_FIN	0.870357	0.277238	3.139387	0.0019	
		ASVI_GLO	-2.019694	0.531953	-3.796753	0.0002	
		ABRETURNS	-3.153284	0.758561	-4.156927	0.0000	
17	lemminkainen	C	-0.014543	0.041628	-0.349361	0.7271	0.009415
		ASVI_FIN	0.778654	0.945068	0.823913	0.4108	
		ASVI_GLO	-0.595458	0.900901	-0.660958	0.5093	
		ABRETURNS	3.504889	2.807487	1.248408	0.2131	
18	martela	C	0.354358	0.083690	4.234187	0.0000	0.007474
		ASVI_FIN	1.272433	0.966483	1.316559	0.1893	
		ASVI_GLO	-1.604074	1.484040	-1.080883	0.2809	
		ABRETURNS	0.990600	5.256394	0.188456	0.8507	
19	metso	C	0.005940	0.014114	0.420835	0.6743	0.002014
		ASVI_FIN	0.085100	0.196183	0.433778	0.6648	
		ASVI_GLO	-0.206616	0.321782	-0.642099	0.5214	
		ABRETURNS	-0.070755	0.546979	-0.129356	0.8972	
20	neste oil	C	0.014356	0.013664	1.050686	0.2945	0.057664
		ASVI_FIN	0.328625	0.219287	1.498604	0.1353	
		ASVI_GLO	-0.030076	0.287873	-0.104475	0.9169	
		ABRETURNS	0.288898	0.570594	0.506312	0.6131	
21	nordea bank	C	0.006863	0.012001	0.571896	0.5679	0.016771
		ASVI_FIN	-0.076500	0.432301	-0.176959	0.8597	

		ASVI_GLO	0.501507	0.434710	1.153659	0.2498	
		ABRETURNS	-1.397947	0.969422	-1.442042	0.1506	
22	nurminen logistics	C	1.049403	0.197308	5.318602	0.0000	0.023860
		ASVI_FIN	3.798469	2.504241	1.516815	0.1307	
		ASVI_GLO	-3.556233	2.347133	-1.515139	0.1311	
		ABRETURNS	13.30221	8.319410	1.598937	0.1112	
23	olvi	C	0.017021	0.029070	0.585538	0.5587	0.024864
		ASVI_FIN	-0.330645	0.264215	-1.251424	0.2120	
		ASVI_GLO	0.357942	0.295338	1.211977	0.2267	
		ABRETURNS	4.967984	2.608520	1.904522	0.0581	
24	oriola	C	0.034203	0.025475	1.342591	0.1807	0.005940
		ASVI_FIN	-0.084904	0.245487	-0.345861	0.7298	
		ASVI_GLO	-0.313024	0.390232	-0.802150	0.4233	
		ABRETURNS	0.171102	1.718932	0.099540	0.9208	
25	outokumpu	C	0.029854	0.022361	1.335077	0.1831	0.019640
		ASVI_FIN	0.283635	0.421195	0.673406	0.5014	
		ASVI_GLO	-0.326750	0.460158	-0.710082	0.4784	
		ABRETURNS	-0.625354	0.302348	-2.068326	0.0397	
26	outotec	C	-0.004847	0.015304	-0.316689	0.7518	0.081047
		ASVI_FIN	0.275480	0.146596	1.879177	0.0615	
		ASVI_GLO	-0.275188	0.194648	-1.413769	0.1588	
		ABRETURNS	-2.677220	0.654307	-4.091688	0.0001	
27	ponsse	C	0.045546	0.033482	1.360314	0.1750	0.043597
		ASVI_FIN	0.225261	0.356440	0.631975	0.5280	
		ASVI_GLO	0.003736	0.537068	0.006956	0.9945	
		ABRETURNS	5.900366	1.887932	3.125306	0.0020	
28	ramirent	C	0.018316	0.020450	0.895683	0.3713	0.018970
		ASVI_FIN	-0.295991	0.281343	-1.052066	0.2938	
		ASVI_GLO	0.084778	0.383769	0.220908	0.8254	

		ABRETURNS	1.401979	1.076747	1.302050	0.1942	
29	rapala	C	0.072515	0.056406	1.285603	0.1998	0.006141
		ASVI_FIN	0.307181	0.457915	0.670827	0.5030	
		ASVI_GLO	-0.331039	2.110347	-0.156865	0.8755	
		ABRETURNS	5.310349	5.161695	1.028799	0.3046	
30	rautaruukki	C	0.001665	0.013110	0.126983	0.8991	0.232262
		ASVI_FIN	-0.043576	0.132674	-0.328442	0.7429	
		ASVI_GLO	-0.041951	0.126122	-0.332625	0.7397	
		ABRETURNS	3.830386	0.460842	8.311704	0.0000	
31	sampo	C	0.013084	0.011545	1.133254	0.2583	0.010266
		ASVI_FIN	0.209566	0.216579	0.967621	0.3342	
		ASVI_GLO	-0.173195	0.333616	-0.519146	0.6041	
		ABRETURNS	-1.203667	1.019790	-1.180309	0.2391	
32	SRV	C	0.001562	0.037528	0.041623	0.9668	0.064197
		ASVI_FIN	0.143816	0.075200	1.912457	0.0570	
		ASVI_GLO	0.370061	0.268276	1.379403	0.1691	
		ABRETURNS	5.236674	1.737124	3.014565	0.0029	
33	stockmann	C	0.033016	0.020198	1.634658	0.1035	0.077042
		ASVI_FIN	-0.951295	0.811530	-1.172225	0.2423	
		ASVI_GLO	1.091887	0.875770	1.246774	0.2137	
		ABRETURNS	4.769340	1.137114	4.194248	0.0000	
34	stora enso	C	0.010633	0.011306	0.940475	0.3479	0.018088
		ASVI_FIN	0.212751	0.136333	1.560526	0.1200	
		ASVI_GLO	-0.242664	0.171244	-1.417066	0.1578	
		ABRETURNS	0.840008	0.671831	1.250327	0.2124	
35	talvivaara	C	0.017076	0.021727	0.785922	0.4327	0.016959
		ASVI_FIN	-0.489496	0.344653	-1.420258	0.1569	
		ASVI_GLO	0.577244	0.337096	1.712401	0.0881	
		ABRETURNS	0.058469	0.198998	0.293818	0.7692	

36	technopolis	C	0.015154	0.014349	1.056084	0.2920	0.012931
		ASVI_FIN	0.232143	0.185386	1.252214	0.2117	
		ASVI_GLO	-0.037915	0.246707	-0.153685	0.8780	
		ABRETURNS	-1.370379	1.252402	-1.094200	0.2750	
37	teliasonera	C	0.015154	0.014349	1.056084	0.2920	0.012931
		ASVI_FIN	0.232143	0.185386	1.252214	0.2117	
		ASVI_GLO	-0.037915	0.246707	-0.153685	0.8780	
		ABRETURNS	-1.370379	1.252402	-1.094200	0.2750	
38	tikkurila	C	0.028451	0.028655	0.992871	0.3218	0.015050
		ASVI_FIN	-0.417975	1.004315	-0.416179	0.6777	
		ASVI_GLO	0.423979	1.052839	0.402701	0.6875	
		ABRETURNS	3.558497	1.971092	1.805343	0.0723	
39	tulikivi	C	0.083539	0.054005	1.546879	0.1232	0.015083
		ASVI_FIN	-0.730367	0.556170	-1.313208	0.1904	
		ASVI_GLO	0.246463	0.618347	0.398583	0.6906	
		ABRETURNS	2.675196	2.148932	1.244895	0.2144	
40	uponor	C	0.014915	0.023806	0.626527	0.5316	0.016104
		ASVI_FIN	-0.439869	0.290983	-1.511666	0.1320	
		ASVI_GLO	-0.055210	0.408699	-0.135088	0.8927	
		ABRETURNS	0.736383	1.391995	0.529013	0.5973	
41	viking line	C	0.650049	0.118930	5.465797	0.0000	0.002526
		ASVI_FIN	-1.359494	4.206646	-0.323178	0.7468	
		ASVI_GLO	2.498863	4.312746	0.579413	0.5629	
		ABRETURNS	-1.432167	7.320243	-0.195645	0.8451	
42	tikkurila	C	0.028451	0.028655	0.992871	0.3218	0.015050
		ASVI_FIN	-0.417975	1.004315	-0.416179	0.6777	
		ASVI_GLO	0.423979	1.052839	0.402701	0.6875	
		ABRETURNS	3.558497	1.971092	1.805343	0.0723	

APPENDIX 3 ATRIA

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.018865	0.018092	-1.042738	0.2971
SVI_FIN	0.696616	0.244146	2.853279	0.0043
SVI_GLO	-2.140435	0.515857	-4.149283	0.0000
AR(2)	-0.196540	0.001152	-170.5638	0.0000
MA(2)	0.354740	0.046813	7.577804	0.0000
Variance Equation				
C	0.002223	0.000813	2.733019	0.0063
RESID(-1)^2	0.175081	0.083293	2.101988	0.0356
RESID(-2)^2	-0.225736	0.084519	-2.670840	0.0076
GARCH(-1)	0.966116	0.157304	6.141721	0.0000
GARCH(-2)	0.062700	0.158853	0.394701	0.6931
R-squared	0.021359	Mean dependent var		-0.003071
Adjusted R-squared	0.004486	S.D. dependent var		0.310332

APPENDIX 4 FORTUM

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.028377	0.008284	-3.425567	0.0006
SVI_FIN	-0.273801	0.154363	-1.773749	0.0761
SVI_GLO	0.829859	0.223463	3.713623	0.0002
Variance Equation				
C	0.004288	0.000341	12.57874	0.0000
RESID(-1)^2	0.013934	0.012826	1.086323	0.2773
RESID(-2)^2	-0.009595	0.012141	-0.790304	0.4294
GARCH(-1)	1.847543	0.016781	110.0966	0.0000
GARCH(-2)	-0.998506	0.014926	-66.89840	0.0000
R-squared	0.011761	Mean dependent var		0.006270
Adjusted R-squared	0.003386	S.D. dependent var		0.175160

APPENDIX 5 KEMIRA

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.014482	0.017756	-0.815615	0.4147
SVI_FIN	0.340225	0.165106	2.060638	0.0393
SVI_GLO	0.144046	0.303848	0.474073	0.6354
Variance Equation				
C	0.016947	0.175031	0.096823	0.9229
RESID(-1)^2	0.112466	0.085015	1.322905	0.1859
RESID(-2)^2	-0.050870	0.609772	-0.083424	0.9335
GARCH(-1)	0.788571	5.654597	0.139457	0.8891
GARCH(-2)	-0.127730	2.192214	-0.058265	0.9535
R-squared	0.048635	Mean dependent var		-0.003443
Adjusted R-squared	0.040573	S.D. dependent var		0.259576

APPENDIX 6 KESKISUOMALAINEN

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.169215	0.068568	2.467843	0.0136
SVI_FIN	5.265797	3.405495	1.546265	0.1220
SVI_GLO	-8.105870	2.959078	-2.739323	0.0062
Variance Equation				
C	0.169762	0.026685	6.361714	0.0000
RESID(-1)^2	0.043760	0.007599	5.758726	0.0000
RESID(-2)^2	-0.047359	0.008030	-5.897755	0.0000
GARCH(-1)	-0.057844	0.015954	-3.625588	0.0003
GARCH(-2)	0.896390	0.012811	69.97156	0.0000

APPENDIX 7 KONE

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.011535	0.011854	0.973086	0.3305
SVI_FIN	0.808568	0.281566	2.871685	0.0041
SVI_GLO	-2.070548	0.542033	-3.819968	0.0001
Variance Equation				
C	0.020191	0.012172	1.658778	0.0972
RESID(-1)^2	0.230219	0.109887	2.095046	0.0362
RESID(-2)^2	0.162940	0.121006	1.346546	0.1781
GARCH(-1)	-0.544600	0.275543	-1.976458	0.0481
GARCH(-2)	0.435430	0.259939	1.675125	0.0939
R-squared	0.068630	Mean dependent var		0.009925
Adjusted R-squared	0.060737	S.D. dependent var		0.172653

APPENDIX 8 GRANGER SVI_GLO

Wald Test Atria: White's robust heteroscedastic adjusted standard errors.		
Null Hypothesis: SVI_GLO(-1)= SVI_GLO(-2)= SVI_GLO(-3)= SVI_GLO(-4)= SVI_GLO(-5)		
Null Hypothesis Summary:		
Normalized Restriction (= 0)	Value	Std. Err.
SVI_GLO(-1) - SVI_GLO(-5)	-0.000000	0.000000
SVI_GLO(-2) - SVI_GLO(-5)	-0.000000	0.000000
SVI_GLO(-3) - SVI_GLO(-5)	0.000000	0.000000
SVI_GLO(-4) - SVI_GLO(-5)	0.11E+93	0.000000

Wald Test Fortum, TV		
Null Hypothesis: SVI_GLO(-1)= SVI_GLO(-2)= SVI_GLO(-3)= SVI_GLO(-4)= SVI_GLO(-5)		
Normalized Restriction (= 0)	Value	Std. Err.
SVI_GLO(-1) - SVI_GLO(-5)	-0.000000	0.000000
SVI_GLO(-2) - SVI_GLO(-5)	-0.000000	0.000000
SVI_GLO(-3) - SVI_GLO(-5)	0.000000	0.000000
SVI_GLO(-4) - SVI_GLO(-5)	0.11E+93	0.000000

Wald Test: Keskisuomalainen, TV		
Null Hypothesis: SVI_GLO(-1)= SVI_GLO(-2)= SVI_GLO(-3)= SVI_GLO(-4)= SVI_GLO(-5)		
Normalized Restriction (= 0).		
SVI_GLO(-1) - SVI_GLO(-5)	-0.000000	0.000000
SVI_GLO(-2) - SVI_GLO(-5)	-0.000000	0.000000
SVI_GLO(-3) - SVI_GLO(-5)	0.000000	0.000000
SVI_GLO(-4) - SVI_GLO(-5)	1.12E+93	0.000000
Restrictions are linear in coefficients. F>15		

Wald Test: Kone, TV		
Null Hyp: SVI_GLO(-1)= SVI_GLO(-2)= SVI_GLO(-3)= SVI_GLO(-4)=VI_GLO(-5)		
Null Hypothesis Summary:	Value	Std. Err.
SVI_GLO(-1) - SVI_GLO(-5)	-0.000000	0.000000
SVI_GLO(-2) - SVI_GLO(-5)	-0.000000	0.000000
SVI_GLO(-3) - SVI_GLO(-5)	0.000000	0.000000
SVI_GLO(-4) - SVI_GLO(-5)	0.11E+93	0.000000
Restrictions are linear in coefficients. F>15		

APPENDIX 9 GRANGER SVI_FIN

Kemira	Dependent Variable: TV			
Variable	Coefficient	Std. Error	t-Statistic	Prob.
SVI_FIN	0.310798	0.264230	1.176240	0.2407
SVI_FIN(-1)	0.352955	0.337319	1.046354	0.2965
SVI_FIN(-2)	0.038682	0.332928	0.116188	0.9076
SVI_FIN(-3)	-0.182358	0.332928	-0.547740	0.5844
SVI_FIN(-4)	-0.158535	0.332928	-0.476185	0.6344
SVI_FIN(-5)	0.135993	0.258610	0.525862	0.5995
R-squared	0.056361	Mean dependent var		-0.003443
Adjusted R-squared	0.036111	S.D. dependent var		0.259576
S.E. of regression	0.254846	Akaike info criterion		0.128471
Sum squared resid	15.13255	Schwarz criterion		0.215746
Log likelihood	-9.352270	Hannan-Quinn criter.		0.163640
Durbin-Watson stat	1.171139			

Wald Test: Kemira, TV		null hyp. accepted F>15
Null Hypothesis: SVI_FIN(-1)= SVI_FIN(-2)= SVI_FIN(-3)= SVI_FIN(-4)= SVI_FIN(-5)		
Normalized Restriction (= 0)	Value	Std. Err.
SVI_FIN(-1) - SVI_FIN(-5)	-0.000000	0.000000
SVI_FIN(-2) - SVI_FIN(-5)	-0.000000	0.000000
SVI_FIN(-3) - SVI_FIN(-5)	0.000000	0.000000
SVI_FIN(-4) - SVI_FIN(-5)	0.000001	0.000000

Kone Oyj				Prob.
Variable	Coefficient	Std. Error	t-Statistic	0.4897
SVI_FIN(-1) - SVI_FIN(-5)	-0.028070	0.501017	-0.056026	0.3883
SVI_FIN(-2) - SVI_FIN(-5)	0.447699	0.647091	0.691864	0.7501
SVI_FIN(-3) - SVI_FIN(-5)	0.553731	0.640665	0.864305	0.6712
SVI_FIN(-4) - SVI_FIN(-5)	-0.204309	0.640665	-0.318902	0.6717
R-squared	0.025780	Mean dependent		0.009925
Adjusted R-squared	0.004873	S.D. dependent		0.172653
S.E. of regression	0.172232	Akaike info criterion		-0.655169
Sum squared resid	6.911650	Schwarz criterion		-0.567894
Log likelihood	84.29269	Hannan-Quinn criter.		-0.619999
Durbin-Watson stat	1.227864			

Null Hypothesis Summary:		
Wald Test:Kone TV	null hyp. accepted F>15	
Null Hypothesis: SVI_FIN(-1)= SVI_FIN(-2)= SVI_FIN(-3)= SVI_FIN(-4)= SVI_FIN(-5)		
Normalized Restriction (= 0)	Value	Std. Err.
SVI_FIN(-1) - SVI_FIN(-5)	-0.000000	0.000000
SVI_FIN(-2) - SVI_FIN(-5)	-0.000000	0.000000
SVI_FIN(-3) - SVI_FIN(-5)	0.000000	0.000000
SVI_FIN(-4) - SVI_FIN(-5)	.12E+92	0.000000

APPENDIX 10 OMXH ALL SHARE CONSTITUENTS FOR THIS STUDY

OMX Helsinki All Share Index approved Constituents for this study		
Company name		Ticker symbol
1	Aktia Bank PLC	AKTRV:HEX
2	Apetit Oyj	APETI:HEX
3	Atria Oyj	ATRAV:HEX
4	Basware Oyj	BAS1V:HEX
5	Cramo Oyj	CRA1V:HEX
6	Digia Oyj	DIG1V:HEX
7	Elisa Oyj	ELI1V:HEX
8	F-Secure Oyj	FSC1V:HEX
9	Finnair	FIA1S:HEX
10	Finnlines Oyj	FLG1S:HEX
11	Fiskars Oyj	FIS1V:HEX
12	Fortum	FUM1V:HEX
13	Honkarakenne Oyj	HONBS:HEX
14	Kemira Oyj	KRA1V:HEX
15	Keskisuomalainen Oyj	KSLAV:HEX
16	Kesko Oyj	KESAV:HEX
17	Kone Oyj	KNEBV:HEX
18	Konecranes Abp	KCR1V:HEX
19	Lemminkainen Oyj	LEM1S:HEX
20	M-real Oyj	MRLAV:HEX
21	Marimekko Oyj	MMO1V:HEX
22	Martela Oyj	MARAS:HEX
23	Metso Oyj	MEO1V:HEX
24	Neste Oil Corporation	NES1V:HEX
25	NOKIA	NOK1V:HEX

26	Nordea Bank AB	NDA1V:HEX
27	Nurminen Logistics Oyj	NLG1V:HEX
28	Olvi Oyj	OLVAS:HEX
29	Oriola KD Oyj	OKDAV:HEX
30	Outokumpu	OUT1V:HEX
31	Outotec Oyj	OTE1V:HEX
32	Ponsse Oyj	PON1V:HEX
33	Ramirent Oyj	RMR1V:HEX
34	Rapala VMC Corp	RAP1V:HEX
35	Rautaruukki Oyj	RTRKS:HEX
36	Sampo Oyj	SAMAS:HEX
37	SRV Yhtiot Oyj	SRV1V:HEX
38	Stockmann Oyj	STCAS:HEX
39	Stora Enso	STERV:HEX
40	Talvivaaran Kaivososakeyhtio Oyj	TLV1V:HEX
41	Technopolis Oyj	TPS1V:HEX
42	TeliaSonera	TLS1V:HEX
43	Tieto Oyj	TIE1V:HEX
44	Tikkurila Oyj	TIK1V:HEX
45	Tulikivi Oyj	TULAV:HEX
46	Uponor Oyj	UNR1V:HEX
47	Vacon Oyj	VAC1V:HEX
48	Vaisala Oyj	VAIAS:HEX
49	Valmet Corp	VALMT:HEX
50	Viking Line Abp	VIK1V:HEX
51	Yit Oyj	YTY1V:HEX